

The Bourgain–Gamburd Method
Volume I: Expansion, Random Walks, and
Finite-Group Harmonic Analysis

Preface

This volume separates the analytic and representation-theoretic foundations of expansion from the noncommutative additive combinatorics that drives flattening. The purpose of Volume I is to close the first dependency chain:

Cayley graphs $\leftrightarrow$ Markov operators $\leftrightarrow$ random walks

and

finite-group Fourier analysis $\implies$ quasirandom product mixing.

Every major theorem used later from these subjects is proved here at the level required by the Bourgain–Gamburd method. Historical citations identify the original sources; they do not replace proofs.

Foundational background used without repeated proof. The series treats as baseline only standard material from a first rigorous course in real analysis and linear algebra: Cauchy–Schwarz, Hölder and Jensen inequalities; the monotone, Fatou and dominated convergence theorems; Fubini–Tonelli; the Riesz representation theorem for positive functionals on $C(X)$ for compact metric X ; and Euclidean Fourier inversion/Plancherel in the classical $L^1 \cap L^2$ setting. Elementary finite-dimensional linear algebra, basic ring/field manipulations, and Zorn’s lemma are used in the usual way. Whenever the argument needs a substantially stronger theorem—for example noncommutative BSG, product growth, algebraic-geometry uniformity, random-matrix contraction, or a sieve theorem—that result is proved in the series before it is used essentially. Later volumes may cite this paragraph rather than silently enlarging the prerequisite list. **Series position.** This volume uses basic finite-group probability and linear algebra. Its main outputs are spectral-gap normalizations, finite-group Fourier analysis, and quasirandomness.

Quantifier and asymptotic conventions. Unless a statement explicitly says otherwise, every constant occurring in a quantitative theorem is chosen in the order in which it is quantified in that theorem. Thus a phrase such as “for all sufficiently small δ ” means: there is a displayed or locally quantified threshold $\delta_0 > 0$, depending only on the parameters listed before that phrase, such that the assertion holds for every $0 < \delta < \delta_0$. The phrases “sufficiently large N ” and “sufficiently large p ” have the analogous meaning.

If $X \ll_{\theta} Y$ or $X = O_{\theta}(Y)$, the implicit constant depends only on θ and on the fixed ambient data of the current volume; it is independent of the variables quantified after θ . The words *absolute*, *uniform*, and *bounded* never introduce an unlisted dependence. Whenever an inequality between exponents or scales is essential to the argument, that inequality is written explicitly in the proof; these conventions are not a substitute for such a calculation.

Contents

Preface	iii
Part 1. Expansion and Probability	1
Chapter 1. Expansion, Spectral Gap, and Random Walk	2
1. Why expansion is an operator-theoretic statement	2
2. Regular graphs and Cayley graphs	3
3. The averaging operator	4
4. The Dirichlet form	5
5. Spectral gap and the Poincaré inequality	6
6. Edge expansion	7
7. The Cheeger inequalities	8
7.1. The easy direction: a sparse cut gives a small spectral gap	8
7.2. A coarea inequality	8
7.3. The median argument	10
8. Expander families	12
9. Random walks and convolution	13
9.1. Two L^2 conventions	13
10. Ordinary versus absolute spectral gap	14
11. L^2 mixing from the spectrum	15
12. Total variation and mixing time	16
13. Examples: what expansion does and does not look like	17
13.1. Cycles are not expanders	17
13.2. The hypercube is highly connected but not an expander family	18
13.3. Complete graphs	18
14. Diameter consequences	19
15. Cayley graphs and translation symmetry	20
16. A first formulation of the expansion problem	20
17. Sanity checks on normalizations	22
18. Worked derivations	23
19. Research-facing questions	24
20. Exercises and mastery checks	25
21. Chapter summary	27

Chapter 2. Convolution, Concentration, and Probability Measures on Groups	28
1. Convolution conventions revisited	28
2. Counting norms and elementary inequalities	30
3. Collision probability and effective support	31
4. Heavy sets and level-set extraction	32
5. Entropy as a complementary measure of spreading	33
6. Convolution can only flatten in ℓ^2	33
7. Subgroups and cosets as model measures	34
8. Mass on cosets and non-concentration	35
9. Symmetrization	36
10. Return probabilities and even times	36
11. A quantitative notion of flattening	37
12. What the norm alone cannot tell us	38
13. A practical dyadic calculus	38
14. Mastery checks	39
15. Chapter summary	40
Chapter 3. Property (T) , Property (τ) , and Uniform Expansion	41
1. Almost invariant vectors	41
2. Property (T)	42
3. Finite quotients and the regular representation	43
4. Property (τ)	44
5. Why property (T) is stronger	44
6. Qualitative generation versus quantitative expansion	45
7. Normal versus nonnormal subgroups	45
8. A compactness characterization of property (T)	47
9. Tensor products and why finite-dimensional representations matter	47
10. Comparison of spectral gaps under bounded words	48
11. Expansion as a family property	49
12. From a Kazhdan constant to a spectral gap	49
13. Exercises and conceptual checks	50
14. Chapter summary	51
Part 2. Fourier Analysis and Quasirandomness	52
Chapter 4. Finite-Group Representation Theory and Fourier Analysis	53
1. Representations and invariant inner products	53
2. Complete reducibility	54
3. Intertwiners and Schur's lemma	55
4. Averaging an operator	56
5. Orthogonality of matrix coefficients	56
6. The regular representation	57

7. The nonabelian Fourier transform	58
8. Convolution becomes matrix multiplication	59
9. Convolution operators and block diagonalization	59
10. Trace and Hilbert–Schmidt identities	60
11. The trivial block and mean-zero functions	61
12. Abelian groups as a boundary case	61
13. Worked example: the cycle	61
14. Exercises and mastery checks	62
15. Chapter summary	62
Chapter 5. Quasirandom Groups and Product Mixing	63
1. Definition and first examples	63
2. The basic quasirandom convolution inequality	64
3. Product mixing for three sets	65
4. A mixing inequality for probability measures	66
5. Multiplicity amplification	67
6. Elementary quasirandomness of $SL_2(\mathbb{F}_p)$	68
7. Quasirandomness is not expansion	71
8. Worked thresholds: when does quasirandomness actually become useful?	71
9. Exercises and mastery checks	72
10. Chapter summary	73
Part 3. Technical Foundations	74
Appendix A. Finite-Dimensional Operator Theory for Expansion	75
1. Finite-dimensional Hilbert spaces	75
2. Rayleigh quotients and the spectral gap	76
3. Averaging operators are contractions	77
4. Projection onto invariant vectors	78
5. Powers and logarithmic time	78
6. Trace and Hilbert–Schmidt norm	79
7. Hilbert–Schmidt norm of convolution	80
8. Nonsymmetric measures and singular values	81
9. Reversible Markov chains	82
10. A compact operator checklist	82
11. Exercises	83
12. Appendix summary	83
Appendix B. Word Metrics, Free Groups, and Schreier Coverings	84
1. Words and word length	84
2. Cayley graphs realize the word metric	85
3. The free group and reduced words	85
4. The Cayley graph of a free group is a tree	86

5. Relations, girth, and injectivity radius	87
6. Local lifting of random walks	87
7. Graph coverings	88
8. Group actions and Schreier graphs	88
9. Stabilizers and orbit size	89
10. Ping-pong	89
11. A quantitative ping-pong viewpoint	90
12. Arithmetic height and short relations	91
13. Schreier return probabilities and subgroup intersections	92
14. A geometric checklist for initial spreading	92
15. Exercises	93
16. Appendix summary	93
Appendix C. Worked Finite-Group Laboratories	94
1. Cyclic groups: the boundary case	94
2. The cycle random walk	95
3. The group S_3	95
4. A central random walk on S_3	96
5. Subgroup-uniform measures	97
6. Direct products	97
7. Quotients and one-dimensional representations	98
8. The finite Heisenberg group	99
9. A product-mixing calculation	101
10. Quasirandomness and the trace endgame in numbers	102
11. Exercises	103
12. Appendix summary	103
Bibliography	104

Part 1

Expansion and Probability

CHAPTER 1

Expansion, Spectral Gap, and Random Walk

The Bourgain–Gamburd method is a mechanism for proving a uniform spectral gap for a family of finite Cayley graphs. Before any product theorem, sum–product estimate, or flattening argument can be used, one must be precise about what a spectral gap means, how it is measured, and why it is equivalent to rapid escape across every cut of the graph. This chapter develops those foundations from first principles. We deliberately keep track of normalizations, distinguish the ordinary from the absolute spectral gap, prove the Cheeger inequalities in the regular setting, and translate the operator language into random-walk language. These conventions will remain fixed throughout the book.

1. Why expansion is an operator-theoretic statement

The phrase *expander graph* suggests a geometric property: every set of vertices that is not too large has many edges leaving it. The Bourgain–Gamburd method, however, is designed to prove an analytic statement: the averaging operator of the graph contracts mean-zero functions by a definite amount. The first task is therefore to understand why the two statements are essentially equivalent.

A useful way to organize the subject is the triangle

$$\boxed{\text{edge expansion}} \longleftrightarrow \boxed{\text{spectral gap}} \longleftrightarrow \boxed{\text{random-walk mixing}}.$$

The first equivalence is quantified by the Cheeger inequalities. The second is elementary spectral theory, but it contains an important subtlety: for a non-lazy walk, the ordinary spectral gap does not by itself rule out an eigenvalue close to -1 . Thus one must distinguish the ordinary spectral gap from the *absolute* spectral gap.

The same distinction is easy to miss when one moves from graphs to probability measures on groups. We shall therefore set the conventions once and for all.

GUIDING IDEA 1.1. A Bourgain–Gamburd argument does not attempt to estimate every eigenvalue directly. It proves that convolution makes probability measures flatter, and then uses representation theory to convert that flattening into contraction of the averaging operator. The language of spectral gap is therefore the endpoint of the method, while random-walk measures are the objects manipulated during the proof.

2. Regular graphs and Cayley graphs

We begin with finite regular graphs. Allowing multiple edges is convenient because a generating multiset can repeat a generator. To keep the graph-theoretic degree convention completely unambiguous, graphs in the Cheeger discussion are loopless. Holding probabilities are introduced separately at the level of Markov operators.

DEFINITION 1.2 (Finite regular multigraph). A finite undirected d -regular multigraph $X = (V, E)$ consists of a finite vertex set V and possibly repeated undirected edges, but no loops, such that every vertex has degree d . Equivalently, we specify symmetric nonnegative integers $m(x, y) = m(y, x)$ satisfying

$$m(x, x) = 0, \quad \sum_{y \in V} m(x, y) = d \quad (x \in V).$$

Here $m(x, y)$ is the multiplicity of the edge joining x and y .

For most of the chapter one can simply think of a finite simple d -regular graph. The multiplicity notation is useful only to make formulas invariant under repeated generators.

DEFINITION 1.3 (Cayley multigraph). Let G be a finite group and let S be a finite multiset in $G \setminus \{e\}$ whose multiplicity function satisfies $m_S(s) = m_S(s^{-1})$ for every $s \in G$. The Cayley multigraph

$$\text{Cay}(G, S)$$

has vertex set G and, for every $x \in G$ and every occurrence of $s \in S$, an oriented step from x to xs . Symmetry of S makes the underlying multigraph undirected. Its degree is

$$d = |S|,$$

where multiplicity is counted.

REMARK 1.4. The graph $\text{Cay}(G, S)$ is connected if and only if the support of S generates G . Indeed, a path from e to g is exactly an expression of g as a word in the elements of S .

EXAMPLE 1.5 (The cycle). For $G = \mathbb{Z}/n\mathbb{Z}$ and $S = \{1, -1\}$, the graph $\text{Cay}(G, S)$ is the cycle C_n .

EXAMPLE 1.6 (The hypercube). For $G = (\mathbb{Z}/2\mathbb{Z})^m$ and $S = \{e_1, \dots, e_m\}$, where e_j is the j th coordinate vector, the Cayley graph is the m -dimensional hypercube. Since each e_j is its own inverse, S is symmetric.

EXAMPLE 1.7 (Complete graphs as Cayley graphs). Let G be any finite group and take $S = G \setminus \{e\}$. Then $\text{Cay}(G, S)$ is the complete graph on $|G|$ vertices. Its degree grows with $|G|$, so it is not the model relevant to bounded-degree expander theory, but it is a useful spectral check.

3. The averaging operator

Let $X = (V, E)$ be a finite d -regular graph and write

$$n = |V|.$$

We equip functions $f : V \rightarrow \mathbb{C}$ with the normalized inner product

$$(1) \quad \langle f, g \rangle_V = \frac{1}{n} \sum_{x \in V} f(x) \overline{g(x)}.$$

The corresponding norm is denoted by $\|f\|_{2,V}$. The normalization by $1/n$ is natural from the point of view of probability: it is the L^2 norm with respect to the uniform probability measure on V .

DEFINITION 1.8 (Averaging or Markov operator). The normalized adjacency operator of X is

$$(2) \quad (Pf)(x) = \frac{1}{d} \sum_{y \in V} m(x, y) f(y).$$

For a Cayley graph $X = \text{Cay}(G, S)$ this becomes

$$(3) \quad (P_S f)(x) = \frac{1}{|S|} \sum_{s \in S} f(xs).$$

The operator P preserves constants:

$$P\mathbf{1} = \mathbf{1}.$$

It is also self-adjoint, because the graph is undirected.

LEMMA 1.9 (Basic operator properties). *Let P be the averaging operator of a finite d -regular undirected multigraph. Then:*

- (i) P is self-adjoint on $L^2(V)$;
- (ii) $\|P\|_{2 \rightarrow 2} \leq 1$;
- (iii) every eigenvalue of P is real and belongs to $[-1, 1]$;
- (iv) 1 is an eigenvalue with eigenfunction $\mathbf{1}$;
- (v) the multiplicity of the eigenvalue 1 is the number of connected components of X .

PROOF. For self-adjointness, symmetry of $m(x, y)$ gives

$$\begin{aligned} \langle Pf, g \rangle_V &= \frac{1}{nd} \sum_{x, y \in V} m(x, y) f(y) \overline{g(x)} \\ &= \frac{1}{nd} \sum_{x, y \in V} m(x, y) f(x) \overline{g(y)} = \langle f, Pg \rangle_V. \end{aligned}$$

Next, Jensen's inequality yields

$$|Pf(x)|^2 \leq \frac{1}{d} \sum_y m(x, y) |f(y)|^2.$$

Averaging over x and using regularity gives

$$\|Pf\|_{2,V}^2 \leq \|f\|_{2,V}^2.$$

Thus $\|P\| \leq 1$. Since P is self-adjoint, all its eigenvalues are real and have absolute value at most 1. The equality $P\mathbf{1} = \mathbf{1}$ is immediate.

Finally, suppose $Pf = f$. Since P has real coefficients, the real and imaginary parts of f are also fixed, so it suffices to consider a real-valued fixed function. On any connected component choose a vertex x where f is maximal. The identity

$$f(x) = Pf(x) = \frac{1}{d} \sum_y m(x, y) f(y)$$

expresses this maximum as the average of neighboring values, all of which are at most $f(x)$. Hence every neighbor of x has the same value. Propagating along paths shows that f is constant on the component. Conversely, every function constant on connected components is visibly fixed by P . $\square$

For a connected graph we therefore order the eigenvalues as

$$(4) \quad 1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n \geq -1.$$

The strict inequality $\lambda_1 > \lambda_2$ follows from connectedness.

4. The Dirichlet form

The operator $I - P$ measures oscillation. Its quadratic form has a concrete edge interpretation.

DEFINITION 1.10 (Dirichlet form). For functions $f, g : V \rightarrow \mathbb{C}$, define

$$\mathcal{E}(f, g) = \langle (I - P)f, g \rangle_V.$$

Write $\mathcal{E}(f) = \mathcal{E}(f, f)$.

PROPOSITION 1.11 (Dirichlet-form identity). *For every $f : V \rightarrow \mathbb{C}$,*

$$(5) \quad \mathcal{E}(f) = \frac{1}{2nd} \sum_{x,y \in V} m(x, y) |f(x) - f(y)|^2.$$

For a Cayley graph this can be written as

$$(6) \quad \mathcal{E}(f) = \frac{1}{2|G||S|} \sum_{x \in G} \sum_{s \in S} |f(x) - f(xs)|^2.$$

PROOF. Expanding the square in Equation (5) gives

$$\begin{aligned} & \frac{1}{2nd} \sum_{x,y} m(x,y) (|f(x)|^2 + |f(y)|^2 - 2\Re(f(x)\overline{f(y)})) \\ &= \frac{1}{n} \sum_x |f(x)|^2 - \Re \left(\frac{1}{nd} \sum_{x,y} m(x,y) f(y)\overline{f(x)} \right). \end{aligned}$$

The second term is $\Re \langle Pf, f \rangle_V$. Since P is self-adjoint, $\langle Pf, f \rangle_V$ is real, so the expression equals

$$\langle f, f \rangle_V - \langle Pf, f \rangle_V = \mathcal{E}(f).$$

□

The identity is the basic bridge between the geometry of edges and the spectrum of P .

DEFINITION 1.12 (Mean and variance). For $f : V \rightarrow \mathbb{C}$, set

$$\mathbb{E}_V f = \frac{1}{n} \sum_{x \in V} f(x), \quad \text{Var}_V(f) = \frac{1}{n} \sum_{x \in V} |f(x) - \mathbb{E}_V f|^2.$$

If $f \perp \mathbf{1}$, then $\text{Var}_V(f) = \|f\|_{2,V}^2$.

5. Spectral gap and the Poincaré inequality

DEFINITION 1.13 (Ordinary spectral gap). Let X be connected and let its normalized adjacency eigenvalues be ordered as in Equation (4). The *spectral gap* is

$$(7) \quad \gamma(X) = 1 - \lambda_2.$$

This is the convention most natural for expansion. The quantity can lie in $(0, 2]$; it need not be at most 1 for a non-lazy graph.

PROPOSITION 1.14 (Variational characterization). For a connected regular graph,

$$(8) \quad \gamma(X) = \inf_{\substack{f \neq 0 \\ f \perp \mathbf{1}}} \frac{\mathcal{E}(f)}{\|f\|_{2,V}^2} = \inf_{f \text{ nonconstant}} \frac{\mathcal{E}(f)}{\text{Var}_V(f)}.$$

Equivalently, $\gamma(X)$ is the largest constant γ for which the Poincaré inequality

$$(9) \quad \mathcal{E}(f) \geq \gamma \text{Var}_V(f)$$

holds for every f .

PROOF. Choose an orthonormal eigenbasis $\phi_1, \dots, \phi_n$ for P , with $\phi_1 = \mathbf{1}$ and $P\phi_j = \lambda_j\phi_j$. If $f \perp \mathbf{1}$ and

$$f = \sum_{j=2}^n a_j \phi_j,$$

then

$$\mathcal{E}(f) = \sum_{j=2}^n (1 - \lambda_j) |a_j|^2 \geq (1 - \lambda_2) \sum_{j=2}^n |a_j|^2.$$

Equality is attained by an eigenfunction for λ_2 . This proves the first identity. Subtracting the mean from a general function does not change its Dirichlet energy and converts its squared L^2 norm into the variance, giving the second identity. $\square$

REMARK 1.15. The Poincaré inequality is often the cleanest formulation of a spectral gap. It says that a function cannot have substantial global variance while changing very little along edges. An expander is therefore a graph on which every nonconstant function is forced to oscillate significantly across the edge set.

6. Edge expansion

For $A \subseteq V$, let

$$b(A) = \sum_{x \in A} \sum_{y \notin A} m(x, y).$$

Thus $b(A)$ counts the oriented incidences leaving A . In a simple undirected graph this equals the usual number of undirected edges with one endpoint in A and one in A^c : each crossing edge has exactly one endpoint in A .

DEFINITION 1.16 (Normalized edge expansion). For a finite d -regular graph X , define

$$(10) \quad h(X) = \min_{\substack{A \subseteq V \\ 0 < |A| \leq n/2}} \frac{b(A)}{d|A|}.$$

The normalization makes $0 \leq h(X) \leq 1$. The denominator $d|A|$ is the number of possible one-step moves starting from points of A , counted with multiplicity. Hence

$$\frac{b(A)}{d|A|}$$

is exactly the probability that a one-step random walk exits A when its starting point is chosen uniformly from A .

LEMMA 1.17 (Connectedness and expansion). *A finite regular graph is connected if and only if $h(X) > 0$.*

PROOF. If X is disconnected, take a connected component of size at most $n/2$; its boundary is empty. Conversely, if $h(X) = 0$, some nonempty A of size at most $n/2$ has no edge to its complement, so the graph is disconnected. $\square$

7. The Cheeger inequalities

The following theorem is the fundamental comparison between geometry and spectrum.

THEOREM 1.18 (Cheeger inequalities for regular graphs). *Let X be a finite connected d -regular undirected multigraph. Then*

$$(11) \quad \frac{h(X)^2}{2} \leq \gamma(X) \leq 2h(X).$$

The upper inequality is easy: a sparse cut produces a test function with small Rayleigh quotient. The lower inequality is subtler: a function with small Dirichlet energy must have a level set with small edge boundary. We prove both directions carefully.

7.1. The easy direction: a sparse cut gives a small spectral gap.

PROOF OF $\gamma(X) \leq 2h(X)$. Choose a nonempty set $A \subseteq V$ with $|A| \leq n/2$ attaining $h(X)$. Let

$$f = \mathbf{1}_A.$$

Subtracting a constant does not change the Dirichlet energy, so by the variational characterization,

$$\gamma(X) \leq \frac{\mathcal{E}(\mathbf{1}_A)}{\text{Var}_V(\mathbf{1}_A)}.$$

The Dirichlet-form identity gives

$$(12) \quad \mathcal{E}(\mathbf{1}_A) = \frac{b(A)}{nd}.$$

Indeed, every crossing edge contributes twice to the oriented sum in Equation (5), and the factor $1/2$ removes this double counting. Also

$$\text{Var}_V(\mathbf{1}_A) = \frac{|A|}{n} \left(1 - \frac{|A|}{n} \right) \geq \frac{|A|}{2n}.$$

Therefore

$$\gamma(X) \leq \frac{b(A)/(nd)}{|A|/(2n)} = 2 \frac{b(A)}{d|A|} = 2h(X).$$

□

7.2. A coarea inequality. For the lower Cheeger inequality we first prove a discrete coarea estimate.

LEMMA 1.19 (Coarea estimate). *Let $u : V \rightarrow [0, \infty)$ and suppose*

$$|\text{supp } u| \leq n/2.$$

Then

$$(13) \quad \sum_{\{x,y\} \in E} |u(x)^2 - u(y)^2| \geq h(X)d \sum_{x \in V} u(x)^2,$$

where edges are counted with multiplicity, each undirected edge once.

PROOF. For $t \geq 0$, define the superlevel set

$$A_t = \{x \in V : u(x) > t\}.$$

Since $A_t \subseteq \text{supp } u$, we have $|A_t| \leq n/2$. Hence, whenever $A_t \neq \emptyset$,

$$b(A_t) \geq h(X)d|A_t|.$$

For nonnegative real numbers a, b ,

$$|a^2 - b^2| = \int_0^\infty 2t |\mathbf{1}_{\{a>t\}} - \mathbf{1}_{\{b>t\}}| dt.$$

Summing over undirected edges and interchanging the finite sum with the integral gives

$$\begin{aligned} \sum_{\{x,y\} \in E} |u(x)^2 - u(y)^2| &= \int_0^\infty 2t b(A_t) dt \\ &\geq h(X)d \int_0^\infty 2t |A_t| dt. \end{aligned}$$

Finally,

$$\int_0^\infty 2t |A_t| dt = \sum_{x \in V} \int_0^{u(x)} 2t dt = \sum_{x \in V} u(x)^2.$$

This proves the claim. $\square$

LEMMA 1.20 (Half-support Poincaré estimate). *If $u : V \rightarrow [0, \infty)$ and $|\text{supp } u| \leq n/2$, then*

$$(14) \quad \mathcal{E}(u) \geq \frac{h(X)^2}{2} \|u\|_{2,V}^2.$$

PROOF. Write the sums below over undirected edges, counted with multiplicity. By Lemma 1.19,

$$(15) \quad hd \sum_x u(x)^2 \leq \sum_{\{x,y\} \in E} |u(x)^2 - u(y)^2|.$$

Factor

$$|u(x)^2 - u(y)^2| = |u(x) - u(y)| |u(x) + u(y)|$$

and apply Cauchy–Schwarz:

$$\left(\sum_{\{x,y\} \in E} |u(x)^2 - u(y)^2| \right)^2 \leq \left(\sum_{\{x,y\} \in E} |u(x) - u(y)|^2 \right) \left(\sum_{\{x,y\} \in E} |u(x) + u(y)|^2 \right).$$

Using $(a + b)^2 \leq 2a^2 + 2b^2$ and d -regularity,

$$\sum_{\{x,y\} \in E} |u(x) + u(y)|^2 \leq 2d \sum_x u(x)^2.$$

Combining this with Equation (15) yields

$$h^2 d^2 \left(\sum_x u(x)^2 \right)^2 \leq 2d \left(\sum_x u(x)^2 \right) \left(\sum_{\{x,y\} \in E} |u(x) - u(y)|^2 \right).$$

If $u \neq 0$, cancel $d \sum_x u(x)^2$ to obtain

$$\sum_{\{x,y\} \in E} |u(x) - u(y)|^2 \geq \frac{h^2 d}{2} \sum_x u(x)^2.$$

Since

$$\mathcal{E}(u) = \frac{1}{nd} \sum_{\{x,y\} \in E} |u(x) - u(y)|^2,$$

we obtain

$$\mathcal{E}(u) \geq \frac{h^2}{2n} \sum_x u(x)^2 = \frac{h^2}{2} \|u\|_{2,V}^2.$$

The zero function is trivial. □

7.3. The median argument. We now pass from half-supported nonnegative functions to arbitrary mean-zero functions. The use of a median is essential because a mean-zero function need not have either its positive or its negative support of size at most $n/2$ after a fixed threshold at zero.

DEFINITION 1.21 (Median). A real number m is a median of a real-valued function $f : V \rightarrow \mathbb{R}$ if

$$|\{x : f(x) > m\}| \leq n/2, \quad |\{x : f(x) < m\}| \leq n/2.$$

Every function on a finite set has a median.

For $g = f - m$, define its positive and negative parts by

$$g_+(x) = \max(g(x), 0), \quad g_-(x) = \max(-g(x), 0).$$

Then

$$g = g_+ - g_-, \quad |g| = g_+ + g_-, \quad g_+ g_- = 0$$

pointwise, and each support has size at most $n/2$.

LEMMA 1.22 (Energy decreases under positive/negative splitting). *For every real-valued function g ,*

$$(16) \quad \mathcal{E}(g_+) + \mathcal{E}(g_-) \leq \mathcal{E}(g).$$

PROOF. It suffices to check the pointwise inequality

$$|a_+ - b_+|^2 + |a_- - b_-|^2 \leq |a - b|^2$$

for real a, b . If a and b have the same sign, equality holds. If, say, $a \geq 0 \geq b$, the left side is $a^2 + b^2$, while the right side is $(a - b)^2 = a^2 + b^2 - 2ab \geq a^2 + b^2$. Summing over edges and using the Dirichlet identity proves the result. $\square$

PROOF OF $h(X)^2/2 \leq \gamma(X)$. By the variational characterization it is enough to consider a real-valued function $f \perp \mathbf{1}$; indeed the minimum Rayleigh quotient is attained by a real eigenfunction for λ_2 . Let m be a median of f and put

$$g = f - m.$$

Both g_+ and g_- have support of size at most $n/2$. Therefore Lemma 1.20 gives

$$\begin{aligned} \mathcal{E}(g_+) + \mathcal{E}(g_-) &\geq \frac{h^2}{2} (\|g_+\|_{2,V}^2 + \|g_-\|_{2,V}^2) \\ &= \frac{h^2}{2} \|g\|_{2,V}^2. \end{aligned}$$

On the other hand, Lemma 1.22 and invariance of Dirichlet energy under addition of constants imply

$$\mathcal{E}(g_+) + \mathcal{E}(g_-) \leq \mathcal{E}(g) = \mathcal{E}(f).$$

Because f has mean zero,

$$\begin{aligned} \|g\|_{2,V}^2 &= \frac{1}{n} \sum_x |f(x) - m|^2 \\ &= \frac{1}{n} \sum_x |f(x)|^2 + m^2 \geq \|f\|_{2,V}^2. \end{aligned}$$

Consequently

$$\mathcal{E}(f) \geq \frac{h^2}{2} \|f\|_{2,V}^2.$$

Taking the infimum over $f \perp \mathbf{1}$ proves

$$\gamma(X) \geq \frac{h(X)^2}{2}.$$

$\square$

WARNING 1.23 (A common loss in the median proof). It is tempting to keep only whichever of g_+ or g_- has larger L^2 norm. Doing so introduces an unnecessary additional factor of 2. The clean argument uses *both* parts and the inequality

$$\mathcal{E}(g_+) + \mathcal{E}(g_-) \leq \mathcal{E}(g).$$

The shift by the median does not harm the denominator because a mean-zero function minimizes its L^2 distance from constants at the constant 0.

8. Expander families

DEFINITION 1.24 (Expander family). A sequence (X_i) of finite connected d_i -regular graphs is an *edge-expander family* if

$$|V(X_i)| \longrightarrow \infty$$

and there is $h_0 > 0$ such that

$$h(X_i) \geq h_0$$

for every i . It is a *spectral expander family* if there is $\gamma_0 > 0$ such that

$$\gamma(X_i) \geq \gamma_0$$

for every i .

By the Cheeger inequalities these two definitions are equivalent, with quantitative conversion of constants.

Usually one also assumes bounded degree:

$$\sup_i d_i < \infty.$$

Without this requirement, complete graphs give a trivial family with excellent expansion. The remarkable feature of expander graphs is that a uniformly bounded number of neighbors per vertex can nevertheless force global connectivity properties comparable to those of dense graphs.

COROLLARY 1.25 (Equivalence of expansion notions). *For a family of finite regular graphs of arbitrary degrees, the following are equivalent:*

- (i) $\inf_i h(X_i) > 0$;
- (ii) $\inf_i \gamma(X_i) > 0$.

More precisely,

$$\frac{h(X_i)^2}{2} \leq \gamma(X_i) \leq 2h(X_i)$$

for every i .

PROOF. This is an immediate quantitative consequence of the two Cheeger inequalities proved above: for every finite regular graph,

$$\frac{h(X)^2}{2} \leq \gamma(X) \leq 2h(X).$$

Thus a uniform positive lower bound for either quantity gives one for the other, with the displayed constants. $\square$

9. Random walks and convolution

We now translate the operator language into the probability-measure language used throughout the Bourgain–Gamburd method.

Let G be a finite group. A probability measure on G is a function

$$\mu : G \rightarrow [0, 1]$$

with $\sum_g \mu(g) = 1$. For two functions $\nu, \mu : G \rightarrow \mathbb{C}$, define convolution by

$$(17) \quad (\nu * \mu)(x) = \sum_{y \in G} \nu(y) \mu(y^{-1}x).$$

If Y and Z are independent G -valued random variables with laws ν and μ , then YZ has law $\nu * \mu$.

Write

$$\mu^{(k)} = \underbrace{\mu * \cdots * \mu}_{k \text{ factors}}.$$

If $X_1, X_2, \dots$ are independent with law μ , then

$$X_1 X_2 \cdots X_k$$

has law $\mu^{(k)}$.

For a symmetric multiset S , let

$$(18) \quad \mu_S = \frac{1}{|S|} \sum_{s \in S} \delta_s,$$

where repetitions in S are retained. Then

$$P_S f(x) = \sum_{s \in G} \mu_S(s) f(xs).$$

Starting from e , the position after k steps has law $\mu_S^{(k)}$.

9.1. Two L^2 conventions. A normalization issue deserves to be isolated now.

CONVENTION 1.26 (Functions versus measures). For functions regarded as elements of L^2 of the uniform probability space G , we write

$$\|f\|_{2,G}^2 = \frac{1}{|G|} \sum_{x \in G} |f(x)|^2.$$

For probability measures and convolution estimates, we use the counting ℓ^2 norm

$$\|\nu\|_2^2 = \sum_{x \in G} |\nu(x)|^2.$$

The subscript 2, G will therefore indicate normalized L^2 , while the unsubscripted $\|\cdot\|_2$ applied to a measure means counting ℓ^2 .

Let u_G denote the uniform probability measure:

$$u_G(x) = \frac{1}{|G|}.$$

The density of a probability measure ν relative to u_G is

$$F_\nu(x) = |G|\nu(x).$$

Then

$$(19) \quad \|F_\nu - 1\|_{2,G}^2 = |G| \|\nu - u_G\|_2^2.$$

This elementary identity prevents a recurring source of factors of $|G|$.

REMARK 1.27. The uniform measure has

$$\|u_G\|_2 = |G|^{-1/2},$$

which is the smallest possible ℓ^2 norm of a probability measure on G . Thus the quantity $\|\nu\|_2$ is a direct measure of concentration. A Dirac mass has norm 1; the uniform measure has norm $|G|^{-1/2}$. Flattening arguments seek to drive the random-walk measure from the former scale toward the latter.

10. Ordinary versus absolute spectral gap

The spectral gap $1 - \lambda_2$ controls expansion, but the convergence of P^k to the projection onto constants depends on

$$\max_{j \geq 2} |\lambda_j|.$$

This distinction matters whenever the graph is close to bipartite.

DEFINITION 1.28 (Absolute spectral gap). For a connected regular graph with eigenvalues as in Equation (4), define

$$(20) \quad \gamma_{\text{abs}}(X) = 1 - \max_{2 \leq j \leq n} |\lambda_j|.$$

We can have $\gamma(X) > 0$ but $\gamma_{\text{abs}}(X) = 0$.

EXAMPLE 1.29 (Bipartite obstruction). Let X be a connected bipartite regular graph with parts V_+ and V_- . The function

$$f = \mathbf{1}_{V_+} - \mathbf{1}_{V_-}$$

satisfies $Pf = -f$. Thus -1 is an eigenvalue. The ordinary gap $1 - \lambda_2$ can be positive, but the walk alternates between the two parts and cannot converge to the uniform distribution at every time.

The standard cure is laziness.

DEFINITION 1.30 (Lazy walk). Given P , define

$$(21) \quad P_L = \frac{I + P}{2}.$$

At each step, the lazy walk stays put with probability $1/2$ and otherwise makes one step according to P .

PROPOSITION 1.31 (Spectrum of the lazy walk). *If $\lambda_1, \dots, \lambda_n$ are the eigenvalues of P , then those of P_L are*

$$\frac{1 + \lambda_1}{2}, \dots, \frac{1 + \lambda_n}{2}.$$

In particular they all lie in $[0, 1]$, and

$$(22) \quad \gamma_{\text{abs}}(P_L) = \gamma(P) = \frac{\gamma(P)}{2}.$$

PROOF. Every eigenvector of P is an eigenvector of $(I + P)/2$ with the displayed eigenvalue. Since $\lambda_j \in [-1, 1]$, all lazy eigenvalues lie in $[0, 1]$. The second largest is $(1 + \lambda_2)/2$, and the conclusion follows. $\square$

11. L^2 mixing from the spectrum

Let Π denote orthogonal projection onto the constants:

$$\Pi f = (\mathbb{E}_V f) \mathbf{1}.$$

Since P fixes constants and preserves their orthogonal complement,

$$P^k - \Pi = (P - \Pi)^k.$$

PROPOSITION 1.32 (L^2 contraction). *For every f with $f \perp \mathbf{1}$,*

$$(23) \quad \|P^k f\|_{2,V} \leq (1 - \gamma_{\text{abs}})^k \|f\|_{2,V}.$$

If P is lazy, this becomes

$$(24) \quad \|P^k f\|_{2,V} \leq (1 - \gamma)^k \|f\|_{2,V}.$$

PROOF. Expand f in an orthonormal eigenbasis of the mean-zero subspace. Every eigenvalue there has absolute value at most $1 - \gamma_{\text{abs}}$, so the result follows from Parseval's identity. For a lazy walk all eigenvalues are nonnegative, hence the ordinary and absolute gaps agree. $\square$

For a Cayley graph we can apply this to the density of the random-walk distribution.

PROPOSITION 1.33 (Random-walk L^2 estimate). *Let G be finite, let μ be a symmetric probability measure whose support generates G , and let P_μ be right averaging by μ . Let*

$$\rho = \max_{j \geq 2} |\lambda_j(P_\mu)|.$$

Then for every $k \geq 0$,

$$(25) \quad \left\| \mu^{(k)} - u_G \right\|_2 \leq \rho^k \sqrt{1 - \frac{1}{|G|}}.$$

Equivalently,

$$(26) \quad \left\| |G| \mu^{(k)} - 1 \right\|_{2,G} \leq \rho^k \sqrt{|G| - 1}.$$

PROOF. Consider the initial density $F_0 = |G| \delta_e$. Its mean with respect to the uniform probability measure is 1. Because μ is symmetric, the Markov kernel is self-adjoint with respect to uniform measure; hence the forward evolution of densities is governed by the same operator that acts on test functions. Therefore

$$F_k = |G| \mu^{(k)} = P_\mu^k F_0.$$

Consequently

$$F_k - 1 = P_\mu^k (F_0 - 1).$$

The vector $F_0 - 1$ is mean zero, and

$$\|F_0 - 1\|_{2,G}^2 = |G| - 1.$$

Apply [Proposition 1.32](#) and then use [Equation \(19\)](#). □

REMARK 1.34 (Why the prefactor is large). The estimate starts from a point mass, whose density relative to uniform has L^2 norm of order $|G|^{1/2}$. Thus a fixed spectral gap requires on the order of $\log |G|$ steps before the right side becomes small. This logarithmic mixing scale is exactly what one expects in a bounded-degree expander: after k steps the walk can reach at most exponentially many words, so reaching a positive fraction of a group of size $|G|$ cannot happen in substantially fewer than $c \log |G|$ steps.

12. Total variation and mixing time

For probability measures ν, η on G , define

$$\|\nu - \eta\|_{\text{TV}} = \frac{1}{2} \sum_{x \in G} |\nu(x) - \eta(x)|.$$

Cauchy–Schwarz gives

$$(27) \quad \|\nu - u_G\|_{\text{TV}} \leq \frac{1}{2} |G|^{1/2} \|\nu - u_G\|_2.$$

Combining this with [Proposition 1.33](#) yields a standard spectral mixing bound.

COROLLARY 1.35 (Spectral bound for total variation). *Under the hypotheses of [Proposition 1.33](#),*

$$(28) \quad \left\| \mu^{(k)} - u_G \right\|_{\text{TV}} \leq \frac{1}{2} |G|^{1/2} \rho^k.$$

If the walk is lazy with ordinary spectral gap γ , then $\rho = 1 - \gamma$ and

$$(29) \quad \left\| \mu^{(k)} - u_G \right\|_{\text{TV}} \leq \frac{1}{2} |G|^{1/2} (1 - \gamma)^k \leq \frac{1}{2} |G|^{1/2} e^{-\gamma k}.$$

PROOF. Apply Equation (27) with $\nu = \mu^{(k)}$ and then Proposition 1.33:

$$\left\| \mu^{(k)} - u_G \right\|_{\text{TV}} \leq \frac{1}{2} |G|^{1/2} \left\| \mu^{(k)} - u_G \right\|_2 \leq \frac{1}{2} |G|^{1/2} \rho^k.$$

For a lazy reversible walk, all nontrivial eigenvalues lie in $[0, 1 - \gamma]$, so $\rho = 1 - \gamma$. The final bound follows from $1 - x \leq e^{-x}$ for $x \geq 0$. $\square$

DEFINITION 1.36 (Mixing time). For $0 < \varepsilon < 1$, define

$$t_{\text{mix}}(\varepsilon) = \min \left\{ k : \sup_{x \in G} \left\| \delta_x * \mu^{(k)} - u_G \right\|_{\text{TV}} \leq \varepsilon \right\}.$$

For a Cayley graph the supremum over starting points is redundant by translation invariance.

COROLLARY 1.37 (Logarithmic mixing on a spectral expander). Suppose a family of finite Cayley graphs carries lazy walks with a common spectral gap $\gamma_0 > 0$. Then for every fixed $0 < \varepsilon < 1$,

$$t_{\text{mix}}(\varepsilon) \leq \frac{1}{\gamma_0} \left(\frac{1}{2} \log |G| + \log \frac{1}{2\varepsilon} \right) + 1.$$

In particular,

$$t_{\text{mix}}(\varepsilon) = O_{\gamma_0, \varepsilon}(\log |G|).$$

PROOF. It suffices by Equation (29) to require

$$\frac{1}{2} |G|^{1/2} e^{-\gamma_0 k} \leq \varepsilon.$$

Solving for k gives the result, with the final +1 accounting for integrality. $\square$

13. Examples: what expansion does and does not look like

Concrete spectra are useful for calibrating the definitions.

13.1. Cycles are not expanders. Let $X = C_n = \text{Cay}(\mathbb{Z}/n\mathbb{Z}, \{\pm 1\})$. For

$$\chi_j(x) = e^{2\pi i j x / n}, \quad 0 \leq j < n,$$

we have

$$P\chi_j = \frac{1}{2}(\chi_j(x+1) + \chi_j(x-1)) = \cos\left(\frac{2\pi j}{n}\right) \chi_j.$$

Thus

$$\lambda_2 = \cos\left(\frac{2\pi}{n}\right),$$

so

$$(30) \quad \gamma(C_n) = 1 - \cos\left(\frac{2\pi}{n}\right) \sim \frac{2\pi^2}{n^2}.$$

The spectral gap tends to zero quadratically.

The edge expansion also tends to zero. If A is an interval of $\lfloor n/2 \rfloor$ consecutive vertices, then $b(A) = 2$ and $d = 2$, so

$$h(C_n) \leq \frac{1}{\lfloor n/2 \rfloor} \asymp \frac{1}{n}.$$

This illustrates an important feature of Cheeger's inequalities: they are uniform equivalences for the positivity of a gap, but the lower bound $\gamma \geq h^2/2$ can have the correct order while the upper bound $\gamma \leq 2h$ is much weaker.

13.2. The hypercube is highly connected but not an expander family. Let $Q_m = \text{Cay}((\mathbb{Z}/2\mathbb{Z})^m, \{e_1, \dots, e_m\})$. The characters are indexed by $a \in (\mathbb{Z}/2\mathbb{Z})^m$:

$$\chi_a(x) = (-1)^{a \cdot x}.$$

If $|a|$ denotes Hamming weight, then

$$P\chi_a = \left(1 - \frac{2|a|}{m}\right)\chi_a.$$

Hence

$$\gamma(Q_m) = \frac{2}{m} \longrightarrow 0.$$

There is also a visible sparse cut: take

$$A = \{x : x_1 = 0\}.$$

Exactly one of the m possible moves exits A , so

$$h(Q_m) \leq \frac{1}{m}.$$

The hypercube has many paths and strong combinatorial symmetry, but its degree grows and its normalized expansion still vanishes.

13.3. Complete graphs. For the complete graph K_n , regarded as $\text{Cay}(G, G \setminus \{e\})$ for any group of order n ,

$$(Pf)(x) = \frac{1}{n-1} \sum_{y \neq x} f(y).$$

If $f \perp \mathbf{1}$, then $\sum_y f(y) = 0$, and therefore

$$Pf(x) = -\frac{1}{n-1} f(x).$$

Thus the nontrivial spectrum consists entirely of $-1/(n-1)$, and

$$\gamma(K_n) = 1 + \frac{1}{n-1} = \frac{n}{n-1}.$$

For $|A| \leq n/2$,

$$\frac{b(A)}{d|A|} = \frac{n-|A|}{n-1},$$

so

$$h(K_n) = \frac{n - \lfloor n/2 \rfloor}{n-1},$$

which is bounded below by $1/2$. The graphs have excellent expansion, but the degree is $n-1$.

14. Diameter consequences

Uniform expansion forces logarithmic diameter. This is another geometric manifestation of the same phenomenon.

PROPOSITION 1.38 (Expansion implies logarithmic diameter). *Let X be a finite connected d -regular graph with $h(X) \geq h_0 > 0$. Then*

$$(31) \quad \text{diam}(X) \leq 2 \left\lceil \frac{\log(n/2)}{\log(1+h_0)} \right\rceil + 2.$$

In particular, for fixed h_0 the diameter is $O(\log n)$.

PROOF. Fix a vertex x and let $B_r(x)$ be the ball of radius r . If $|B_r(x)| \leq n/2$, then at least $h_0 d |B_r(x)|$ edge incidences leave $B_r(x)$. Each vertex outside $B_r(x)$ is incident to at most d such edges, so the number of new vertices reached at the next step is at least $h_0 |B_r(x)|$. Hence

$$|B_{r+1}(x)| \geq (1+h_0)|B_r(x)|$$

until the ball has size greater than $n/2$. Therefore every ball has size greater than $n/2$ once

$$r \geq \left\lceil \frac{\log(n/2)}{\log(1+h_0)} \right\rceil + 1.$$

Two subsets of V of size greater than $n/2$ intersect. Thus balls of this radius about any two vertices intersect, giving the displayed diameter bound. $\square$

REMARK 1.39. The converse is false. A graph can have logarithmic diameter and still contain a sparse bottleneck. Expansion is a uniform statement about *every* set up to half the graph, not merely about the existence of short paths between individual points.

15. Cayley graphs and translation symmetry

Cayley graphs have more structure than arbitrary regular graphs. Left translation by any $g \in G$,

$$L_g(x) = gx,$$

is a graph automorphism of $\text{Cay}(G, S)$ because

$$L_g(xs) = gxs = (gx)s.$$

Thus the graph is vertex-transitive.

On functions, let

$$(L_g f)(x) = f(g^{-1}x).$$

The right-averaging operator commutes with every left translation:

$$P_S L_g = L_g P_S.$$

This elementary identity is the doorway to representation theory. It implies that every eigenspace of P_S is invariant under the left-regular representation of G . Consequently, any lower bound on the dimension of a nontrivial representation can force multiplicity in the nonconstant spectrum; this is one of the basic ways group structure constrains a Cayley graph.

We record the observation now without using representation theory.

PROPOSITION 1.40 (Translation equivariance). *For every finite group G and probability measure μ on G , the right-convolution operator*

$$(P_\mu f)(x) = \sum_{s \in G} \mu(s) f(xs)$$

commutes with the left-regular action of G .

PROOF. For $g, x \in G$,

$$(P_\mu L_g f)(x) = \sum_s \mu(s) f(g^{-1}xs),$$

$$(L_g P_\mu f)(x) = (P_\mu f)(g^{-1}x) = \sum_s \mu(s) f(g^{-1}xs).$$

□

16. A first formulation of the expansion problem

Suppose Γ is an infinite finitely generated group with a fixed finite symmetric generating set S . Let

$$\pi_q : \Gamma \longrightarrow G_q$$

be a family of finite quotients, and write

$$S_q = \pi_q(S).$$

One obtains finite Cayley graphs

$$X_q = \text{Cay}(G_q, S_q).$$

The expansion problem asks whether there is a constant $\gamma_0 > 0$, independent of q , such that

$$(32) \quad \gamma(X_q) \geq \gamma_0.$$

Equivalently, by Cheeger,

$$h(X_q) \geq h_0 > 0$$

uniformly in q .

For a fixed finite graph, positivity of the gap is automatic once the graph is connected. The content lies entirely in *uniformity as the quotient grows*. In an arithmetic application the group orders $|G_q|$ tend to infinity, while the generating set has fixed cardinality. The challenge is therefore to produce a global mixing phenomenon from a bounded amount of local data.

GUIDING IDEA 1.41 (The scale of the problem). A fixed generating set gives only $|S|^k$ words of length k . If $|G_q|$ is large, the random walk cannot spread over a positive proportion of the group until k is at least a constant multiple of $\log |G_q|$. Uniform expansion says, roughly, that this unavoidable logarithmic obstruction is the only obstruction: after $O(\log |G_q|)$ steps, the walk is already close to uniform.

This viewpoint explains why random walks are central. To prove Equation (32), one wants to understand the sequence

$$\mu_q, \quad \mu_q^{(2)}, \quad \mu_q^{(4)}, \quad \mu_q^{(8)}, \dots$$

and show that its ℓ^2 norm falls rapidly toward the uniform scale $|G_q|^{-1/2}$. The Bourgain–Gamburd method supplies a robust mechanism for forcing precisely that descent, provided algebraic concentration can be ruled out.

We can already isolate three logically different obstacles to rapid mixing.

- (1) **Periodic obstruction.** The walk can oscillate because the graph is bipartite or nearly bipartite. Laziness removes this issue without changing expansion by more than an absolute factor.
- (2) **Subgroup obstruction.** The walk can spend too much mass inside a proper subgroup or a coset-like algebraic region. A proof of expansion must show that this does not happen at the relevant scales.
- (3) **Approximate-subgroup obstruction.** Even when no proper subgroup captures the walk, a set with unusually small multiplicative growth can prevent rapid L^2 flattening. Product-growth theorems are designed to rule out this possibility except in controlled algebraic situations.

The first obstruction is completely understood in this chapter. The other two are the genuinely nontrivial inputs in the Bourgain–Gamburd architecture.

17. Sanity checks on normalizations

Before closing the chapter, we collect several identities that will be used repeatedly. They are elementary, but a mistake in any one of them propagates into incorrect powers of $|G|$ or incorrect spectral constants.

PROPOSITION 1.42 (Normalization checklist). *Let G be finite and S a symmetric generating multiset of cardinality d . Let $P = P_S$, let u_G be uniform measure, and let $A \subseteq G$. Then:*

(i)

$$\mathcal{E}(f) = \frac{1}{2|G|d} \sum_{x \in G} \sum_{s \in S} |f(x) - f(xs)|^2.$$

(ii)

$$\mathcal{E}(\mathbf{1}_A) = \frac{b(A)}{|G|d}.$$

(iii)

$$\text{Var}_G(\mathbf{1}_A) = \frac{|A|}{|G|} \left(1 - \frac{|A|}{|G|} \right).$$

(iv)

$$\|u_G\|_2 = |G|^{-1/2}.$$

(v) *For every probability measure ν ,*

$$|G|^{-1/2} \leq \|\nu\|_2 \leq 1.$$

The left equality holds exactly for u_G , and the right equality holds exactly for a point mass.

(vi)

$$\| |G|\nu - \mathbf{1} \|_{2,G}^2 = |G| \|\nu - u_G\|_2^2.$$

(vii) *If P has ordinary spectral gap γ , then $(I + P)/2$ has spectral gap $\gamma/2$ and has no negative spectrum.*

PROOF. Items (i)–(iii) have already been proved. For (iv),

$$\|u_G\|_2^2 = |G| \cdot |G|^{-2} = |G|^{-1}.$$

For (v), Cauchy–Schwarz gives

$$1 = \left(\sum_x \nu(x) \right)^2 \leq |G| \sum_x \nu(x)^2,$$

with equality exactly when ν is constant. The upper bound follows from

$$\sum_x \nu(x)^2 \leq \left(\sum_x \nu(x) \right)^2 = 1,$$

with equality exactly when at most one value is nonzero. Item (vi) is Equation (19), and (vii) is Proposition 1.31. $\square$

WARNING 1.43 (Three conventions that should never be merged). There are three superficially similar quantities:

$$1 - \lambda_2, \quad 1 - \max_{j \geq 2} |\lambda_j|, \quad \left\| \mu^{(k)} \right\|_2.$$

The first measures expansion, the second controls non-lazy L^2 mixing, and the third measures concentration of a probability measure in counting ℓ^2 . A large part of the Bourgain–Gamburd method consists of proving decay of the third quantity and then converting it into a lower bound for the first. Treating these quantities as interchangeable obscures the actual logic of the method.

18. Worked derivations

The goal of these examples is not merely computational practice. Each one tests a convention that is easy to mishandle inside more elaborate arguments.

EXAMPLE 1.44 (A two-point group). Let $G = \mathbb{Z}/2\mathbb{Z}$ and $S = \{1\}$. Then

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

with eigenvalues $1, -1$. Hence

$$\gamma = 2, \quad \gamma_{\text{abs}} = 0.$$

The graph has $h = 1$, so Cheeger gives

$$\frac{1}{2} \leq 2 \leq 2,$$

and the upper bound is sharp. The random walk never mixes because it alternates deterministically between the two vertices. After lazification,

$$P_L = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

whose spectrum is $1, 0$; one lazy step produces the uniform measure exactly.

EXAMPLE 1.45 (Indicator Rayleigh quotient). Let $A \subseteq V$ with $0 < |A| < n$. Put $\alpha = |A|/n$ and

$$f = \mathbf{1}_A - \alpha.$$

Then $f \perp \mathbf{1}$ and

$$\|f\|_{2,V}^2 = \alpha(1 - \alpha).$$

Because subtraction of constants leaves $\mathcal{E}$ unchanged,

$$\frac{\mathcal{E}(f)}{\|f\|_{2,V}^2} = \frac{b(A)/(nd)}{\alpha(1 - \alpha)} = \frac{b(A)}{d|A|} \frac{1}{1 - |A|/n}.$$

For $|A| \leq n/2$, this lies between

$$\frac{b(A)}{d|A|} \quad \text{and} \quad 2 \frac{b(A)}{d|A|}.$$

Thus the upper Cheeger inequality is already visible in a single exact formula.

EXAMPLE 1.46 (How ℓ^2 detects support size). Suppose a probability measure ν is supported on a set A of size M . Then Cauchy–Schwarz gives

$$1 = \left(\sum_{x \in A} \nu(x) \right)^2 \leq M \sum_x \nu(x)^2,$$

so

$$(33) \quad \|\nu\|_2 \geq M^{-1/2}.$$

Equality holds when ν is uniform on A . Thus a bound such as

$$\|\nu\|_2 \leq |G|^{-1/2+\eta}$$

forces

$$|\text{supp } \nu| \geq |G|^{1-2\eta}.$$

This is the simplest reason L^2 flattening implies combinatorial spreading.

EXAMPLE 1.47 (Collision probability). Let X, Y be independent G -valued random variables with the same law ν . Then

$$\mathbb{P}(X = Y) = \sum_{g \in G} \nu(g)^2 = \|\nu\|_2^2.$$

Thus ℓ^2 mass is also a collision probability. A flattening estimate says that two independent copies of the walk are becoming progressively less likely to land at the same point. This probabilistic interpretation is often helpful when translating between energy estimates and convolution norms.

19. Research-facing questions

The definitions above lead to several questions that already point toward the deeper theory.

QUESTION 1.48 (Uniformity in quotients). Let Γ be generated by a fixed finite symmetric set S and let $\pi_i : \Gamma \rightarrow G_i$ be finite quotients with $|G_i| \rightarrow \infty$. Which structural properties of the images $\pi_i(\Gamma)$ and of the ambient algebraic groups force

$$\inf_i \gamma(\text{Cay}(G_i, \pi_i(S))) > 0?$$

Connectivity alone gives a positive gap for each individual quotient, but gives no uniform lower bound.

QUESTION 1.49 (How much flattening is enough?). Suppose one can prove at some logarithmic time $k \asymp \log |G|$ that

$$\left\| \mu^{(k)} \right\|_2 \leq |G|^{-1/2+\eta}$$

for a small fixed $\eta > 0$. Under what additional representation-theoretic information can this be upgraded to a spectral gap? The answer depends on the minimum dimension of a nontrivial irreducible representation of G .

QUESTION 1.50 (What causes failure of flattening?). If

$$\|v * v\|_2$$

is unexpectedly close to $\|v\|_2$, then convolution has failed to spread mass. What algebraic structure must be present? In abelian groups this is related to additive energy and small doubling. In noncommutative groups the correct objects are approximate subgroups and high multiplicative energy.

These questions isolate the architecture of the method using only the material developed in this chapter.

20. Exercises and mastery checks

The exercises are chosen to reinforce proof mechanisms rather than routine calculation.

EXERCISE 1.51 (Disconnected graphs). Let X be a finite d -regular graph, not assumed connected. Prove directly from the variational characterization that $\gamma = 0$ if λ_2 is defined with multiplicity and X has at least two connected components. Reconcile this with $h(X) = 0$.

EXERCISE 1.52 (Bipartiteness and the eigenvalue -1). Let X be a connected d -regular graph. Prove that -1 is an eigenvalue of P if and only if X is bipartite. Do not use Perron–Frobenius theory; use the equality case in

$$|Pf(x)| \leq \frac{1}{d} \sum_{y \sim x} |f(y)|.$$

EXERCISE 1.53 (Vertex expansion from edge expansion). For $A \subseteq V$, define its external vertex boundary

$$\partial_V A = \{y \notin A : \exists x \in A \text{ with } x \sim y\}.$$

Show that

$$|\partial_V A| \geq \frac{b(A)}{d}.$$

Deduce that a uniform lower bound on normalized edge expansion implies a uniform lower bound on vertex expansion. Investigate the converse when the degree is uniformly bounded.

EXERCISE 1.54 (Diameter from the spectral gap). Use Cheeger's lower or upper inequality together with Proposition 1.38 to derive an $O_\gamma(\log n)$ diameter bound from a uniform spectral gap $\gamma > 0$. Track an explicit constant.

EXERCISE 1.55 (A direct spectral diameter estimate). Let P be lazy with spectral gap γ . Show that if

$$k > \frac{\log n}{\gamma},$$

then $P^k(x, y) > 0$ for every pair x, y once the numerical constant is chosen correctly. Deduce a logarithmic diameter bound without using Cheeger. Compare the constant with the one obtained from expansion.

EXERCISE 1.56 (Return probability). For a symmetric probability measure μ on a finite group G , prove

$$\mu^{(2k)}(e) = \left\| \mu^{(k)} \right\|_2^2.$$

Interpret this as a collision identity for two independent k -step walks. Explain why symmetry is essential in the displayed form.

EXERCISE 1.57 (Support growth lower bound). Let S be a generating set of cardinality d and suppose the lazy walk has spectral gap at least $\gamma_0 > 0$. Use total variation mixing to prove that for some $C = C(\gamma_0)$ and all sufficiently large G ,

$$|S^{\lfloor C \log |G| \rfloor}| \geq \frac{|G|}{2}.$$

Here S^k denotes products of at most k generators, with the identity allowed.

EXERCISE 1.58 (Exact expansion of the cycle). Determine $h(C_n)$ exactly for even and odd n . Compare $h(C_n)^2/2$ with the exact spectral gap in Equation (30) and compute the limiting ratio as $n \rightarrow \infty$.

EXERCISE 1.59 (Hypercube cut). Show that $h(Q_m) = 1/m$. One direction was obtained from a coordinate cut. For the reverse inequality, prove an edge-isoperimetric estimate sufficient for the claim, for example by induction on m .

EXERCISE 1.60 (Changing the generating multiset). Let S be a symmetric generating multiset and let $S' = S \sqcup \{e, e, \dots, e\}$ be obtained by adding r copies of the identity. Express $P_{S'}$ in terms of P_S and I , and determine exactly how the ordinary spectral gap changes. Explain why adding holding probability never creates expansion from nothing but can eliminate periodicity.

CHECKPOINT 1.61. At this point the reader should be able to move fluently among the following statements for a finite regular Cayley graph:

- every set of size at most half the group has a definite proportion of steps leaving it;
- every mean-zero function has Dirichlet energy comparable to its variance;

- the second normalized adjacency eigenvalue is bounded away from 1;
- after lazification, the random walk contracts mean-zero L^2 at a uniform exponential rate;
- mixing from a point mass occurs in $O(\log |G|)$ steps when the gap is uniform.

The reader should also be able to explain why the last statement requires control of negative spectrum unless the walk is lazy.

21. Chapter summary

We have fixed the conventions that underlie the entire expansion method. For a finite d -regular graph,

$$P = \frac{1}{d}A$$

is the normalized adjacency operator, and

$$\gamma = 1 - \lambda_2(P)$$

is the ordinary spectral gap. The normalized edge-expansion constant is

$$h = \min_{0 < |A| \leq |V|/2} \frac{b(A)}{d|A|}.$$

The Cheeger inequalities give

$$\boxed{\frac{h^2}{2} \leq \gamma \leq 2h}.$$

For mixing one must instead use the absolute gap

$$\gamma_{\text{abs}} = 1 - \max_{j \geq 2} |\lambda_j|,$$

unless the chain has been made lazy. Lazification replaces P by $(I + P)/2$, removes negative spectrum, and halves the ordinary gap.

For a probability measure ν on a finite group, the counting norm

$$\|\nu\|_2 = \left(\sum_g \nu(g)^2 \right)^{1/2}$$

runs from 1 for a point mass down to $|G|^{-1/2}$ for uniform measure. This norm measures concentration, support size, and collision probability. It is therefore the natural analytic quantity for studying how convolution spreads a random walk.

The central expansion problem can now be stated precisely: given a fixed generating set whose images generate larger and larger finite quotients, prove a spectral gap bounded below independently of the quotient. In Bourgain–Gamburd arguments this is achieved by forcing the corresponding random-walk measures to flatten rapidly enough that a uniform gap becomes unavoidable.

CHAPTER 2

Convolution, Concentration, and Probability Measures on Groups

The Bourgain–Gamburd method manipulates probability measures rather than eigenfunctions. This chapter develops a quantitative calculus for doing so. We prove the convolution inequalities that are repeatedly used in flattening arguments, relate ℓ^2 mass to collision probability and effective support, study level sets and dyadic decompositions, and isolate the special role of subgroups and cosets as exact obstructions to flattening. The main theme is that the single number $\|\mu\|_2$ contains useful but incomplete information: one must learn how to convert it into statements about where the mass lies and how convolution can change that geometry.

GUIDING IDEA 2.1 (What the norm is trying to detect). A probability measure on a finite group can fail to be uniform for two very different reasons. It may simply be *too concentrated*—for example, a point mass has $\|\mu\|_2 = 1$ —or it may have already spread over many points but remain trapped by algebraic structure, as happens for the uniform measure on a large subgroup. The first phenomenon is visible from the size of $\|\mu\|_2$ alone; the second is not. The Bourgain–Gamburd method works by alternating between these two questions:

how concentrated is the walk? and where can that concentration live?

This chapter develops the first language. Later, BSG will convert a failure of L^2 -flattening into a structured set, and the group-theoretic part of the argument will rule out that structure. It is useful to keep this division of labour in mind: the inequalities below do not know what a subgroup or an approximate group “means”; they only measure collisions and concentration.

1. Convolution conventions revisited

Let G be a finite group. For functions $f, g : G \rightarrow \mathbb{C}$ we use

$$(34) \quad (f * g)(x) = \sum_{y \in G} f(y)g(y^{-1}x).$$

The involution is

$$(35) \quad f^*(x) = \overline{f(x^{-1})}.$$

For a set $A \subseteq G$, 1_A denotes its indicator, and

$$\mu_A = \frac{1_A}{|A|}$$

denotes the uniform probability measure on A when $A \neq \emptyset$.

The definition in Equation (34) is chosen so that independent random variables multiply in the same order as their laws convolve. If X has law f and Y has law g , with f, g probability measures, then XY has law $f * g$.

LEMMA 2.2 (Algebra of convolution). *For functions on a finite group:*

- (i) *convolution is associative;*
- (ii) *δ_e is the identity for convolution;*
- (iii) *$(f * g)^* = g^* * f^*$;*
- (iv) *$\sum_x (f * g)(x) = (\sum_x f(x)) (\sum_x g(x))$;*
- (v) *if $f, g \geq 0$, then $f * g \geq 0$.*

Hence probability measures are closed under convolution.

PROOF. Only associativity merits writing out. We have

$$\begin{aligned} ((f * g) * h)(x) &= \sum_z \sum_y f(y) g(y^{-1}z) h(z^{-1}x) \\ &= \sum_y f(y) \sum_w g(w) h(w^{-1}y^{-1}x) = (f * (g * h))(x), \end{aligned}$$

where $w = y^{-1}z$. The other identities follow by direct substitutions; for the involution, put $z = x^{-1}y$ after taking complex conjugates. $\square$

WARNING 2.3 (Left and right conventions). The graph operator in Chapter 1 was written

$$P_S f(x) = \frac{1}{|S|} \sum_{s \in S} f(xs),$$

which is right averaging on functions. The convolution operator

$$L_\mu f = \mu * f$$

is left convolution. For symmetric measures the spectra relevant to Cayley expansion agree, but the two actions should not be identified term by term. We keep the convolution convention Equation (34) because it makes products of independent random variables transparent and makes the Fourier transform multiplicative in the natural order.

2. Counting norms and elementary inequalities

For $1 \leq p < \infty$ write

$$\|f\|_p = \left(\sum_{x \in G} |f(x)|^p \right)^{1/p}, \quad \|f\|_\infty = \max_x |f(x)|.$$

These are counting norms, not normalized norms.

LEMMA 2.4 (Young inequalities). *For functions $f, g : G \rightarrow \mathbb{C}$,*

$$(36) \quad \|f * g\|_1 \leq \|f\|_1 \|g\|_1,$$

$$(37) \quad \|f * g\|_2 \leq \|f\|_1 \|g\|_2,$$

$$(38) \quad \|f * g\|_2 \leq \|f\|_2 \|g\|_1,$$

$$(39) \quad \|f * g\|_\infty \leq \|f\|_2 \|g\|_2.$$

In particular, convolution by a probability measure is a contraction on $\ell^2(G)$.

PROOF. The ℓ^1 inequality follows from the triangle inequality and the bijection $x \mapsto y^{-1}x$. For Equation (37), write

$$f * g = \sum_y f(y) \tau_y g, \quad (\tau_y g)(x) = g(y^{-1}x),$$

and use the triangle inequality in ℓ^2 together with $\|\tau_y g\|_2 = \|g\|_2$. The other ℓ^2 inequality follows by applying this to $(f * g)^* = g^* * f^*$. Finally, for each x ,

$$|(f * g)(x)| \leq \left(\sum_y |f(y)|^2 \right)^{1/2} \left(\sum_y |g(y^{-1}x)|^2 \right)^{1/2} = \|f\|_2 \|g\|_2.$$

□

The following identity is elementary but fundamental.

LEMMA 2.5 (Inner-product/convolution identity). *With counting inner product $\langle f, g \rangle = \sum_x f(x) \overline{g(x)}$,*

$$\langle f * g, h \rangle = \langle g, f^* * h \rangle = \langle f, h * g^* \rangle.$$

Consequently

$$(40) \quad \|f\|_2^2 = (f^* * f)(e) = (f * f^*)(e).$$

PROOF. Expand the left side and interchange the finite sums:

$$\begin{aligned} \langle f * g, h \rangle &= \sum_{x,y} f(y) g(y^{-1}x) \overline{h(x)} \\ &= \sum_{y,z} f(y) g(z) \overline{h(yz)}. \end{aligned}$$

The two asserted forms are obtained by rewriting the same double sum. Setting the appropriate functions to point masses gives Equation (40) directly as well. $\square$

3. Collision probability and effective support

If μ is a probability measure and X, Y are independent with law μ , then

$$\|\mu\|_2^2 = \mathbb{P}(X = Y).$$

For a nonsymmetric law there is another useful form. If X and Y are independent with law μ , then $X^{-1}Y$ has law $\mu^* * \mu$, and

$$(\mu^* * \mu)(e) = \|\mu\|_2^2.$$

Thus ℓ^2 mass is exactly a return/collision statistic.

DEFINITION 2.6 (Effective support size). For a probability measure μ define

$$(41) \quad N_2(\mu) = \|\mu\|_2^{-2}.$$

If μ is uniform on a set A , then $N_2(\mu) = |A|$. For general μ the number $N_2(\mu)$ is the support size suggested by its collision probability.

PROPOSITION 2.7 (Basic support bounds). *For every probability measure μ ,*

$$1 \leq N_2(\mu) \leq |\text{supp } \mu| \leq |G|.$$

Moreover,

$$|G|^{-1/2} \leq \|\mu\|_2 \leq 1,$$

with equality on the left precisely for the uniform measure and equality on the right precisely for a point mass.

PROOF. Cauchy–Schwarz on the support gives

$$1 = \left(\sum_{x \in \text{supp } \mu} \mu(x) \right)^2 \leq |\text{supp } \mu| \|\mu\|_2^2.$$

The remaining claims follow from this and the same inequality with the full group as support. Equality in Cauchy–Schwarz forces constancy on the relevant support. $\square$

GUIDING IDEA 2.8. The heuristic “ μ occupies about $N_2(\mu)$ points” is useful but dangerous. Two measures can have the same ℓ^2 norm while one is nearly uniform on a set and the other has a heavy spike plus a diffuse tail. Flattening proofs therefore supplement ℓ^2 information by decomposing the measure into level sets.

4. Heavy sets and level-set extraction

A tempting but false inference is that the single number $M = \|\mu\|_2^{-2}$ forces a fixed proportion of the probability mass onto $O(M)$ points. Measures with one moderate spike and a very long diffuse tail show that no such universal statement holds. What one can always extract is a level set carrying a controlled portion of the ℓ^2 norm, after discarding a small pointwise tail. This is exactly the form needed in flattening arguments.

WARNING 2.9 (A common false “effective support” heuristic). The identity $N_2(\mu) = \|\mu\|_2^{-2}$ does not mean that, say, half the mass must lie on $O(N_2(\mu))$ points. The quantity N_2 measures collision scale, not a hard quantile of the distribution. Whenever a proof needs a nearly uniform set, it must obtain one by a level-set or energy argument rather than by this heuristic alone.

LEMMA 2.10 (Truncated dyadic decomposition). *Let μ be a probability measure on a set of size N , and let $0 < \eta < 1$. Set*

$$J = \left\lceil \log_2 \frac{N}{\eta} \right\rceil.$$

For $0 \leq j < J$ define

$$A_j = \{x : 2^{-j-1} < \mu(x) \leq 2^{-j}\},$$

and put $R = \{x : \mu(x) \leq 2^{-J}\}$. Then

$$(42) \quad \mu = \sum_{j=0}^{J-1} \mu 1_{A_j} + \mu 1_R,$$

with

$$\|\mu 1_R\|_1 \leq \eta, \quad \|\mu 1_R\|_2 \leq \eta N^{-1/2}.$$

Consequently some $j < J$ satisfies

$$(43) \quad \|\mu 1_{A_j}\|_2 \geq \frac{\|\mu\|_2 - \eta N^{-1/2}}{J}.$$

PROOF. Since $|R| \leq N$ and every point of R has mass at most $2^{-J} \leq \eta/N$,

$$\|\mu 1_R\|_1 \leq \eta, \quad \|\mu 1_R\|_2^2 \leq N(\eta/N)^2 = \eta^2/N.$$

The triangle inequality in ℓ^2 applied to Equation (42) gives

$$\|\mu\|_2 \leq \sum_{j < J} \|\mu 1_{A_j}\|_2 + \eta N^{-1/2},$$

which implies Equation (43). □

REMARK 2.11. The logarithmic factor in a dyadic decomposition is usually harmless in a Bourgain–Gamburd argument. If all key estimates have power savings $|G|^{-c}$, then $\log |G|$ can be absorbed into $|G|^\varepsilon$ after decreasing an exponent. What is not harmless is silently replacing a logarithmic loss by an absolute constant.

5. Entropy as a complementary measure of spreading

For a probability measure μ define the Shannon entropy

$$H(\mu) = - \sum_x \mu(x) \log \mu(x),$$

with $0 \log 0 = 0$. The collision or Rényi entropy of order 2 is

$$H_2(\mu) = - \log \sum_x \mu(x)^2 = \log N_2(\mu).$$

PROPOSITION 2.12 (Entropy comparison). *For every probability measure on a finite set,*

$$H_2(\mu) \leq H(\mu) \leq \log |\text{supp } \mu|.$$

Equivalently,

$$\|\mu\|_2^{-2} \leq e^{H(\mu)} \leq |\text{supp } \mu|.$$

PROOF. Let $m = |\text{supp } \mu|$. Concavity of $\log$ gives the upper bound directly:

$$H(\mu) = \sum_{\mu(x)>0} \mu(x) \log \frac{1}{\mu(x)} \leq \log \left(\sum_{\mu(x)>0} \mu(x) \frac{1}{\mu(x)} \right) = \log m.$$

For the lower bound, apply Jensen's inequality to the convex function $-\log t$ with respect to the probability weights $\mu(x)$:

$$H(\mu) = \sum_x \mu(x) (-\log \mu(x)) \geq - \log \left(\sum_x \mu(x)^2 \right) = H_2(\mu).$$

□

REMARK 2.13. Entropy is conceptually useful because convolution of independent variables cannot decrease Shannon entropy in many natural settings, but the ℓ^2 norm interacts more directly with energy and representation theory. The Bourgain–Gamburd method is therefore naturally a Rényi-2 method.

6. Convolution can only flatten in ℓ^2

If μ and ν are probability measures, Young's inequality gives

$$(44) \quad \|\mu * \nu\|_2 \leq \min(\|\mu\|_2, \|\nu\|_2).$$

Hence

$$\|\mu^{(n+1)}\|_2 \leq \|\mu^{(n)}\|_2.$$

The question is when the inequality is quantitatively strict.

PROPOSITION 2.14 (Equality for convolution by a probability measure). *Let ν be a probability measure and $f \in \ell^2(G)$. Then*

$$\|\nu * f\|_2 = \|f\|_2$$

if and only if all translates $\tau_y f$, $y \in \text{supp } \nu$, that occur with positive weight are identical. Equivalently,

$$f(y^{-1}x) = f(z^{-1}x)$$

for all x and all $y, z \in \text{supp } \nu$.

PROOF. In the proof of Young's inequality we wrote

$$\nu * f = \sum_y \nu(y) \tau_y f.$$

This is a convex combination of vectors all having norm $\|f\|_2$. The norm is strictly convex in a Hilbert space: equality in the triangle inequality for a convex combination of equal-norm vectors occurs exactly when all nonzero-weight vectors agree. $\square$

COROLLARY 2.15 (Exact subgroup obstruction). *Put*

$$H = \langle (\text{supp } \nu)^{-1} (\text{supp } \nu) \rangle.$$

Then equality in Proposition 2.14 holds exactly when f is left H -invariant.

PROOF. The equality condition says $\tau_{z^{-1}y} f = f$ for all y, z in the support. These elements generate H . $\square$

GUIDING IDEA 2.16. This exact equality statement predicts the approximate theory. If convolution hardly decreases ℓ^2 norm, then many translates of the measure must be nearly parallel. Approximate parallelism creates many multiplicative coincidences; those coincidences are measured by multiplicative energy; high energy produces an approximate subgroup. The flattening lemma is a quantitative stability version of this simple Hilbert-space observation.

7. Subgroups and cosets as model measures

Let $H \leq G$. The uniform measure

$$u_H = \frac{1_H}{|H|}$$

is idempotent:

$$(45) \quad \nu_H * \nu_H = \nu_H.$$

Indeed, for $x \in H$ there are exactly $|H|$ pairs $(h_1, h_2) \in H^2$ with $h_1 h_2 = x$, and there are none for $x \notin H$.

For left and right cosets one has

$$\delta_g * \nu_H = \mu_{gH}, \quad \nu_H * \delta_g = \mu_{Hg}.$$

Thus coset-uniform measures are the simplest nonuniform probability measures whose convolution can remain highly concentrated.

PROPOSITION 2.17 (Projection associated with a subgroup). *Left convolution by ν_H is the orthogonal projection in counting $\ell^2(G)$ onto the subspace of left H -invariant functions:*

$$(\nu_H * f)(x) = \frac{1}{|H|} \sum_{h \in H} f(h^{-1}x).$$

In particular $(\nu_H^*)^2 = \nu_H^*$ as operators.

PROOF. The formula shows invariance under left multiplication by H . If f is already left H -invariant, the average equals f . Since $\nu_H^* = \nu_H$, the operator is self-adjoint. An idempotent self-adjoint operator is the orthogonal projection onto its image. $\square$

8. Mass on cosets and non-concentration

A random walk can fail to mix because it puts too much mass on a proper subgroup or one of its cosets. Several equivalent-looking formulations are useful.

DEFINITION 2.18 (Coset concentration parameter). For a probability measure μ and a subgroup $H \leq G$, define

$$\alpha_H(\mu) = \max_{g \in G} \mu(gH).$$

LEMMA 2.19 (Coset concentration controls subgroup smoothing). *For every probability measure μ and subgroup H ,*

$$\|\mu * \nu_H\|_\infty = \frac{\alpha_H(\mu)}{|H|}.$$

Moreover

$$\|\mu * \nu_H\|_2^2 = \frac{1}{|H|} \sum_{gH \in G/H} \mu(gH)^2 \leq \frac{\alpha_H(\mu)}{|H|}.$$

PROOF. For $x \in G$,

$$\begin{aligned} (\mu * \nu_H)(x) &= \frac{1}{|H|} \sum_y \mu(y) 1_H(y^{-1}x) \\ &= \frac{1}{|H|} \mu(xH), \end{aligned}$$

where xH is the right coset containing the possible y . The function is constant on right cosets. Summing its square over the $|H|$ elements of each coset gives the identity, and

$$\sum_C \mu(C)^2 \leq \alpha_H(\mu) \sum_C \mu(C) = \alpha_H(\mu).$$

□

REMARK 2.20. For left versus right cosets, one should follow the side dictated by the convolution convention. Symmetric random walks often allow one to pass between the two, but a proof should not use this silently.

9. Symmetrization

Nonsymmetric measures can be reduced to symmetric positive-definite objects by convolving with the involution.

DEFINITION 2.21 (Symmetrized measure). For a probability measure μ , set

$$\sigma_\mu = \mu^* * \mu.$$

Then σ_μ is a symmetric probability measure and

$$\sigma_\mu(e) = \|\mu\|_2^2.$$

LEMMA 2.22 (Norm relation under symmetrization). For every μ ,

$$\|\mu^* * \mu\|_2 = \|\mu * \mu^*\|_2.$$

If T_μ denotes left convolution by μ on $\ell^2(G)$, then

$$T_\mu^* T_\mu = T_{\mu^* * \mu}, \quad T_\mu T_\mu^* = T_{\mu * \mu^*}.$$

Thus these two positive operators have the same nonzero eigenvalues, counted with multiplicity.

PROOF. The adjoint identity follows from Lemma 2.5. The two products $T_\mu^* T_\mu$ and $T_\mu T_\mu^*$ have the same nonzero singular-value squares. Their Hilbert–Schmidt norms are therefore equal. Since the Hilbert–Schmidt norm of a convolution operator is $|G|^{1/2}$ times the counting ℓ^2 norm of its kernel, the asserted equality follows. □

10. Return probabilities and even times

For a symmetric probability measure $\mu = \mu^*$,

$$(46) \quad \mu^{(2n)}(e) = \|\mu^{(n)}\|_2^2.$$

This identity is often the cleanest bridge between random-walk estimates and flattening.

PROPOSITION 2.23 (Log-convexity of even return probabilities). Let μ be symmetric. Then the sequence

$$r_n = \mu^{(2n)}(e)$$

is nonincreasing and log-convex in the sense that

$$r_n^2 \leq r_{n-1} r_{n+1} \quad (n \geq 1).$$

PROOF. Monotonicity is Young's inequality applied to $\mu^{(n+1)} = \mu^{(n)} * \mu$. For log-convexity, use the self-adjoint convolution operator $T = T_\mu$ and the point mass δ_e :

$$r_n = \langle T^{2n} \delta_e, \delta_e \rangle = \|T^n \delta_e\|_2^2.$$

By the spectral theorem there is a finite positive measure ω on $[-1, 1]$ such that

$$r_n = \int t^{2n} d\omega(t).$$

Cauchy–Schwarz applied to t^{n-1} and t^{n+1} gives the claim. $\square$

11. A quantitative notion of flattening

Suppose μ is a probability measure with

$$|G|^{-1/2} \ll \|\mu\|_2 \ll 1.$$

We say that one convolution step produces a power-type flattening if

$$(47) \quad \|\mu * \mu\|_2 \leq \|\mu\|_2^{1+\delta}$$

for some fixed $\delta > 0$.

The exponent form is useful because iteration is extremely fast. If $a_j = -\log \|\mu_j\|_2$ and

$$\|\mu_{j+1}\|_2 \leq \|\mu_j\|_2^{1+\delta},$$

then

$$a_{j+1} \geq (1 + \delta)a_j.$$

Thus a logarithmic amount of entropy becomes a constant fraction of $\log |G|$ after only $O_\delta(\log \log |G|)$ successful flattening stages.

PROPOSITION 2.24 (Iteration of a power flattening estimate). *Let $0 < a < 1$ and suppose a sequence $r_j \in (0, 1]$ satisfies*

$$r_{j+1} \leq r_j^{1+\delta}$$

for $0 \leq j < m$. Then

$$r_m \leq a^{(1+\delta)^m}$$

whenever $r_0 \leq a$. Consequently, if $a \leq c < 1$ is bounded away from 1, then $m = O_\delta(\log \log N)$ suffices to force $r_m \leq N^{-c_1}$ for any prescribed fixed $c_1 > 0$.

PROOF. Induction gives the first assertion. Taking logarithms, one requires

$$(1 + \delta)^m |\log a| \geq c_1 \log N,$$

which is equivalent to $m \geq (\log \log N + O(1))/\log(1 + \delta)$. $\square$

12. What the norm alone cannot tell us

Three examples delimit the information content of $\|\mu\|_2$.

EXAMPLE 2.25 (Uniform on a subgroup). If $H < G$ is proper and $\mu = \nu_H$, then

$$\|\mu\|_2 = |H|^{-1/2}, \quad \mu * \mu = \mu.$$

The measure can have very small ℓ^2 norm while being permanently trapped far from uniform on G .

EXAMPLE 2.26 (A spike plus a uniform background). Let

$$\mu = \theta \delta_e + (1 - \theta)u_G.$$

Then

$$\|\mu\|_2^2 = \theta^2 + \frac{1 - \theta^2}{|G|}.$$

For fixed θ , the norm is bounded away from both 1 and $|G|^{-1/2}$ even though almost all points carry mass approximately $(1 - \theta)/|G|$. A single effective-support number does not describe this two-scale geometry.

EXAMPLE 2.27 (Two measures with the same effective support). Take $M \ll N$. The uniform measure on M points has norm $M^{-1/2}$. One can choose a two-level measure supported on all N points with the same ℓ^2 norm by placing a suitable small spike on one point and spreading the remaining mass uniformly. Their convolution behavior can be completely different.

13. A practical dyadic calculus

For flattening arguments one often writes, schematically,

$$(48) \quad \mu = \sum_i 2^{-i} 1_{A_i} + \text{small error},$$

where the A_i are disjoint level sets. The exact coefficients need only be comparable to the pointwise values.

LEMMA 2.28 (Dyadic majorant and minorant). *Let μ be a probability measure and $0 < \eta < 1$. With J and A_j as in Lemma 2.10, define*

$$\mu_- = \sum_{j < J} 2^{-j-1} 1_{A_j}, \quad \mu_+ = \sum_{j < J} 2^{-j} 1_{A_j}.$$

Then pointwise

$$\mu_- \leq \mu \leq \mu_+ + \frac{\eta}{|G|} 1_G, \quad \mu_+ \leq 2\mu \quad \text{on } G \setminus R.$$

Hence for every nonnegative $F : G \rightarrow \mathbb{R}$,

$$\langle \mu, F \rangle \leq \langle \mu_+, F \rangle + \eta \|F\|_\infty.$$

PROOF. All assertions are immediate from the definition of the dyadic level sets and the bound $\mu \leq \eta/|G|$ on R . $\square$

This device turns an analytic measure problem into finitely many set problems. The cost is $O(\log(|G|/\eta))$ levels.

14. Mastery checks

EXERCISE 2.29 (Convolution and translations). Verify directly that left convolution commutes with right translation and right convolution commutes with left translation. Explain why this is the representation-theoretic reason Cayley operators admit a decomposition indexed by irreducible representations.

EXERCISE 2.30 (Equality in Young). Prove [Proposition 2.14](#) using the parallelogram identity rather than strict convexity. Quantify the defect: if $\|\nu * f\|_2^2 \geq (1 - \varepsilon)\|f\|_2^2$, show that the average squared distance between two independent translates $\tau_Y f$ and $\tau_Z f$ is at most $2\varepsilon\|f\|_2^2$.

EXERCISE 2.31 (Coset smoothing). Prove the left-coset analogue of [Lemma 2.19](#) for $\nu_H * \mu$. Check carefully which cosets appear.

EXERCISE 2.32 (Rényi entropies). For $q > 1$, define

$$H_q(\mu) = \frac{1}{1-q} \log \sum_x \mu(x)^q.$$

Show that H_q is maximized by the uniform distribution and compute it for a uniform measure on a set. Investigate which parts of the flattening method are special to $q = 2$.

EXERCISE 2.33 (Near-idempotence). Suppose μ is a symmetric probability measure and $\|\mu * \mu - \mu\|_2$ is very small. Use [Proposition 2.17](#) to explain why subgroup-uniform measures are the exact models. Do not attempt a full stability theorem; identify the obstruction that a stability theorem would have to classify.

CHECKPOINT 2.34. The reader should now be able to explain, without invoking representation theory, why the following chain is plausible:

little L^2 decay under convolution $\Rightarrow$ many nearly coincident translates
 $\Rightarrow$ many multiplicative coincidences
 $\Rightarrow$ structured support.

The missing quantitative bridge is multiplicative energy and Balog–Szemerédi–Gowers theory.

15. Chapter summary

For probability measures on a finite group, counting ℓ^2 is the central concentration norm. Its reciprocal square is an effective support size, and its square is a collision probability. Convolution by a probability measure never increases this norm. Exact failure to decrease is explained by translation invariance, with subgroup-uniform measures giving the model obstruction.

The norm does not by itself describe the geometry of the measure. A rigorous flattening argument therefore uses level sets, usually through a truncated dyadic decomposition. Logarithmic losses from this decomposition must be displayed rather than silently suppressed. The next structural layers of the method will identify what approximate invariance under many translations forces algebraically.

CHAPTER 3

Property (T), Property (τ), and Uniform Expansion

Expansion of finite quotients can be viewed either as a sequence of graph inequalities or as a uniform representation-theoretic isolation of the trivial representation. Property (T) imposes such isolation simultaneously for all unitary representations of an infinite group; property (τ) imposes it only for a specified family of finite quotients. The Bourgain–Gamburd method is important precisely because it proves property-(τ)-type conclusions in settings where property (T) is unavailable. This chapter develops the equivalence between property (τ) and expansion with complete normalization, explains dependence on the generating set, and separates the qualitative fact of generation from the quantitative fact of a uniform Kazhdan constant.

GUIDING IDEA 3.1 (Three levels of uniformity). There are three logically different statements that are easy to blur together. A set S may generate a group; a family of finite Cayley graphs may have a uniform spectral gap; or an infinite group may have property (T). Generation is qualitative and contains no numerical information. Property (τ) is a uniform statement only for a specified family of finite quotients. Property (T) controls *all* unitary representations and is therefore stronger. The Bourgain–Gamburd method is valuable precisely because it can prove the middle statement without requiring the strongest one.

When reading this chapter, translate every occurrence of “almost invariant vector” into the geometric picture of a function on a Cayley graph that changes very little under each generator. Property (T) or (τ) says that such a function must be close to an invariant one. In a finite transitive action, invariant functions are the constants, so this is exactly a Poincaré inequality and hence a spectral-gap statement.

1. Almost invariant vectors

Let Γ be a finitely generated group and let S be a finite symmetric generating set. If $\pi : \Gamma \rightarrow U(\mathcal{H})$ is a unitary representation, define the averaging operator

$$(49) \quad A_{\pi, S} = \frac{1}{|S|} \sum_{s \in S} \pi(s).$$

It is self-adjoint because $S = S^{-1}$.

A vector v is invariant if $\pi(g)v = v$ for every $g \in \Gamma$. Since S generates, this is equivalent to invariance under every $s \in S$.

DEFINITION 3.2 (Kazhdan displacement). For a unitary representation π with no nonzero invariant vectors, set

$$\kappa_S(\pi) = \inf_{\|v\|=1} \max_{s \in S} \|\pi(s)v - v\|.$$

For a class $\mathcal{C}$ of unitary representations define

$$\kappa_S(\mathcal{C}) = \inf_{\pi \in \mathcal{C}} \kappa_S(\pi).$$

LEMMA 3.3 (Displacement and the averaging operator). *For every unit vector v ,*

$$(50) \quad 1 - \langle A_{\pi, S} v, v \rangle = \frac{1}{2|S|} \sum_{s \in S} \|\pi(s)v - v\|^2.$$

Consequently, if $\kappa_S(\pi) \geq \kappa$, then

$$\langle A_{\pi, S} v, v \rangle \leq 1 - \frac{\kappa^2}{2|S|}$$

for every unit vector v . Conversely, if

$$\langle A_{\pi, S} v, v \rangle \leq 1 - \gamma$$

for every unit vector v , then

$$\max_{s \in S} \|\pi(s)v - v\| \geq \sqrt{2\gamma}$$

for every unit vector v .

PROOF. Since $\pi(s)$ is unitary,

$$\|\pi(s)v - v\|^2 = 2\|v\|^2 - 2\Re \langle \pi(s)v, v \rangle.$$

Average over S . Symmetry makes $\langle A_{\pi, S} v, v \rangle$ real and gives Equation (50). If one summand is at least κ^2 , the average of the squares is at least $\kappa^2/|S|$. Conversely, if the average square is at least 2γ , its maximum is at least that large. $\square$

REMARK 3.4. The constants in the two directions are intentionally not optimized into a false equivalence. A lower bound on the *maximum* displacement yields only one large summand, while a spectral gap controls the *average* squared displacement.

2. Property (T)

DEFINITION 3.5 (Property (T)). A group Γ has Kazhdan's property (T) if there exist a finite set $Q \subset \Gamma$ and $\kappa > 0$ such that every unitary representation π with a unit vector v satisfying

$$\max_{q \in Q} \|\pi(q)v - v\| < \kappa$$

has a nonzero invariant vector.

For finitely generated groups, once property (T) holds, every finite generating set is a Kazhdan set with some positive constant. The constant depends on the generating set.

PROPOSITION 3.6 (Change of generating set). *Let S and T be finite generating sets of Γ . Suppose every $t \in T$ is a word of length at most L in $S \cup S^{-1}$. Then for every unitary representation and vector v ,*

$$\max_{t \in T} \|\pi(t)v - v\| \leq L \max_{s \in S} \|\pi(s)v - v\|.$$

Hence positivity of a Kazhdan constant is independent of the chosen finite generating set.

PROOF. If $t = s_1 \cdots s_m$ with $m \leq L$, telescope:

$$\pi(t)v - v = \sum_{j=1}^m \pi(s_1 \cdots s_{j-1})(\pi(s_j)v - v).$$

Unitary invariance of the norm gives the estimate. □

3. Finite quotients and the regular representation

Let $N \triangleleft \Gamma$ be a finite-index normal subgroup and put $G = \Gamma/N$. The right regular representation of G on $L^2(G)$ is

$$(R_g f)(x) = f(xg).$$

The Cayley averaging operator associated with the image S_N of S is

$$P_{S_N} = \frac{1}{|S|} \sum_{s \in S} R_{\bar{s}}.$$

The invariant vectors are exactly the constant functions when S_N generates G .

Thus the graph spectral gap is exactly the failure of the regular representation on mean-zero functions to have almost invariant vectors.

PROPOSITION 3.7 (Gap as a quotient Kazhdan inequality). *For $G = \Gamma/N$, let $L_0^2(G)$ denote mean-zero functions. The ordinary spectral gap γ_N of $\text{Cay}(G, S_N)$ is the largest γ such that*

$$\frac{1}{2|S|} \sum_{s \in S} \|\bar{R}_s f - f\|_{2,G}^2 \geq \gamma \|f\|_{2,G}^2 \quad (f \in L_0^2(G)).$$

PROOF. This is exactly the Dirichlet-form identity and variational characterization from [Chapter 1](#), written for the quotient regular representation. □

4. Property (τ)

Let $\mathcal{N}$ be a family of finite-index normal subgroups of Γ .

DEFINITION 3.8 (Property (τ) relative to a family). The group Γ has property (τ) with respect to $\mathcal{N}$ if the trivial representation is uniformly isolated among the nontrivial irreducible representations that factor through some quotient Γ/N , $N \in \mathcal{N}$. Equivalently, for one, hence every, finite generating set S , there is $\kappa > 0$ such that every representation in this quotient class without invariant vectors satisfies

$$\max_{s \in S} \|\pi(s)v - v\| \geq \kappa \|v\|$$

for all v .

THEOREM 3.9 (Property (τ) and expansion). *Let Γ be generated by a fixed finite symmetric set S , and let $\mathcal{N}$ be a family of finite-index normal subgroups such that the images of S generate every quotient. Then the following are equivalent:*

- (i) Γ has property (τ) with respect to $\mathcal{N}$;
- (ii) there exists $\gamma > 0$ such that every Cayley graph

$$\text{Cay}(\Gamma/N, S_N), \quad N \in \mathcal{N},$$

has ordinary spectral gap at least γ ;

- (iii) the same family has a uniform positive Cheeger constant.

PROOF. The equivalence of (ii) and (iii) is Cheeger's inequality from [Chapter 1](#). If property (τ) holds with displacement constant κ , apply [Lemma 3.3](#) to every irreducible constituent of $L_0^2(\Gamma/N)$ to obtain a uniform gap at least $\kappa^2/(2|S|)$. Conversely, a uniform gap γ in the quotient regular representations implies a displacement lower bound $\sqrt{2\gamma}$ by the other direction of [Lemma 3.3](#). Every irreducible representation factoring through a quotient embeds in the regular representation of that quotient, so the same lower bound applies to the entire quotient class. $\square$

5. Why property (T) is stronger

If Γ has property (T), then it has property (τ) with respect to every family of finite-index normal subgroups by the proposition below. The same conclusion for arbitrary finite-index subgroups is true as well, but the relevant finite object is then a Schreier graph rather than a quotient Cayley graph; we prove that extension explicitly in [Theorem 3.17](#). The converse is false: property (τ) controls only a prescribed family of finite quotient or finite permutation representations.

PROPOSITION 3.10 (Property (T) implies all quotient expansion). *Let Γ be finitely generated and have property (T). For any fixed finite symmetric generating set S , there is $\gamma_S > 0$ such that for every finite quotient $q : \Gamma \rightarrow G$ for which $q(S)$ generates G ,*

$$\gamma(\text{Cay}(G, q(S))) \geq \gamma_S.$$

PROOF. Take a Kazhdan constant for S and apply Lemma 3.3 to $L_0^2(G)$. $\square$

GUIDING IDEA 3.11. The Bourgain–Gamburd method is most striking when the ambient infinite group does not supply property (T). Instead of proving that *every* unitary representation has no almost invariant vectors, one analyzes the random walk in each finite quotient directly. Product growth, subgroup escape, and quasirandomness together manufacture a quotient-specific spectral gap that is nevertheless uniform in the quotient size.

6. Qualitative generation versus quantitative expansion

If S_N generates a finite quotient G_N , then the Cayley graph is connected and has some positive spectral gap. Nothing in connectivity prevents that gap from tending to zero.

EXAMPLE 3.12 (Cyclic quotients). Let $\Gamma = \mathbb{Z}$ with $S = \{1, -1\}$ and $N = n\mathbb{Z}$. Every quotient is generated by the image of S , but the Cayley graph is C_n and

$$\gamma(C_n) = 1 - \cos(2\pi/n) \asymp n^{-2}.$$

Thus residual finiteness plus generation has no quantitative expansion content.

EXAMPLE 3.13 (A generating-set warning). Even for a family of groups known to admit expander generating sets, a poorly chosen varying generating set can destroy uniform expansion. Property (τ) is a statement about a fixed generating set in the infinite group, or equivalently any other fixed finite generating set after a bounded word-length comparison. It is not a claim about arbitrary generating sets whose word complexity may grow with the quotient.

7. Normal versus nonnormal subgroups

Property (τ) is often formulated for finite-index subgroups that need not be normal. Then the relevant representation is the permutation representation on Γ/H , and the graph is a Schreier graph.

DEFINITION 3.14 (Schreier graph). Let $H \leq \Gamma$ have finite index. The Schreier graph $\text{Sch}(\Gamma/H, S)$ has vertex set of right cosets $H\backslash\Gamma$ and edges

$$Hx \longleftrightarrow Hxs \quad (s \in S).$$

For $s \in S$ define the permutation operator on $\ell^2(H\backslash\Gamma)$ by

$$(R_s f)(Hx) = f(Hxs),$$

and put

$$P_H = \frac{1}{|S|} \sum_{s \in S} R_s.$$

The following proposition supplies the operator-theoretic statement that is often hidden in the phrase “the Cayley argument also works for Schreier graphs.”

PROPOSITION 3.15 (Schreier Dirichlet form and spectral gap). *Let Γ be generated by a finite symmetric set S and let $H \leq \Gamma$ have finite index. Then every R_s is unitary, P_H is self-adjoint, and the invariant vectors for the permutation representation on $\ell^2(H \backslash \Gamma)$ are exactly the constants. If γ_H denotes the ordinary spectral gap of P_H , then*

$$\gamma_H = \inf_{\substack{f \perp \mathbf{1} \\ f \neq 0}} \frac{1}{2|S|} \frac{\sum_{s \in S} \|R_s f - f\|_2^2}{\|f\|_2^2}.$$

PROOF. Right multiplication by s permutes the finite coset set $H \backslash \Gamma$, so R_s is unitary and $R_s^* = R_{s^{-1}}$. Symmetry of S therefore makes P_H self-adjoint. Because S generates Γ , its right action on $H \backslash \Gamma$ is transitive. A function fixed by every R_s is constant on each orbit, hence is constant.

For every f ,

$$\begin{aligned} \frac{1}{2|S|} \sum_{s \in S} \|R_s f - f\|_2^2 &= \frac{1}{2|S|} \sum_s (2\|f\|_2^2 - 2\Re \langle R_s f, f \rangle) \\ &= \langle (I - P_H)f, f \rangle, \end{aligned}$$

where the last equality uses self-adjointness of P_H . Restricting to the orthogonal complement of the constants and applying the variational characterization of the largest nontrivial eigenvalue gives the displayed formula. $\square$

DEFINITION 3.16 (Property (τ) for an arbitrary subgroup family). Let $\mathcal{H}$ be a family of finite-index subgroups of Γ . We say that Γ has property (τ) with respect to $\mathcal{H}$ if, for one (hence every) fixed finite generating set S , there is $\kappa > 0$ such that for every $H \in \mathcal{H}$ and every $f \perp \mathbf{1}$ in $\ell^2(H \backslash \Gamma)$,

$$\max_{s \in S} \|R_s f - f\|_2 \geq \kappa \|f\|_2.$$

Equivalently, by Proposition 3.15, the Schreier graphs

$$\text{Sch}(H \backslash \Gamma, S), \quad H \in \mathcal{H},$$

have a uniform positive ordinary spectral gap.

THEOREM 3.17 (Property (T) gives uniform Schreier expansion). *Suppose Γ has property (T) and let S be a fixed finite symmetric generating set with Kazhdan constant $\kappa_S > 0$. Then for every finite-index subgroup $H \leq \Gamma$,*

$$\gamma(\text{Sch}(H \backslash \Gamma, S)) \geq \frac{\kappa_S^2}{2|S|}.$$

Consequently Γ has property (τ) with respect to every family of finite-index subgroups, normal or not.

PROOF. The permutation representation of Γ on $\ell^2(H \setminus \Gamma)$ is unitary. By transitivity its invariant subspace is the constants, so the mean-zero subspace contains no nonzero invariant vector. Property (T) therefore gives, for every $f \perp \mathbf{1}$,

$$\max_{s \in S} \|R_s f - f\|_2 \geq \kappa_S \|f\|_2.$$

At least one term in the sum of squared displacements is therefore at least $\kappa_S^2 \|f\|_2^2$. Applying Proposition 3.15 yields

$$\frac{1}{2|S|} \sum_{s \in S} \|R_s f - f\|_2^2 \geq \frac{\kappa_S^2}{2|S|} \|f\|_2^2,$$

and taking the infimum over mean-zero f proves the claimed uniform spectral gap. $\square$

Normality is therefore needed only if one wants the finite transitive Γ -set to carry a quotient-group structure and the graph to be a Cayley graph. The operator and Kazhdan arguments themselves require only a finite transitive action.

8. A compactness characterization of property (T)

For context, we record a standard equivalent formulation.

THEOREM 3.18 (Spectral isolation formulation). *For a finitely generated group Γ , the following are equivalent:*

- (i) Γ has property (T);
- (ii) for some finite generating set S there is $\gamma > 0$ such that, for every unitary representation π , the restriction of $A_{\pi, S}$ to the orthogonal complement of invariant vectors has spectrum contained in $[-1, 1 - \gamma]$.

PROOF. If property (T) holds with Kazhdan constant κ for S , Lemma 3.3 gives the upper spectral bound $1 - \kappa^2/(2|S|)$ on the orthogonal complement. Conversely, a uniform upper spectral bound gives a uniform displacement bound by the same lemma. The lower endpoint -1 is irrelevant to ordinary property (T); if desired it can be removed by adding laziness or the identity to the averaging measure. $\square$

9. Tensor products and why finite-dimensional representations matter

Suppose π is a nontrivial finite-dimensional representation of a quotient G . It appears inside the regular representation, so a spectral gap must contract it. Conversely, a single irreducible representation on which the averaging operator has norm close to 1 produces an almost invariant vector and destroys expansion.

This simple observation motivates two very different ways to prove expansion:

- (1) control the average $\frac{1}{|S|} \sum_s \pi(s)$ directly in every nontrivial irreducible representation;

- (2) avoid individual representation estimates, prove that the random-walk convolution powers become flat, and then use only coarse information about dimensions of irreducible representations.

The second route is the Bourgain–Gamburd route.

10. Comparison of spectral gaps under bounded words

The qualitative independence of the generating set can be strengthened quantitatively.

PROPOSITION 3.19 (Comparison under word substitution). *Let S, T be finite symmetric generating sets of a finite group G . Assume every $t \in T$ is a word of length at most L in S . Then the universally valid comparison*

$$\gamma_S \geq \frac{1}{L^2|S|} \gamma_T$$

holds. More precisely, for every mean-zero f ,

$$\frac{1}{|T|} \sum_{t \in T} \|R_t f - f\|_2^2 \leq L^2 \sum_{s \in S} \|R_s f - f\|_2^2 = L^2 |S| \left(\frac{1}{|S|} \sum_{s \in S} \|R_s f - f\|_2^2 \right).$$

PROOF. For $t = s_1 \cdots s_m$, telescoping and Cauchy–Schwarz give

$$\|R_t f - f\|_2^2 \leq m \sum_{j=1}^m \|R_{s_j} f - f\|_2^2 \leq L \sum_{j=1}^m \|R_{s_j} f - f\|_2^2.$$

Summing over t and bounding the total number of occurrences of any generator by the total word length $L|T|$ yields

$$\sum_t \|R_t f - f\|_2^2 \leq L^2 |T| \sum_s \|R_s f - f\|_2^2,$$

and hence, after division by $|T|$, gives

$$\frac{1}{|T|} \sum_t \|R_t f - f\|_2^2 \leq L^2 \sum_s \|R_s f - f\|_2^2 = L^2 |S| \left(\frac{1}{|S|} \sum_s \|R_s f - f\|_2^2 \right).$$

Thus

$$\gamma_T \leq L^2 |S| \gamma_S,$$

and therefore

$$(51) \quad \gamma_S \geq \frac{\gamma_T}{L^2 |S|}.$$

□

WARNING 3.20. This proof illustrates a useful normalization warning: when comparing generating sets, it is easy to insert an unjustified factor $|T|^{-1}$. A word-length bound controls total occurrences but not how evenly those occurrences are distributed among generators.

11. Expansion as a family property

For a fixed group G , every connected finite Cayley graph has positive gap. The substance is uniformity over a growing family.

DEFINITION 3.21 (Uniform quotient spectral gap). For finite quotients $q_i : \Gamma \rightarrow G_i$ and a fixed S , define

$$\gamma_{\mathcal{G}}(S) = \inf_i \gamma(\text{Cay}(G_i, q_i(S))).$$

The family expands exactly when $\gamma_{\mathcal{G}}(S) > 0$.

The dependence on S disappears at the qualitative level for fixed finite generating sets of Γ , by [Proposition 3.6](#) or [Proposition 3.19](#). The numerical value need not be stable.

12. From a Kazhdan constant to a spectral gap

The relation between almost invariant vectors and expansion is so important that it is worth deriving the numerical conversion once. Let $S = S^{-1}$ be a finite generating set of cardinality d , let π be a unitary representation, and define

$$P_S = \frac{1}{d} \sum_{s \in S} \pi(s).$$

For every vector v ,

$$\begin{aligned} \frac{1}{d} \sum_{s \in S} \|\pi(s)v - v\|^2 &= \frac{1}{d} \sum_{s \in S} (2\|v\|^2 - 2\Re\langle \pi(s)v, v \rangle) \\ &= 2\langle (I - P_S)v, v \rangle. \end{aligned}$$

The symmetry of S makes P_S self-adjoint, so the right-hand side is real.

Suppose now that S has Kazhdan constant at least κ on a class of representations: every unit vector orthogonal to invariant vectors satisfies

$$\max_{s \in S} \|\pi(s)v - v\| \geq \kappa.$$

At least one summand in the preceding average is then at least $\kappa^2\|v\|^2$, and hence

$$\langle (I - P_S)v, v \rangle \geq \frac{\kappa^2}{2d} \|v\|^2.$$

Thus the top of the nontrivial spectrum of P_S is at most

$$1 - \frac{\kappa^2}{2d}.$$

This elementary calculation is the quantitative bridge from a representation-theoretic almost-invariance statement to a graph-theoretic spectral gap.

REMARK 3.22 (Why the dependence on d appears). The definition above uses a maximum over generators, while the averaging operator uses an average. Passing from a maximum to an average costs the factor d . If one starts instead with an averaged Kazhdan inequality

$$\frac{1}{d} \sum_{s \in S} \|\pi(s)v - v\|^2 \geq \kappa_{\text{av}}^2 \|v\|^2,$$

then the spectral gap is directly at least $\kappa_{\text{av}}^2/2$.

EXAMPLE 3.23 (Finite quotients). Let Γ act by left translation on $\ell^2(\Gamma/N)$, where $N \triangleleft \Gamma$ has finite index. The invariant subspace consists exactly of the constants. A uniform Kazhdan inequality on the mean-zero subspace therefore gives a uniform Poincaré inequality for the Cayley graphs of the quotients Γ/N . This is the concrete mechanism behind property (τ) .

WARNING 3.24 (Generation is not a quantitative substitute). If S generates a finite group, then the only vectors fixed by every $\pi(s)$ in the regular representation are the constants. This says merely that the nontrivial eigenvalues are strictly below 1. It gives no uniform distance from 1 as the group varies. A family of cycles illustrates the gap between these assertions: two generators suffice for every cycle, yet the spectral gap tends to zero quadratically in the cycle length.

13. Exercises and conceptual checks

EXERCISE 3.25 (RMS displacement). Improve the constants in Lemma 3.3 when displacement is measured by the root-mean-square

$$\left(\frac{1}{|S|} \sum_s \|\pi(s)v - v\|^2 \right)^{1/2}$$

rather than the maximum.

EXERCISE 3.26 (Schreier formulation). Prove the analogue of Theorem 3.9 for a family of finite-index, not necessarily normal, subgroups using the permutation representations on $\ell^2(H \backslash \Gamma)$.

EXERCISE 3.27 (Adding the identity). Let $S' = S \sqcup \{e\}$ as a multiset. Express the quotient averaging operator for S' in terms of that for S and compute the new ordinary spectral gap exactly. Compare this with the lazy walk from Chapter 1.

EXERCISE 3.28 (Property (T) inheritance). Show directly from the definition that a quotient of a property- (T) group has property (T) . Explain why this is consistent with uniform expansion of all finite quotients.

EXERCISE 3.29 (Why abelian quotients obstruct expansion). Suppose a fixed finitely generated group has finite cyclic quotients of unbounded order in such a way

that the image of each fixed generator is represented by an integer of size $O(1)$ in the cyclic quotient. Construct a nontrivial character whose average under the generators tends to 1, and deduce failure of uniform expansion.

CHECKPOINT 3.30. The reader should be able to distinguish the following statements sharply:

- every quotient Cayley graph is connected;
- a specified family of quotient Cayley graphs has a uniform spectral gap;
- the group has property (τ) for that quotient family;
- the group has property (T) .

Only the middle two are equivalent. Property (T) is a substantially stronger statement about all unitary representations.

14. Chapter summary

A finite quotient Cayley graph is a concrete unitary representation: its averaging operator acts in the quotient regular representation. Uniform expansion is therefore uniform isolation of the trivial representation inside a family of quotient representations. This is property (τ) . Property (T) imposes the same phenomenon on all unitary representations and hence automatically implies expansion of every finite quotient family.

The Bourgain–Gamburd method should be viewed as a mechanism for proving property- (τ) -type conclusions without assuming property (T) . Its analytic input is not a priori control of all almost invariant vectors, but rapid flattening of a random walk in each finite quotient.

Part 2

Fourier Analysis and Quasirandomness

CHAPTER 4

Finite-Group Representation Theory and Fourier Analysis

The final step of the Bourgain–Gamburd method is representation-theoretic, but it uses remarkably little detailed character theory. What is essential is a precise Fourier decomposition of convolution and a lower bound on the dimensions of nontrivial irreducible representations. This chapter develops the necessary representation theory from first principles: unitarization, complete reducibility, Schur’s lemma, orthogonality of matrix coefficients, the regular representation, Fourier inversion, Plancherel, and block diagonalization of convolution operators. We keep the counting-normalization convention compatible with the probability measures of the preceding chapters.

GUIDING IDEA 4.1 (Fourier analysis as block diagonalization). The nonabelian Fourier transform is best understood not as a collection of formulas but as a change of coordinates for convolution. In an abelian group, every irreducible representation has dimension one, so convolution is diagonalized by characters. In a nonabelian group, the irreducible representations have higher dimension, and “diagonalization” becomes block diagonalization:

$$L^2(G) \simeq \bigoplus_{\rho \in \text{Irr}(G)} V_\rho \otimes \mathbb{C}^{d_\rho}, \quad T_\mu \simeq \bigoplus_{\rho} \widehat{\mu}(\rho) \otimes I_{d_\rho}.$$

Everything in the chapter is aimed at making this formula rigorous. The multiplicity factor d_ρ is not decorative: it is exactly what the Bourgain–Gamburd endgame later exploits. A large eigenvalue in a nontrivial block cannot occur once; it occurs at least d_ρ times in the regular representation, and quasirandomness makes d_ρ large.

1. Representations and invariant inner products

A complex representation of a finite group G is a homomorphism

$$\rho : G \rightarrow \text{GL}(V)$$

on a finite-dimensional complex vector space V . Its degree is $d_\rho = \dim V$.

The nonabelian Fourier endgame as a dependency chain

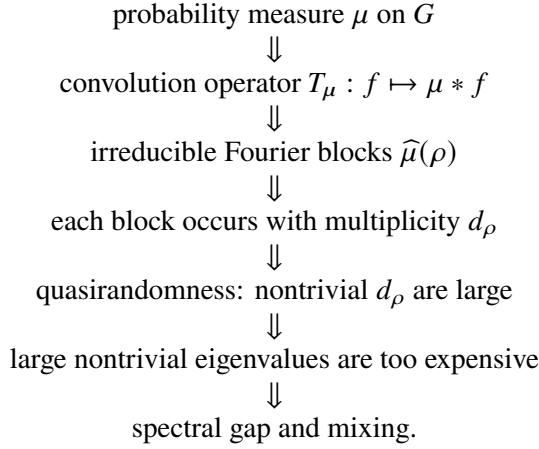


FIGURE 1. The representation-theoretic part of the Bourgain–Gamburd method. The chapter proves each arrow up to the final quasirandomness input, which is developed in the next chapter.

PROPOSITION 4.2 (Every finite-group representation is unitarizable). *Every finite-dimensional complex representation of a finite group admits a G -invariant positive-definite Hermitian inner product. Relative to an orthonormal basis for that inner product, every $\rho(g)$ is unitary.*

PROOF. Begin with any Hermitian inner product $\langle \cdot, \cdot \rangle_0$ and average it:

$$\langle v, w \rangle_G = \frac{1}{|G|} \sum_{g \in G} \langle \rho(g)v, \rho(g)w \rangle_0.$$

This remains positive definite. For $h \in G$,

$$\begin{aligned}
 \langle \rho(h)v, \rho(h)w \rangle_G &= \frac{1}{|G|} \sum_g \langle \rho(gh)v, \rho(gh)w \rangle_0 \\
 &= \langle v, w \rangle_G,
 \end{aligned}$$

using the bijection $g \mapsto gh$. □

2. Complete reducibility

A subspace $W \leq V$ is invariant if $\rho(g)W \subseteq W$ for every g . A nonzero representation is irreducible if its only invariant subspaces are 0 and V .

THEOREM 4.3 (Maschke). *Every finite-dimensional complex representation of a finite group is a direct sum of irreducible representations.*

PROOF. Use a G -invariant Hermitian inner product from [Proposition 4.2](#). If W is invariant, then its orthogonal complement $W^\perp$ is invariant because

$$\langle \rho(g)v, w \rangle = \langle v, \rho(g^{-1})w \rangle = 0$$

for $v \in W^\perp$ and $w \in W$. Thus $V = W \oplus W^\perp$. Induct on dimension. $\square$

3. Intertwiners and Schur's lemma

For representations $\rho : G \rightarrow \text{GL}(V)$ and $\sigma : G \rightarrow \text{GL}(W)$, an intertwiner is a linear map $T : V \rightarrow W$ satisfying

$$T\rho(g) = \sigma(g)T \quad (g \in G).$$

Write $\text{Hom}_G(V, W)$ for the space of intertwiners.

LEMMA 4.4 (Schur). *Let V, W be irreducible complex representations.*

- (i) *If V and W are inequivalent, then $\text{Hom}_G(V, W) = 0$.*
- (ii) *Every element of $\text{End}_G(V)$ is a scalar multiple of the identity.*

PROOF. For an intertwiner $T : V \rightarrow W$, both $\ker T$ and $\text{im } T$ are invariant. Irreducibility therefore makes a nonzero T an isomorphism, proving (i). For (ii), choose an eigenvalue λ of T , possible over $\mathbb{C}$. Then $T - \lambda I$ is a noninvertible intertwiner and hence must be zero. $\square$

PROPOSITION 4.5 (Representations of finite abelian groups). *Let A be a finite abelian group.*

- (i) *Every irreducible complex representation of A is one-dimensional.*
- (ii) *Every finite-dimensional complex representation of A is a direct sum of one-dimensional characters; after unitarization this may be taken to be an orthogonal direct sum of character spaces.*
- (iii) *If A is nontrivial, then A has a nontrivial one-dimensional character.*

PROOF. Let $\rho : A \rightarrow \text{GL}(V)$ be irreducible. Because A is abelian, for every $a \in A$ the operator $\rho(a)$ commutes with every $\rho(b)$. Thus $\rho(a) \in \text{End}_A(V)$, and Schur's lemma gives

$$\rho(a) = \chi(a)I_V$$

for a scalar $\chi(a) \in \mathbb{C}^\times$. If $\dim V > 1$, then every one-dimensional subspace of V would be invariant under all these scalar operators, contradicting irreducibility. Hence $\dim V = 1$, proving (i).

Part (ii) follows from complete reducibility, [Theorem 4.3](#): decompose the representation into irreducible summands and apply (i). If the representation is unitary, orthogonal complements of invariant subspaces are invariant, so the decomposition may be chosen orthogonal.

For (iii), consider the regular representation of A on $L^2(A)$. By complete reducibility it is a direct sum of irreducibles, hence by (i) of one-dimensional

characters. If every constituent were trivial, then every element of A would act as the identity on $L^2(A)$. But for $a \neq e$, right translation sends δ_e to $\delta_{a^{-1}} \neq \delta_e$. Therefore at least one constituent is a nontrivial character. $\square$

4. Averaging an operator

The most efficient route to orthogonality is to average conjugates of an arbitrary linear map.

LEMMA 4.6 (Averaging lemma). *Let ρ on V and σ on W be representations and let $A : V \rightarrow W$ be linear. Define*

$$\mathcal{A}(A) = \frac{1}{|G|} \sum_{g \in G} \sigma(g) A \rho(g)^{-1}.$$

Then $\mathcal{A}(A) \in \text{Hom}_G(V, W)$.

PROOF. For $h \in G$,

$$\begin{aligned} \sigma(h) \mathcal{A}(A) \rho(h)^{-1} &= \frac{1}{|G|} \sum_g \sigma(hg) A \rho(hg)^{-1} \\ &= \mathcal{A}(A). \end{aligned}$$

$\square$

If $\rho = \sigma$ is irreducible, Schur gives

$$\mathcal{A}(A) = cI.$$

Taking traces and using invariance of trace under conjugation gives

$$c = \frac{\text{tr } A}{d_\rho}.$$

Hence

$$(52) \quad \frac{1}{|G|} \sum_g \rho(g) A \rho(g)^{-1} = \frac{\text{tr } A}{d_\rho} I.$$

5. Orthogonality of matrix coefficients

Fix unitary irreducible representations ρ and σ , with orthonormal bases. Write $\rho_{ij}(g)$ for the matrix coefficients.

THEOREM 4.7 (Schur orthogonality). *If ρ and σ are irreducible unitary representations, then*

$$(53) \quad \frac{1}{|G|} \sum_{g \in G} \rho_{ij}(g) \overline{\sigma_{kl}(g)} = \begin{cases} 0, & \rho \neq \sigma, \\ \frac{1}{d_\rho} \delta_{ik} \delta_{jl}, & \rho \simeq \sigma \end{cases}$$

when equivalent representations have been identified in the chosen model.

PROOF. Let $E_{j\ell} : W \rightarrow V$ be the rank-one map sending the ℓ th basis vector to the j th and annihilating the others. Average

$$\frac{1}{|G|} \sum_g \rho(g) E_{j\ell} \sigma(g)^{-1}.$$

Its (i, k) matrix entry is the left side of Equation (53). If $\rho \neq \sigma$, the averaged map is zero by Schur. If $\rho = \sigma$, Equation (52) gives

$$\frac{\text{tr } E_{j\ell}}{d_\rho} I = \frac{\delta_{j\ell}}{d_\rho} I.$$

With indices aligned, this is exactly the stated Kronecker-delta formula. Equivalently one may take the rank-one map $E_{\ell j}$ depending on the row/column convention; direct calculation yields the displayed indexing. $\square$

COROLLARY 4.8 (Linear independence of inequivalent coefficient spaces). *The subspaces of $L^2(G)$ spanned by matrix coefficients of inequivalent irreducible representations are orthogonal. Within an irreducible representation, the functions*

$$\sqrt{d_\rho} \rho_{ij}$$

are orthonormal for normalized $L^2(G)$.

PROOF. Schur orthogonality gives zero inner product between coefficients belonging to inequivalent irreducibles. For a fixed irreducible ρ of dimension d_ρ , the same formula gives

$$\langle \rho_{ij}, \rho_{k\ell} \rangle_{L^2(G)} = \frac{1}{d_\rho} \delta_{ik} \delta_{j\ell}.$$

Multiplying every coefficient by $\sqrt{d_\rho}$ therefore gives an orthonormal family. $\square$

6. The regular representation

Let L_g act on functions by

$$(L_g f)(x) = f(g^{-1}x).$$

This is the left regular representation. Its character is especially simple:

$$\chi_{\text{reg}}(e) = |G|, \quad \chi_{\text{reg}}(g) = 0 \quad (g \neq e).$$

THEOREM 4.9 (Decomposition of the regular representation). *Choose one representative of each irreducible complex representation of G . Then*

$$(54) \quad \mathbb{C}[G] \simeq \bigoplus_{\rho \in \text{Irr}(G)} d_\rho V_\rho.$$

In particular,

$$(55) \quad \sum_{\rho \in \text{Irr}(G)} d_\rho^2 = |G|.$$

PROOF. Schur orthogonality gives d_ρ^2 orthogonal matrix coefficients for each ρ , so at least it shows that the sum of their dimensions does not exceed $|G|$. To prove completeness without invoking character theory as a black box, suppose $f \in L^2(G)$ is orthogonal to every matrix coefficient of every irreducible representation. Consider the left-invariant subspace

$$W = \text{span}\{L_g f : g \in G\}.$$

If $f \neq 0$, Maschke gives a nonzero irreducible subrepresentation $U \leq W$. Choose $0 \neq u \in U$. Evaluation at the identity is a linear functional on U ; translating u if necessary, it is nonzero. By Riesz representation, this functional is a matrix coefficient of the representation on U . Since u lies in the translate span of f , orthogonality of f to all translates of that coefficient contradicts nonzero evaluation. Hence the matrix coefficients span $L^2(G)$.

There are d_ρ^2 coefficients for each irreducible, so their total number is $|G|$, proving Equation (55). Under left translation, fixing the column index j and varying the row i gives a copy of V_ρ ; there are d_ρ column choices, yielding Equation (54). $\square$

REMARK 4.10. A more standard proof computes the multiplicity of ρ in the regular representation as the character inner product $\langle \chi_{\text{reg}}, \chi_\rho \rangle = d_\rho$. The proof above keeps matrix coefficients visible because they are exactly what Fourier analysis will use.

7. The nonabelian Fourier transform

For $f : G \rightarrow \mathbb{C}$ and an irreducible unitary representation ρ , define

$$(56) \quad \widehat{f}(\rho) = \sum_{x \in G} f(x) \rho(x) \in \text{End}(V_\rho).$$

This transform is matrix-valued. In an abelian group every irreducible representation is one-dimensional, recovering the usual scalar Fourier transform up to the convention of whether characters are conjugated.

THEOREM 4.11 (Fourier inversion). *For every $f : G \rightarrow \mathbb{C}$,*

$$(57) \quad f(x) = \frac{1}{|G|} \sum_{\rho \in \text{Irr}(G)} d_\rho \text{tr}(\widehat{f}(\rho) \rho(x^{-1})).$$

PROOF. Expand the right side:

$$\frac{1}{|G|} \sum_y f(y) \sum_\rho d_\rho \text{tr}(\rho(yx^{-1})).$$

The inner sum is the character of the regular representation evaluated at yx^{-1} by Theorem 4.9. It equals $|G|$ when $y = x$ and 0 otherwise. $\square$

THEOREM 4.12 (Plancherel). For $f, g : G \rightarrow \mathbb{C}$,

$$(58) \quad \sum_x f(x) \overline{g(x)} = \frac{1}{|G|} \sum_{\rho \in \text{Irr}(G)} d_\rho \text{tr}(\widehat{f}(\rho) \widehat{g}(\rho)^*).$$

In particular,

$$(59) \quad \|f\|_2^2 = \frac{1}{|G|} \sum_\rho d_\rho \|\widehat{f}(\rho)\|_{\text{HS}}^2.$$

PROOF. Apply Fourier inversion to g , conjugate, multiply by $f(x)$, and sum over x . The inner x -sum becomes $\widehat{f}(\rho)$ in the trace after cyclic permutation. Taking $g = f$ gives the norm identity. $\square$

8. Convolution becomes matrix multiplication

PROPOSITION 4.13 (Convolution theorem). With convolution as in Equation (34),

$$(60) \quad \widehat{f * g}(\rho) = \widehat{f}(\rho) \widehat{g}(\rho).$$

Moreover

$$\widehat{f^*}(\rho) = \widehat{f}(\rho)^*.$$

PROOF. For the first identity,

$$\begin{aligned} \widehat{f * g}(\rho) &= \sum_x \sum_y f(y) g(y^{-1}x) \rho(x) \\ &= \sum_{y,z} f(y) g(z) \rho(yz) \\ &= \left(\sum_y f(y) \rho(y) \right) \left(\sum_z g(z) \rho(z) \right). \end{aligned}$$

For the involution, substitute $y = x^{-1}$ and use unitarity $\rho(y^{-1}) = \rho(y)^*$. $\square$

COROLLARY 4.14 (Symmetric measures give Hermitian Fourier blocks). If a real probability measure satisfies $\mu(x) = \mu(x^{-1})$, then every matrix $\widehat{\mu}(\rho)$ is Hermitian and has operator norm at most 1.

PROOF. Symmetry says $\mu^* = \mu$, so the first claim follows from Proposition 4.13. The matrix is an average of unitary matrices $\rho(x)$ with nonnegative weights summing to one, hence its operator norm is at most one. $\square$

9. Convolution operators and block diagonalization

Let

$$T_\mu f = \mu * f.$$

The convolution theorem says that Fourier transform turns T_μ into left multiplication by $\widehat{\mu}(\rho)$ on the ρ -block.

PROPOSITION 4.15 (Spectrum of convolution). *Let μ be symmetric. The spectrum of T_μ on $\ell^2(G)$ is the union, with multiplicities, of the spectra of the Hermitian matrices $\widehat{\mu}(\rho)$, each eigenvalue in the ρ block occurring with multiplicity d_ρ . The trivial representation contributes the eigenvalue 1.*

PROOF. Under the decomposition of the regular representation, the ρ -isotypic component is naturally $V_\rho \otimes \mathbb{C}^{d_\rho}$. Left convolution commutes with right translations and acts as $\widehat{\mu}(\rho) \otimes I_{d_\rho}$ on this component. The multiplicity statement follows. $\square$

COROLLARY 4.16 (Spectral gap in Fourier language). *If $\text{supp } \mu$ generates G and μ is symmetric, then the ordinary spectral gap of the corresponding left-convolution Markov operator is*

$$1 - \max_{\rho \neq 1} \lambda_{\max}(\widehat{\mu}(\rho)).$$

The absolute spectral gap is

$$1 - \max_{\rho \neq 1} \|\widehat{\mu}(\rho)\|_{\text{op}}.$$

PROOF. By the preceding proposition, the spectrum on the ρ -isotypic component is the spectrum of $\widehat{\mu}(\rho)$, repeated d_ρ times. The trivial representation contributes the eigenvalue 1. Symmetry of μ makes every Fourier block self-adjoint. Hence the largest nontrivial eigenvalue gives the ordinary gap, while the largest absolute nontrivial eigenvalue, equivalently the maximum operator norm of the nontrivial blocks, gives the absolute gap. $\square$

10. Trace and Hilbert–Schmidt identities

The flattening endgame often avoids bounding every Fourier block. Instead it bounds a trace or Hilbert–Schmidt quantity and uses multiplicity.

PROPOSITION 4.17 (Hilbert–Schmidt norm of convolution). *Let T_μ be left convolution on counting $\ell^2(G)$. Then*

$$(61) \quad \|T_\mu\|_{\text{HS}}^2 = |G| \|\mu\|_2^2.$$

If μ is symmetric, then

$$(62) \quad \text{tr}(T_\mu^2) = |G| \|\mu\|_2^2.$$

PROOF. The matrix kernel of T_μ is $K(x, y) = \mu(xy^{-1})$ or the equivalent orientation dictated by the chosen basis. For each fixed y , the squared entries sum to $\|\mu\|_2^2$, and there are $|G|$ columns. This gives Equation (61). If μ is symmetric, T_μ is self-adjoint, so $\|T_\mu\|_{\text{HS}}^2 = \text{tr}(T_\mu^2)$. $\square$

COROLLARY 4.18 (Fourier form of the same identity). *For every μ ,*

$$|G| \|\mu\|_2^2 = \sum_{\rho} d_{\rho} \|\widehat{\mu}(\rho)\|_{\text{HS}}^2.$$

PROOF. Multiply Plancherel Equation (59) by $|G|$. $\square$

11. The trivial block and mean-zero functions

For the trivial representation 1_G ,

$$\widehat{f}(1_G) = \sum_x f(x).$$

Thus f has mean zero exactly when its trivial Fourier block vanishes.

For a probability measure,

$$\widehat{\mu}(1_G) = 1.$$

If u_G is uniform, then

$$\widehat{u_G}(1_G) = 1, \quad \widehat{u_G}(\rho) = 0 \quad (\rho \neq 1).$$

Therefore

$$\widehat{\mu - u_G}(1_G) = 0, \quad \widehat{\mu - u_G}(\rho) = \widehat{\mu}(\rho) \quad (\rho \neq 1).$$

This is the Fourier reason the distance to uniform is controlled entirely by nontrivial representations.

12. Abelian groups as a boundary case

If G is finite abelian, every irreducible representation has degree 1. Hence

$$\|f\|_2^2 = \frac{1}{|G|} \sum_{\chi \in \widehat{G}} |\widehat{f}(\chi)|^2,$$

and convolution is scalar multiplication on each character.

This makes clear why the Bourgain–Gamburd endgame needs nonabelian groups with large irreducible representations: in an abelian group there is no multiplicity gain from representation dimension. A single nontrivial character can carry an eigenvalue arbitrarily close to 1.

13. Worked example: the cycle

For $G = \mathbb{Z}/n\mathbb{Z}$ and

$$\mu = \frac{1}{2}(\delta_1 + \delta_{-1}),$$

the irreducible characters are

$$\chi_k(x) = e^{2\pi i k x / n}.$$

Our Fourier convention gives

$$\widehat{\mu}(\chi_k) = \frac{1}{2} \left(e^{2\pi i k / n} + e^{-2\pi i k / n} \right) = \cos(2\pi k / n).$$

Thus the spectral calculation from [Chapter 1](#) is simply the abelian Fourier transform of the step measure.

14. Exercises and mastery checks

EXERCISE 4.19 (Character orthogonality). Deduce from [Theorem 4.7](#) that irreducible characters are orthonormal in normalized $L^2(G)$ among class functions.

EXERCISE 4.20 (Central measures). Suppose μ is constant on conjugacy classes. Show by Schur's lemma that

$$\widehat{\mu}(\rho) = c_\rho I_{d_\rho}.$$

Find c_ρ in terms of χ_ρ and μ . Explain why central random walks are representation-theoretically much easier than arbitrary Cayley walks.

EXERCISE 4.21 (Fourier inversion from coefficient orthogonality). Prove [Theorem 4.11](#) directly by expanding f in the orthonormal basis of matrix coefficients, without using the regular character.

EXERCISE 4.22 (Operator norm versus Fourier blocks). Show for an arbitrary, not necessarily symmetric, probability measure that

$$\|T_\mu\|_{L_0^2 \rightarrow L_0^2} = \max_{\rho \neq 1} \|\widehat{\mu}(\rho)\|_{\text{op}}.$$

Interpret this as a singular-value statement when T_μ is not self-adjoint.

EXERCISE 4.23 (Return probability through Fourier analysis). For symmetric μ , prove

$$\mu^{(2n)}(e) = \frac{1}{|G|} \sum_{\rho} d_\rho \text{tr}(\widehat{\mu}(\rho)^{2n}).$$

Separate the trivial contribution and explain how a spectral gap forces the remainder to decay.

CHECKPOINT 4.24. At this point the reader should be able to translate any convolution estimate into matrix-valued Fourier language, while also knowing when doing so is strategically unhelpful. The essential identities are

$$\widehat{f * g}(\rho) = \widehat{f}(\rho)\widehat{g}(\rho), \quad \|f\|_2^2 = \frac{1}{|G|} \sum_{\rho} d_\rho \|\widehat{f}(\rho)\|_{\text{HS}}^2.$$

The dimension factors d_ρ are the source of quasirandom mixing.

15. Chapter summary

Finite-group Fourier analysis is the decomposition of the regular representation into irreducibles written in convolution coordinates. Convolution becomes matrix multiplication, symmetry becomes Hermitianity of every Fourier block, and the trivial representation is exactly the constant part. The sum-of-squares identity $\sum d_\rho^2 = |G|$ and the multiplicity d_ρ of each irreducible in the regular representation will be the quantitative leverage behind the quasirandom endgame.

CHAPTER 5

Quasirandom Groups and Product Mixing

A finite group is quasirandom when all of its nontrivial complex representations have large dimension. This apparently representation-theoretic condition has strong combinatorial consequences: dense sets multiply almost as if they were independent, product-free sets must be small, and sufficiently large sets fill the group after a bounded number of products. For the Bourgain–Gamburd method, quasirandomness supplies the final amplification step: once a random walk has been flattened enough, high representation multiplicity makes an eigenvalue near 1 impossible. We prove the basic Gowers mixing inequality and an elementary quasirandomness bound for $\mathrm{SL}_2(\mathbb{F}_p)$.

GUIDING IDEA 5.1 (Why representation dimension creates mixing). Quasirandomness does not say that a particular generating set mixes. It says that *if* a nonconstant Fourier mode survives, then that mode must occupy a large matrix block. Plancherel then makes such survival expensive. The basic inequality in this chapter is therefore a dimension-counting principle disguised as an L^2 estimate.

A useful mental model is this. In an abelian group, a single one-dimensional character can carry a large Fourier coefficient at very little Plancherel cost. In a D -quasirandom group, every nontrivial frequency has at least D^2 matrix coefficients and appears with multiplicity at least D in the regular representation. The factor $D^{-1/2}$ in product mixing is the analytic shadow of that representation-theoretic bulk. This is why quasirandomness is the *last* input in the Bourgain–Gamburd machine, not the first: before it becomes effective, the walk must already have been flattened to a scale at which this dimensional penalty dominates.

1. Definition and first examples

DEFINITION 5.2 (D -quasirandom group). A finite group G is D -quasirandom if every nontrivial irreducible complex representation of G has dimension at least D .

This terminology was introduced by Gowers [1]. The parameter D measures how difficult it is for G to act linearly in low dimension.

EXAMPLE 5.3 (Abelian groups). By [Proposition 4.5](#), every irreducible complex representation of a finite abelian group is one-dimensional, and a nontrivial finite

abelian group has a nontrivial character. Thus a nontrivial abelian group is exactly 1-quasirandom and never D -quasirandom for $D > 1$.

EXAMPLE 5.4 (A quotient obstruction). If G has a nontrivial quotient Q with a nontrivial irreducible representation of degree d , then composing with the quotient map gives a nontrivial representation of G of degree d . Hence

$$D(G) \leq D(Q),$$

where $D(G)$ denotes the least degree of a nontrivial irreducible representation. In particular, any nontrivial abelian quotient forces $D(G) = 1$.

2. The basic quasirandom convolution inequality

The following estimate is one of the cleanest points in the entire method.

THEOREM 5.5 (Gowers convolution inequality). *Let G be a finite D -quasirandom group of order N . If $f : G \rightarrow \mathbb{C}$ has mean zero, then for every $g : G \rightarrow \mathbb{C}$,*

$$(63) \quad \|f * g\|_2 \leq \sqrt{\frac{N}{D}} \|f\|_2 \|g\|_2.$$

The same conclusion holds if g rather than f has mean zero.

PROOF. Because $\widehat{f}(1_G) = 0$, Plancherel and the convolution theorem give

$$\begin{aligned} \|f * g\|_2^2 &= \frac{1}{N} \sum_{\rho \neq 1} d_\rho \|\widehat{f}(\rho) \widehat{g}(\rho)\|_{\text{HS}}^2 \\ &\leq \frac{1}{N} \sum_{\rho \neq 1} d_\rho \|\widehat{f}(\rho)\|_{\text{HS}}^2 \|\widehat{g}(\rho)\|_{\text{op}}^2. \end{aligned}$$

For each nontrivial ρ , Plancherel applied to g implies

$$d_\rho \|\widehat{g}(\rho)\|_{\text{HS}}^2 \leq N \|g\|_2^2,$$

so

$$\|\widehat{g}(\rho)\|_{\text{op}}^2 \leq \|\widehat{g}(\rho)\|_{\text{HS}}^2 \leq \frac{N}{d_\rho} \|g\|_2^2 \leq \frac{N}{D} \|g\|_2^2.$$

Substituting gives

$$\begin{aligned} \|f * g\|_2^2 &\leq \frac{N}{D} \|g\|_2^2 \left(\frac{1}{N} \sum_{\rho \neq 1} d_\rho \|\widehat{f}(\rho)\|_{\text{HS}}^2 \right) \\ &= \frac{N}{D} \|f\|_2^2 \|g\|_2^2. \end{aligned}$$

The other-sided version follows either symmetrically or by involution. □

REMARK 5.6 (Normalized form). If one uses normalized convolution

$$(f \star g)(x) = \frac{1}{N} \sum_y f(y)g(y^{-1}x)$$

and normalized L^2 , then the same theorem reads

$$\|f \star g\|_{2,G} \leq D^{-1/2} \|f\|_{2,G} \|g\|_{2,G}$$

for mean-zero f . The counting form Equation (63) is more compatible with probability measures.

3. Product mixing for three sets

For $A, B, C \subseteq G$, let

$$N(A, B, C) = |\{(a, b, c) \in A \times B \times C : ab = c\}|.$$

Equivalently,

$$N(A, B, C) = \langle 1_A * 1_B, 1_C \rangle.$$

THEOREM 5.7 (Product mixing). *If G is D -quasirandom of order N , then*

$$(64) \quad \left| N(A, B, C) - \frac{|A||B||C|}{N} \right| \leq \sqrt{\frac{N}{D}} (|A||B||C|)^{1/2}.$$

Consequently, if

$$(65) \quad |A||B||C| > \frac{N^3}{D},$$

then there is at least one solution $ab = c$ with $a \in A, b \in B, c \in C$.

PROOF. Write

$$f = 1_A - \frac{|A|}{N} 1_G.$$

Then f has mean zero and

$$1_A * 1_B = \frac{|A||B|}{N} 1_G + f * 1_B.$$

Therefore the error is

$$|\langle f * 1_B, 1_C \rangle| \leq \|f * 1_B\|_2 \|1_C\|_2.$$

By Theorem 5.5,

$$\|f * 1_B\|_2 \leq \sqrt{\frac{N}{D}} \|f\|_2 |B|^{1/2}.$$

Since

$$\|f\|_2^2 = |A| - \frac{|A|^2}{N} \leq |A|,$$

we obtain Equation (64). If the main term exceeds the error, a solution exists. Squaring the inequality

$$\frac{|A||B||C|}{N} > \sqrt{\frac{N}{D}} (|A||B||C|)^{1/2}$$

gives exactly Equation (65). $\square$

COROLLARY 5.8 (Product-free sets). *If $A \subseteq G$ is product-free, meaning there are no $a, b, c \in A$ with $ab = c$, then*

$$|A| \leq \frac{N}{D^{1/3}}.$$

PROOF. Apply Theorem 5.7 with $A = B = C$. $\square$

COROLLARY 5.9 (Large sets have full triple product). *If*

$$|A| > \frac{N}{D^{1/3}},$$

then

$$A^3 = G.$$

PROOF. Fix $x \in G$ and apply Theorem 5.7 to A, A, xA^{-1} . The three sets have equal cardinality, so there are $a_1, a_2, a_3 \in A$ with

$$a_1 a_2 = x a_3^{-1}.$$

Thus $x = a_1 a_2 a_3$. $\square$

4. A mixing inequality for probability measures

Let $u = u_G$ be uniform. If μ is a probability measure, then $\mu - u$ has mean zero.

PROPOSITION 5.10 (Quasirandom smoothing). *If G is D -quasirandom, then for probability measures μ, ν ,*

$$(66) \quad \|\mu * \nu - u\|_2 \leq \sqrt{\frac{N}{D}} \|\mu - u\|_2 \|\nu - u\|_2.$$

In particular,

$$\|\mu * \nu - u\|_2 \leq \sqrt{\frac{N}{D}} \|\mu\|_2 \|\nu\|_2.$$

PROOF. Since $u * \nu = \mu * u = u$,

$$\mu * \nu - u = (\mu - u) * (\nu - u).$$

Apply Theorem 5.5. The second estimate follows from

$$\|\mu - u\|_2^2 = \|\mu\|_2^2 - N^{-1} \leq \|\mu\|_2^2.$$

$\square$

COROLLARY 5.11 (Flatness threshold for immediate mixing). *Suppose G is D -quasirandom and*

$$\|\mu\|_2 \leq N^{-1/2} D^{1/4} \eta^{1/2}.$$

Then

$$\|\mu * \mu - u\|_2 \leq \eta N^{-1/2}.$$

Equivalently, the density $N(\mu * \mu)$ is within η of 1 in normalized L^2 .

PROOF. Use [Proposition 5.10](#):

$$\|\mu * \mu - u\|_2 \leq \sqrt{N/D} \|\mu\|_2^2 \leq \eta N^{-1/2}.$$

□

5. Multiplicity amplification

There is another route from flatness to a spectral bound that is especially close to the original Bourgain–Gamburd argument. If an eigenvalue occurs in a high-dimensional irreducible block, it occurs many times in the regular representation. A large eigenvalue therefore forces a large trace of a high convolution power.

PROPOSITION 5.12 (Eigenvalue multiplicity bound). *Let G be D -quasirandom and μ a symmetric probability measure. Every eigenvalue of T_μ on $L_0^2(G)$ occurs with multiplicity at least D .*

PROOF. By [Proposition 4.15](#), an eigenvalue arising from the irreducible block ρ is repeated d_ρ times in the regular representation. Nonconstant spectrum comes only from nontrivial ρ , for which $d_\rho \geq D$. □

THEOREM 5.13 (Trace-to-gap principle). *Let μ be symmetric on a D -quasirandom group G of order N . Let*

$$\lambda = \max\{\lambda_j(T_\mu) : \lambda_j \neq 1\}$$

be the largest nontrivial eigenvalue. For every integer $n \geq 1$,

$$(67) \quad D\lambda^{2n} \leq N\|\mu^{(n)}\|_2^2 - 1.$$

Consequently, if

$$\|\mu^{(n)}\|_2^2 \leq N^{-1} + DN^{-1}e^{-2cn},$$

then $\lambda \leq e^{-c}$.

PROOF. The operator $T_\mu^{2n} = T_{\mu^{(2n)}}$ is positive semidefinite because T_μ is self-adjoint and the exponent is even. Its trace is

$$\mathrm{tr}(T_\mu^{2n}) = N\mu^{(2n)}(e) = N\|\mu^{(n)}\|_2^2.$$

The trivial eigenvalue contributes 1. If λ is a nontrivial eigenvalue, its multiplicity is at least D , so

$$1 + D\lambda^{2n} \leq \mathrm{tr}(T_\mu^{2n}),$$

proving Equation (67). The consequence follows by division by D and taking $2n$ th roots. $\square$

REMARK 5.14. The precise flatness threshold needed in an expansion proof depends on how D grows with N . If $D \geq N^\beta$ for fixed $\beta > 0$, then a power saving beyond the uniform scale can be converted into a constant spectral gap at logarithmic time. This is why polynomial quasirandomness is so effective.

6. Elementary quasirandomness of $\mathrm{SL}_2(\mathbb{F}_p)$

We now prove the rank-one estimate that is sufficient for the classical expansion argument.

Let p be an odd prime and

$$G = \mathrm{SL}_2(\mathbb{F}_p).$$

Define the upper unipotent subgroup

$$U = \left\{ u(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in \mathbb{F}_p \right\}.$$

Then $U \simeq (\mathbb{F}_p, +)$. Let

$$a(t) = \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix}, \quad t \in \mathbb{F}_p^\times.$$

A direct calculation gives

$$(68) \quad a(t)u(x)a(t)^{-1} = u(t^2x).$$

We shall use, and now prove, the elementary generation fact that prevents the trivial U -weight from being the only weight of a nontrivial irreducible representation.

LEMMA 5.15 (Upper and lower unipotents generate SL_2). *Let F be a field. Put*

$$u(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \quad \ell(y) = \begin{pmatrix} 1 & 0 \\ y & 1 \end{pmatrix}.$$

Then the subgroup of $\mathrm{SL}_2(F)$ generated by all $u(x)$ and $\ell(y)$ is the whole group.

PROOF. Let

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(F).$$

If $a \neq 0$, left multiplication by $\ell(-c/a)$ clears the lower-left entry:

$$\ell(-c/a)g = \begin{pmatrix} a & b \\ 0 & d - cb/a \end{pmatrix}.$$

The determinant is 1, so the lower-right entry is a^{-1} . Right multiplication by $u(-b/a)$ then clears the upper-right entry and gives $\mathrm{diag}(a, a^{-1})$. It therefore suffices to express diagonal matrices by unipotents. A direct multiplication gives, for $t \in F^\times$,

$$(69) \quad \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} = u(t-1) \ell(1) u(t^{-1}-1) \ell(-t).$$

Indeed the first three factors equal $\begin{pmatrix} t & 0 \\ 1 & t^{-1} \end{pmatrix}$, and the last factor clears the lower-left entry.

If $a = 0$, then $c \neq 0$ because $\det g = 1$. Multiplying on the left by $u(1)$ changes the upper-left entry to $c \neq 0$, so the preceding argument applies to $u(1)g$. Since $u(1)$ itself is one of the generators, it applies to g as well. $\square$

REMARK 5.16. The proof is Gaussian elimination carried out inside SL_2 : every elementary operation is represented by multiplication by one of the determinant-one unipotent generators. We included the calculation because the quasirandomness proof uses normal generation by U ; leaving the generation assertion as an exercise would hide a logical step in the first group-specific theorem of the series.

LEMMA 5.17 (Order of $\mathrm{SL}_2(\mathbb{F}_p)$). *For every prime p ,*

$$|\mathrm{SL}_2(\mathbb{F}_p)| = p(p^2 - 1).$$

PROOF. A matrix in $\mathrm{SL}_2(\mathbb{F}_p)$ is determined by its two column vectors

$$v = \begin{pmatrix} a \\ c \end{pmatrix}, \quad w = \begin{pmatrix} b \\ d \end{pmatrix},$$

subject to $\det(v, w) = ad - bc = 1$. The first column v may be any nonzero vector, giving $p^2 - 1$ choices. For a fixed nonzero $v = (a, c)$, the map

$$L_v : \mathbb{F}_p^2 \rightarrow \mathbb{F}_p, \quad L_v(b, d) = ad - bc,$$

is a nonzero linear functional. Hence it is surjective and every fiber has cardinality p : its kernel is a one-dimensional subspace of the two-dimensional vector space $\mathbb{F}_p^2$. In particular the equation $L_v(b, d) = 1$ has exactly p solutions. Multiplying the numbers of choices gives

$$|\mathrm{SL}_2(\mathbb{F}_p)| = (p^2 - 1)p.$$

$\square$

THEOREM 5.18 (Rank-one quasirandomness). *Let p be an odd prime. Every nontrivial irreducible complex representation of $\mathrm{SL}_2(\mathbb{F}_p)$ has dimension at least*

$$\frac{p-1}{2}.$$

Hence $\mathrm{SL}_2(\mathbb{F}_p)$ is $(p-1)/2$ -quasirandom.

PROOF. Let $\rho : G \rightarrow U(V)$ be nontrivial and irreducible. Restrict ρ to the abelian group U . By [Proposition 4.5](#), V decomposes into U -weight spaces

$$V = \bigoplus_{r \in \mathbb{F}_p} V_r,$$

where

$$\rho(u(x))v = e^{2\pi i r x/p} v \quad (v \in V_r),$$

after identifying $\mathbb{F}_p$ with residues modulo p for the additive characters.

At least one nonzero weight $r \neq 0$ occurs. Indeed, if only V_0 occurred, then U would act trivially. The kernel of ρ is normal and would contain U . The conjugates of U contain the lower unipotent subgroup as well (conjugate by $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$), and upper and lower unipotents generate $\mathrm{SL}_2(\mathbb{F}_p)$ by [Lemma 5.15](#). Hence the normal closure of U is all of G , so ρ would be trivial, a contradiction.

Choose $0 \neq v \in V_r$ with $r \neq 0$. From [Equation \(68\)](#), the vector $\rho(a(t))v$ has weight rt^{-2} (the inverse square rather than square depends only on whether one moves $a(t)$ to the other side; the orbit is the same). As t varies over $\mathbb{F}_p^\times$, the values rt^{-2} run through the coset $r(\mathbb{F}_p^\times)^2$ of the subgroup of nonzero squares. This orbit has cardinality $(p-1)/2$.

Distinct characters of U have orthogonal weight spaces in a unitary representation. Therefore V contains at least $(p-1)/2$ nonzero mutually orthogonal weight spaces, and

$$\dim V \geq \frac{p-1}{2}.$$

□

REMARK 5.19 (Why the proof is useful). The argument reveals the mechanism rather than just the numerical bound. A large abelian subgroup U supplies many characters, while another subgroup acts on those characters with large orbits. Any nontrivial representation containing one nontrivial U -weight is forced to contain the whole orbit. Similar weight-orbit ideas recur throughout representation theory of groups of Lie type, although sharp lower bounds in higher rank require deeper theory.

REMARK 5.20 (Order comparison). By [Lemma 5.17](#), for $G = \mathrm{SL}_2(\mathbb{F}_p)$,

$$|G| = p(p^2 - 1) \asymp p^3.$$

Thus [Theorem 5.18](#) gives

$$D(G) \gg |G|^{1/3}.$$

This polynomial relation is exactly the scale needed for the multiplicity amplification in the expansion method.

7. Quasirandomness is not expansion

It is essential not to confuse a property of the group with a property of a generating measure.

WARNING 5.21. A large lower bound for dimensions of nontrivial representations does not by itself show that a given Cayley graph expands. Quasirandomness says that an offending Fourier block, if it exists, is large. It does not prevent the averaging matrix $\widehat{\mu}(\rho)$ from having an eigenvalue near 1. The random-walk/flattening part of the Bourgain–Gamburd method is what rules out such an eigenvalue.

This separation of roles is worth remembering:

quasirandomness amplifies a spectral obstruction

but

flattening eliminates the obstruction.

8. Worked thresholds: when does quasirandomness actually become useful?

The statements above are cleanest when one translates them into exponents. Write

$$|G| = N, \quad D(G) \geq N^\beta.$$

Suppose a probability measure has reached the scale

$$\|\mu\|_2 \leq N^{-1/2+\theta}.$$

Then [Proposition 5.10](#) gives

$$\|\mu * \mu - u\|_2 \leq N^{1/2-\beta/2} N^{-1+2\theta} = N^{-1/2-\beta/2+2\theta}.$$

Thus a single further convolution is already smaller than the uniform L^2 scale by a power of N as soon as

$$2\theta < \frac{\beta}{2}, \quad \text{equivalently} \quad \theta < \frac{\beta}{4}.$$

This explains the recurrent target $N^{-1/2+\beta/4}$ in abstract formulations of the machine. The exact fraction $1/4$ is not sacred; what matters is leaving fixed positive slack below the threshold $\beta/4$.

EXAMPLE 5.22 (The rank-one scale). For $G = \mathrm{SL}_2(\mathbb{F}_p)$, [Lemma 5.17](#) and [Theorem 5.18](#) give

$$|G| = p(p^2 - 1) \asymp p^3, \quad D(G) \geq \frac{p-1}{2} \asymp p.$$

Hence one may think of $D(G)$ as roughly $N^{1/3}$. In exponent language, β is therefore approximately $1/3$. The crude threshold $\theta < \beta/4$ becomes

$$\theta < \frac{1}{12}.$$

So the flattening part of the proof does not need to drive the walk all the way to the uniform norm $N^{-1/2}$; reaching, say, $N^{-5/12}$ already enters a range where quasirandomness can finish the argument after a bounded amount of additional smoothing.

EXAMPLE 5.23 (The three-set threshold). The product-mixing condition

$$|A||B||C| > \frac{N^3}{D}$$

becomes particularly transparent when the three sets have the same size M . Then it is enough that

$$M > ND^{-1/3}.$$

If $D \asymp N^\beta$, this is

$$M > N^{1-\beta/3}.$$

For $\mathrm{SL}_2(\mathbb{F}_p)$, where $\beta \approx 1/3$, the threshold is of order $N^{8/9} \asymp p^{8/3}$. This is the characteristic exponent appearing in the original Gowers product-mixing application to rank-one groups.

GUIDING IDEA 5.24 (Why this is only an endgame). The threshold $N^{1-\beta/3}$ is far too large to prove growth for a medium set. Product mixing says that *already dense* sets cover the group rapidly. Helfgott growth and the flattening mechanism are what move the problem from sparse or medium scales into this dense regime. Quasirandomness then turns density into global mixing.

9. Exercises and mastery checks

EXERCISE 5.25 (Gowers inequality with normalized norms). Rewrite Theorem 5.5 entirely in normalized $L^2(G)$ and normalized convolution. Verify every factor of N .

EXERCISE 5.26 (Two-set product size). Let $A, B \subseteq G$ in a D -quasirandom group. Use product mixing to obtain a lower bound for $|AB|$ in terms of $|A|, |B|, N, D$. Compare the bound with the trivial inequality $|AB| \geq \max(|A|, |B|)$.

EXERCISE 5.27 (Triple products). Prove Corollary 5.9 directly from the convolution inequality, keeping track of strict versus non-strict inequalities at the threshold.

EXERCISE 5.28 (Quasirandom quotient obstruction). Suppose G is D -quasirandom with $D > 1$. Prove that G is perfect, i.e. $G = [G, G]$. Hint: any nontrivial abelianization has a one-dimensional character.

EXERCISE 5.29 (Elementary matrices generate SL_2). Supply the generation step used in Theorem 5.18: show using row operations that the upper and lower unipotent subgroups generate $\mathrm{SL}_2(\mathbb{F}_p)$.

EXERCISE 5.30 (A trace endgame). Suppose $|G| = N$, $D \geq N^\beta$, and for some $n \leq C \log N$ one has

$$\|\mu^{(n)}\|_2^2 \leq N^{-1} + N^{-1-\eta}$$

for fixed $\beta, \eta > 0$. Use Theorem 5.13 to determine for which relation between η and β this estimate alone yields a uniform positive spectral gap.

CHECKPOINT 5.31. The reader should be able to prove the product-mixing estimate from Plancherel in a few lines and to explain the conceptual role of each factor:

$$\sqrt{\frac{|G|}{D}}$$

is the penalty for converting a Hilbert–Schmidt bound into an operator-norm bound on a nontrivial Fourier block. Large D makes this penalty small enough to force random-like multiplication for dense sets.

10. Chapter summary

A D -quasirandom group has no nontrivial representation below dimension D . Fourier analysis turns this into the fundamental inequality

$$\|f * g\|_2 \leq \sqrt{|G|/D} \|f\|_2 \|g\|_2$$

when one factor has mean zero. The resulting product-mixing theorem shows that sets with $|A||B||C| > |G|^3/D$ cannot avoid solutions to $ab = c$.

For expansion, the same dimension lower bound appears as an eigenvalue multiplicity lower bound in the regular representation. The elementary weight-orbit argument for $\mathrm{SL}_2(\mathbb{F}_p)$ gives $D \geq (p-1)/2 \asymp |G|^{1/3}$. This supplies the representation-theoretic endgame, but not the flattening needed to reach it.

Part 3

Technical Foundations

APPENDIX A

Finite-Dimensional Operator Theory for Expansion

Spectral-gap arguments on finite groups are finite-dimensional Hilbert-space arguments, but several normalizations recur so often that it is useful to place them in one self-contained reference. We prove the spectral theorem in the form needed for averaging operators, the Rayleigh–Ritz variational principle, projection onto invariant vectors, trace and Hilbert–Schmidt identities, the relation between eigenvalues and powers, and the singular-value substitute for nonsymmetric walks. The point is not functional-analytic generality; it is to make every operator-theoretic transition in the expansion machine explicit and checkable.

1. Finite-dimensional Hilbert spaces

Throughout this appendix, V is a finite-dimensional complex inner-product space. We write the inner product linearly in the first variable. For a linear map $T : V \rightarrow V$, the adjoint T^* is characterized by

$$\langle Tv, w \rangle = \langle v, T^*w \rangle.$$

The operator is *self-adjoint* if $T = T^*$ and *unitary* if $T^*T = TT^* = I$.

LEMMA A.1 (Orthogonality of distinct eigenspaces). *If $T = T^*$ and $Tv = \lambda v$, $Tw = \mu w$ with $\lambda \neq \mu$, then $v \perp w$.*

PROOF. We have

$$\lambda \langle v, w \rangle = \langle Tv, w \rangle = \langle v, Tw \rangle = \bar{\mu} \langle v, w \rangle.$$

The eigenvalues of a self-adjoint map are real, since

$$\lambda \|v\|^2 = \langle Tv, v \rangle = \overline{\langle Tv, v \rangle} = \bar{\lambda} \|v\|^2.$$

Thus $(\lambda - \mu) \langle v, w \rangle = 0$. □

THEOREM A.2 (Finite-dimensional spectral theorem). *Every self-adjoint operator T has an orthonormal basis of eigenvectors. In particular, all eigenvalues are real and*

$$\|T\|_{\text{op}} = \max_{\lambda \in \text{Spec}(T)} |\lambda|.$$

PROOF. The characteristic polynomial has a complex root λ , so T has a unit eigenvector v . By the preceding calculation, $\lambda \in \mathbb{R}$. If $w \perp v$, then

$$\langle Tw, v \rangle = \langle w, Tv \rangle = \lambda \langle w, v \rangle = 0,$$

so $v^\perp$ is T -invariant. Induction on $\dim V$ gives an orthonormal eigenbasis. In such a basis, if $x = \sum_i x_i e_i$ and $Te_i = \lambda_i e_i$, then

$$\|Tx\|^2 = \sum_i |\lambda_i|^2 |x_i|^2 \leq \left(\max_i |\lambda_i|^2 \right) \|x\|^2,$$

and equality is attained on an eigenvector of maximal absolute eigenvalue. $\square$

2. Rayleigh quotients and the spectral gap

For $v \neq 0$ define the Rayleigh quotient

$$\mathcal{R}_T(v) = \frac{\langle Tv, v \rangle}{\|v\|^2}.$$

PROPOSITION A.3 (Rayleigh–Ritz). *Let $T = T^*$ have eigenvalues*

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m.$$

Then

$$\lambda_1 = \max_{v \neq 0} \mathcal{R}_T(v), \quad \lambda_m = \min_{v \neq 0} \mathcal{R}_T(v).$$

If e_1 is a unit eigenvector for λ_1 , then

$$\lambda_2 = \max_{0 \neq v \perp e_1} \mathcal{R}_T(v)$$

provided the top eigenspace is one-dimensional.

PROOF. Expand $v = \sum_i c_i e_i$ in an orthonormal eigenbasis. Then

$$\mathcal{R}_T(v) = \frac{\sum_i \lambda_i |c_i|^2}{\sum_i |c_i|^2},$$

a convex combination of the eigenvalues. This lies between the smallest and largest eigenvalues. If $v \perp e_1$, then $c_1 = 0$, and the same argument gives the final assertion. $\square$

COROLLARY A.4 (Poincaré form of a gap). *Suppose $P = P^*$, $P\mathbf{1} = \mathbf{1}$, and the constants form the entire 1-eigenspace. Then the ordinary spectral gap*

$$\gamma = 1 - \lambda_2(P)$$

is the largest constant such that

$$\langle (I - P)f, f \rangle \geq \gamma \|f\|^2$$

for every $f \perp \mathbf{1}$.

PROOF. On $\mathbf{1}^\perp$, the largest Rayleigh quotient of P is λ_2 . Therefore

$$\frac{\langle (I - P)f, f \rangle}{\|f\|^2} = 1 - \mathcal{R}_P(f) \geq 1 - \lambda_2.$$

An eigenvector for λ_2 shows sharpness. $\square$

REMARK A.5 (Ordinary versus absolute gap). The Poincaré inequality sees only the top of the nontrivial spectrum. The contraction estimate

$$\|P^n f\| \leq (1 - \gamma)^n \|f\| \quad (f \perp \mathbf{1})$$

requires the *absolute* gap

$$1 - \max_{i \geq 2} |\lambda_i|.$$

A large negative eigenvalue does not damage the Poincaré inequality but does damage non-lazy mixing. This is the operator-theoretic reason for the laziness convention in Chapter 1.

3. Averaging operators are contractions

LEMMA A.6 (Convex averages of unitaries). *If $U_1, \dots, U_r$ are unitary and $a_i \geq 0$ with $\sum_i a_i = 1$, then*

$$T = \sum_i a_i U_i$$

satisfies $\|T\|_{\text{op}} \leq 1$.

PROOF. For every v ,

$$\|Tv\| \leq \sum_i a_i \|U_i v\| = \|v\|.$$

$\square$

LEMMA A.7 (Equality in a unitary average). *With the notation above, suppose all $a_i > 0$. If $\|Tv\| = \|v\|$ for some $v \neq 0$, then*

$$U_1 v = U_2 v = \dots = U_r v.$$

If one of the U_i is the identity, then every U_i fixes v .

PROOF. The triangle inequality gave

$$\left\| \sum_i a_i U_i v \right\| \leq \sum_i a_i \|U_i v\| = \|v\|.$$

Equality in the triangle inequality in a Hilbert space forces the nonzero vectors $U_i v$ to point in the same direction. Since they also have equal norm, they are equal. If one is v , all are v . $\square$

GUIDING IDEA A.8. This exact equality statement is the linear shadow of the flattening philosophy. Exact failure to contract means genuine invariance. Near-failure to contract means approximate invariance, and the Bourgain–Gamburd method converts that analytic near-invariance into approximate algebraic structure.

4. Projection onto invariant vectors

Let $\pi : G \rightarrow U(V)$ be a unitary representation of a finite group.

PROPOSITION A.9 (Averaging projection). *The operator*

$$\Pi_G = \frac{1}{|G|} \sum_{g \in G} \pi(g)$$

is the orthogonal projection onto

$$V^G = \{v : \pi(g)v = v \text{ for all } g \in G\}.$$

PROOF. For $h \in G$,

$$\pi(h)\Pi_G = \frac{1}{|G|} \sum_g \pi(hg) = \Pi_G,$$

so the image lies in V^G . Conversely, $\Pi_G v = v$ for $v \in V^G$. Finally

$$\Pi_G^* = \frac{1}{|G|} \sum_g \pi(g^{-1}) = \Pi_G,$$

so the idempotent Π_G is orthogonal projection. $\square$

COROLLARY A.10 (Mean-zero projection in the regular representation). *For the right regular representation on $L^2(G)$, $\Pi_G f$ is the constant function equal to the uniform average of f . Thus*

$$L^2(G) = \mathbb{C}\mathbf{1} \oplus L_0^2(G).$$

PROOF. For the right regular representation, $\Pi_G f(x) = |G|^{-1} \sum_{g \in G} f(xg)$. As $g \mapsto xg$ is a bijection of G , this is the uniform average of f and is independent of x . Hence the invariant subspace is exactly $\mathbb{C}\mathbf{1}$ and its orthogonal complement is the mean-zero subspace. $\square$

5. Powers and logarithmic time

PROPOSITION A.11 (Power norm for self-adjoint operators). *If $T = T^*$, then*

$$\|T^n\|_{\text{op}} = \|T\|_{\text{op}}^n$$

for every $n \geq 1$.

PROOF. In an orthonormal eigenbasis, the eigenvalues of T^n are λ_i^n . Apply Theorem A.2. $\square$

COROLLARY A.12 (Gap gives exponential decay). *If a self-adjoint Markov operator has absolute gap $\gamma_{\text{abs}} > 0$, then on the mean-zero subspace*

$$\|P^n\|_{\text{op}} \leq (1 - \gamma_{\text{abs}})^n \leq e^{-\gamma_{\text{abs}} n}.$$

PROOF. The first inequality is the definition of the absolute gap plus [Proposition A.11](#). The second uses $1 - x \leq e^{-x}$. $\square$

REMARK A.13 (Why logarithmic time is the right scale). If the stationary mass of a point is N^{-1} , then reducing an L^2 error of order 1 to order N^{-c} under a uniform gap takes $O(\log N)$ steps. Conversely, the Bourgain–Gamburd machine tries to prove strong enough spreading by $O(\log N)$ time that quasirandomness forces a uniform gap. The logarithm is therefore not an incidental feature of the proof; it is the natural mixing scale of bounded-degree expanders.

6. Trace and Hilbert–Schmidt norm

For an orthonormal basis (e_i) define

$$\text{tr } T = \sum_i \langle T e_i, e_i \rangle, \quad \|T\|_{\text{HS}}^2 = \sum_i \|T e_i\|^2.$$

LEMMA A.14 (Basis independence). *Both quantities are independent of the orthonormal basis, and*

$$\|T\|_{\text{HS}}^2 = \text{tr}(T^* T).$$

PROOF. Let U be the unitary change-of-basis matrix. Matrix representatives transform by $A \mapsto U^* A U$. The elementary identity $\text{tr}(BC) = \text{tr}(CB)$ yields

$$\text{tr}(U^* A U) = \text{tr}(A U U^*) = \text{tr } A.$$

Similarly,

$$\text{tr}((U^* A U)^* (U^* A U)) = \text{tr}(U^* A^* A U) = \text{tr}(A^* A).$$

The formula for the Hilbert–Schmidt norm follows from

$$\text{tr}(T^* T) = \sum_i \langle T^* T e_i, e_i \rangle = \sum_i \|T e_i\|^2.$$

$\square$

PROPOSITION A.15 (Eigenvalue counting by trace). *If $T = T^*$ has eigenvalues $\lambda_1, \dots, \lambda_m$, then*

$$\text{tr}(T^{2n}) = \sum_{i=1}^m \lambda_i^{2n}.$$

Consequently, if an eigenvalue λ occurs with multiplicity at least D , then

$$D|\lambda|^{2n} \leq \text{tr}(T^{2n}).$$

PROOF. Use an orthonormal eigenbasis. Every term in the even-power trace is nonnegative, so the contribution of a multiplicity- D eigenvalue alone is $D|\lambda|^{2n}$. $\square$

GUIDING IDEA A.16. The Bourgain–Gamburd endgame is precisely this one-line inequality applied to a convolution operator. Quasirandomness supplies D ; a random-walk L^2 bound supplies the trace; the inequality then forbids a large nontrivial eigenvalue.

7. Hilbert–Schmidt norm of convolution

Let G have order N , and let

$$T_\mu f = \mu * f$$

act on counting $\ell^2(G)$.

PROPOSITION A.17 (Kernel formula). *The matrix kernel of T_μ is*

$$K(x, z) = \mu(xz^{-1})$$

up to the left/right convention, and in either convention

$$\|T_\mu\|_{\text{HS}}^2 = N\|\mu\|_2^2.$$

If μ is symmetric, then

$$\text{tr}(T_\mu^{2n}) = N\|\mu^{(n)}\|_2^2.$$

PROOF. For the convolution convention of this book,

$$(T_\mu f)(x) = \sum_y \mu(y)f(y^{-1}x).$$

Writing $z = y^{-1}x$ gives $y = xz^{-1}$ and hence kernel $K(x, z) = \mu(xz^{-1})$. Therefore

$$\|T_\mu\|_{\text{HS}}^2 = \sum_{x, z} |\mu(xz^{-1})|^2.$$

For fixed z , the map $x \mapsto xz^{-1}$ is a bijection, so the inner sum is $\|\mu\|_2^2$. There are N choices of z .

If μ is symmetric, T_μ is self-adjoint and $T_\mu^{2n} = T_{\mu^{(2n)}}$. The diagonal kernel value is $\mu^{(2n)}(e)$, so

$$\text{tr}(T_\mu^{2n}) = N\mu^{(2n)}(e).$$

Symmetry gives

$$\mu^{(2n)}(e) = \sum_x \mu^{(n)}(x)\mu^{(n)}(x^{-1}) = \|\mu^{(n)}\|_2^2.$$

□

8. Nonsymmetric measures and singular values

The main text usually symmetrizes the measure because self-adjointness makes the spectral argument transparent. For completeness we record the nonsymmetric substitute.

DEFINITION A.18 (Singular values). The singular values of T are the nonnegative square roots of the eigenvalues of T^*T .

PROPOSITION A.19 (Operator norm from singular values). *For every finite-dimensional operator T ,*

$$\|T\|_{\text{op}}^2 = \|T^*T\|_{\text{op}}.$$

Thus $\|T\|_{\text{op}}$ is the largest singular value.

PROOF. For every unit vector v ,

$$\|Tv\|^2 = \langle T^*Tv, v \rangle.$$

Since T^*T is positive self-adjoint, Rayleigh–Ritz gives

$$\sup_{\|v\|=1} \langle T^*Tv, v \rangle = \|T^*T\|_{\text{op}}.$$

□

PROPOSITION A.20 (Adjoint convolution). *If*

$$\tilde{\mu}(x) = \overline{\mu(x^{-1})},$$

then

$$T_{\mu}^* = T_{\tilde{\mu}}.$$

For a probability measure, therefore,

$$T_{\mu}^*T_{\mu} = T_{\tilde{\mu}*\mu}.$$

PROOF. A direct change of variables gives

$$\begin{aligned} \langle T_{\mu}f, h \rangle &= \sum_{x,y} \mu(y) f(y^{-1}x) \overline{h(x)} \\ &= \sum_{z,y} f(z) \overline{\mu(y)} \overline{h(yz)}. \end{aligned}$$

Replacing y by u^{-1} identifies the bracketed expression with $T_{\tilde{\mu}}h(z)$. □

REMARK A.21. A nonsymmetric random walk can therefore be studied through the symmetric measure $\tilde{\mu} * \mu$. What one obtains directly is a singular-value gap. For Cayley expanders one normally begins with a symmetric generating multiset, so this additional layer is unnecessary; nevertheless it is useful in directed or nonreversible variants of the method.

9. Reversible Markov chains

The uniform measure is stationary for a Cayley walk, but some applications naturally live on a finite state space Ω with a nonuniform stationary measure π .

DEFINITION A.22 (Detailed balance). A transition kernel $P(x, y)$ is reversible with respect to π if

$$\pi(x)P(x, y) = \pi(y)P(y, x)$$

for all x, y .

Equip functions with

$$\langle f, g \rangle_\pi = \sum_x \pi(x) f(x) \overline{g(x)}.$$

PROPOSITION A.23 (Reversibility equals self-adjointness). A Markov kernel satisfying detailed balance is self-adjoint on $L^2(\pi)$.

PROOF. Using detailed balance,

$$\begin{aligned} \langle Pf, g \rangle_\pi &= \sum_{x,y} \pi(x) P(x, y) f(y) \overline{g(x)} \\ &= \sum_{x,y} \pi(y) P(y, x) f(y) \overline{g(x)} \\ &= \langle f, Pg \rangle_\pi. \end{aligned}$$

□

PROPOSITION A.24 (Dirichlet form for a reversible chain). If P is reversible, then

$$\langle (I - P)f, f \rangle_\pi = \frac{1}{2} \sum_{x,y} \pi(x) P(x, y) |f(x) - f(y)|^2.$$

PROOF. Expand the square. The $|f(x)|^2$ term contributes $\sum_x \pi(x) |f(x)|^2$ because rows sum to 1. Detailed balance and column summation give the same for the $|f(y)|^2$ term. The cross terms combine to $-2\Re \langle Pf, f \rangle_\pi$, which equals $-2\langle Pf, f \rangle_\pi$ because P is self-adjoint. □

10. A compact operator checklist

When an expansion proof invokes spectral theory, the following points should be checked explicitly.

- (1) Which Hilbert-space normalization is being used: counting ℓ^2 or normalized L^2 ?
- (2) Is the averaging operator self-adjoint? If not, is the argument really about singular values?
- (3) Does “spectral gap” mean $1 - \lambda_2$ or $1 - \max_{i \geq 2} |\lambda_i|$?

- (4) Is the trivial representation one-dimensional, or are there several invariant components because the graph is disconnected?
- (5) When a trace estimate is used, what is the multiplicity of the eigenvalue under consideration?
- (6) When a power P^n is estimated, has periodicity been removed or controlled?

These are elementary questions, but a wrong answer to any one of them can alter the conclusion by a power of $|G|$ or can turn a mixing statement into a false statement.

11. Exercises

EXERCISE A.25 (Lazy spectrum). If $P = P^*$ has spectrum in $[-1, 1]$, prove that

$$P_L = \frac{1}{2}(I + P)$$

has spectrum in $[0, 1]$ and ordinary spectral gap exactly half that of P .

EXERCISE A.26 (Trace without symmetry). Show that for an arbitrary probability measure μ ,

$$\|T_\mu\|_{\text{HS}}^2 = N\|\mu\|_2^2$$

still holds, but the identity $\text{tr}(T_\mu^{2n}) = N\|\mu^{(n)}\|_2^2$ need not. Find the correct identity involving $T_\mu^*T_\mu$.

EXERCISE A.27 (Variational comparison). Let P, Q be self-adjoint Markov operators with the same invariant constants. Suppose

$$\langle (I - P)f, f \rangle \geq c \langle (I - Q)f, f \rangle$$

for all f . Prove that the ordinary gap of P is at least c times the ordinary gap of Q .

EXERCISE A.28 (Multiplicity endgame). Let $T = T^*$ be a Markov operator on an N -dimensional space, with constants as the 1-eigenspace. Suppose every nontrivial eigenspace arising from a group representation has dimension at least D , and suppose

$$\text{tr}(T^{2n}) \leq 1 + D\rho^{2n}$$

for some $0 < \rho < 1$. Prove that every nontrivial eigenvalue has absolute value at most ρ .

12. Appendix summary

The operator theory used by the Bourgain–Gamburd method is finite-dimensional and completely explicit. Self-adjointness gives an orthonormal eigenbasis; the Rayleigh quotient turns graph inequalities into spectral gaps; averaging projects onto invariant vectors; and trace plus multiplicity converts random-walk return estimates into eigenvalue bounds. For nonsymmetric walks, singular values replace eigenvalues. These facts are simple enough to prove once and important enough never to leave implicit.

APPENDIX B

Word Metrics, Free Groups, and Schreier Coverings

The initial-spreading part of the Bourgain–Gamburd method is geometric before it is analytic: short words in the generators should remain distinct after passage to a finite quotient, and proper-subgroup concentration is naturally measured on Schreier graphs. This appendix develops that geometry from first principles. We construct word metrics and Cayley graphs, prove the normal-form theorem for free groups, identify the regular tree, formulate injectivity radius precisely, explain graph coverings of Cayley and Schreier graphs, and prove the ping-pong lemma. These tools make the phrases “free up to logarithmic scale” and “escape from cosets” mathematically exact.

1. Words and word length

Let S be a finite symmetric generating set of a group Γ ; thus $S = S^{-1}$. A word over S is an expression

$$w = s_1 s_2 \cdots s_n, \quad s_i \in S.$$

It represents the product of its letters in Γ .

DEFINITION B.1 (Word length). The word length of $g \in \Gamma$ is

$$|g|_S = \min\{n : g = s_1 \cdots s_n, s_i \in S\}.$$

The associated word metric is

$$d_S(g, h) = |g^{-1}h|_S.$$

PROPOSITION B.2 (Word metric). *The function d_S is a left-invariant metric. Moreover,*

$$d_S(kg, kh) = d_S(g, h)$$

for all $g, h, k \in \Gamma$.

PROOF. Positivity and symmetry follow from $S = S^{-1}$. If $g^{-1}h$ has a word of length m and $h^{-1}u$ has one of length n , concatenation gives a word of length $m + n$ for $g^{-1}u$, proving the triangle inequality. Finally

$$(kg)^{-1}(kh) = g^{-1}h.$$

□

2. Cayley graphs realize the word metric

The undirected Cayley graph $\text{Cay}(\Gamma, S)$ has vertex set Γ and an edge between g and gs for every $s \in S$.

PROPOSITION B.3 (Graph distance equals word distance). *For every $g, h \in \Gamma$, the graph distance in $\text{Cay}(\Gamma, S)$ is $d_S(g, h)$.*

PROOF. A path

$$g = g_0, g_1, \dots, g_n = h$$

with $g_i = g_{i-1}s_i$ gives

$$g^{-1}h = s_1 \cdots s_n,$$

so $d_S(g, h) \leq n$. Conversely, a shortest word for $g^{-1}h$ gives a path of that length from g to h . $\square$

COROLLARY B.4 (Balls are word sets). *The ball of radius r around the identity is*

$$B_S(r) = \{g : |g|_S \leq r\}.$$

Every other radius- r ball is a left translate of $B_S(r)$.

PROOF. The first assertion is exactly the preceding proposition with $g = e$. Left invariance of the word metric gives $B_S(g, r) = gB_S(r)$. $\square$

3. The free group and reduced words

Fix formal symbols $a_1, \dots, a_k$ and their formal inverses. A word is *reduced* if it contains no adjacent pair $a_i a_i^{-1}$ or $a_i^{-1} a_i$.

LEMMA B.5 (Reduction terminates). *Starting from a finite word and repeatedly deleting adjacent inverse pairs always terminates after finitely many deletions.*

PROOF. Every deletion lowers the word length by 2. $\square$

Termination alone does not imply uniqueness. For that we use a stack algorithm.

PROPOSITION B.6 (Unique reduced form). *Every word in the alphabet $\{a_i^{\pm 1}\}$ reduces to a unique reduced word under the relations $a_i a_i^{-1} = a_i^{-1} a_i = e$.*

PROOF. Read the word from left to right while maintaining a stack. When a new letter is the inverse of the top letter, remove the top; otherwise append the new letter. The output is reduced.

We show that inserting or deleting an adjacent inverse pair anywhere in a word does not change the final stack. Indeed, processing ss^{-1} consecutively either appends s and immediately removes it, or first removes a top s^{-1} and then restores it. In either case the stack after the pair is exactly what it was before the pair. Thus the stack output is invariant under all elementary cancellations and insertions. Any reduced word is fixed by the stack algorithm, so two equivalent reduced words must coincide. $\square$

DEFINITION B.7 (Free group). The free group $F_k = F(a_1, \dots, a_k)$ is the set of reduced words with multiplication defined by concatenation followed by reduction.

THEOREM B.8 (Universal property). *Given any group G and elements $g_1, \dots, g_k \in G$, there is a unique homomorphism*

$$\phi : F_k \rightarrow G$$

with $\phi(a_i) = g_i$.

PROOF. For a reduced word

$$w = a_{i_1}^{\varepsilon_1} \cdots a_{i_n}^{\varepsilon_n},$$

define

$$\phi(w) = g_{i_1}^{\varepsilon_1} \cdots g_{i_n}^{\varepsilon_n}.$$

Cancellation of adjacent inverse letters does not change this product, so multiplication after reduction is respected. Uniqueness follows because the a_i generate F_k . $\square$

4. The Cayley graph of a free group is a tree

Take the symmetric generating set

$$S = \{a_1^{\pm 1}, \dots, a_k^{\pm 1}\}.$$

THEOREM B.9 (Free Cayley tree). *The Cayley graph $\text{Cay}(F_k, S)$ is the $(2k)$ -regular tree.*

PROOF. Each vertex has one neighbor gs for each $s \in S$; these neighbors are distinct, since $gs = gt$ implies $s = t$. Thus the graph is $2k$ -regular.

Suppose there were a simple cycle based at some g . Left translation moves it to a simple cycle based at the identity. Reading its edge labels produces a nonempty word

$$s_1 \cdots s_n = e.$$

Because the cycle has no immediate backtracking, the word is reduced. But the only reduced representative of the identity in F_k is the empty word, a contradiction. A connected graph with no cycles is a tree. $\square$

COROLLARY B.10 (Sphere sizes). *For $n \geq 1$,*

$$|\{g \in F_k : |g|_S = n\}| = 2k(2k - 1)^{n-1}.$$

Hence

$$|B_S(r)| = 1 + 2k \sum_{j=0}^{r-1} (2k - 1)^j.$$

PROOF. There are $2k$ choices for the first letter and then $2k - 1$ choices at each subsequent position, since only the inverse of the preceding letter is forbidden. $\square$

5. Relations, girth, and injectivity radius

Let $\phi : F_k \twoheadrightarrow G$ send the free generators to a symmetric generating tuple of a finite group G .

DEFINITION B.11 (Relation length). A nontrivial reduced word $w \in F_k$ is a relation in G if $\phi(w) = e$. The relation length is $|w|$.

PROPOSITION B.12 (Relations and Cayley girth). *The girth of the Cayley graph of G is the minimum length of a nontrivial cyclically reduced relation. In particular, if there is no nontrivial reduced relation of length at most $2r$, then the quotient map*

$$F_k \rightarrow G$$

is injective on $B_S(r)$.

PROOF. A simple closed path gives a cyclically reduced relation by reading labels around the path. Conversely, a shortest nontrivial cyclically reduced relation gives a closed nonbacktracking path; if a vertex repeated before the end, one of the resulting subloops would give a shorter relation, contradicting minimality.

For the final assertion, if $u, v \in B_S(r)$ have the same image, then $u^{-1}v$ lies in the kernel and has word length at most $2r$. After reduction it is either empty, giving $u = v$, or a forbidden nontrivial relation. $\square$

DEFINITION B.13 (Injectivity radius). The injectivity radius of the quotient $F_k \rightarrow G$ is the largest r for which the map is injective on $B_S(r)$.

REMARK B.14. For a Cayley graph of girth g , the injectivity radius is at least $\lfloor (g-1)/2 \rfloor$. This factor of 2 matters: equality of two words of length at most r produces a relation of length at most $2r$, not r .

6. Local lifting of random walks

Let μ_F be the simple random-walk measure on F_k and μ_G its pushforward to G .

PROPOSITION B.15 (Exact local lifting). *If $F_k \rightarrow G$ is injective on $B_S(n)$, then for every $x \in B_S(n)$,*

$$\mu_G^{(n)}(\phi(x)) = \mu_F^{(n)}(x).$$

In particular,

$$\mu_G^{(2n)}(e) = \mu_F^{(2n)}(e)$$

whenever the quotient is injective on $B_S(2n)$.

PROOF. Every length- n path from e ends in $B_S(n)$. Under injectivity on that ball, two endpoints upstairs cannot collide downstairs. The quotient map preserves edge labels and path probabilities, so endpoint masses agree. Apply this with endpoint e at time $2n$ for the second assertion. $\square$

WARNING B.16. To compare a return probability at time $2n$, injectivity on $B(2n)$ is sufficient and clean. It is sometimes possible to use only a girth bound slightly larger than $2n$ by analyzing the closed path directly, but one should not silently replace a $2n$ condition by an n condition.

7. Graph coverings

A graph map $p : X \rightarrow Y$ is a covering if for every vertex $x \in X$, the map from edges incident to x onto edges incident to $p(x)$ is a bijection, respecting labels when labels are present.

PROPOSITION B.17 (Free Cayley graph covers every generated Cayley graph). *The natural map*

$$\text{Cay}(F_k, S) \rightarrow \text{Cay}(G, \phi(S))$$

is a label-preserving graph covering.

PROOF. From each vertex $w \in F_k$ there is exactly one outgoing edge labeled by each $s \in S$, namely $w \rightarrow ws$. Its image is the corresponding labeled edge $\phi(w) \rightarrow \phi(w)\phi(s)$. Thus the star at every vertex maps bijectively onto the labeled star below. $\square$

REMARK B.18. The covering need not be injective globally; its deck transformations are related to the kernel of $F_k \rightarrow G$. The injectivity radius is precisely the radius on which the covering looks like an isomorphism around a chosen lift.

8. Group actions and Schreier graphs

Let a group Γ act on a set Ω on the right. For a symmetric generating set S , the Schreier graph $\text{Sch}(\Omega, S)$ has vertices Ω and edges

$$\omega \longleftrightarrow \omega s.$$

EXAMPLE B.19 (Coset action). If $H \leq \Gamma$, the right action on left cosets $H \backslash \Gamma$ is

$$(Hg)s = H(gs).$$

The corresponding Schreier graph records movement among left cosets. If instead one works with right cosets and left multiplication, the convention changes. Expansion arguments must keep this choice synchronized with the convolution convention.

PROPOSITION B.20 (Coset probability as a Schreier transition probability). *Let μ be a probability measure on Γ , and let $X_n = X_1 \cdots X_n$ with independent increments of law μ . Then for the right action on $H \backslash \Gamma$,*

$$\mathbb{P}(HX_n = Hg) = \mu^{(n)}(Hg) := \sum_{x \in Hg} \mu^{(n)}(x).$$

PROOF. The event $HX_n = Hg$ is exactly the event $X_n \in Hg$. Summing the endpoint law over the coset gives the formula. $\square$

COROLLARY B.21 (Escape is a mixing statement on every proper Schreier graph).
A uniform estimate

$$\sup_{g \in \Gamma} \mu^{(n)}(Hg) \leq \varepsilon$$

for a subgroup H is exactly an L^∞ upper bound for the n -step distribution of the induced walk on $H \backslash \Gamma$.

PROOF. By the preceding proposition, the probability assigned by the induced walk to the coset Hg is exactly $\mu^{(n)}(Hg)$. Taking the supremum over cosets is therefore the ℓ^∞ norm of the induced probability distribution. $\square$

9. Stabilizers and orbit size

PROPOSITION B.22 (Orbit–stabilizer). *If a finite group G acts on Ω , then*

$$|G \cdot \omega| = [G : \text{Stab}_G(\omega)].$$

PROOF. The map

$$G/\text{Stab}_G(\omega) \rightarrow G \cdot \omega, \quad g \text{Stab}_G(\omega) \mapsto g\omega$$

is well-defined and bijective. $\square$

REMARK B.23. Subgroups appearing in rank-one groups are often stabilizers of simple geometric data: a point or pair of points of a projective line, a line in a vector space, or a decomposition preserved by a torus. Orbit–stabilizer then converts subgroup size estimates into geometry of the action.

10. Ping-pong

The ping-pong lemma is a robust way to prove that two transformations generate a free subgroup or free semigroup.

THEOREM B.24 (Ping-pong lemma, free-group form). *Let a group act on a set X . Suppose a, b have infinite order and there exist four nonempty pairwise disjoint subsets*

$$A_+, A_-, B_+, B_- \subset X$$

such that

$$\begin{aligned} a(X \setminus A_-) &\subset A_+, & a^{-1}(X \setminus A_+) &\subset A_-, \\ b(X \setminus B_-) &\subset B_+, & b^{-1}(X \setminus B_+) &\subset B_-. \end{aligned}$$

Then $\langle a, b \rangle$ is freely generated by a and b .

PROOF. For $s \in \{a, a^{-1}, b, b^{-1}\}$ write D_s for its attracting domain, so

$$D_a = A_+, \quad D_{a^{-1}} = A_-, \quad D_b = B_+, \quad D_{b^{-1}} = B_-.$$

The hypotheses say uniformly that

$$(70) \quad s(X \setminus D_{s^{-1}}) \subseteq D_s.$$

Take a nonempty reduced word

$$w = s_1 s_2 \cdots s_n.$$

Choose a label $t \in \{a, a^{-1}, b, b^{-1}\}$ with

$$t \neq s_n^{-1} \quad \text{and} \quad t \neq s_1.$$

Such a label exists because at most two of the four labels are forbidden. Pick $x \in D_t$.

Since the four domains are pairwise disjoint and $t \neq s_n^{-1}$, we have $x \notin D_{s_n^{-1}}$, so Equation (70) gives

$$s_n x \in D_{s_n}.$$

Because the word is reduced, $s_{n-1} \neq s_n^{-1}$; hence D_{s_n} is disjoint from $D_{s_{n-1}^{-1}}$, and another application of Equation (70) gives

$$s_{n-1} s_n x \in D_{s_{n-1}}.$$

Continuing inductively yields

$$wx \in D_{s_1}.$$

But $x \in D_t$ and $t \neq s_1$, so pairwise disjointness gives $x \notin D_{s_1}$. Thus $wx \neq x$, hence $w \neq e$. Every nonempty reduced word is nontrivial, so a, b freely generate a free group. $\square$

REMARK B.25 (A cleaner basepoint variant). A commonly used formulation assumes only two disjoint sets A, B and the inclusions

$$a^n(B) \subset A, \quad b^n(A) \subset B$$

for every nonzero integer n . Then a reduced word written in alternating nonzero powers sends a basepoint alternately into A and B , making the proof even shorter. The four-domain version above is better adapted to proximal dynamics, where attracting and repelling neighborhoods are naturally distinct.

11. A quantitative ping-pong viewpoint

For finite quotients, freeness itself is impossible once the quotient is finite. What is useful is *short-word freeness*.

DEFINITION B.26 (Free to radius R). A tuple $(g_1, \dots, g_k)$ in a group G is free to radius R if the induced map $F_k \rightarrow G$ is injective on $B(R)$.

PROPOSITION B.27 (Quantitative consequence). *If a k -generator tuple is free to radius $2n$, then the $2n$ -step return probability of the simple random walk on G equals that on F_k . Consequently,*

$$\|\mu_G^{(n)}\|_2^2 = \mu_G^{(2n)}(e) = \mu_{F_k}^{(2n)}(e).$$

PROOF. Use Proposition B.15 and symmetry of the simple random walk. $\square$

12. Arithmetic height and short relations

Let $S \subset \mathrm{GL}_d(\mathbb{Z})$ be finite, and put

$$M = \max_{s \in S} \max_{i,j} |s_{ij}|.$$

For a matrix A write $\|A\|_{\max} = \max_{i,j} |a_{ij}|$.

LEMMA B.28 (Elementary height growth). *For $A, B \in M_d(\mathbb{Z})$,*

$$\|AB\|_{\max} \leq d\|A\|_{\max}\|B\|_{\max}.$$

Hence a word $w = s_1 \cdots s_n$ in S satisfies

$$\|w\|_{\max} \leq d^{n-1}M^n \leq (dM)^n.$$

PROOF. Each entry of AB is a sum of d products, each of absolute value at most $\|A\|_{\max}\|B\|_{\max}$. Iterate. $\square$

PROPOSITION B.29 (Congruence identity forces a height threshold). *Suppose $w \in \mathrm{GL}_d(\mathbb{Z})$ satisfies*

$$w \equiv I \pmod{p}$$

but $w \neq I$. Then

$$p \leq \|w - I\|_{\max} \leq (dM)^n + 1$$

if w is represented by a word of length n in S .

PROOF. Some entry of $w - I$ is a nonzero multiple of p , hence has absolute value at least p . The upper bound follows from Lemma B.28. $\square$

COROLLARY B.30 (Logarithmic injectivity from absence of integral relations). *Assume the chosen generators satisfy no nontrivial integral relation of word length at most n . If*

$$(dM)^n + 1 < p,$$

then reduction modulo p introduces no nontrivial relation of length at most n . In particular, one obtains injectivity radius $c \log p$ for every fixed

$$c < \frac{1}{2 \log(dM)}$$

once p is large, provided the corresponding integral words are nontrivial.

PROOF. A relation modulo p would give a word matrix $w \neq I$ congruent to I modulo p , contradicting Proposition B.29. The factor $1/2$ is included when converting a relation-length statement to injectivity on a ball via Proposition B.12. $\square$

WARNING B.31 (What height does not prove). The height estimate detects whether a fixed short word becomes the identity modulo p . It does not by itself show that a random walk avoids every proper subgroup. A word may land in a Borel subgroup, a torus normalizer, or another proper algebraic subgroup without becoming the identity. Subgroup escape therefore requires additional geometry or subgroup classification.

13. Schreier return probabilities and subgroup intersections

Let $H \leq G$ and let μ be symmetric.

PROPOSITION B.32 (Coset collision identity). *For every n ,*

$$\sum_{Hg \in H \backslash G} \mu^{(n)}(Hg)^2 = \mu^{(2n)}(H),$$

where

$$\mu^{(2n)}(H) = \sum_{h \in H} \mu^{(2n)}(h).$$

PROOF. Expand the left side:

$$\sum_{Hg} \sum_{x, y \in Hg} \mu^{(n)}(x) \mu^{(n)}(y) = \sum_{x, y} \mu^{(n)}(x) \mu^{(n)}(y) 1_{xy^{-1} \in H}.$$

By symmetry, $\mu^{(n)}(y) = \mu^{(n)}(y^{-1})$. For a fixed $h \in H$, the contribution of pairs with $xy^{-1} = h$ is

$$\sum_y \mu^{(n)}(hy) \mu^{(n)}(y^{-1}) = \mu^{(2n)}(h).$$

Summing over $h \in H$ gives the result. $\square$

COROLLARY B.33 (From subgroup return to worst coset). *For every left coset Hg ,*

$$\mu^{(n)}(Hg) \leq \mu^{(2n)}(H)^{1/2}.$$

PROOF. Each summand in the nonnegative sum of squares in Proposition B.32 is at most the whole sum. $\square$

GUIDING IDEA B.34. This identity is often the cleanest way to pass from an estimate on the probability of a two-walk difference landing in H to a uniform estimate on every coset. It also explains why even times and symmetrization appear so naturally in escape arguments.

14. A geometric checklist for initial spreading

Before using a free-group or girth argument, check the following chain with the constants shown:

$$\begin{aligned} \text{height bound} &\Rightarrow \text{no relation of length at most } c \log |G|, \\ &\Rightarrow \text{injectivity on a ball of radius } \frac{c}{2} \log |G|, \\ &\Rightarrow \text{the free return estimate at the chosen even time,} \\ &\Rightarrow \|\mu^{(n)}\|_2 \leq |G|^{-\alpha}. \end{aligned}$$

Every arrow loses a specific factor or requires a specific convention. In particular, the random-walk time must remain within the local injectivity range.

15. Exercises

EXERCISE B.35 (Free sphere growth). Derive the exact formula for $|B_S(r)|$ in F_k and show that

$$|B_S(r)| \asymp_k (2k - 1)^r.$$

Explain why cardinality growth alone does not determine the simple random-walk return probability.

EXERCISE B.36 (Girth under changing generators). Let T be another generating set in which every $t \in T$ is an S -word of length at most L . Show that an S -relation of length r gives a T -statement only after controlling the reverse word comparison. Explain why girth is much less stable under changing generators than qualitative generation.

EXERCISE B.37 (Schreier stationary distribution). For a finite transitive G -set Ω and a symmetric probability measure μ on G , prove that the induced Markov chain on Ω is reversible with respect to the uniform measure.

EXERCISE B.38 (Ping-pong with powers). Suppose a, b act on a compact space with disjoint attracting and repelling neighborhoods and sufficiently large powers satisfy the four ping-pong inclusions. Prove that those powers freely generate a free group. Identify precisely where compactness is used, if at all, after the inclusions have been established.

EXERCISE B.39 (Coset collision without symmetry). Find the correct analogue of Proposition B.32 for a nonsymmetric measure μ , using $\tilde{\mu}^{(n)} * \mu^{(n)}$.

16. Appendix summary

Word metrics turn algebraic expressions into graph distance. Free groups have unique reduced words, so their Cayley graphs are regular trees. A finite quotient agrees with this tree up to its injectivity radius, allowing exact transfer of short-time random-walk estimates. Schreier graphs encode coset probabilities, while ping-pong and arithmetic height provide mechanisms for proving short-word freeness. These geometric facts are the correct foundation for the “initial spreading” and “subgroup escape” inputs of the expansion machine.

APPENDIX C

Worked Finite-Group Laboratories

Abstract machinery becomes easier to understand and test once one has small groups on which every object can be computed. This appendix develops several laboratories: cyclic groups show the abelian obstruction to quasirandomness; S_3 exhibits nonabelian Fourier blocks explicitly; subgroup-uniform measures show exact failure of flattening; direct products and quotients illustrate how quasirandomness behaves under basic constructions; and the finite Heisenberg group shows that large-dimensional representations can coexist with many one-dimensional representations. The examples are chosen to clarify mechanisms rather than to catalogue character tables.

1. Cyclic groups: the boundary case

Let

$$G = C_n = \mathbb{Z}/n\mathbb{Z}.$$

For $0 \leq j < n$ define

$$\chi_j(x) = e^{2\pi i j x / n}.$$

PROPOSITION C.1 (All irreducibles of C_n are one-dimensional). *The n characters χ_j are pairwise distinct irreducible representations and exhaust $\text{Irr}(C_n)$.*

PROOF. Each χ_j is a homomorphism to the unit circle, hence a one-dimensional irreducible representation. Orthogonality gives

$$\frac{1}{n} \sum_{x=0}^{n-1} \chi_j(x) \overline{\chi_k(x)} = 1_{j=k}.$$

The sum of squares of the dimensions of all irreducibles equals $|G|$ by the regular-representation decomposition. The n characters already contribute n , so there are no others. $\square$

COROLLARY C.2. *Every nontrivial finite abelian group has quasirandomness parameter $D(G) = 1$.*

PROOF. By [Proposition 4.5](#), every irreducible complex representation of a finite abelian group is one-dimensional and every nontrivial finite abelian group has a

nontrivial character. Hence the least degree of a nontrivial irreducible representation is 1. $\square$

GUIDING IDEA C.3. This is why the quasirandom endgame is inherently nonabelian. Abelian groups can certainly have expanding Cayley graphs if the degree grows, but a bounded-degree family of abelian Cayley graphs cannot obtain a Bourgain–Gamburd endgame from representation multiplicity: there is no multiplicity to exploit.

2. The cycle random walk

Take the symmetric measure

$$\mu = \frac{1}{2}(\delta_1 + \delta_{-1})$$

on C_n . Fourier transformation diagonalizes convolution:

$$\widehat{\mu}(j) = \frac{1}{2} \left(e^{2\pi i j/n} + e^{-2\pi i j/n} \right) = \cos \frac{2\pi j}{n}.$$

PROPOSITION C.4 (Cycle spectral gap). *The ordinary spectral gap of the nearest-neighbor walk on C_n is*

$$1 - \cos \frac{2\pi}{n} = \frac{2\pi^2}{n^2} + O(n^{-4}).$$

In particular the cycles are not an expander family.

PROOF. The nontrivial eigenvalues are $\cos(2\pi j/n)$. The largest is attained at $j = 1$ or $n - 1$. Taylor expansion of cosine gives the asymptotic. $\square$

REMARK C.5. The walk spreads, but only diffusively: distance of order $\sqrt{t}$ is typical at time t before the finite quotient wraps around. Reaching macroscopic scale therefore requires $t \asymp n^2$, in sharp contrast with the $O(\log |G|)$ scale of expanders.

3. The group S_3

The symmetric group S_3 has three conjugacy classes:

$$\{e\}, \quad \{(12), (13), (23)\}, \quad \{(123), (132)\}.$$

It has the trivial representation, the sign representation, and a two-dimensional standard representation.

PROPOSITION C.6 (The standard representation). *Let S_3 act on*

$$W = \{(z_1, z_2, z_3) \in \mathbb{C}^3 : z_1 + z_2 + z_3 = 0\}$$

by permuting coordinates. Then W is a two-dimensional irreducible representation.

PROOF. The coordinate-sum-zero subspace is invariant and has dimension 2. Suppose it contained a one-dimensional invariant subspace L . Since transpositions have order 2, each transposition would act on L by ± 1 . Thus the action on L would define either the trivial or sign character.

There is no nonzero S_3 -fixed vector in W , because an invariant vector has all coordinates equal and sum zero. There is also no sign line in W : a vector v satisfying $(12)v = -v$ and $(23)v = -v$ would have the form $(a, -a, 0)$ from the first relation and $(0, b, -b)$ from the second, forcing $v = 0$. Hence W has no invariant line and is irreducible. $\square$

The dimension identity

$$1^2 + 1^2 + 2^2 = 6 = |S_3|$$

shows these are all irreducibles.

4. A central random walk on S_3

Let μ be uniform on the three transpositions. Since the support is a conjugacy class, μ is central. By Schur's lemma, $\widehat{\mu}(\rho)$ is scalar on every irreducible representation.

PROPOSITION C.7 (Spectrum of the transposition walk). *For the transposition walk on S_3 , the Fourier scalars are*

1 on the trivial representation,

−1 on the sign representation,

and

0 on the standard representation.

Thus the ordinary gap is 1, but the absolute gap is 0.

PROOF. The trivial scalar is 1. Every transposition has sign -1 , so the sign scalar is -1 .

For an irreducible ρ of dimension d , a central class average has scalar

$$\frac{\chi_\rho(C)}{d},$$

where $\chi_\rho(C)$ is the common character value on the class. In the standard representation, a transposition acts on the coordinate-sum-zero plane as a reflection, with eigenvalues $1, -1$, so its trace is 0. Hence the scalar is 0. $\square$

GUIDING IDEA C.8. This tiny example makes the ordinary/absolute distinction unavoidable. The chain is bipartite: parity alternates at every step, producing the eigenvalue -1 . Laziness replaces the three eigenvalues $1, -1, 0$ by $1, 0, 1/2$, after which ordinary and absolute mixing are compatible.

5. Subgroup-uniform measures

Let $H \leq G$ and

$$\mu_H = |H|^{-1} 1_H.$$

PROPOSITION C.9 (Idempotence). *One has*

$$\mu_H * \mu_H = \mu_H.$$

Hence

$$\|\mu_H^{(n)}\|_2 = |H|^{-1/2}$$

for every $n \geq 1$.

PROOF. For $x \in H$, every $h \in H$ gives a unique factorization

$$x = h(h^{-1}x)$$

with both factors in H , so

$$(1_H * 1_H)(x) = |H|.$$

For $x \notin H$ there are no such factorizations. Therefore

$$1_H * 1_H = |H| 1_H,$$

and division by $|H|^2$ gives idempotence. The norm formula is immediate. $\square$

COROLLARY C.10 (Coset translates). *If xH is a left coset and $\nu = |H|^{-1} 1_{xH}$, then $\|\nu\|_2 = |H|^{-1/2}$. Convolution by subgroup-uniform measures transports mass among cosets without creating internal flattening beyond the subgroup scale.*

PROOF. Left translation preserves counting ℓ^2 , so $|H|^{-1} 1_{xH}$ has the same norm as $u_H = |H|^{-1} 1_H$, namely $|H|^{-1/2}$. Moreover convolution by u_H averages over a right H -coset and convolution by a translate of u_H merely translates the resulting coset-uniform measure. Thus the measure can move between cosets but cannot become flatter inside a coset than the uniform H -scale. $\square$

REMARK C.11. This is the exact algebraic obstruction that the subgroup-escape hypothesis excludes. An approximate subgroup carrying substantial mass is an approximate version of the same phenomenon.

6. Direct products

Let G_1, G_2 be finite groups.

PROPOSITION C.12 (Irreducibles of a direct product). *Every irreducible complex representation of $G_1 \times G_2$ is of the form*

$$\rho_1 \boxtimes \rho_2, \quad (\rho_1 \boxtimes \rho_2)(g_1, g_2) = \rho_1(g_1) \otimes \rho_2(g_2),$$

with $\rho_i \in \text{Irr}(G_i)$. Consequently

$$D(G_1 \times G_2) = \min(D(G_1), D(G_2)).$$

PROOF. The tensor products listed are irreducible: if an operator commutes with all $\rho_1(g_1) \otimes I$ and all $I \otimes \rho_2(g_2)$, Schur's lemma twice forces it to be scalar, and complete reducibility then gives irreducibility. Their squared dimensions sum to

$$\left(\sum_{\rho_1} d_{\rho_1}^2 \right) \left(\sum_{\rho_2} d_{\rho_2}^2 \right) = |G_1| |G_2|,$$

so they exhaust the regular representation.

A nontrivial representation can be $\rho_1 \boxtimes 1$ or $1 \boxtimes \rho_2$, showing

$$D(G_1 \times G_2) \leq \min(D(G_1), D(G_2)).$$

Every other nontrivial tensor product has dimension at least both relevant minima, giving the reverse inequality. $\square$

WARNING C.13. Polynomial quasirandomness can be lost under products if one factor stays small. For example, $G_p \times C_2$ always has the sign character of the C_2 factor, so its quasirandomness parameter is 1 no matter how quasirandom G_p becomes.

7. Quotients and one-dimensional representations

PROPOSITION C.14 (Quotients can only lower quasirandomness). *If $\pi : G \twoheadrightarrow Q$ is a nontrivial quotient, then*

$$D(G) \leq D(Q).$$

In particular, if G has a nontrivial abelian quotient, then $D(G) = 1$.

PROOF. Every irreducible representation ρ of Q pulls back to an irreducible representation $\rho \circ \pi$ of G of the same dimension. If Q is nontrivial abelian, it has a nontrivial one-dimensional character. $\square$

COROLLARY C.15 (Perfectness is necessary for quasirandomness beyond 1). *If $D(G) > 1$, then*

$$G = [G, G].$$

PROOF. The abelianization $G/[G, G]$ must be trivial by the proposition. $\square$

REMARK C.16. This elementary observation foreshadows why perfectness appears in super-approximation. It is not sufficient by itself for expansion, but a large abelian quotient is an immediate representation-theoretic obstruction to the quasirandom endgame.

8. The finite Heisenberg group

Let p be prime and define

$$H_p = \left\{ \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} : x, y, z \in \mathbb{F}_p \right\}.$$

We identify an element with (x, y, z) , for which multiplication is

$$(x, y, z)(x', y', z') = (x + x', y + y', z + z' + xy').$$

PROPOSITION C.17 (Center and abelianization). *The center is*

$$Z(H_p) = \{(0, 0, z) : z \in \mathbb{F}_p\},$$

and

$$H_p/Z(H_p) \cong \mathbb{F}_p^2$$

is abelian. Hence $D(H_p) = 1$.

PROOF. A direct computation gives

$$[(x, y, z), (x', y', z')] = (0, 0, xy' - x'y).$$

Thus every central element has $x = y = 0$, and such elements are visibly central. Modulo the center, multiplication is componentwise addition in (x, y) . Since the quotient is nontrivial abelian, Proposition C.14 gives $D(H_p) = 1$. $\square$

Nevertheless H_p also has genuinely large irreducible representations.

PROPOSITION C.18 (The Schrödinger representation is irreducible). *Fix a nontrivial additive character*

$$\psi : \mathbb{F}_p \rightarrow \mathbb{C}^\times.$$

On $V = \mathbb{C}^{\mathbb{F}_p}$ define

$$(\rho(x, y, z)f)(t) = \psi(z + yt)f(t + x).$$

Then ρ is a p -dimensional irreducible representation of H_p .

PROOF. First we check the representation law with the multiplication convention in Proposition C.17. For $h = (x, y, z)$ and $h' = (x', y', z')$,

$$\begin{aligned} (\rho(h)\rho(h')f)(t) &= \psi(z + yt)\psi(z' + y'(t + x))f(t + x + x') \\ &= \psi(z + z' + xy' + (y + y')t)f(t + x + x') \\ &= (\rho(hh')f)(t). \end{aligned}$$

Thus ρ is a representation.

We use the elementary orthogonality relation for the fixed nontrivial additive character:

$$(71) \quad \sum_{y \in \mathbb{F}_p} \psi(yu) = \begin{cases} p, & u = 0, \\ 0, & u \neq 0. \end{cases}$$

The first case is immediate. If $u \neq 0$, multiplication by u permutes $\mathbb{F}_p$, so the sum equals $S = \sum_y \psi(y)$. Choose a with $\psi(a) \neq 1$. Translation by a also permutes $\mathbb{F}_p$, hence

$$S = \sum_y \psi(y + a) = \psi(a)S,$$

forcing $S = 0$.

Let $0 \neq W \leq V$ be H_p -invariant and choose $0 \neq f \in W$. Pick $s \in \mathbb{F}_p$ with $f(s) \neq 0$. Since

$$\rho(0, y, 0)f(t) = \psi(yt)f(t),$$

the operator

$$P_s = \frac{1}{p} \sum_{y \in \mathbb{F}_p} \psi(-ys)\rho(0, y, 0)$$

preserves W . By (71),

$$(P_s f)(t) = f(t) \frac{1}{p} \sum_y \psi(y(t-s)) = f(s) \delta_s(t).$$

Hence $\delta_s \in W$. The translation operators

$$\rho(x, 0, 0)\delta_s = \delta_{s-x}$$

(up to the harmless indexing convention for basis delta functions) produce every basis vector δ_t . Therefore $W = V$, proving irreducibility. $\square$

PROPOSITION C.19 (One-dimensional characters of H_p). *The group H_p has exactly p^2 one-dimensional complex characters.*

PROOF. Every one-dimensional representation kills the commutator subgroup because $\mathbb{C}^\times$ is abelian. By Proposition C.17 the commutator subgroup equals the center: the displayed commutator formula gives $[H_p, H_p] \subseteq Z(H_p)$, while taking $(x, y) = (1, 0)$ and $(x', y') = (0, z)$ realizes every $(0, 0, z)$ as a commutator. Hence every one-dimensional character factors uniquely through

$$H_p/Z(H_p) \cong \mathbb{F}_p^2.$$

For $(r, s) \in \mathbb{F}_p^2$ define

$$\chi_{r,s}(x, y) = \psi(rx + sy).$$

These are p^2 distinct characters of the additive group $\mathbb{F}_p^2$: if $(r, s) \neq (r', s')$, then the nonzero linear functional $(r - r')x + (s - s')y$ takes every value in $\mathbb{F}_p$, so the two characters differ because ψ is nontrivial.

Conversely, let $\chi : \mathbb{F}_p^2 \rightarrow \mathbb{C}^\times$ be a character. Its restrictions to the two coordinate copies of $(\mathbb{F}_p, +)$ are determined by the images of 1. Every such image is a p th root of unity, and because the fixed nontrivial ψ maps 1 to a primitive p th root, each restriction is $x \mapsto \psi(rx)$ for a unique $r \in \mathbb{F}_p$. Thus $\chi = \chi_{r,s}$ for a unique pair (r, s) . There are exactly p^2 characters. $\square$

EXAMPLE C.20 (Large and small irreducibles coexist). By Propositions C.18 and C.19, H_p has an irreducible representation of dimension p and also p^2 one-dimensional characters. This is a useful warning: the existence of some large irreducibles is irrelevant to quasirandomness; one needs a lower bound for *every* nontrivial irreducible.

9. A product-mixing calculation

Suppose G is D -quasirandom. The Gowers mixing inequality from Chapter 5 says that for mean-zero f and arbitrary g ,

$$\|f * g\|_2 \leq \sqrt{\frac{|G|}{D}} \|f\|_2 \|g\|_2$$

under counting normalization in the form used there.

For sets $A, B, C \subseteq G$, write

$$1_A = \alpha \mathbf{1} + f_A, \quad \alpha = |A|/|G|,$$

and similarly for B, C . The expected number of solutions to $ab = c$ under perfect independence is

$$\frac{|A||B||C|}{|G|}.$$

The quasirandom error is controlled by the mean-zero convolution term.

PROPOSITION C.21 (Concrete triple-product threshold). *If*

$$|A||B||C| > \frac{|G|^3}{D},$$

then there exist $a \in A, b \in B, c \in C$ with

$$ab = c.$$

Equivalently,

$$AB \cap C \neq \emptyset.$$

PROOF. This is the product-mixing theorem from the main text; we reproduce the short calculation. The number of solutions is

$$\langle 1_A * 1_B, 1_C \rangle.$$

The constant part contributes $|A||B||C|/|G|$. The nonconstant error has absolute value at most

$$\|f_A * f_B\|_2 \|f_C\|_2 \leq \sqrt{\frac{|G|}{D}} \|f_A\|_2 \|f_B\|_2 \|f_C\|_2.$$

Since $\|f_A\|_2 \leq |A|^{1/2}$ and similarly,

$$|\text{error}| \leq \sqrt{\frac{|G|}{D}} (|A||B||C|)^{1/2}.$$

The main term is larger than the error precisely under the displayed strict inequality. $\square$

EXAMPLE C.22 (Dense triple products). If

$$|A| > |G|D^{-1/3},$$

then $A^3 = G$ provided the same bound is available uniformly after translating the target. Indeed, for any $g \in G$, apply the proposition to

$$A, \quad A, \quad gA^{-1},$$

whose three sizes equal $|A|$. A solution $ab = ga^{-1}$ rearranges to $aba = g$ after relabeling the last element. More carefully, if $c = ga_3^{-1} \in gA^{-1}$ and $a_1a_2 = c$, then

$$g = a_1a_2a_3.$$

Thus every g lies in A^3 .

10. Quasirandomness and the trace endgame in numbers

Suppose $|G| = N$, every nontrivial irreducible has dimension at least

$$D = N^\beta,$$

and a symmetric random walk satisfies at time n

$$\|\mu^{(n)}\|_2 \leq N^{-1/2+\beta/4}.$$

Then

$$N\|\mu^{(n)}\|_2^2 \leq N^{\beta/2}.$$

If λ is a nontrivial eigenvalue of T_μ , multiplicity amplification gives

$$N^\beta |\lambda|^{2n} \leq N\|\mu^{(n)}\|_2^2 \leq N^{\beta/2}.$$

Hence

$$|\lambda|^{2n} \leq N^{-\beta/2}.$$

If $n \leq C \log N$, then

$$|\lambda| \leq \exp\left(-\frac{\beta}{4C}\right).$$

This is a uniform constant below 1.

GUIDING IDEA C.23. This numerical exercise is the whole endgame stripped of notation. The random walk need not be microscopically close to uniform. It only has to beat the square-root uniform scale by a fixed power supplied by quasirandom multiplicity.

11. Exercises

EXERCISE C.24 (Lazy S_3 walk). For the transposition measure on S_3 , compute the spectrum of

$$\mu_L = \frac{1}{2}\delta_e + \frac{1}{2}\mu.$$

Compute its absolute gap and the exact n -step total-variation distance from uniform when starting at the identity.

EXERCISE C.25 (A subgroup obstruction in S_3). Let $H = \{e, (12)\}$. Compute $\mu_H * \mu_H$ directly and compare its Fourier transforms on the three irreducible representations of S_3 .

EXERCISE C.26 (Direct-product obstruction). Let (G_n) be a family with $D(G_n) \rightarrow \infty$. Show that $G_n \times C_2$ is never D_n -quasirandom for any $D_n > 1$. Explain which exact step of the abstract Bourgain–Gamburd theorem fails if no additional quotient restriction is imposed.

EXERCISE C.27 (Heisenberg commutator). Verify the commutator formula for H_p and show that

$$[H_p, H_p] = Z(H_p).$$

Count the one-dimensional representations from the abelianization.

EXERCISE C.28 (Triple-product threshold). Derive the threshold in [Proposition C.21](#) using normalized $L^2(G)$ instead of counting $\ell^2(G)$. Check that the final cardinality condition is unchanged.

12. Appendix summary

Cyclic groups show that quasirandomness is absent in the abelian world; the cycle makes the resulting slow spectral scale explicit. The group S_3 shows how nonabelian Fourier blocks, periodicity, and laziness interact in a completely computable case. Uniform subgroup measures are exact nonflattening obstructions. Direct products and quotients show how small representations survive structural operations, while the Heisenberg group warns that a few large irreducibles do not imply quasirandomness. Finally, the numerical trace calculation makes clear why polynomial quasirandomness plus L^2 flattening at logarithmic time yields a uniform spectral gap.

Bibliography

1. W. T. Gowers, *Quasirandom groups*, Combinatorics, Probability and Computing **17** (2008), no. 3, 363–387.

The Bourgain–Gamburd Method
Volume II: Noncommutative Additive
Combinatorics and Flattening

Preface

The compressed edition of this series treated noncommutative Balog–Szemerédi–Gowers theory and the small-tripling theorem as named external inputs. That is incompatible with the goal of understanding the Bourgain–Gamburd mechanism from first principles. This volume therefore opens those black boxes completely in the finite discrete setting needed for expansion.

The first branch of the logical chain is

$$\begin{aligned} \text{Ruzsa calculus} &\implies \text{small tripling} \\ &\implies \text{approximate groups} \\ &\implies \text{small-product control.} \end{aligned}$$

Independently, the energy branch is

$$\text{large energy} \implies \text{weak BSG} \implies \text{full BSG.}$$

The two chains then meet in the energy–approximate-group equivalence. A complete dyadic argument extends the set inverse theorem to arbitrary probability measures, and the final chapter proves the abstract L^2 -flattening criterion.

The policy of this series is strict: Tao’s product-set paper is cited for attribution and historical organization, but every theorem from it that is essential to the flattening mechanism is proved in this volume. [1] **Series position.** This volume uses Volume I and elementary group theory. Its main outputs are noncommutative Ruzsa calculus, BSG theory, approximate groups, and the flattening inverse mechanism.

Quantifier and asymptotic conventions. Unless a statement explicitly says otherwise, every constant occurring in a quantitative theorem is chosen in the order in which it is quantified in that theorem. Thus a phrase such as “for all sufficiently small δ ” means: there is a displayed or locally quantified threshold $\delta_0 > 0$, depending only on the parameters listed before that phrase, such that the assertion holds for every $0 < \delta < \delta_0$. The phrases “sufficiently large N ” and “sufficiently large p ” have the analogous meaning.

If $X \ll_\theta Y$ or $X = O_\theta(Y)$, the implicit constant depends only on θ and on the fixed ambient data of the current volume; it is independent of the variables quantified after θ . The words *absolute*, *uniform*, and *bounded* never introduce an unlisted dependence. Whenever an

inequality between exponents or scales is essential to the argument, that inequality is written explicitly in the proof; these conventions are not a substitute for such a calculation.

Contents

Preface	iii
Part 1. Ruzsa Calculus and Approximate Groups	1
Chapter 1. Ruzsa Calculus and Multiplicative Energy	2
1. Product sets and representation functions	2
2. Multiplicative energy	3
3. Energy of uniform measures	4
4. The Ruzsa triangle inequality	4
5. Ruzsa distance	5
6. The Ruzsa covering lemma	5
7. Popular products	6
8. Energy and quotient sets	7
9. Small product set implies controlled quotient sets	7
10. Why energy is the correct inverse statistic for flattening	7
11. A worked collision calculation	8
12. Exercises and mastery checks	9
13. Chapter summary	10
Chapter 2. Small Tripling and Approximate Groups	11
1. The left Ruzsa distance revisited	11
2. Every bounded signed word is controlled	11
3. Approximate groups	13
4. Small tripling produces an approximate group	14
5. Why this theorem is enough for flattening	15
6. Dependencies and output	15
Chapter 3. Small Doubling, Almost-Periods, and Approximate-Group Control	16
1. Right almost-periods of a set	16
2. From a small product to one approximate group	18
3. Dependencies	20
Part 2. Balog–Szemerédi–Gowers Theory	21

Chapter 4. The Weak Balog–Szemerédi–Gowers Lemma	22
1. Partial product control	22
2. What has been extracted?	24
3. A concrete finite-group check	25
4. Dependencies	25
Chapter 5. The Full Noncommutative Balog–Szemerédi–Gowers Theorem	26
1. Statement	26
1.1. Stage I: popular products carry a fixed share of the energy	27
1.2. Stage II: regularize the A -side	27
1.3. Stage III: apply the weak BSG lemma and prune bad quotient degree	28
1.4. Stage IV: regularize the B -side and bound the complete product set	29
2. The proof architecture after the calculation	30
3. Dependencies	30
Chapter 6. Energy, Small Product Sets, and Approximate Groups: The Exact Equivalence	31
1. The four manifestations of multiplicative structure	31
2. The one-set version	33
3. Uniform probability measures	33
4. How the inverse step is assembled	35
Part 3. Measures and Flattening	36
Chapter 7. From General Measures to Approximate Groups	37
1. Truncating tiny atoms	37
2. Dyadic levels	38
3. A form tailored to Bourgain–Gamburd	41
4. Dependencies	41
Chapter 8. The L^2 -Flattening Theorem	42
1. Power-scale form of the weighted inverse theorem	42
2. The algebraic and probabilistic hypotheses	43
3. Structural flattening	43
4. Iteration and logarithmic time	45
5. A failure mode worth remembering	45
6. Dependencies	46
Part 4. Technical Laboratories	47
Appendix A. Dense-Graph Mechanics Behind Balog–Szemerédi–Gowers	48
1. Restricted product graphs	48
2. Popularity from energy	48
3. Pruning without losing the density scale	49

4. Common neighbors encode quotients	50
5. Why “most pairs are good” is not yet BSG	50
6. The main proof translated into graph language	51
7. A subgroup model	51
8. What the graph proof teaches	51
Appendix B. Dyadic Measures, Effective Support, and the Anti-Atom Condition	52
1. Effective support is an averaged cardinality	52
2. A two-level calculation	53
3. What the anti-atom inequality forbids	53
4. Why only logarithmically many levels occur	54
5. A numerical laboratory	54
Appendix C. An Exponent Ledger for the Flattening Machine	55
1. The four exponents	55
2. Entering the medium interval	56
3. Subgroup escape must dominate the capture loss	56
4. How many flattening stages?	56
5. A sample consistent choice	57
Bibliography	58

Part 1

Ruzsa Calculus and Approximate Groups

CHAPTER 1

Ruzsa Calculus and Multiplicative Energy

Failure of convolution to spread is a statement about multiplicative coincidences. The language that records those coincidences is multiplicative energy; the language that manipulates the resulting small product sets is Ruzsa calculus. This chapter develops both in an arbitrary finite, possibly nonabelian, group. We prove the Ruzsa triangle inequality and covering lemma, establish the basic energy identities, and explain carefully which familiar abelian simplifications fail once order matters.

GUIDING IDEA 1.1 (Energy is a count of collisions). The multiplicative energy

$$E(A, B) = \#\{(a, b, a', b') : \\ a, a' \in A, \quad b, b' \in B, \\ ab = a'b'\}.$$

should be read literally as the number of collisions of the multiplication map $A \times B \rightarrow G$. If AB is small, many pairs must collide; conversely, abnormally many collisions suggest that a large part of $A \times B$ is governed by a low-complexity multiplication pattern. Balog–Szemerédi–Gowers theory is the theorem that turns this suggestion into a structured subset.

Ruzsa calculus supplies the deterministic bookkeeping around that inverse theorem. It lets us pass from control of one product or quotient set to control of several related ones without pretending that multiplication is commutative. The order of factors in every displayed set is therefore part of the statement. Whenever a familiar abelian identity seems to be used, one should check whether it survives with the factors in exactly that order.

1. Product sets and representation functions

For finite nonempty sets $A, B \subseteq G$, write

$$AB = \{ab : a \in A, b \in B\}, \quad A^{-1} = \{a^{-1} : a \in A\}.$$

For $x \in G$, define

$$r_{A,B}(x) = |\{(a, b) \in A \times B : ab = x\}|.$$

Then

$$(1) \quad r_{A,B} = 1_A * 1_B.$$

Two immediate identities are

$$\sum_x r_{A,B}(x) = |A||B|$$

and

$$\text{supp } r_{A,B} = AB.$$

2. Multiplicative energy

DEFINITION 1.2 (Multiplicative energy). The multiplicative energy of A and B is

$$(2) \quad E(A, B) = |\{(a_1, b_1, a_2, b_2) \in A \times B \times A \times B : a_1 b_1 = a_2 b_2\}|.$$

Write $E(A) = E(A, A)$.

PROPOSITION 1.3 (Energy identities). *For finite A, B ,*

$$(3) \quad E(A, B) = \sum_x r_{A,B}(x)^2 = \|1_A * 1_B\|_2^2.$$

Also

$$(4) \quad E(A, B) = |\{(a_1, a_2, b_1, b_2) : a_2^{-1} a_1 = b_2 b_1^{-1}\}|.$$

PROOF. The first equality groups solutions to $a_1 b_1 = a_2 b_2$ by their common product. The second is [Equation \(1\)](#). Rearranging the equality $a_1 b_1 = a_2 b_2$ gives $a_2^{-1} a_1 = b_2 b_1^{-1}$; no commutativity is used. $\square$

PROPOSITION 1.4 (Cauchy–Schwarz energy lower bound). *For finite nonempty A, B ,*

$$(5) \quad E(A, B) \geq \frac{|A|^2 |B|^2}{|AB|}.$$

In particular, small product set implies large energy.

PROOF. By Cauchy–Schwarz,

$$|A|^2 |B|^2 = \left(\sum_{x \in AB} r_{A,B}(x) \right)^2 \leq |AB| \sum_x r_{A,B}(x)^2.$$

$\square$

The converse is false without passing to large subsets: a set can have large energy because a substantial portion of it is structured while the rest is arbitrary. The Balog–Szemerédi–Gowers theorem is precisely the extraction theorem that repairs this failure.

3. Energy of uniform measures

For nonempty A, B ,

$$(6) \quad \|\mu_A * \mu_B\|_2^2 = \frac{E(A, B)}{|A|^2|B|^2}.$$

If $|A| \asymp |B| \asymp M$, then the non-flattening condition

$$\|\mu_A * \mu_B\|_2 \gg M^{-1/2}$$

is equivalent, up to constants, to

$$E(A, B) \gg M^3,$$

the maximal possible energy scale.

EXAMPLE 1.5 (Subgroup energy). If $A = H$ is a subgroup, then every triple (h_1, h_2, h_3) determines a unique h_4 satisfying $h_1 h_2 = h_3 h_4$. Hence

$$E(H) = |H|^3,$$

which is maximal. Correspondingly $\mu_H * \mu_H = \mu_H$.

4. The Ruzsa triangle inequality

The noncommutative Ruzsa triangle inequality has exactly the same cardinality form as in the abelian setting, provided the order of factors is respected.

THEOREM 1.6 (Ruzsa triangle inequality). *For finite nonempty $A, B, C \subseteq G$,*

$$(7) \quad |AC^{-1}| |B| \leq |AB^{-1}| |BC^{-1}|.$$

PROOF. For each $x \in AC^{-1}$ choose once and for all a representation

$$x = a_x c_x^{-1} \quad (a_x \in A, c_x \in C).$$

Define

$$\Phi : (AC^{-1}) \times B \longrightarrow (AB^{-1}) \times (BC^{-1})$$

by

$$\Phi(x, b) = (a_x b^{-1}, b c_x^{-1}).$$

The product of the two coordinates is $a_x c_x^{-1} = x$, so the image determines x . Once x is known, the chosen a_x is known, and the first coordinate determines

$$b^{-1} = a_x^{-1} (a_x b^{-1}).$$

Thus Φ is injective. Counting domain and codomain proves the inequality. $\square$

COROLLARY 1.7 (Some useful specializations). *For finite nonempty A, B ,*

$$(8) \quad |AA^{-1}| \leq \frac{|AB^{-1}|^2}{|B|},$$

$$(9) \quad |A^{-1}A| \leq \frac{|B^{-1}A|^2}{|B|}.$$

PROOF. Apply [Theorem 1.6](#) with $C = A$ for the first inequality and then invert/relabel for the second. $\square$

5. Ruzsa distance

Define

$$(10) \quad d(A, B) = \log \frac{|AB^{-1}|}{\sqrt{|A||B|}}.$$

Since $(AB^{-1})^{-1} = BA^{-1}$, this quantity is symmetric. The Ruzsa triangle inequality gives

$$d(A, C) \leq d(A, B) + d(B, C).$$

However,

$$d(A, A) = \log \frac{|AA^{-1}|}{|A|}$$

need not vanish. Thus d is not literally a metric; it is a useful multiplicative measure of relative product growth.

WARNING 1.8 (Order matters). In an abelian group the sets $A - B$, $B - A$, $A + B$, and their translates are tightly interchangeable. In a nonabelian group the expressions

$$AB^{-1}, \quad A^{-1}B, \quad BA^{-1}, \quad B^{-1}A$$

can have different geometry even when some have the same cardinality by inversion. A noncommutative proof should preserve the order of factors until an explicit bijection justifies changing it.

6. The Ruzsa covering lemma

THEOREM 1.9 (Ruzsa covering). *Let A, B be finite nonempty subsets of a group and suppose*

$$|AB| \leq K|B|.$$

Then there is $X \subseteq A$ with $|X| \leq K$ such that

$$(11) \quad A \subseteq XBB^{-1}.$$

PROOF. Choose $X \subseteq A$ maximal subject to the translates xB , $x \in X$, being pairwise disjoint. Since all these translates lie in AB ,

$$|X||B| \leq |AB| \leq K|B|,$$

so $|X| \leq K$.

For any $a \in A$, maximality gives some $x \in X$ such that $aB \cap xB \neq \emptyset$. Hence $ab_1 = xb_2$ for some $b_1, b_2 \in B$, so

$$a = xb_2b_1^{-1} \in XBB^{-1}.$$

$\square$

REMARK 1.10. The conclusion covers A , not AB , by translates of the symmetric quotient set BB^{-1} . This asymmetry is exactly what makes the lemma effective in noncommutative arguments.

COROLLARY 1.11 (Right-handed Ruzsa covering). *If A, B are finite nonempty subsets of a group and*

$$|AB| \leq K|A|,$$

then there is $Y \subseteq B$ with $|Y| \leq K$ such that

$$B \subseteq A^{-1}AY.$$

PROOF. Choose $Y \subseteq B$ maximal subject to the left translates Ay , $y \in Y$, being pairwise disjoint. They all lie in AB , so

$$|Y||A| \leq |AB| \leq K|A|,$$

and therefore $|Y| \leq K$. For any $b \in B$, maximality gives $y \in Y$ with

$$Ab \cap Ay \neq \emptyset.$$

Thus $a_1b = a_2y$ for some $a_1, a_2 \in A$, whence

$$b = a_1^{-1}a_2y \in A^{-1}AY.$$

□

7. Popular products

If energy is large, a substantial portion of the energy must occur on products with large representation multiplicity.

LEMMA 1.12 (Popularity truncation). *Let A, B be finite and set $M = \min(|A|, |B|)$. For $0 < \theta < 1$ define*

$$P_\theta = \{x \in AB : r_{A,B}(x) \geq \theta M\}.$$

Then

$$\sum_{x \notin P_\theta} r_{A,B}(x)^2 < \theta M|A||B|.$$

Consequently, if

$$E(A, B) \geq \eta M|A||B|,$$

then choosing $\theta = \eta/2$ gives

$$\sum_{x \in P_{\eta/2}} r_{A,B}(x)^2 \geq \frac{\eta}{2} M|A||B|.$$

Moreover

$$|P_\theta| \leq \frac{|A||B|}{\theta M}.$$

PROOF. For $x \notin P_\theta$,

$$r(x)^2 < \theta M r(x).$$

Summing and using $\sum_x r(x) = |A||B|$ gives the first estimate. The second follows by subtraction. Finally,

$$\theta M |P_\theta| \leq \sum_{x \in P_\theta} r(x) \leq |A||B|.$$

□

8. Energy and quotient sets

The equality form [Equation \(4\)](#) can be rewritten using quotient representation functions. Define

$$r_{A^{-1},A}(x) = |\{(a_1, a_2) \in A^2 : a_1^{-1}a_2 = x\}|.$$

Then

$$(12) \quad E(A, B) = \sum_x r_{A^{-1},A}(x) r_{B,B^{-1}}(x).$$

This form says that high energy means the left quotient geometry of A correlates strongly with the right quotient geometry of B .

9. Small product set implies controlled quotient sets

PROPOSITION 1.13 (A first Ruzsa consequence). *If $|AB^{-1}| \leq K\sqrt{|A||B|}$, then*

$$|AA^{-1}| \leq K^2|A|, \quad |BB^{-1}| \leq K^2|B|.$$

PROOF. The first follows from [Equation \(8\)](#):

$$|AA^{-1}| \leq \frac{|AB^{-1}|^2}{|B|} \leq K^2|A|.$$

The second follows by symmetry. □

This is the elementary beginning of the principle “small mixed product implies internal structure.” To get a genuine approximate group one needs a stronger package of product-set estimates.

10. Why energy is the correct inverse statistic for flattening

Suppose A is finite and μ_A uniform. Young gives

$$\|\mu_A * \mu_A\|_2 \leq \|\mu_A\|_2 = |A|^{-1/2}.$$

By [Equation \(6\)](#), near equality means

$$E(A) \approx |A|^3.$$

Thus failure of L^2 flattening is exactly near-maximal multiplicative energy for uniform measures. The nonuniform case is reduced to this picture by a dyadic decomposition, developed from scratch in Chapter 7.

GUIDING IDEA 1.14. The logical direction relevant to expansion is the inverse one:

large energy $\implies$ large structured subset $\implies$ approximate subgroup.

The first arrow is Balog–Szemerédi–Gowers. The second is Ruzsa/approximate-group theory.

11. A worked collision calculation

It is useful to see exactly what energy measures before invoking an inverse theorem. Let $A, B \subseteq G$ and choose independent uniform random variables

$$X, X' \in A, \quad Y, Y' \in B.$$

Then

$$\mathbb{P}(XY = X'Y') = \frac{E(A, B)}{|A|^2|B|^2} = \|\mu_A * \mu_B\|_2^2.$$

Thus the squared L^2 norm of the product distribution is literally a collision probability. Compare three extreme models.

- (i) If multiplication on $A \times B$ were perfectly injective, then $E(A, B) = |A||B|$ and

$$\|\mu_A * \mu_B\|_2 = (|A||B|)^{-1/2}.$$

This is the flattest distribution compatible with $|A||B|$ input pairs.

- (ii) If $A = B = H$ is a subgroup of size M , then XY is uniform on H and

$$E(H) = M^3, \quad \|\mu_H * \mu_H\|_2 = M^{-1/2}.$$

Convolution has produced no L^2 gain at all.

- (iii) If A and B both have size M and their product set has size roughly M^2 , with representations roughly equidistributed, then

$$E(A, B) \asymp M^2, \quad \|\mu_A * \mu_B\|_2 \asymp M^{-1}.$$

This is a full square-root improvement over the subgroup model.

The BSG regime lies between (ii) and (iii). If the collision probability is much larger than the random-looking scale, many pairs (a, b) participate in products having many representations. Popularity isolates these products; a dense graph records which pairs participate; the graph-theoretic extraction then finds large subsets on which quotient sets are controlled.

REMARK 1.15 (Why a small product set is not enough). Cauchy–Schwarz gives

$$E(A, B) \geq \frac{|A|^2|B|^2}{|AB|}.$$

So small AB implies high energy. The converse is false without discarding elements: a small portion of $A \times B$ may carry a disproportionate number of collisions. BSG is precisely the theorem that allows one to discard a controlled fraction of the vertices and recover a genuinely small-product core.

12. Exercises and mastery checks

EXERCISE 1.16 (Energy upper bound). Show

$$E(A, B) \leq |A||B| \min(|A|, |B|).$$

Characterize equality when $A = B = H$ is a subgroup, and investigate what equality forces in general.

EXERCISE 1.17 (Left-right energy identity). Starting from $a_1b_1 = a_2b_2$, derive at least three equivalent quotient identities. Record the order of factors carefully and identify which ones are related by inversion.

EXERCISE 1.18 (Ruzsa triangle as injection). Reconstruct the injection in [Theorem 1.6](#) without looking at the proof. Explain why choosing one representation $x = a_x c_x^{-1}$ for each x is necessary.

EXERCISE 1.19 (Covering with a symmetric set). Under the hypothesis $|AB| \leq K|B|$, show that A is contained in at most K left translates of the symmetric set BB^{-1} . Give an example where replacing BB^{-1} by B is impossible.

EXERCISE 1.20 (Energy and convolution). Let A, B have comparable sizes $M \leq |A|, |B| \leq 2M$. Show that

$$\|\mu_A * \mu_B\|_2 \geq K^{-1} M^{-1/2}$$

implies

$$E(A, B) \geq cK^{-2}M^3$$

with an explicit absolute $c > 0$.

CHECKPOINT 1.21. The reader should now be able to recognize large energy as the exact combinatorial signature of non-flattening for uniform measures, and should be able to move from a small mixed product set to small quotient sets using only Ruzsa’s triangle inequality. What remains is the extraction step from large energy to small product structure.

13. Chapter summary

Multiplicative energy is the squared ℓ^2 norm of a set convolution:

$$E(A, B) = \|1_A * 1_B\|_2^2.$$

Small product sets force high energy by Cauchy–Schwarz. Ruzsa’s triangle inequality

$$|AC^{-1}||B| \leq |AB^{-1}||BC^{-1}|$$

and the covering lemma convert cardinality bounds into controlled quotient sets and coverings. These elementary tools are insensitive to the detailed algebraic nature of the ambient group and therefore belong to the reusable core of the Bourgain–Gamburd method.

CHAPTER 2

Small Tripling and Approximate Groups

In a noncommutative group, small doubling does not control higher products: a set may satisfy $|A^2| \asymp |A|$ while A^3 is much larger. Small tripling is the stable threshold. This chapter proves, from the Ruzsa triangle inequality and covering lemma alone, that $|A^3| \leq K|A|$ controls every bounded signed word in A and places A inside an approximate group of polynomial size. The proof is given completely because this stability statement will be used repeatedly later.

GUIDING IDEA 2.1 (Why tripling, rather than doubling?). If H is a subgroup and x does not normalize H , then $A = H \cup \{x\}$ can have $|A^2| = O(|A|)$ while the word HxH already contributes many new elements to A^3 . Thus a noncommutative Plünnecke theory cannot begin from small doubling alone. The hypothesis $|A^3| \leq K|A|$ is stronger: it simultaneously controls enough left- and right-handed quotients to make the Ruzsa triangle inequality iterable.

1. The left Ruzsa distance revisited

For finite nonempty subsets X, Y of a group G , define

$$d(X, Y) = \log \frac{|XY^{-1}|}{\sqrt{|X||Y|}}.$$

By [Theorem 1.6](#), d is symmetric and obeys the triangle inequality. It is left-translation invariant, but generally neither right-translation invariant nor invariant under inversion of both arguments.

The point of the logarithm is that a polynomial cardinality loss becomes an additive $O(\log K)$ error. We shall repeatedly use the elementary equivalence

$$(13) \quad d(X, Y) \leq C \log K \iff |XY^{-1}| \leq K^C \sqrt{|X||Y|}.$$

2. Every bounded signed word is controlled

THEOREM 2.2 (Signed-word control from small tripling). *Let $A \subseteq G$ be finite and nonempty and suppose*

$$(14) \quad |A^3| \leq K|A|, \quad K \geq 1.$$

For every fixed $m \geq 1$ and every choice $\varepsilon_1, \dots, \varepsilon_m \in \{-1, 1\}$,

$$(15) \quad |A^{\varepsilon_1} \dots A^{\varepsilon_m}| \leq K^{C_m} |A|$$

for a constant C_m depending only on m .

PROOF. Since multiplication by a fixed element is a bijection,

$$|A| \leq |A^2| \leq |A^3| \leq K|A|.$$

We first establish every signed word of length at most three. This base case is where factor order matters, so we record each of the three nontrivial patterns rather than hiding them in “and similarly.”

First,

$$d(A, A^{-1}) = \log \frac{|A^2|}{|A|} \leq \log K,$$

and

$$d(A^2, A^{-1}) = \log \frac{|A^3|}{\sqrt{|A^2||A|}} \leq \log K.$$

The triangle inequality therefore gives

$$d(A^2, A) \leq 2 \log K,$$

which, after converting the distance back to a cardinality, yields

$$(16) \quad |A^2 A^{-1}| \leq K^{O(1)} |A|.$$

Equivalently $d(A, AA^{-1}) = O(\log K)$, because

$$A(AA^{-1})^{-1} = A^2 A^{-1}.$$

Hence another triangle inequality gives

$$d(AA^{-1}, A^{-1}) = O(\log K),$$

and therefore

$$(17) \quad |AA^{-1}A| \leq K^{O(1)} |A|.$$

The third genuinely different order is $A^{-1}AA$. To obtain it without an illicit commutation, observe directly that

$$d(A, A^{-1}A^{-1}) = \log \frac{|A^3|}{\sqrt{|A||A^2|}} \leq \log K.$$

Together with $d(A^{-1}, A) = d(A, A^{-1}) \leq \log K$ this gives

$$d(A^{-1}, A^{-1}A^{-1}) \leq 2 \log K.$$

But

$$A^{-1}(A^{-1}A^{-1})^{-1} = A^{-1}AA,$$

so

$$(18) \quad |A^{-1}AA| \leq K^{O(1)}|A|.$$

Taking inverses of (16)–(18), together with the original bound for A^3 , gives all eight sign patterns of length three. The length-one and length-two cases follow by fixing unused factors, so the theorem is proved for $m \leq 3$.

We now induct on $m \geq 4$. Suppose the assertion is known for every word of length smaller than m . Put

$$X = A^{\varepsilon_1} \dots A^{\varepsilon_{m-2}}, \quad Y = A^{-\varepsilon_m} A^{-\varepsilon_{m-1}}.$$

The desired product is XY^{-1} . By the inductive hypothesis, $|X|$ and $|Y|$ are both $K^{O_m(1)}|A|$. More importantly,

$$d(X, A) = O_m(\log K),$$

because XA^{-1} is a signed word of length $m - 1$, while

$$d(A, Y) = O(\log K)$$

because AY^{-1} is a signed word of length three. Thus

$$d(X, Y) \leq d(X, A) + d(A, Y) = O_m(\log K).$$

Converting back to cardinalities gives

$$|XY^{-1}| \leq K^{O_m(1)} \sqrt{|X||Y|} \leq K^{O_m(1)}|A|,$$

which is exactly $|A^{\varepsilon_1} \dots A^{\varepsilon_m}| \leq K^{O_m(1)}|A|$. This closes the induction. $\square$

REMARK 2.3 (What the induction actually does). The proof is not a disguised commutative Plünnecke inequality. At each step we insert the one fixed anchor A into the Ruzsa triangle inequality. The length-three calculation is therefore the irreducible base case: without simultaneous control of all signed words of length three, the induction would have no stable anchor.

3. Approximate groups

DEFINITION 2.4 (Approximate group). A finite set $H \subseteq G$ is a K -approximate group if

- (i) $e \in H$ and $H^{-1} = H$;
- (ii) there exists a symmetric set $X \subseteq H^2$ with $|X| \leq K$ such that

$$H^2 \subseteq XH.$$

Because H and X are symmetric, inversion also gives $H^2 \subseteq HX$.

LEMMA 2.5 (Bounded powers of an approximate group). *If H is a K -approximate group, then for every $n \geq 1$,*

$$H^n \subseteq X^{n-1}H \quad \text{and} \quad |H^n| \leq K^{n-1}|H|.$$

PROOF. Induct on n . If $H^n \subseteq X^{n-1}H$, then

$$H^{n+1} \subseteq X^{n-1}H^2 \subseteq X^nH.$$

The cardinality estimate follows from $|X^{n-1}| \leq |X|^{n-1}$. $\square$

4. Small tripling produces an approximate group

THEOREM 2.6 (Small tripling $\Rightarrow$ approximate group). *Assume (14). Set*

$$H_0 = A \cup A^{-1} \cup \{e\}, \quad H = H_0^3.$$

Then

$$|H| \leq K^{O(1)}|A|,$$

$A \subseteq H$, and H is a $K^{O(1)}$ -approximate group. In particular, the polynomial approximate-group conclusion is a consequence of the elementary Ruzsa calculus proved in this volume; no structural classification theorem is being used here.

PROOF. By Theorem 2.2, every product obtained from a bounded number of copies of A , A^{-1} and $\{e\}$ has size $K^{O(1)}|A|$. Hence

$$|H| = |H_0^3| \leq K^{O(1)}|A|$$

and also

$$(19) \quad |H_0H^2| = |H_0^7| \leq K^{O(1)}|A|.$$

Since $|H_0| \geq |A|$, (19) gives

$$|H_0H^2| \leq K^{O(1)}|H_0|.$$

Apply the right-handed Ruzsa covering lemma, Corollary 1.11, to the pair (H_0, H^2) . There is a set $Y \subseteq H^2$ with $|Y| \leq K^{O(1)}$ such that

$$H^2 \subseteq H_0^{-1}H_0Y.$$

Because H_0 is symmetric and $H_0^2 \subseteq H_0^3 = H$,

$$H^2 \subseteq HY.$$

Taking inverses gives $H^2 \subseteq Y^{-1}H$. Put

$$X = Y \cup Y^{-1} \cup \{e\}.$$

Then X is symmetric, $|X| \leq K^{O(1)}$, and $H^2 \subseteq XH \cap HX$.

It remains only to ensure the conventional requirement $X \subseteq H^2$. We chose $Y \subseteq H^2$, hence $Y^{-1} \subseteq H^2$ because H is symmetric, and $e \in H^2$. Thus $X \subseteq H^2$. Therefore H is a $K^{O(1)}$ -approximate group. $\square$

COROLLARY 2.7 (Symmetric form). *If $A = A^{-1}$, $e \in A$, and $|A^3| \leq K|A|$, then A^3 is a $K^{O(1)}$ -approximate group and every bounded power A^m has cardinality $K^{O_m(1)}|A|$.*

PROOF. Here $H_0 = A$, so [Theorem 2.6](#) gives the first assertion. The second already follows directly from [Theorem 2.2](#); alternatively use [Lemma 2.5](#) for A^3 . $\square$

5. Why this theorem is enough for flattening

The theorem above is intentionally weaker than the Breuillard–Green–Tao structure theorem for arbitrary approximate groups. The Bourgain–Gamburd inverse step does not need a classification at this stage. It needs only the implication

small tripling $\implies$ a symmetric object with controlled bounded powers.

The ambient product theorem—for $\mathrm{SL}_2(\mathbb{F}_p)$ in the original argument—will later decide what such an object can be.

6. Dependencies and output

The proof of [Theorem 2.6](#) uses only:

- (1) the Ruzsa triangle inequality, proved in [Chapter 1](#);
- (2) the Ruzsa covering lemma, proved there as well;
- (3) elementary induction and cardinality estimates.

Thus the small-tripling-to-approximate-group step is now fully internal to the series. Tao’s product-set paper remains the historical source for this organization of the argument, but it is no longer a proof dependency. [\[1\]](#)

CHAPTER 3

Small Doubling, Almost-Periods, and Approximate-Group Control

Small tripling directly produces an approximate group, but the Balog–Szemerédi–Gowers argument naturally creates sets with a small product set or a small quotient set. This chapter proves the missing bridge. Starting from $|AA^{-1}| \leq K|A|$, we construct many right almost-periods of A , prove quantitative control of $AS^n A^{-1}$ for every fixed n , and use the small-tripling theorem to build an approximate group. We then deduce the two-set theorem: if $|AB|$ is small on the geometric-mean scale, both A and B are efficiently covered by translates of one approximate group.

1. Right almost-periods of a set

For $x \in G$ put

$$\rho_A(x) = |A \cap Ax|.$$

The function ρ_A measures how much right multiplication by x^{-1} preserves A . It satisfies

$$(20) \quad \sum_{x \in G} \rho_A(x) = |A|^2,$$

because every ordered pair $(a, a') \in A^2$ contributes to the unique quotient $x = a'^{-1}a$ for which $a \in Ax$.

A second identity is

$$(21) \quad \sum_x \rho_A(x)^2 = E(A^{-1}, A) = E(A, A^{-1}).$$

The last equality follows by the bijection

$$a_1^{-1}a_2 = a_3^{-1}a_4 \iff a_2a_4^{-1} = a_1a_3^{-1}.$$

This elementary identity is the weak substitute for commutativity that makes the argument work.

PROPOSITION 3.1 (A large set of almost-periods). *Let $A \subseteq G$ be finite and nonempty and suppose*

$$(22) \quad |AA^{-1}| \leq K|A|.$$

Define

$$S = \left\{ x \in G : \rho_A(x) > \frac{|A|}{2K} \right\}.$$

Then $S = S^{-1}$,

$$(23) \quad |S| \geq \frac{|A|}{2K},$$

and for every integer $n \geq 1$,

$$(24) \quad |AS^n A^{-1}| \leq 2^n K^{2n+1} |A|.$$

PROOF. The symmetry follows from

$$|A \cap Ax| = |Ax^{-1} \cap A| = |A \cap Ax^{-1}|.$$

By Proposition 1.4 applied to (A, A^{-1}) and (22),

$$\sum_x \rho_A(x)^2 = E(A, A^{-1}) \geq \frac{|A|^4}{|AA^{-1}|} \geq \frac{|A|^3}{K}.$$

On $G \setminus S$ we have $\rho_A(x) \leq |A|/(2K)$; hence by (20),

$$\sum_{x \notin S} \rho_A(x)^2 \leq \frac{|A|}{2K} \sum_x \rho_A(x) = \frac{|A|^3}{2K}.$$

Thus

$$\sum_{x \in S} \rho_A(x)^2 \geq \frac{|A|^3}{2K}.$$

Since $\rho_A(x) \leq |A|$, this gives

$$|S||A|^2 \geq \frac{|A|^3}{2K},$$

which is (23).

We turn to (24). Consider

$$Q = \sum_{y_0, \dots, y_n \in AA^{-1}} 1_{AS^n A^{-1}}(y_0 \cdots y_n).$$

The trivial upper bound is

$$(25) \quad Q \leq |AA^{-1}|^{n+1} \leq K^{n+1} |A|^{n+1}.$$

We shall prove the lower bound

$$(26) \quad Q \geq |AS^n A^{-1}| \left(\frac{|A|}{2K} \right)^n.$$

Fix $x \in AS^n A^{-1}$. Write

$$x = a_0 s_1 \cdots s_n b_{n+1}^{-1}$$

with $a_0, b_{n+1} \in A$ and $s_i \in S$. Choose $b_1, \dots, b_n \in A$ such that also $b_i s_i \in A$. There are exactly

$$\prod_{i=1}^n |A \cap A s_i| > \left(\frac{|A|}{2K} \right)^n$$

such choices. For each choice put

$$y_0 = a_0 b_1^{-1}, \quad y_i = b_i s_i b_{i+1}^{-1} \quad (1 \leq i \leq n-1), \quad y_n = b_n s_n b_{n+1}^{-1}.$$

Every y_i lies in AA^{-1} , and telescoping gives $y_0 \cdots y_n = x$. Moreover the tuple $(y_0, \dots, y_n)$ determines the successive b_i once the fixed representation of x is chosen, so distinct choices give distinct tuples. Hence each x contributes more than $(|A|/(2K))^n$ to Q , proving (26).

Comparing (25) and (26) yields

$$|AS^n A^{-1}| \leq \frac{K^{n+1} |A|^{n+1}}{(|A|/(2K))^n} = 2^n K^{2n+1} |A|.$$

□

GUIDING IDEA 3.2 (Why the telescoping count matters). The proposition says that a small quotient set AA^{-1} creates many transformations s for which A and As overlap substantially. One such s is weak information. The telescoping construction shows that a string $s_1 \cdots s_n$ can be threaded through independent overlap choices inside A ; the huge number of threads forces the ambient set $AS^n A^{-1}$ to stay small. This is the noncommutative replacement for the easier additive statement that many almost-periods add together.

2. From a small product to one approximate group

THEOREM 3.3 (Small product set theorem). *Let $A, B \subseteq G$ be finite and nonempty and suppose*

$$(27) \quad |AB| \leq K \sqrt{|A||B|}.$$

Then there is a $K^{O(1)}$ -approximate group H satisfying

$$(28) \quad |H| \leq K^{O(1)} \sqrt{|A||B|},$$

and sets X, Y with $|X|, |Y| \leq K^{O(1)}$ such that

$$(29) \quad A \subseteq XH, \quad B \subseteq HY.$$

Equivalently, after replacing $X \cup Y^{-1}$ by one set of polynomial size, both A and B are controlled by left/right translates of the same approximate group.

PROOF. The hypothesis is exactly

$$d(A, B^{-1}) \leq \log K.$$

By the Ruzsa triangle inequality,

$$d(A, A) \leq 2d(A, B^{-1}) \leq 2 \log K,$$

so

$$(30) \quad |AA^{-1}| \leq K^2|A|.$$

Apply [Proposition 3.1](#) with K^2 in place of K . We obtain a symmetric set S such that

$$|S| \geq \frac{|A|}{2K^2}$$

and

$$|AS^3A^{-1}| \leq K^{O(1)}|A|.$$

In particular

$$|S^3| \leq K^{O(1)}|A| \leq K^{O(1)}|S|.$$

By [Theorem 2.6](#), there exists a $K^{O(1)}$ -approximate group H containing S with

$$(31) \quad |H| \leq K^{O(1)}|S| \leq K^{O(1)}|A|.$$

Since $S \subseteq H$ and H^2 has polynomial size,

$$|SH| \leq |H^2| \leq K^{O(1)}|H| \leq K^{O(1)}|A|.$$

Thus

$$d(S, H) = O(\log K).$$

To obtain the corresponding estimate between A and S , one small point should be made explicit. Choose fixed elements $s_2, s_3 \in S$ and $a_0 \in A$ (all these sets are nonempty). Right multiplication by the fixed element $s_2s_3a_0^{-1}$ is injective, and

$$(AS)s_2s_3a_0^{-1} \subseteq AS^3A^{-1}.$$

Hence

$$|AS| \leq |AS^3A^{-1}| \leq K^{O(1)}|A|.$$

Since $S = S^{-1}$ and $|S| \geq K^{-O(1)}|A|$, this gives

$$d(A, S) = O(\log K).$$

Hence

$$d(A, H) = O(\log K).$$

Therefore

$$(32) \quad |AH| \leq K^{O(1)}\sqrt{|A||H|} = K^{O(1)}|A|,$$

where we used (31) and the reverse bound $|H| \geq |S| \geq K^{-O(1)}|A|$.

Apply Ruzsa covering to (A, H) using (32). There is $X_0 \subseteq A$ with $|X_0| \leq K^{O(1)}$ such that

$$A \subseteq X_0HH^{-1} = X_0H^2.$$

Because H is an approximate group, H^2 is covered by $K^{O(1)}$ left translates of H ; absorbing those translates into X_0 gives

$$A \subseteq XH, \quad |X| \leq K^{O(1)}.$$

We now treat B . From $d(A, B^{-1}) \leq \log K$ and $d(A, H) = O(\log K)$,

$$d(B^{-1}, H) = O(\log K).$$

Thus $|B^{-1}H| \leq K^{O(1)}|B|$, after noting that the triangle inequalities also imply $|A|$ and $|B|$ are polynomially comparable: indeed $|AB| \geq \max(|A|, |B|)$ and (27) give

$$K^{-2}|A| \leq |B| \leq K^2|A|.$$

Apply Ruzsa covering to (B^{-1}, H) to obtain

$$B^{-1} \subseteq Y_0 H^2$$

with $|Y_0| \leq K^{O(1)}$. Invert and absorb the bounded-power cover of H^2 to get

$$B \subseteq HY$$

with $|Y| \leq K^{O(1)}$.

Finally (31) and $|A| \leq K\sqrt{|A||B|}$ imply (28) after enlarging the polynomial exponent. $\square$

COROLLARY 3.4 (Small doubling). *If $|A^2| \leq K|A|$, then A is covered by $K^{O(1)}$ left translates and $K^{O(1)}$ right translates of a $K^{O(1)}$ -approximate group H with $|H| \leq K^{O(1)}|A|$.*

PROOF. Apply [Theorem 3.3](#) with $B = A$. $\square$

3. Dependencies

The entire chapter is proved from the material already internal to the book:

- (1) energy and Cauchy–Schwarz from [Chapter 1](#);
- (2) Ruzsa triangle and covering from the same chapter;
- (3) the fully proved small-tripling theorem [Theorem 2.6](#).

Thus the small-product/approximate-group theorem, historically organized in this form by Tao [\[1\]](#), is no longer an external input.

Part 2

Balog–Szemerédi–Gowers Theory

CHAPTER 4

The Weak Balog–Szemerédi–Gowers Lemma

The first BSG extraction starts from a dense set of pairs whose products all lie in a relatively small set C . Two Cauchy–Schwarz/pigeonhole steps produce a large subset A' in which almost every quotient aa'^{-1} belongs to a set D of controlled size. We prove this carefully, with the constants visible. No graph-extraction theorem is hidden here: the argument is a direct counting proof.

1. Partial product control

Let $A, B, C \subseteq G$ be finite and define the edge set

$$E_C(A, B) = \{(a, b) \in A \times B : ab \in C\}.$$

The weak BSG lemma starts from two quantitative facts:

$$(33) \quad |C| \leq K' \sqrt{|A||B|}, \quad |E_C(A, B)| \geq \frac{|A||B|}{K}.$$

The first says that the allowed products live in a small target; the second says that the restricted multiplication graph is dense.

THEOREM 4.1 (Weak noncommutative BSG). *Assume (33), with $K, K' \geq 1$, and let $0 < \eta < 1$. Then there exist $A' \subseteq A$ and $D \subseteq G$ such that*

$$(34) \quad |A'| \geq \frac{|A|}{\sqrt{2} K},$$

$$(35) \quad |D| \leq \frac{2(KK')^2}{\eta} |A|,$$

and at least a proportion $1 - \eta$ of the ordered pairs $(a, a') \in A'^2$ satisfy

$$(36) \quad aa'^{-1} \in D.$$

Equivalently,

$$|\{(a, a') \in A'^2 : aa'^{-1} \notin D\}| \leq \eta |A'|^2.$$

PROOF. For each $b \in B$ define

$$m(b) = |\{a \in A : ab \in C\}|.$$

The density assumption gives

$$\sum_{b \in B} m(b) = |E_C(A, B)| \geq \frac{|A||B|}{K}.$$

Cauchy–Schwarz therefore yields

$$(37) \quad \sum_{b \in B} m(b)^2 \geq \frac{1}{|B|} \left(\sum_b m(b) \right)^2 \geq \frac{|A|^2|B|}{K^2}.$$

Expand the square:

$$(38) \quad \sum_{a, a' \in A} |\{b \in B : ab \in C, a'b \in C\}| \geq \frac{|A|^2|B|}{K^2}.$$

Call a pair $(a, a') \in A^2$ *bad* if

$$(39) \quad |\{b \in B : ab \in C, a'b \in C\}| \leq \frac{\eta}{2K^2}|B|.$$

Let Ω be the set of bad pairs. The total contribution of bad pairs to the left side of (38) is at most

$$|A|^2 \frac{\eta|B|}{2K^2}.$$

Subtracting η^{-1} times this bound from (38), we obtain

$$(40) \quad \sum_{b \in B} \sum_{a, a' \in A} \left(1 - \frac{1_{\Omega}(a, a')}{\eta} \right) 1_C(ab) 1_C(a'b) \geq \frac{|A|^2|B|}{2K^2}.$$

Hence some $b_0 \in B$ satisfies

$$(41) \quad \sum_{a, a' \in A} \left(1 - \frac{1_{\Omega}(a, a')}{\eta} \right) 1_C(ab_0) 1_C(a'b_0) \geq \frac{|A|^2}{2K^2}.$$

Set

$$A' = \{a \in A : ab_0 \in C\}.$$

Then the left side of (41) is

$$|A'|^2 - \frac{1}{\eta} |\Omega \cap (A' \times A')|.$$

In particular it is at most $|A'|^2$, so

$$|A'|^2 \geq \frac{|A|^2}{2K^2},$$

which proves (34). The same identity together with the nonnegativity of the right side gives

$$|\Omega \cap (A' \times A')| \leq \eta |A'|^2.$$

Thus all but an η fraction of pairs in A'^2 are good.

Define

$$D = \{aa'^{-1} : a, a' \in A', (a, a') \notin \Omega\}.$$

Then every good pair satisfies (36). It remains to bound $|D|$.

Fix $d \in D$ and choose $a, a' \in A'$ with $d = aa'^{-1}$ and $(a, a') \notin \Omega$. By the negation of (39),

$$|\{b \in B : ab \in C, a'b \in C\}| > \frac{\eta}{2K^2}|B|.$$

Put $c = a'b$. Then $c \in C$ and

$$dc = aa'^{-1}a'b = ab \in C.$$

Consequently every $d \in D$ has more than $\eta|B|/(2K^2)$ witnesses $c \in C$ to the relation $c, dc \in C$. Summing over d gives

$$(42) \quad |D| \frac{\eta|B|}{2K^2} < \sum_{d \in D} \sum_{c \in G} 1_C(c) 1_C(dc).$$

The right side is at most the same sum over all $d \in G$. But every ordered pair $(c, c') \in C^2$ determines exactly one $d = c'c^{-1}$ with $dc = c'$. Therefore

$$\sum_{d \in G} \sum_c 1_C(c) 1_C(dc) = |C|^2.$$

By (33),

$$|C|^2 \leq (K')^2 |A| |B|.$$

Combining with (42) yields

$$|D| \leq \frac{2K^2(K')^2}{\eta} |A|,$$

which is (35). □

2. What has been extracted?

The conclusion is intentionally asymmetric. We started with a dense graph between A and B and extracted one large subset $A' \subseteq A$. Inside A' almost every quotient aa'^{-1} lies in one controlled set D . This is not yet small doubling, because the exceptional pairs may be distributed adversarially. The full BSG theorem will prune the exceptional pairs and use the original product graph once more to force a genuinely small complete product set.

WARNING 4.2 (Why “almost every quotient is in D ” is not enough). Even if 99% of the pairs $(a, a') \in A'^2$ have $aa'^{-1} \in D$, the complete quotient set $A'A'^{-1}$ may still be enormous: the exceptional 1% may realize many distinct quotients. The next chapter does not simply discard those quotients. It selects vertices for which the exceptional degree is small and then intersects that information with another dense product graph.

3. A concrete finite-group check

If $A = B = H$ is a subgroup and $C = H$, then $K = K' = 1$. Every pair is an edge, and

$$|\{b \in H : ab \in H, a'b \in H\}| = |H|$$

for every a, a' . Thus $\Omega = \emptyset$, we may take $A' = H$, and $D = HH^{-1} = H$. The weak BSG lemma is therefore exactly sharp at the subgroup model: the structured target it discovers is the subgroup itself.

4. Dependencies

The proof above uses only Cauchy–Schwarz, double counting, Markov/pigeonhole arguments, and group cancellation. It is the discrete specialization of Tao’s weak noncommutative BSG lemma [1], but the citation is now historical rather than logical.

CHAPTER 5

The Full Noncommutative Balog–Szemerédi–Gowers Theorem

Large multiplicative energy means that many pairs share the same product. The full BSG theorem turns that averaged collision statement into large subsets with a small complete product set. We give the complete discrete proof. The argument has four stages: isolate popular products, regularize one side of the product graph, apply the weak BSG lemma to obtain many controlled quotients, and then intersect the two pieces of information so that every product in the final set has many factorizations through two small auxiliary sets.

1. Statement

THEOREM 5.1 (Noncommutative Balog–Szemerédi–Gowers). *There is an absolute constant C_0 such that the following holds. Let $A, B \subseteq G$ be finite and nonempty and suppose*

$$(43) \quad E(A, B) \geq K^{-1} |A|^{3/2} |B|^{3/2}, \quad K \geq 1.$$

Then there are subsets $A_* \subseteq A$ and $B_* \subseteq B$ satisfying

$$(44) \quad |A_*| \geq \frac{|A|}{C_0 K},$$

$$(45) \quad |B_*| \geq \frac{|B|}{C_0 K},$$

$$(46) \quad |A_* B_*| \leq C_0 K^8 \sqrt{|A| |B|}.$$

The exponent 8 is not important for the expansion machine; what matters is that the loss is polynomial in K . We keep an explicit power so that every polynomial loss remains visible.

PROOF. We proceed in four stages, separating the popularity, regularization, quotient-control, and final counting steps so that each polynomial loss is visible.

1.1. Stage I: popular products carry a fixed share of the energy. Write

$$r(x) = r_{A,B}(x) = |\{(a, b) \in A \times B : ab = x\}|.$$

We have

$$\sum_x r(x) = |A||B|, \quad r(x) \leq \min(|A|, |B|) \leq \sqrt{|A||B|}.$$

Define the popular-product set

$$(47) \quad C = \left\{ x : r(x) > \frac{\sqrt{|A||B|}}{2K} \right\}.$$

On $G \setminus C$,

$$r(x)^2 \leq \frac{\sqrt{|A||B|}}{2K} r(x),$$

so

$$\sum_{x \notin C} r(x)^2 \leq \frac{|A|^{3/2}|B|^{3/2}}{2K}.$$

By (43),

$$(48) \quad \sum_{x \in C} r(x)^2 \geq \frac{|A|^{3/2}|B|^{3/2}}{2K}.$$

Since $r(x) \leq \sqrt{|A||B|}$, this implies

$$(49) \quad \sum_{x \in C} r(x) \geq \frac{|A||B|}{2K}.$$

On the other hand, every $x \in C$ has $r(x) > \sqrt{|A||B|}/(2K)$, while the total number of pairs is $|A||B|$. Hence

$$(50) \quad |C| \leq 2K\sqrt{|A||B|}.$$

Thus a positive fraction of all pairs has product in a target set of essentially the smallest scale compatible with the energy hypothesis.

1.2. Stage II: regularize the A-side. For $a \in A$ let

$$d_C(a) = |\{b \in B : ab \in C\}|.$$

Equation (49) says

$$\sum_{a \in A} d_C(a) \geq \frac{|A||B|}{2K}.$$

Define

$$A' = \left\{ a \in A : d_C(a) > \frac{|B|}{4K} \right\}.$$

The contribution from $A \setminus A'$ is at most $|A||B|/(4K)$, hence

$$(51) \quad \sum_{a \in A'} d_C(a) \geq \frac{|A||B|}{4K}.$$

Since $d_C(a) \leq |B|$,

$$(52) \quad |A'| \geq \frac{|A|}{4K}.$$

Write

$$(53) \quad |A| = L|A'|, \quad 1 \leq L \leq 4K.$$

Then (50) becomes

$$(54) \quad |C| \leq 2KL^{1/2}\sqrt{|A'||B|},$$

and (51) becomes

$$(55) \quad |E_C(A', B)| \geq \frac{L}{4K}|A'||B|.$$

1.3. Stage III: apply the weak BSG lemma and prune bad quotient degree.

Apply Theorem 4.1 to A', B, C with

$$K_1 = \frac{4K}{L}, \quad K'_1 = 2KL^{1/2}, \quad \eta = \frac{1}{32K}.$$

The hypotheses are exactly (54) and (55). We obtain $A'' \subseteq A'$ and $D \subseteq G$ such that

$$(56) \quad |A''| \geq \frac{|A'|}{\sqrt{2K_1}} = \frac{L|A'|}{4\sqrt{2K}} = \frac{|A|}{4\sqrt{2K}},$$

and

$$(57) \quad |D| \leq CK^5 \frac{|A'|}{L}$$

for an absolute C . Since (56) gives

$$|A''| \geq \frac{L}{4\sqrt{2K}}|A'|,$$

we may rewrite (57) as the simpler polynomial bound

$$(58) \quad |D| \leq C_1 K^6 |A''|.$$

Moreover at most a proportion $1/(32K)$ of the ordered pairs $(a, a') \in (A'')^2$ have $aa'^{-1} \notin D$.

For $a \in A''$, define its bad quotient degree

$$\beta(a) = |\{a' \in A'' : aa'^{-1} \notin D\}|.$$

Then

$$\sum_{a \in A''} \beta(a) \leq \frac{|A''|^2}{32K}.$$

Set

$$A''' = \left\{ a \in A'' : \beta(a) \leq \frac{|A''|}{16K} \right\}.$$

Markov's inequality gives

$$(59) \quad |A'''| \geq \frac{1}{2}|A''| \geq \frac{|A|}{8\sqrt{2}K}.$$

1.4. Stage IV: regularize the B -side and bound the complete product set.

Because $A'' \subseteq A'$, every $a \in A''$ has more than $|B|/(4K)$ neighbors b with $ab \in C$. Therefore

$$\sum_{b \in B} |\{a \in A'' : ab \in C\}| > \frac{|A''||B|}{4K}.$$

Define

$$B''' = \left\{ b \in B : |\{a \in A'' : ab \in C\}| > \frac{|A''|}{8K} \right\}.$$

The complement contributes at most $|A''||B|/(8K)$, so the contribution of B''' is still at least $|A''||B|/(8K)$. Since each b has at most $|A''|$ neighbors,

$$(60) \quad |B'''| \geq \frac{|B|}{8K}.$$

We claim that $A'''B'''$ is small. Fix

$$c = ab, \quad a \in A''', \quad b \in B'''.$$

By the definition of B''' , there are more than $|A''|/(8K)$ elements $a' \in A''$ with $a'b \in C$. By the definition of A''' , at most $|A''|/(16K)$ elements $a' \in A''$ violate $aa'^{-1} \in D$. Hence more than

$$(61) \quad \frac{|A''|}{16K}$$

elements $a' \in A''$ satisfy simultaneously

$$x := a'b \in C, \quad aa'^{-1} \in D.$$

Since $c = ab$,

$$aa'^{-1} = ab(a'b)^{-1} = cx^{-1}.$$

Thus every $c \in A'''B'''$ has more than $|A''|/(16K)$ representations of the form

$$(62) \quad x \in C, \quad cx^{-1} \in D.$$

Summing (61) over $c \in A'''B'''$ gives

$$(63) \quad |A'''B'''| \frac{|A''|}{16K} \leq \sum_{c \in G} \sum_{x \in G} 1_C(x) 1_D(cx^{-1}).$$

For each pair $(x, d) \in C \times D$ there is a unique $c = dx$ contributing to the right side, so the double sum equals $|C||D|$. Therefore

$$|A'''B'''| \leq \frac{16K|C||D|}{|A''|}.$$

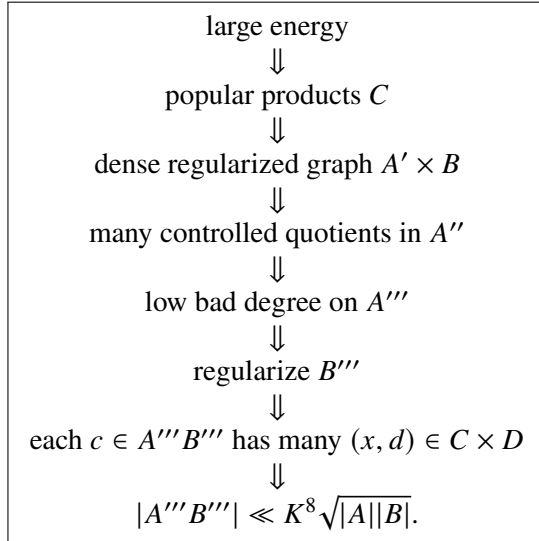
Using (50) and (58),

$$|A'''B'''| \leq C_2 K^8 \sqrt{|A||B|}.$$

Take $A_* = A'''$, $B_* = B'''$, and choose C_0 large enough to dominate $8\sqrt{2}$, 8, and C_2 . Equations (59), (60), and the last display prove the theorem. $\square$

2. The proof architecture after the calculation

The proof is worth compressing only *after* it has been seen in full:



The final inequality is a many-representations argument: a large set of products would require more pairs in $C \times D$ than exist.

3. Dependencies

The proof of [Theorem 5.1](#) invokes only the weak BSG theorem proved in [Chapter 4](#), plus Cauchy–Schwarz, thresholding, and double counting. Thus the noncommutative BSG theorem, historically due in this form to Tao [\[1\]](#), is now a proved theorem of the series rather than an external interface.

CHAPTER 6

Energy, Small Product Sets, and Approximate Groups: The Exact Equivalence

The previous chapters proved the two difficult directions separately: energy produces large subsets with small product, and a small product set is controlled by an approximate group. We now assemble these into the form actually used in inverse theorems. Large energy is equivalent, up to polynomial losses, to substantial portions of A and B lying in opposite translates of one approximate group of the geometric-mean size. Every implication is proved from earlier chapters.

1. The four manifestations of multiplicative structure

THEOREM 6.1 (Energy–approximate-group equivalence). *Let $A, B \subseteq G$ be finite and nonempty and let $K \geq 2$. The following assertions are equivalent up to replacing K by a fixed power K^C .*

(i) **Large energy:**

$$E(A, B) \geq K^{-1}(|A||B|)^{3/2}.$$

(ii) **A dense restricted product has small image:** *there is $E \subseteq A \times B$ with*

$$|E| \geq K^{-C}|A||B|$$

and

$$|\{ab : (a, b) \in E\}| \leq K^C \sqrt{|A||B|}.$$

(iii) **Large subsets have a small complete product:** *there are $A' \subseteq A$, $B' \subseteq B$ such that*

$$|A'| \geq K^{-C}|A|, \quad |B'| \geq K^{-C}|B|,$$

and

$$|A'B'| \leq K^C \sqrt{|A||B|}.$$

(iv) **Approximate-group capture:** *there is a K^C -approximate group H , with*

$$K^{-C} \sqrt{|A||B|} \leq |H| \leq K^C \sqrt{|A||B|},$$

and elements $x, y \in G$ such that

$$|A \cap xH| \geq K^{-C}|A|, \quad |B \cap Hy| \geq K^{-C}|B|.$$

PROOF. We prove a cycle of implications.

(i) $\Rightarrow$ (iii). This is exactly [Theorem 5.1](#). The theorem gives subsets of polynomial relative size and a complete product set of polynomially controlled geometric-mean size.

(iii) $\Rightarrow$ (iv). First note that A' and B' are polynomially comparable in size. Indeed

$$|A'B'| \geq \max(|A'|, |B'|)$$

and the lower bounds on $|A'|, |B'|$ imply

$$K^{-O(1)}|A| \leq |A'| \leq K^{O(1)}\sqrt{|A||B|},$$

with the analogous estimate for B' . In particular $|A|$ and $|B|$ themselves are polynomially comparable.

Apply the small-product theorem [Theorem 3.3](#) to A', B' . We obtain a $K^{O(1)}$ -approximate group H with

$$|H| \leq K^{O(1)}\sqrt{|A'||B'|} \leq K^{O(1)}\sqrt{|A||B|}$$

and coverings

$$A' \subseteq XH, \quad B' \subseteq HY,$$

where $|X|, |Y| \leq K^{O(1)}$. By pigeonhole, some $x \in X$ satisfies

$$|A' \cap xH| \geq |A'|/|X| \geq K^{-O(1)}|A|,$$

and similarly some $y \in Y$ satisfies

$$|B' \cap Hy| \geq K^{-O(1)}|B|.$$

Since the first intersection has at most $|H|$ elements,

$$|H| \geq K^{-O(1)}|A|,$$

and the analogous B estimate gives

$$|H| \geq K^{-O(1)}\sqrt{|A||B|}$$

after polynomial comparability. This is (iv).

(iv) $\Rightarrow$ (ii). Let

$$A_0 = A \cap xH, \quad B_0 = B \cap Hy.$$

Take $E = A_0 \times B_0$. Its density is polynomially large. Moreover

$$A_0B_0 \subseteq xH^2y.$$

Since H is a K^C -approximate group,

$$|H^2| \leq K^C|H| \leq K^{O(1)}\sqrt{|A||B|}.$$

Thus the restricted product image has the required size.

(ii) $\Rightarrow$ (i). Let

$$P = \{ab : (a, b) \in E\}.$$

For $z \in P$ define

$$r_E(z) = |\{(a, b) \in E : ab = z\}|.$$

Then

$$\sum_z r_E(z) = |E|.$$

By Cauchy–Schwarz,

$$\sum_z r_E(z)^2 \geq \frac{|E|^2}{|P|}.$$

Every collision counted by $r_E(z)^2$ is also a collision counted by $E(A, B)$, so

$$E(A, B) \geq \frac{|E|^2}{|P|} \geq K^{-O(1)} \frac{|A|^2|B|^2}{\sqrt{|A||B|}} = K^{-O(1)} (|A||B|)^{3/2}.$$

This is (i), after changing the polynomial exponent. $\square$

2. The one-set version

COROLLARY 6.2 (Large energy forces a large approximate-group piece). *If*

$$E(A) \geq K^{-1}|A|^3,$$

then there is a $K^{O(1)}$ -approximate group H with

$$K^{-O(1)}|A| \leq |H| \leq K^{O(1)}|A|$$

and some $x \in G$ such that

$$|A \cap xH| \geq K^{-O(1)}|A|.$$

PROOF. Apply [Theorem 6.1](#) with $B = A$. The theorem gives left and right captures; either one yields the stated conclusion. $\square$

3. Uniform probability measures

For the uniform measure $\mu_A = |A|^{-1}1_A$,

$$\|\mu_A * \mu_A\|_2^2 = \frac{E(A)}{|A|^4}, \quad \|\mu_A\|_2 = |A|^{-1/2}.$$

Hence

$$\|\mu_A * \mu_A\|_2 \geq K^{-1}\|\mu_A\|_2$$

implies

$$E(A) \geq K^{-2}|A|^3.$$

By [Corollary 6.2](#), a polynomial fraction of A lies in a coset of an approximate group whose size is polynomially comparable to $|A|$.

This is already the inverse theorem for *uniform* measures. The remaining analytic difficulty is that a random-walk measure is not uniform. The next chapter supplies the dyadic decomposition that converts a general measure into a pair of nearly uniform levels without losing more than powers of a logarithm.

EXAMPLE 6.3 (A complete nonabelian inverse-step calculation in S_3). Let

$$G = S_3, \quad A = \{e, (12), (13), (23)\}.$$

This set is symmetric, contains the identity, and is not a subgroup. Nevertheless

$$A^2 = S_3, \quad A^3 = S_3,$$

so its tripling constant is only

$$\frac{|A^3|}{|A|} = \frac{6}{4} = \frac{3}{2}.$$

The representation function $r(z) = |\{(a, b) \in A^2 : ab = z\}|$ can be written down exactly. The identity has four representations, each transposition has two, and each of the two 3-cycles has three. Hence

$$\sum_{z \in S_3} r(z) = 4 + 3 \cdot 2 + 2 \cdot 3 = 16 = |A|^2$$

and

$$E(A) = \sum_z r(z)^2 = 4^2 + 3 \cdot 2^2 + 2 \cdot 3^2 = 46.$$

For the uniform measure $\mu_A = \frac{1}{4} 1_A$,

$$\|\mu_A * \mu_A\|_2^2 = \frac{46}{4^4} = \frac{23}{128}, \quad \|\mu_A\|_2^2 = \frac{1}{4},$$

and therefore

$$\frac{\|\mu_A * \mu_A\|_2}{\|\mu_A\|_2} = \sqrt{\frac{23}{32}} \approx 0.848.$$

Thus one convolution has flattened the measure only modestly. The inverse conclusion is visible without any asymptotics: the small-tripling construction gives

$$H = (A \cup A^{-1} \cup \{e\})^3 = S_3.$$

The set H is a 1-approximate group, has size $6 \asymp |A|$, and captures all of A .

This toy model should not be mistaken for the hard direction of BSG: here the ambient group is so small that we can enumerate everything. Its value is that every object in the abstract inverse theorem—collision energy, failure of L^2 -flattening, small tripling, and approximate-group capture—can be seen numerically in one genuinely noncommutative example.

Two routes meet in the inverse theorem

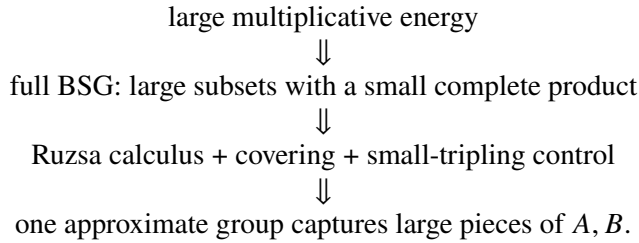


FIGURE 1. The noncommutative inverse machine. The upper route converts collisions into a small complete product; the lower structural route turns product control into an approximate group.

4. How the inverse step is assembled

The structural arrow

large energy $\Rightarrow$ approximate group

is obtained from the following internal dependency chain:

Ruzsa triangle + covering

$\Downarrow$

small tripling $\Rightarrow$ approximate group

$\Downarrow$

small product $\Rightarrow$ approximate-group control,

while independently

energy $\Rightarrow$ full BSG $\Rightarrow$ small complete product.

Combining the two proves the inverse step. Tao's paper [1] is the historical source for this noncommutative BSG architecture; the chain needed here has been proved in the preceding chapters.

Part 3

Measures and Flattening

CHAPTER 7

From General Measures to Approximate Groups

A random-walk measure is not uniform on a set, so the set BSG theorem cannot be applied directly. We give a complete dyadic reduction. If convolution fails to decrease the ℓ^2 norm, two dyadic levels have large mixed energy; the full BSG theorem then produces an approximate group carrying definite mass. The proof also isolates an important point suppressed in many schematic accounts: an upper bound $|H| \ll \|\mu\|_2^{-2}$ is unconditional, but a matching lower bound requires an anti-atom estimate such as $\|\mu\|_\infty \ll \|\mu\|_2^2$.

1. Truncating tiny atoms

Let G be finite of order N and let μ be a probability measure. Put

$$s = \|\mu\|_2.$$

Since $\|\mu\|_1 = 1$, Cauchy–Schwarz gives

$$(64) \quad N^{-1/2} \leq s \leq 1.$$

Fix $K \geq 2$ and set

$$\tau = \frac{s}{32KN}.$$

Let

$$r = \mu 1_{\{x: \mu(x) \leq \tau\}}, \quad \mu_0 = \mu - r.$$

Then

$$(65) \quad \|r\|_1 \leq N\tau = \frac{s}{32K} \leq \frac{1}{32K}.$$

Young’s inequality therefore gives

$$(66) \quad \|r * \mu\|_2 + \|\mu_0 * r\|_2 \leq \frac{s}{16K}.$$

2. Dyadic levels

Choose J so that $2^{-J} \leq \tau < 2^{-J+1}$ and define, for $0 \leq j < J$,

$$A_j = \{x : 2^{-j-1} < \mu(x) \leq 2^{-j}\}, \quad f_j = \mu 1_{A_j}.$$

Then

$$\mu_0 = \sum_{j=0}^{J-1} f_j,$$

and

$$(67) \quad J \leq C \log(2NK)$$

for an absolute C , using (64). For $m_j = |A_j|$ and $\alpha_j = 2^{-j}$,

$$(68) \quad \frac{\alpha_j}{2} m_j \leq \|f_j\|_1 \leq \alpha_j m_j,$$

$$(69) \quad \frac{\alpha_j}{2} \sqrt{m_j} \leq \|f_j\|_2 \leq \alpha_j \sqrt{m_j}.$$

THEOREM 7.1 (Weighted inverse theorem for failed ℓ^2 flattening). *Let μ be a probability measure on a finite group G , put $s = \|\mu\|_2$, and suppose*

$$(70) \quad \|\mu * \mu\|_2 \geq K^{-1}s, \quad K \geq 2.$$

Set

$$L = K \log(2|G|K).$$

Then there are an $L^{O(1)}$ -approximate group H and $x, y \in G$ such that

$$(71) \quad |H| \leq L^{O(1)} s^{-2},$$

$$(72) \quad \mu(xH) \geq L^{-O(1)},$$

$$(73) \quad \mu(Hy) \geq L^{-O(1)}.$$

If, in addition,

$$(74) \quad \|\mu\|_\infty \leq Ms^2$$

for some $M \geq 1$, then H may be chosen so that also

$$(75) \quad |H| \geq M^{-2} L^{-O(1)} s^{-2}.$$

PROOF. By (66),

$$\|\mu_0 * \mu_0\|_2 \geq \|\mu * \mu\|_2 - \|r * \mu\|_2 - \|\mu_0 * r\|_2 \geq \frac{15}{16K} s.$$

Since

$$\mu_0 * \mu_0 = \sum_{i,j < J} f_i * f_j,$$

the triangle inequality and pigeonhole principle yield indices i, j with

$$(76) \quad \|f_i * f_j\|_2 \geq qs, \quad q := \frac{15}{16KJ^2}.$$

Step 1: the selected levels carry definite ℓ^1 and ℓ^2 mass. Young's inequality gives

$$\|f_i * f_j\|_2 \leq \|f_i\|_1 \|f_j\|_2 \leq s \|f_i\|_1,$$

so (76) implies

$$(77) \quad \|f_i\|_1 \geq q.$$

By symmetry of the argument, $\|f_j\|_1 \geq q$. Likewise

$$\|f_i * f_j\|_2 \leq \|f_i\|_2 \|f_j\|_1 \leq \|f_i\|_2,$$

so

$$(78) \quad \|f_i\|_2, \|f_j\|_2 \geq qs.$$

Step 2: the selected levels have large multiplicative energy. Since $f_i \leq \alpha_i 1_{A_i}$ and $f_j \leq \alpha_j 1_{A_j}$,

$$\|f_i * f_j\|_2 \leq \alpha_i \alpha_j \|1_{A_i} * 1_{A_j}\|_2 = \alpha_i \alpha_j E(A_i, A_j)^{1/2}.$$

We compare the normalizing factor with s . From (68)–(69),

$$\alpha_i m_i^{3/4} = (\alpha_i m_i)^{1/2} (\alpha_i \sqrt{m_i})^{1/2} \leq (2\|f_i\|_1)^{1/2} (2\|f_i\|_2)^{1/2} \leq 2s^{1/2},$$

and similarly for j . Hence

$$(79) \quad \alpha_i \alpha_j (m_i m_j)^{3/4} \leq 4s.$$

Combining (76) and (79),

$$E(A_i, A_j)^{1/2} \geq \frac{q}{4} (m_i m_j)^{3/4},$$

so

$$(80) \quad E(A_i, A_j) \geq \frac{q^2}{16} (m_i m_j)^{3/2}.$$

Step 3: apply the proved set inverse theorem. Apply Theorem 6.1 to A_i, A_j with energy parameter

$$R = 16q^{-2} = (KJ)^{O(1)}.$$

We obtain an $R^{O(1)}$ -approximate group H and $x, y \in G$ such that

$$(81) \quad |H| \leq R^{O(1)} \sqrt{m_i m_j},$$

$$(82) \quad |A_i \cap xH| \geq R^{-O(1)} m_i,$$

$$(83) \quad |A_j \cap Hy| \geq R^{-O(1)} m_j.$$

Step 4: convert level capture to μ -mass. Every point of A_i has mass $> \alpha_i/2$. Thus

$$\mu(xH) \geq \frac{\alpha_i}{2} |A_i \cap xH| \geq R^{-O(1)} \alpha_i m_i.$$

But (77) and $\|f_i\|_1 \leq \alpha_i m_i$ give $\alpha_i m_i \geq q$. Therefore

$$\mu(xH) \geq (KJ)^{-O(1)}.$$

The same argument gives the right-coset estimate. Since $J = O(\log(2NK))$, these are (72)–(73).

Step 5: upper bound the size of H . From (68) and (69),

$$\sqrt{m_i} \leq 2 \frac{\|f_i\|_1}{\|f_i\|_2} \leq \frac{2}{qs},$$

where we used $\|f_i\|_1 \leq 1$ and (78). Hence

$$m_i, m_j \leq 4q^{-2}s^{-2}.$$

Substituting into (81) yields

$$|H| \leq (KJ)^{O(1)} s^{-2},$$

which is (71).

Step 6: lower bound under the anti-atom hypothesis. Assume (74). If A_i is nonempty, then $\alpha_i < 2\|\mu\|_\infty \leq 2Ms^2$. By (69) and (78),

$$\alpha_i \sqrt{m_i} \geq \|f_i\|_2 \geq qs,$$

so

$$m_i \geq \frac{q^2 s^2}{\alpha_i^2} \geq \frac{q^2}{4M^2 s^2}.$$

The same holds for m_j . The lower-size conclusion in Theorem 6.1 therefore gives

$$|H| \geq R^{-O(1)} \sqrt{m_i m_j} \geq M^{-2} (KJ)^{-O(1)} s^{-2},$$

which is (75). □

WARNING 7.2 (The lower scale is not automatic). The anti-atom assumption is logically real. The effective-support number s^{-2} is an averaged notion. A measure can contain a small heavy level and a diffuse tail; failed flattening may be detected on the heavy level, whose cardinality is much smaller than s^{-2} . The proof above therefore gives the unconditional upper bound (71), while (74) rules out precisely this heavy-level pathology and restores the lower bound.

3. A form tailored to Bourgain–Gamburd

In expansion arguments one usually applies the theorem after an initial-spreading stage that supplies a point-mass estimate. A convenient formulation is the following.

COROLLARY 7.3 (Flattening obstruction at a controlled scale). *Suppose (70) holds and, for some fixed $M \geq 1$,*

$$\|\mu\|_\infty \leq M\|\mu\|_2^2.$$

Then failed ℓ^2 flattening produces an $L^{O(1)}$ -approximate group H satisfying

$$M^{-2}L^{-O(1)}\|\mu\|_2^{-2} \leq |H| \leq L^{O(1)}\|\mu\|_2^{-2}$$

and a left and a right coset of H carrying $L^{-O(1)}$ μ -mass.

PROOF. This is exactly [Theorem 7.1](#). □

4. Dependencies

The weighted theorem uses no external inverse theorem. Its only non-elementary input is [Theorem 6.1](#), already proved in this volume. The remaining ingredients are truncation, dyadic decomposition, Young’s inequality, and the elementary comparison between ℓ^1 and ℓ^2 norms of a nearly constant function. This is the analytic-to-combinatorial bridge usually summarized as “weighted BSG.”

CHAPTER 8

The L^2 -Flattening Theorem

We now turn the measure inverse theorem into the analytic engine of Bourgain–Gamburd. The theorem is conditional only on algebraic information about approximate groups and probabilistic non-concentration on proper subgroups; all additive-combinatorial input has already been proved. We state the scale hypotheses explicitly, including the anti-atom estimate needed to know that the approximate group produced by BSG really lies in the medium range.

GUIDING IDEA 8.1 (Flattening is a contradiction theorem). Convolution by a probability measure is merely nonexpanding on ℓ^2 :

$$\|\mu * \nu\|_2 \leq \|\mu\|_1 \|\nu\|_2 = \|\nu\|_2.$$

A power saving is therefore extra information. The inverse theorem says exactly what obstructs it: a translate of an approximate group carrying noticeable mass. The rest of the Bourgain–Gamburd method is designed to make that obstruction impossible.

1. Power-scale form of the weighted inverse theorem

Let $N = |G|$. For fixed $a > 0$ and sufficiently large N ,

$$\log(2N^2) \leq N^a.$$

Thus, for any fixed $\varepsilon > 0$, if $K = N^\varepsilon$, then the parameter

$$L = K \log(2NK)$$

from [Theorem 7.1](#) satisfies $L \leq N^{2\varepsilon}$ for sufficiently large N (with the threshold allowed to depend on the fixed ε).

COROLLARY 8.2 (Power-scale inverse theorem). *There is an absolute constant C such that the following holds. Let ν be a probability measure on a finite group G of order N , let $s = \|\nu\|_2$, and suppose*

$$\|\nu * \nu\|_2 \geq N^{-\varepsilon} s.$$

Then some $N^{C\varepsilon}$ -approximate group H has a left and a right coset carrying $N^{-C\varepsilon}$ ν -mass and

$$|H| \leq N^{C\varepsilon} s^{-2}.$$

If also

$$\|v\|_\infty \leq N^{C\varepsilon}s^2,$$

then

$$N^{-C\varepsilon}s^{-2} \leq |H| \leq N^{C\varepsilon}s^{-2}.$$

PROOF. Apply Theorem 7.1 with $K = N^\varepsilon$ and absorb the logarithm into a slightly larger multiple of ε . $\square$

2. The algebraic and probabilistic hypotheses

DEFINITION 8.3 (Medium-range approximate-subgroup dichotomy). Fix $0 < \delta < 1/2$ and $\eta > 0$. A finite group G of order N satisfies $\text{ASD}(\delta, \eta)$ if every N^η -approximate group H with

$$N^\delta \leq |H| \leq N^{1-\delta}$$

is contained in a union of at most N^η left cosets of some proper subgroup $L < G$.

DEFINITION 8.4 (Subgroup non-concentration). A probability measure ν satisfies $\text{NC}(\kappa)$ if

$$\nu(gL) \leq N^{-\kappa}$$

for every proper subgroup $L < G$ and every $g \in G$.

LEMMA 8.5 (Approximate-group capture contradicts subgroup escape). Suppose $H \subseteq XL$ with $L < G$ proper and $|X| \leq N^\eta$. If ν satisfies $\text{NC}(\kappa)$, then

$$\nu(gH) \leq N^{\eta-\kappa}$$

for every $g \in G$.

PROOF. The set gH is contained in the union of the $|X|$ cosets gxL , $x \in X$. Sum the non-concentration bound. $\square$

3. Structural flattening

THEOREM 8.6 (Abstract Bourgain–Gamburd L^2 flattening). Fix $0 < \alpha, \theta, \kappa < 1/2$ and choose

$$0 < \delta < \min(2\alpha, 2\theta).$$

There exist an absolute constant $D \geq 1$ and a fixed exponent

$$\varepsilon = \varepsilon(\alpha, \theta, \kappa, \delta) > 0$$

such that the following holds for all sufficiently large N . Let G have order N and let ν be a probability measure satisfying

$$(84) \quad N^{-1/2+\theta} \leq \|v\|_2 \leq N^{-\alpha},$$

$$(85) \quad \|v\|_\infty \leq N^\varepsilon \|v\|_2^2,$$

and $\text{NC}(\kappa)$. Assume moreover that G satisfies $\text{ASD}(\delta, D\varepsilon)$. Then

$$(86) \quad \|\nu * \nu\|_2 < N^{-\varepsilon} \|\nu\|_2.$$

PROOF. Let C_1 be an absolute constant large enough to dominate every exponent implicit in [Corollary 8.2](#), and choose $D \geq C_1$. Now choose a *fixed* positive ε so small that

$$(87) \quad C_1\varepsilon < \min(2\alpha - \delta, 2\theta - \delta) \quad \text{and} \quad (C_1 + D)\varepsilon < \kappa.$$

Because ε is fixed, the logarithm in the weighted inverse theorem is absorbed into $N^{O(\varepsilon)}$ once N is sufficiently large.

Suppose (86) fails. Apply [Corollary 8.2](#) with this same ε . The anti-atom hypothesis (85) is at the same power scale, so after increasing C_1 if necessary we obtain an $N^{C_1\varepsilon}$ -approximate group H and a coset gH such that

$$(88) \quad \nu(gH) \geq N^{-C_1\varepsilon}$$

and

$$(89) \quad N^{-C_1\varepsilon} \|\nu\|_2^{-2} \leq |H| \leq N^{C_1\varepsilon} \|\nu\|_2^{-2}.$$

The matching power scale is essential: it is what makes both losses in (89) proportional to ε .

From (84),

$$N^{2\alpha - C_1\varepsilon} \leq |H| \leq N^{1 - 2\theta + C_1\varepsilon}.$$

The first inequalities in (87) place H in the medium range of $\text{ASD}(\delta, D\varepsilon)$. Since $D \geq C_1$, H is also an $N^{D\varepsilon}$ -approximate group. Hence

$$H \subseteq XL$$

for a proper subgroup $L < G$ and $|X| \leq N^{D\varepsilon}$. By [Lemma 8.5](#),

$$\nu(gH) \leq N^{D\varepsilon - \kappa}.$$

The second inequality in (87) gives $D\varepsilon - \kappa < -C_1\varepsilon$, contradicting (88). This proves the theorem. $\square$

REMARK 8.7 (Where the group theory enters). Everything in the proof up to ASD is now internal to Volumes I–II. The statement ASD is not being asserted without proof for any concrete group here; it is an abstract hypothesis that isolates exactly what a group-specific application must establish. Thus [Theorem 8.6](#) is a complete conditional theorem: once product-growth/approximate-subgroup rigidity and subgroup nonconcentration are verified for a chosen family, no additional combinatorial inverse theorem is needed.

4. Iteration and logarithmic time

Let μ be a symmetric step distribution and set

$$\nu_0 = \mu^{(n_0)}, \quad \nu_{j+1} = \nu_j * \nu_j = \mu^{(2^{j+1}n_0)}.$$

PROPOSITION 8.8 (Only finitely many flattening stages are needed). *Suppose $\|\nu_0\|_2 \leq N^{-\alpha}$ and whenever*

$$\|\nu_j\|_2 > N^{-1/2+\theta}$$

one has

$$\|\nu_{j+1}\|_2 \leq N^{-\varepsilon} \|\nu_j\|_2.$$

Then after at most

$$J_* = \left\lceil \frac{1/2 - \theta - \alpha}{\varepsilon} \right\rceil_+$$

stages,

$$\|\nu_{J_*}\|_2 \leq N^{-1/2+\theta}.$$

PROOF. Until the target is reached,

$$\|\nu_j\|_2 \leq N^{-\alpha-j\varepsilon}.$$

The exponent reaches $1/2 - \theta$ after the stated number of increments. $\square$

If $n_0 = O(\log N)$ and all exponents are fixed independently of N , then J_* is an absolute constant. Therefore

$$2^{J_*} n_0 = O(\log N).$$

This is why a power-saving flattening estimate is exactly the right strength: it preserves logarithmic random-walk time.

5. A failure mode worth remembering

A bound of the form

$$\|\nu * \nu\|_2 \leq 0.99 \|\nu\|_2$$

would not suffice. To improve the exponent of N by a fixed amount, one would need $\asymp \log N$ such stages; doubling the time at each stage would then produce polynomial rather than logarithmic time. Uniform expansion requires a power of N improvement per successful stage, not merely a fixed numerical contraction.

6. Dependencies

The theorem of this chapter uses only:

- (1) the weighted inverse theorem proved in [Chapter 7](#);
- (2) the abstract group-theoretic hypothesis ASD;
- (3) the explicitly stated non-concentration hypothesis NC.

No BSG theorem, approximate-group extraction, or dyadic-measure lemma is left external. This completes the combinatorial/analytic core needed before the group-specific growth theory begins.

Part 4

Technical Laboratories

APPENDIX A

Dense-Graph Mechanics Behind Balog–Szemerédi–Gowers

Chapters 4–5 prove the noncommutative Balog–Szemerédi–Gowers theorem by a deliberately economical counting argument. This appendix slows that proof down and recasts it as a graph problem. We prove the popularity, pruning, common-neighbor, and path-counting lemmas separately; work through a model example; and then identify exactly where each lemma occurs in the main proof. No additional BSG theorem is invoked here.

1. Restricted product graphs

Let A, B be finite nonempty subsets of a group G and let $\Gamma \subseteq A \times B$. Define the restricted product set

$$A \cdot_{\Gamma} B = \{ab : (a, b) \in \Gamma\}.$$

The BSG situation is the simultaneous occurrence of

$$|\Gamma| \gg |A||B| \quad \text{and} \quad |A \cdot_{\Gamma} B| \ll \sqrt{|A||B|}.$$

The first inequality says that many pairs survive. The second says that their products collide heavily. The proof consists of turning those collisions into short paths and then using the labels on those paths to control quotients.

2. Popularity from energy

Write

$$r(x) = r_{A,B}(x) = |\{(a, b) \in A \times B : ab = x\}|, \quad E = \sum_x r(x)^2.$$

LEMMA A.1 (Energy on popular products). *For $0 < \theta < 1$ put*

$$P_{\theta} = \left\{ x : r(x) \geq \frac{\theta E}{|A||B|} \right\}.$$

Then

$$\sum_{x \in P_{\theta}} r(x)^2 \geq (1 - \theta)E$$

and

$$|P_{\theta}| \leq \frac{|A|^2 |B|^2}{\theta^2 E}.$$

PROOF. On the complement of P_θ ,

$$r(x)^2 < \frac{\theta E}{|A||B|} r(x).$$

Summing and using $\sum_x r(x) = |A||B|$ gives a contribution less than θE . This proves the first assertion. For the second, every $x \in P_\theta$ contributes at least $\theta^2 E^2 / (|A|^2 |B|^2)$ to the energy; summing these lower bounds and comparing with E proves the result. $\square$

COROLLARY A.2 (Energy creates a dense small-target graph). *Suppose*

$$E(A, B) \geq K^{-1} |A|^{3/2} |B|^{3/2}.$$

Then there is $\Gamma \subseteq A \times B$ such that

$$|\Gamma| \geq (2K)^{-1} |A||B|$$

and

$$|A \cdot_\Gamma B| \leq 4K \sqrt{|A||B|}.$$

PROOF. Interchange the two sides after inversion if necessary so that $|A| \leq |B|$. Apply the lemma with $\theta = 1/2$ and let Γ consist of pairs whose product lies in $P_{1/2}$. Since $r(x) \leq |A|$,

$$|A||\Gamma| \geq \sum_{x \in P_{1/2}} r(x)^2 \geq E/2.$$

Thus

$$|\Gamma| \geq \frac{1}{2K} |A|^{1/2} |B|^{3/2}.$$

The energy hypothesis together with the trivial upper bound $E(A, B) \leq |A|^2 |B|$ implies $|B|/|A| \leq K^2$, and hence the displayed bound is at least $|A||B|/(2K)$. The bound for the restricted product set is the second conclusion of Lemma A.1. $\square$

3. Pruning without losing the density scale

LEMMA A.3 (Two-sided pruning). *If $\Gamma \subseteq A \times B$ has*

$$|\Gamma| \geq \delta |A||B|,$$

then one may delete vertices and obtain nonempty $A_0 \subseteq A$, $B_0 \subseteq B$ and a subgraph Γ_0 with at least $\delta |A||B|/2$ edges such that every surviving $a \in A_0$ has degree at least $\delta |B|/4$ and every surviving $b \in B_0$ has degree at least $\delta |A|/4$.

PROOF. Repeatedly delete any A -vertex whose current degree is less than $\delta |B|/4$ and any B -vertex whose current degree is less than $\delta |A|/4$. Deletions on the A side remove in total fewer than $\delta |A||B|/4$ edges, and the B side removes fewer than the same number. At least half the original lower bound therefore survives. The degree conditions hold by construction. $\square$

4. Common neighbors encode quotients

For $a \in A$ write $N(a) = \{b : (a, b) \in \Gamma\}$.

LEMMA A.4 (Second moment of degrees). *If $|\Gamma| = \delta|A||B|$, then*

$$\sum_{a, a' \in A} |N(a) \cap N(a')| = \sum_{b \in B} \deg(b)^2 \geq \delta^2 |A|^2 |B|.$$

PROOF. Both sums count triples (a, a', b) with $(a, b), (a', b) \in \Gamma$. Apply Cauchy–Schwarz to the degrees on the B side:

$$\sum_b \deg(b)^2 \geq \frac{(\sum_b \deg(b))^2}{|B|} = \delta^2 |A|^2 |B|.$$

□

Let $C = A \cdot_{\Gamma} B$. A common neighbor has an algebraic meaning:

$$b \in N(a) \cap N(a') \implies aa'^{-1} = (ab)(a'b)^{-1} \in CC^{-1}.$$

More is true: distinct common neighbors give distinct pairs $(ab, a'b) \in C^2$.

LEMMA A.5 (Path multiplicity becomes quotient multiplicity). *For fixed $a, a' \in A$, the number $|N(a) \cap N(a')|$ is at most the number of representations*

$$aa'^{-1} = cd^{-1}, \quad c, d \in C.$$

PROOF. Send b to $(ab, a'b)$. The displayed quotient identity is immediate, and cancellation shows that the map is injective. □

5. Why “most pairs are good” is not yet BSG

The moment estimate guarantees many pairs with many common neighbors. It does not give a large set on which *every* pair is good. Nor is such a statement necessary. The main proof uses a more efficient route: it chooses one fiber b_0 , obtaining a large A' for which only a small proportion of quotients are bad, and then prunes vertices of excessive bad degree. The second regularization, on the B side, supplies enough alternative factorizations to bound the complete product set.

The distinction is worth seeing on a toy graph. Partition

$$A = A_1 \sqcup A_2, \quad |A_2| = \eta|A|,$$

and join every point of A_1 to every point of B , while giving the vertices of A_2 arbitrary neighborhoods. A proportion $1 - O(\eta)$ of pairs on the A side have maximal common-neighbor count, but the quotients involving A_2 can be completely uncontrolled. Deleting vertices of high bad degree is therefore not a cosmetic step: it prevents a small exceptional vertex set from realizing an enormous complete quotient set.

6. The main proof translated into graph language

We now match the stages of [Theorem 5.1](#) with the graph picture. Suppose

$$E(A, B) \geq K^{-1}|A|^{3/2}|B|^{3/2}.$$

- Stage 1.** *Popularity.* The set C of products represented more than $\sqrt{|A||B|}/(2K)$ times supports at least $|A||B|/(2K)$ edges and has $|C| \leq 2K\sqrt{|A||B|}$.
- Stage 2.** *One-sided regularization.* Keep the $a \in A$ incident to more than $|B|/(4K)$ popular edges. The surviving set has size at least $|A|/(4K)$.
- Stage 3.** *Common-neighbor extraction.* The weak BSG theorem is exactly a fiber-averaged version of [Lemma A.4](#): it chooses one b_0 for which a large fiber A'' has only a small proportion of bad quotients. Each accepted quotient has many witnesses $c, dc \in C$.
- Stage 4.** *Bad-degree pruning.* Remove the a for which too many partners produce quotients outside the controlled set D . At least half of A'' survives.
- Stage 5.** *Regularize the opposite side.* Keep those $b \in B$ incident to many popular edges from A'' . The resulting B''' has polynomially large size.
- Stage 6.** *Complete product bound.* For every $c = ab \in A'''B'''$, many choices a' simultaneously satisfy $a'b \in C$ and $aa'^{-1} \in D$. Hence c has many representations $c = dx$ with $(d, x) \in D \times C$. Since there are only $|D||C|$ such pairs, the number of possible c is small.

This list is not a substitute for the proof in Chapter 5; it is a map of the proof after every quantitative step has already been justified.

7. A subgroup model

Let $H < G$ be a finite subgroup and take $A = B = C = H$ with the complete graph. Then

$$E(H, H) = |H|^3, \quad r_{H, H}(h) = |H|, \quad HH = H.$$

Every pair of left vertices has all of H as common neighbors, every quotient lies in H , and every stage of the BSG proof returns H itself up to harmless choices. This is the exact model that the inverse theorem is trying to detect: large energy is the softened statistical shadow of subgroup-like multiplication.

8. What the graph proof teaches

Three principles survive in later applications.

First, an inverse theorem begins with *multiplicity*, not with a small set. The small target is produced by thresholding a representation function. Second, local structure does not automatically become complete-set structure; pruning is what makes exceptional pairs harmless. Third, the final product-set bound is always a counting contradiction: every output has many witnesses, but there are only a controlled number of witness pairs available.

APPENDIX B

Dyadic Measures, Effective Support, and the Anti-Atom Condition

The weighted inverse theorem is often remembered as “apply BSG to a dyadic level.” That phrase hides the only analytic subtlety in passing from sets to measures: the level that witnesses non-flattening need not have the same cardinality as the effective support $\|\mu\|_2^{-2}$. This appendix works through the dyadic decomposition numerically, proves the scale comparisons in a reusable form, and gives examples showing why an anti-atom hypothesis is needed for a two-sided size estimate.

1. Effective support is an averaged cardinality

For a probability measure μ on a finite set define

$$S_2(\mu) = \|\mu\|_2^{-2}.$$

If μ is uniform on a set of cardinality m , then $S_2(\mu) = m$. In general S_2 need not be the cardinality of any level set.

PROPOSITION B.1 (Basic bounds for effective support). *If G has order N , then*

$$1 \leq S_2(\mu) \leq N.$$

Moreover

$$|\text{supp } \mu| \geq S_2(\mu), \quad \|\mu\|_\infty \geq S_2(\mu)^{-1}.$$

PROOF. The first bounds are equivalent to $N^{-1/2} \leq \|\mu\|_2 \leq 1$. Cauchy–Schwarz gives

$$1 = \|\mu\|_1 \leq |\text{supp } \mu|^{1/2} \|\mu\|_2,$$

which yields the support bound. Finally

$$\|\mu\|_2^2 = \sum_x \mu(x)^2 \leq \|\mu\|_\infty \sum_x \mu(x) = \|\mu\|_\infty.$$

□

2. A two-level calculation

Let H and R be disjoint, with $|H| = h$ and $|R| = r$, and put

$$\mu = \frac{\alpha}{h} 1_H + \frac{1-\alpha}{r} 1_R, \quad 0 < \alpha < 1.$$

Then

$$(90) \quad \|\mu\|_2^2 = \frac{\alpha^2}{h} + \frac{(1-\alpha)^2}{r}.$$

If $r \gg h/\alpha^2$, the first term dominates and

$$S_2(\mu) \asymp \frac{h}{\alpha^2}.$$

The heavy level itself has only h points, smaller than effective support by the factor α^2 . If α is a negative power of $|G|$, this is a genuine power discrepancy.

Suppose now that H is a subgroup and the diffuse component is negligible for one convolution step. The heavy part satisfies

$$\left(\frac{\alpha}{h} 1_H\right) * \left(\frac{\alpha}{h} 1_H\right) = \frac{\alpha^2}{h} 1_H.$$

Thus non-flattening may be detected on H even though $|H|$ is substantially smaller than $S_2(\mu)$. This is the mechanism behind the warning in Chapter 7.

3. What the anti-atom inequality forbids

Assume

$$\|\mu\|_\infty \leq M \|\mu\|_2^2.$$

For any dyadic level

$$A_j = \{x : 2^{-j-1} < \mu(x) \leq 2^{-j}\}$$

that is nonempty, write $\alpha_j = 2^{-j}$. Then

$$\alpha_j < 2 \|\mu\|_\infty \leq 2M \|\mu\|_2^2.$$

If the level has ℓ^2 norm at least $q \|\mu\|_2$, then

$$q \|\mu\|_2 \leq \|\mu 1_{A_j}\|_2 \leq \alpha_j \sqrt{|A_j|},$$

so

$$(91) \quad |A_j| \geq \frac{q^2}{4M^2} \|\mu\|_2^{-2}.$$

Thus anti-concentration at individual points turns a lower bound on level ℓ^2 mass into a lower bound on level cardinality at the effective-support scale.

GUIDING IDEA B.2. The role of the anti-atom condition is not to prove that μ is close to uniform. It only rules out the possibility that the entire ℓ^2 signal is stored on a very small set of very heavy atoms. Once that pathology is excluded, the dyadic level detected by failed flattening has the correct cardinality scale for the algebraic product theorem.

4. Why only logarithmically many levels occur

Let $s = \|\mu\|_2$ and truncate atoms below

$$\tau = \frac{s}{32KN}.$$

Since $s \geq N^{-1/2}$,

$$\tau \geq \frac{1}{32KN^{3/2}}.$$

Hence the number of dyadic intervals between 1 and τ is at most

$$1 + \log_2(32KN^{3/2}) = O(\log(2NK)).$$

The logarithm in the weighted inverse theorem comes from this bookkeeping and nowhere deeper. When the flattening parameter is a power $K = N^\varepsilon$, this logarithm may be absorbed into $N^{c\varepsilon}$ for large N , which is why the expansion machine is naturally formulated with power losses.

5. A numerical laboratory

Take $N = 10^{12}$ and imagine a measure with

$$\|\mu\|_2 = N^{-0.3} = 10^{-3.6}.$$

Then

$$S_2(\mu) = N^{0.6} = 10^{7.2}.$$

If failed flattening is tested with $K = N^{0.01}$, the dyadic theorem loses only $N^{O(0.01)}$ after the logarithm is absorbed. Under an anti-atom estimate of the same power strength, the extracted approximate group therefore has size

$$N^{0.6 \pm O(0.01)}.$$

This is precisely a medium scale: far larger than a bounded subgroup and far smaller than the whole group. The product theorem can then rule it out unless it is controlled by a proper subgroup, while subgroup non-concentration rules out that remaining possibility.

The example explains why the exponents in the abstract flattening theorem are organized around $\|\mu\|_2^{-2}$ rather than around $\|\mu\|_1$ or the literal support size.

APPENDIX C

An Exponent Ledger for the Flattening Machine

Power notation such as $N^{O(\varepsilon)}$ is indispensable in expansion arguments, but it can conceal which inequalities actually require slack. This appendix rewrites the abstract flattening theorem as an exponent ledger. The purpose is not to optimize constants; it is to make the logical direction of every inequality visible and to show how much room is needed between the sparse, medium, and quasirandom scales.

1. The four exponents

Let $|G| = N$ and suppose a probability measure ν has

$$\|\nu\|_2 = N^{-\sigma}.$$

Uniform measure corresponds to $\sigma = 1/2$ and a point mass to $\sigma = 0$. The effective-support exponent is 2σ because

$$\|\nu\|_2^{-2} = N^{2\sigma}.$$

Suppose failed flattening at exponent ε produces an approximate group H with

$$|H| = N^{2\sigma + O(\varepsilon)}$$

and a coset carrying mass $N^{-O(\varepsilon)}$. The group-specific product theorem is available only in a medium interval

$$N^\delta \leq |H| \leq N^{1-\delta}.$$

Thus we need

$$(92) \quad 2\sigma - C\varepsilon > \delta$$

and

$$(93) \quad 2\sigma + C\varepsilon < 1 - \delta.$$

2. Entering the medium interval

If the initial spreading stage gives

$$\|v\|_2 \leq N^{-\alpha},$$

then $\sigma \geq \alpha$, so the lower inequality follows from

$$2\alpha - C\varepsilon > \delta.$$

If flattening is stopped once

$$\|v\|_2 \leq N^{-1/2+\theta},$$

then before stopping one has $\sigma < 1/2 - \theta$, and the upper inequality follows from

$$2\theta - C\varepsilon > \delta.$$

These are the two inequalities hidden in the phrase “choose ε sufficiently small.”

3. Subgroup escape must dominate the capture loss

Suppose the product theorem controls H by at most $N^{C\varepsilon}$ cosets of a proper subgroup L , while non-concentration gives

$$v(gL) \leq N^{-\kappa}.$$

Then

$$v(gH) \leq N^{C\varepsilon-\kappa}.$$

The inverse theorem gives the opposite bound

$$v(gH) \geq N^{-C\varepsilon}.$$

Contradiction follows as soon as

$$(94) \quad 2C\varepsilon < \kappa.$$

Thus the non-concentration exponent need not be large; it only has to dominate the polynomial losses accumulated by the inverse theorem and product theorem.

4. How many flattening stages?

If every successful squaring improves

$$\|v\|_2 \mapsto N^{-\varepsilon}\|v\|_2,$$

then the exponent σ increases by ε . Starting at $\sigma \geq \alpha$ and aiming for $1/2 - \theta$ requires at most

$$J \leq \frac{1/2 - \theta - \alpha}{\varepsilon} + 1$$

stages. All parameters are fixed independently of N , so $J = O(1)$. If the starting walk has length $n_0 = O(\log N)$, repeated squaring changes the time to at most

$$2^J n_0 = O(\log N).$$

This is the quantitative reason that a *power* improvement is needed. A fixed contraction such as 0.99 would require $\asymp \log N$ stages and would turn logarithmic time into a polynomial time scale.

5. A sample consistent choice

Suppose, for illustration, that initial spreading yields $\alpha = 1/20$, the quasirandom endgame will start at $\theta = 1/30$, and subgroup escape gives $\kappa = 1/100$. Choose a medium-scale margin $\delta = 1/40$. Then

$$2\alpha - \delta = \frac{3}{40}, \quad 2\theta - \delta = \frac{1}{24}.$$

If C denotes the largest absolute polynomial-loss constant in the inverse and control statements, it is enough to choose ε satisfying

$$C\varepsilon < \min\left(\frac{3}{80}, \frac{1}{48}, \frac{1}{200}\right).$$

No numerical value of C is conceptually important. What matters is the ordering of choices:

$$\alpha, \theta, \kappa \implies \delta \implies \varepsilon.$$

Choosing the flattening exponent before fixing the geometric margins is a common source of circular bookkeeping.

Bibliography

1. Terence Tao, *Product set estimates for non-commutative groups*, Combinatorica **28** (2008), no. 5, 547–594.

The Bourgain–Gamburd Method
Volume III: Sum–Product in Finite Fields

Preface

Finite-field sum–product is not a decorative lemma in the Bourgain–Gamburd method. In rank one it is the arithmetic principle that prevents a set of matrix traces from behaving simultaneously like an additive object and a multiplicative object. Some compressed accounts quote a sum–product theorem before using it inside Helfgott’s growth argument. Here we reverse that order: the theorem is proved here before any matrix argument is allowed to use it.

The route is deliberately elementary in spirit. We prove Cauchy–Davenport, the dilation-growth lemma, the representation-strength form of Balog–Szemerédi–Gowers needed by Katz and Tao, the Katz–Tao mixed refinement, and the polynomial-closure lemma. Only after those pieces are visible do we run the Bourgain–Katz–Tao contradiction. The paper of Bourgain, Katz and Tao [1] remains the historical source of the architecture, but the results used later in this series are proved in the present volume.

A recurring principle is worth stating at the outset:

small $A + A$ and small $AA \implies$ an approximate finite ring $\implies$ too much of $\mathbb{F}_p$.

The last implication is where the assumption that the field has prime order enters. A genuine proper subfield would be an exact counterexample, so the proof must at some point use the absence of such a subfield rather than merely manipulate cardinalities.

Series position. This volume uses Volumes I–II and elementary finite-field algebra.

Its main outputs are finite-field sum–product and the trace-shaped arithmetic inputs used in rank-one growth.

Quantifier and asymptotic conventions. Unless a statement explicitly says otherwise, every constant occurring in a quantitative theorem is chosen in the order in which it is quantified in that theorem. Thus a phrase such as “for all sufficiently small δ ” means: there is a displayed or locally quantified threshold $\delta_0 > 0$, depending only on the parameters listed before that phrase, such that the assertion holds for every $0 < \delta < \delta_0$. The phrases “sufficiently large N ” and “sufficiently large p ” have the analogous meaning.

If $X \ll_\theta Y$ or $X = O_\theta(Y)$, the implicit constant depends only on θ and on the fixed ambient data of the current volume; it is independent of the variables quantified after θ . The words *absolute*, *uniform*,

and *bounded* never introduce an unlisted dependence. Whenever an inequality between exponents or scales is essential to the argument, that inequality is written explicitly in the proof; these conventions are not a substitute for such a calculation.

Contents

Preface	iii
Part 1. Prime-Field Combinatorics	1
Chapter 1. Prime-Field Additive Tools from First Principles	2
1. A finite interpolation identity	2
2. Cauchy–Davenport with proof	3
3. A generic-dilation growth lemma	4
4. Bounded additive combinations from the earlier inverse theory	6
5. Multiplicative specialization	6
Chapter 2. Representation BSG and the Katz–Tao Refinement	8
1. The representation-strength BSG theorem	8
2. The Katz–Tao mixed refinement	12
3. What the lemma means	15
4. Dependencies	15
Part 2. Approximate Rings and Polynomial Closure	16
Chapter 3. Approximate Rings and Polynomial Closure	17
1. Essential containment	17
2. Good scalars	18
3. Why pointwise goodness is not enough	19
4. Polynomial closure theorem	20
5. Dependencies	21
Chapter 4. Linear Filling and Rank Reduction	22
1. From a moderately large set to a bounded-rank surjection	22
2. Eliminating one coordinate	23
3. Iterating the reduction	25
Part 3. The Sum–Product Theorem	26
Chapter 5. The Bourgain–Katz–Tao Sum–Product Theorem	27
1. The proof architecture in one diagram	28

2. Where every hypothesis is used	28
3. Source concordance and dependencies	30
Chapter 6. General Finite Fields and the Subfield Obstruction	31
1. A sharper generic-dilation lemma	31
2. Bounded-rank filling over an arbitrary finite field	32
3. Collision rank reduction without a primality assumption	33
4. When injectivity forces a subfield	35
5. The inverse theorem: subfields are the only obstruction	36
6. A direct sum–product corollary away from subfields	38
7. Dependencies	40
Part 4. Interfaces and Trace-Shaped Expansion	41
Chapter 7. Interfaces for Growth in SL_2	42
1. Safe normalizations and unsafe ones	42
2. A trace-shaped laboratory	43
3. Range bookkeeping	44
4. Worked numerical ledger	45
5. What has been proved, and what is left	46
6. Source concordance	46
Chapter 8. Quadratic Extensions and the Norm-One Torus	47
1. The only proper subfield	47
2. Quadratic-extension sum–product from the inverse theorem	47
3. The norm-one group	49
4. Why this is the natural nonsplit-torus interface	51
5. Dependencies	52
Chapter 9. Trace-Shaped Expansion Functions	53
1. Elementary fibers of the trace parameter	53
2. From w -products to sum–product	54
3. A coefficient-transfer lemma from common approximate-group control	56
4. Square cosets and the transverse trace expression	58
5. The exact interface with the transverse torus calculation	61
6. Concordance with Helfgott’s expanding-function step	61
Part 5. Technical Laboratories	62
Appendix A. The Katz–Tao Telescoping Argument Under a Microscope	63
1. Why a representation-rich set is needed	63
2. The four increments	63
3. Why each increment lies in the same controlled set	64
4. The injectivity ledger	65

5. A symbolic exponent ledger	65
6. What would fail without representation richness	66
Appendix B. Rank Reduction in a Concrete Extension Field	67
1. The field $\mathbb{F}_9$	67
2. A bijective rank-two map with no collision	67
3. Perturb the subfield by one point	68
4. The norm-one group in $\mathbb{F}_9$	68
Appendix C. Generic Projections and Filling: Worked Finite-Field Examples	69
1. A complete table in $\mathbb{F}_7$	69
2. Projection energy	69
3. Sparse and dense regimes	70
4. Why coefficients are allowed to change	71
5. A small-field fallback is not a loophole	71
Bibliography	72

Part 1

Prime-Field Combinatorics

CHAPTER 1

Prime-Field Additive Tools from First Principles

We establish the elementary field-combinatorial inputs used by the sum–product proof. The main points are Cauchy–Davenport, proved from a finite interpolation identity rather than quoted; a random-dilation lemma that makes a Cartesian product grow under a generic linear projection; and the bounded-combination consequences of the Ruzsa/approximate-group theory proved in Volume II. These tools explain the two ways in which the prime field fights concentration: addition eventually fills the field, while a generic change of slope turns two moderately large sets into a much larger sumset.

Throughout this volume, p is prime and $F = \mathbb{F}_p$. For subsets $A, B \subseteq F$ we use additive notation

$$A + B = \{a + b : a \in A, b \in B\}, \quad A - B = \{a - b : a \in A, b \in B\},$$

and multiplicative notation $AB = \{ab : a \in A, b \in B\}$.

1. A finite interpolation identity

The usual polynomial proof of Cauchy–Davenport is sometimes presented by invoking the Combinatorial Nullstellensatz. That would simply move the work to another theorem. We need only the following coefficient identity, whose proof is one-variable Lagrange interpolation iterated across coordinates.

LEMMA 1.1 (Coefficient extraction on a finite grid). *Let K be a field. For $1 \leq i \leq n$, let $S_i \subset K$ have cardinality $d_i + 1$, and put*

$$\phi_i(t) = \prod_{s \in S_i} (t - s).$$

If $P \in K[x_1, \dots, x_n]$ has total degree at most $d_1 + \dots + d_n$, then the coefficient of $x_1^{d_1} \dots x_n^{d_n}$ in P equals

$$(1) \quad \sum_{(s_1, \dots, s_n) \in S_1 \times \dots \times S_n} \frac{P(s_1, \dots, s_n)}{\phi_1'(s_1) \dots \phi_n'(s_n)}.$$

PROOF. First consider one variable. If $S = \{s_0, \dots, s_d\}$ and Q has degree at most d , Lagrange interpolation gives

$$Q(t) = \sum_{s \in S} Q(s) L_s(t), \quad L_s(t) = \prod_{u \in S \setminus \{s\}} \frac{t - u}{s - u}.$$

The coefficient of t^d in L_s is

$$\frac{1}{\prod_{u \neq s} (s - u)} = \frac{1}{\phi'(s)}.$$

Hence

$$(2) \quad [t^d]Q(t) = \sum_{s \in S} \frac{Q(s)}{\phi'(s)}.$$

For several variables, apply the one-variable interpolation operator successively in $x_1, \dots, x_n$. Equivalently, expand P as a sum of monomials. The right side of (1) factors on a monomial:

$$\prod_{i=1}^n \left(\sum_{s_i \in S_i} \frac{s_i^{e_i}}{\phi'_i(s_i)} \right).$$

By (2), the i th factor is 0 for $e_i < d_i$ and is 1 for $e_i = d_i$. Because the total degree of P is at most $\sum_i d_i$, a monomial different from $x_1^{d_1} \cdots x_n^{d_n}$ cannot have $e_i \geq d_i$ for every i . Thus every other monomial contributes zero, while the target monomial contributes one. This proves the formula. $\square$

COROLLARY 1.2 (Grid nonvanishing criterion). *Under the hypotheses of Lemma 1.1, if the coefficient of $x_1^{d_1} \cdots x_n^{d_n}$ is nonzero, then P does not vanish identically on $S_1 \times \cdots \times S_n$.*

PROOF. If P vanished on the entire grid, the right side of (1) would be zero. $\square$

2. Cauchy–Davenport with proof

THEOREM 1.3 (Cauchy–Davenport). *For nonempty $A, B \subseteq \mathbb{F}_p$,*

$$|A + B| \geq \min(p, |A| + |B| - 1).$$

PROOF. Put $m = |A|$ and $n = |B|$. If $m + n - 1 > p$, choose subsets $A_0 \subseteq A$ and $B_0 \subseteq B$ with

$$|A_0| + |B_0| - 1 = p.$$

The case of equality therefore implies $|A_0 + B_0| = p$ and hence $A + B = F$. It is enough to treat $m + n - 1 \leq p$.

Suppose for contradiction that

$$|A + B| \leq m + n - 2.$$

Choose a set $D \subseteq F$ containing $A + B$ and having exactly $m + n - 2$ elements. Consider

$$P(x, y) = \prod_{d \in D} (x + y - d).$$

It has total degree $m + n - 2$ and vanishes on $A \times B$. The coefficient of $x^{m-1}y^{n-1}$ is the same as in $(x + y)^{m+n-2}$, namely

$$\binom{m+n-2}{m-1}.$$

Because $m + n - 2 < p$, none of the factorials occurring in this binomial coefficient is divisible by p , so the coefficient is nonzero in $\mathbb{F}_p$. Applying [Corollary 1.2](#) with $d_1 = m - 1$ and $d_2 = n - 1$ contradicts the fact that P vanishes on $A \times B$. $\square$

REMARK 1.4 (Where primality is used). The proof needs the top binomial coefficient to be nonzero modulo the characteristic. The theorem itself is special to the cyclic group of prime order. More importantly for sum–product, repeated Cauchy–Davenport lets any subset of positive density additively fill $\mathbb{F}_p$ after boundedly many summations; a proper subfield in $\mathbb{F}_{p^k}$ would prevent the analogous ring argument from forcing the whole field.

3. A generic-dilation growth lemma

The next lemma is a finite-field projection statement. Think of $A \times B \subseteq F^2$ and project it by

$$(a, b) \mapsto a + \xi b.$$

For a generic slope ξ , most pairs should not collide.

LEMMA 1.5 (Dilation growth). *Let $A, B \subseteq F$ be finite and nonempty. There exists $\xi \in F^\times$ such that*

$$(3) \quad |A + \xi B| \geq \min \left(\frac{5}{8} |A| |B|, \frac{1}{10} p \right).$$

PROOF. Write $M = |A| |B|$. We first assume

$$(4) \quad M \leq \frac{p}{2}.$$

For fixed $\xi \neq 0$,

$$A + \xi B = \bigcup_{a \in A} (a + \xi B).$$

The elementary second-order inclusion–exclusion inequality gives

$$|A + \xi B| \geq |A| |B| - \frac{1}{2} \sum_{\substack{a, a' \in A \\ a \neq a'}} |(a + \xi B) \cap (a' + \xi B)|$$

$$= M - \frac{1}{2} \sum_{\substack{a \neq a' \\ b \neq b'}} 1_{\{\xi = (a-a')/(b'-b)\}}.$$

Average over $\xi \in F^\times$. Every quadruple with $a \neq a'$ and $b \neq b'$ contributes for exactly one nonzero ξ , hence

$$(5) \quad \frac{1}{p-1} \sum_{\xi \in F^\times} |A + \xi B| \geq M - \frac{|A|(|A|-1)|B|(|B|-1)}{2(p-1)}$$

$$(6) \quad \geq M - \frac{M^2}{2(p-1)}.$$

If $p \geq 3$ and $M \leq p/2$, then

$$\frac{M}{2(p-1)} \leq \frac{p}{4(p-1)} \leq \frac{3}{8},$$

so the last member of (5) is at least $5M/8$. Thus some ξ gives

$$|A + \xi B| \geq \frac{5}{8}M.$$

The case $p = 2$ is immediate.

It remains to remove the auxiliary hypothesis. Suppose $M > p/2$. If $p < 11$, then the right-hand side of (3) is at most $p/10 < 1$, whereas $A + \xi B$ is nonempty for every $\xi \in F^\times$; hence the desired inequality is automatic. We may therefore assume $p \geq 11$. We claim that there are nonempty subsets $A_0 \subseteq A$ and $B_0 \subseteq B$ such that

$$(7) \quad \frac{p}{5} \leq |A_0||B_0| \leq \frac{p}{2}.$$

Indeed, put $a = |A|$. If $a \geq \lceil p/5 \rceil$, take $|B_0| = 1$ and $|A_0| = \lceil p/5 \rceil$. Otherwise take $A_0 = A$ and

$$|B_0| = \left\lfloor \frac{p}{2a} \right\rfloor.$$

The assumption $a|B| > p/2$ ensures that this integer is at most $|B|$ and is positive. Moreover

$$|A_0||B_0| > \frac{p}{2} - a > \frac{p}{2} - \left(\frac{p}{5} + 1\right) \geq \frac{p}{5}$$

for $p \geq 11$, while the upper bound in (7) is immediate. Apply the already proved case to A_0, B_0 . For some $\xi \neq 0$,

$$|A + \xi B| \geq |A_0 + \xi B_0| \geq \frac{5}{8}|A_0||B_0| \geq \frac{p}{8} > \frac{p}{10}.$$

This proves (3) in all cases. $\square$

GUIDING IDEA 1.6 (Why this lemma is stronger than Cauchy–Davenport). Cauchy–Davenport forces linear growth under a fixed addition. Randomizing the slope produces essentially multiplicative growth, $|A + \xi B| \asymp |A||B|$, until the ambient field size becomes the obstruction. The sum–product proof will arrange that the slope ξ

itself is generated by A ; then this forced growth conflicts with the assumption that every algebraic expression generated by A stays small.

4. Bounded additive combinations from the earlier inverse theory

We record explicitly how Volume II enters. Its small-product theorem was proved in an arbitrary group. Specialize to the additive group of F . If

$$(8) \quad |A + B| \leq K \min(|A|, |B|),$$

then the theorem provides a $K^{O(1)}$ -approximate subgroup H of $(F, +)$ and $K^{O(1)}$ translates of H covering A and B , with $|H| \leq K^{O(1)} \min(|A|, |B|)$. Since bounded powers of an approximate group have polynomial size, we obtain the following result without invoking an external Plünnecke theorem.

PROPOSITION 1.7 (Bounded additive-word estimate). *Assume (8). For each fixed $r \geq 1$ and every choice of signs $\varepsilon_1, \dots, \varepsilon_r \in \{-1, 1\}$,*

$$|\varepsilon_1 A + \dots + \varepsilon_r A| \leq K^{C_r} |A|,$$

and similarly with B or with a fixed bounded mixture of A and B . Here C_r depends only on the length of the expression.

PROOF. By the proved small-product theorem of Volume II, write $A \subseteq X + H$ with $|X| \leq K^{O(1)}$ and H a $K^{O(1)}$ -approximate subgroup satisfying $|H| \leq K^{O(1)} |A|$. Then

$$\varepsilon_1 A + \dots + \varepsilon_r A \subseteq (\varepsilon_1 X + \dots + \varepsilon_r X) + rH,$$

where symmetry of H absorbs signs. The first factor has at most $K^{O_r(1)}$ elements and the second has at most $K^{O_r(1)} |H|$ elements by the bounded-power lemma proved in Volume II. Multiplying the bounds gives the claim. The mixed version is identical because the same approximate subgroup controls both A and B . $\square$

5. Multiplicative specialization

If $0 \notin A$ we may apply the same theorem to the abelian multiplicative group $F^\times$. Thus

$$(9) \quad |AA| \leq K|A|$$

implies, for every fixed signed multiplicative word,

$$|A^{\varepsilon_1} \dots A^{\varepsilon_r}| \leq K^{O_r(1)} |A|.$$

Removing 0 changes the size of a nontrivial set by at most one; throughout asymptotic arguments we shall either remove it explicitly or isolate the tiny-set case first.

CHECKPOINT 1.8. At this point every elementary additive-combinatorial tool used later is internal: finite interpolation, Cauchy–Davenport, generic dilation, Ruzsa/covering theory, and bounded additive or multiplicative words. The genuinely mixed step—turning simultaneous smallness of $A + A$ and AA into control of $AA - AA$ on a large subset—is the subject of the next chapter.

CHAPTER 2

Representation BSG and the Katz–Tao Refinement

The mixed additive–multiplicative step in the Bourgain–Katz–Tao proof is often compressed into the words “by the Katz–Tao lemma.” Here we prove it. The key strengthening of BSG is not merely that a large subset has a small complete sumset: every element of that complete sumset is forced to possess many alternating three-term representations by popular restricted sums. This representation multiplicity is exactly what allows multiplication by a whole quotient set without paying a factor equal to its cardinality. We then repeat the argument multiplicatively and obtain a large subset A' for which $A'A' - A'A'$ is small.

We use the notation

$$X \lesssim_L Y$$

to mean $X \leq L^C Y$ for some absolute exponent C that may change from line to line. A subscript involving further fixed parameters has the analogous meaning. In this chapter L will itself be a polynomial in the original doubling parameter, so this notation never hides a non-polynomial loss.

1. The representation-strength BSG theorem

The next theorem is a consequence of the proof of BSG, not merely of its usual final statement. We give the extraction again because the extra conclusion is what the Katz–Tao argument consumes.

THEOREM 2.1 (Representation BSG). *Let G be an abelian group, written multiplicatively for the moment. Let $X, Y \subseteq G$ and suppose that for some $N \geq 1$ and $L \geq 2$,*

$$(10) \quad L^{-1}N \leq |X|, |Y| \leq N, \quad |XY| \leq LN.$$

Then there are subsets $X_* \subseteq X$, $Y_* \subseteq Y$ with

$$(11) \quad |X_*|, |Y_*| \geq L^{-C}N$$

for an absolute C , such that every $z \in X_*Y_*$ has at least

$$(12) \quad L^{-C}N^5$$

representations of the form

$$(13) \quad z = (x_1y_1)(x_2y_2)^{-1}(x_3y_3), \quad x_i \in X, y_i \in Y.$$

All constants are absolute.

PROOF. We replay the fully proved BSG mechanism from Volume II while keeping the representation counts that are normally discarded at the end.

Step 1: energy and popular products. By Cauchy–Schwarz,

$$E(X, Y) \geq \frac{|X|^2|Y|^2}{|XY|}.$$

Using (10),

$$E(X, Y) \geq L^{-2}|X|^{3/2}|Y|^{3/2}.$$

Thus the BSG energy parameter may be taken to be

$$K_0 = L^2.$$

Let

$$C = \left\{ c : r_{X,Y}(c) > \frac{\sqrt{|X||Y|}}{2K_0} \right\}.$$

Exactly as in the popular-product calculation of Volume II,

$$(14) \quad |C| \leq 2K_0\sqrt{|X||Y|} \leq 2L^2N,$$

$$(15) \quad \sum_{c \in C} r_{X,Y}(c) \geq \frac{|X||Y|}{2K_0}.$$

Every $c \in C$ therefore has at least

$$(16) \quad \frac{\sqrt{|X||Y|}}{2K_0} \geq \frac{N}{2L^3}$$

representations $c = xy$ with $x \in X, y \in Y$.

Step 2: regularize the X side. For $x \in X$ let

$$d_C(x) = |\{y \in Y : xy \in C\}|.$$

From (15),

$$\sum_{x \in X} d_C(x) \geq \frac{|X||Y|}{2K_0}.$$

Set

$$X' = \left\{ x \in X : d_C(x) > \frac{|Y|}{4K_0} \right\}.$$

The complement contributes at most $|X||Y|/(4K_0)$, hence

$$\sum_{x \in X'} d_C(x) \geq \frac{|X||Y|}{4K_0}, \quad |X'| \geq \frac{|X|}{4K_0} \geq \frac{N}{4L^3}.$$

Write $|X| = M|X'|$, so $1 \leq M \leq 4K_0$.

Step 3: weak BSG and a quotient set with witnesses. Apply the weak BSG theorem proved in Volume II to X', Y, C , with exceptional proportion

$$\eta = \frac{1}{32K_0}.$$

Let K_w denote the density parameter in that theorem and K'_w its target-size parameter. Since

$$|E_C(X', Y)| = \sum_{x \in X'} d_C(x) \geq \frac{|X||Y|}{4K_0} \geq \frac{|X'||Y|}{4K_0},$$

we may take $K_w \leq 4K_0$. Also, because $|X|/|X'| \leq 4K_0$ and (14) holds,

$$|C| \leq 2K_0\sqrt{|X||Y|} \leq 4K_0^{3/2}\sqrt{|X'||Y|},$$

so $K'_w \leq 4K_0^{3/2}$. Thus every parameter entering weak BSG is polynomial in L . Inspecting its proved construction gives a set $X'' \subseteq X'$ and a set $\mathcal{D}$ such that

$$(17) \quad |X''| \geq L^{-C_1}N, \quad |\mathcal{D}| \leq L^{C_1}N,$$

all but an η fraction of pairs $(x, x') \in (X'')^2$ satisfy $xx'^{-1} \in \mathcal{D}$, and—this is the information we retain—every $d \in \mathcal{D}$ has at least

$$(18) \quad L^{-C_1}N$$

witnesses $c_0 \in C$ such that

$$c_0 \in C, \quad dc_0 \in C.$$

For completeness, this last assertion is not an extra theorem: in the weak BSG proof a quotient $d = xx'^{-1}$ is admitted to $\mathcal{D}$ only when there are more than $\eta|Y|/(2K_w^2)$ common neighbors y with $xy, x'y \in C$. Putting $c_0 = x'y$ turns those neighbors injectively into witnesses $c_0, dc_0 \in C$. Since $|Y| \geq N/L$ and $K_1 \leq L^{O(1)}$, the lower bound is (18).

Prune X'' exactly as in the full BSG proof: remove points whose bad quotient degree exceeds twice its average. The remaining set X''' satisfies

$$(19) \quad |X'''| \geq \frac{1}{2}|X''| \geq L^{-C_2}N,$$

and every $x \in X'''$ has at most $|X''|/(16K_0)$ points $x' \in X''$ with $xx'^{-1} \notin \mathcal{D}$.

Step 4: regularize the Y side. Every $x' \in X'' \subseteq X'$ has more than $|Y|/(4K_0)$ neighbors y with $x'y \in C$. Double counting and a popularity cutoff give

$$Y''' = \left\{ y \in Y : |\{x' \in X'' : x'y \in C\}| > \frac{|X''|}{8K_0} \right\}$$

with

$$(20) \quad |Y'''| \geq \frac{|Y|}{8K_0} \geq L^{-C_3}N.$$

Set $X_* = X'''$ and $Y_* = Y'''$.

Step 5: every complete product has many decompositions through C and $\mathcal{D}$. Fix

$$z = xy, \quad x \in X_*, y \in Y_*.$$

There are more than $|X''|/(8K_0)$ elements $x' \in X''$ with

$$c_2 := x'y \in C.$$

At most $|X''|/(16K_0)$ of them have

$$d := xx'^{-1} \notin \mathcal{D}.$$

Hence at least $L^{-C_4}N$ choices of x' satisfy both conditions. For each such choice,

$$z = dc_2.$$

By (18), the element d has at least $L^{-C_1}N$ choices of $c_0 \in C$ for which

$$c_1 := dc_0 \in C.$$

Therefore every $z \in X_*Y_*$ has at least

$$(21) \quad L^{-C_5}N^2$$

representations

$$z = c_1c_0^{-1}c_2, \quad c_0, c_1, c_2 \in C.$$

The choices counted above are genuinely distinct at the level of (c_0, c_1, c_2) : for fixed z and c_2 , $d = zc_2^{-1}$ is determined, while the chosen witness c_0 determines $c_1 = dc_0$.

Finally each $c_i \in C$ has at least $N/(2L^3)$ representations as x_iy_i by (16). Choosing these three representations independently and inserting them into

$$z = c_1c_0^{-1}c_2$$

produces at least

$$L^{-C_5}N^2 \left(\frac{N}{2L^3} \right)^3 \geq L^{-C}N^5$$

representations of the form (13). A six-tuple determines the three c_i , so no overcount has been introduced in the final multiplication of choices. This proves the theorem. $\square$

REMARK 2.2 (Additive form). In an additive group, apply the theorem with multiplication replaced by addition. Then every $z \in X_* + Y_*$ has at least $L^{-C}N^5$ representations

$$z = (x_1 + y_1) - (x_2 + y_2) + (x_3 + y_3).$$

Taking $Y = -X$ produces the alternating-difference representation used below.

2. The Katz–Tao mixed refinement

THEOREM 2.3 (Katz–Tao refinement, fully proved). *Let $A \subseteq F$ be finite, $|A| = N \geq 2$, and suppose*

$$(22) \quad |A + A| \leq KN, \quad |AA| \leq KN, \quad K \geq 2.$$

Then there exists $A' \subseteq A$ with

$$(23) \quad |A'| \geq K^{-C}N$$

and

$$(24) \quad |A'A' - A'A'| \leq K^C|A'|$$

for an absolute constant C .

PROOF. We may assume N is larger than an absolute constant, since otherwise the claim follows after enlarging C . Removing 0 loses at most one element; apply the argument to $A \setminus \{0\}$ and then regard the resulting set as a subset of A . Thus we assume $0 \notin A$.

Step 1: create an additively representation-rich difference set. By the bounded additive-word estimate of [Proposition 1.7](#),

$$|A - A| \leq K^{C_0}N.$$

Apply the additive form of [Theorem 2.1](#) to $X = A$ and $Y = -A$, taking $L = K^{C_0+2}$. If the theorem returns $C \subseteq A$ and $Y_* \subseteq -A$, write $Y_* = -D$ with $D \subseteq A$. Thus we obtain subsets $C, D \subseteq A$ with

$$(25) \quad |C|, |D| \geq K^{-C_1}N$$

such that every $z \in C - D$ has at least

$$(26) \quad K^{-C_1}N^5$$

representations

$$(27) \quad z = (a_1 - a_2) - (a_3 - a_4) + (a_5 - a_6), \quad a_i \in A.$$

Set

$$M = A^3/A^2 = \{a_1a_2a_3/(a_4a_5) : a_i \in A\}.$$

The multiplicative bounded-word estimate from Volume II gives

$$(28) \quad |M| \leq K^{C_2}N.$$

We claim

$$(29) \quad |(C - D)M| \leq K^{C_3}N.$$

Indeed, choose any $w \in (C - D)M$ and one factorization $w = zm$ with $z \in C - D$ and $m \in M$. Multiplying each of the representations (27) by m yields at least $K^{-C_1}N^5$ representations

$$w = b_1 - b_2 - b_3 + b_4 + b_5 - b_6,$$

after regrouping the signs, with every $b_i \in A^4/A^2$. The set

$$T = A^4/A^2$$

has cardinality at most $K^{C_4}N$ by multiplicative bounded-word control. Hence the total number of possible six-tuples is at most $K^{6C_4}N^6$. Since every w has at least $K^{-C_1}N^5$ such tuples,

$$|(C - D)M| \leq K^{C_1+6C_4}N,$$

which proves (29).

Step 2: refine C, D multiplicatively. Because $C, D \subseteq A$,

$$|CD| \leq |AA| \leq KN,$$

while (25) makes $|C|$ and $|D|$ polynomially comparable with N . Apply [Theorem 2.1](#) in the abelian multiplicative group $F^\times$. We obtain $C' \subseteq C$ and $D' \subseteq D$ with

$$(30) \quad |C'|, |D'| \geq K^{-C_5}N$$

such that every $u \in C'D'$ has at least

$$(31) \quad K^{-C_5}N^5$$

representations

$$(32) \quad u = \frac{c_1 d_1 c_3 d_3}{c_2 d_2}, \quad c_i \in C, d_i \in D.$$

Step 3: subtract two products and telescope through the small set $(C - D)M$. Fix $u = cd \in C'D'$ and $v = c'd' \in C'D'$. Among the at least $K^{-C_5}N^5$ representations (32) of u , there are only $|C||D| \leq N^2$ possibilities for (c_2, d_2) . By pigeonhole, some fixed pair (c_2, d_2) occurs in at least

$$(33) \quad K^{-C_5}N^3$$

representations, with c_1, c_3, d_1, d_3 still varying.

For each such representation define

$$\begin{aligned} x_1 &= (c_1 - d') \frac{d_1 c_3 d_3}{c_2 d_2}, \\ x_2 &= d' (c' - d_1) \frac{c_3 d_3}{c_2 d_2}, \\ x_3 &= d' c' (c_3 - d_2) \frac{d_3}{c_2 d_2}, \\ x_4 &= d' c' d_2 (c_2 - d_3) \frac{1}{c_2 d_2}. \end{aligned}$$

A direct telescoping calculation gives

$$(34) \quad u - v = x_1 - x_2 + x_3 - x_4.$$

Every x_i belongs to $(C - D)M$. For x_1, x_2, x_3 this is immediate from the displayed formulas. For x_4 , the multiplicative coefficient is $d'c'/c_2 \in A^2/A$; multiplying numerator and denominator by any fixed element of A places it in $A^3/A^2 = M$.

For fixed c, d, c', d', c_2, d_2 , the map

$$(c_1, c_3, d_1, d_3) \mapsto (x_1, x_2, x_3, x_4)$$

is injective. Indeed x_1 determines c_1 because all other factors in its formula are nonzero and fixed except c_3, d_1, d_3 ; more transparently, one may recover the variables successively from the four telescoping increments: x_4 determines d_3 , then x_3 determines c_3 , then x_2 determines d_1 , and finally x_1 determines c_1 . Thus (33) gives at least $K^{-C_5}N^3$ distinct four-tuples representing $u - v$ as in (34).

Put $W = (C - D)M$. By (29), $|W| \leq K^{C_3}N$. Counting all possible four-tuples in W^4 gives

$$|C'D' - C'D'| K^{-C_5}N^3 \leq |W|^4 \leq K^{4C_3}N^4.$$

Hence

$$(35) \quad |C'D' - C'D'| \leq K^{C_6}N.$$

Step 4: align C' and D' by a common dilation. From (35), in particular $|C'D'| \leq K^{C_6}N$. Apply the multiplicative Ruzsa triangle inequality in $F^\times$ to C' and $(D')^{-1}$; together with (30) it gives

$$(36) \quad |C'/D'| \leq K^{C_7}N.$$

The quotient map $C' \times D' \rightarrow C'/D'$ has $|C'||D'| \geq K^{-2C_5}N^2$ input pairs. By pigeonhole some $\lambda \in F^\times$ has at least $K^{-C_8}N$ representations $c/d = \lambda$. Equivalently,

$$A' := C' \cap \lambda D'$$

satisfies (23).

If $a, b \in A'$, write $b = \lambda d$ with $d \in D'$ and $a \in C'$. Then

$$ab \in \lambda C'D'.$$

Therefore

$$A'A' - A'A' \subseteq \lambda(C'D' - C'D').$$

Using (35) and $N \leq K^{C_8}|A'|$ gives

$$|A'A' - A'A'| \leq K^{C_6+C_8}|A'|.$$

This is (24). □

WARNING 2.4 (A sign correction worth making explicit). In the telescoping step the correct identity is $u - v = x_1 - x_2 + x_3 - x_4$ with x_4 defined using the factor $(c_2 - d_3)$. Writing a plus sign before x_4 would fail the cancellation. Because this four-term identity is exactly where additive structure meets multiplicative structure, we display it rather than hiding it under “a simple rearrangement.”

3. What the lemma means

The conclusion

$$|A'A' - A'A'| \leq K^{O(1)} |A'|$$

is much stronger than separate smallness of $A' + A'$ and $A'A'$. It says that the mixed bilinear expression

$$(a, b, c, d) \mapsto ab - cd$$

has only linearly many values on a large Cartesian power of A' . In the next chapter this becomes an approximate-ring closure statement: every fixed polynomial expression in A' has only polynomially many values.

4. Dependencies

The proof uses only results already proved in the series: Ruzsa calculus, weak and full BSG, and bounded powers of approximate groups from Volume II; and the elementary additive/multiplicative specializations recorded in [Chapter 1](#). The Katz–Tao refinement is therefore no longer an imported lemma. The architecture follows Bourgain–Katz–Tao [1], but all logical steps needed later have been supplied here.

Part 2

Approximate Rings and Polynomial Closure

CHAPTER 3

Approximate Rings and Polynomial Closure

The Katz–Tao refinement produces a large set B for which $BB - BB$ is small. This chapter explains why that single mixed estimate is the correct notion of an approximate ring. After normalizing $1 \in B$, we prove that every fixed polynomial expression generated by B has only polynomially many values. The proof uses a covering relation rather than an appeal to an inverse theorem: multiplication by a “good” scalar preserves a controlled additive model, good scalars are closed under addition, subtraction, and multiplication, and a separate induction controls the union over all monomials.

Let F be a field and $B \subseteq F$ finite and nonempty. Assume throughout the chapter that

$$(37) \quad 1 \in B, \quad |BB - BB| \leq K|B|, \quad K \geq 2.$$

Put

$$\Delta = B - B.$$

Since $B \subseteq BB$ by $1 \in B$,

$$(38) \quad |\Delta| \leq |BB - BB| \leq K|B|.$$

1. Essential containment

DEFINITION 3.1 (Polynomial covering relation). For sets $X, Y \subseteq F$ write

$$X \preccurlyeq_K Y$$

if there exists a set $T \subseteq F$ with $|T| \leq K^C$ for some absolute constant C —allowed to depend on the fixed expression under discussion—such that

$$X \subseteq T + Y.$$

When several applications occur in one fixed proof, the exponent may increase; only polynomial dependence on K matters.

This relation is transitive up to changing the polynomial exponent. It is also stable under a fixed finite number of additions and subtractions provided the reference set has controlled bounded additive words.

LEMMA 3.2 (Bounded additive combinations of B and Δ). *For every fixed $r \geq 1$,*

$$|\varepsilon_1 B + \cdots + \varepsilon_r B| \leq K^{O_r(1)} |B|,$$

and the same is true for bounded additive combinations of Δ . Moreover every such bounded combination is $\preccurlyeq_K \Delta$.

PROOF. The small difference bound (38) is a small-product hypothesis in the additive group: $|B + (-B)| \leq K|B|$. The small-product theorem proved in Volume II therefore places B inside polynomially many translates of one $K^{O(1)}$ -approximate subgroup H with $|H| \leq K^{O(1)} |B|$. Bounded signed sums of B and hence of $\Delta = B - B$ have size $K^{O_r(1)} |B|$.

To obtain covering by Δ , let X be any such bounded combination. Then $|X + B| \leq K^{O_r(1)} |B|$. The additive Ruzsa covering lemma, proved in Volume II, gives

$$X \subseteq T + (B - B) = T + \Delta$$

with $|T| \leq K^{O_r(1)}$. □

2. Good scalars

DEFINITION 3.3 (Good scalar). An element $x \in F$ is *good* if

$$(39) \quad xB \preccurlyeq_K \Delta.$$

The covering exponent is allowed to depend on the algebraic complexity by which x has been constructed from B .

LEMMA 3.4 (Every element of B is good). *If $b \in B$, then b is good.*

PROOF. We first cover the whole product set BB . Because $B \subseteq BB$,

$$BB - B \subseteq BB - BB,$$

so (37) gives $|BB - B| \leq K|B|$. Apply Ruzsa covering to the pair $(-B, BB)$, or equivalently to the small sum $BB + (-B)$. We obtain

$$BB \subseteq T + (B - B) = T + \Delta, \quad |T| \leq K.$$

For $b \in B$, $bB \subseteq BB$, hence $bB \preccurlyeq_K \Delta$. □

PROPOSITION 3.5 (Algebraic closure of good scalars). *If x and y are good, then $x + y$, $x - y$, and xy are good.*

PROOF. Suppose

$$xB \subseteq T_x + \Delta, \quad yB \subseteq T_y + \Delta, \quad |T_x|, |T_y| \leq K^{O(1)}.$$

Then

$$(x + y)B \subseteq xB + yB \subseteq T_x + T_y + \Delta + \Delta.$$

By Lemma 3.2, $\Delta + \Delta \preccurlyeq_K \Delta$, so $x + y$ is good. Replacing y by $-y$ gives $x - y$; note that $-yB = -(yB)$ and symmetry of Δ makes $-y$ good whenever y is.

For multiplication, start from

$$yB \subseteq T_y + \Delta.$$

Multiplying by x gives

$$xyB \subseteq xT_y + x\Delta.$$

The first term is a set of at most $K^{O(1)}$ translations. Since $\Delta = B - B$,

$$x\Delta = xB - xB.$$

Goodness of x gives $xB \preccurlyeq_K \Delta$, hence

$$xB - xB \preccurlyeq_K \Delta - \Delta \preccurlyeq_K \Delta$$

by Lemma 3.2. Therefore $xyB \preccurlyeq_K \Delta$. $\square$

COROLLARY 3.6 (Polynomially generated scalars are good). *Let $P \in \mathbb{Z}[t_1, \dots, t_m]$ be fixed. Every element of $P(B, \dots, B)$ is good, with a covering exponent depending only on P .*

PROOF. Build P from the variables, integer constants, addition, subtraction and multiplication. Variables take values in B and are good by Lemma 3.4. The element $1 \in B$ is good, so fixed integer constants are obtained by bounded additions and sign changes. Apply Proposition 3.5 inductively along the expression tree of P . $\square$

3. Why pointwise goodness is not enough

A subtlety remains. From “every $x \in X$ is good” one cannot immediately conclude that the union $XB = \bigcup_{x \in X} xB$ is small when X itself is large. The original proof needs a second induction showing that entire multiplicative powers of B are covered by polynomially many translates of Δ .

PROPOSITION 3.7 (Multiplicative powers are essentially additive). *For every fixed $k \geq 0$,*

$$(40) \quad B^k \preccurlyeq_K \Delta.$$

PROOF. For $k = 0$, $B^0 = \{1\}$ and the assertion is trivial. For $k = 1$, it is also trivial because for any fixed $b_0 \in B$,

$$B \subseteq b_0 + \Delta.$$

For $k = 2$, Lemma 3.4 actually proved $BB \preccurlyeq_K \Delta$.

Assume $k \geq 3$ and that

$$B^{k-1} \subseteq T + \Delta, \quad |T| \leq K^{O_k(1)}.$$

We may replace T by a subset of

$$B^{k-1} - \Delta$$

without increasing its size: for each translate $t + \Delta$ that meets B^{k-1} , choose a point of intersection $z = t + \delta$ and replace t by $z - \delta$. Consequently every $t \in T$ is a fixed

polynomial expression in elements of B . By [Corollary 3.6](#), every t is good with exponent depending only on k .

Multiply the covering of B^{k-1} by B :

$$B^k \subseteq TB + \Delta B = TB + (B - B)B \subseteq TB + BB - BB.$$

Because each $t \in T$ is good and $|T| \leq K^{O_k(1)}$,

$$TB \preccurlyeq_K \Delta.$$

Also $BB \preccurlyeq_K \Delta$, hence

$$BB - BB \preccurlyeq_K \Delta - \Delta \preccurlyeq_K \Delta.$$

Adding the two covered sets still gives a bounded additive combination of Δ , which by [Lemma 3.2](#) is $\preccurlyeq_K \Delta$. Thus $B^k \preccurlyeq_K \Delta$. $\square$

4. Polynomial closure theorem

THEOREM 3.8 (Polynomial closure of an approximate ring). *Assume (37). Let*

$$P \in \mathbb{Z}[t_1, \dots, t_m]$$

be a fixed polynomial. Then

$$(41) \quad |P(B, \dots, B)| \leq K^{C_P} |B|$$

for a constant C_P depending only on P .

PROOF. Write P as a fixed integer linear combination of finitely many monomials. A monomial of degree k takes values in B^k , and by [Proposition 3.7](#),

$$B^k \preccurlyeq_K \Delta.$$

Taking a fixed number of sums and differences therefore shows

$$P(B, \dots, B) \preccurlyeq_K \varepsilon_1 \Delta + \dots + \varepsilon_r \Delta$$

for some fixed r depending on P . By [Lemma 3.2](#), the right side has cardinality at most $K^{O_P(1)} |B|$. The covering set contributes another polynomial factor, proving (41). $\square$

GUIDING IDEA 3.9 (Approximate ring versus genuine ring). The hypothesis $|BB - BB| \ll K|B|$ says that the bilinear map $(b_1, b_2, b_3, b_4) \mapsto b_1 b_2 - b_3 b_4$ has only linearly many outputs. The theorem shows that this one mixed closure relation propagates to *every fixed algebraic expression*. In an infinite ring this resembles the first step of constructing a subring. In a prime field, the next chapters show that enough approximate closure forces the generated expressions to fill the whole field, contradicting their small cardinality.

5. Dependencies

The proof used only the small difference hypothesis, Ruzsa covering and the additive approximate-group consequences proved in Volume II, and elementary induction. No Freiman theorem and no classification of approximate rings has been invoked.

CHAPTER 4

Linear Filling and Rank Reduction

Polynomial closure says that every fixed expression in an approximate ring is small. To contradict this in a prime field we must manufacture one fixed expression that equals the entire field. The construction has two stages. Generic dilations give a bounded-rank linear surjection $B^k \rightarrow \mathbb{F}_p$. A collision in any surjective map of rank $k > 1$ then eliminates one coordinate, at the price of replacing B by a fixed polynomial expression in itself. Repeating the elimination reaches rank one, so one polynomial expression of B must already have p values.

1. From a moderately large set to a bounded-rank surjection

LEMMA 4.1 (Bounded-rank additive filling). *Let $0 < \delta \leq 1$ and let $B \subseteq F = \mathbb{F}_p$ satisfy*

$$|B| \geq p^\delta.$$

Then there is an integer $k \leq C/\delta$ and nonzero coefficients $\xi_1, \dots, \xi_k \in F^\times$ such that

$$(42) \quad F = \xi_1 B + \dots + \xi_k B.$$

Equivalently, the linear map

$$B^k \longrightarrow F, \quad (b_1, \dots, b_k) \longmapsto \sum_{j=1}^k \xi_j b_j$$

is surjective.

PROOF. We first build a sumset of positive density. Start with $S_1 = B$ and $\xi_1 = 1$. If $|S_j| \geq p/10$, stop. Otherwise apply Lemma 1.5 to S_j and B . There is $\xi_{j+1} \neq 0$ such that

$$|S_j + \xi_{j+1} B| \geq \min \left(\frac{1}{2} |S_j| |B|, \frac{p}{10} \right).$$

Set $S_{j+1} = S_j + \xi_{j+1} B$. Until the stopping threshold is reached, the second alternative in Lemma 1.5 has not yet occurred, and therefore

$$(43) \quad |S_{j+1}| \geq \frac{5}{8} |B| |S_j|.$$

Because $p^\delta > 1$ and $|B|$ is an integer, $|B| \geq 2$, so every nonterminal step grows by at least the factor $5/4$.

We now verify the claimed $O(1/\delta)$ number of steps rather than hiding it in the notation. If $p^\delta \geq (8/5)^2$, then

$$\log\left(\frac{5}{8}p^\delta\right) \geq \frac{1}{2}\delta \log p,$$

so (43) reaches size $p/10$ after at most $2/\delta + O(1)$ steps. If instead $p^\delta < (8/5)^2$, then

$$\log p < \frac{2 \log(8/5)}{\delta},$$

and growth by the fixed factor $5/4$ reaches $p/10$ after at most C/δ steps. Thus in every case the number of generic-dilation steps is bounded by C/δ with an absolute C .

We now have

$$S = \xi_1 B + \cdots + \xi_r B, \quad |S| \geq p/10, \quad r = O(1/\delta).$$

Repeated Cauchy–Davenport gives

$$|mS| \geq \min(p, m|S| - (m - 1)).$$

Taking $m = 101$ works uniformly for every prime. Indeed, if $p \geq 11$, then $|S| \geq p/10$ gives

$$101|S| - 100 \geq \frac{101}{10}p - 100 \geq p,$$

while if $p < 11$, the standing inequality $|B| \geq p^\delta > 1$ gives $|S| \geq 2$ and hence $101|S| - 100 \geq 102 > p$. Repeated Cauchy–Davenport therefore gives $101S = F$ in all cases. Expanding these copies of S produces (42) with $k = 101r = O(1/\delta)$. $\square$

REMARK 4.2 (Why Cauchy–Davenport appears only at the end). The generic-dilation lemma rapidly moves from size p^δ to positive density, but it cannot exceed the ambient size p . Cauchy–Davenport then converts positive density to exact equality with the field. These are two genuinely different growth mechanisms: random slope for multiplicative-size growth, then deterministic addition for saturation.

2. Eliminating one coordinate

LEMMA 4.3 (Rank reduction). *Let $B \subseteq \mathbb{F}_p$ be nonempty, and suppose $k > 1$ and*

$$(44) \quad F = \xi_1 B + \cdots + \xi_k B$$

for coefficients $\xi_j \in F$. Define

$$(45) \quad \tilde{B} = B(B - B) + B(B - B).$$

Then there exist coefficients $\tilde{\xi}_1, \dots, \tilde{\xi}_{k-1}$ such that

$$(46) \quad F = \tilde{\xi}_1 \tilde{B} + \cdots + \tilde{\xi}_{k-1} \tilde{B}.$$

Thus a surjection of rank k produces one of rank $k - 1$ from a fixed polynomial expression in B .

PROOF. The map

$$\Phi : B^k \rightarrow F, \quad \Phi(b_1, \dots, b_k) = \sum_{j=1}^k \xi_j b_j$$

is surjective. It cannot be injective. If it were, then $|B|^k = p$. Since p is prime and $k > 1$, no integer $|B| > 1$ has k th power equal to p ; while $|B| = 1$ cannot surject onto a field with more than one element.

Choose two distinct tuples $b = (b_1, \dots, b_k)$ and $b' = (b'_1, \dots, b'_k)$ with

$$(47) \quad \sum_{j=1}^k (b_j - b'_j) \xi_j = 0.$$

After reordering coordinates assume

$$\Delta_k := b_k - b'_k \neq 0.$$

Multiplication by Δ_k is a bijection of F , so (44) gives

$$F = \sum_{j=1}^k B \xi_j \Delta_k.$$

From (47),

$$\xi_k \Delta_k = - \sum_{j < k} \xi_j \Delta_j, \quad \Delta_j = b_j - b'_j.$$

Therefore

$$\begin{aligned} F &= \sum_{j < k} B \xi_j \Delta_k + B \xi_k \Delta_k \\ &\subseteq \sum_{j < k} \xi_j (B \Delta_k - B \Delta_j). \end{aligned}$$

Each bracket is contained in

$$B(B - B) + B(B - B) = \widetilde{B},$$

because a minus sign can be absorbed by reversing the difference in the second copy of $B - B$. Hence

$$F \subseteq \xi_1 \widetilde{B} + \dots + \xi_{k-1} \widetilde{B}.$$

The reverse inclusion is automatic because the right side lies in F . Thus equality holds, proving (46); one may take $\widetilde{\xi}_j = \xi_j$. $\square$

3. Iterating the reduction

Define recursively

$$B_0 = B, \quad B_{r+1} = B_r(B_r - B_r) + B_r(B_r - B_r).$$

For each fixed r , B_r is the value set of some integer polynomial P_r in finitely many variables evaluated on B . The complexity of P_r grows rapidly with r , but when $r \leq k = O(1/\delta)$ it depends only on δ .

PROPOSITION 4.4 (A polynomial expression fills the prime field). *Let $B \subseteq F$ satisfy $|B| \geq p^\delta$. Then there is an integer polynomial P_δ in $m(\delta)$ variables, depending only on δ and not on p or B , such that*

$$(48) \quad P_\delta(B, \dots, B) = F.$$

PROOF. By [Lemma 4.1](#), there is a surjection of rank $k = O(1/\delta)$ from B^k to F . Apply [Lemma 4.3](#) repeatedly. After at most $k - 1$ steps we obtain

$$F = \xi B_{k-1}$$

for some coefficient $\xi \neq 0$; a zero coefficient could not give a nontrivial surjection. Hence $B_{k-1} = F$ after multiplication by the fixed nonzero scalar ξ^{-1} .

Multiplication by ξ is a bijection of F , so the equality $F = \xi B_{k-1}$ implies

$$|B_{k-1}| = p.$$

But $B_{k-1} \subseteq F$, hence $B_{k-1} = F$. By construction B_{k-1} is the value set of a single integer polynomial obtained by iterating

$$X \mapsto X(X - X) + X(X - X)$$

at most $k - 1 = O(1/\delta)$ times. Its formula therefore depends only on δ , not on p , on B , or on the coefficients ξ_j chosen during the filling argument. Calling this polynomial P_δ proves (48). $\square$

WARNING 4.5 (What primality does here). The only obstruction to reducing the rank is injectivity of a surjective map $B^k \rightarrow F$. In a prime field, $|B|^k = p$ is impossible for $k > 1$. Over $\mathbb{F}_{p^m}$, it can happen that $|B| = p^r$ and $k = m/r$; a genuine subfield is exactly the algebraic structure that can survive the reduction. Thus the proof is already detecting the subfield obstruction before we name it explicitly.

Part 3

The Sum–Product Theorem

CHAPTER 5

The Bourgain–Katz–Tao Sum–Product Theorem

All ingredients are now internal, so the sum–product theorem becomes a short contradiction rather than a black box. Assuming both $A + A$ and AA are nearly as small as A , the Katz–Tao refinement produces a large approximate ring B . Polynomial closure says every fixed expression in B has only $|B|^{1+o(1)}$ values, while rank reduction constructs one expression that fills all of $\mathbb{F}_p$. The size range $p^\delta < |A| < p^{1-\delta}$ makes these statements incompatible.

THEOREM 5.1 (Finite-field sum–product). *For every $\delta \in (0, 1/2]$ there exist constants $\varepsilon = \varepsilon(\delta) > 0$ and $c = c(\delta) > 0$ such that for every prime p and every $A \subseteq \mathbb{F}_p$ satisfying*

$$(49) \quad p^\delta \leq |A| \leq p^{1-\delta},$$

one has

$$(50) \quad \max(|A + A|, |AA|) \geq c|A|^{1+\varepsilon}.$$

PROOF. Write $N = |A|$. Suppose the conclusion fails for an exponent $\varepsilon > 0$ to be chosen. After decreasing $c(\delta)$ if necessary, it is enough to rule out

$$(51) \quad |A + A| \leq KN, \quad |AA| \leq KN, \quad K = N^\varepsilon.$$

All harmless absolute constants may be absorbed by replacing K with CK , which changes powers of K only by constants once N is beyond a fixed threshold. The finitely many smaller cases are handled by choosing $c(\delta)$ sufficiently small.

Apply the fully proved Katz–Tao refinement [Theorem 2.3](#). There is $A' \subseteq A$ such that

$$(52) \quad |A'| \geq K^{-C_0}N, \quad |A'A' - A'A'| \leq K^{C_0}|A'|.$$

The set A' contains a nonzero element once N is sufficiently large compared with K^{C_0} ; otherwise (52) is impossible. Choose $a_0 \in A' \setminus \{0\}$ and normalize

$$B = a_0^{-1}A'.$$

Then $1 \in B$, $|B| = |A'|$, and scaling (52) gives

$$(53) \quad |BB - BB| \leq K^{C_0}|B|.$$

The set B is still large on the p scale. From $N \geq p^\delta$ and $K = N^\varepsilon$,

$$|B| \geq N^{1-C_0\varepsilon} \geq p^{\delta(1-C_0\varepsilon)}.$$

Choose initially

$$(54) \quad \varepsilon \leq \frac{1}{2C_0}.$$

Then

$$(55) \quad |B| \geq p^{\delta/2}.$$

Apply [Proposition 4.4](#) with $\delta/2$. There is a fixed polynomial P_δ , depending only on δ , such that

$$(56) \quad |P_\delta(B, \dots, B)| = p.$$

On the other hand, the polynomial-closure theorem [Theorem 3.8](#), applied to (53), gives

$$(57) \quad |P_\delta(B, \dots, B)| \leq (K^{C_0})^{C_\delta} |B| \leq N^{1+C_0C_\delta\varepsilon}.$$

Here C_δ depends only on the fixed polynomial produced by at most $O(1/\delta)$ rank-reduction steps.

The upper range in (49) is equivalent to

$$p \geq N^{1/(1-\delta)}.$$

Choose $\varepsilon(\delta) > 0$ still smaller, if necessary, so that besides (54),

$$(58) \quad 1 + C_0C_\delta\varepsilon < \frac{1}{1-\delta}.$$

Then (56) and (57) imply

$$N^{1/(1-\delta)} \leq p \leq N^{1+C_0C_\delta\varepsilon},$$

contradicting (58) for $N > 1$. This proves the theorem. $\square$

1. The proof architecture in one diagram

The proof can now be compressed without concealing anything: This contradicts $|A| \leq p^{1-\delta}$ when ε is sufficiently small.

2. Where every hypothesis is used

The lower size bound $|A| \geq p^\delta$ ensures that only $O_\delta(1)$ generic dilations are needed before a sumset has positive density. Without it, the polynomial P_δ would have complexity depending on $|A|$ and the closure estimate would no longer yield a fixed exponent.

The upper size bound $|A| \leq p^{1-\delta}$ creates the final numerical contradiction. Sets occupying almost all of F can of course have both $A + A$ and AA equal to the whole field, which is only a constant-factor expansion.

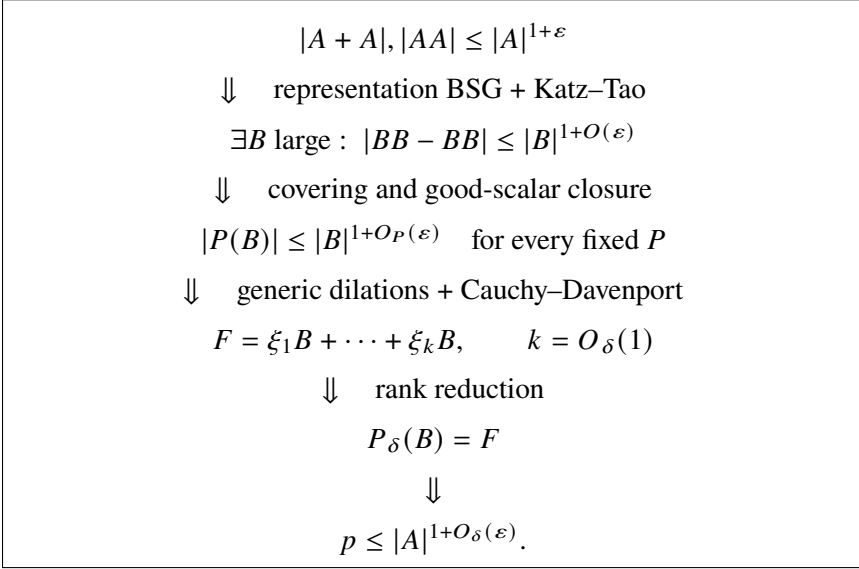


FIGURE 1. The Bourgain–Katz–Tao proof architecture in the prime-field regime. Each downward arrow has been proved earlier in the volume.

Primality enters twice: Cauchy–Davenport is strongest in the prime cyclic group, and the rank-reduction map cannot be a bijection $B^k \rightarrow \mathbb{F}_p$ for $k > 1$. Both manifestations express the same algebraic fact: there is no proper finite subfield.

EXAMPLE 5.2 (The subfield obstruction in $\mathbb{F}_9$, and what one transverse point does). Write

$$\mathbb{F}_9 = \mathbb{F}_3[\omega]/(\omega^2 + 1), \quad \omega^2 = -1 = 2.$$

The embedded prime field

$$A = \mathbb{F}_3 = \{0, 1, 2\}$$

has no sum–product growth at all:

$$A + A = A, \quad AA = A.$$

Thus

$$|A + A| = |AA| = |A| = 3.$$

The same phenomenon occurs for every scalar copy $\xi\mathbb{F}_3$: after rescaling it is just the proper subfield again. This is exactly why a finite-field sum–product theorem over a nonprime field must exclude concentration in scalar copies of proper subfields.

Now add one genuinely transverse element and set

$$B = \mathbb{F}_3 \cup \{\omega\}.$$

A direct enumeration gives

$$B + B = \{0, 1, 2, \omega, 2\omega, 1 + \omega, 2 + \omega\}, \quad |B + B| = 7,$$

while

$$BB = \{0, 1, 2, \omega, 2\omega\}, \quad |BB| = 5.$$

So one point outside the subfield already destroys simultaneous closure under both operations. The general inverse theorem of [Chapter 6](#) is a quantitative version of this toy calculation: persistent failure of sum–product growth forces the set back toward a scalar copy of a proper subfield.

3. Source concordance and dependencies

The theorem and the proof architecture originate in Bourgain–Katz–Tao [[1](#)]. In their paper several additive-combinatorial facts are recalled and the Katz–Tao mixed lemma is presented in compressed form. In this volume those dependencies have been reconstructed: Cauchy–Davenport, the required BSG representation multiplicity, the mixed refinement, polynomial closure, and rank reduction all have complete proofs. Thus [Theorem 5.1](#) is now a theorem proved within the series, not a citation interface.

CHAPTER 6

General Finite Fields and the Subfield Obstruction

Prime fields have no proper subfields, which is why approximate ring structure must eventually contradict the ambient size. Over a general finite field that conclusion is false: a subfield is simultaneously closed under addition and multiplication. This chapter proves the precise inverse theorem behind that obstruction. We fill in the argument that is only sketched in the original Bourgain–Katz–Tao paper: small sumset and product set force all but polynomially many elements of A into a scalar multiple of a genuine subfield. The proof isolates three principles: generic linear projections create a bounded-rank parametrization of the field; every collision lowers the rank while preserving polynomial complexity; and an injective parametrization can persist only when the current approximate ring has become an actual subfield.

Throughout this chapter F is a finite field of cardinality

$$q = |F|.$$

No assumption is made that q is prime.

1. A sharper generic-dilation lemma

For the general-field argument we need a version of the dilation lemma that approaches full density rather than stopping at a fixed small fraction of the field.

LEMMA 6.1 (Mean-energy dilation bound). *Let $X, Y \subseteq F$ be finite and nonempty, and put $M = |X||Y|$. There exists $\xi \in F^\times$ such that*

$$(59) \quad |X + \xi Y| \geq \frac{M(q-1)}{q-1+M}.$$

PROOF. For $\xi \in F^\times$ let

$$E_\xi = \left| \{(x, y, x', y') \in X \times Y \times X \times Y : x + \xi y = x' + \xi y'\} \right|.$$

Cauchy–Schwarz applied to the representation function of $X + \xi Y$ gives

$$(60) \quad |X + \xi Y| \geq \frac{M^2}{E_\xi}.$$

We average E_ξ over nonzero ξ . If $y = y'$, then the equality forces $x = x'$, so the diagonal quadruples contribute exactly

$$(q - 1)M$$

to $\sum_{\xi \neq 0} E_\xi$. If $y \neq y'$, then for fixed (x, y, x', y') there is at most one possible value

$$\xi = \frac{x' - x}{y - y'}.$$

It is nonzero only if $x \neq x'$. Consequently

$$\sum_{\xi \neq 0} E_\xi \leq (q - 1)M + |X|(|X| - 1)|Y|(|Y| - 1) \leq (q - 1)M + M^2.$$

Hence some $\xi \neq 0$ satisfies

$$E_\xi \leq M + \frac{M^2}{q - 1}.$$

Substituting this into (60) yields

$$|X + \xi Y| \geq \frac{M^2}{M + M^2/(q - 1)} = \frac{M(q - 1)}{q - 1 + M},$$

as claimed. $\square$

REMARK 6.2 (Two regimes in one formula). If $M \leq q - 1$, then (59) is at least $M/2$: a generic slope gives essentially multiplicative growth. If $M \gg q$, the right side is $(1 - o(1))q$: the same collision calculation drives the image to high density. This second regime is what allows the prime-field filling argument to survive without Cauchy–Davenport.

2. Bounded-rank filling over an arbitrary finite field

LEMMA 6.3 (Bounded-rank linear filling). *For every $\eta > 0$ there is an integer $k_0 = k_0(\eta)$ with the following property. If $S \subseteq F$ and*

$$|S| \geq q^\eta,$$

then for some $1 \leq k \leq k_0$ and some coefficients $\xi_1, \dots, \xi_k \in F^\times$,

$$(61) \quad F = \xi_1 S + \dots + \xi_k S.$$

PROOF. Write $N = |S|$. We first treat the range $q^\eta \geq 4$. Starting with $S_1 = S$, recursively choose

$$S_{j+1} = S_j + \xi_{j+1} S$$

using Lemma 6.1. As long as $|S_j|N \leq q - 1$, we have

$$|S_{j+1}| \geq \frac{1}{2}|S_j|N.$$

Because $N \geq q^\eta$ and $N/2 \geq q^{\eta/2}$ under the present assumption, after at most $2/\eta + 2$ steps we reach a set S_r for which

$$|S_r|N > q - 1.$$

One further application of Lemma 6.1 gives a set

$$T = S_r + \xi S$$

with

$$|T| > \frac{q-1}{2}.$$

Apply the same lemma to T and T . For $q \geq 7$ the lower bound

$$\frac{|T|^2(q-1)}{q-1+|T|^2}$$

is strictly larger than $q/2$. Thus for some $\zeta \neq 0$ the set

$$U = T + \zeta T$$

has more than half the elements of F . For every $x \in F$, the two sets U and $x - U$ both have cardinality greater than $q/2$, so they intersect. Hence

$$U + U = F.$$

Expanding $U + U$ expresses F as a sum of at most $4(r+1) = O_\eta(1)$ scalar copies of S .

It remains only to absorb the finite range excluded above. If $q^\eta < 4$ or $q < 7$, then $q \leq Q(\eta)$ for a constant depending only on η . Since $q^\eta > 1$, the set S contains distinct elements a, b . Put $d = b - a \neq 0$. Choose an $\mathbb{F}_p$ -basis $e_1, \dots, e_m$ of F , where $q = p^m$. For each i , repeat the coefficient e_i/d exactly $p-1$ times. By choosing in each corresponding input coordinate either a or b , the contribution in the e_i direction can be any element of $\mathbb{F}_p e_i$, up to one fixed global translation. Therefore this linear map from $S^{m(p-1)}$ onto F is surjective. Since $m(p-1) \leq q \leq Q(\eta)$, this also has bounded rank depending only on η . $\square$

GUIDING IDEA 6.4 (Why no additive classification is needed). The coefficients ξ_j are allowed to vary in the whole field. A set may lie in a proper additive $\mathbb{F}_p$ -subspace, but a generic scalar rotates that subspace. Repeating this boundedly many times makes a linear image of S^k fill F . The obstruction that eventually survives rank reduction is therefore not merely additive: it must be closed under multiplication as well, and hence be a subfield.

3. Collision rank reduction without a primality assumption

The algebraic elimination in the prime-field proof never used primality; only the guarantee of a collision did. We isolate the general statement.

LEMMA 6.5 (Collision rank reduction). *Let $T \subseteq F$ be nonempty and let $r > 1$. Suppose*

$$F = \xi_1 T + \cdots + \xi_r T$$

for some coefficients $\xi_j \in F$, and suppose the associated map

$$\Phi : T^r \rightarrow F, \quad \Phi(t_1, \dots, t_r) = \sum_{j=1}^r \xi_j t_j$$

is not injective. Put

$$\mathcal{R}(T) = T(T - T) + T(T - T).$$

Then there are coefficients $\tilde{\xi}_1, \dots, \tilde{\xi}_{r-1}$ such that

$$F = \tilde{\xi}_1 \mathcal{R}(T) + \cdots + \tilde{\xi}_{r-1} \mathcal{R}(T).$$

If moreover $0, 1 \in T$, then

$$T \subseteq \mathcal{R}(T) \quad \text{and} \quad 0, 1 \in \mathcal{R}(T).$$

PROOF. Choose distinct $t = (t_1, \dots, t_r)$ and $t' = (t'_1, \dots, t'_r)$ with $\Phi(t) = \Phi(t')$. After permuting coordinates assume

$$\Delta_r = t_r - t'_r \neq 0.$$

Thus

$$(62) \quad \sum_{j=1}^r \xi_j \Delta_j = 0, \quad \Delta_j = t_j - t'_j.$$

Multiplication by Δ_r is a bijection of F , so

$$F = \sum_{j=1}^r \xi_j T \Delta_r.$$

Using (62) to eliminate the last coefficient gives

$$\begin{aligned} F &= \sum_{j < r} \xi_j T \Delta_r + T \xi_r \Delta_r \\ &\subseteq \sum_{j < r} \xi_j (T \Delta_r - T \Delta_j) \subseteq \sum_{j < r} \xi_j \mathcal{R}(T). \end{aligned}$$

The reverse inclusion is automatic because the right side lies in F . This proves the rank reduction.

If $0, 1 \in T$, then for every $t \in T$,

$$t = 1 \cdot (t - 0) + 0 \cdot (0 - 0) \in \mathcal{R}(T),$$

so $T \subseteq \mathcal{R}(T)$. The same formula gives $0, 1 \in \mathcal{R}(T)$. □

4. When injectivity forces a subfield

LEMMA 6.6 (Finite semiring criterion). *Let $T \subseteq F$ be finite with $0, 1 \in T$. If*

$$T + T = T \quad \text{and} \quad TT = T,$$

then T is a subfield of F .

PROOF. The identity $T + T = T$ makes T a finite additive monoid containing 0. For $t \in T$, all positive integer multiples of t lie in T . Since the additive group of F is finite, t has finite additive order n , and

$$-t = (n - 1)t \in T.$$

Thus T is an additive subgroup.

Now let $t \in T \setminus \{0\}$. The identity $TT = T$ implies that all positive powers of t lie in T . Since $F^\times$ is finite, some positive power of t equals 1; hence t^{-1} is another positive power and belongs to T . Therefore the nonzero elements of T form a multiplicative group. Distributivity is inherited from F , so T is a subfield. $\square$

PROPOSITION 6.7 (Rank reduction stops only at a subfield). *Let $S \subseteq F$ contain 0 and 1. Suppose:*

- (i) *for some $k = O(1)$ there is a surjective linear map $S^k \rightarrow F$;*
- (ii) *every set obtained from S by a fixed number of the operations*

$$X \mapsto X + X, \quad X \mapsto XX, \quad X \mapsto X(X - X) + X(X - X)$$

has cardinality at most M .

Then S is contained in a subfield $G \leq F$ with $|G| \leq M$.

PROOF. We maintain a set T containing $S, 0, 1$ and a surjective linear map

$$\Phi : T^r \rightarrow F, \quad 1 \leq r \leq k.$$

Initially take $T = S$. If $r = 1$, then $F = \xi T$ for some $\xi \neq 0$, so $|T| = q$. Since $T \subseteq F$, we have $T = F$; take $G = F$. Hypothesis (ii) gives $q = |T| \leq M$.

Assume $r > 1$. If Φ is not injective, apply Lemma 6.5. Replace T by $\mathcal{R}(T)$ and reduce r by one. The new set still contains $S, 0, 1$.

Suppose instead that Φ is injective. Because it is also surjective,

$$(63) \quad |T|^r = q.$$

If $|T + T| > |T|$, then $T \subseteq T + T$ because $0 \in T$, so the same linear formula defines a surjection

$$(T + T)^r \rightarrow F.$$

Its domain has cardinality strictly larger than $|T|^r = q$, hence it cannot be injective. Applying Lemma 6.5 with $T + T$ in place of T lowers the rank.

If $|T + T| = |T|$ but $|TT| > |T|$, use $T \subseteq TT$ (because $1 \in T$) in exactly the same way. The map $(TT)^r \rightarrow F$ is surjective and noninjective, so another collision lowers the rank.

The only remaining possibility is

$$|T + T| = |T| \quad \text{and} \quad |TT| = |T|.$$

The containments $T \subseteq T + T$ and $T \subseteq TT$ therefore imply equality. By Lemma 6.6, T is a subfield. Take $G = T$. Hypothesis (ii) gives $|G| \leq M$.

Each nonterminal step lowers the rank, so after at most $k - 1$ steps one of the two terminal cases occurs. Every replacement is among the permitted operations, completing the proof. $\square$

5. The inverse theorem: subfields are the only obstruction

THEOREM 6.8 (Approximate fields are controlled by subfields). *Fix $0 < \delta < 1$. There is $C = C(\delta)$ such that the following holds. Let $A \subseteq F$ satisfy*

$$(64) \quad |A| > |F|^\delta, \quad |A + A| \leq K|A|, \quad |AA| \leq K|A|, \quad K \geq 2.$$

Then there exist a subfield $G \leq F$, a nonzero scalar $\xi \in F^\times$, and a set $X \subseteq F$ such that

$$(65) \quad |G| \leq K^C |A|, \quad |X| \leq K^C, \quad A \subseteq \xi G \cup X.$$

PROOF. Write $N = |A|$ and $q = |F|$. All constants below may depend on δ . If

$$q \leq K^{C_0/\delta} N$$

for a sufficiently large absolute C_0 , then the conclusion is immediate with $G = F$, $\xi = 1$, and $X = \emptyset$. Thus we may assume throughout that q is larger than every fixed polynomial in K that occurs below. In particular, losses of the form K^{-C_1} in cardinality preserve a power lower bound $q^{\delta/2}$ after increasing C_0 .

Step 1: extract and normalize an approximate ring. Apply the fully proved Katz–Tao refinement, Theorem 2.3. There is $A' \subseteq A \setminus \{0\}$ such that

$$(66) \quad |A'| \geq K^{-C_1} N$$

and

$$(67) \quad |A'A' - A'A'| \leq K^{C_1} |A'|.$$

Choose $a_0 \in A'$ and scale

$$\overline{A} = a_0^{-1} A, \quad B = a_0^{-1} A'.$$

Then $1 \in B \subseteq \overline{A}$, and scaling preserves the cardinalities of sumsets and product sets. Moreover

$$|BB - BB| \leq K^{C_1} |B|.$$

By the polynomial-closure theorem, Theorem 3.8, every fixed polynomial value set $P(B, \dots, B)$ has cardinality at most

$$(68) \quad K^{C_P} |B|.$$

Put

$$S = B \cup \{0\}.$$

For a fixed polynomial P , the set $P(S, \dots, S)$ is a union of only finitely many polynomial value sets on B , obtained by deciding which input variables are 0. Thus (68) remains valid, with a changed exponent, when B is replaced by S . Notice also that $0, 1 \in S$ and, under our nontriviality assumption on q ,

$$|S| \geq |B| \geq q^{\delta/2}.$$

Step 2: fill the field with bounded rank and reduce that rank. By Lemma 6.3, there is a surjective linear map

$$S^k \rightarrow F$$

with $k = k(\delta)$. During at most k rank-reduction steps, every current set is obtained from S by a bounded number, depending only on δ , of the operations appearing in Proposition 6.7. Each such set is the value set of a fixed integer polynomial in elements of S . Therefore (68) gives a common bound

$$M \leq K^{C_2(\delta)} |B| \leq K^{C_2(\delta)} N$$

for all current sets. Apply Proposition 6.7. We obtain a subfield $G \leq F$ such that

$$(69) \quad B \subseteq G, \quad |G| \leq K^{C_2(\delta)} N.$$

Step 3: cover the original normalized set by few additive cosets of G . Since $B \subseteq \bar{A}$,

$$|\bar{A} + B| \leq |\bar{A} + \bar{A}| = |A + A| \leq KN.$$

Together with (66), the additive Ruzsa covering lemma gives a set $Y \subseteq F$ with

$$(70) \quad |Y| \leq K^{C_3}$$

and

$$\bar{A} \subseteq Y + (B - B) \subseteq Y + G.$$

Thus $\bar{A}$ lies in polynomially many additive cosets of G .

Step 4: every coset other than G contains only polynomially many points of $\bar{A}$. Fix $y \notin G$ and distinct

$$x, x' \in \bar{A} \cap (G + y).$$

We claim that

$$x(B \setminus \{0\}) \quad \text{and} \quad x'(B \setminus \{0\})$$

are disjoint. Otherwise $xb = x'b'$ for nonzero $b, b' \in B \subseteq G$. If $b = b'$, then $x = x'$, contrary to choice. Hence $b \neq b'$, and

$$x = \frac{(x' - x)b'}{b - b'}.$$

Because $x' - x \in G$ and $b', b - b' \in G$ with $b - b' \neq 0$, the right side lies in G . This contradicts $x \in G + y$ with $y \notin G$.

The disjoint sets above all lie in

$$\overline{AB} \subseteq \overline{A} \overline{A}.$$

Therefore, writing

$$m_y = |\overline{A} \cap (G + y)|,$$

we have

$$m_y(|B| - 1) \leq |\overline{A} \overline{A}| = |AA| \leq KN.$$

If $|B| \leq 2$, then (66) gives $N \leq 2K^{C_1}$, while $N > q^\delta$ gives $q < (2K^{C_1})^{1/\delta}$. This contradicts the nontriviality assumption at the beginning of the proof once C_0 is chosen larger than $C_1 + 1$. Hence this case cannot occur here. Therefore $|B| > 2$, so $|B| - 1 \geq |B|/2 \geq K^{-C_1}N/2$, and

$$(71) \quad m_y \leq K^{C_4}.$$

There are at most $|Y| \leq K^{C_3}$ relevant cosets. Let X_0 be the union of the points of $\overline{A}$ lying in all cosets $G + y$ with $y \notin G$. By (71),

$$|X_0| \leq K^{C_5}.$$

Hence

$$\overline{A} \subseteq G \cup X_0.$$

Multiplying by a_0 gives

$$A \subseteq a_0G \cup a_0X_0.$$

Taking $\xi = a_0$ and $X = a_0X_0$, and combining the cardinality bounds above, proves (65). $\square$

GUIDING IDEA 6.9 (The stopping mechanism). The rank-reduction process is a finite-field analogue of an inverse theorem. A collision says that one coordinate of a linear parametrization was redundant; the polynomial operation $T(T - T) + T(T - T)$ records the algebra needed to remove it. If no collision exists, the parametrization is bijective. Enlarging T additively or multiplicatively would then create more domain points than field points and force a collision after all. Thus injectivity can persist only when neither enlargement is possible, which in a finite field means exactly that T has become a subfield.

6. A direct sum-product corollary away from subfields

COROLLARY 6.10 (Growth under quantitative subfield avoidance). *Fix $0 < \delta, \rho, \kappa < 1$. Then there exists $\varepsilon = \varepsilon(\delta, \rho, \kappa) > 0$ such that the following holds for all sufficiently large finite fields F . Let $A \subseteq F$ satisfy*

$$|F|^\delta < |A| < |F|^{1-\rho}$$

and suppose that for every proper subfield $G < F$ and every $\xi \in F^\times$,

$$(72) \quad |A \cap \xi G| \leq |A|^{1-\kappa}.$$

Then

$$\max(|A + A|, |AA|) \geq |A|^{1+\varepsilon}.$$

PROOF. Suppose instead that both sumset and product set are at most $|A|^{1+\varepsilon}$, and put $K = |A|^\varepsilon$. Apply [Theorem 6.8](#). We obtain

$$A \subseteq \xi G \cup X, \quad |X| \leq K^C = |A|^{C\varepsilon}, \quad |G| \leq |A|^{1+C\varepsilon}.$$

Choose $\varepsilon > 0$ so small that

$$C\varepsilon < \min\left(\frac{\kappa}{2}, \frac{\rho}{2(1-\rho)}\right).$$

If G is proper, then

$$|A| \leq |A \cap \xi G| + |X| \leq |A|^{1-\kappa} + |A|^{C\varepsilon},$$

which is impossible for sufficiently large $|A|$.

If $G = F$, then

$$|F| \leq |A|^{1+C\varepsilon} < |F|^{(1-\rho)(1+C\varepsilon)}.$$

Our choice of ε makes $(1-\rho)(1+C\varepsilon) < 1$, another contradiction. Thus at least one of $A + A$ and AA has cardinality at least $|A|^{1+\varepsilon}$. $\square$

WARNING 6.11 (Two distinct obstructions). Subfield avoidance excludes concentration in proper subfields but says nothing against A itself occupying most of F . This is why [Corollary 6.10](#) also imposes $|A| < |F|^{1-\rho}$. The inverse theorem cleanly separates the two obstructions: proper subfields are algebraic, whereas the full-field alternative is purely an ambient-cardinality obstruction.

EXAMPLE 6.12 (A genuine subfield obstruction). Let $F = \mathbb{F}_{p^2}$ and let $A = \mathbb{F}_p \subset F$. Then

$$A + A = A, \quad AA = A,$$

so neither addition nor multiplication produces any growth at all. This is not a defect of the sum-product theorem: A is literally a proper subfield. More generally, if $\xi \in F^\times$, the set $\xi\mathbb{F}_p$ has the same cardinality as $\mathbb{F}_p$ and is the basic model for the concentration excluded by the hypothesis

$$|A \cap \xi G| \leq |A|^{1-\kappa}.$$

This example explains why the arbitrary-finite-field theorem cannot have the same statement as the prime-field theorem. Over $\mathbb{F}_p$ there is no proper subfield to absorb both additive and multiplicative structure; over $\mathbb{F}_{p^2}$ there is.

7. Dependencies

The proof of [Theorem 6.8](#) uses only results already proved in the first three volumes: the Katz–Tao refinement, polynomial closure, Ruzsa covering, elementary collision counting, and finite-field linear algebra. Bourgain–Katz–Tao state this inverse theorem and give a sketch of the same stopping argument; the purpose of this chapter is precisely to supply the details hidden by that sketch. No classification of approximate fields or finite subfields has been imported.

Part 4

Interfaces and Trace-Shaped Expansion

CHAPTER 7

Interfaces for Growth in SL_2

In rank-one growth arguments, sum-product enters through matrix identities rather than in the bare form $|A + A|$ versus $|AA|$. This chapter develops the translations needed for that use. We prove robust variants under scaling and bounded algebraic changes of variables, explain how the theorem interacts with trace sets, and isolate the exact range bookkeeping that must be checked before invoking sum-product in a Helfgott-type argument.

1. Safe normalizations and unsafe ones

The sum-product theorem is invariant under a common nonzero dilation, and this is the normalization that will be used repeatedly in the matrix argument.

COROLLARY 7.1 (Dilation invariance). *Fix $\delta > 0$, and let c, ε be the constants from [Theorem 5.1](#). If*

$$p^\delta \leq |A| \leq p^{1-\delta}$$

and $\lambda \in F^\times$, then for $B = \lambda A$ one has

$$\max(|B + B|, |BB|) = \max(|A + A|, |AA|) \geq c|B|^{1+\varepsilon}.$$

PROOF. Multiplication by λ and by λ^2 are bijections of F , so

$$|B + B| = |\lambda(A + A)| = |A + A|, \quad |BB| = |\lambda^2 AA| = |AA|,$$

and $|B| = |A|$. □

WARNING 7.2 (Do not invent a two-coefficient sum-product theorem). There is no consequence of [Theorem 5.1](#) asserting that $|\alpha A + \beta A|$ must be large for every pair of nonzero coefficients α, β . Even $|A + A|$ may be small; sum-product says that this cannot happen *simultaneously* with small multiplicative growth in the medium-size regime. When a trace calculation produces coefficients such as ad and bc , a rigorous application must either normalize them by an actual bijective change of variables or derive their control from the approximate-ring calculus. We will not hide that step under a vague “scaled sum-product” slogan.

2. A trace-shaped laboratory

The basic split torus in $\mathrm{SL}_2(F)$ is

$$T = \left\{ t(x) = \begin{pmatrix} x & 0 \\ 0 & x^{-1} \end{pmatrix} : x \in F^\times \right\}.$$

Its trace is

$$\mathrm{tr} \, t(x) = x + x^{-1}.$$

The two coordinate lines

$$L_+ = F \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad L_- = F \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

are the two eigendirections of the diagonal torus (here the statement is made after extending scalars to whatever field contains the torus parameters). We make the word “transverse” precise before using it.

DEFINITION 7.3 (Transverse to the diagonal eigendirections). For $g \in \mathrm{SL}_2(F)$, say that g is *transverse to the two diagonal eigendirections* if

$$gL_+, gL_-$$

are both different from both L_+ and L_- . Equivalently, the eigendirection sets of the diagonal torus and its conjugate by g are disjoint.

LEMMA 7.4 (Transversality and matrix entries). *Let*

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(F).$$

Then g is transverse in the sense of Definition 7.3 if and only if

$$a, b, c, d \neq 0.$$

In particular, under transversality both coefficients ad and bc in the transverse trace formula are nonzero.

PROOF. The images of the coordinate eigendirections are

$$gL_+ = K \begin{pmatrix} a \\ c \end{pmatrix}, \quad gL_- = K \begin{pmatrix} b \\ d \end{pmatrix}.$$

The line $F(a, c)^T$ equals L_+ exactly when $c = 0$, and equals L_- exactly when $a = 0$. Hence gL_+ avoids both coordinate lines exactly when $a, c \neq 0$. Likewise gL_- avoids both coordinate lines exactly when $b, d \neq 0$. Combining the two conditions proves the equivalence. The final statement is immediate. $\square$

If a subset $X \subseteq F^\times$ has controlled multiplicative growth, then products inside the torus correspond to products in X . To escape the torus one conjugates by a matrix

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad ad - bc = 1,$$

and, when the eigendirections are transverse in the sense of [Definition 7.3](#), studies the resulting product. The matrix identity itself is valid for every g , and a direct multiplication gives

$$(73) \quad \text{tr}(t(x)gt(y)g^{-1}) = ad(xy + (xy)^{-1}) - bc\left(\frac{x}{y} + \frac{y}{x}\right).$$

We prove the identity because its algebra is the doorway through which sum-product enters matrix growth.

VERIFICATION OF (73). Since

$$g^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

we have

$$t(x)gt(y) = \begin{pmatrix} xay & xby^{-1} \\ x^{-1}cy & x^{-1}dy^{-1} \end{pmatrix}.$$

Multiplying by g^{-1} , the two diagonal entries are

$$adxy - bcx/y$$

and

$$-bcy/x + ad(xy)^{-1}.$$

Their sum is (73). □

GUIDING IDEA 7.5 (Why transversality creates two operations). Inside one torus, multiplication only sees xy . After conjugating one torus slice transversely, the trace depends simultaneously on

$$xy \quad \text{and} \quad x/y.$$

Equivalently, after a change of variables it mixes multiplicative and additive-type combinations of trace parameters. A set that stayed small under all such products would therefore behave like an approximate ring, and the sum-product theorem rules this out in the medium-size range.

3. Range bookkeeping

The sum-product theorem applies only when

$$p^\delta \leq |A| \leq p^{1-\delta}.$$

In the matrix argument, A will often be a trace set or a parameter set extracted from a torus slice rather than the original generating set. Before invoking [Theorem 5.1](#), the proof must therefore establish two inequalities:

- (i) the extracted parameter set is not too small, usually by a centralizer or rich-torus count;

- (ii) it is not too large, or else the ambient matrix set is already in the large-set regime where a different completion argument applies.

Any later group-theoretic application must keep these alternatives explicit rather than writing “apply sum–product” without checking its range.

PROPOSITION 7.6 (Polynomial losses preserve a power range). *Fix $0 < \delta < 1/2$ and $C > 0$. There exists $\eta = \eta(\delta, C) > 0$ such that if*

$$p^\delta \leq N \leq p^{1-\delta}, \quad K \leq N^\eta,$$

and a set B satisfies

$$K^{-C}N \leq |B| \leq K^C N,$$

then, after possibly replacing δ by $\delta/2$,

$$p^{\delta/2} \leq |B| \leq p^{1-\delta/2}.$$

PROOF. Since $K^C \leq N^{C\eta}$,

$$|B| \geq N^{1-C\eta} \geq p^{\delta(1-C\eta)}.$$

Choose $\eta \leq 1/(2C)$ to obtain the lower bound $p^{\delta/2}$. For the upper bound,

$$|B| \leq N^{1+C\eta} \leq p^{(1-\delta)(1+C\eta)}.$$

Choose η still smaller so that

$$(1-\delta)(1+C\eta) \leq 1-\delta/2.$$

Both choices depend only on δ, C . □

4. Worked numerical ledger

Suppose, schematically, that a matrix set $S \subseteq \mathrm{SL}_2(\mathbb{F}_p)$ has size

$$|S| = p^\alpha, \quad 0 < \alpha < 3,$$

and a rich-torus argument extracts a parameter set A with

$$|A| \geq |S|^{1/3} K^{-C} = p^{\alpha/3} K^{-C}.$$

If $K \leq |S|^\eta$, then

$$|A| \geq p^{\alpha(1/3-C\eta)}.$$

Thus whenever α stays in a compact subinterval of $(0, 3)$ and η is sufficiently small, A lies in a genuine power range $[p^{\delta'}, p^{1-\delta'}]$. This is the quantitative reason the “one-dimensional torus inside a three-dimensional group” heuristic produces a sum–product-sized object.

5. What has been proved, and what is left

This volume has now internalized the arithmetic engine required by the rank-one product theorem:

small sum and product $\Rightarrow$ Katz–Tao approximate ring $\Rightarrow$ polynomial closure
 $\Rightarrow$ prime-field saturation $\Rightarrow$ sum–product growth.

The remaining chapters of this volume extend the arithmetic engine to $\mathbb{F}_{p^2}$ and prove the trace-shaped two-variable expansion statement that a rank-one growth proof consumes. This closes the arithmetic side of the interface. Separate group-theoretic work is still needed to establish the structure of $\mathrm{SL}_2(\mathbb{F}_p)$, the relevant subgroup classification, production of regular semisimple elements, concentration on tori, and the pivot mechanism; none of those facts is being silently assumed in the present arithmetic volume.

6. Source concordance

The core architecture of Chapters 2–5 follows the proof in Bourgain–Katz–Tao [1]: the mixed $AA - AA$ refinement, propagation to polynomial expressions, bounded-rank filling, and rank reduction. In this presentation the combinatorial statements that their paper recalls from earlier work have been replaced by results already proved in Volumes I–II or proved directly here. Citations therefore record provenance rather than logical dependencies.

CHAPTER 8

Quadratic Extensions and the Norm-One Torus

The nonsplit torus of $\mathrm{SL}_2(\mathbb{F}_p)$ diagonalizes over $\mathbb{F}_{p^2}$. Rather than build a second ad hoc sum–product proof for that field, we now use the inverse theorem of the preceding chapter. It says that any failure of expansion must come from a scalar copy of a subfield. In $\mathbb{F}_{p^2}$ there is only one proper subfield, namely $\mathbb{F}_p$; if a set is largely contained in such a scalar line, the prime-field theorem applies after rescaling. For the norm-one torus the situation is even cleaner: every scalar copy of $\mathbb{F}_p$ meets the norm-one group in at most two points. Thus the algebraic obstruction is automatically transverse to the torus.

Throughout this chapter

$$K = \mathbb{F}_{p^2}, \quad |K| = p^2.$$

1. The only proper subfield

LEMMA 8.1 (Subfields of a quadratic finite extension). *Every subfield $G \leq K$ is either $\mathbb{F}_p$ or K .*

PROOF. Every subfield contains the prime field $\mathbb{F}_p$. By the tower law,

$$[K : \mathbb{F}_p] = [K : G][G : \mathbb{F}_p].$$

The left side equals 2, so $[G : \mathbb{F}_p]$ is either 1 or 2. In the first case $G = \mathbb{F}_p$; in the second case $G = K$. □

REMARK 8.2 (What replaces the “characteristic barrier”). The earlier compressed proof detected the cardinality p as the point at which a rank-two parametrization could become bijective. The general inverse theorem identifies the structure behind that numerical event: a set of size p can be the genuine subfield $\mathbb{F}_p$. The subfield formulation is stronger because it distinguishes an algebraic obstruction from an accidental equality of cardinalities.

2. Quadratic-extension sum–product from the inverse theorem

THEOREM 8.3 (Power-range sum–product in $\mathbb{F}_{p^2}$). *Fix $0 < \delta < 1/2$. There exist $c = c(\delta) > 0$ and $\varepsilon = \varepsilon(\delta) > 0$ such that every set $A \subseteq K$ satisfying*

$$(74) \quad p^\delta \leq |A| \leq p^{1-\delta}$$

obeys

$$(75) \quad \max(|A + A|, |AA|) \geq c|A|^{1+\varepsilon}.$$

PROOF. It is enough to prove the result for all sufficiently large p , because the remaining finitely many fields can be absorbed by decreasing c . Suppose for contradiction that

$$|A + A|, |AA| \leq K_0|A|, \quad K_0 = |A|^\varepsilon,$$

where $\varepsilon > 0$ will be chosen in terms of δ .

Apply the general inverse theorem [Theorem 6.8](#) with the ambient field K . Since

$$|A| \geq p^\delta = |K|^{\delta/2},$$

we obtain a subfield $G \leq K$, a scalar $\xi \neq 0$, and a set X such that

$$(76) \quad A \subseteq \xi G \cup X, \quad |X| \leq K_0^{C_1}, \quad |G| \leq K_0^{C_1}|A|,$$

where $C_1 = C_1(\delta)$.

By [Lemma 8.1](#), either $G = K$ or $G = \mathbb{F}_p$.

Case 1: $G = K$. Then (76) gives

$$p^2 \leq |A|^{1+C_1\varepsilon} \leq p^{(1-\delta)(1+C_1\varepsilon)}.$$

Choose ε so small that

$$(1-\delta)(1+C_1\varepsilon) < 2.$$

This is automatic for sufficiently small ε (indeed the left side starts below 1), and yields a contradiction.

Case 2: $G = \mathbb{F}_p$. Let

$$B = A \cap \xi \mathbb{F}_p.$$

From (76),

$$|B| \geq |A| - K_0^{C_1}.$$

Since $|A| \geq p^\delta$ and $K_0^{C_1} = |A|^{C_1\varepsilon}$, choosing $C_1\varepsilon < 1/2$ gives, for large p ,

$$(77) \quad |B| \geq \frac{1}{2}|A|.$$

Write $B = \xi C$ with $C \subseteq \mathbb{F}_p$. Then

$$|C + C| = |B + B| \leq |A + A| \leq K_0|A| \leq 2K_0|B| = 2K_0|C|,$$

and likewise

$$|CC| = |BB| \leq |AA| \leq 2K_0|C|,$$

because multiplication by the nonzero scalars ξ and ξ^2 preserves cardinality.

The size comparison (77) and (74) give, after decreasing the exponent slightly and taking p large,

$$p^{\delta/2} \leq |C| \leq p^{1-\delta/2}.$$

Apply the prime-field sum–product theorem [Theorem 5.1](#), already proved in this volume. There is $\eta = \eta(\delta) > 0$ such that

$$\max(|C + C|, |CC|) \geq c_0|C|^{1+\eta}.$$

On the other hand the preceding upper bounds are at most

$$2|A|^\varepsilon|C| \leq 2^{1+\varepsilon}|C|^{1+\varepsilon}.$$

Taking $\varepsilon < \eta/2$ makes these inequalities incompatible for sufficiently large $|C|$. This contradiction completes the proof. $\square$

GUIDING IDEA 8.4 (The induction on subfields is visible in degree two). If small sum and product force A near a proper subfield, one does not lose: one descends to that smaller field and applies the theorem already proved there. For $\mathbb{F}_{p^2}$ the descent has only one step, from $\mathbb{F}_{p^2}$ to $\mathbb{F}_p$. In larger extensions the same inverse-theorem viewpoint organizes the obstructions by the lattice of subfields.

3. The norm-one group

We first supply the finite-field fact that identifies the norm-one equation.

LEMMA 8.5 (Frobenius and the quadratic norm). *Let $K = \mathbb{F}_{p^2}$. The Frobenius map*

$$\sigma : K \rightarrow K, \quad \sigma(z) = z^p,$$

is the nontrivial $\mathbb{F}_p$ -automorphism of K , its fixed field is exactly $\mathbb{F}_p$, and for every $z \in K$ the field norm, defined by

$$N_{K/\mathbb{F}_p}(z) = \det_{\mathbb{F}_p}(m_z), \quad m_z(u) = zu,$$

is

$$N_{K/\mathbb{F}_p}(z) = z\sigma(z) = z^{p+1}.$$

In particular, for $t \in \mathbb{F}_p$ one has $N_{K/\mathbb{F}_p}(t) = t^2$.

PROOF. In characteristic p , the binomial theorem gives $(x + y)^p = x^p + y^p$, and clearly $(xy)^p = x^p y^p$. Thus σ is a field homomorphism. It is injective, hence bijective because K is finite. Every element of $\mathbb{F}_p$ is fixed by σ .

Conversely, the fixed points of σ are precisely the roots in K of $X^p - X$. This polynomial has degree p , while its p elements in $\mathbb{F}_p$ are already distinct roots. Hence the fixed field is exactly $\mathbb{F}_p$. Moreover, for $z \neq 0$, Lagrange's theorem in the multiplicative group $K^\times$ of order $p^2 - 1$ gives $z^{p^2-1} = 1$, and hence $z^{p^2} = z$; the same is trivial for $z = 0$. Therefore $\sigma^2 = 1$, and since the fixed field is proper, σ is the nontrivial automorphism of the quadratic extension.

We now compute the determinant norm directly. If $z \in \mathbb{F}_p$, multiplication by z is the scalar operator zI on the two-dimensional $\mathbb{F}_p$ -space K , so

$$N_{K/\mathbb{F}_p}(z) = z^2 = z^{p+1}.$$

Suppose $z \notin \mathbb{F}_p$. Then $(1, z)$ is an $\mathbb{F}_p$ -basis of K . Put

$$T = z + z^p, \quad P = z^{p+1}.$$

Because $\sigma^2 = 1$, both T and P are fixed by σ , hence lie in $\mathbb{F}_p$. The factorization

$$(X - z)(X - z^p) = X^2 - TX + P$$

gives $z^2 = Tz - P$. In the basis $(1, z)$ the multiplication operator m_z therefore has matrix

$$\begin{pmatrix} 0 & -P \\ 1 & T \end{pmatrix},$$

whose determinant is $P = z^{p+1}$. This proves the formula in all cases. $\square$

Let

$$U = \{z \in K^\times : z^{p+1} = 1\}.$$

By Lemma 8.5, this is exactly the norm-one subgroup of $K^\times$.

LEMMA 8.6 (Scalar subfield lines meet U in at most two points). *For every $\xi \in K^\times$,*

$$|U \cap \xi\mathbb{F}_p| \leq 2.$$

When $p = 2$ the bound is interpreted literally and remains true.

PROOF. Suppose $z \in U \cap \xi\mathbb{F}_p$. Then $z = \xi t$ for some $t \in \mathbb{F}_p^\times$. Taking norms gives

$$1 = N_{K/\mathbb{F}_p}(z) = N_{K/\mathbb{F}_p}(\xi) t^2,$$

because $t^p = t$. Thus t is a root in $\mathbb{F}_p$ of the quadratic equation

$$N_{K/\mathbb{F}_p}(\xi) t^2 = 1.$$

A nonzero quadratic polynomial over a field has at most two roots. Hence there are at most two possible values of t , and therefore at most two points z . $\square$

COROLLARY 8.7 (Sum-product on the norm-one torus). *Fix $0 < \delta < 1/2$. There exist $c, \varepsilon > 0$, depending only on δ , such that if $A \subseteq U$ and*

$$p^\delta \leq |A| \leq p^{1-\delta},$$

then

$$\max(|A + A|, |AA|) \geq c|A|^{1+\varepsilon}.$$

PROOF. One may simply apply Theorem 8.3. It is nevertheless useful to see directly how the inverse theorem reads on U . If both $A + A$ and AA were at most $K_0|A|$ with $K_0 = |A|^\varepsilon$, then Theorem 6.8 would place all but $K_0^{O_\delta(1)}$ points of A in a scalar copy of either K or $\mathbb{F}_p$. The full-field alternative contradicts the upper power range exactly as in Theorem 8.3. The proper-subfield alternative is much stronger here: by Lemma 8.6, that scalar copy contains at most two points of A . For small enough ε this cannot account for a set of cardinality p^δ . Hence growth is forced. $\square$

4. Why this is the natural nonsplit-torus interface

We now prove the group-theoretic interface that was previously only asserted.

PROPOSITION 8.8 (Quadratic model for a nonsplit centralizer). *Let $h \in \mathrm{SL}_2(\mathbb{F}_p)$ have irreducible characteristic polynomial over $\mathbb{F}_p$, and let*

$$T = C_{\mathrm{SL}_2(\mathbb{F}_p)}(h).$$

Then h is diagonalizable over $K = \mathbb{F}_{p^2}$, and there is a canonical eigenvalue-parameter isomorphism

$$T \cong U = \{x \in K^\times : x^{p+1} = 1\}.$$

More precisely, after choosing one K -eigenline of h , an element $t \in T$ acts on that line by an eigenvalue $x(t)$, on the Frobenius-conjugate line by $x(t)^p$, and the map $t \mapsto x(t)$ is a group isomorphism from T onto U .

PROOF. Write

$$P_h(X) = X^2 - \tau X + 1 \in \mathbb{F}_p[X].$$

Because P_h is irreducible of degree two, a root λ generates the quadratic field $K = \mathbb{F}_p(\lambda)$. Frobenius sends roots to roots. Since $\lambda \notin \mathbb{F}_p$, [Lemma 8.5](#) gives $\lambda^p \neq \lambda$, so the two roots are the distinct elements λ and λ^p . Hence h is diagonalizable over K .

We next identify its full centralizer over $\mathbb{F}_p$. Choose any nonzero vector $v \in \mathbb{F}_p^2$. If v and $h v$ were linearly dependent over $\mathbb{F}_p$, then v would be an $\mathbb{F}_p$ -eigenvector of h , contradicting irreducibility of P_h . Thus $(v, h v)$ is an $\mathbb{F}_p$ -basis. Under the $\mathbb{F}_p$ -linear identification

$$\Phi : K \longrightarrow \mathbb{F}_p^2, \quad \Phi(a + b\lambda) = av + b h v,$$

the operator h corresponds to multiplication by λ . Indeed, from $P_h(h) = 0$ and $P_h(\lambda) = 0$ we have $h^2 = \tau h - I$ and $\lambda^2 = \tau \lambda - 1$, so

$$h\Phi(a + b\lambda) = -bv + (a + b\tau)hv = \Phi(\lambda(a + b\lambda)).$$

Let A be an $\mathbb{F}_p$ -linear endomorphism commuting with h . Since $(v, h v)$ is a basis, there are unique $a, b \in \mathbb{F}_p$ such that

$$Av = av + b h v.$$

Commutation gives

$$A(hv) = h(Av) = a h v + b h^2 v,$$

so on the basis $(v, h v)$ the operator A agrees with $aI + bh$. Hence every endomorphism commuting with h is a polynomial $aI + bh$, and under Φ it corresponds exactly to multiplication by

$$x = a + b\lambda \in K.$$

Thus the invertible centralizer of h in $\mathrm{GL}_2(\mathbb{F}_p)$ is naturally $K^\times$.

By [Lemma 8.5](#), the determinant of multiplication by x is $N_{K/\mathbb{F}_p}(x) = x^{p+1}$. Therefore the determinant-one centralizer is precisely

$$T \cong \{x \in K^\times : N_{K/\mathbb{F}_p}(x) = 1\} = U.$$

Finally choose a λ -eigenvector w of h in K^2 . If $t = aI + bh$ corresponds to $x = a + b\lambda$, then

$$tw = (a + b\lambda)w = xw.$$

Because t has entries in $\mathbb{F}_p$, Frobenius gives $t(w^p) = x^p w^p$. Thus the two eigenvalues are exactly x and x^p , which proves the stated eigenvalue parameter and completes the proof. $\square$

Whenever the rank-one argument reaches the determinant-one centralizer of a regular semisimple element with irreducible characteristic polynomial—the nonsplit case relevant here—the proposition proves directly that this centralizer is the norm-one group. Thus the use of norm-one parameters in this arithmetic interface is no longer an unproved structural assertion; no classification of all maximal tori is being assumed in this volume. Together with [Lemma 8.6](#), it explains geometrically why the prime subfield is not an obstruction there: a scalar $\mathbb{F}_p$ -line meets the norm-one conic in at most two points.

This observation is more informative than a bare “characteristic- p barrier.” The obstruction to sum–product in an extension field is not cardinality p by itself; it is the existence of a subfield of cardinality p . Nonsplit torus parameters avoid its scalar copies automatically.

5. Dependencies

The quadratic-extension theorem now uses no new sum–product input. It is deduced from the general subfield inverse theorem proved in [Chapter 6](#) and the prime-field sum–product theorem proved in [Chapter 5](#). The norm-one corollary adds the explicitly proved Frobenius/norm calculation and the nonsplit-torus eigenvalue interface above. Thus the extension-field interface is logically downstream of the complete finite-field theory rather than a second independent black box.

CHAPTER 9

Trace-Shaped Expansion Functions

The sum-product theorem does not enter SL_2 through the literal pair of sets $A + A$ and AA . The trace of a torus element is $w(x) = x + x^{-1}$, and a transverse torus product produces a two-variable expression built from $w(xy)$ and $w(x/y)$. This chapter proves, from the sum-product theorems already established and the BSG/Ruzsa machinery of Volume II, the two expansion statements that the rank-one proof actually consumes. The square-coset covering step is written out because it is the mechanism that makes the change of variables $r = xy$, $s = x/y$ legitimate.

Let K denote either $\mathbb{F}_p$ or $\mathbb{F}_{p^2}$. For a set $A \subseteq K^\times$ define its signed multiplicative word ball

$$A_{\pm, r} = \{a_1 \cdots a_m : 0 \leq m \leq r, a_i \in A \cup A^{-1}\}.$$

Thus $1 \in A_{\pm, r}$ and $A_{\pm, r}^{-1} = A_{\pm, r}$. We reserve

$$A^{(2)} = \{a^2 : a \in A\}$$

for the pointwise square set. Finally put

$$w(z) = z + z^{-1}.$$

1. Elementary fibers of the trace parameter

LEMMA 9.1 (The map w is at most two-to-one). *For every $t \in K$, the equation $w(z) = t$ has at most two solutions $z \in K^\times$. Consequently*

$$|w(X)| \geq \frac{1}{2}|X|$$

for every finite $X \subseteq K^\times$.

PROOF. The equation $z + z^{-1} = t$ is equivalent to

$$z^2 - tz + 1 = 0.$$

A nonzero polynomial of degree two over a field has at most two roots. Summing the fiber bound gives the cardinality estimate. $\square$

LEMMA 9.2 (The two trace coordinates have bounded fibers). *The map*

$$\Psi : (K^\times)^2 \rightarrow K^2, \quad \Psi(x, y) = (w(xy), w(x/y))$$

has fibers of cardinality at most 16.

PROOF. Fix a target (s, t) . Put $u = xy$ and $v = x/y$. By Lemma 9.1, there are at most two possibilities for u from $w(u) = s$ and at most two possibilities for v from $w(v) = t$. For each fixed pair (u, v) we have

$$x^2 = uv, \quad y^2 = u/v.$$

Each equation has at most two solutions. Hence the total fiber size is at most

$$2 \cdot 2 \cdot 2 \cdot 2 = 16.$$

No assertion that these four choices are independent is needed; 16 is only an upper bound. $\square$

The elementary identity

$$(78) \quad w(x)w(y) = w(xy) + w(x/y)$$

is obtained by expanding both sides. It is the bridge between multiplicative trace products and additive combinatorics.

2. From w -products to sum-product

PROPOSITION 9.3 (Expansion of the product of trace parameters). *Fix $0 < \delta < 1/2$. There exist $c, \eta > 0$, depending only on δ , with the following property. Let $K = \mathbb{F}_p$ or $\mathbb{F}_{p^2}$ and let $A \subseteq K^\times$ satisfy*

$$(79) \quad p^\delta \leq |A| \leq p^{1-\delta}.$$

Then

$$(80) \quad |w(A_{\pm,2}) w(A_{\pm,2})| \geq c|A|^{1+\eta}.$$

PROOF. Write $N = |A|$ and suppose, toward a contradiction, that

$$(81) \quad |w(A_{\pm,2})w(A_{\pm,2})| \leq N^{1+\theta},$$

where $\theta > 0$ will be chosen sufficiently small in terms of δ . Put

$$P = A_{\pm,2}, \quad U = w(P).$$

If every element of U were zero, then Lemma 9.1 would give $|P| \leq 2$, impossible for large N . Choose $u_0 \in U \setminus \{0\}$. Since multiplication by u_0 is injective,

$$|U| = |u_0 U| \leq |UU| \leq N^{1+\theta}.$$

Together with $A \subseteq P$ and Lemma 9.1, this gives

$$(82) \quad \frac{N}{2} \leq |U| \leq N^{1+\theta}.$$

It also gives $|P| \leq 2|U| \leq 2N^{1+\theta}$.

Consider the relation

$$\mathcal{R} = \{(w(xy), w(x/y)) : x, y \in A\} \subseteq U \times U.$$

By Lemma 9.2,

$$(83) \quad |\mathcal{R}| \geq \frac{N^2}{16}.$$

For $(r, s) \in \mathcal{R}$, identity (78) shows that

$$r + s = w(x)w(y) \in UU.$$

Thus the restricted sumset of the relation has size at most $N^{1+\theta}$.

We now convert the restricted relation into the hypotheses of the BSG theorem proved in Volume II; no separate “graph BSG” result is needed. Set $M = |U|$ and let

$$S = \{r + s : (r, s) \in \mathcal{R}\}.$$

From (82) and (83),

$$|\mathcal{R}| \geq M^2 N^{-C\theta}$$

for an absolute C , while $|S| \leq MN^{C\theta}$. For $t \in S$ let

$$\rho(t) = |\{(r, s) \in \mathcal{R} : r + s = t\}|.$$

Every collision counted by $\sum_t \rho(t)^2$ is also an additive-energy collision of the full pair U, U . Hence Cauchy–Schwarz gives

$$E_+(U, U) \geq \sum_t \rho(t)^2 \geq \frac{|\mathcal{R}|^2}{|S|} \geq M^3 N^{-C'\theta}.$$

Apply the energy–approximate-group equivalence proved in Volume II, in additive notation. It yields an $N^{C''\theta}$ -approximate additive group H with

$$N^{-C''\theta} M \leq |H| \leq N^{C''\theta} M$$

and a translate $x + H$ satisfying

$$|U \cap (x + H)| \geq N^{-C''\theta} M.$$

Put

$$U' = U \cap (x + H).$$

Then, using $M \geq N/2$ from (82),

$$(84) \quad |U'| \geq N^{1-C_1\theta}$$

for a suitable absolute C_1 . Moreover

$$U' + U' \subseteq 2x + (H + H),$$

and bounded powers of an approximate group give

$$(85) \quad |U' + U'| \leq |H + H| \leq N^{C_1\theta} M \leq N^{1+C_1\theta}.$$

Thus the complete-sumset conclusion has been derived explicitly from the proved energy inverse theorem.

Because $U' \subseteq U$, the multiplicative contradiction hypothesis also gives

$$(86) \quad |U'U'| \leq |UU| \leq N^{1+\theta}.$$

Choose θ sufficiently small in terms of δ . The lower bound (84) and $N \geq p^\delta$ give

$$|U'| \geq p^{\delta/2}.$$

For the upper bound, $U' \subseteq U$ and $|U| \leq N^{1+\theta}$, so

$$|U'| \leq p^{(1-\delta)(1+\theta)} \leq p^{1-\delta/2}$$

after decreasing θ once more and absorbing bounded p into c .

If $K = \mathbb{F}_p$, apply the prime-field sum–product theorem proved in Chapter 5; if $K = \mathbb{F}_{p^2}$, apply Theorem 8.3. In either case there is an exponent $\varepsilon_0(\delta) > 0$ such that

$$\max(|U' + U'|, |U'U'|) \geq c_0|U'|^{1+\varepsilon_0}.$$

But (84), (85), and (86) bound the left side by $N^{1+C_2\theta}$ while the right side is at least

$$c_0N^{(1-C_1\theta)(1+\varepsilon_0)}.$$

For sufficiently small $\theta = \theta(\delta)$ these exponents are incompatible. This contradiction proves (80), after renaming a fixed fraction of θ as η and adjusting c for bounded N . $\square$

GUIDING IDEA 9.4 (What BSG accomplishes here). The large relation $\mathcal{R}$ says that many pairs of trace coordinates have their sum inside the small set $w(P)w(P)$. BSG converts this *restricted* additive structure into a large set U' on which the *complete* sumset is small. Because U' already lies inside $w(P)$, its product set is also small under the contradiction hypothesis. We have therefore manufactured an honest sum–product counterexample, which the preceding chapters rule out.

3. A coefficient-transfer lemma from common approximate-group control

We next need to replace the equal coefficients in $X + X$ by arbitrary prescribed nonzero coefficients. It is tempting to quote one of the classical abelian Ruzsa inequalities here. We instead derive exactly what is needed from the two-set small-product theorem proved in Volume II, so no additional additive-combinatorial fact is hidden.

LEMMA 9.5 (Large equal-coefficient sum forces every mixed coefficient sum to grow). *There is an absolute constant $c_0 > 0$ with the following property. Let $X \subseteq K$ be finite and nonempty, let $a_1, a_2 \in K^\times$, and suppose*

$$|X + X| \geq |X|L$$

for some $L \geq 1$. Then

$$(87) \quad |a_1X + a_2X| \geq |X|L^{c_0}.$$

PROOF. Put

$$M = \frac{|a_1X + a_2X|}{|X|}.$$

If M is larger than a fixed power of L , there is nothing to prove. We show conversely that

$$(88) \quad L \leq M^C$$

for an absolute constant C ; then $c_0 = 1/C$ works.

Work in the additive group of K and set

$$A = a_1X, \quad B = a_2X.$$

Then $|A| = |B| = |X|$ and

$$|A + B| = M|X| = M\sqrt{|A||B|}.$$

Apply the two-set small-product theorem from Volume II (in additive notation). It produces an M^{C_1} -approximate subgroup H and sets of at most M^{C_1} translates covering A and B , with

$$|H| \leq M^{C_1}|X|.$$

In particular

$$A \subseteq Y + H, \quad |Y| \leq M^{C_1}.$$

Because bounded powers of an approximate group have polynomial size,

$$|H + H| \leq M^{C_2}|H| \leq M^{C_3}|X|.$$

Therefore

$$|A + A| \leq |Y + Y| |H + H| \leq M^{C_4}|X|.$$

Multiplication by a_1 is an additive bijection, so

$$|A + A| = |X + X| \geq L|X|.$$

This proves (88). Taking $c_0 = 1/C_4$ gives (87). □

REMARK 9.6 (What is really being transferred). The lemma does not say that the coefficient change preserves a numerical doubling constant. It says something more structural: a small mixed sum would force the two dilates a_1X and a_2X into the same approximate additive group. Once one dilate is controlled in this way, its self-sum is automatically small. Thus a genuinely large self-sum rules out every small mixed sum by an inverse-theorem argument.

4. Square cosets and the transverse trace expression

LEMMA 9.7 (A fixed transverse parameter has at most two fibers). *Let $a_1, a_2 \in K^\times$. For all but at most four $y \in K^\times$, the map*

$$x \mapsto a_1 w(xy) + a_2 w(x/y)$$

has fibers of cardinality at most two on $K^\times$.

PROOF. Expand the expression as

$$\alpha_y x + \beta_y x^{-1},$$

where

$$\alpha_y = a_1 y + a_2 y^{-1}, \quad \beta_y = a_1 y^{-1} + a_2 y.$$

If $(\alpha_y, \beta_y) \neq (0, 0)$, then the equation

$$\alpha_y x + \beta_y x^{-1} = t$$

is either linear in x^{-1} when $\alpha_y = 0$, or becomes the nonzero quadratic

$$\alpha_y x^2 - tx + \beta_y = 0$$

when $\alpha_y \neq 0$. Thus every fiber has size at most two.

If both coefficients vanish, multiplying the two equations by y gives

$$a_1 y^2 + a_2 = 0, \quad a_1 + a_2 y^2 = 0.$$

Eliminating a_1, a_2 implies $y^4 = 1$. A degree-four polynomial has at most four roots in a field. Hence there are at most four exceptional y . $\square$

PROPOSITION 9.8 (Trace-shaped two-variable expansion). *Fix $0 < \delta < 1/2$ and nonzero $a_1, a_2 \in K$. There exist $c, \varepsilon > 0$, depending only on δ , such that whenever*

$$p^\delta \leq |A| \leq p^{1-\delta}, \quad A \subseteq K^\times,$$

one has

$$(89) \quad \left| \left\{ a_1(xy + x^{-1}y^{-1}) + a_2(x^{-1}y + xy^{-1}) : x, y \in A_{\pm, 20} \right\} \right| \geq c|A|^{1+\varepsilon}.$$

PROOF. Write $N = |A|$. By Proposition 9.3, for some $\eta = \eta(\delta) > 0$,

$$(90) \quad R_0 := |w(A_{\pm, 2})w(A_{\pm, 2})| \geq c_0 N^{1+\eta}.$$

Choose a small parameter

$$\gamma = \eta/4.$$

First alternative: the multiplicative word ball has already grown. Suppose

$$(91) \quad |A_{\pm, 6}| > N^{1+\gamma}.$$

Since $A \subseteq A_{\pm, 20}$ and N is larger than an absolute constant, Lemma 9.7 lets us choose

$$y_0 \in A_{\pm, 20}$$

outside the at most four exceptional values. Restrict the target in (89) to $y = y_0$ and $x \in A_{\pm,6}$. The fiber bound gives at least

$$\frac{1}{2}|A_{\pm,6}| > \frac{1}{2}N^{1+\gamma}$$

distinct values. Thus the proposition follows in this alternative, with any $\varepsilon < \gamma$ after adjusting c .

Second alternative: all relevant multiplicative words remain small. Assume henceforth

$$(92) \quad |A_{\pm,6}| \leq N^{1+\gamma}.$$

Let

$$Q = A^{(2)} = \{a^2 : a \in A\}.$$

By Lemma 9.1 applied to the simpler polynomial $z^2 - t$, the squaring map has fibers of size at most two, so

$$(93) \quad |Q| \geq N/2.$$

Apply the multiplicative Ruzsa covering lemma, proved in Volume II, to

$$X = A_{\pm,4}, \quad Y = Q.$$

It covers $A_{\pm,4}$ by at most

$$(94) \quad J \leq \frac{|A_{\pm,4}Q|}{|Q|} \leq \frac{2|A_{\pm,6}|}{N} \leq 2N^\gamma$$

translates

$$a_j Q Q^{-1}.$$

Because the multiplicative group is abelian,

$$Q Q^{-1} = \{(u/v)^2 : u, v \in A\} \subseteq (A_{\pm,2})^{(2)}.$$

Choose one such cover and, for each distinct value in the product set from (90), choose one representing pair $x, y \in A_{\pm,2}$. Put

$$r = xy, \quad s = x/y.$$

Then $r \in A_{\pm,4}$, so assign the pair to an index j with

$$r \in a_j Q Q^{-1} \subseteq a_j (A_{\pm,2})^{(2)}.$$

Since $s = r y^{-2}$ and $y^{-1} \in A_{\pm,2}$, commutativity gives

$$r, s \in B_j := a_j (A_{\pm,4})^{(2)}.$$

Identity (78) then places the corresponding product value in

$$w(B_j) + w(B_j).$$

By pigeonhole and (94), some index j satisfies

$$(95) \quad |w(B_j) + w(B_j)| \geq \frac{c_0 N^{1+\eta}}{2N^\gamma} \geq c_1 N^{1+3\eta/4}.$$

Fix this j and write $B = B_j$, $X = w(B)$.

We need both upper and lower bounds for $|X|$. Because $A \subseteq A_{\pm,4}$,

$$a_j A^{(2)} \subseteq B,$$

so (93) gives $|B| \geq N/2$ and hence, by Lemma 9.1,

$$(96) \quad |X| \geq N/4.$$

On the other hand, pointwise squaring cannot increase cardinality and multiplication by a_j is a bijection, so

$$(97) \quad |X| \leq |B| \leq |A_{\pm,4}| \leq |A_{\pm,6}| \leq N^{1+\gamma}.$$

Combining (95) and (97),

$$|X + X| \geq c_1 |X| N^{\eta-2\gamma} = c_1 |X| N^{\eta/2}.$$

For N beyond a δ -dependent threshold, absorb c_1 into the exponent and obtain

$$|X + X| \geq |X| N^{\eta/3}.$$

Apply Lemma 9.5:

$$(98) \quad |a_1 X + a_2 X| \geq |X| N^{c_0 \eta/3} \geq \frac{1}{4} N^{1+c_0 \eta/3}.$$

It remains to check that every value counted in (98) is genuinely a value of the two-variable expression in (89). Take $r, s \in B$. By definition there are $u, v \in A_{\pm,4}$ with

$$r = a_j u^2, \quad s = a_j v^2.$$

Set

$$y = u/v \in A_{\pm,8}, \quad x = r/y = a_j uv.$$

The covering representatives satisfy $a_j \in A_{\pm,4}$, so $x \in A_{\pm,12}$. Hence $x, y \in A_{\pm,20}$, and a direct calculation gives

$$xy = r, \quad x/y = s.$$

Therefore

$$a_1 w(r) + a_2 w(s) = a_1 w(xy) + a_2 w(x/y)$$

is one of the target values. Thus the target set contains $a_1 X + a_2 X$, and (98) proves the proposition with, for example,

$$\varepsilon = \min\left(\frac{\gamma}{2}, \frac{c_0 \eta}{4}\right)$$

after adjusting the constant for bounded N . □

5. The exact interface with the transverse torus calculation

For

$$t(x) = \begin{pmatrix} x & 0 \\ 0 & x^{-1} \end{pmatrix}, \quad g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(K),$$

Chapter 7 proved

$$\mathrm{tr}(t(x)gt(y)g^{-1}) = ad w(xy) - bc w(x/y).$$

If g is transverse to the two eigendirections of the diagonal torus in the sense of Definition 7.3, then Lemma 7.4 gives $a, b, c, d \neq 0$, so

$$a_1 = ad, \quad a_2 = -bc$$

are nonzero. Proposition 9.8 therefore applies exactly to the trace set produced by a transverse torus product. No vague “sum–product-like” assertion remains between the finite-field theorem and the matrix calculation.

6. Concordance with Helfgott's expanding-function step

The two propositions above give the arithmetic mechanism of Helfgott's Propositions 3.1 and 3.3 [2] in the power range actually used in the growth proof. The logical ingredients are all internal to the series:

- (i) the full BSG theorem, the energy–approximate-group equivalence, and Ruzsa calculus were proved in Volume II;
- (ii) prime-field sum–product was proved in Chapters 5 and earlier chapters of this volume;
- (iii) the quadratic-extension sum–product theorem was proved in Chapter 8;
- (iv) the square-coset covering and the substitution $r = xy$, $s = x/y$ were proved explicitly above.

The citation records the historical source; it is not a proof dependency.

Part 5

Technical Laboratories

APPENDIX A

The Katz–Tao Telescoping Argument Under a Microscope

The Katz–Tao refinement is the point at which additive and multiplicative structure are fused. Its four-term telescoping identity is short enough to be written on one line and delicate enough that a sign or factor-order error destroys the proof. This appendix reconstructs that step slowly, explains why representation multiplicity rather than mere set cardinality is needed, and records an exponent ledger showing how polynomial losses propagate.

1. Why a representation-rich set is needed

Suppose $C, D \subseteq A$ and every element of $C - D$ has many alternating representations

$$z = (a_1 - a_2) - (a_3 - a_4) + (a_5 - a_6), \quad a_i \in A.$$

If $m \in A^3/A^2$, then every such representation of z gives a representation of zm as a bounded signed sum of elements of A^4/A^2 . The point is not simply that $C - D$ is small. A small set multiplied by a large quotient set could still become huge. The many representations force every output zm to consume many tuples from one common bounded ambient set, so a global counting argument limits the number of possible outputs.

This is the first place where BSG must provide more than a subset of small doubling. The representation-strength formulation proved in Chapter 2 is designed exactly for this multiplication step.

2. The four increments

Use the notation of the proof of [Theorem 2.3](#). Fix

$$u = \frac{c_1 d_1 c_3 d_3}{c_2 d_2}, \quad v = c' d'.$$

Define

$$\begin{aligned} x_1 &= (c_1 - d') \frac{d_1 c_3 d_3}{c_2 d_2}, \\ x_2 &= d' (c' - d_1) \frac{c_3 d_3}{c_2 d_2}, \end{aligned}$$

$$x_3 = d'c'(c_3 - d_2) \frac{d_3}{c_2 d_2},$$

$$x_4 = d'c'd_2(c_2 - d_3) \frac{1}{c_2 d_2}.$$

The identity to prove is

$$(99) \quad u - v = x_1 - x_2 + x_3 - x_4.$$

STEP-BY-STEP VERIFICATION. Expand only one increment at a time. First,

$$x_1 = u - d' \frac{d_1 c_3 d_3}{c_2 d_2}.$$

Next,

$$x_2 = d'c' \frac{c_3 d_3}{c_2 d_2} - d' \frac{d_1 c_3 d_3}{c_2 d_2},$$

so

$$x_1 - x_2 = u - d'c' \frac{c_3 d_3}{c_2 d_2}.$$

Similarly,

$$x_3 = d'c' \frac{c_3 d_3}{c_2 d_2} - d'c' \frac{d_3}{c_2},$$

and hence

$$x_1 - x_2 + x_3 = u - d'c' \frac{d_3}{c_2}.$$

Finally,

$$x_4 = d'c' \frac{c_2 - d_3}{c_2} = d'c' - d'c' \frac{d_3}{c_2}.$$

Subtracting this last expression gives

$$x_1 - x_2 + x_3 - x_4 = u - d'c' = u - v.$$

This is (99). □

WARNING A.1 (The sign of the last increment). The cancellation requires $-x_4$. Replacing it by $+x_4$ leaves two uncanceled terms. The identity is not a cosmetic rearrangement: it is the precise algebraic device that turns a difference of two products into four elements of the controlled set $(C - D)(A^3/A^2)$.

3. Why each increment lies in the same controlled set

Put

$$M = A^3/A^2, \quad W = (C - D)M.$$

For x_1 , the additive factor $c_1 - d'$ lies in $C - D$, and the remaining multiplicative coefficient lies in A^3/A^2 . The same is true for x_2 and x_3 .

For x_4 , after cancellation of d_2 the multiplicative coefficient is

$$\frac{d'c'}{c_2} \in A^2/A.$$

Choose any fixed $a_0 \in A$. Then

$$\frac{d'c'}{c_2} = \frac{d'c'a_0}{c_2a_0} \in A^3/A^2.$$

The difference $c_2 - d_3$ lies in $C - D$. Thus every x_i lies in the same set W .

This common target is what makes the four-tuple count useful: one has at most $|W|^4$ possible increment tuples, regardless of which product difference $u - v$ they represent.

4. The injectivity ledger

After pigeonholing the denominator pair (c_2, d_2) , the variables still free are

$$(c_1, c_3, d_1, d_3).$$

For fixed c', d', c_2, d_2 , the increment tuple determines them in reverse order:

(1) x_4 determines d_3 from

$$x_4 = \frac{d'c'}{c_2}(c_2 - d_3);$$

(2) knowing d_3 , the value x_3 determines c_3 ;

(3) knowing c_3, d_3 , the value x_2 determines d_1 ;

(4) knowing d_1, c_3, d_3 , the value x_1 determines c_1 .

All coefficients divided by here are nonzero because the Katz–Tao refinement was first applied after deleting 0 from A . Thus the map to (x_1, x_2, x_3, x_4) is injective.

5. A symbolic exponent ledger

Let $N = |A|$. Suppose the additive representation theorem provides at least

$$K^{-a}N^5$$

representations of every element of $C - D$, and bounded multiplicative words have size at most

$$K^bN.$$

Then

$$|(C - D)M| \leq K^{a+6b}N.$$

Suppose the multiplicative representation theorem later provides at least

$$K^{-c}N^5$$

representations of each $u \in C'D'$. Pigeonholing two variables leaves at least

$$K^{-c}N^3$$

representations with a fixed denominator pair. The telescoping injection therefore gives

$$|C'D' - C'D'| \leq \frac{|W|^4}{K^{-c}N^3} \leq K^{4a+24b+c}N.$$

Nothing exponential in N has appeared: all deterioration is a fixed power of the original doubling parameter K . This polynomial-loss feature, rather than the displayed exponent itself, is what later survives the choice $K = N^{o(1)}$.

6. What would fail without representation richness

If one knew only

$$|C - D| \leq K^{O(1)}N,$$

then the trivial estimate

$$|(C - D)M| \leq |C - D| |M|$$

would be of order N^2 , too large by a full factor of N . The alternating representation theorem recovers exactly that missing factor. This is why the Katz–Tao step should be viewed as a *representation-transfer theorem*, not merely another small-set estimate.

APPENDIX B

Rank Reduction in a Concrete Extension Field

The general subfield theorem is easiest to understand by watching the rank-reduction algorithm stop on an actual subfield. We work in $\mathbb{F}_9 = \mathbb{F}_3(\omega)$ with $\omega^2 = -1$. The prime subfield gives a bijective rank-two parametrization of the whole field, and precisely because it is closed under both addition and multiplication there is no enlargement that can force a collision. A one-point perturbation immediately destroys this rigidity.

1. The field $\mathbb{F}_9$

Over $\mathbb{F}_3$, the polynomial $t^2 + 1$ has no root: for $t = 0, 1, 2$ its values are 1, 2, 2. Hence it is irreducible, and

$$F = \mathbb{F}_3[\omega]/(\omega^2 + 1)$$

is a field with nine elements. Every element has a unique expression

$$a + b\omega, \quad a, b \in \mathbb{F}_3,$$

and $\omega^2 = -1 = 2$.

The unique proper subfield is

$$G = \mathbb{F}_3 = \{0, 1, 2\}.$$

2. A bijective rank-two map with no collision

Consider

$$\Phi : G^2 \rightarrow F, \quad \Phi(a, b) = a + \omega b.$$

Uniqueness of the basis expansion shows immediately that Φ is bijective. Thus

$$F = G + \omega G$$

is a rank-two linear filling, but there is no collision from which the rank-reduction lemma could eliminate a coordinate.

This is exactly the exceptional stopping case of Chapter 6. Indeed,

$$G + G = G, \quad GG = G.$$

The current set cannot be enlarged either additively or multiplicatively. The finite-semiring criterion therefore recognizes G as a subfield and stops.

GUIDING IDEA B.1 (Cardinality is a symptom, not the cause). Here $|G|^2 = 3^2 = |F|$, so a bijection $G^2 \rightarrow F$ is numerically possible. But a random three-element subset of F would not make the algorithm stop. What matters is that the set is simultaneously closed under $+$ and $\cdot$. The equality of cardinalities merely allows injectivity to expose that algebraic structure.

3. Perturb the subfield by one point

Let

$$T = G \cup \{\omega\}.$$

Then $|T| = 4$. We can compute the first two enlargements completely.

For addition,

$$G + G = G, \quad G + \omega = \{\omega, 1 + \omega, 2 + \omega\}, \quad \omega + \omega = 2\omega.$$

These pieces are disjoint except for the obvious internal repetitions, so

$$|T + T| = 7 > 4.$$

For multiplication,

$$GG = G, \quad G\omega = \{0, \omega, 2\omega\}, \quad \omega^2 = 2 \in G,$$

and therefore

$$TT = \{0, 1, 2, \omega, 2\omega\}, \quad |TT| = 5 > 4.$$

Thus the moment a point transverse to the subfield is added, both algebraic operations detect growth.

If a surjective map $T^r \rightarrow F$ happened to be injective, replacing T by either $T + T$ or TT would produce a larger domain still mapping onto the same nine-point field. A collision would then be unavoidable, and rank reduction would restart. This is the concrete content of the stopping argument in [Proposition 6.7](#).

4. The norm-one group in $\mathbb{F}_9$

The norm to $\mathbb{F}_3$ is

$$N(z) = z^{1+3} = z^4.$$

Since $F^\times$ has order eight, the norm-one group has order four. Directly,

$$U = \{1, 2, \omega, 2\omega\}.$$

Indeed $1^4 = 2^4 = 1$ and $\omega^4 = (\omega^2)^2 = 2^2 = 1$.

The two scalar copies of the prime-subfield line visible from the coordinate axes are

$$G = \{0, 1, 2\}, \quad \omega G = \{0, \omega, 2\omega\}.$$

Each meets U in exactly two points. Other scalar copies also meet it in at most two points, as proved abstractly in [Lemma 8.6](#). This tiny example makes the geometric transversality concrete: the norm-one conic cannot contain a full scalar copy of the prime subfield.

APPENDIX C

Generic Projections and Filling: Worked Finite-Field Examples

The dilation lemma is the elementary geometric engine behind the filling argument. This appendix computes it in small fields, then rewrites the collision average as a projection-energy identity. The examples show the two regimes used in the proofs: nearly multiplicative growth when the Cartesian product is sparse, and near-saturation when the product size exceeds the ambient field.

1. A complete table in $\mathbb{F}_7$

Take

$$X = \{0, 1, 2\}, \quad Y = \{0, 1\} \subseteq \mathbb{F}_7.$$

For each nonzero slope ξ , the projection set $X + \xi Y$ is:

ξ	$X + \xi Y$	$ X + \xi Y $
1	$\{0, 1, 2, 3\}$	4
2	$\{0, 1, 2, 3, 4\}$	5
3	$\{0, 1, 2, 3, 4, 5\}$	6
4	$\{0, 1, 2, 4, 5, 6\}$	6
5	$\{0, 1, 2, 5, 6\}$	5
6	$\{0, 1, 2, 6\}$	4

Here $|X||Y| = 6$. The best slopes lose no cardinality at all: every point of the Cartesian product has a different projection. The worst slopes have two collisions, but the average is still large.

2. Projection energy

For general finite $X, Y \subseteq F$, define

$$r_\xi(t) = |\{(x, y) \in X \times Y : x + \xi y = t\}|.$$

Then

$$\sum_t r_\xi(t) = |X||Y|$$

and

$$E_\xi := \sum_t r_\xi(t)^2$$

is the number of ordered pairs of Cartesian-product points lying on the same fiber of the projection $(x, y) \mapsto x + \xi y$.

Cauchy–Schwarz gives

$$|X + \xi Y| \geq \frac{|X|^2 |Y|^2}{E_\xi}.$$

Averaging E_ξ over $\xi \neq 0$ counts each off-diagonal quadruple with $y \neq y'$ for at most one slope. Thus

$$\frac{1}{q-1} \sum_{\xi \neq 0} E_\xi \leq |X||Y| + \frac{|X|^2 |Y|^2}{q-1}.$$

This is the calculation behind [Lemma 6.1](#). It is useful to read it probabilistically: a random slope has the unavoidable diagonal collision contribution $|X||Y|$, plus only the ambient-field fraction $1/(q-1)$ of all possible off-diagonal pairs.

3. Sparse and dense regimes

Put $M = |X||Y|$. The lower bound

$$|X + \xi Y| \geq \frac{M(q-1)}{q-1+M}$$

has two transparent asymptotics.

If $M \ll q$, then

$$\frac{M(q-1)}{q-1+M} = M \left(1 + O\left(\frac{M}{q}\right) \right)^{-1},$$

so a good projection is nearly injective on $X \times Y$.

If $M \gg q$, then

$$\frac{M(q-1)}{q-1+M} = (q-1) \left(1 + O\left(\frac{q}{M}\right) \right)^{-1},$$

so a good projection occupies almost the entire field.

The bounded-rank filling lemma alternates these regimes. At first, each new scalar copy of S multiplies the cardinality of the current sumset. Once positive density is reached, one more generic projection drives the set above half density; a final addition then fills the field exactly.

4. Why coefficients are allowed to change

The filling identity

$$F = \xi_1 S + \cdots + \xi_k S$$

should not be confused with a statement about repeated ordinary addition kS . A proper additive subspace of F_{p^m} may satisfy $kS = S$ for every k and therefore never fill F . Multiplication by a generic ξ rotates the subspace relative to itself. The freedom to choose coefficients is what makes the procedure sensitive to field multiplication and ultimately turns the surviving obstruction into a subfield rather than an arbitrary additive subspace.

5. A small-field fallback is not a loophole

In the proof of [Lemma 6.3](#), fields with q bounded in terms of the exponent η are treated by a direct basis construction. This is logically harmless because the theorem asks for a rank bound depending only on η . It is also constructive: if $a, b \in S$ are distinct and $d = b - a$, then repeating the coefficient e_i/d exactly $p - 1$ times allows the e_i -coordinate to take every value in $\mathbb{F}_p$. Doing this for an $\mathbb{F}_p$ -basis $e_1, \dots, e_m$ gives an explicit surjection with at most $m(p - 1) \leq q$ coordinates.

Bibliography

1. Jean Bourgain, Nets Katz, and Terence Tao, *A sum-product estimate in finite fields, and applications*, Geometric and Functional Analysis **14** (2004), no. 1, 27–57.
2. Harald A. Helfgott, *Growth and generation in $\mathrm{SL}_2(F/pF)$* , Annals of Mathematics **167** (2008), no. 2, 601–623.

The Bourgain–Gamburd Method
Volume IV: Helfgott Growth in $\mathrm{SL}_2(\mathbb{F}_p)$

Preface

Volume IV begins the second arc of the Bourgain–Gamburd series. The first three volumes established the analytic and additive-combinatorial machinery: finite-group harmonic analysis and quasirandomness, noncommutative Balog–Szemerédi–Gowers theory and L^2 -flattening, and finite-field sum–product. We now ask the first genuinely group-specific question:

why must a generating subset of $\mathrm{SL}_2(\mathbb{F}_p)$ grow?

In this volume the complete growth theorem is proved over the prime field $\mathbb{F}_p$; the structural calculations are often stated over $\mathbb{F}_q$ because they cost nothing extra. The general type- A finite-field theorem, including $\mathrm{SL}_2(\mathbb{F}_q)$, is proved at the end of Volume VI.

The answer is Helfgott’s rank-one product-growth mechanism [1]. We develop it from the structure of SL_2 , its projective action, tori and centralizers, trace identities, regular semisimple production, bounded escape from eigenline degeneracies, rich torus slices, transverse trace maps, and a final large-set quasirandom completion. The proof is reconstructed rather than quoted. In particular, the subgroup dichotomy needed in prime characteristic is proved directly from the action on the projective line instead of being hidden behind the name “Dickson classification.”

There is one arithmetic subtlety that deserves emphasis. The power-range sum–product theorem of Volume III does not by itself suffice for Helfgott’s rich-torus slice, because that slice can be larger than a constant yet smaller than every fixed power of p . Chapter 4 therefore proves the missing constant-threshold small-set estimate by a Stepanov–Konyagin argument, following the architecture of Konyagin [2], and then transfers it to the trace-shaped expressions required in the nonsplit as well as split torus cases.

The proof rule is the same as in the first arc: a non-basic theorem is used only if it was proved in an earlier volume or is proved here. Historical citations identify origins; they do not replace proof dependencies. **Series position.** This volume uses Volumes II–III and elementary structure of SL_2 . Its main outputs are Helfgott product growth and the rank-one algebraic growth engine.

Quantifier and asymptotic conventions. Unless a statement explicitly says otherwise, every constant occurring in a quantitative theorem is

chosen in the order in which it is quantified in that theorem. Thus a phrase such as “for all sufficiently small δ ” means: there is a displayed or locally quantified threshold $\delta_0 > 0$, depending only on the parameters listed before that phrase, such that the assertion holds for every $0 < \delta < \delta_0$. The phrases “sufficiently large N ” and “sufficiently large p ” have the analogous meaning.

If $X \ll_\theta Y$ or $X = O_\theta(Y)$, the implicit constant depends only on θ and on the fixed ambient data of the current volume; it is independent of the variables quantified after θ . The words *absolute*, *uniform*, and *bounded* never introduce an unlisted dependence. Whenever an inequality between exponents or scales is essential to the argument, that inequality is written explicitly in the proof; these conventions are not a substitute for such a calculation.

Contents

Preface	iii
Part 1. Rank-One Structure	1
Chapter 1. The Rank-One Group SL_2	2
1. Why begin with explicit matrices?	2
2. Basic structure and cardinality	3
3. The trace controls the characteristic polynomial	3
4. Centralizers of regular semisimple elements	4
5. Conjugacy classes and their sizes	6
6. Trace and conjugacy	7
7. Unipotent elements and Borels	7
8. The projective line appears already	8
9. The normalizer of a split torus	8
10. A first dimension-producing map	9
11. A classification table for individual elements	10
12. What this chapter will be used for	11
Chapter 2. The Projective Line, Bruhat Cells, and Proper Subgroups	12
1. The action on $\mathbb{P}^1$	12
2. Bruhat decomposition in rank one	13
3. Intersections of Borels	13
4. Counting split tori	14
5. Nonsplit tori as conjugate pairs over $\mathbb{F}_{q^2}$	14
6. Fixed points and element types	16
7. Geometric families of proper subgroups	16
8. A self-contained subgroup dichotomy over a prime field	16
9. Consequences that we will actually use	20
10. Generation and escape from geometric traps	20
11. A worked example: two unipotent directions generate	21
Chapter 3. Trace Identities and the Geometry of Conjugation	22
1. The fundamental trace identity	22
2. The Fricke commutator identity	23

3. When does trace distinguish elements inside a torus?	24
4. Simultaneously diagonalizable elements	24
5. Conjugation maps and centralizer fibers	25
6. A multiplication map with controlled fibers	26
7. Trace ± 2 is exceptional but small	26
8. The trace map on triples	27
9. A worked trace-coordinate calculation	28
10. From identities to growth questions	29
Part 2. Arithmetic Bridge	30
Chapter 4. The Small-Set Sum–Product Supplement	31
1. The Stepanov difference estimate for a multiplicative subgroup	31
2. Konyagin’s small-set lower bound	36
3. Uniform prime-field sum–product below the top scale	40
4. Uniform trace-shaped expansion for rank-one torus parameters	42
Part 3. Tori, Trace Growth, and Escape	45
Chapter 5. From Torus Slices to Trace Growth	46
1. Eigenvalue coordinates for a torus	46
2. Moving a torus into transverse position	47
3. The decisive trace computation	48
4. Deriving the transverse trace formula by hand	48
5. A trace-growth lemma from a torus slice	50
6. From many traces back to many matrices	51
7. Trace lower bounds from a generating set	52
8. The transfer chain	52
Chapter 6. Centralizers, Escape, and the Production of Torus Slices	54
1. A conjugacy-class counting principle	54
2. Producing many regular semisimple elements	56
3. Escape from a union of linear subspaces	58
4. Escaping eigenline degeneracies	59
5. A large simultaneously diagonalizable subset	60
6. The dimension heuristic behind the centralizer count	61
7. A precise cubic-growth lemma	61
8. Trace size controls group size	63
9. A fully proved cube-root trace bound	63
Chapter 7. Small Tripling, Pivots, and Propagation of Torus Structure	66
1. From small tripling to an approximate group	66
2. Rich tori	68

3. The pivot map	69
4. Distinct tori overlap only in exceptional elements	70
5. What the local pivot argument proves	71
Part 4. Helfgott Growth	73
Chapter 8. Helfgott Growth: the Small and Medium Regimes	74
1. Statement of the theorem	74
2. Step 0: the legitimate symmetrization	75
3. Step 1: produce many traces	75
4. Step 2: extract a torus slice of the correct dimension	76
5. Step 3: obtain a transverse conjugate	76
6. Step 4: sum-product forces trace growth	77
7. Step 5: convert trace growth into group growth	77
8. A dependency ledger for the rank-one proof	78
9. Tracking the growth exponent	79
10. What is genuinely rank one?	79
11. Comparison with the original word-exponent proof	79
Chapter 9. Large Sets, Completion, and the Full Rank-One Product Theorem	81
1. Quasirandomness of $\mathrm{SL}_2(\mathbb{F}_p)$	81
2. The quasirandom product lemma	82
3. Historical note on the large-set completion	84
4. The absolute product theorem	84
5. Polylogarithmic diameter	85
6. Product growth versus expansion	86
Bibliography	87

Part 1

Rank-One Structure

CHAPTER 1

The Rank-One Group SL_2

The product theorem for $\mathrm{SL}_2(\mathbb{F}_p)$ is not a theorem about four matrix entries independently. Its geometry is controlled by a small collection of rank-one objects: traces, one-dimensional tori, one-dimensional unipotent groups, and two-dimensional Borels. This chapter develops those objects explicitly. The goal is not classification for its own sake; every calculation will later become a counting statement for product sets.

GUIDING IDEA 1.1 (Rank one is a laboratory, not merely an example). The group SL_2 is three-dimensional, its maximal tori are one-dimensional, and the generic conjugacy classes are two-dimensional. The exponents $1/3$ and $2/3$ that appear later are therefore already encoded in the geometry developed here. The point of doing the matrix algebra explicitly is to make those dimension ratios visible before algebraic geometry repackages them.

There are three element types to keep separate: regular semisimple elements, whose centralizers are tori; nontrivial unipotent elements, whose repeated eigenvalue creates a larger degeneracy; and the central elements $\pm I$. Most growth arguments first manufacture many regular semisimple elements because their centralizers have the smallest possible dimension and therefore give the strongest fiber bounds.

1. Why begin with explicit matrices?

The group

$$G = \mathrm{SL}_2(k) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : ad - bc = 1 \right\}$$

should be regarded simultaneously in three ways:

- (i) as a concrete matrix group;
- (ii) as a three-dimensional algebraic variety carrying multiplication and inversion;
- (iii) when $k = \mathbb{F}_q$, as a finite group of cardinality comparable with q^3 .

The whole growth argument moves back and forth between these descriptions. A maximal torus has algebraic dimension one and therefore about q rational points;

a Borel subgroup has dimension two and about q^2 points; the ambient group has dimension three and about q^3 points. Later, inequalities of the form

$$|A \cap V| \lesssim |A|^{\dim V / \dim G}$$

will make these dimension heuristics quantitative. In rank one, however, we can see the mechanism directly in coordinates.

CONVENTION 1.2. Unless a finite-field hypothesis is stated, k denotes an arbitrary field. When characteristic 2 creates a genuine distinction we say so explicitly. The Bourgain–Gamburd application concerns odd primes, so later chapters usually assume $p \geq 5$.

2. Basic structure and cardinality

PROPOSITION 1.3 (Order of $\mathrm{SL}_2(\mathbb{F}_q)$). *For every finite field $\mathbb{F}_q$,*

$$|\mathrm{SL}_2(\mathbb{F}_q)| = q(q^2 - 1) = q(q - 1)(q + 1).$$

Consequently

$$|\mathrm{PSL}_2(\mathbb{F}_q)| = \frac{q(q^2 - 1)}{\gcd(2, q - 1)}.$$

PROOF. Choose the first column $v \in \mathbb{F}_q^2 \setminus \{0\}$; there are $q^2 - 1$ choices. For fixed v , the map

$$\mathbb{F}_q^2 \longrightarrow \mathbb{F}_q, \quad w \longmapsto \det(v, w)$$

is a nonzero linear functional, hence every fiber has cardinality q . Exactly q vectors w satisfy $\det(v, w) = 1$. Thus $|\mathrm{SL}_2(\mathbb{F}_q)| = (q^2 - 1)q$.

The center consists of scalar matrices λI with $\lambda^2 = 1$. It therefore has cardinality $\gcd(2, q - 1)$, and quotienting by it gives the second formula. $\square$

REMARK 1.4. The proof already exhibits the scale q^3 : one chooses a nonzero vector, contributing two parameters, and then one affine parameter in the determinant-one fiber. This elementary count is the finite-field shadow of $\dim \mathrm{SL}_2 = 3$.

3. The trace controls the characteristic polynomial

For $g \in \mathrm{SL}_2(k)$, Cayley–Hamilton gives

$$(1) \quad g^2 - (\mathrm{tr} g)g + I = 0.$$

Thus the characteristic polynomial is

$$\chi_g(X) = X^2 - (\mathrm{tr} g)X + 1.$$

Its discriminant is

$$\Delta(g) = (\mathrm{tr} g)^2 - 4.$$

When $\mathrm{char}(k) \neq 2$, this separates the familiar cases.

DEFINITION 1.5. Let $g \in \mathrm{SL}_2(k)$.

- (i) g is *regular semisimple* if $\Delta(g) \neq 0$.
- (ii) A regular semisimple g is *split* over k if $\Delta(g)$ is a square in k .
- (iii) It is *nonsplit* if $\Delta(g)$ is not a square in k .
- (iv) If $\Delta(g) = 0$, then $\text{tr}(g) = \pm 2$ when $\text{char}(k) \neq 2$; aside from the scalar matrices $\pm I$, these are the nontrivial unipotent or negative-unipotent elements.

LEMMA 1.6 (Trace recurrence). *Let $g \in \text{SL}_2(k)$ and write $t = \text{tr}(g)$. Then for every $n \geq 1$ there is a polynomial $P_n \in \mathbb{Z}[T]$ such that*

$$\text{tr}(g^n) = P_n(t),$$

where $P_0(T) = 2$, $P_1(T) = T$, and

$$P_{n+1}(T) = TP_n(T) - P_{n-1}(T).$$

PROOF. Multiply (1) by g^{n-1} and take traces:

$$\text{tr}(g^{n+1}) = t \text{tr}(g^n) - \text{tr}(g^{n-1}).$$

The stated recurrence follows by induction. □

GUIDING IDEA 1.7. The trace is useful because it is both *conjugacy invariant* and *algebraically productive*. Conjugating a large set may create many group elements without changing traces; multiplying commuting diagonal elements turns trace expressions into sums of monomials. Helfgott's argument exploits the tension between those two facts.

4. Centralizers of regular semisimple elements

A decisive rank-one fact is that a noncentral semisimple element has a one-dimensional centralizer. We first prove this without algebraic-group terminology.

PROPOSITION 1.8 (Centralizer as a quadratic algebra). *Let $g \in \text{SL}_2(k)$ be non-scalar. Then every matrix in $\text{Mat}_2(k)$ commuting with g belongs to the two-dimensional k -algebra*

$$k[g] = \{xI + yg : x, y \in k\}.$$

Consequently

$$C_{\text{SL}_2(k)}(g) = k[g]^\times \cap \text{SL}_2(k).$$

PROOF. Because g is non-scalar, its minimal polynomial has degree 2. Hence there exists $v \in k^2$ such that v and gv are linearly independent. Indeed, if every nonzero v were an eigenvector, then comparing $v, w, v + w$ for independent v, w would force a common eigenvalue and g would be scalar.

Let $T \in \text{Mat}_2(k)$ commute with g . Express

$$Tv = xv + ygv$$

for unique $x, y \in k$. Then

$$T(gv) = g(Tv) = xgv + yg^2v.$$

Thus T and $xI + yg$ agree on the basis $\{v, gv\}$, so they are equal. The statement about the group centralizer follows by imposing invertibility and determinant one. $\square$

COROLLARY 1.9. *The centralizer in $\mathrm{SL}_2(k)$ of every non-scalar element is abelian.*

PROOF. It is contained in the commutative algebra $k[g]$. $\square$

For regular semisimple elements one can identify the centralizer explicitly.

PROPOSITION 1.10 (Split centralizer). *Suppose $g \in \mathrm{SL}_2(k)$ has two distinct eigenvalues in k . Then g is conjugate in $\mathrm{GL}_2(k)$ to $\mathrm{diag}(\lambda, \lambda^{-1})$ with $\lambda \neq \pm 1$, and*

$$C_{\mathrm{SL}_2(k)}(g) \cong \{\mathrm{diag}(t, t^{-1}) : t \in k^\times\}.$$

In particular, if $k = \mathbb{F}_q$, then $|C_G(g)| = q - 1$.

PROOF. Distinct eigenvalues give an eigenbasis. In that basis, any commuting matrix must preserve the two eigenspaces and is therefore diagonal. The determinant-one condition leaves matrices $\mathrm{diag}(t, t^{-1})$. Over $\mathbb{F}_q$ there are $q - 1$ choices for t . $\square$

PROPOSITION 1.11 (Nonsplit centralizer). *Let $k = \mathbb{F}_q$ and let $g \in \mathrm{SL}_2(k)$ have irreducible characteristic polynomial. Then*

$$C_{\mathrm{SL}_2(k)}(g) \cong \{z \in \mathbb{F}_{q^2}^\times : N_{\mathbb{F}_{q^2}/\mathbb{F}_q}(z) = 1\},$$

and hence $|C_G(g)| = q + 1$.

PROOF. The algebra $k[g]$ is a field with q^2 elements, hence isomorphic to $\mathbb{F}_{q^2}$. View $k[g]$ acting on the two-dimensional k -vector space k^2 by multiplication. For $z \in k[g]^\times$, the determinant of this k -linear multiplication map is the field norm $N_{\mathbb{F}_{q^2}/\mathbb{F}_q}(z)$. Therefore the determinant-one elements are exactly the norm-one subgroup.

The norm map $\mathbb{F}_{q^2}^\times \rightarrow \mathbb{F}_q^\times$ is $z \mapsto z^{q+1}$. Since $\mathbb{F}_{q^2}^\times$ is cyclic of order $q^2 - 1$, its kernel has order $(q^2 - 1)/(q - 1) = q + 1$. $\square$

DEFINITION 1.12. A subgroup of $\mathrm{SL}_2(k)$ conjugate over an algebraic closure to

$$T = \{\mathrm{diag}(t, t^{-1}) : t \in \bar{k}^\times\}$$

is a *maximal torus*. Over a finite field, it is split if it is already diagonalizable over k , and nonsplit otherwise.

Thus every regular semisimple element belongs to a unique maximal torus: its centralizer.

PROPOSITION 1.13 (Uniqueness of the torus). *Let $g \in \mathrm{SL}_2(k)$ be regular semisimple. Then its centralizer is a maximal torus. If g belongs to a maximal torus T , then $T = C_G(g)$. In particular, two maximal tori containing the same regular semisimple element are equal.*

PROOF. Over $\bar{k}$, diagonalize g . Its centralizer is then exactly the diagonal determinant-one subgroup by [Proposition 1.10](#), so it is a one-dimensional torus. Any torus containing g is abelian and hence lies in $C_G(g)$. Since both are maximal one-dimensional tori, equality follows. $\square$

5. Conjugacy classes and their sizes

Orbit–stabilizer for conjugation gives

$$|g^G| = [G : C_G(g)].$$

Combining this with [Propositions 1.3](#), [1.10](#) and [1.11](#) immediately produces the correct dimension heuristic.

COROLLARY 1.14 (Regular semisimple conjugacy classes). *Let $G = \mathrm{SL}_2(\mathbb{F}_q)$.*

(i) *If g is split regular semisimple, then*

$$|g^G| = q(q + 1).$$

(ii) *If g is nonsplit regular semisimple, then*

$$|g^G| = q(q - 1).$$

Thus every regular semisimple conjugacy class has cardinality $\asymp q^2$.

PROOF. For every finite group, orbit–stabilizer for conjugation gives

$$|g^G| = [G : C_G(g)] = \frac{|G|}{|C_G(g)|}.$$

By [Proposition 1.3](#), $|G| = q(q^2 - 1)$. If g is split regular semisimple, [Proposition 1.10](#) gives $|C_G(g)| = q - 1$, hence

$$|g^G| = q(q + 1).$$

If g is nonsplit regular semisimple, [Proposition 1.11](#) gives $|C_G(g)| = q + 1$, hence

$$|g^G| = q(q - 1).$$

$\square$

REMARK 1.15. This is “dimension two” inside a “dimension three” group. Later, if A behaves like a three-dimensional object of size N , one should expect roughly $N^{2/3}$ points on a typical conjugacy-class-type surface and $N^{1/3}$ points on a torus. The Larsen–Pink inequality makes exactly this philosophy rigorous for approximate groups.

6. Trace and conjugacy

For regular semisimple elements of SL_2 over a finite field, the trace almost determines the conjugacy class; the slight qualification comes from rational versus geometric conjugacy. For our counting arguments, what matters first is simpler.

LEMMA 1.16. *If $g, h \in \mathrm{SL}_2(k)$ have the same trace, then they have the same characteristic polynomial. If that polynomial has distinct roots and both matrices split over k , then g and h are conjugate in $\mathrm{GL}_2(k)$.*

PROOF. The first assertion follows from $\chi_g(X) = X^2 - (\mathrm{tr} g)X + 1$. In the split regular case, both matrices diagonalize with the same unordered pair of eigenvalues $\{\lambda, \lambda^{-1}\}$, so they are $\mathrm{GL}_2(k)$ -conjugate. $\square$

WARNING 1.17. One should not silently replace $\mathrm{GL}_2(k)$ -conjugacy by $\mathrm{SL}_2(k)$ -conjugacy in arbitrary fields. The distinction is harmless in several finite-field counting arguments but matters if one is classifying rational conjugacy classes exactly. Whenever only class-size or centralizer bounds are needed, we will state those directly and avoid an unnecessary conjugacy-class assertion.

7. Unipotent elements and Borels

The upper triangular subgroup

$$B = \left\{ \begin{pmatrix} t & x \\ 0 & t^{-1} \end{pmatrix} : t \in k^\times, x \in k \right\}$$

is the standard Borel subgroup. It contains the split torus

$$T = \{\mathrm{diag}(t, t^{-1}) : t \in k^\times\}$$

and the unipotent subgroup

$$U = \left\{ u(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in k \right\}.$$

The relation

$$(2) \quad \mathrm{diag}(t, t^{-1})u(x)\mathrm{diag}(t^{-1}, t) = u(t^2x)$$

will recur throughout the subject. It is the smallest example of a torus acting on a root subgroup.

PROPOSITION 1.18 (Structure of the standard Borel). *The map*

$$U \rtimes T \longrightarrow B, \quad \left(u(y), \mathrm{diag}(t, t^{-1}) \right) \longmapsto u(y) \mathrm{diag}(t, t^{-1})$$

is an isomorphism, where T acts on $U \cong (k, +)$ by $x \mapsto t^2x$. If $k = \mathbb{F}_q$, then

$$|U| = q, \quad |T| = q - 1, \quad |B| = q(q - 1).$$

PROOF. Every upper triangular determinant-one matrix admits a unique factorization

$$\begin{pmatrix} t & x \\ 0 & t^{-1} \end{pmatrix} = \begin{pmatrix} 1 & xt \\ 0 & 1 \end{pmatrix} \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix}.$$

The conjugation rule is (2); the cardinalities are immediate. $\square$

8. The projective line appears already

The action of $\mathrm{SL}_2(k)$ on nonzero vectors descends to an action on the projective line

$$\mathbb{P}^1(k) = k \cup \{\infty\}.$$

For

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

the action is the fractional-linear transformation

$$x \mapsto \frac{ax + b}{cx + d}$$

with the usual conventions at poles and at ∞ .

PROPOSITION 1.19. *The stabilizer in $\mathrm{SL}_2(k)$ of the point $\infty = [1 : 0]$ is the upper triangular Borel subgroup B . Consequently, for $k = \mathbb{F}_q$,*

$$[G : B] = q + 1 = |\mathbb{P}^1(\mathbb{F}_q)|.$$

PROOF. The line $k(1, 0)$ is preserved precisely by upper triangular matrices. The index formula follows from Propositions 1.3 and 1.18. $\square$

This observation later explains why proper subgroups of “Borel type” are large—of order q^2 —yet still too thin to capture a genuinely spreading random walk or a generating approximate subgroup.

9. The normalizer of a split torus

Let

$$w = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Then

$$w \operatorname{diag}(t, t^{-1}) w^{-1} = \operatorname{diag}(t^{-1}, t).$$

Thus w normalizes T but does not centralize it unless $t = \pm 1$.

PROPOSITION 1.20. *Assume $|k| > 3$. The normalizer of the standard split torus in $\mathrm{SL}_2(k)$ is*

$$N_G(T) = T \sqcup wT.$$

In particular, for $k = \mathbb{F}_q$ with $q > 3$,

$$|N_G(T)| = 2(q - 1).$$

PROOF. The torus T has two common eigenspaces in k^2 , namely the coordinate lines. An element normalizing T must permute these two lines, because it sends the set of T -weight spaces to itself. If it fixes both lines it is diagonal and belongs to T . If it swaps them, multiplying by w^{-1} fixes both and hence lies in T . The two cosets are disjoint because w is not diagonal. $\square$

GUIDING IDEA 1.21. The quotient $N_G(T)/T$ has order two: this is the Weyl group in rank one. In later product-growth arguments, a set concentrated in a torus can fail to grow, but a generating set cannot remain forever inside the torus normalizer. An element outside the normalizer moves the torus to a genuinely different torus, and multiplication between the two directions creates dimension.

10. A first dimension-producing map

Fix a split torus T and an element $g \notin N_G(T)$. Consider

$$\Phi : T \times T \longrightarrow G, \quad (t_1, t_2) \longmapsto t_1 g t_2.$$

At the set-theoretic level this map is almost injective.

LEMMA 1.22 (Double-coset fiber bound). *Let $T \leq G$ be any subgroup and $g \in G$. If*

$$t_1 g t_2 = t'_1 g t'_2$$

with $t_i, t'_i \in T$, then

$$t_1^{-1} t'_1 \in T \cap g T g^{-1}.$$

Every fiber of $T \times T \rightarrow T g T$ therefore has cardinality $|T \cap g T g^{-1}|$, and

$$|T g T| = \frac{|T|^2}{|T \cap g T g^{-1}|}.$$

PROOF. Rearranging the displayed equality gives

$$t_1^{-1} t'_1 = g(t_2 t'^{-1}_2) g^{-1} \in T \cap g T g^{-1}.$$

Conversely, if $h \in T \cap g T g^{-1}$ and $h = g s g^{-1}$ with $s \in T$, then

$$(t_1 h^{-1}) g (s t_2) = t_1 g t_2,$$

so the intersection acts freely and transitively on each fiber. The formula follows. $\square$

For maximal tori in SL_2 , distinct tori intersect only in the center at the regular semisimple level; in particular their intersection is uniformly bounded in the situations we shall use. Thus $T g T$ has about q^2 elements. This is our first concrete example of multiplying two one-dimensional algebraic directions to create a two-dimensional set.

11. A classification table for individual elements

Assume for this discussion that q is odd. For

$$g \in \mathrm{SL}_2(\mathbb{F}_q), \quad \chi_g(X) = X^2 - tX + 1, \quad t = \mathrm{tr}(g),$$

the discriminant

$$\Delta(g) = t^2 - 4$$

organizes the generic possibilities.

Condition on $\Delta(g)$	Element type	Generic centralizer size
nonzero square in $\mathbb{F}_q$	split regular semisimple	$q - 1$
non-square in $\mathbb{F}_q$	nonsplit regular semisimple	$q + 1$
0 and $g \neq \pm I$	nontrivial unipotent up to sign	$2q$ in SL_2
$g = \pm I$	central	$ \mathrm{SL}_2(\mathbb{F}_q) $

The first two rows are the ones used most heavily in growth arguments. A regular semisimple element has two distinct eigenlines over $\mathbb{F}_q$, and its centralizer is the unique maximal torus containing it. In the split case the eigenlines are already defined over $\mathbb{F}_q$; in the nonsplit case Frobenius exchanges them and the torus becomes diagonal only over $\mathbb{F}_{q^2}$.

EXAMPLE 1.23 (A split element). Choose $\lambda \in \mathbb{F}_q^\times$ with $\lambda \neq \pm 1$ and set

$$g = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}.$$

Then

$$\Delta(g) = (\lambda - \lambda^{-1})^2$$

is a nonzero square. The centralizer consists of the diagonal determinant-one matrices and has $q - 1$ elements.

EXAMPLE 1.24 (A nonsplit element). Choose $t \in \mathbb{F}_q$ such that $t^2 - 4$ is a nonsquare and consider

$$g = \begin{pmatrix} 0 & -1 \\ 1 & t \end{pmatrix}.$$

Its characteristic polynomial is $X^2 - tX + 1$, irreducible over $\mathbb{F}_q$. Over $\mathbb{F}_{q^2}$ it has eigenvalues $\lambda, \lambda^q = \lambda^{-1}$, so $\lambda^{q+1} = 1$. The centralizer is the norm-one torus and has $q + 1$ rational points.

GUIDING IDEA 1.25 (Centralizers as dimension signals). Since $|\mathrm{SL}_2(\mathbb{F}_q)| \asymp q^3$ while a regular torus has $\asymp q$ points, a regular conjugacy class has $\asymp q^2$ points. These are the finite-field shadows of the algebraic dimensions

$$\dim \mathrm{SL}_2 = 3, \quad \dim T = 1, \quad \dim(\text{regular class}) = 2.$$

The later 1/3 and 2/3 exponents are already visible in this table.

12. What this chapter will be used for

The following dictionary should be kept in mind.

Object	Size over $\mathbb{F}_q$	Growth role
$G = \mathrm{SL}_2(\mathbb{F}_q)$	$\asymp q^3$	ambient space
maximal torus T	$q \pm 1$	possible one-dimensional concentration
Borel B	$q(q - 1)$	largest geometric proper subgroup
regular conjugacy class	$\asymp q^2$	conjugation creates two dimensions
$N_G(T)$	$\asymp q$	obstruction to moving a torus

CHECKPOINT 1.26. At this stage the reader should be able to prove directly, without invoking algebraic groups, that a regular semisimple element has an abelian centralizer of order $q - 1$ or $q + 1$, that the corresponding conjugacy class has order $\asymp q^2$, and that an element outside a torus normalizer creates a double coset of order $\asymp q^2$. These are the rank-one prototypes of the dimension estimates used later.

CHAPTER 2

The Projective Line, Bruhat Cells, and Proper Subgroups

The action on $\mathbb{P}^1$ turns matrix calculations into geometry. Borels become point stabilizers, the Bruhat decomposition becomes the statement that a pair of projective points either coincide or do not, and torus normalizers become stabilizers of unordered pairs. These observations explain the main geometric proper subgroups of $\mathrm{SL}_2(\mathbb{F}_p)$. We prove the structural facts needed later and prove, from projective fixed-point geometry, the precise subgroup dichotomy needed later. No classification theorem is left as a black box.

1. The action on $\mathbb{P}^1$

Write points of $\mathbb{P}^1(k)$ as one-dimensional subspaces of k^2 . The natural linear action of $\mathrm{SL}_2(k)$ descends to a projective action because scalar multiplication of a vector does not change the line it spans. In the affine chart $[x : 1] \leftrightarrow x$,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot x = \frac{ax + b}{cx + d}.$$

The center acts trivially, so this action factors through $\mathrm{PSL}_2(k)$.

PROPOSITION 2.1 (Transitivity and two-point stabilizers). *Let k be a field.*

- (i) $\mathrm{SL}_2(k)$ acts transitively on $\mathbb{P}^1(k)$.
- (ii) The stabilizer of ∞ is the standard Borel B .
- (iii) The subgroup fixing both 0 and ∞ is the standard split torus T .
- (iv) The subgroup preserving the unordered pair $\{0, \infty\}$ is $N_G(T)$.

PROOF. Given a line ℓ , choose a nonzero vector $v \in \ell$ and complete v to a determinant-one basis (v, w) ; the corresponding matrix sends ∞ to ℓ . The remaining assertions follow by inspecting which matrices preserve the coordinate lines individually or as an unordered pair. The last assertion agrees with [Proposition 1.20](#). $\square$

When $k = \mathbb{F}_q$, transitivity and [Proposition 1.19](#) give the familiar identification

$$G/B \cong \mathbb{P}^1(\mathbb{F}_q), \quad |G/B| = q + 1.$$

GUIDING IDEA 2.2 (Subgroups become incidence geometry on $\mathbb{P}^1$). A matrix in $\mathrm{SL}_2(\mathbb{F}_p)$ acts on the projective line by a fractional linear transformation. Under this

action, a Borel subgroup is exactly a point stabilizer, a split torus fixes an ordered pair of distinct points, and its normalizer stabilizes the corresponding unordered pair. This turns subgroup questions into geometry of one or two projective points.

The classification theorem later says that these geometric families, together with a short list of exceptional cases, account for every proper subgroup. For the Bourgain–Gamburd method we rarely need the classification in its most refined form. We need to know what kind of algebraic relation a random word would have to satisfy in order to remain trapped in a proper subgroup, and how many such relations can persist at logarithmic word length.

2. Bruhat decomposition in rank one

The group $G = \mathrm{SL}_2(k)$ has exactly two B -double cosets.

THEOREM 2.3 (Bruhat decomposition for SL_2). *Let*

$$w = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Then

$$G = B \sqcup BwB.$$

More explicitly, every matrix with lower-left entry $c \neq 0$ can be written

$$(3) \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & a/c \\ 0 & 1 \end{pmatrix} \begin{pmatrix} c^{-1} & 0 \\ 0 & c \end{pmatrix} w \begin{pmatrix} 1 & d/c \\ 0 & 1 \end{pmatrix}.$$

PROOF. If $c = 0$, the matrix is upper triangular and lies in B . If $c \neq 0$, direct multiplication on the right-hand side of (3) gives

$$\begin{pmatrix} a & (ad - 1)/c \\ c & d \end{pmatrix}.$$

Since $ad - bc = 1$, the upper-right entry is b . Thus every matrix lies in one of the two displayed double cosets. They are disjoint because matrices in B stabilize ∞ , whereas an element of BwB sends ∞ away from ∞ . $\square$

GUIDING IDEA 2.4. Bruhat decomposition is the first place where “rank one” becomes visible. Relative to a Borel, there are only two positions: the same projective point, or a different one. In higher rank, the Weyl group contains many relative positions, and the bookkeeping in product theorems becomes correspondingly more elaborate.

3. Intersections of Borels

For a projective point x , let B_x denote its stabilizer. Distinct Borels intersect in a torus.

PROPOSITION 2.5. *Let $x \neq y$ in $\mathbb{P}^1(k)$. Then $B_x \cap B_y$ is a maximal split torus over k . Conversely every split maximal torus arises as the pointwise stabilizer of two distinct points of $\mathbb{P}^1(k)$.*

PROOF. Choose $g \in G$ sending (x, y) to $(\infty, 0)$. Conjugation by g identifies $B_x \cap B_y$ with the subgroup fixing both coordinate lines, namely T . Conversely a split torus is diagonalizable over k , and its two eigenspaces give the two fixed points. $\square$

For q odd, the split regular elements are precisely the noncentral elements lying in such intersections. This gives a useful geometric interpretation: a split torus records an ordered pair of eigendirections, up to swapping.

4. Counting split tori

PROPOSITION 2.6 (Number of split maximal tori). *Let $G = \mathrm{SL}_2(\mathbb{F}_q)$ with $q > 3$. The number of split maximal tori is*

$$[G : N_G(T)] = \frac{q(q+1)}{2}.$$

PROOF. All split maximal tori are conjugate: choose their two eigendirections and send them to the coordinate lines. Their stabilizer under conjugation is the normalizer. Hence the number is the stated index. Using [Propositions 1.3](#) and [1.20](#),

$$[G : N_G(T)] = \frac{q(q^2 - 1)}{2(q - 1)} = \frac{q(q + 1)}{2}.$$

$\square$

The same number can be seen projectively: a split torus corresponds to an unordered pair of distinct points of $\mathbb{P}^1(\mathbb{F}_q)$, and there are

$$\binom{q+1}{2} = \frac{q(q+1)}{2}$$

such pairs.

5. Nonsplit tori as conjugate pairs over $\mathbb{F}_{q^2}$

A nonsplit torus has no eigenline over $\mathbb{F}_q$, but over $\mathbb{F}_{q^2}$ it has two eigenlines exchanged by Frobenius. Thus it corresponds to a Frobenius-conjugate unordered pair in

$$\mathbb{P}^1(\mathbb{F}_{q^2}) \setminus \mathbb{P}^1(\mathbb{F}_q).$$

The normalizer again has index two over the torus.

PROPOSITION 2.7. *Let T_{ns} be a nonsplit maximal torus of $\mathrm{SL}_2(\mathbb{F}_q)$, q odd. Then*

$$|T_{ns}| = q + 1, \quad |N_G(T_{ns})| = 2(q + 1).$$

Consequently the number of nonsplit maximal tori is

$$\frac{q(q-1)}{2}.$$

PROOF. The first formula is [Proposition 1.11](#). We prove the normalizer statement explicitly in the quadratic-field model used there. Let

$$K = \mathbb{F}_{q^2},$$

viewed as a two-dimensional vector space over $\mathbb{F}_q$, and identify

$$T_{ns} = \{m_z : z \in K^\times, N_{K/\mathbb{F}_q}(z) = 1\},$$

where m_z denotes multiplication by z . The Frobenius map

$$\sigma : K \rightarrow K, \quad x \mapsto x^q,$$

is $\mathbb{F}_q$ -linear, satisfies $\sigma^2 = 1$, and is not the identity. Since q is odd, its eigenvalues are 1 and -1 , so

$$\det_{\mathbb{F}_q}(\sigma) = -1.$$

The norm map $K^\times \rightarrow \mathbb{F}_q^\times$ is onto: its kernel has $q+1$ elements by [Proposition 1.11](#), while $|K^\times| = (q-1)(q+1)$, so its image has $q-1$ elements. Choose $c \in K^\times$ with $N(c) = -1$ and put

$$w = m_c \sigma.$$

Then

$$\det(w) = N(c) \det(\sigma) = 1,$$

so $w \in \mathrm{SL}_2(\mathbb{F}_q)$, and for every norm-one z ,

$$wm_z w^{-1} = m_{z^q} = m_{z^{-1}}.$$

Thus w normalizes T_{ns} and induces inversion.

It remains to show that these are all normalizer elements. Over K , the torus has two common eigendirections, exchanged by Frobenius. Any $g \in N_G(T_{ns})$ permutes this unordered pair. If it fixes both eigendirections, it centralizes every element of the diagonalized torus and hence lies in $C_G(T_{ns}) = T_{ns}$. If it swaps them, then $w^{-1}g$ fixes both and therefore lies in T_{ns} . Consequently

$$N_G(T_{ns}) = T_{ns} \sqcup wT_{ns}$$

and has order $2(q+1)$. Finally orbit-stabilizer gives

$$[G : N_G(T_{ns})] = \frac{q(q^2-1)}{2(q+1)} = \frac{q(q-1)}{2}.$$

□

6. Fixed points and element types

The projective action gives a useful classification without mentioning eigenvalues explicitly.

PROPOSITION 2.8. *Assume q odd and let $1 \neq \bar{g} \in \mathrm{PSL}_2(\mathbb{F}_q)$.*

- (i) *A split regular semisimple element fixes exactly two points of $\mathbb{P}^1(\mathbb{F}_q)$.*
- (ii) *A nontrivial unipotent element fixes exactly one point.*
- (iii) *A nonsplit regular semisimple element fixes no point of $\mathbb{P}^1(\mathbb{F}_q)$.*

PROOF. A projective fixed point is exactly an eigenline over $\mathbb{F}_q$. The three cases correspond respectively to two distinct eigenlines, one repeated eigenline, and no rational eigenline. $\square$

This simple observation is very useful when identifying possible stabilizers of small configurations.

7. Geometric families of proper subgroups

Three natural proper subgroups are already visible:

- (i) point stabilizers B_x , of order $q(q-1)$ in $\mathrm{SL}_2(\mathbb{F}_q)$;
- (ii) normalizers of split tori, of order $2(q-1)$;
- (iii) normalizers of nonsplit tori, of order $2(q+1)$.

Modulo the center these are, respectively, affine/triangular groups and dihedral groups. They are the positive-dimensional proper algebraic subgroups that matter most in rank one.

8. A self-contained subgroup dichotomy over a prime field

For the applications in this series we do not need the full classification of all subgroups of $\mathrm{PSL}_2(q)$. What we need is the following prime-field consequence: a proper subgroup is either geometrically trapped in a Borel or a torus normalizer, or it has bounded order. We now prove this directly. The argument is worth knowing because it exposes the geometry used later in random-walk escape.

LEMMA 2.9 (Normalizer of a Sylow p -subgroup). *Let $p \geq 5$, let $\bar{G} = \mathrm{PSL}_2(\mathbb{F}_p)$, and let $U < \bar{G}$ be the image of*

$$\left\{ \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} : t \in \mathbb{F}_p \right\}.$$

Then $|U| = p$, U is a Sylow p -subgroup, and $N_{\bar{G}}(U)$ is the image of the upper triangular Borel.

PROOF. The order formula from [Proposition 1.3](#) gives

$$|\bar{G}| = \frac{p(p^2 - 1)}{2},$$

so the p -part of $|\bar{G}|$ is exactly p . Hence U is Sylow. Every nonidentity element of U is unipotent and fixes the unique point ∞ of $\mathbb{P}^1(\mathbb{F}_p)$. Consequently an element normalizing U must preserve the common fixed point of U , hence lies in the point stabilizer B_∞ . Conversely B_∞ normalizes U by direct multiplication. Thus $N_{\bar{G}}(U) = B_\infty$. $\square$

PROPOSITION 2.10 (Elementary Gauss decomposition). *If $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(k)$ with $c \neq 0$, then*

$$g = u\left(\frac{a-1}{c}\right) \ell(c) u\left(\frac{d-1}{c}\right).$$

PROOF. Multiplying the right-hand side gives

$$\begin{pmatrix} a & (ad-1)/c \\ c & d \end{pmatrix},$$

and $(ad-1)/c = b$ because $ad-bc = 1$. $\square$

LEMMA 2.11 (Two Sylow p -subgroups generate). *If $P, Q < \mathrm{PSL}_2(\mathbb{F}_p)$ are distinct Sylow p -subgroups, then*

$$\langle P, Q \rangle = \mathrm{PSL}_2(\mathbb{F}_p).$$

PROOF. Each Sylow p -subgroup fixes exactly one projective point, and distinct Sylow subgroups have distinct fixed points: a nontrivial p -subgroup fixing a point is contained in the corresponding unipotent radical, which already has order p . Conjugate so that $P = U_\infty$. Let $x \neq \infty$ be the fixed point of Q . An element of the Borel B_∞ acts transitively on the affine line $\mathbb{F}_p = \mathbb{P}^1(\mathbb{F}_p) \setminus \{\infty\}$, so after a further conjugation by an element of B_∞ we may assume that $Q = U_0$, the lower unipotent subgroup. The elementary Gauss decomposition [Proposition 2.10](#) shows that U_∞ and U_0 generate every matrix in the big Bruhat cell; together with U_∞ they therefore generate $\mathrm{SL}_2(\mathbb{F}_p)$, and hence their images generate $\mathrm{PSL}_2(\mathbb{F}_p)$. $\square$

REMARK 2.12 (The one elementary group-theoretic input). For completeness, Cauchy's theorem in the form just used has a two-line counting proof. If a prime p divides $|H|$, let

$$X = \{(h_1, \dots, h_p) \in H^p : h_1 \cdots h_p = 1\}.$$

There are $|H|^{p-1}$ such tuples, hence $p \mid |X|$. Cyclic rotation acts on X with orbits of size 1 or p . Therefore the number of fixed tuples is divisible by p . The fixed tuples are $(h, \dots, h)$ with $h^p = 1$; since the identity is one of them, there must be a nonidentity such h .

PROPOSITION 2.13 (p -torsion forces a Borel). *Let $H < \mathrm{PSL}_2(\mathbb{F}_p)$ be proper. If $p \mid |H|$, then H is contained in a Borel subgroup.*

PROOF. By Cauchy's theorem, proved in Remark 2.12 by counting cyclic shifts of p -tuples, H contains an element of order p , and hence a Sylow p -subgroup P of $\mathrm{PSL}_2(\mathbb{F}_p)$. If P were not the unique Sylow p -subgroup of H , then H would contain a distinct conjugate Q and Lemma 2.11 would force $H = \mathrm{PSL}_2(\mathbb{F}_p)$. Thus $P \triangleleft H$, so

$$H \leq N_{\bar{G}}(P),$$

and Lemma 2.9 identifies this normalizer with a Borel. $\square$

We now treat subgroups of order prime to p . The key observation replaces the usual appeal to the classification of finite subgroups of PGL_2 by a fixed point count.

LEMMA 2.14 (Cyclic point stabilizers in the tame case). *Let k be algebraically closed of characteristic p and let $H < \mathrm{PGL}_2(k)$ be finite with $p \nmid |H|$. Then the stabilizer H_x of every $x \in \mathbb{P}^1(k)$ is cyclic.*

PROOF. Conjugate x to ∞ . Every element fixing ∞ acts on the affine line as

$$z \mapsto az + b.$$

The multiplier map $H_\infty \rightarrow k^\times$, $(a, b) \mapsto a$, is a homomorphism. Its kernel consists of translations $z \mapsto z + b$. Every nontrivial finite translation has order p , so the kernel is trivial because $p \nmid |H|$. Thus H_∞ embeds in the finite subgroup of $k^\times$ consisting of its multipliers. A finite subgroup of the multiplicative group of a field is cyclic: if its exponent is m , all of its elements are roots of $X^m - 1$, so there are at most m elements, while the exponent of a finite abelian group is attained by an element after combining its Sylow components. Hence H_∞ is cyclic. $\square$

LEMMA 2.15 (The fixed-point identity). *Let $H < \mathrm{PGL}_2(k)$ be a finite p' -subgroup as above. Let $O_1, \dots, O_r$ be the H -orbits in $\mathbb{P}^1(k)$ whose points have nontrivial stabilizer, and let e_i be the common order of a stabilizer on O_i . Then*

$$(4) \quad \sum_{i=1}^r \left(1 - \frac{1}{e_i}\right) = 2 - \frac{2}{|H|}.$$

In particular $r \leq 3$.

PROOF. Every nonidentity element of $\mathrm{PGL}_2(k)$ whose order is prime to p is semisimple with two distinct eigenlines over k , and therefore fixes exactly two points of $\mathbb{P}^1(k)$. Count pairs

$$(h, x), \quad 1 \neq h \in H, \quad hx = x,$$

in two ways. Counting first by h gives $2(|H| - 1)$. Counting by x gives

$$\sum_x (|H_x| - 1) = \sum_{i=1}^r \frac{|H|}{e_i} (e_i - 1) = |H| \sum_{i=1}^r \left(1 - \frac{1}{e_i}\right).$$

Dividing by $|H|$ proves (4). Since every $e_i \geq 2$, each summand is at least $1/2$, whereas the right-hand side is strictly smaller than 2. Hence $r \leq 3$. $\square$

PROPOSITION 2.16 (Tame subgroup dichotomy). *Let $H < \mathrm{PGL}_2(k)$ be finite with $p \nmid |H|$. Then one of the following holds:*

- (i) H is cyclic and fixes a pair of points;
- (ii) H contains a cyclic subgroup of index two and preserves an unordered pair of points;
- (iii) $|H| \leq 60$.

PROOF. If H is trivial, alternative (i) holds after choosing any two distinct points. Assume henceforth that $|H| > 1$. If there are no exceptional orbits, the left side of (4) is zero, impossible for $|H| > 1$. If $r = 1$, then $1 - 1/e_1 = 2 - 2/|H|$, which forces $1/e_1 = 2/|H| - 1 \leq 0$ for $|H| \geq 2$, again impossible. Thus $r = 2$ or 3.

If $r = 2$, (4) becomes

$$\frac{1}{e_1} + \frac{1}{e_2} = \frac{2}{|H|}.$$

Since $e_i \leq |H|$, each term on the left is at least $1/|H|$, so equality forces $e_1 = e_2 = |H|$. Hence every element fixes the two corresponding points, and H is cyclic by Lemma 2.14.

Assume $r = 3$ and order $2 \leq e_1 \leq e_2 \leq e_3$. Then

$$(5) \quad \frac{1}{e_1} + \frac{1}{e_2} + \frac{1}{e_3} = 1 + \frac{2}{|H|} > 1.$$

Thus $e_1 = 2$. If $e_2 = 2$, the same equation gives $1/e_3 = 2/|H|$, so $|H| = 2e_3$. The stabilizer C of a point in the third exceptional orbit has order e_3 and therefore index two; it is normal. Its two fixed points are consequently permuted by H , so H preserves their unordered pair. This is alternative (ii).

If $e_2 \geq 3$, then (5) forces $e_2 = 3$ and $e_3 < 6$. Therefore $e_3 \in \{3, 4, 5\}$. Substituting into (5) yields respectively

$$|H| = 12, \quad 24, \quad 60.$$

This is alternative (iii). □

THEOREM 2.17 (Prime-field subgroup dichotomy). *Let $p \geq 5$ be prime and let $H < \mathrm{PSL}_2(\mathbb{F}_p)$ be proper. Then one of the following holds:*

- (i) H is contained in a Borel subgroup;
- (ii) H is contained in the normalizer of a split or nonsplit maximal torus;
- (iii) $|H| \leq 60$.

PROOF. If $p \mid |H|$, apply Proposition 2.13. Assume $p \nmid |H|$ and view H inside $\mathrm{PGL}_2(\overline{\mathbb{F}_p})$. Apply Proposition 2.16. In alternative (i), the two common fixed points are either both rational over $\mathbb{F}_p$ or are exchanged by Frobenius, because the unordered pair is H -invariant and H is defined over $\mathbb{F}_p$. The corresponding cyclic group lies in a split or nonsplit torus. In alternative (ii), the same Frobenius argument shows that the preserved unordered pair defines a split or nonsplit torus, and H lies in its normalizer. Alternative (iii) is already the bounded case. □

REMARK 2.18 (Relation with Dickson’s theorem). The classical Dickson classification identifies the bounded groups in the last alternative more precisely as subgroups of A_4 , S_4 , or A_5 (with small-prime coincidences). Nothing in the growth or expansion arguments of this series uses that finer identification. The theorem above is therefore both sufficient and, unlike a citation to the full classification, completely proved here.

9. Consequences that we will actually use

COROLLARY 2.19 (Large proper subgroups are essentially Borel). *There is an absolute constant C such that for every prime $p \geq 5$ and every proper subgroup $H < \mathrm{PSL}_2(\mathbb{F}_p)$, either*

$$|H| \leq Cp,$$

or H is contained in a Borel subgroup, in which case $|H| \ll p^2$.

PROOF. A torus normalizer has order $p - 1$ or $p + 1$ in $\mathrm{PSL}_2(\mathbb{F}_p)$ up to the harmless parity convention, hence $O(p)$; the bounded alternative has order at most 60. A Borel has order $p(p - 1)/2$. Apply Theorem 2.17. $\square$

COROLLARY 2.20 (No large proper subgroup of dimension three). *For every proper $H < \mathrm{SL}_2(\mathbb{F}_p)$,*

$$|H| \ll p^2.$$

PROOF. Pass to the image in $\mathrm{PSL}_2(\mathbb{F}_p)$ and use Corollary 2.19. Lifting through the center changes the order by at most two. $\square$

10. Generation and escape from geometric traps

Suppose $A \subseteq G$ generates G . Then A cannot be contained in $N_G(T)$ for any torus T , and it cannot be contained in any Borel. This apparently trivial observation becomes powerful when combined with bounded word length.

LEMMA 2.21 (Bounded escape from a subgroup). *Let G be generated by a finite set A containing 1, and let $H < G$ have index $m < \infty$. Then there exists $a \in A^{m-1}$ with $a \notin H$.*

PROOF. If $A^{m-1} \subseteq H$, then every word of length at most $m - 1$ lies in H . Consider a shortest word $a_1 \cdots a_r \notin H$. The cosets

$$H, Ha_1, Ha_1a_2, \dots, Ha_1 \cdots a_r$$

are all distinct: equality of two earlier cosets would let us delete an intervening segment and obtain a shorter word outside H . Thus $r + 1 \leq m$, so $r \leq m - 1$. $\square$

This lemma is useless when $[G : H]$ grows with p and one needs an absolute word length. Later we will prove stronger “escape from subvarieties” statements: bounded complexity, rather than finite index, controls the escape time.

11. A worked example: two unipotent directions generate

Let

$$u(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \quad \ell(y) = \begin{pmatrix} 1 & 0 \\ y & 1 \end{pmatrix}.$$

Then

$$u(x)\ell(y) = \begin{pmatrix} 1 + xy & x \\ y & 1 \end{pmatrix}.$$

The two root directions already produce a two-dimensional family, and a third multiplication fills the group.

GUIDING IDEA 2.22. This formula is a miniature product theorem. One-dimensional root groups in transverse directions multiply to create an open cell of the three-dimensional group. The difficulty in Helfgott's theorem is that an arbitrary finite set A need not contain long pieces of root groups. The trace and torus arguments are devices for extracting transverse algebraic directions from the much weaker hypothesis that A generates and does not grow quickly.

CHECKPOINT 2.23. You should now be able to translate between: point stabilizers and Borels; unordered pairs and split torus normalizers; eigenline counts and semisimple/unipotent type; and the two Bruhat cells. The subgroup dichotomy needed later has now been proved internally; the finer names of the bounded exceptional groups play no role in the argument.

CHAPTER 3

Trace Identities and the Geometry of Conjugation

Helfgott's original proof is driven by trace. This chapter develops the elementary identities behind that proof and, more importantly, explains which maps have small fibers. The recurring principle is that a failure of growth must force many algebraic coincidences; in rank one those coincidences usually mean that elements commute or lie in a common torus.

GUIDING IDEA 3.1 (Trace is the coordinate that survives conjugation). Matrix entries change violently under conjugation, but trace does not. That makes trace the natural scalar coordinate through which a noncommutative growth problem can communicate with additive combinatorics in the field. The central mechanism of the next several chapters is therefore

many matrices $\longrightarrow$ many controlled conjugates
 $\longrightarrow$ many trace values
 $\longrightarrow$ sum-product growth in $\mathbb{F}_p$.

The price for passing to trace is that it forgets information. The fiber calculations in this chapter quantify how much is forgotten. A trace map is useful only when one can prove that its fibers are not so large that all cardinality growth disappears.

1. The fundamental trace identity

For every $y \in \mathrm{SL}_2(k)$, Cayley–Hamilton gives

$$y + y^{-1} = (\mathrm{tr} \, y)I.$$

Multiplying by x and taking traces yields the basic identity

$$(6) \quad \mathrm{tr}(xy) + \mathrm{tr}(xy^{-1}) = \mathrm{tr}(x) \mathrm{tr}(y).$$

LEMMA 3.2 (Trace identities). *For $x, y \in \mathrm{SL}_2(k)$:*

$$(7) \quad \mathrm{tr}(x) = \mathrm{tr}(x^{-1}),$$

$$(8) \quad \mathrm{tr}(xy) = \mathrm{tr}(yx),$$

$$(9) \quad \mathrm{tr}(xy) + \mathrm{tr}(xy^{-1}) = \mathrm{tr}(x) \mathrm{tr}(y).$$

PROOF. The first follows because x and x^{-1} have eigenvalues interchanged, or directly from $x^{-1} = (\operatorname{tr} x)I - x$. The second is the usual cyclicity of matrix trace. The third is (6). $\square$

The identity (9) is the bridge between multiplicative structure in G and additive/multiplicative structure in the field.

2. The Fricke commutator identity

Let

$$a = \operatorname{tr} x, \quad b = \operatorname{tr} y, \quad c = \operatorname{tr}(xy).$$

Then the trace of the commutator is a polynomial in a, b, c .

PROPOSITION 3.3 (Fricke identity). *For $x, y \in \operatorname{SL}_2(k)$,*

$$(10) \quad \operatorname{tr}([x, y]) = a^2 + b^2 + c^2 - abc - 2, \quad [x, y] = xyx^{-1}y^{-1}.$$

PROOF. From $x^{-1} = aI - x$ and $y^{-1} = bI - y$,

$$[x, y] = xy(aI - x)(bI - y).$$

Expand and take traces:

$$\operatorname{tr}([x, y]) = ab \operatorname{tr}(xy) - a \operatorname{tr}(xy^2) - b \operatorname{tr}(x^2y) + \operatorname{tr}(xyxy).$$

Use Cayley–Hamilton in the forms $y^2 = by - I$ and $x^2 = ax - I$:

$$\operatorname{tr}(xy^2) = bc - a, \quad \operatorname{tr}(x^2y) = ac - b.$$

Finally $xy \in \operatorname{SL}_2(k)$ has trace c , so

$$\operatorname{tr}((xy)^2) = c^2 - 2.$$

Substitution gives

$$abc - a(bc - a) - b(ac - b) + (c^2 - 2) = a^2 + b^2 + c^2 - abc - 2.$$

$\square$

REMARK 3.4. The polynomial

$$a^2 + b^2 + c^2 - abc - 2$$

is the coordinate expression for the character variety of a rank-two free group in SL_2 . We will not need that theory, but it is useful to recognize that the trace manipulations in expansion proofs sit inside a much richer geometric structure.

3. When does trace distinguish elements inside a torus?

Let

$$T = \{t(\lambda) = \text{diag}(\lambda, \lambda^{-1}) : \lambda \in k^\times\}.$$

Then

$$\text{tr}(t(\lambda)) = \lambda + \lambda^{-1}.$$

LEMMA 3.5 (Two-to-one trace on a split torus). *For $\lambda, \mu \in k^\times$,*

$$\lambda + \lambda^{-1} = \mu + \mu^{-1}$$

if and only if $\mu \in \{\lambda, \lambda^{-1}\}$. Hence every trace fiber on T has cardinality at most two.

PROOF. Rearranging gives

$$(\lambda - \mu) \left(1 - \frac{1}{\lambda\mu}\right) = 0.$$

Thus $\lambda = \mu$ or $\lambda\mu = 1$. □

COROLLARY 3.6. *For every finite $S \subseteq T$,*

$$|\text{tr}(S)| \geq \frac{1}{2}|S|.$$

PROOF. The preceding lemma shows that each trace value has at most two preimages in T . Therefore $|S| \leq 2|\text{tr}(S)|$, which is equivalent to the claimed lower bound. □

This elementary two-to-one property is one reason trace is especially effective in rank one. In higher rank, the Weyl group is larger and one studies several conjugation-invariant functions at once.

4. Simultaneously diagonalizable elements

Suppose x, y lie in a common split torus and write

$$x = t(\lambda), \quad y = t(\mu).$$

Then

$$\text{tr}(x) = \lambda + \lambda^{-1},$$

$$\text{tr}(y) = \mu + \mu^{-1},$$

$$\text{tr}(xy) = \lambda\mu + (\lambda\mu)^{-1},$$

$$\text{tr}(xy^{-1}) = \lambda\mu^{-1} + \lambda^{-1}\mu.$$

The basic trace identity becomes

$$(\lambda + \lambda^{-1})(\mu + \mu^{-1}) = \text{tr}(xy) + \text{tr}(xy^{-1}).$$

Thus multiplication of field-valued trace sets is controlled by sums of traces of group products.

PROPOSITION 3.7 (Trace product containment). *Let $S \subseteq T$ be a finite subset of a split torus and put $X = \text{tr}(S)$. Then*

$$X \cdot X \subseteq \text{tr}(S^2) + \text{tr}(S^2),$$

where the right side is a sumset in k .

PROOF. For $x, y \in S$, (9) gives

$$\text{tr}(x) \text{tr}(y) = \text{tr}(xy) + \text{tr}(xy^{-1}).$$

If S is symmetric then both products lie in S^2 . In general replace S by $S \cup S^{-1}$; trace is unchanged under inversion. $\square$

This alone does not give sum–product growth, because we also need to control additive combinations of traces. The missing ingredient comes from conjugating one torus relative to another.

5. Conjugation maps and centralizer fibers

Let $A \subseteq G$ and fix $g \in G$. Consider

$$\phi_g : A \rightarrow G, \quad a \mapsto aga^{-1}.$$

Two elements a, b have the same image precisely when $b^{-1}a \in C_G(g)$.

LEMMA 3.8 (Conjugation fiber lemma). *For finite $A \subseteq G$ and $g \in G$,*

$$|\{aga^{-1} : a \in A\}| \geq \frac{|A|}{\max_{x \in G} |A \cap xC_G(g)|}.$$

More precisely, each fiber of ϕ_g is an intersection of A with a right coset of $C_G(g)$.

PROOF. We have

$$aga^{-1} = bgb^{-1} \iff b^{-1}a \in C_G(g) \iff a \in bC_G(g).$$

The fiber through b is therefore $A \cap bC_G(g)$. Dividing $|A|$ by the largest possible fiber gives the result. $\square$

A form better adapted to product sets uses $A^{-1}A$.

LEMMA 3.9 (Centralizer intersection from small conjugacy orbit). *Let $A \subseteq G$ be finite and nonempty, and let*

$$\Lambda_A(g) = \{aga^{-1} : a \in A\}.$$

Then

$$|C_G(g) \cap A^{-1}A| \geq \frac{|A|}{|\Lambda_A(g)|}.$$

PROOF. Partition A into the fibers of ϕ_g . Some fiber has size at least $|A|/|\Lambda_A(g)|$. If a_0 lies in that fiber, then for every other a in the same fiber,

$$a_0^{-1}a \in C_G(g) \cap A^{-1}A.$$

The map $a \mapsto a_0^{-1}a$ is injective. $\square$

GUIDING IDEA 3.10. This lemma is a basic growth dichotomy. Either conjugating g by A creates many new elements, or $A^{-1}A$ contains many elements of the one-dimensional centralizer of g . The first alternative is direct growth; the second produces a structured torus slice that sum-product can attack.

6. A multiplication map with controlled fibers

A closely related device is the map

$$A \times S \rightarrow G, \quad (a, s) \mapsto asa^{-1},$$

where S lies in a torus. Distinct traces separate the s -coordinate because conjugation preserves trace.

PROPOSITION 3.11 (Conjugating a trace-separated torus slice). *Let $S \subseteq G$ be a set of pairwise distinct traces, and let $A \subseteq G$ be finite. Then*

$$|ASA^{-1}| \geq \sum_{s \in S} \frac{|A|}{\max_{x \in G} |A \cap xC_G(s)|}.$$

If all $s \in S$ are regular semisimple in one torus T , then $C_G(s) = T$, and hence

$$|ASA^{-1}| \geq |S| \frac{|A|}{\max_x |A \cap xT|}.$$

PROOF. For distinct $s, s' \in S$, the conjugacy images $\{asa^{-1} : a \in A\}$ and $\{as'a^{-1} : a \in A\}$ are disjoint because trace is constant on each and the traces differ. Apply Lemma 3.8 to each s and sum. If the elements are regular semisimple in T , uniqueness of the torus from Proposition 1.13 gives $C_G(s) = T$. $\square$

This estimate makes the exponent bookkeeping transparent. If $|A| = N$ and a torus slice has about $N^{1/3}$ elements while no torus coset captures more than $N^{1/3}$ elements, conjugation can create about N elements. To force a power gain, one needs either a larger slice, many tori, or a second operation. Helfgott's proof arranges precisely such an imbalance.

7. Trace ± 2 is exceptional but small

The hypersurfaces $\text{tr}(g) = 2$ and $\text{tr}(g) = -2$ contain the nonregular elements. In $\text{SL}_2(\mathbb{F}_q)$ they have only $O(q^2)$ points.

PROPOSITION 3.12 (Counting a trace fiber). *Let $t \in \mathbb{F}_q$. The number of matrices in $\mathrm{SL}_2(\mathbb{F}_q)$ with trace t is at most $q(q+1)$. In particular*

$$|\{g \in G : \mathrm{tr}(g) = \pm 2\}| \leq 2q(q+1).$$

PROOF. Write

$$g = \begin{pmatrix} a & b \\ c & t-a \end{pmatrix}.$$

The determinant condition is

$$bc = a(t-a) - 1.$$

For each a , if the right-hand side is nonzero there are $q-1$ choices for $b \neq 0$ and then a unique c ; if it is zero, the number of pairs (b, c) with $bc = 0$ is $2q-1$. Hence the total is at most

$$q \max(q-1, 2q-1) < 2q^2.$$

A slightly sharper inspection gives the stated $q(q+1)$ bound; for our applications any absolute multiple of q^2 suffices. To verify the displayed constant directly, let r be the number (at most two) of roots of $a(t-a) - 1 = 0$. The count equals

$$(q-r)(q-1) + r(2q-1) = q(q-1) + rq \leq q(q+1).$$

□

WARNING 3.13. A set A of size much smaller than q^2 could still lie entirely in the trace-2 hypersurface. Therefore ambient counting alone does not show that A contains many regular semisimple elements. What saves the growth proof is generation: a generating set cannot remain trapped in a bounded-complexity exceptional variety under boundedly many products. We will prove an explicit rank-one escape lemma in [Chapter 6](#).

8. The trace map on triples

A classical fact is that the three numbers

$$\mathrm{tr} x, \quad \mathrm{tr} y, \quad \mathrm{tr}(xy)$$

generically determine a pair (x, y) up to conjugation. We record a concrete version because it explains why three trace coordinates are the right number for a three-dimensional group.

PROPOSITION 3.14 (Generic reconstruction). *Let $x \in \mathrm{SL}_2(k)$ be split regular semisimple, with eigenvalues $\lambda \neq \lambda^{-1}$. Conjugate so that $x = \mathrm{diag}(\lambda, \lambda^{-1})$. If*

$$y = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then a and d are uniquely determined by $\mathrm{tr}(y)$ and $\mathrm{tr}(xy)$. The product bc is then determined by $\det y = 1$. If $bc \neq 0$, all such y with the same three trace data form a single orbit under conjugation by the diagonal torus.

PROOF. The equations

$$a + d = \text{tr}(y), \quad \lambda a + \lambda^{-1}d = \text{tr}(xy)$$

form a nonsingular 2×2 linear system because $\lambda \neq \lambda^{-1}$. Thus a, d are determined. Then $bc = ad - 1$.

Conjugation by $t(s) = \text{diag}(s, s^{-1})$ sends (b, c) to $(s^2b, s^{-2}c)$. If $bc \neq 0$, any two pairs (b, c) and (b', c') with the same nonzero product differ by such a scaling over an algebraically closed field; over k one must check whether b'/b is a square. The geometric statement, which is what later dimension arguments use, is therefore exact; rationally there may be a bounded splitting into orbits. $\square$

9. A worked trace-coordinate calculation

Let

$$x = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}, \quad \lambda \neq \lambda^{-1},$$

and write

$$y = \begin{pmatrix} r & s \\ t & u \end{pmatrix}, \quad ru - st = 1.$$

Then

$$\text{tr}(y) = r + u, \quad \text{tr}(xy) = \lambda r + \lambda^{-1}u.$$

Because $\lambda - \lambda^{-1} \neq 0$, the pair

$$(\text{tr}(y), \text{tr}(xy))$$

determines r and u uniquely. Once r, u are fixed, the determinant equation becomes

$$st = ru - 1.$$

If $ru \neq 1$, there are exactly $q - 1$ choices for (s, t) ; if $ru = 1$, there are $2q - 1$. Thus the generic fiber of the two-trace map has size $q - 1$, the size of a split torus.

This calculation is a model for the dimension principle used throughout the book. Two scalar coordinates cut the three-dimensional group down to a one-dimensional fiber. In the explicit rank-one proof one can see that fiber and count it directly. In the general Larsen–Pink framework one instead records that the fiber dimension has dropped from 3 to 1.

WARNING 3.15 (Do not count three dependent trace coordinates). Cayley–Hamilton gives

$$x^{-1} = (\text{tr } x)I - x,$$

and hence

$$\text{tr}(x^{-1}y) = (\text{tr } x)(\text{tr } y) - \text{tr}(xy).$$

Therefore the three quantities $\text{tr } y$, $\text{tr}(xy)$ and $\text{tr}(x^{-1}y)$ contain only two independent scalar coordinates. Any argument claiming a cubic cardinality gain from treating all three as independent is overcounting.

10. From identities to growth questions

There are now three mechanisms available:

- (1) *Trace multiplication*: for commuting semisimple elements, products of traces become sums of traces of products.
- (2) *Conjugation growth*: if an A -orbit under conjugation is small, then $A^{-1}A$ has a large intersection with a centralizer.
- (3) *Transversality*: moving a torus by an element outside its normalizer creates a new algebraic direction.

The finite-field sum–product theorem will show that a trace set cannot have simultaneously small additive and multiplicative complexity unless it is extremely small or extremely large. The remaining work is to manufacture the relevant trace set from A without losing too much cardinality. That is the purpose of the pivot machinery developed after the sum–product chapters.

CHECKPOINT 3.16. The identities in this chapter should be reproducible from Cayley–Hamilton alone. More importantly, the reader should know what each identity is *for*: (9) converts group multiplication into field sum–product, while Lemmas 3.8 and 3.9 convert a failure of conjugacy growth into concentration on a torus.

Part 2

Arithmetic Bridge

CHAPTER 4

The Small-Set Sum–Product Supplement

The finite-field theory developed in Volume III proves the Bourgain–Katz–Tao power-range theorem and the trace-shaped consequences needed once a set has size at least a fixed power of the characteristic. Helfgott’s rank-one argument also encounters torus slices whose size is only known to exceed an absolute constant. This appendix closes that gap from first principles. We prove the Stepanov estimate for differences in multiplicative subgroups, derive Konyagin’s small-set lower bound, combine it with the approximate-ring machinery already proved in Volume III, and finally obtain a uniform trace-shaped expansion statement for split and nonsplit torus eigenvalue sets.

GUIDING IDEA 4.1 (Why another arithmetic argument is needed). The power-range sum–product theorem has a lower hypothesis of the form

$$|A| \geq p^\rho.$$

A rich torus produced from a very small generating set need not satisfy any such lower bound. It is therefore logically invalid to feed that slice directly into the power-range theorem. The missing input is a genuinely small-set estimate over the prime field. Konyagin’s argument supplies it, but its heart is a nontrivial multiplicative-subgroup incidence estimate. We prove that estimate rather than treating it as an external lemma.

1. The Stepanov difference estimate for a multiplicative subgroup

Throughout this appendix p is an odd prime and $F = \mathbb{F}_p$. For a subset X of a multiplicative group we use the notation

$$X_{\pm, r} = (X \cup X^{-1} \cup \{1\})^r.$$

Let $\Gamma \leq F^\times$ be a multiplicative subgroup of cardinality h . If $u \in F^\times$, put

$$C(u) = \{x \in \Gamma : x - u \in \Gamma\}.$$

The cardinality of $C(u)$ depends only on the multiplicative coset $u\Gamma$: if $v = \gamma u$ with $\gamma \in \Gamma$, multiplication by γ is a bijection $C(u) \rightarrow C(v)$.

Choose representatives $u_1, u_2, \dots$ of the nonzero cosets of Γ and order them so that

$$N_1 \geq N_2 \geq \dots, \quad N_j := |C(u_j)|.$$

LEMMA 4.2 (Sparse-polynomial nonvanishing). *Let*

$$P(X) = \sum_{j=1}^N c_j X^{m_j} \in F[X]$$

be nonzero, with $c_j \neq 0$, distinct exponents $0 \leq m_j < p$, and $N \geq 1$. Then

$$(X - 1)^N \nmid P(X).$$

PROOF. We induct on N . The assertion is immediate for $N = 1$. Suppose $N > 1$ and choose one exponent, say m_N . Put

$$Q(X) = XP'(X) - m_N P(X).$$

The m_N -term cancels, while for $j < N$ the coefficient of X^{m_j} is

$$c_j(m_j - m_N),$$

which is nonzero in F because the exponents are distinct integers in $[0, p-1]$. Hence Q has exactly $N - 1$ nonzero monomials, all of degree $< p$.

If $(X - 1)^N$ divided P , then $(X - 1)^{N-1}$ would divide P' , and therefore also $XP' - m_N P = Q$. This contradicts the induction hypothesis applied to Q . Thus $(X - 1)^N \nmid P$. $\square$

THEOREM 4.3 (Stepanov cumulative coset bound). *With the notation above, let $T \geq 1$. If*

$$h^4 T < p^3,$$

then

$$\sum_{j=1}^T N_j \ll (hT)^{2/3},$$

with an absolute implied constant.

PROOF. Take the first T representatives $U = \{u_1, \dots, u_T\}$. For $u \in U$ define

$$D(u) = u^{-1}C(u) = \{y \in F : uy \in \Gamma, u(y-1) \in \Gamma\}.$$

If u, v represent distinct cosets of Γ , then $D(u)$ and $D(v)$ are disjoint: from a common y we would have $uy, vy \in \Gamma$, hence $u/v \in \Gamma$. Consequently, for

$$E = \bigcup_{u \in U} D(u)$$

we have

$$(11) \quad |E| = \sum_{u \in U} |C(u)| = \sum_{j=1}^T N_j.$$

We prove $|E| \ll (hT)^{2/3}$.

Choose positive integers A, B, D , to be specified later, and seek a nonzero polynomial

$$\Phi(X, Y, Z) = \sum_{\substack{0 \leq a < A \\ 0 \leq b, c < B}} \lambda_{a,b,c} X^a Y^b Z^c$$

such that

$$\Psi(X) := \Phi(X, X^h, (X-1)^h)$$

has a zero of multiplicity at least D at every $x \in E$.

For $n \geq 0$, multiplication by $\{X(X-1)\}^n$ clears the denominators that would otherwise appear when differentiating powers of X and $X-1$. More precisely, for every monomial

$$f(X) = X^a X^{hb} (X-1)^{hc}$$

there is a polynomial $P_{n,a,b,c}(X)$ of degree at most $a+n$ such that

$$\{X(X-1)\}^n f^{(n)}(X) = P_{n,a,b,c}(X) X^{hb} (X-1)^{hc}.$$

This assertion itself is elementary and we record the induction because the degree bound is important. For $n=0$ take $P_0 = X^a$. If the formula holds at level n , then

$$f^{(n)} = P_n(X) X^{hb-n} (X-1)^{hc-n}.$$

Differentiating and multiplying by $\{X(X-1)\}^{n+1}$ gives the recurrence

$$P_{n+1} = X(X-1)P'_n + ((hb-n)(X-1) + (hc-n)X)P_n.$$

Hence

$$\deg P_{n+1} \leq \deg P_n + 1 \leq a + n + 1,$$

which proves the claim for every n . If $x \in D(u)$, then

$$x^h = (x-1)^h = u^{-h}.$$

Hence, for fixed n and u , the value at $x \in D(u)$ of

$$\{X(X-1)\}^n \Psi^{(n)}(X)$$

is a polynomial $P_{n,u}(x)$ of degree $< A+n$, whose coefficients are linear forms in the AB^2 unknowns $\lambda_{a,b,c}$.

It is therefore sufficient to impose the stronger requirement

$$P_{n,u} \equiv 0 \quad (0 \leq n < D, u \in U).$$

The number of homogeneous linear conditions is at most

$$\sum_{n < D} T(A+n) \leq D(A+D)T.$$

Thus a nonzero choice of coefficients exists whenever

$$(12) \quad D(A+D)T < AB^2.$$

For such a choice, every point of E is a zero of Ψ of multiplicity at least D , provided Ψ is not the zero polynomial.

We now prove that $\Psi \not\equiv 0$ under two simple numerical conditions. Write

$$\Phi(X, Y, Z) = \sum_{c=0}^{B-1} \Phi_c(X, Y) Z^c$$

and let c_0 be the least index for which $\Phi_{c_0} \neq 0$. If $\Psi \equiv 0$, then after factoring $(X-1)^{hc_0}$ we obtain

$$\Phi_{c_0}(X, X^h) \equiv 0 \pmod{(X-1)^h}.$$

If $A \leq h$, the exponents $a + hb$ with $0 \leq a < A$, $0 \leq b < B$ are distinct. Thus $\Phi_{c_0}(X, X^h)$ is a nonzero polynomial with at most AB monomials. If moreover

$$(13) \quad AB \leq h, \quad A + hB < p,$$

then its degree is $< p$, and Lemma 4.2 says that $(X-1)^h$ cannot divide it. Hence Ψ is nonzero.

Since

$$\deg \Psi < A + 2hB,$$

root multiplicity gives

$$(14) \quad D|E| \leq A + 2hB.$$

It remains to choose parameters. If $T > h^2/8$, the trivial identity

$$\sum_{u \in \Gamma} |C(u)| = h - 1$$

already gives

$$|E| \leq h \ll (hT)^{2/3}.$$

We may therefore assume $T \leq h^2/8$. Set

$$A = \lfloor c_1 h^{2/3} T^{-1/3} \rfloor, \quad B = \lfloor c_1 h^{1/3} T^{1/3} \rfloor$$

with a sufficiently small absolute $c_1 > 0$. The assumption $T \leq h^2/8$ ensures $A, B \geq 1$ after harmless adjustment in the finitely many tiny cases, while $AB \leq h$. The hypothesis $h^4 T < p^3$ gives

$$hB \ll h^{4/3} T^{1/3} < p,$$

so decreasing c_1 ensures $A + hB < p$.

Finally choose

$$D = \lfloor c_2 h^{2/3} T^{-1/3} \rfloor$$

with $c_2 > 0$ sufficiently small compared with c_1 . Then

$$D(A + D)T \ll c_2(c_1 + c_2)h^{4/3}T^{1/3},$$

whereas

$$AB^2 \gg c_1^3 h^{4/3} T^{1/3},$$

so (12) holds. Substituting in (14) yields

$$|E| \ll \frac{h^{4/3}T^{1/3}}{h^{2/3}T^{-1/3}} = (hT)^{2/3}.$$

Together with (11), this proves the theorem. $\square$

REMARK 4.4 (What Stepanov is doing). The auxiliary polynomial is not an arbitrary trick. Every $y \in D(u)$ satisfies two multiplicative equations,

$$(uy)^h = 1, \quad (u(y-1))^h = 1.$$

The variables $Y = X^h$ and $Z = (X-1)^h$ record those two constraints. The coefficients of Φ are chosen so that many derivatives vanish simultaneously on many cosets. A sparse-polynomial argument prevents the auxiliary polynomial from vanishing identically, so multiplicity cannot exceed degree. This is the finite-field analogue of forcing too many high-order zeros of a low-degree function.

COROLLARY 4.5 (Difference growth inside a multiplicative subgroup). *Let $\Gamma \leq F^\times$ and $B \subseteq \Gamma$. If $|B| < p^{1/2}$, then*

$$|B - B| \gg \frac{|B|^{5/2}}{|\Gamma|}.$$

PROOF. Put $h = |\Gamma|$. For each nonzero coset Γ_j of Γ in $F^\times$, let N_j be the corresponding difference multiplicity as above, arranged in decreasing order. Define

$$L_j = |\{(b_1, b_2) \in B^2 : b_1 - b_2 \in \Gamma_j\}|$$

and

$$M_j = |\{(b_1, b_2, b_3, b_4) \in B^4 : b_1 - b_2 = b_3 - b_4 \in \Gamma_j\}|.$$

For fixed b_1 , there are at most N_j possible b_2 , hence

$$L_j \leq N_j |B|.$$

For a fixed difference $s \in \Gamma_j$, there are at most N_j pairs of elements of Γ with difference s , and therefore at most N_j pairs from B . Thus

$$M_j \leq N_j L_j \leq N_j^2 |B|.$$

Also

$$(15) \quad \sum_j L_j = |B|(|B| - 1).$$

If $h \geq |B|^{3/2}$, the asserted lower bound is at most a constant multiple of $|B|$, and the result is trivial because $|B - B| \geq |B|$. Assume henceforth

$$h < |B|^{3/2}.$$

Take

$$T = \left\lfloor \frac{|B|^{3/2}}{h} \right\rfloor.$$

Then $T \geq 1$. Since $|B| < p^{1/2}$,

$$h^4 T \ll h^3 |B|^{3/2} < |B|^6 < p^3,$$

so Theorem 4.3 applies. In particular

$$N_T \ll h^{2/3} T^{-1/3}$$

and, by summation of the squared pointwise bound $N_j \ll h^{2/3} j^{-1/3}$,

$$\sum_{j \leq T} N_j^2 \ll h^{4/3} T^{1/3}.$$

Hence

$$\sum_{j \leq T} M_j \ll |B| h^{4/3} T^{1/3}.$$

For the tail, monotonicity and (15) give

$$\sum_{j > T} M_j \leq N_T \sum_j L_j \ll h^{2/3} T^{-1/3} |B|^2.$$

With the chosen value of T , both terms are

$$\ll |B|^{3/2} h.$$

Therefore the additive energy of B satisfies

$$E_+(B) = |B|^2 + \sum_j M_j \ll |B|^{3/2} h + |B|^2 \ll |B|^{3/2} h,$$

where the last step uses $h \geq |B|$; if $h < |B|$ then $B \subseteq \Gamma$ is impossible. Cauchy–Schwarz now gives

$$|B - B| \geq \frac{|B|^4}{E_+(B)} \gg \frac{|B|^{5/2}}{h}.$$

□

2. Konyagin's small-set lower bound

For $A \subseteq F$ define

$$I(A) = A(A - A) + A(A - A) = \{a_1(a_2 - a_3) + a_4(a_5 - a_6) : a_i \in A\}.$$

LEMMA 4.6 (A collision turns a linear image into $I(A)$). For $\xi \in F^\times$ put

$$S_\xi(A) = A + \xi A.$$

If $|S_\xi(A)| < |A|^2$, then

$$|I(A)| \geq |S_\xi(A)|.$$

PROOF. The map $(a, b) \mapsto a + \xi b$ from A^2 is not injective, so there are distinct pairs with

$$(a_1 - a_2) + (b_1 - b_2)\xi = 0.$$

Necessarily $b_1 \neq b_2$. Multiplication of $S_\xi(A)$ by $b_1 - b_2$ is injective and gives

$$(b_1 - b_2)a + (b_1 - b_2)b\xi = (b_1 - b_2)a + (a_2 - a_1)b \in I(A).$$

Thus a set of cardinality $|S_\xi(A)|$ is contained in $I(A)$. $\square$

LEMMA 4.7 (A useful dilation inside a prescribed multiplicative set). *Let $A \subseteq F$ and let $G \subseteq F^\times$ be nonempty. There exists $\xi \in G$ with*

$$|A + \xi A| \geq \frac{|A|^2|G|}{|A|^2 + |G|}.$$

PROOF. For $\xi \in G$ and $s \in F$ let

$$r_\xi(s) = |\{(a, b) \in A^2 : a + \xi b = s\}|.$$

Summing collision counts over $\xi \in G$, every quadruple with $b_1 \neq b_2$ determines at most one value

$$\xi = \frac{a_2 - a_1}{b_1 - b_2}.$$

The diagonal pairs contribute $|A|^2|G|$. Hence

$$\sum_{\xi \in G} \sum_s r_\xi(s)^2 \leq |A|^2|G| + |A|^4.$$

For some $\xi \in G$,

$$\sum_s r_\xi(s)^2 \leq |A|^2 + \frac{|A|^4}{|G|}.$$

Since $\sum_s r_\xi(s) = |A|^2$, Cauchy–Schwarz gives

$$|A + \xi A| \geq \frac{|A|^4}{|A|^2 + |A|^4/|G|} = \frac{|A|^2|G|}{|A|^2 + |G|}.$$

$\square$

Assume now $A \subseteq F^\times$. Let

$$r_{A/A}(s) = |\{(a, b) \in A^2 : a/b = s\}|$$

and define the popular-ratio set

$$H = \left\{ s \in F^\times : r_{A/A}(s) \geq \frac{|A|^2}{5|A|} \right\}.$$

Let $G = \langle H \rangle$. Notice that H is symmetric and contains 1. Indeed

$$r_{A/A}(s) = r_{A/A}(s^{-1})$$

by interchanging the two entries of a ratio representation, while $r_{A/A}(1) = |A| \geq |A|^2/(5|AA|)$ because $|AA| \geq |A|$. Consequently every element of G is a product of elements of H without introducing a separate inverse alphabet. This elementary observation is needed in the propagation argument below.

LEMMA 4.8 (Popular ratios force concentration in one coset). *Some multiplicative coset gG satisfies*

$$|A \cap gG| \geq |A|/3.$$

PROOF. Suppose not. Partition A by the cosets of G and choose a minimal union A' of nonempty coset pieces with $|A'| \geq |A|/3$. By minimality and the assumed bound on every individual piece,

$$|A|/3 \leq |A'| < 2|A|/3.$$

Put $A'' = A \setminus A'$. Then

$$(16) \quad |A'| |A''| \geq 2|A|^2/9.$$

No ratio a'/a'' with $a' \in A'$, $a'' \in A''$ belongs to G , and hence none belongs to H . Thus, if

$$f(s) = |\{(a', a'') \in A' \times A'' : a'/a'' = s\}|,$$

then

$$f(s) < \frac{|A|^2}{5|AA|}.$$

Consequently

$$\sum_s f(s)^2 \leq \frac{|A|^2}{5|AA|} |A'| |A''|.$$

On the other hand, if

$$g(s) = |\{(a', a'') \in A' \times A'' : a'a'' = s\}|,$$

Cauchy–Schwarz gives

$$\sum_s g(s)^2 \geq \frac{(|A'| |A''|)^2}{|AA|}.$$

The two energies are equal: both count quadruples

$$a'_1 a'_2 = a'_2 a'_1.$$

Hence

$$|A'| |A''| \leq |A|^2/5,$$

contradicting (16). □

LEMMA 4.9 (A nonmaximal dilation from the popular-ratio group). *If $|A| > 1$, there exists $\xi \in G$ such that*

$$\min \left(\frac{|A|^3}{5|AA|}, \frac{|A|^2|G|}{|A|^2 + |G|} \right) \leq |A + \xi A| < |A|^2.$$

PROOF. If every $g \in G$ already satisfies $|A + gA| < |A|^2$, the conclusion follows from Lemma 4.7.

Otherwise choose $g \in G$ with $|A + gA| = |A|^2$. We claim that some $\xi \in G$ satisfies

$$\frac{|A|^3}{5|AA|} \leq |A + \xi A| < |A|^2.$$

If not, take any $h \in H$ and put

$$A_h = \{b \in A : bh \in A\}.$$

By definition of H ,

$$|A_h| \geq \frac{|A|^2}{5|AA|}.$$

The injectivity of $(a, b) \mapsto a + gb$ on A^2 implies injectivity of

$$(a, b) \mapsto a + g(hb) = a + (gh)b$$

on $A \times A_h$. Hence

$$|A + ghA| \geq |A||A_h| \geq \frac{|A|^3}{5|AA|}.$$

Under the negation of the claim this forces $|A + ghA| = |A|^2$. Thus the property of having maximal image propagates from g to gh for every $h \in H$, and hence to every element of the group generated by H , namely all of G . But $1 \in G$ and

$$|A + A| \leq |A|(|A| + 1)/2 < |A|^2$$

for $|A| > 1$, a contradiction. $\square$

THEOREM 4.10 (Konyagin's small-set inequality in the nontrivial product regime).
Let $A \subseteq F^\times$ with $1 < |A| < p^{1/2}$ and

$$|AA| \leq |A|^{3/2}.$$

Then

$$|A - A| |I(A)| \gg |A|^{5/2}.$$

In particular

$$|I(A)| \gg |A|^{5/4}.$$

PROOF. The hypothesis $|AA| \leq |A|^{3/2}$ is exactly the regime in which the product set has not already supplied a power gain. Let G be generated by the popular-ratio set above. By Lemma 4.9 and Lemma 4.6, there is $\xi \in G$ for which

$$(17) \quad |I(A)| \geq \min \left(\frac{|A|^3}{5|AA|}, \frac{|A|^2|G|}{|A|^2 + |G|} \right).$$

If $|G| > |A|^{3/2}$, then

$$\frac{|A|^2|G|}{|A|^2 + |G|} \gg |A|^{3/2}, \quad \frac{|A|^3}{5|AA|} \geq \frac{1}{5}|A|^{3/2},$$

so (17) gives $|I(A)| \gg |A|^{3/2}$. Since $|A - A| \geq |A|$, the result follows.

It remains to consider $|G| \leq |A|^{3/2}$. By the hypothesis $|AA| \leq |A|^{3/2}$, both terms in (17) are $\gg |G|$: the second because $|G| \leq |A|^2$, and the first because $|G| \leq |A|^{3/2} \leq |A|^3/(5|AA|)$ up to an absolute constant. Hence

$$|I(A)| \gg |G|.$$

By Lemma 4.8, some coset gG contains a set $A_0 = A \cap gG$ of size at least $|A|/3$. Scaling by g^{-1} gives $B \subseteq G$ with $|B| = |A_0|$. Since $|B| < p^{1/2}$, Corollary 4.5 yields

$$|A - A| \geq |A_0 - A_0| = |B - B| \gg \frac{|A|^{5/2}}{|G|}.$$

Multiplying by $|I(A)| \gg |G|$ proves the theorem.

Finally, $A - A \subseteq c^{-1}I(A)$ for any fixed $c \in A$: choose $a_1 = c$, set $a_4 = a_5 = a_6$ in the definition of $I(A)$. Thus $|I(A)| \geq |A - A|$, and

$$|I(A)|^2 \geq |A - A| |I(A)| \gg |A|^{5/2},$$

which gives $|I(A)| \gg |A|^{5/4}$. $\square$

REMARK 4.11 (Why the product-regime hypothesis is harmless). For the sum-product theorem, the complementary case $|AA| > |A|^{3/2}$ is already a much stronger conclusion than required. The Stepanov–Konyagin mechanism is needed only when multiplication itself has not already produced a power gain.

3. Uniform prime-field sum-product below the top scale

We now combine the preceding lower bound with the approximate-ring machinery already proved in Volume III. We use the following consequence of that volume: if

$$|A + A|, |AA| \leq K|A|,$$

then there is a subset $A' \subseteq A \setminus \{0\}$ of size at least $K^{-C}|A|$ for which every fixed polynomial expression obtained from A' by boundedly many additions, subtractions, and multiplications has cardinality at most $K^C|A'|$. This is exactly the proved Katz–Tao refinement followed by polynomial closure; no new source theorem is being invoked here.

THEOREM 4.12 (Uniform prime-field sum-product). *For every $\delta > 0$ there are constants $c = c(\delta) > 0$ and $\varepsilon = \varepsilon(\delta) > 0$ such that, for every prime p and every finite $A \subseteq \mathbb{F}_p$ with*

$$1 < |A| < p^{1-\delta},$$

one has

$$\max(|A + A|, |AA|) \geq c|A|^{1+\varepsilon}.$$

PROOF. If $|A| \geq p^{1/2}$, apply the power-range prime-field sum-product theorem proved in Volume III with

$$\rho = \min(\delta, 1/2).$$

Indeed $|A| \geq p^\rho$ and $|A| < p^{1-\rho}$ after weakening the upper bound when necessary. Thus it remains to treat $|A| < p^{1/2}$.

If $0 \in A$, replace A by $A_0 = A \setminus \{0\}$. For $|A| \geq 3$ one has $|A_0| \geq (2/3)|A|$, while $A_0 + A_0 \subseteq A + A$ and $A_0 A_0 \subseteq AA$; thus a lower bound for A_0 implies the same bound for A after changing the constant. The cases $|A| \leq 2$ are absorbed by decreasing c . We therefore assume $A \subseteq \mathbb{F}_p^\times$ and put

$$K = \frac{\max(|A + A|, |AA|)}{|A|}.$$

Let $C_0, C_1 \geq 1$ be absolute exponents supplied by the Katz–Tao refinement and polynomial-closure theorem proved in Volume III: there is $A' \subseteq A$ with

$$|A'| \geq K^{-C_0}|A|$$

and

$$|A' - A'|, |A'A'|, |I(A')| \leq K^{C_1}|A'|.$$

Choose once and for all an absolute number $\varepsilon_0 > 0$ so small that

$$C_1 \varepsilon_0 < \frac{1 - C_0 \varepsilon_0}{2} \quad \text{and} \quad \varepsilon_0 \left(2C_1 + \frac{C_0}{2} \right) < \frac{1}{4}.$$

We claim that $K \gg |A|^{\varepsilon_0}$. Suppose contrariwise that, after enlarging the harmless absolute threshold for $|A|$,

$$K < |A|^{\varepsilon_0}.$$

Then

$$|A'| \geq |A|^{1 - C_0 \varepsilon_0}.$$

The choice of ε_0 gives

$$K^{C_1} \leq |A|^{C_1 \varepsilon_0} \leq |A'|^{1/2},$$

so in particular

$$|A'A'| \leq |A'|^{3/2}.$$

Moreover $|A'| < p^{1/2}$. Therefore the Stepanov–Konyagin inequality proved in Theorem 4.10 applies and yields

$$|A' - A'| |I(A')| \gg |A'|^{5/2}.$$

The two polynomial-closure upper bounds give

$$K^{2C_1} |A'|^2 \gg |A'|^{5/2},$$

hence

$$|A'|^{1/2} \ll K^{2C_1}.$$

Using $|A'| \geq K^{-C_0}|A|$ we obtain

$$|A|^{1/2} \ll K^{2C_1 + C_0/2}.$$

For sufficiently large $|A|$ this contradicts $K < |A|^{\varepsilon_0}$ by the second inequality in the choice of ε_0 . Thus

$$K \gg |A|^{\varepsilon_0}.$$

Absorbing the implicit constant by reducing the exponent slightly for large sets and then decreasing c to handle the finitely many bounded cardinalities proves the small-set range. Combining it with the power-range theorem from Volume III proves the result for all $1 < |A| < p^{1-\delta}$. $\square$

4. Uniform trace-shaped expansion for rank-one torus parameters

Let

$$w(x) = x + x^{-1}.$$

Call $X \subseteq \mathbb{F}_{p^2}^\times$ *toral over $\mathbb{F}_p$* if either

$$X \subseteq \mathbb{F}_p^\times$$

or

$$X \subseteq U_1 := \{x \in \mathbb{F}_{p^2}^\times : x^{p+1} = 1\}.$$

In both cases

$$w(X) \subseteq \mathbb{F}_p :$$

for $x \in U_1$, one has $x^{-1} = x^p$, so $w(x) = x + x^p$ is Frobenius-fixed.

PROPOSITION 4.13 (Uniform expansion of products of toral trace parameters). *Fix $\delta > 0$. There are $c, \eta > 0$, depending only on δ , such that if X is toral over $\mathbb{F}_p$ and*

$$C(\delta) < |X| < p^{1-\delta},$$

then

$$|w(X_{\pm,2})w(X_{\pm,2})| \geq c|X|^{1+\eta}.$$

PROOF. The argument is the one proved in the trace-shaped chapter of Volume III; we indicate every place at which the range hypothesis changes. Put $N = |X|$, $P = X_{\pm,2}$ and $U = w(P) \subseteq \mathbb{F}_p$. Suppose

$$|UU| \leq N^{1+\theta}.$$

The map w has fibers of size at most two, so

$$N/2 \leq |U| \leq N^{1+\theta}.$$

The relation

$$\mathcal{R} = \{(w(xy), w(x/y)) : x, y \in X\} \subseteq U^2$$

has cardinality at least $N^2/16$ by the sixteen-fiber calculation proved in Volume III, and every $(r, s) \in \mathcal{R}$ satisfies

$$r + s = w(x)w(y) \in UU.$$

Exactly as in the proved BSG reduction of Volume III, Cauchy–Schwarz and the energy–approximate-group theorem of Volume II produce a subset $U' \subseteq U$ with

$$|U'| \geq N^{1-C_1\theta},$$

and

$$|U' + U'|, |U'U'| \leq N^{1+C_2\theta}.$$

The only point where the proof in Volume III used the lower bound $N \geq p^\rho$ was to ensure that U' lay in the power range of the old sum–product theorem. Here that is unnecessary: $U' \subseteq \mathbb{F}_p$, and Theorem 4.12 has no power lower threshold. For θ small in terms of δ , the upper bound

$$|U'| \leq |U| \leq N^{1+\theta} < p^{1-\delta/2}$$

holds, while $|U'|$ exceeds the constant threshold once N does. Therefore

$$\max(|U' + U'|, |U'U'|) \geq c_0|U'|^{1+\varepsilon_0}$$

by Theorem 4.12. The lower bound for $|U'|$ makes this incompatible with $N^{1+C_2\theta}$ when θ is sufficiently small. Thus $|UU| \geq cN^{1+\eta}$. $\square$

LEMMA 4.14 (Prime-field coefficients for a nonsplit transverse calculation). *Let $T < \mathrm{SL}_2(\mathbb{F}_p)$ be nonsplit. Choose an eigenbasis (v, v^p) over $\mathbb{F}_{p^2}$ so that T becomes diagonal. If $g \in \mathrm{SL}_2(\mathbb{F}_p)$ has matrix*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

in this basis, then

$$d = a^p, \quad c = b^p.$$

Consequently

$$ad = N(a) \in \mathbb{F}_p, \quad bc = N(b) \in \mathbb{F}_p.$$

If g is transverse to the two eigendirections, both are nonzero.

PROOF. Frobenius fixes the linear transformation g and exchanges the basis vectors v and v^p . From

$$gv = av + cv^p$$

we obtain by applying Frobenius

$$gv^p = a^p v^p + c^p v.$$

But the second column of the matrix is $gv^p = bv + dv^p$. Hence $b = c^p$ and $d = a^p$; applying Frobenius once more gives $c = b^p$. The coefficient claims follow. Transversality means all four entries are nonzero. $\square$

THEOREM 4.15 (Uniform trace-shaped expansion for torus slices). *Fix $\delta > 0$. There are $C = C(\delta)$ and $\varepsilon = \varepsilon(\delta) > 0$ with the following property. Let $X \subseteq \mathbb{F}_{p^2}^\times$ be toral over $\mathbb{F}_p$, with*

$$C < |X| < p^{1-\delta},$$

and let $a_1, a_2 \in \mathbb{F}_p^\times$. Then

$$\left| \{a_1(xy + x^{-1}y^{-1}) + a_2(x^{-1}y + xy^{-1}) : x, y \in X_{\pm,20}\} \right| \geq |X|^{1+\varepsilon}.$$

PROOF. The coefficient-transfer lemma, the fixed- y quadratic fiber lemma, and the square-coset covering argument were all proved in the trace-shaped chapter of Volume III without using any lower power bound on $|X|$. The lower power bound entered that proof only through the preliminary expansion estimate for

$$w(X_{\pm,2})w(X_{\pm,2}).$$

Replace that preliminary proposition by [Proposition 4.13](#). The remainder of the earlier proof applies verbatim: either $|X_{\pm,6}|$ has already grown by a power, in which case a fixed nonexceptional y and the quadratic fiber bound give the conclusion, or multiplicative Ruzsa covering places the relevant products in $O(|X|^\gamma)$ square-cosets. Pigeonholing one square-coset produces a set $Y = w(B) \subseteq \mathbb{F}_p$ with a power-large self-sum. The coefficient-transfer lemma then gives a power-large set $a_1Y + a_2Y$, and the explicit substitutions

$$r = a_j u^2, \quad s = a_j v^2, \quad y = u/v, \quad x = a_j uv$$

show that every such mixed sum is a value of the displayed two-variable trace expression with $x, y \in X_{\pm,20}$. All word exponents and fiber constants are the same absolute constants as in the proved Volume III argument. This yields the asserted exponent after decreasing it to absorb bounded-size cases. $\square$

CHECKPOINT 4.16. The arithmetic gap needed for the rank-one growth theorem is now closed internally. The chain is

$$\begin{aligned} & \text{Stepanov multiplicative-coset estimate} \\ & \implies \text{Konyagin small-set inequality} \\ & \implies \text{uniform prime-field sum-product} \\ & \implies \text{uniform toral trace expansion.} \end{aligned}$$

The middle implication uses only the Katz–Tao refinement and polynomial closure already proved in Volume III; the last implication uses only the BSG and coefficient-transfer machinery already proved in Volumes II–III. No small-set source theorem remains as a black box.

Part 3

Tori, Trace Growth, and Escape

CHAPTER 5

From Torus Slices to Trace Growth

The sum–product theorem lives in a field, whereas the product theorem lives in a noncommutative group. This chapter builds the transfer mechanism between them. A large slice of a product set inside a maximal torus is encoded by its eigenvalues; conjugating that torus into a transverse position produces a two-parameter trace polynomial; and the finite-field sum–product theorem forces that polynomial to take many values. Every loss in passing from matrices to eigenvalues and back is recorded.

GUIDING IDEA 5.1 (The transverse-torus mechanism). A one-dimensional torus slice by itself can only produce one-dimensional growth. The key move is to place a conjugate torus in transverse position. Multiplying elements from the two directions creates a map with generically small fibers, and taking trace converts that multiplication map into a two-variable field expression. Sum–product then prevents the image from remaining one-dimensional unless the original set has already grown.

This is the first place where the geometric slogan “dimension adds in transverse directions” becomes a cardinality argument. In SL_2 , the matrix calculations are explicit enough that no abstract dimension theorem is needed. Later Larsen–Pink theory will recognize the same pattern without coordinates.

1. Eigenvalue coordinates for a torus

Let T be a maximal torus of $G = \mathrm{SL}_2(\mathbb{F}_p)$. Over $K = \mathbb{F}_{p^2}$ every torus splits, so after conjugating in $\mathrm{GL}_2(K)$ we may write

$$T(K) = \{t(x) = \mathrm{diag}(x, x^{-1}) : x \in K^\times\}.$$

If T is split over $\mathbb{F}_p$, then its $\mathbb{F}_p$ -points correspond to $x \in \mathbb{F}_p^\times$; if T is nonsplit, its $\mathbb{F}_p$ -points correspond to norm-one elements of $\mathbb{F}_{p^2}^\times$.

DEFINITION 5.2. For $V \subseteq T(\mathbb{F}_p)$, an *eigenvalue set* for V is

$$\Lambda(V) = \{x \in \mathbb{F}_{p^2}^\times : t(x) \text{ is the chosen diagonal form of some } v \in V\}.$$

The choice of eigenline ordering replaces x by x^{-1} ; all cardinality estimates below are invariant under this ambiguity.

LEMMA 5.3 (No loss in eigenvalue coordinates). *For every $V \subseteq T(\mathbb{F}_p)$,*

$$|\Lambda(V)| = |V|.$$

Moreover

$$\Lambda(V^m) \supseteq \Lambda(V)^m$$

for every $m \geq 1$, after fixing one diagonalization of T over $\mathbb{F}_{p^2}$.

PROOF. The map $x \mapsto t(x)$ is injective. Once the torus is simultaneously diagonalized, multiplication corresponds to multiplication of eigenvalues:

$$t(x_1) \cdots t(x_m) = t(x_1 \cdots x_m).$$

□

The trace coordinate is

$$\mathrm{tr}(t(x)) = w(x) = x + x^{-1}.$$

As noted in Corollary 3.6, the map w has fibers of size at most two.

COROLLARY 5.4. *For $V \subseteq T(\mathbb{F}_p)$,*

$$\frac{1}{2}|V| \leq |\mathrm{tr}(V)| \leq |V|.$$

PROOF. The upper bound is immediate because trace is a function on V . The lower bound is Corollary 3.6, applied after identifying T with its eigenvalue parameter torus over the splitting field. □

2. Moving a torus into transverse position

Fix the diagonal torus over a field K and let

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(K).$$

The condition that g preserve or interchange the two coordinate eigenlines is exactly $g \in N(T)$. Thus if $g \notin N(T)$, the entries are not arranged in one of the two monomial patterns. For the cleanest counting argument we want all four entries nonzero.

REMARK 5.5 (How transversality will be produced). For the arithmetic statement of this chapter, g is an input: we assume that it sends neither eigenline of T to either eigenline. In an eigenbasis this is equivalent to all four matrix entries being nonzero. The existence of such a g in a bounded word set is a separate escape problem. It is proved directly, over both split and nonsplit tori, in Corollary 6.9. Keeping the two tasks separate prevents a qualitative “ $g \notin N(T)$ ” statement from being mistaken for the stronger four-entry transversality actually needed by the trace polynomial.

After choosing an eigenbasis for T , the conclusion says that the four entries a, b, c, d of g are all nonzero. Indeed, $a = 0$ or $c = 0$ means the first eigenline is sent to one of the coordinate eigenlines, and similarly for the second column.

3. The decisive trace computation

PROPOSITION 5.6 (Two-torus trace polynomial). *Let*

$$t(x) = \begin{pmatrix} x & 0 \\ 0 & x^{-1} \end{pmatrix}, \quad g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(K).$$

Then for all $x, y \in K^\times$,

$$(18) \quad \mathrm{tr}(t(x)gt(y)g^{-1}) = ad(xy + x^{-1}y^{-1}) - bc(xy^{-1} + x^{-1}y).$$

If a, b, c, d are all nonzero, then the two coefficients ad and $-bc$ are nonzero.

PROOF. Since $\det g = 1$,

$$g^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

A direct multiplication gives

$$t(x)gt(y)g^{-1} = \begin{pmatrix} x(ady - bcy^{-1}) & xab(y^{-1} - y) \\ x^{-1}cd(y - y^{-1}) & x^{-1}(ady^{-1} - bcy) \end{pmatrix}.$$

Taking the trace and regrouping terms yields (18). The last assertion is immediate. $\square$

GUIDING IDEA 5.7. This formula is the exact point at which noncommutative geometry creates a sum–product expression. If g normalized T , one of the coefficients would disappear and the expression would collapse to a one-dimensional function. Transversality of g is therefore not cosmetic: it is what forces two genuinely different monomials to coexist.

4. Deriving the transverse trace formula by hand

Because this calculation is the exact bridge from group multiplication to field arithmetic, it is worth doing once without suppressing any step. Put

$$t(x) = \begin{pmatrix} x & 0 \\ 0 & x^{-1} \end{pmatrix}, \quad g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad ad - bc = 1.$$

Then

$$g^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

and a direct multiplication gives

$$\mathrm{tr}(t(x)gt(y)g^{-1}) = ad \left(xy + \frac{1}{xy} \right) - bc \left(\frac{x}{y} + \frac{y}{x} \right).$$

Introduce

$$\alpha = ad, \quad \beta = -bc.$$

Since $ad - bc = 1$, one has

$$\alpha + \beta = 1.$$

Thus the trace is a weighted combination of two different multiplicative coordinates:

$$\alpha w(xy) + \beta w(x/y), \quad w(z) = z + z^{-1}.$$

If g lies in the normalizer of the diagonal torus, the geometry degenerates: in a suitable monomial pattern one of the two coefficients disappears, and the trace depends on only one multiplicative combination of x and y . If all four entries of g are nonzero, then $\alpha\beta \neq 0$ and both combinations survive. This is the concrete meaning of transversality in rank one.

EXAMPLE 5.8 (A particularly symmetric transverse matrix). If the field contains an element r with $2r^2 = 1$, take

$$g = \begin{pmatrix} r & r \\ -r & r \end{pmatrix}.$$

Then $\det g = 1$ and

$$ad = -bc = \frac{1}{2}.$$

The trace polynomial becomes

$$\frac{1}{2} \left(xy + (xy)^{-1} + x/y + y/x \right).$$

The four monomials make visible why neither purely additive nor purely multiplicative structure can keep the image small over a prime field.

EXAMPLE 5.9 (The transverse trace polynomial numerically in $\mathbb{F}_7$). In $\mathbb{F}_7$, the element $r = 2$ satisfies $2r^2 = 1$. Therefore the preceding symmetric choice becomes

$$g = \begin{pmatrix} 2 & 2 \\ 5 & 2 \end{pmatrix}, \quad ad = -bc = 4 = \frac{1}{2} \quad \text{in } \mathbb{F}_7.$$

Take $x = 2$ and $y = 3$. Since

$$x^{-1} = 4, \quad y^{-1} = 5,$$

the four multiplicative coordinates are

$$xy = 6, \quad (xy)^{-1} = 6, \quad x/y = 3, \quad y/x = 5.$$

Hence Equation (18) gives

$$\mathrm{tr}(t(2)g t(3)g^{-1}) = 4(6 + 6 + 3 + 5) \equiv 4 \cdot 6 \equiv 3 \pmod{7}.$$

The calculation exhibits both coordinates xy and x/y in one concrete matrix product. If g normalized T , one coefficient would vanish and this two-directional arithmetic would collapse.

5. A trace-growth lemma from a torus slice

We now package the uniform arithmetic supplement proved in Chapter 4 into a group-theoretic statement. Unlike the power-range theorem of Volume III, this version has only a constant lower threshold, which is essential because a rich torus slice in Helfgott's argument can have subpower cardinality.

PROPOSITION 5.10 (Torus slice forces uniform trace growth). *Fix $\delta > 0$. There are constants $C = C(\delta)$, $c = c(\delta) > 0$, and $\varepsilon = \varepsilon(\delta) > 0$ with the following property. Let p be prime, $T < \mathrm{SL}_2(\mathbb{F}_p)$ a maximal torus, and $V \subseteq T(\mathbb{F}_p)$ a finite set with*

$$C < |V| < p^{1-\delta}.$$

Suppose $g \in \mathrm{SL}_2(\mathbb{F}_p)$ sends neither eigenline of T to either eigenline of T over $\mathbb{F}_{p^2}$. Put

$$V_{\pm} = V \cup V^{-1} \cup \{1\}.$$

Then

$$|\mathrm{tr}(V_{\pm}^{20} g V_{\pm}^{20} g^{-1})| \geq c|V|^{1+\varepsilon}.$$

PROOF. Diagonalize T over its splitting field K , which is either $\mathbb{F}_p$ or $\mathbb{F}_{p^2}$, and let $X = \Lambda(V)$. By Lemma 5.3, $|X| = |V|$. Write

$$X_{\pm,20} = (X \cup X^{-1} \cup \{1\})^{20};$$

then $X_{\pm,20}$ parameterizes $V_{\pm}^{20}$.

In the chosen eigenbasis, the hypothesis on g means all four entries of its matrix are nonzero. By Proposition 5.6,

$$\mathrm{tr}(t(x)gt(y)g^{-1}) = a_1w(xy) + a_2w(x/y), \quad w(z) = z + z^{-1},$$

with $a_1a_2 \neq 0$. In the split case $a_1, a_2 \in \mathbb{F}_p^{\times}$ immediately. In the nonsplit case Lemma 4.14 proves, using Frobenius conjugacy of the eigenbasis, that the same coefficients again lie in $\mathbb{F}_p^{\times}$. Moreover X is toral over $\mathbb{F}_p$: it lies in $\mathbb{F}_p^{\times}$ in the split case and in the norm-one subgroup of $\mathbb{F}_{p^2}^{\times}$ in the nonsplit case.

The uniform trace-shaped theorem Theorem 4.15 therefore applies and gives at least $c|X|^{1+\varepsilon}$ values of

$$a_1w(xy) + a_2w(x/y)$$

as x, y range over $X_{\pm,20}$. These are precisely traces of matrices in $V_{\pm}^{20}gV_{\pm}^{20}g^{-1}$, and $|X| = |V|$. $\square$

REMARK 5.11 (Why the quadratic splitting field causes no new obstruction). For a nonsplit torus, the eigenvalue parameter lies in the norm-one group

$$U_1 = \{x \in \mathbb{F}_{p^2}^{\times} : x^{p+1} = 1\},$$

so $w(x) = x + x^{-1} = x + x^p \in \mathbb{F}_p$. Chapter 4 proves the needed trace-shaped expansion by reducing these prime-field trace parameters to the uniform prime-field

sum–product theorem. Thus no unproved “sum–product over $\mathbb{F}_{p^2}$ ” statement is being used here.

6. From many traces back to many matrices

Many trace values trivially imply many matrices, but in the growth proof we need a quantitative lower bound strong enough to offset earlier losses. The following controlled-fiber map is the rank-one substitute for a dimension argument.

PROPOSITION 5.12 (Two trace coordinates and centralizer-scale fibers). *Let $x \in \mathrm{SL}_2(\mathbb{F}_p)$ be split regular semisimple. The map*

$$\Psi_x : \mathrm{SL}_2(\mathbb{F}_p) \rightarrow \mathbb{F}_p^2, \quad y \mapsto (\mathrm{tr} y, \mathrm{tr}(xy))$$

has fibers of cardinality at most $2p$. Away from a degenerate locus its nonempty fibers have exactly $p - 1$ points.

PROOF. Diagonalize

$$x = \mathrm{diag}(\lambda, \lambda^{-1}), \quad \lambda \neq \lambda^{-1},$$

and write

$$y = \begin{pmatrix} r & s \\ t & u \end{pmatrix}.$$

The two trace equations

$$r + u = \alpha, \quad \lambda r + \lambda^{-1}u = \beta$$

form a nonsingular linear system in r, u , because $\lambda - \lambda^{-1} \neq 0$. Thus a fiber fixes r and u . The determinant condition becomes

$$st = ru - 1.$$

If $ru - 1 \neq 0$, there are exactly $p - 1$ choices, one for every $s \in \mathbb{F}_p^\times$, with $t = (ru - 1)s^{-1}$. If $ru - 1 = 0$, the solutions are $s = 0$ or $t = 0$, hence there are $2p - 1$ choices. In all cases the fiber has at most $2p$ points. $\square$

WARNING 5.13 (The third trace is redundant). It is tempting to add $\mathrm{tr}(x^{-1}y)$ as a third coordinate and then speak of three independent trace parameters. That is false. Cayley–Hamilton for $x \in \mathrm{SL}_2$ gives

$$x + x^{-1} = (\mathrm{tr} x)I,$$

so

$$\mathrm{tr}(x^{-1}y) = (\mathrm{tr} x)(\mathrm{tr} y) - \mathrm{tr}(xy).$$

The genuinely three-dimensional count in the growth argument comes instead from multiplying three transverse one-dimensional torus pieces. This distinction is important: trace coordinates detect conjugacy, whereas transversality creates dimension.

GUIDING IDEA 5.14 (Where the cubic count really comes from). If T is a maximal torus and g is sufficiently transverse to T , then the multiplication map

$$T \times gTg^{-1} \times T \longrightarrow \mathrm{SL}_2, \quad (t_1, t_2, t_3) \mapsto t_1 t_2 t_3,$$

has domain and target of dimension three and is generically finite. In rank one this can be checked by an explicit matrix calculation, yielding a lower bound of order $|V|^3$ for a finite slice $V \subseteq T$. We will use only the explicit bounded-fiber form of this statement, not the dimension slogan by itself.

7. Trace lower bounds from a generating set

The next proposition is the other direction: before sum–product can act, we must know that some bounded product of A already has many trace values.

GUIDING IDEA 5.15 (The cube-root target). The later growth argument needs a quantitative statement of the following shape. Under a small-tripling hypothesis, a bounded product of a generating set should have at least $\gg |A|^{1/3}$ trace values. The exponent is forced by

$$\frac{\dim T}{\dim \mathrm{SL}_2} = \frac{1}{3}.$$

There are two possible ways to obtain such a bound. Either many regular semisimple conjugacy classes are already represented, giving many traces directly, or one conjugacy class is represented so heavily that the corresponding centralizer contains a large torus slice. A transverse conjugate of that slice then produces cubic growth. This dichotomy is proved quantitatively in the rank-one growth arc; at this stage we record it only as the target that explains why the arithmetic trace estimates above have the correct scale.

WARNING 5.16 (A dimension heuristic is not yet a proof). The exponent $1/3$ comes from $\dim T / \dim G = 1/3$, but one may not infer the inequality by simply dividing dimensions. The proof needs two quantitative ingredients: bounded escape from the nonregular variety and a controlled-fiber multiplication map. Neither is replaced by dimension-counting notation in the proof of the product theorem.

8. The transfer chain

The rank-one arithmetic mechanism can now be summarized without suppressing any conceptual step:

$$\begin{aligned} & \boxed{A \text{ has small tripling}} \implies \boxed{A^m \cap T \text{ is large}} \\ \implies & \boxed{\Lambda(A^m \cap T) \subseteq \mathbb{F}_{p^2}^\times \text{ is large}} \implies \boxed{\text{sum–product creates many trace values}} \\ \implies & \boxed{\text{controlled fibers create many matrices}} \implies \boxed{A \text{ grows}}, \end{aligned}$$

a contradiction. The subsequent rank-one argument supplies the escape and fiber estimates needed to turn this transfer chain into a proof.

CHAPTER 6

Centralizers, Escape, and the Production of Torus Slices

Small product growth must manufacture structure. In rank one, the relevant structure is a large intersection with a maximal torus. We derive it in three stages: conjugacy classes force a large centralizer intersection; a generating set escapes the trace- ± 2 and eigenline degeneracies in bounded word length; and a large trace set therefore yields a quantitatively large simultaneously diagonalizable subset. The escape lemma is proved from first principles because its quantitative form is reused throughout the product theorem.

CONVENTION 6.1 (Symmetric word-set convention). Whenever a result in this chapter uses a literal product power A^m , the set A is assumed symmetric and to contain 1. A general finite set is first replaced by

$$A_{\pm} = A \cup A^{-1} \cup \{1\}.$$

This changes its size by at most an absolute factor and preserves the generated subgroup.

GUIDING IDEA 6.2 (Where the exponent $1/3$ comes from). A generic maximal torus in SL_2 has dimension 1 inside a group of dimension 3. If a finite set A behaves like a roughly equidistributed three-dimensional object of size M , one expects a typical torus slice to have size about $M^{1/3}$. The centralizer and conjugacy-class arguments in this chapter make that heuristic rigorous up to controlled powers of the tripling constant.

The proof has two logically separate jobs. It must first produce enough regular semisimple elements so that torus centralizers are available. It must then show that at least one such centralizer meets a bounded word in A in the correct quantitative range. Skipping either step hides the origin of the cubic exponent that powers the later pivot argument.

1. A conjugacy-class counting principle

Let G be an arbitrary group and $A \subseteq G$ finite and nonempty. Denote by $C(A)$ the set of G -conjugacy classes meeting A .

PROPOSITION 6.3 (Conjugacy classes versus centralizers). *There exists $g \in A$ such that*

$$|C_G(g) \cap A^{-1}A| \geq \frac{|C(A)| |A|}{|AAA^{-1}|}.$$

PROOF. Choose one representative $g_C \in A \cap C$ for every conjugacy class $C \in C(A)$. For fixed $g = g_C$, the set

$$\{hgh^{-1} : h \in A\}$$

is contained in AAA^{-1} . Every fiber of the map $h \mapsto hgh^{-1}$ has size at most

$$|C_G(g) \cap A^{-1}A|,$$

because if $h_1gh_1^{-1} = h_2gh_2^{-1}$ then $h_2^{-1}h_1 \in C_G(g) \cap A^{-1}A$. Hence

$$|\{hgh^{-1} : h \in A\}| \geq \frac{|A|}{|C_G(g) \cap A^{-1}A|}.$$

The conjugacy orbits arising from distinct classes C are disjoint. Summing over $C \in C(A)$ gives

$$|AAA^{-1}| \geq \sum_C \frac{|A|}{|C_G(g_C) \cap A^{-1}A|}.$$

If every denominator were smaller than

$$\frac{|C(A)| |A|}{|AAA^{-1}|},$$

the sum would exceed $|AAA^{-1}|$. Therefore at least one g_C satisfies the claimed inequality. $\square$

COROLLARY 6.4 (Trace classes produce a torus slice). *Let q be an odd prime power, put $G = \mathrm{SL}_2(\mathbb{F}_q)$, let $A \subseteq G$, and let*

$$A_{\mathrm{reg}} = \{a \in A : \mathrm{tr}(a) \neq \pm 2\}.$$

Then there exists $g \in A_{\mathrm{reg}}$ such that, with $T = C_G(g)$,

$$|T \cap A^{-1}A| \geq \frac{|\mathrm{tr}(A_{\mathrm{reg}})| |A_{\mathrm{reg}}|}{|A_{\mathrm{reg}}A_{\mathrm{reg}}A_{\mathrm{reg}}^{-1}|}.$$

The subgroup T is a maximal torus.

PROOF. Regular semisimple elements with distinct traces cannot be conjugate, so

$$|C(A_{\mathrm{reg}})| \geq |\mathrm{tr}(A_{\mathrm{reg}})|.$$

Apply Proposition 6.3 and use Proposition 1.13. $\square$

2. Producing many regular semisimple elements

A generating set can initially contain many unipotents, so we need a quantitative bounded-word statement. The following is Helfgott's elementary lemma, rewritten under our symmetric-set convention.

LEMMA 6.5 (Helfgott: many non-unipotent elements). *Let K be a field with $\text{char } K \neq 2$, and let $A \subseteq \text{SL}_2(K)$ be finite, symmetric, contain 1, and not be contained in a proper subgroup. Then A^2 contains at least*

$$\frac{1}{4}|A| - 1$$

elements whose trace is different from ± 2 .

PROOF. If A already contains at least $|A|/4 - 1$ such elements there is nothing to prove. Otherwise there must be a noncentral element of trace ± 2 : if every noncentral element were regular semisimple, then all but the at most two central elements $\pm I$ would already be regular, which is more than $|A|/4 - 1$ once $|A| > 4$; the remaining bounded cases are immediate. Choose therefore

$$g \in A \setminus \{\pm I\}$$

with $\text{tr } g = \pm 2$. Such a noncentral element has a unique eigenline ℓ over K . Let $B \subseteq A$ consist of the elements with trace ± 2 that have ℓ as an eigenline.

First suppose

$$|B| \leq \frac{1}{4}|A| + 3.$$

Take $h \in A \setminus B$. If h is already regular semisimple, keep h . Otherwise h has trace ± 2 but does not share the eigenline of g . For the trace calculation only, multiply g and h by suitable central scalars in $\{\pm I\}$ so that the normalized matrices have trace 2. This does not assert that the normalized matrices lie in A ; it merely changes the traces of the actual products gh and $g^{-1}h$ by signs, so regularity of the normalized products is equivalent to regularity of the corresponding products formed from the original elements. In a basis with the eigenline of g equal to Ke_1 we then have

$$g = \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}, \quad h = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad u \neq 0, \quad c \neq 0, \quad a + d = 2.$$

Here $c \neq 0$ is exactly the assertion that h does not preserve the eigenline of g . Direct multiplication gives

$$\text{tr}(gh) = 2 + uc, \quad \text{tr}(g^{-1}h) = 2 - uc.$$

If both numbers belonged to $\{\pm 2\}$, then their sum 4 would force both to equal 2; hence $uc = 0$, contradicting $u, c \neq 0$. Thus at least one of

$$h, \quad gh, \quad g^{-1}h$$

is regular semisimple. All three lie in A^2 . A fixed output can arise from at most three choices of h through these three maps, so A^2 contains at least

$$\frac{1}{3}|A \setminus B| \geq \frac{1}{3} \left(\frac{3}{4}|A| - 3 \right) = \frac{1}{4}|A| - 1$$

regular semisimple elements.

Now suppose

$$|B| > \frac{1}{4}|A| + 3.$$

Since A is not contained in the Borel stabilizing ℓ , choose $h \in A$ which does not preserve ℓ . In a basis with $\ell = Ke_1$, every $b \in B$ has the form

$$b = \varepsilon \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \quad \varepsilon \in \{\pm 1\}.$$

Writing $h = (h_{ij})$ with $h_{21} \neq 0$, we have

$$\text{tr}(bh) = \varepsilon(\text{tr } h + th_{21}).$$

The condition $\text{tr}(bh) \in \{\pm 2\}$ permits at most two values of t , hence at most four elements $b \in B$. Right multiplication by h is injective, so $Bh \subseteq A^2$ contains at least

$$|B| - 4 > \frac{1}{4}|A| - 1$$

regular semisimple elements. □

COROLLARY 6.6 (Bounded production of a regular element). *Under the same hypotheses, a regular semisimple element occurs in A^{16} . If $|A| > 4$, one may take A^2 . For a general nonsymmetric generating set the same conclusion holds in an absolute bounded power of its symmetric closure.*

PROOF. If $|A| > 4$, this is immediate from [Lemma 6.5](#). Suppose instead $2 \leq |A| \leq 4$. Because A generates $\text{SL}_2(K)$ and the latter has more than four elements in every case in which this corollary is used, A is not a subgroup. More generally, whenever a finite symmetric set B contains 1, is not the whole generated group, and satisfies $B^2 = B$, it is a subgroup: closure under multiplication follows from $B^2 = B$, while inverses and the identity are already present. Hence, as long as $|A^{2^j}| \leq 4$, strict inclusion

$$A^{2^j} \subsetneq A^{2^{j+1}}$$

must hold. Starting from a set of cardinality at least two, after at most three squarings we obtain a symmetric set

$$B = A^{2^j}, \quad j \leq 3,$$

with $|B| > 4$. It still contains 1 and generates the same group. Applying [Lemma 6.5](#) to B produces a regular semisimple element in $B^2 \subseteq A^{16}$. Thus an absolute bounded word length suffices in all cases. □

3. Escape from a union of linear subspaces

We now prove the quantitative escape lemma in a form adapted from the linear argument underlying Helfgott's proof.

LEMMA 6.7 (Linear-subspace escape). *Let a group Γ act linearly on a d -dimensional vector space W over a field. Let*

$$Z = Z_1 \cup \cdots \cup Z_n$$

be a union of proper linear subspaces. Let $A \subseteq \Gamma$ be finite, symmetric, contain 1, and suppose the $\langle A \rangle$ -orbit O of $w \in W$ is not contained in Z .

Then there exist integers $r = r(d, n)$ and $L = L(d, n)$ and elements

$$\gamma_1, \dots, \gamma_L \in A^r$$

such that for every $x \in O$, at least one $\gamma_i x$ lies outside Z . Consequently

$$|\{g \in A^{r+1} : gw \notin Z\}| \geq \frac{|A|}{L}.$$

PROOF. We induct on the lexicographically ordered pair

$$(e, s),$$

where e is the largest dimension of a component Z_i and s is the number of components having dimension e . The empty union is the base case.

Let Z^+ be the union of the components of maximal dimension e . If $O \cap Z^+ = \emptyset$, delete those components and apply induction to $Z \setminus Z^+$.

Otherwise choose $x_0 \in O \cap Z^+$. Since O is not contained in Z^+ , choose a shortest word

$$a_m \cdots a_1, \quad a_i \in A = A^{-1},$$

with $a_m \cdots a_1 x_0 \notin Z^+$. Put

$$x = a_{m-1} \cdots a_1 x_0 \in Z^+, \quad a = a_m;$$

then $ax \notin Z^+$. Choose a maximal-dimensional component Z_i containing x . For every maximal-dimensional Z_j , the intersection

$$Z_i \cap a^{-1}Z_j$$

has dimension $< e$: otherwise two e -dimensional linear subspaces with an e -dimensional intersection would be equal, so $Z_i = a^{-1}Z_j$, forcing $ax \in Z_j \subseteq Z^+$, a contradiction.

Consider

$$Z' = Z \cap a^{-1}Z.$$

Every component of Z' lies in an intersection $Z_r \cap a^{-1}Z_j$. No new component of dimension e can appear: an e -dimensional intersection must equal both of the participating e -dimensional linear subspaces. Moreover the particular component Z_i contributes no e -dimensional component at all by the preceding paragraph. Thus,

after discarding redundant components, either the maximal dimension drops or the number of maximal-dimensional components strictly decreases. Hence the lexicographic pair (e, s) decreases. The number of candidate intersections grows at most quadratically at each step; since the pair (e, s) can decrease only finitely many times, all resulting bounds depend only on d and the original n .

By induction there are $\eta_1, \dots, \eta_M \in A^{r'}$ such that for every $x \in \mathcal{O}$, some $\eta_i x \notin Z'$. For that i , either $\eta_i x \notin Z$ or $a\eta_i x \notin Z$. Thus the enlarged list

$$\eta_1, \dots, \eta_M, \quad a\eta_1, \dots, a\eta_M$$

works for Z . This proves the first assertion with bounds depending only on d, n .

For the second assertion apply the first part to each aw with $a \in A$. For every a choose an index $i(a)$ such that $\gamma_{i(a)}aw \notin Z$. There are only L possible indices, so some i occurs for at least $|A|/L$ values of a . The corresponding elements $\gamma_i a \in A^{r+1}$ are distinct and all move w outside Z . $\square$

REMARK 6.8. The proof is qualitative in the constants but fully uniform: r and L depend only on the ambient dimension and the number of forbidden subspaces, not on the field, the group, or $|A|$. That uniformity is what the growth theorem needs.

4. Escaping eigenline degeneracies

We apply Lemma 6.7 after extending scalars when necessary; this is essential for nonsplit tori.

COROLLARY 6.9 (A transverse matrix in bounded word length). *Let $q > 5$ be a prime power, let $A \subseteq \mathrm{SL}_2(\mathbb{F}_q)$ be finite, symmetric, contain 1, and generate $\mathrm{SL}_2(\mathbb{F}_q)$. Let L be either $\mathbb{F}_q$ or $\mathbb{F}_{q^2}$, and let $\ell_1, \ell_2 \subseteq L^2$ be two distinct lines which are either both $\mathbb{F}_q$ -rational or are exchanged by Frobenius. There is an absolute r and some $g \in A^r$ such that*

$$g\ell_i \neq \ell_j \quad (i, j \in \{1, 2\}).$$

Equivalently, in a basis spanning ℓ_1, ℓ_2 , all four entries of g are nonzero.

PROOF. Apply Lemma 6.7 over the vector space $\mathrm{Mat}_2(L)$ to the union of the four L -linear subspaces

$$Z_{ij} = \{h : h\ell_i \subseteq \ell_j\}.$$

It remains only to verify that the orbit $\mathrm{SL}_2(\mathbb{F}_q) \cdot I = \mathrm{SL}_2(\mathbb{F}_q)$ is not contained in their union.

If the two lines are $\mathbb{F}_q$ -rational, the set of matrices sending a fixed source line to a fixed target line is a coset of a Borel and has $q(q-1)$ elements. Hence the four forbidden sets contain at most $4q(q-1) < q(q^2-1) = |\mathrm{SL}_2(\mathbb{F}_q)|$ elements for $q > 3$.

If the lines are Frobenius-conjugate, a matrix over $\mathbb{F}_q$ that fixes one fixes the other as well; the stabilizer is a nonsplit torus of order $q+1$. A set of matrices sending one prescribed line to one prescribed conjugate line is either empty or a coset of

this torus, so the four forbidden sets have at most $4(q + 1) < q(q^2 - 1)$ elements for $q > 3$. Thus some element of $\mathrm{SL}_2(\mathbb{F}_q)$ avoids all four subspaces, and the escape lemma gives a bounded word in A with the same property. $\square$

A positive-proportion version will be useful for trace reconstruction.

COROLLARY 6.10 (Many words move both eigenlines). *Under the hypotheses of Corollary 6.9, there are absolute r and $c > 0$ such that*

$$|\{g \in A^r : g\ell_i \neq \ell_i \text{ for } i = 1, 2\}| \geq c|A|.$$

In a basis spanning ℓ_1, ℓ_2 , every such g has both off-diagonal entries nonzero.

PROOF. The two conditions $g\ell_i = \ell_i$ define two proper L -linear subspaces of $\mathrm{Mat}_2(L)$. Their union does not contain $\mathrm{SL}_2(\mathbb{F}_q)$ by the same counting argument as in Corollary 6.9. Apply the positive-proportion conclusion of Lemma 6.7. $\square$

5. A large simultaneously diagonalizable subset

We can now combine regularity and centralizer counting.

PROPOSITION 6.11 (Torus concentration from traces). *Let q be an odd prime power, let $A \subseteq \mathrm{SL}_2(\mathbb{F}_q)$ be finite, symmetric, contain 1, and generate $\mathrm{SL}_2(\mathbb{F}_q)$. Assume $|A| \geq 4$ and $|\mathrm{tr}(A)| \geq 2$. Then there exists a maximal torus T and a simultaneously diagonalizable set*

$$V \subseteq A^4 \cap T$$

satisfying

$$(19) \quad |V| \geq \frac{(|\mathrm{tr}(A)| - 2)(|A|/4 - 1)}{|A^6|}.$$

PROOF. Let

$$B = \{b \in A^2 : \mathrm{tr}(b) \neq \pm 2\}.$$

By Lemma 6.5,

$$|B| \geq |A|/4 - 1.$$

Because $A \subseteq A^2$ and the only trace values removed in forming B are ± 2 ,

$$|\mathrm{tr}(B)| \geq |\mathrm{tr}(A)| - 2.$$

Apply Corollary 6.4 to B . For some $g \in B$,

$$|C_G(g) \cap B^{-1}B| \geq \frac{|\mathrm{tr}(B)| |B|}{|BBB^{-1}|}.$$

Now

$$B^{-1}B \subseteq A^4, \quad BBB^{-1} \subseteq A^6.$$

Since g is regular semisimple, $C_G(g)$ is a maximal torus and all of its elements are simultaneously diagonalizable with g over the splitting field. Substituting the two lower bounds gives (19). $\square$

REMARK 6.12. The exponent bookkeeping is already visible. If bounded powers of A all have size comparable with $|A|$ and $|\operatorname{tr}(A)| \asymp |A|^{1/3}$, then Proposition 6.11 gives a torus slice of order $|A|^{1/3}$, exactly the dimension ratio $1/3$.

6. The dimension heuristic behind the centralizer count

The centralizer argument is easier to remember if one first performs the expected dimension calculation. Suppose a set A of size M behaves like a roughly three-dimensional cloud inside SL_2 . A regular conjugacy class has codimension one, so one expects about $M^{2/3}$ points of A to contribute distinct conjugacy directions. A regular centralizer is one-dimensional, so one expects a slice of size about $M^{1/3}$ in one of those centralizers.

The combinatorial identity behind this heuristic is the orbit–stabilizer relation for conjugation. For fixed regular g ,

$$|\{aga^{-1} : a \in A\}| \geq \frac{|A|}{|A^{-1}A \cap C_G(g)|}.$$

If many conjugates of g occur inside a bounded product of A , then either the conjugacy image is large or the denominator is large. Under a small-tripling hypothesis the conjugacy image cannot grow arbitrarily, so one is forced into the second alternative: a bounded word in A has a large intersection with $C_G(g)$, hence with a torus.

GUIDING IDEA 6.13 (Why the exponent is robust). The exact word length and powers of the tripling constant vary from one implementation to another, but the exponent $1/3$ is geometric. It is the ratio

$$\frac{\dim T}{\dim \operatorname{SL}_2} = \frac{1}{3}.$$

Larsen–Pink theory later replaces this hand-built centralizer count by the general estimate

$$|A \cap V| \lesssim K^{O(1)} |A|^{\dim V / \dim G}.$$

7. A precise cubic-growth lemma

We now prove the transverse torus-product estimate by an explicit calculation; this is the genuinely cubic bounded-fiber step foreshadowed by the dimension heuristic.

LEMMA 6.14 (Helfgott’s transverse torus product estimate). *Let K be a field, let V be a finite set of diagonal matrices*

$$V \subseteq \{\operatorname{diag}(r, r^{-1}) : r \in K^\times\},$$

and let

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \operatorname{SL}_2(K)$$

have $abcd \neq 0$. Then

$$|VgVg^{-1}V| \geq \frac{1}{2} \left(\frac{|V|}{4} - 5 \right) |V|^2.$$

In particular $|VgVg^{-1}V| \gg |V|^3$ for $|V|$ larger than an absolute constant.

PROOF. For $r \in K^\times$ set $t(r) = \text{diag}(r, r^{-1})$. Direct multiplication gives

$$gt(r)g^{-1} = \begin{pmatrix} rad - r^{-1}bc & (r^{-1} - r)ab \\ (r - r^{-1})cd & r^{-1}ad - rbc \end{pmatrix}.$$

The product of its off-diagonal entries is

$$-(r - r^{-1})^2 abcd.$$

For a prescribed value of this quantity, r satisfies a polynomial equation of degree at most four, so there are at most four possible r . Hence the set of values of $h_{12}h_{21}$ obtained from $h \in gVg^{-1}$ has size at least $|V|/4$.

At most four values of r can make one of the two diagonal entries vanish simultaneously with the nondegenerate off-diagonal condition. Therefore there is a subset $U \subseteq gVg^{-1}$ for which all four entries needed below are nonzero and

$$|\{h_{12}h_{21} : h \in U\}| \geq |V|/4 - 5.$$

Choose one h for each distinct value of $h_{12}h_{21}$. For s, t corresponding to elements of V , consider

$$t(s)ht(t) = \begin{pmatrix} sth_{11} & st^{-1}h_{12} \\ s^{-1}th_{21} & s^{-1}t^{-1}h_{22} \end{pmatrix}.$$

For fixed h , write the resulting matrix as $M = (M_{ij})$. Because all four entries of h are nonzero,

$$\frac{M_{11}}{M_{21}} = s^2 \frac{h_{11}}{h_{21}}, \quad \frac{M_{11}}{M_{12}} = t^2 \frac{h_{11}}{h_{12}}.$$

Thus M determines s^2 and t^2 . It also determines

$$st = \frac{M_{11}}{h_{11}}.$$

Among the at most four choices of square roots of s^2, t^2 , fixing the product st leaves at most two pairs (s, t) . Distinct chosen h have distinct invariants $h_{12}h_{21}$, which are preserved by left and right diagonal multiplication, so their images are disjoint. We therefore obtain at least

$$\frac{1}{2} \left(\frac{|V|}{4} - 5 \right) |V|^2$$

distinct elements of $VgVg^{-1}V$. □

8. Trace size controls group size

PROPOSITION 6.15 (From traces to a bounded product set). *Let $q > 5$ be a prime power and let $A \subseteq \mathrm{SL}_2(\mathbb{F}_q)$ be finite, symmetric, contain 1, and generate $\mathrm{SL}_2(\mathbb{F}_q)$. Assume $|A| \geq 8$ and $|\mathrm{tr}(A)| \geq 3$. Then for an absolute integer m ,*

$$|A^m| \gg \left(\frac{(|\mathrm{tr}(A)| - 2)|A|}{|A^6|} \right)^3,$$

provided the quantity in parentheses is larger than an absolute constant.

PROOF. By Proposition 6.11, there is $V \subseteq A^4$ inside a maximal torus T with

$$|V| \gg \frac{(|\mathrm{tr}(A)| - 2)|A|}{|A^6|}.$$

By Corollary 6.9, there is $g \in A^r$ transverse to the two eigenlines of T , so in a diagonalizing basis all four entries of g are nonzero. Apply Lemma 6.14. The set

$$VgVg^{-1}V$$

lies in A^{12+2r} and has cardinality $\gg |V|^3$. Absorb $12 + 2r$ into the absolute integer m . $\square$

9. A fully proved cube-root trace bound

We now prove the cube-root trace lower bound that the transfer mechanism requires.

PROPOSITION 6.16 (Cube-root trace bound). *There are absolute constants $m \geq 1$ and $c > 0$ such that, for every prime power $q > 5$ and every finite symmetric $A \subseteq \mathrm{SL}_2(\mathbb{F}_q)$ containing 1 and generating $\mathrm{SL}_2(\mathbb{F}_q)$,*

$$|\mathrm{tr}(A^m)| \geq c|A|^{1/3}.$$

PROOF. First find a regular semisimple h in a bounded power A^{m_1} by Corollary 6.6. Its two eigenlines are either defined over $\mathbb{F}_q$ or are exchanged by Frobenius over $\mathbb{F}_{q^2}$. Work over that splitting field and choose an eigenbasis of h , with eigenvalues r, r^{-1} and $r \neq r^{-1}$.

By Corollary 6.10, there is a set

$$X \subseteq A^{m_2}$$

with $|X| \geq c_0|A|$ such that every $x \in X$ has both off-diagonal entries nonzero in this basis. Put

$$D = \{(x_{11}, x_{22}) : x \in X\}.$$

We claim

$$(20) \quad |\mathrm{tr}(XX^{-1})| \geq \frac{|X|}{2|D|}.$$

Indeed, choose a fiber $X_{a,d} = \{x \in X : (x_{11}, x_{22}) = (a, d)\}$ of maximal size, so $|X_{a,d}| \geq |X|/|D|$. Fix

$$x = \begin{pmatrix} a & b_0 \\ c_0 & d \end{pmatrix} \in X_{a,d}.$$

The construction of X gives $b_0 c_0 \neq 0$. Since $\det x = 1$, put

$$k = ad - 1 = b_0 c_0 \neq 0.$$

Every $y \in X_{a,d}$ has the form

$$y = \begin{pmatrix} a & k/t \\ t & d \end{pmatrix}, \quad t = y_{21} \in K^\times.$$

As $\det y = 1$,

$$y^{-1} = \begin{pmatrix} d & -k/t \\ -t & a \end{pmatrix},$$

and therefore

$$\mathrm{tr}(xy^{-1}) = 2ad - b_0 t - c_0 k/t.$$

For a prescribed trace value τ , multiplication by the nonzero parameter t gives the quadratic equation

$$b_0 t^2 - (2ad - \tau)t + c_0 k = 0.$$

Its leading coefficient is nonzero, so it has at most two roots over the splitting field. Hence a given trace value has at most two preimages in the fiber, proving (20).

Now partition D by the ordinary trace $a + d$. Some trace fiber D_t has size at least

$$|D|/|\mathrm{tr}(X)|.$$

For $(a, d) \in D_t$, the trace of hx is

$$ra + r^{-1}d = (r - r^{-1})a + r^{-1}t.$$

Since $r \neq r^{-1}$, this is injective as a function of a on D_t . Therefore

$$(21) \quad |\mathrm{tr}(hX)| \geq \frac{|D|}{|\mathrm{tr}(X)|}.$$

Multiply (20) and (21):

$$|\mathrm{tr}(XX^{-1})| |\mathrm{tr}(hX)| |\mathrm{tr}(X)| \geq \frac{|X|}{2}.$$

All three trace sets are contained in $\mathrm{tr}(A^m)$ for one absolute m after increasing m to dominate $2m_2$, $m_1 + m_2$, and m_2 . Thus

$$|\mathrm{tr}(A^m)|^3 \geq |X|/2 \geq c_0 |A|/2.$$

Taking cube roots proves the result. $\square$

CHECKPOINT 6.17. We have now proved, without using algebraic geometry, the two rank-one dimension estimates that later become Larsen–Pink inequalities:

- (i) some bounded power of a generating set has at least $\gg |A|^{1/3}$ trace values;

- (ii) a torus slice much larger than the $|A|^{1/3}$ scale forces cubic growth through a transverse multiplication map.

The only substantial arithmetic input in the small-set proof is finite-field sum–product: its power-range form was proved in Volume III, while Chapter 4 proves the additional constant-threshold small-set supplement needed by the rich-torus argument.

CHAPTER 7

Small Tripling, Pivots, and Propagation of Torus Structure

The growth theorem is proved by contradiction: assume A^3 is only slightly larger than A and propagate that hypothesis through every bounded product appearing later. Once this is done, a large intersection with one torus cannot remain an isolated accident. Conjugation and controlled fibers transport the same scale of concentration to other tori. This is the rank-one form of the pivot principle that later becomes the “involved torus” argument of Breuillard–Green–Tao and Pyber–Szabó.

GUIDING IDEA 7.1 (A pivot is a transversality certificate). Suppose T is a rich torus. For an element a , the map built from multiplying an element of T by a conjugate or translate involving a is useful only if its fibers are small. A *pivot* is precisely a choice of a for which this transversality holds. If many pivots exist, counting the image forces growth. If pivots systematically fail, the failures imply that many conjugates of T are themselves involved with the approximate group. That abundance of involved tori creates a different contradiction.

Thus the pivot argument is a dichotomy inside a dichotomy: either transversality gives immediate growth, or failure of transversality produces so much algebraic structure that normality becomes unavoidable. This is the rank-one ancestor of the “rich tori plus conjugation” argument in the general product theorem.

1. From small tripling to an approximate group

The original product theorem is naturally stated for an arbitrary generating set, while the matrix arguments are cleaner for symmetric sets containing the identity. The following theorem is the rigorous bridge between those two formulations.

THEOREM 7.2 (Small tripling controls all signed bounded products). *Let A be a finite nonempty subset of a group and put*

$$K = \frac{|A^3|}{|A|}.$$

Then there is a K^C -approximate subgroup H with

$$A \subseteq H, \quad |H| \leq K^C |A|$$

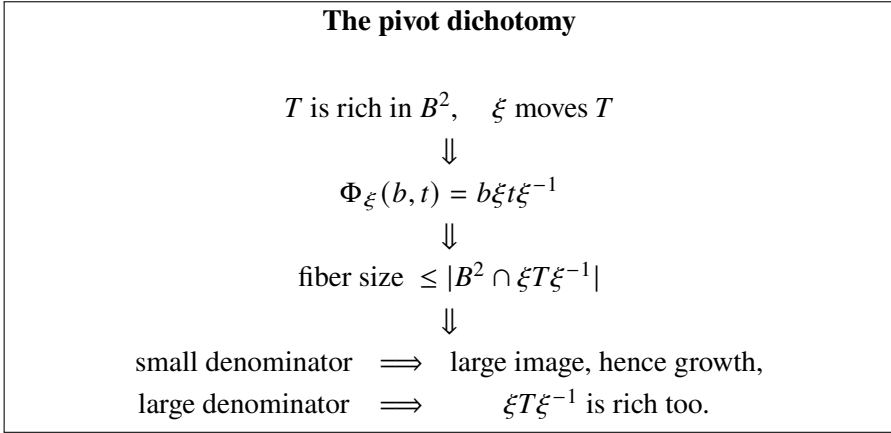


FIGURE 1. Why a pivot either creates growth or propagates torus structure. The proof is the fiber count in [Lemma 7.8](#).

for an absolute constant C . Consequently, if

$$A_\pm = A \cup A^{-1} \cup \{1\},$$

then for every fixed $m \geq 1$,

$$|A_\pm^m| \leq K^{C_m} |A|,$$

where C_m depends only on m .

PROOF. Volume II proved the noncommutative small-tripling theorem from Ruzsa calculus, covering, and the full Balog–Szemerédi–Gowers argument: from

$$|A^3| = K|A|$$

one obtains a K^C -approximate subgroup H satisfying

$$A \subseteq H, \quad |H| \leq K^C |A|.$$

This is an earlier proved result of the series, not an external input at this point. Since H is symmetric and contains the identity,

$$A_\pm \subseteq H.$$

Choose X with $|X| \leq K^C$ and $H^2 \subseteq XH$. Induction gives

$$H^m \subseteq X^{m-1}H,$$

and therefore

$$|A_\pm^m| \leq |H^m| \leq |X|^{m-1} |H| \leq K^{C_m} |A|.$$

All constants are independent of the ambient group. □

REMARK 7.3 (Historical note). The small-tripling theorem used here originated in Tao’s noncommutative product-set theory. Its proof, including the polynomial losses needed above, is proved in Volume II; the citation is therefore historical rather than a proof dependency.

REMARK 7.4. This theorem is the formal reason that the large word exponents in Helfgott’s original proof do not matter. Under a contradiction hypothesis on $|A^3|$, every fixed signed word set in A has cardinality at most a fixed power of the same tripling factor times $|A|$. Thus a contradiction of the form

$$K^C \geq |A|^\eta$$

still yields genuine power growth for the original, possibly nonsymmetric, triple product A^3 .

COROLLARY 7.5 (Bounded signed-word growth transfers back to A^3). *Fix $m \geq 1$ and $\eta > 0$. Suppose*

$$|A_\pm^m| \geq c|A|^{1+\eta}.$$

Then

$$|A^3| \geq c'|A|^{1+\eta'},$$

where $c', \eta' > 0$ depend only on c, η, m .

PROOF. Put $K = |A^3|/|A|$. By Theorem 7.2,

$$|A_\pm^m| \leq K^{C_m}|A|.$$

Comparison with the assumed lower bound gives

$$K \geq c^{1/C_m}|A|^{\eta/C_m},$$

which is the required conclusion. $\square$

2. Rich tori

DEFINITION 7.6. Let $B \subseteq \mathrm{SL}_2(\mathbb{F}_p)$ be symmetric with $1 \in B$. A maximal torus T is called ρ -rich for B if

$$|B^2 \cap T| \geq \rho.$$

It is called *involved* with B if $B^2 \cap T$ contains a regular semisimple element.

The two notions serve different purposes. “Rich” is quantitative; “involved” is stable under recognizing a torus from one regular element.

LEMMA 7.7 (Regular elements identify their tori). *If x is regular semisimple in $\mathrm{SL}_2(\overline{\mathbb{F}}_p)$, then the unique maximal torus containing x is $C_G(x)$. Consequently, if $x \in B^2$ is regular semisimple and $a \in G$, then*

$$a^{-1}xa$$

is regular semisimple and belongs to the unique torus $a^{-1}Ta$.

PROOF. This is Proposition 1.13, and regular semisimplicity is preserved by conjugation. $\square$

3. The pivot map

Fix a maximal torus T and $\xi \in G$. Consider

$$\Phi_\xi : B \times (B^2 \cap T) \longrightarrow G, \quad (b, t) \longmapsto b\xi t\xi^{-1}.$$

LEMMA 7.8 (Pivot fiber bound). *Every fiber of Φ_ξ has cardinality at most*

$$|B^2 \cap \xi T \xi^{-1}|.$$

Therefore

$$|B\xi(B^2 \cap T)\xi^{-1}| \geq \frac{|B| |B^2 \cap T|}{|B^2 \cap \xi T \xi^{-1}|}.$$

PROOF. If

$$b_1 \xi t_1 \xi^{-1} = b_2 \xi t_2 \xi^{-1},$$

then

$$b_2^{-1} b_1 = \xi t_2 t_1^{-1} \xi^{-1} \in B^2 \cap \xi T \xi^{-1}.$$

Once one point in the fiber is fixed, the element $b_2^{-1} b_1$ determines the other point, so the fiber has the stated size bound. Divide the size of the domain by the maximal fiber size. $\square$

DEFINITION 7.9. Relative to B and T , call ξ a *pivot* if the conjugate torus $\xi T \xi^{-1}$ has only a small intersection with B^2 compared with T , so that Lemma 7.8 forces the image to be substantially larger than B .

The definition is intentionally qualitative because the threshold depends on the application. The dichotomy is what matters:

PROPOSITION 7.10 (Pivot dichotomy). *Let B be a K -approximate subgroup of $G = \mathrm{SL}_2(\mathbb{F}_p)$ and let T satisfy*

$$|B^2 \cap T| \geq K^{-C_0} |B|^{1/3}.$$

Then for every $\xi \in B$ at least one of the following holds:

- (i) *some bounded power B^m has size at least $K^{-C} |B|^{1+\eta}$ for an absolute $\eta > 0$;*
- (ii) *the conjugate torus $\xi T \xi^{-1}$ is also rich:*

$$|B^2 \cap \xi T \xi^{-1}| \geq K^{-C} |B|^{1/3}.$$

Here C, m are absolute once C_0 is fixed.

PROOF. Because $\xi \in B$ and B is a K -approximate subgroup, the image of Φ_ξ is contained in a bounded power B^m , whose size is at most $K^{O(1)}|B|$ unless alternative (i) already holds. Thus Lemma 7.8 gives

$$K^{O(1)}|B| \geq \frac{|B| K^{-C_0} |B|^{1/3}}{|B^2 \cap \xi T \xi^{-1}|}.$$

Rearranging gives alternative (ii). $\square$

GUIDING IDEA 7.11. A pivot is useful because it creates growth. If there are no useful pivots, richness propagates to every conjugate torus reached by B . That apparent failure is even more rigid: it produces a large conjugation-invariant family of tori. In a simple group such a family cannot be too small, and counting its regular elements eventually forces B to be a substantial fraction of the whole group.

4. Distinct tori overlap only in exceptional elements

LEMMA 7.12 (Intersection of maximal tori). *Let T_1, T_2 be distinct maximal tori of $\mathrm{SL}_2(\overline{\mathbb{F}}_p)$. Then*

$$T_1 \cap T_2 \subseteq \{\pm I\}.$$

In particular, the sets of regular semisimple elements of distinct maximal tori are disjoint.

PROOF. If $x \in T_1 \cap T_2$ is regular semisimple, then Lemma 7.7 says that x belongs to a unique maximal torus, forcing $T_1 = T_2$. Hence every element in the intersection is nonregular. A torus has no nontrivial unipotent elements, so in characteristic different from two the only possibilities are the central elements $\pm I$. $\square$

EXAMPLE 7.13 (A pivot count in $\mathrm{SL}_2(\mathbb{F}_7)$). Let

$$T = \{\mathrm{diag}(\lambda, \lambda^{-1}) : \lambda \in \mathbb{F}_7^\times\}, \quad |T| = 6,$$

and take

$$\xi = \begin{pmatrix} 2 & 2 \\ 5 & 2 \end{pmatrix}.$$

Its determinant is $4 - 10 \equiv 1 \pmod{7}$, and all four entries are nonzero, so ξ does not normalize the diagonal torus. Hence T and $\xi T \xi^{-1}$ are distinct maximal tori. By Lemma 7.12,

$$T \cap \xi T \xi^{-1} = \{\pm I\},$$

so this intersection has size 2.

Apply Lemma 7.8 with $B = T$. Since $B^2 = T$,

$$|T \xi T \xi^{-1}| \geq \frac{|T| |T|}{|T \cap \xi T \xi^{-1}|} = \frac{6 \cdot 6}{2} = 18.$$

Thus two transverse torus directions already create at least three times as many group elements as one torus. This is the finite $p = 7$ shadow of the dimension heuristic

$1 + 1 = 2$: a one-dimensional torus multiplied against a transverse conjugate begins to fill a two-dimensional family.

5. What the local pivot argument proves

The bounded-fiber statement already contains the rank-one pivot principle needed for intuition. Let

$$R_T = |B^2 \cap T|, \quad R_{\xi T \xi^{-1}} = |B^2 \cap \xi T \xi^{-1}|.$$

Then [Lemma 7.8](#) gives the exact inequality

$$(22) \quad |B\xi(B^2 \cap T)\xi^{-1}| \geq \frac{|B|R_T}{R_{\xi T \xi^{-1}}}.$$

If T is rich while the conjugate torus has a much smaller intersection with B^2 , the right-hand side is substantially larger than $|B|$ and a bounded word set grows. Conversely, failure of this growth forces the conjugate torus to be rich as well. This is a completely proved local dichotomy; no classification or algebraic geometry is hidden in it.

PROPOSITION 7.14 (Local pivot dichotomy). *Let B be a symmetric K -approximate subgroup of $\mathrm{SL}_2(\mathbb{F}_p)$ containing 1, let T be a maximal torus, and let $\xi \in B$. If*

$$|B^2 \cap T| \geq R, \quad |B^6| \leq M|B|,$$

then

$$|B^2 \cap \xi T \xi^{-1}| \geq R/M.$$

Equivalently, if the latter intersection is smaller than R/M , then $|B^6| > M|B|$.

PROOF. The image in [Lemma 7.8](#) is contained in B^6 : indeed

$$B\xi(B^2 \cap T)\xi^{-1} \subseteq B B B^2 B \subseteq B^5 \subseteq B^6.$$

Thus

$$M|B| \geq |B^6| \geq \frac{|B|R}{|B^2 \cap \xi T \xi^{-1}|}.$$

Cancel $|B|$ and rearrange. □

GUIDING IDEA 7.15 (Why the global involved-torus argument is postponed). Iterating the preceding local statement over an arbitrarily long word is not legitimate with a fixed loss: the factor M would accumulate with the word length. A genuine global “involved tori” theorem therefore needs an additional algebraic-geometric mechanism that makes the family stable without unbounded quantitative degradation. That mechanism is precisely what the later Larsen–Pink theory supplies. We do not state a global no-pivot theorem here and then conceal that missing step inside the phrase “iterate over the generators.”

CHECKPOINT 7.16. For the classical Helfgott proof used in the next chapter, the only result needed from the present chapter is the signed-word propagation theorem at the beginning. The pivot material is a fully proved rank-one laboratory showing the principle that will later be upgraded by algebraic geometry; it is not used as a black box in the concrete product-growth proof.

Part 4

Helfgott Growth

CHAPTER 8

Helfgott Growth: the Small and Medium Regimes

We now assemble the rank-one machinery into the decisive theorem: a generating set in $\mathrm{SL}_2(\mathbb{F}_p)$ of size below $p^{3-\delta}$ must grow by a fixed power under triple multiplication. The proof is organized as a contradiction from small tripling. The small-tripling theorem proved in Volume II first legitimizes passage from an arbitrary set to symmetric signed word sets; cube-root trace production then gives a torus slice of the correct one-dimensional scale, a transverse conjugate feeds that slice into the uniform trace-shaped sum-product theorem, and a final three-dimensional multiplication estimate converts the arithmetic gain back into group growth.

GUIDING IDEA 8.1 (How to read the Helfgott proof). The theorem is not one long estimate. It is a sequence of changes of language:

$$\begin{aligned}
 A \subset \mathrm{SL}_2(\mathbb{F}_p) &\longrightarrow \text{regular elements and centralizers} \\
 &\longrightarrow \text{a rich torus slice} \\
 &\longrightarrow \text{transverse trace expressions} \\
 &\longrightarrow \text{finite-field sum-product} \\
 &\longrightarrow |A^3| > |A|^{1+\delta}.
 \end{aligned}$$

Small tripling is assumed only to derive enough control to carry information through these changes of language. At the end, the sum-product theorem says that the resulting field set cannot remain simultaneously additive and multiplicative. The contradiction therefore occurs in the field, but the reason the field set exists at the right scale is entirely group-theoretic.

1. Statement of the theorem

THEOREM 8.2 (Helfgott growth below the top scale). *For every $\delta \in (0, 3)$ there exist constants $c = c(\delta) > 0$ and $\varepsilon = \varepsilon(\delta) > 0$ such that the following holds for every prime $p \geq 5$.*

Let $E \subseteq G = \mathrm{SL}_2(\mathbb{F}_p)$ be a finite set generating G . If

$$|E| \leq p^{3-\delta},$$

then

$$|E^3| \geq c|E|^{1+\varepsilon}.$$

REMARK 8.3. We use E for the original, possibly nonsymmetric, set and reserve S for its symmetric closure. This distinction prevents a common logical shortcut: smallness of E^3 does not literally say that the symmetric closure has the same tripling constant. What is true, and sufficient, is the signed-word propagation theorem [Theorem 7.2](#).

PROOF. 2. Step 0: the legitimate symmetrization

The case $p = 5$ is absorbed into the constant $c(\delta)$, so assume $p > 5$. Put

$$N = |E|, \quad K = \frac{|E^3|}{|E|}, \quad S = E \cup E^{-1} \cup \{1\}.$$

Then

$$N \leq |S| \leq 2N + 1.$$

By [Theorem 7.2](#), for every fixed m ,

$$(23) \quad |S^m| \leq K^{C_m} N.$$

Thus the rest of the proof may work entirely with the symmetric generating set S , while the contradiction parameter remains the tripling factor of the original set E .

GUIDING IDEA 8.4. Assume K is small. Equation (23) says that every bounded signed word set in E is still only $K^{O(1)}N$. The rank-one geometry will force one such bounded word set to contain $K^{-O(1)}N^{1+\alpha}$ elements. Comparing the two estimates forces $K \geq N^\eta$.

3. Step 1: produce many traces

Apply the cube-root trace bound [Proposition 6.16](#) to S . There are absolute m_0 and $c_0 > 0$ such that

$$(24) \quad |\operatorname{tr}(S^{m_0})| \geq c_0|S|^{1/3} \geq c_0N^{1/3}.$$

Set

$$B = S^{m_0}.$$

The set B is symmetric, contains the identity, and generates G . Moreover, by (23),

$$(25) \quad |B^r| \leq K^{C'_r} N \quad \text{for every fixed } r.$$

In particular,

$$|B| \geq |S| \geq N.$$

4. Step 2: extract a torus slice of the correct dimension

Apply [Proposition 6.11](#) to B . There is a maximal torus T and a simultaneously diagonalizable set

$$V_0 \subseteq B^4 \cap T$$

with

$$(26) \quad |V_0| \geq \frac{(|\operatorname{tr}(B)| - 2)(|B|/4 - 1)}{|B^6|}$$

$$(27) \quad \gg K^{-C_1} N^{1/3},$$

provided N is larger than an absolute constant. The bounded values of N are absorbed into the final constant.

Choose a subset $V \subseteq V_0$ with

$$(28) \quad c_1 K^{-C_1} N^{1/3} \leq |V| \leq 2c_1 K^{-C_1} N^{1/3},$$

whenever the lower endpoint is at least 1; if it is smaller than 1, then $K^{C_1} \gg N^{1/3}$ and we already have the desired power lower bound for K .

Since $N \leq p^{3-\delta}$,

$$N^{1/3} \leq p^{1-\delta/3}.$$

After decreasing c_1 if necessary,

$$(29) \quad |V| < p^{1-\delta/3}.$$

This is precisely the upper range required by the characteristic-scale arithmetic theorem.

REMARK 8.5 (Split and nonsplit tori). If T is split, diagonalize over $\mathbb{F}_p$. If it is nonsplit, diagonalize over $\mathbb{F}_{p^2}$. In either case the eigenvalue parameter set X has $|X| = |V|$, and (29) is measured against the characteristic p , exactly as in the uniform trace-shaped theorem of Chapter 4.

5. Step 3: obtain a transverse conjugate

Let v_1, v_2 be the two eigenlines of T over its splitting field. By [Corollary 6.9](#), there is an absolute m_1 and

$$g \in S^{m_1}$$

such that

$$gv_i \neq v_j \quad (i, j \in \{1, 2\}).$$

In an eigenbasis of T , all four entries of g are therefore nonzero.

Parameterize V by an eigenvalue set X in the splitting field:

$$t(x) = \operatorname{diag}(x, x^{-1}), \quad x \in X, \quad |X| = |V|.$$

The explicit matrix calculation [Equation \(18\)](#) gives nonzero coefficients a_1, a_2 such that

$$(30) \quad \text{tr}(t(x)gt(y)g^{-1}) = a_1(xy + x^{-1}y^{-1}) + a_2(x^{-1}y + xy^{-1}).$$

The nonvanishing of a_1a_2 is exactly what the four eigenline-avoidance conditions guarantee.

6. Step 4: sum-product forces trace growth

Apply [Proposition 5.10](#) with scale parameter $\delta/3$. If $|V|$ is above its constant lower threshold, then for some $\alpha = \alpha(\delta) > 0$,

$$(31) \quad |\text{tr}(V_{\pm}^{20} g V_{\pm}^{20} g^{-1})| \gg_{\delta} |V|^{1+\alpha}.$$

The matrix set in (31) is contained in S^M for an absolute M . Hence

$$(32) \quad |\text{tr}(S^M)| \gg K^{-C_2} N^{(1+\alpha)/3}.$$

If $|V|$ is below the constant threshold $C(\delta)$, then (28) instead yields

$$N^{1/3} \ll_{\delta} K^{C_1},$$

which already gives $K \gg_{\delta} N^{1/(3C_1)}$. Thus no case is lost by invoking the arithmetic theorem only above its stated lower threshold.

7. Step 5: convert trace growth into group growth

Set

$$D = S^M.$$

Then D is symmetric, contains 1, and generates G . Apply [Proposition 6.15](#):

$$|D^{m_2}| \gg \left(\frac{(|\text{tr}(D)| - 2)|D|}{|D^6|} \right)^3$$

provided the quantity in parentheses is above the absolute threshold in that proposition. If it is not, (32) and the bounds below again already force a power lower bound for K .

By (23),

$$|D| \geq N, \quad |D^6| \leq K^{C_3} N.$$

Together with (32), this yields

$$(33) \quad |S^{Mm_2}| = |D^{m_2}|$$

$$(34) \quad \gg K^{-C_4} |\text{tr}(S^M)|^3$$

$$(35) \quad \gg K^{-C_5} N^{1+\alpha}.$$

On the other hand, signed-word propagation gives

$$(36) \quad |S^{Mm_2}| \leq K^{C_6} N.$$

Comparing (35) and (36),

$$K^{C_5+C_6} \gg N^{\alpha}.$$

Therefore

$$(37) \quad K \gg N^{\alpha/(C_5+C_6)}.$$

Since $K = |E^3|/|E|$, this proves Theorem 8.2. $\square$

8. A dependency ledger for the rank-one proof

Before tracking numerical exponents, it is useful to record which statement feeds which. Under the contradictory assumption

$$|A^3| \leq K|A|,$$

with K intended to be a small power of $|A|$, the proof uses the following chain.

- (1) **Small-tripling control.** The noncommutative small-tripling theorem proved in Volume II controls every bounded signed word in A by $K^{O(1)}|A|$. This legitimizes the later use of symmetric bounded word sets without replacing the original tripling parameter.
- (2) **Regular production and trace count.** Bounded words contain many non-unipotent, hence regular semisimple, elements. Trace fibers are controlled strongly enough to obtain many distinct regular conjugacy directions.
- (3) **Centralizer extraction.** Conjugacy counting forces one regular centralizer to meet a bounded word in A in about the cube-root scale, up to $K^{O(1)}$ losses.
- (4) **Transverse escape.** A bounded word produces an element moving the two eigenlines of the selected torus into generic position. Otherwise the generated group would preserve a forbidden finite configuration in $\mathbb{P}^1$.
- (5) **Arithmetic growth.** Eigenvalue coordinates on the torus and the transverse trace polynomial produce a field set to which the trace-shaped sum-product input applies.
- (6) **Return to matrices.** Distinct trace values give distinct matrices, and controlled trace fibers ensure that the gain survives the earlier quotienting and centralizer losses.

The contradiction is obtained by choosing the growth exponent $\delta > 0$ small enough that every factor $K^{O(1)}$ can be absorbed into $|A|^{c\delta}$ while the sum-product theorem contributes a fixed positive power. The proof therefore needs no optimization of the hidden constants, but it does need every occurrence of $O(1)$ to be uniform in p and in A .

WARNING 8.6 (Where circularity can enter). One must not use the desired product theorem itself to justify control of the bounded words occurring before the contradiction. That control comes from the general small-tripling/Ruzsa theory of Volume II, which is valid in every group. The special SL_2 growth theorem is used only at the final contradiction stage.

9. Tracking the growth exponent

Ignoring polynomial losses in K , the proof has the following dimension profile:

$$|\operatorname{tr}(S^{m_0})| \gtrsim N^{1/3}.$$

This produces a torus slice

$$|V| \gtrsim N^{1/3}.$$

Sum–product improves the one-dimensional exponent:

$$|\operatorname{tr}(S^M)| \gtrsim |V|^{1+\alpha} \gtrsim N^{(1+\alpha)/3}.$$

The transverse three-dimensional multiplication estimate then cubes that gain:

$$|S^{M'}| \gtrsim |\operatorname{tr}(S^M)|^3 \gtrsim N^{1+\alpha}.$$

Thus the engine is the dimension cycle

$$3 \longrightarrow 1 \xrightarrow{\text{sum-product}} 1 + \alpha \longrightarrow 3 + 3\alpha.$$

The signed-word propagation theorem merely guarantees that, under the assumed lack of triple-product growth, every fixed word set on the left-hand side remains at the $K^{O(1)}N$ scale.

10. What is genuinely rank one?

Three ingredients use special properties of SL_2 :

- (i) trace is essentially a complete invariant of regular semisimple conjugacy classes;
- (ii) a maximal torus is one-dimensional and its Weyl group has order two;
- (iii) the explicit transverse trace polynomial reduces the growth problem to a one-dimensional finite-field sum–product statement; see (30).

The logic of the proof is more general. In higher rank, trace is replaced by regular semisimple conjugacy data, torus intersections are controlled by dimension, and the final contradiction comes from algebraic-geometric intersection estimates rather than an explicit two-variable trace polynomial.

11. Comparison with the original word-exponent proof

Helfgott’s original proof keeps explicit track of numerous bounded word powers. The present presentation packages those fixed exponents using the small-tripling theorem proved in Volume II. This changes bookkeeping, not mathematics: under the contradiction hypothesis $|E^3| = K|E|$, [Theorem 7.2](#) controls every fixed signed word set by $K^{O(1)}|E|$. The advantage is that the only exponent that remains conceptually visible is the genuine dimension gain supplied by sum–product.

CHECKPOINT 8.7. Every hypothesis of [Theorem 8.2](#) now has a visible role:

- generation excludes subgroup traps and enables eigenline escape;

- the bound $|E| \leq p^{3-\delta}$ puts the torus eigenvalue set below $p^{1-\delta/3}$;
- the finite-field input gives a uniform one-dimensional power gain;
- small tripling controls all bounded signed words, legitimizing symmetrization;
- transversality restores the ambient three-dimensional scale and closes the contradiction.

CHAPTER 9

Large Sets, Completion, and the Full Rank-One Product Theorem

Sum-product is needed only while the set is substantially below the size of the group. Once a subset of $\mathrm{SL}_2(\mathbb{F}_p)$ reaches the $p^{8/3}$ scale, quasirandomness finishes the argument in a bounded number of multiplications. We prove this completion from representation theory, compare it with Helfgott's original Borel-based completion, and then deduce the full product theorem and the polylogarithmic diameter bound.

1. Quasirandomness of $\mathrm{SL}_2(\mathbb{F}_p)$

Let $G = \mathrm{SL}_2(\mathbb{F}_p)$ with $p \geq 5$. We need only the following rank-one representation-theoretic estimate.

PROPOSITION 9.1 (Minimal representation degree). *Every nontrivial complex irreducible representation of G has dimension at least*

$$\frac{p-1}{2}.$$

Thus G is D -quasirandom with $D = (p-1)/2$.

PROOF. Let

$$U = \left\{ u(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in \mathbb{F}_p \right\} \cong (\mathbb{F}_p, +).$$

For a nontrivial irreducible representation $\rho : G \rightarrow \mathrm{GL}(V)$, the restriction $\rho|_U$ decomposes into characters

$$\psi_a(x) = e^{2\pi i ax/p}, \quad a \in \mathbb{F}_p.$$

If only the trivial character occurred, then U would act trivially. The normal closure of U is all of G : conjugation by the Weyl element produces the lower unipotent subgroup, and the Gauss decomposition proved in Chapter 2 shows that the upper and lower unipotents generate G . Hence if U acted trivially, the entire group would act trivially. Hence some nonzero a occurs.

Now conjugate by

$$h(t) = \mathrm{diag}(t, t^{-1}), \quad t \in \mathbb{F}_p^\times.$$

We have

$$h(t)u(x)h(t)^{-1} = u(t^2x).$$

Therefore, if the weight ψ_a occurs, so does every ψ_{at^2} . The nonzero squares form a subgroup of $\mathbb{F}_p^\times$ of index two, so the orbit of a has $(p-1)/2$ distinct characters. Distinct U -characters have linearly independent weight spaces. Hence

$$\dim V \geq (p-1)/2.$$

□

GUIDING IDEA 9.2 (Why a separate large-set argument is necessary). Sum–product is most effective in an intermediate cardinality range. Once A is already very large, it is wasteful to keep extracting tori and trace sets. Quasirandom product mixing gives a different completion principle: sufficiently dense subsets of a highly quasirandom group must multiply onto the whole group after a bounded number of steps.

This division into small/medium and large regimes is not a defect of Helfgott’s argument. It is a recurring feature of growth theory. Algebraic-combinatorial methods create a power gain away from the ambient scale; representation theory takes over near the ambient scale and turns density into covering.

2. The quasirandom product lemma

We prove the exact finite-group statement needed for completion.

THEOREM 9.3 (Gowers product lemma). *Let G be a finite D -quasirandom group. If $X, Y, Z \subseteq G$ satisfy*

$$|X||Y||Z| > \frac{|G|^3}{D},$$

then

$$XYZ = G.$$

PROOF. Fix $g \in G$. It is enough to prove that there exist $x \in X, y \in Y, z \in Z$ with $xyz = g$. Equivalently,

$$\langle 1_X * 1_Y, 1_{gZ^{-1}} \rangle > 0,$$

where convolution and the inner product use counting measure.

Write

$$1_X = \frac{|X|}{|G|} \mathbf{1} + f_X,$$

with $f_X \perp \mathbf{1}$, and similarly for Y . Then

$$1_X * 1_Y = \frac{|X||Y|}{|G|} \mathbf{1} + f_X * f_Y.$$

The regular representation decomposes into irreducibles. Since every nontrivial constituent has dimension at least D , the following quasirandom convolution estimate holds

$$(38) \quad \|f_X * f_Y\|_2 \leq \sqrt{\frac{|G|}{D}} \|f_X\|_2 \|f_Y\|_2.$$

For completeness, we derive (38) with the present counting-measure normalization. For an irreducible unitary representation π of degree d_π , set

$$\widehat{f}(\pi) = \sum_{x \in G} f(x) \pi(x).$$

Then

$$\widehat{f * g}(\pi) = \widehat{f}(\pi) \widehat{g}(\pi), \quad \|f\|_2^2 = \frac{1}{|G|} \sum_{\pi \in \widehat{G}} d_\pi \|\widehat{f}(\pi)\|_{\text{HS}}^2.$$

Because f_X has mean zero, its trivial Fourier block vanishes. Hence

$$\begin{aligned} \|f_X * f_Y\|_2^2 &= \frac{1}{|G|} \sum_{\pi \neq 1} d_\pi \|\widehat{f}_X(\pi) \widehat{f}_Y(\pi)\|_{\text{HS}}^2 \\ &\leq \frac{1}{|G|} \sum_{\pi \neq 1} d_\pi \|\widehat{f}_X(\pi)\|_{\text{HS}}^2 \|\widehat{f}_Y(\pi)\|_{\text{HS}}^2. \end{aligned}$$

Since $d_\pi \geq D$ for every nontrivial block,

$$d_\pi ab \leq \frac{1}{D} (d_\pi a) (d_\pi b) \quad (a, b \geq 0).$$

Therefore

$$\begin{aligned} \|f_X * f_Y\|_2^2 &\leq \frac{1}{D|G|} \left(\sum_{\pi \neq 1} d_\pi \|\widehat{f}_X(\pi)\|_{\text{HS}}^2 \right) \left(\sum_{\pi \neq 1} d_\pi \|\widehat{f}_Y(\pi)\|_{\text{HS}}^2 \right) \\ &\leq \frac{|G|}{D} \|f_X\|_2^2 \|f_Y\|_2^2, \end{aligned}$$

which is exactly (38) after taking square roots.

Since $\|f_X\|_2 \leq |X|^{1/2}$ and $\|f_Y\|_2 \leq |Y|^{1/2}$,

$$\|f_X * f_Y\|_2 \leq \sqrt{\frac{|G||X||Y|}{D}}.$$

Hence by Cauchy–Schwarz,

$$|\langle f_X * f_Y, 1_{gZ^{-1}} \rangle| \leq \sqrt{\frac{|G||X||Y||Z|}{D}}.$$

The main term is

$$\left\langle \frac{|X||Y|}{|G|} \mathbf{1}, 1_{gZ^{-1}} \right\rangle = \frac{|X||Y||Z|}{|G|}.$$

The hypothesis says that this main term is strictly larger than the error term, because

$$\frac{|X||Y||Z|}{|G|} > \sqrt{\frac{|G||X||Y||Z|}{D}}$$

is equivalent to $|X||Y||Z| > |G|^3/D$. Thus the inner product is positive for every g , proving $XYZ = G$. $\square$

COROLLARY 9.4 (Completion threshold in SL_2). *There is an absolute $C > 0$ such that, if $A \subseteq \mathrm{SL}_2(\mathbb{F}_p)$ satisfies*

$$|A| \geq Cp^{8/3},$$

then

$$A^3 = G.$$

PROOF. We have $|G| = p(p^2 - 1) < p^3$ and $D \geq (p - 1)/2 \gg p$. Thus

$$\left(\frac{|G|^3}{D}\right)^{1/3} \ll p^{8/3}.$$

Apply [Theorem 9.3](#) with $X = Y = Z = A$. $\square$

GUIDING IDEA 9.5. The exponent $8/3$ is not mysterious. The ambient group has size p^3 , while quasirandomness contributes a representation dimension of size p . The three-set mixing threshold is

$$|G|D^{-1/3} \asymp p^3 p^{-1/3} = p^{8/3}.$$

This is exactly where the arithmetic growth phase can hand the problem to representation theory.

3. Historical note on the large-set completion

Helfgott's original argument also contains a concrete large-set completion inside upper and lower Borel subgroups: sufficiently large fibers are converted into the full unipotent radicals, after which elementary Gaussian decomposition fills the group. That route is historically illuminating, but it requires an additional finite-field dilation estimate that plays no role elsewhere in the present series. We therefore use the quasirandom completion proved above, whose representation-theoretic input was already proved in Volume I, and do not introduce a redundant auxiliary theorem merely for a second completion proof.

4. The absolute product theorem

THEOREM 9.6 (Product theorem for $\mathrm{SL}_2(\mathbb{F}_p)$). *There exist absolute constants $\varepsilon_0 > 0$ and $c_0 > 0$ such that, for every prime $p \geq 5$ and every finite generating set $A \subseteq \mathrm{SL}_2(\mathbb{F}_p)$,*

$$\text{either} \quad |A^3| \geq c_0 |A|^{1+\varepsilon_0}, \quad \text{or} \quad A^3 = G.$$

PROOF. No symmetrization is needed here: [Theorem 8.2](#) was proved for the original, possibly nonsymmetric, triple product, and the quasirandom completion lemma also applies to an arbitrary subset. Fix

$$\delta_0 = \frac{1}{4}.$$

If

$$|A| \leq p^{3-\delta_0} = p^{11/4},$$

then [Theorem 8.2](#) gives

$$|A^3| \geq c|A|^{1+\varepsilon(\delta_0)}.$$

On the other hand [Corollary 9.4](#) gives $A^3 = G$ whenever

$$|A| \geq Cp^{8/3}.$$

The two ranges overlap for all sufficiently large p , since

$$\frac{p^{11/4}}{Cp^{8/3}} = \frac{p^{1/12}}{C} \longrightarrow \infty.$$

Thus every sufficiently large prime lies in at least one of the two regimes. The finitely many remaining primes are absorbed by decreasing c_0 and, if necessary, ε_0 . Taking

$$\varepsilon_0 \leq \varepsilon(1/4)$$

proves the theorem. □

WARNING 9.7 (About the transition range). One should not silently set the constant in the quasirandom threshold equal to one. The clean solution is to choose the small-set exponent with slack: $11/4 > 8/3$, so the two regimes overlap for large p .

5. Polylogarithmic diameter

THEOREM 9.8 (Helfgott diameter theorem). *There are absolute constants $C, c > 0$ such that, for every prime $p \geq 5$ and every generating set $S \subseteq \mathrm{SL}_2(\mathbb{F}_p)$,*

$$\mathrm{diam} \mathrm{Cay}(G, S) \leq C(\log p)^c.$$

PROOF. Let

$$A_0 = S \cup S^{-1} \cup \{1\}, \quad A_{j+1} = A_j^3.$$

As long as $A_j^3 \neq G$, [Theorem 9.6](#) gives

$$|A_{j+1}| \geq c_0|A_j|^{1+\varepsilon_0}.$$

After discarding the first bounded number of stages needed to absorb $c_0 < 1$, the logarithms satisfy

$$\log |A_{j+1}| \geq (1 + \varepsilon_0/2) \log |A_j|.$$

Since $|G| < p^3$, the number of stages before reaching the completion regime is $O(\log \log p)$. But A_j consists of words of length at most 3^j in $S \cup S^{-1}$. Therefore the necessary word length is

$$3^{O(\log \log p)} = (\log p)^{O(1)}.$$

One final triple product gives all of G . □

6. Product growth versus expansion

The theorem just proved is not itself a spectral-gap theorem. It says that *every generating set grows under self-multiplication*. The Bourgain–Gamburd expansion machine of Volume I needs a related but more measure-theoretic statement: approximate subgroups cannot trap a random walk at intermediate scales. The implication is

product theorem $\implies$ classification of approximate obstructions $\implies L^2$ -flattening.

Volumes V and VI of this arc repackage the explicit rank-one proof into quantitative algebraic geometry and Larsen–Pink theory, isolating the mechanisms that survive beyond rank one.

CHECKPOINT 9.9. We have completed the concrete $\mathrm{SL}_2(\mathbb{F}_p)$ arc. At this point no algebraic-group theorem beyond elementary rank-one structure has been used as a black box. The finite-field sum–product inputs have been proved in Volume III and Chapter 4, while the representation-theoretic completion input was proved in Volume I. Thus the concrete product theorem now has no non-basic external proof dependency.

Bibliography

1. Harald A. Helfgott, *Growth and generation in $\mathrm{SL}_2(F/pF)$* , Annals of Mathematics **167** (2008), no. 2, 601–623.
2. Sergei V. Konyagin, *A sum-product estimate in fields of prime order*, 2003, arXiv:math/0304217.

The Bourgain–Gamburd Method
Volume V: Quantitative Algebraic Geometry for
Growth

Preface

The product-growth arguments of the preceding volume repeatedly used a rank-one phenomenon: maps that are generically finite produce three-dimensional growth, one-dimensional tori meet an approximate group at a one-third scale, and exceptional equations define lower-dimensional traps from which a generating set must escape. In rank one these facts can be checked directly with 2×2 matrices. In higher rank they must be replaced by a systematic geometry.

This volume develops exactly that geometry, with proofs. Statements such as “the generic fiber has the expected dimension”, “the exceptional set has bounded complexity”, or “a proper subvariety can be escaped in bounded word length” will not be used as slogans. We build the commutative algebra that makes them precise, prove the dimension and fiber theorems needed later, and then translate them into quantitative finite-set statements.

The governing convention is the same as in Volumes I–IV. Basic linear algebra and elementary field theory may be used without reproving them. A non-basic theorem that carries mathematical weight in the growth argument is proved here unless it was established in an earlier volume. Historical citations indicate provenance; they do not replace proof. **Series position.** This volume uses elementary algebra and

the growth questions from Volume IV. Its main outputs are the dimension, degree, escape, and morphism estimates used by Larsen–Pink arguments.

Quantifier and asymptotic conventions. Unless a statement explicitly says otherwise, every constant occurring in a quantitative theorem is chosen in the order in which it is quantified in that theorem. Thus a phrase such as “for all sufficiently small δ ” means: there is a displayed or locally quantified threshold $\delta_0 > 0$, depending only on the parameters listed before that phrase, such that the assertion holds for every $0 < \delta < \delta_0$. The phrases “sufficiently large N ” and “sufficiently large p ” have the analogous meaning.

If $X \ll_\theta Y$ or $X = O_\theta(Y)$, the implicit constant depends only on θ and on the fixed ambient data of the current volume; it is independent of the variables quantified after θ . The words *absolute*, *uniform*, and *bounded* never introduce an unlisted dependence. Whenever an inequality between exponents or scales is essential to the argument, that

inequality is written explicitly in the proof; these conventions are not a substitute for such a calculation.

Contents

Preface	iii
Part 1. Algebra and Affine Geometry	1
Chapter 1. Noetherian Algebra and the Nullstellensatz	2
1. Noetherian rings	2
2. Zariski's lemma	3
3. Weak and strong Nullstellensatz	5
4. Irreducibility and prime ideals	6
Chapter 2. Integral Dependence, Normalization, and Dimension	8
1. Integral elements	8
2. Lying over and closed finite maps	9
3. Noether normalization	11
4. Dimension	12
Chapter 3. Morphisms, Function Fields, and Finite Charts	15
1. Regular maps and coordinate rings	15
2. Clearing denominators into integrality	16
3. A worked finite-chart example	17
4. Quasi-finite maps and finite fibers on a dense chart	18
5. Fiber equations	19
Part 2. Dimension, Degree, and Morphisms	20
Chapter 4. Constructible Images and the Generic Fiber Theorem	21
1. Dominant images contain an open set	21
2. Generic versus exceptional fibers: a laboratory	22
3. Constructible sets	22
4. The generic fiber theorem	23
Chapter 5. Hilbert Functions and Degree	26
1. Graded algebras and Hilbert functions	26
2. Homogeneous normalization and the degree of the Hilbert polynomial	27
3. Projective degree	28

4. What degree will and will not do for us	30
Chapter 6. Quantitative Finite Maps and Slicing	31
1. Fiber size of a finite morphism	31
2. Generic finite-fiber bounds	32
3. Adding the missing parameters	33
4. A product-map laboratory	34
5. From a finite map to a finite-set inequality	34
Part 3. Algebraic Groups and Growth Geometry	36
Chapter 7. Algebraic Groups, Actions, and Orbit Dimension	37
1. Affine algebraic groups	37
2. Algebraic actions and stabilizers	38
3. Orbits are locally closed	39
4. Centralizers and conjugacy classes	39
5. Stabilizers of subvarieties	40
Chapter 8. Escape from Proper Subvarieties	42
1. The bounded-degree equation space	42
2. A bounded list of escaping words	43
3. Positive-proportion escape	45
4. A two-dimensional escape laboratory	45
Chapter 9. Slicing and the Product–Conjugation Phenomenon	47
1. A slicing lemma	47
2. Setwise stabilizers under translation	48
3. The bad-conjugator set is algebraic	48
4. When multiplication fails to gain dimension	49
5. Rank-one shadow of product–conjugation	51
6. Proof architecture for the next volume	51
Appendix A. Principal Ideals and Degree Bookkeeping	52
1. Nakayama’s lemma	52
2. Symbolic powers	52
3. Krull’s principal ideal theorem	53
4. Leading coefficients of unions	55
5. The full reduced hypersurface degree estimate	56
6. Finite-field point counting from degree	57
Appendix B. Uniform Algebraic Geometry by Ultralimits	59
1. Complexity	59
2. A non-principal ultrafilter	59
3. Ultraproducts of fields	60

4.	Bounded polynomials survive the limit	61
5.	Continuity of dimension	61
6.	Weak Bézout as a compactness consequence	62
7.	Degree and bounded equations for irreducible subvarieties	62
8.	Irreducibility and components are continuous	65
9.	Uniformity of the operations used in growth	66

Part 1

Algebra and Affine Geometry

CHAPTER 1

Noetherian Algebra and the Nullstellensatz

Algebraic geometry enters product growth through polynomial equations. Before dimension or degree can be used quantitatively, one needs two finiteness principles: polynomial ideals are finitely generated, and over an algebraically closed field algebraic sets are recovered from their vanishing ideals. We prove both from the ground up.

Throughout this chapter k is a field. When the Nullstellensatz is invoked, k will be algebraically closed.

1. Noetherian rings

DEFINITION 1.1. A commutative ring R is *Noetherian* if every ideal of R is finitely generated.

The form of Noetherianity used in geometry is often the ascending-chain condition.

PROPOSITION 1.2 (Equivalent forms of Noetherianity). *For a commutative ring R , the following are equivalent.*

- (i) *Every ideal of R is finitely generated.*
- (ii) *Every increasing chain of ideals*

$$I_1 \subseteq I_2 \subseteq \cdots$$

stabilizes.

- (iii) *Every nonempty collection of ideals has a maximal element under inclusion.*

PROOF. Assume (i), and let $I_1 \subseteq I_2 \subseteq \cdots$. The union $I = \bigcup_n I_n$ is an ideal. Choose generators $a_1, \dots, a_r$ of I . All of them lie in some I_N , hence $I = I_N$ and the chain stabilizes. Thus (i) implies (ii).

Assume (ii), and suppose a nonempty collection C of ideals has no maximal member. Choose $I_1 \in C$. Since it is not maximal, choose $I_2 \in C$ with $I_1 \subsetneq I_2$, and continue inductively. This contradicts (ii). Hence (ii) implies (iii).

Assume (iii), and let I be an ideal. Let C be the collection of finitely generated ideals contained in I . It is nonempty because $(0) \in C$. Choose a maximal $J \in C$. If $J \neq I$, choose $a \in I \setminus J$; then $J + (a)$ is a strictly larger finitely generated ideal in I , contradiction. Thus $J = I$, proving (i). $\square$

THEOREM 1.3 (Hilbert basis theorem). *If R is Noetherian, then $R[x]$ is Noetherian. Consequently $k[x_1, \dots, x_n]$ is Noetherian for every field k and every $n \geq 0$.*

PROOF. Let $I \triangleleft R[x]$. For each $d \geq 0$, let $L_d \subseteq R$ be the set of leading coefficients of polynomials in I of degree at most d , together with 0. Then L_d is an ideal: sums are handled after padding degrees, and multiplying a polynomial by an element of R multiplies its leading coefficient. Moreover

$$L_0 \subseteq L_1 \subseteq L_2 \subseteq \cdots.$$

Since R is Noetherian, this chain stabilizes at some D .

For each $0 \leq d \leq D$, choose finitely many polynomials

$$f_{d,1}, \dots, f_{d,r_d} \in I$$

of degree at most d whose leading coefficients generate L_d . Let J be the ideal of $R[x]$ generated by all these finitely many polynomials. We show $I = J$ by induction on degree.

Take $f \in I$, $f \neq 0$, of degree m . If $m \leq D$, the leading coefficient of f lies in L_m and is a linear combination of the leading coefficients of the chosen $f_{m,j}$. Multiplying those $f_{m,j}$ by suitable powers of x if necessary, we can subtract an element of J having the same leading term as f , leaving an element of I of smaller degree.

If $m > D$, stabilization gives $L_m = L_D$. Express the leading coefficient of f as a combination of leading coefficients of the $f_{D,j}$ and multiply $f_{D,j}$ by $x^{m-\deg f_{D,j}}$ to cancel the leading term. Again the remainder lies in I and has smaller degree. Repeating and using induction gives $f \in J$. Hence $I = J$.

Starting from the Noetherian field k and iterating the one-variable result proves the final assertion. $\square$

COROLLARY 1.4 (Descending-chain condition for algebraic sets). *Let k be any field. Every descending chain of algebraic subsets of k^n stabilizes.*

PROOF. If $X_1 \supseteq X_2 \supseteq \cdots$, then their vanishing ideals satisfy

$$I(X_1) \subseteq I(X_2) \subseteq \cdots$$

in the Noetherian ring $k[x_1, \dots, x_n]$. The ideals stabilize by [Proposition 1.2](#), and equal vanishing ideals have equal zero sets. $\square$

2. Zariski's lemma

The algebraic heart of the Nullstellensatz is the following finite-generation theorem for fields.

LEMMA 1.5 (A field cannot be a polynomial ring in one transcendental element). *Let K be a field and t transcendental over K . If $u_1, \dots, u_m \in K(t)$, then the K -subalgebra $K[u_1, \dots, u_m]$ is not all of $K(t)$.*

PROOF. Write each $u_i = f_i(t)/g_i(t)$ with $f_i, g_i \in K[t]$ and $g_i \neq 0$. Put $g = \prod_i g_i$. Then

$$K[u_1, \dots, u_m] \subseteq K[t, g(t)^{-1}].$$

Choose an irreducible polynomial $q(t) \in K[t]$ not dividing $g(t)$. Such a q exists: if only finitely many irreducibles existed, their product plus 1 would have an irreducible factor outside the list. Then $1/q(t) \notin K[t, g^{-1}]$, because an identity

$$\frac{1}{q} = \frac{h}{g^N}$$

would imply $g^N = qh$, forcing $q \mid g$. Hence the displayed finitely generated algebra is a proper subring of $K(t)$. $\square$

THEOREM 1.6 (Zariski's lemma). *Let $k \subseteq L$ be fields. If L is finitely generated as a k -algebra, then L is a finite algebraic extension of k .*

PROOF. Write

$$L = k[a_1, \dots, a_n].$$

Choose a maximal algebraically independent subfamily among the a_i , say $a_1, \dots, a_r$. Then every remaining a_j is algebraic over the rational function field

$$K = k(a_1, \dots, a_r).$$

Hence L is a finite extension of K .

We claim $r = 0$. Suppose $r > 0$. Since L/K is finite, each a_j for $j > r$ is integral over a localization

$$A = k[a_1, \dots, a_r, s^{-1}]$$

for some nonzero polynomial $s \in k[a_1, \dots, a_r]$: clear the finitely many denominators in monic algebraic equations for the a_j . Thus L is integral over A .

Because L is a field and integral over A , the ring A must itself be a field. Indeed, for $0 \neq a \in A$, the element $a^{-1} \in L$ is integral over A , so

$$a^{-m} + c_{m-1}a^{-(m-1)} + \dots + c_0 = 0$$

for some $c_i \in A$; multiplying by a^{m-1} expresses a^{-1} as an element of A .

Now regard

$$A = k[a_2, \dots, a_r][a_1, s^{-1}].$$

Its fraction field contains the rational function field $k(a_2, \dots, a_r)(a_1)$. If A were a field, it would equal its fraction field, contradicting Lemma 1.5 after taking $K = k(a_2, \dots, a_r)$ and observing that adjoining the single inverse s^{-1} is a finitely generated K -algebra. Therefore $r = 0$.

Thus every a_i is algebraic over k . A finitely generated algebraic field extension is finite, so $[L : k] < \infty$. $\square$

REMARK 1.7. The only subtle point in the proof is the localization step. If a_j satisfies

$$c_m a_j^m + \cdots + c_0 = 0$$

with $c_i \in k[a_1, \dots, a_r]$ and $c_m \neq 0$, then after inverting c_m we divide by c_m and obtain a monic equation. One common denominator works for the finitely many algebraic generators.

3. Weak and strong Nullstellensatz

For an ideal $I \subseteq k[x_1, \dots, x_n]$, write

$$V(I) = \{x \in k^n : f(x) = 0 \text{ for every } f \in I\}.$$

For $X \subseteq k^n$, write

$$I(X) = \{f \in k[x_1, \dots, x_n] : f|_X = 0\}.$$

THEOREM 1.8 (Weak Nullstellensatz). *Let k be algebraically closed. Every maximal ideal of $k[x_1, \dots, x_n]$ has the form*

$$(x_1 - a_1, \dots, x_n - a_n)$$

for a unique point $a = (a_1, \dots, a_n) \in k^n$. Equivalently, every proper ideal has a common zero in k^n .

PROOF. Let $\mathfrak{m}$ be maximal. Then

$$L = k[x_1, \dots, x_n]/\mathfrak{m}$$

is a field finitely generated as a k -algebra. By Zariski's lemma, L/k is finite algebraic. Since k is algebraically closed, $L = k$. Let a_i be the image of x_i in $L = k$. Then $x_i - a_i \in \mathfrak{m}$ for every i , and maximality gives

$$\mathfrak{m} = (x_1 - a_1, \dots, x_n - a_n).$$

Uniqueness is immediate by evaluating the coordinate functions.

If I is proper, it lies in a maximal ideal by Proposition 1.2 applied to the collection of proper ideals containing I ; the corresponding point is a common zero of I . $\square$

THEOREM 1.9 (Hilbert Nullstellensatz). *Let k be algebraically closed and $I \subseteq k[x_1, \dots, x_n]$ an ideal. Then*

$$I(V(I)) = \sqrt{I}.$$

PROOF. Every element of $\sqrt{I}$ vanishes on $V(I)$, so

$$\sqrt{I} \subseteq I(V(I)).$$

For the reverse inclusion, take $f \in I(V(I))$. Introduce a new variable y and consider the ideal

$$J = (I, 1 - yf) \subseteq k[x_1, \dots, x_n, y].$$

There is no common zero of J : a zero would have $x \in V(I)$ and hence $f(x) = 0$, contradicting $1 - yf(x) = 0$. By the weak Nullstellensatz, J is the whole polynomial ring. Thus

$$1 = \sum_i a_i(x, y)g_i(x) + b(x, y)(1 - yf(x))$$

for finitely many $g_i \in I$.

Substitute formally $y = f(x)^{-1}$ in the localization $k[x_1, \dots, x_n]_f$. The last term vanishes and gives

$$1 \in I k[x_1, \dots, x_n]_f.$$

Therefore for some $N \geq 0$ one has $f^N \in I$. Hence $f \in \sqrt{I}$. $\square$

COROLLARY 1.10 (Algebra–geometry correspondence). *For algebraically closed k , the assignments*

$$X \longmapsto I(X), \quad I \longmapsto V(I)$$

give inverse inclusion-reversing bijections between algebraic subsets of k^n and radical ideals of $k[x_1, \dots, x_n]$.

PROOF. For an algebraic set $X = V(J)$, the Nullstellensatz gives

$$V(I(X)) = V(\sqrt{J}) = V(J) = X.$$

For a radical ideal I , the same theorem gives $I(V(I)) = \sqrt{I} = I$. $\square$

4. Irreducibility and prime ideals

DEFINITION 1.11. A nonempty algebraic set X is *irreducible* if it cannot be written as the union of two proper algebraic subsets. Its coordinate ring is

$$k[X] = k[x_1, \dots, x_n]/I(X).$$

PROPOSITION 1.12 (Irreducibility criterion). *Let k be algebraically closed and $X \subseteq k^n$ algebraic. Then X is irreducible if and only if $I(X)$ is prime, equivalently if and only if $k[X]$ is an integral domain.*

PROOF. Suppose $I(X)$ is not prime. Choose $f, g \notin I(X)$ with $fg \in I(X)$. Then

$$X = (X \cap V(f)) \cup (X \cap V(g)),$$

and neither piece is all of X , so X is reducible.

Conversely, if $X = Y \cup Z$ with $Y, Z \subsetneq X$ closed, choose

$$f \in I(Y) \setminus I(X), \quad g \in I(Z) \setminus I(X).$$

Then fg vanishes on $Y \cup Z = X$, so $fg \in I(X)$ while neither factor belongs to $I(X)$. Thus $I(X)$ is not prime. $\square$

THEOREM 1.13 (Finite irreducible decomposition). *Every algebraic set X is a finite union*

$$X = X_1 \cup \cdots \cup X_r$$

of irreducible closed subsets maximal under inclusion. These maximal irreducible components are uniquely determined up to order.

PROOF. Existence follows from Noetherianity. If some closed set had no finite irreducible decomposition, choose a minimal such set under inclusion. It is reducible, say $X = Y \cup Z$ with $Y, Z \subsetneq X$ closed. By minimality both Y and Z have finite irreducible decompositions, hence so does X , contradiction.

For uniqueness, suppose

$$X = \bigcup_{i=1}^r X_i = \bigcup_{j=1}^s Y_j$$

with all X_i, Y_j irreducible and maximal. Fix i . Since

$$X_i = \bigcup_j (X_i \cap Y_j),$$

irreducibility of X_i implies $X_i \subseteq Y_j$ for some j ; maximality gives equality. Thus each X_i is one of the Y_j , and symmetrically each Y_j is one of the X_i . $\square$

CHECKPOINT 1.14. At this point every affine algebraic set admits a finite irreducible decomposition, and irreducible geometry can be translated into prime-ideal algebra. The next chapter develops integrality and normalization, which turn transcendence degree into a workable notion of geometric dimension.

CHAPTER 2

Integral Dependence, Normalization, and Dimension

Dimension in growth arguments is not a picture; it is the number of algebraically independent parameters. This chapter proves the algebraic machinery that makes that statement precise. The central result is Noether normalization: every irreducible affine variety is finite over an affine space of the same transcendence degree.

Throughout, rings are commutative with identity and algebras are finitely generated unless stated otherwise.

1. Integral elements

DEFINITION 2.1. Let $A \subseteq B$ be rings. An element $b \in B$ is *integral over A* if it satisfies a monic equation

$$b^m + a_{m-1}b^{m-1} + \cdots + a_0 = 0, \quad a_i \in A.$$

The extension B/A is *integral* if every $b \in B$ is integral over A .

PROPOSITION 2.2 (Finite-module criterion). *Let $A \subseteq B$ and $b \in B$. The following are equivalent.*

- (i) b is integral over A .
- (ii) $A[b]$ is finitely generated as an A -module.
- (iii) There exists a faithful finitely generated A -submodule $M \subseteq B$ with $1 \in M$ and $bM \subseteq M$.

PROOF. If b satisfies a monic equation of degree m , every b^n with $n \geq m$ is an A -linear combination of $1, b, \dots, b^{m-1}$, proving (ii). Condition (ii) implies (iii) with $M = A[b]$.

Assume (iii), and choose A -generators $m_1, \dots, m_r$ of M . There is a matrix $C = (c_{ij}) \in M_r(A)$ such that

$$bm_i = \sum_j c_{ij}m_j.$$

Thus

$$(bI - C)(m_1, \dots, m_r)^t = 0.$$

Multiplying by the adjugate gives

$$\det(bI - C)m_i = 0$$

for every i . Since M is faithful and contains 1, $\det(bI - C) = 0$ in B . This determinant is a monic polynomial in b with coefficients in A , proving integrality. $\square$

COROLLARY 2.3 (Transitivity and finite generation). *If $B = A[b_1, \dots, b_r]$ and each b_i is integral over A , then B is a finite A -module. If $A \subseteq B \subseteq C$, B is integral over A , and C is integral over B , then C is integral over A .*

PROOF. Adjoin the b_i one at a time and use [Proposition 2.2](#); a product of finite generating sets generates the final algebra as an A -module.

For transitivity, take $c \in C$. A monic equation for c over B involves finitely many coefficients $b_1, \dots, b_s \in B$. The algebra $A[b_1, \dots, b_s, c]$ is finite over $A[b_1, \dots, b_s]$, which is finite over A , hence finite over A . Apply the finite-module criterion to c . $\square$

LEMMA 2.4 (Integrality survives quotients and localization). *If B is integral over A , then:*

- (i) *for every ideal $I \triangleleft A$, B/IB is integral over A/I ;*
- (ii) *for every multiplicative set $S \subseteq A$, $S^{-1}B$ is integral over $S^{-1}A$.*

PROOF. Reduce a monic equation modulo I for (i), and divide its coefficients by suitable elements of S for (ii). More explicitly, if b satisfies $b^m + \sum a_i b^i = 0$, then b/s satisfies after multiplying by a suitable power of s a monic equation over $S^{-1}A$. $\square$

2. Lying over and closed finite maps

LEMMA 2.5 (Fields in integral extensions). *Let $A \subseteq B$ be an integral extension.*

- (i) *If B is a field, then A is a field.*
- (ii) *If A is a field and B is an integral domain, then B is a field.*

Thus an integral domain integral over a field is automatically a field.

PROOF. For (i), take $0 \neq a \in A$. Since B is a field, $a^{-1} \in B$; integrality gives

$$(a^{-1})^m + c_{m-1}(a^{-1})^{m-1} + \dots + c_0 = 0, \quad c_i \in A.$$

Multiplying by a^{m-1} gives

$$a^{-1} = -(c_{m-1} + c_{m-2}a + \dots + c_0a^{m-1}) \in A.$$

Hence every nonzero element of A is invertible.

For (ii), let $0 \neq b \in B$ and choose a monic equation of least degree

$$b^m + a_{m-1}b^{m-1} + \dots + a_0 = 0, \quad a_i \in A.$$

Minimality and the fact that B is a domain imply $a_0 \neq 0$: otherwise division by b gives a shorter monic relation. Since A is a field,

$$b^{-1} = -a_0^{-1}(b^{m-1} + a_{m-1}b^{m-2} + \cdots + a_1) \in B.$$

So B is a field. □

THEOREM 2.6 (Lying over). *Let B be integral over A . For every prime ideal $\mathfrak{p} \subseteq A$, there exists a prime ideal $\mathfrak{q} \subseteq B$ with*

$$\mathfrak{q} \cap A = \mathfrak{p}.$$

PROOF. Let

$$S = A \setminus \mathfrak{p}.$$

By Lemma 2.4, the localization $S^{-1}B$ is integral over the local ring $A_{\mathfrak{p}} = S^{-1}A$. Choose a maximal ideal $\mathfrak{n}$ of $S^{-1}B$, and let

$$\mathfrak{m} = \mathfrak{n} \cap A_{\mathfrak{p}}.$$

The field $(S^{-1}B)/\mathfrak{n}$ is integral over the domain $A_{\mathfrak{p}}/\mathfrak{m}$. By part (i) of Lemma 2.5, the latter is a field. Hence $\mathfrak{m}$ is maximal in the local ring $A_{\mathfrak{p}}$, so

$$\mathfrak{m} = \mathfrak{p}A_{\mathfrak{p}}.$$

Let $\mathfrak{q}$ be the inverse image of $\mathfrak{n}$ under $B \rightarrow S^{-1}B$. Then $\mathfrak{q}$ is prime, is disjoint from S , and its contraction localizes to $\mathfrak{p}A_{\mathfrak{p}}$. Therefore

$$\mathfrak{q} \cap A = \mathfrak{p}.$$

□

PROPOSITION 2.7 (Finite morphisms are closed). *Let k be algebraically closed, $A \subseteq B$ a finite integral extension of finitely generated reduced k -algebras, and let*

$$\pi : \operatorname{Spec} B(k) \longrightarrow \operatorname{Spec} A(k)$$

be contraction of maximal ideals, equivalently the corresponding map of affine algebraic sets. Then π is surjective and maps closed algebraic subsets to closed algebraic subsets.

PROOF. Surjectivity is lying over, together with the weak Nullstellensatz identifying maximal ideals with k -points.

A closed subset of $\operatorname{Spec} B(k)$ is $V_B(J)$ for some ideal $J \triangleleft B$. The quotient B/J is integral over $A/(J \cap A)$ by Lemma 2.4. Lying over shows that a maximal ideal $\mathfrak{m}$ of A contains $J \cap A$ if and only if some maximal ideal of B above $\mathfrak{m}$ contains J . Hence

$$\pi(V_B(J)) = V_A(J \cap A),$$

which is closed. □

3. Noether normalization

We now prove the central finite-parameter theorem.

LEMMA 2.8 (Making one relation monic). *Let k be an infinite field and*

$$0 \neq f \in k[x_1, \dots, x_n], \quad n \geq 2.$$

There is a triangular linear change of variables

$$x_i = y_i + c_i y_n \quad (1 \leq i < n), \quad x_n = y_n,$$

with $c_i \in k$ such that the transformed polynomial, regarded as an element of

$$k[y_1, \dots, y_{n-1}][y_n],$$

has a nonzero constant leading coefficient in y_n . After scaling, it is monic in y_n .

PROOF. Write $f = f_d + f_{d-1} + \dots$ with f_d its highest homogeneous part. Under the substitution, the coefficient of y_n^d is

$$f_d(c_1, \dots, c_{n-1}, 1).$$

This is not the zero polynomial in the c_i : otherwise f_d would vanish on the affine chart $x_n = 1$, hence by homogeneity everywhere, contradiction. Since k is infinite, choose c_i making it nonzero. Divide the transformed polynomial by that nonzero scalar. $\square$

THEOREM 2.9 (Noether normalization). *Let k be an infinite field and A a finitely generated k -algebra that is an integral domain. If*

$$d = \text{trdeg}_k \text{Frac}(A),$$

then there exist algebraically independent elements $y_1, \dots, y_d \in A$ such that A is finite and integral over

$$k[y_1, \dots, y_d].$$

PROOF. Write $A = k[a_1, \dots, a_n]$. We argue by induction on n . If the a_i are algebraically independent, take them all and there is nothing to prove.

Otherwise there is a nonzero polynomial relation

$$f(a_1, \dots, a_n) = 0.$$

After a triangular linear change among the generators, Lemma 2.8 makes this relation monic in the last new generator, say b_n . Hence b_n is integral over $k[b_1, \dots, b_{n-1}]$. Therefore

$$A = k[b_1, \dots, b_n]$$

is integral over $A_0 = k[b_1, \dots, b_{n-1}]$.

The fraction fields $\text{Frac}(A)$ and $\text{Frac}(A_0)$ have the same transcendence degree, because b_n is algebraic over $\text{Frac}(A_0)$. Apply induction to the domain A_0 to obtain algebraically independent $y_1, \dots, y_d \in A_0$ such that A_0 is finite over $k[y_1, \dots, y_d]$. By transitivity, A is finite over the same polynomial algebra. $\square$

REMARK 2.10. In our geometric applications the ground field is often an algebraic closure of a finite field, which is infinite. Thus the infinitude hypothesis is harmless. Statements over a finite field are interpreted geometrically after base change to its algebraic closure.

4. Dimension

DEFINITION 2.11 (Dimension). Let X be an irreducible affine variety over an algebraically closed field k . Define

$$\dim X = \text{trdeg}_k \text{Frac}(k[X]).$$

For a general algebraic set $X = \bigcup_i X_i$ with irreducible components X_i , define

$$\dim X = \max_i \dim X_i,$$

and set $\dim \emptyset = -\infty$.

LEMMA 2.12 (Products of irreducible affine varieties). *If X and Y are irreducible affine varieties over an algebraically closed field k , then $X \times Y$ is irreducible.*

PROOF. Put $A = k[X]$ and $B = k[Y]$. It is enough to prove that $A \otimes_k B$ is a domain. Let

$$0 \neq F = \sum_{i=1}^r a_i \otimes b_i, \quad 0 \neq G = \sum_{j=1}^s c_j \otimes d_j,$$

where the b_i are k -linearly independent and the d_j are k -linearly independent; such expressions are obtained by choosing bases for the finite-dimensional spans of the second tensor factors.

For $x \in X(k)$, write

$$F_x = \sum_i a_i(x) b_i \in B.$$

If $F_x = 0$ for every x , linear independence of the b_i gives $a_i(x) = 0$ for every x and every i . The strong Nullstellensatz then gives $a_i = 0$, contradiction. Hence

$$U_F = \{x : F_x \neq 0\}$$

is a nonempty open subset of X ; concretely it is the union of the principal opens $D(a_i)$. Similarly U_G is nonempty open. Irreducibility of X gives an $x \in U_F \cap U_G$. Since B is a domain,

$$F_x G_x \neq 0.$$

But $(FG)_x = F_x G_x$, so $FG \neq 0$. Thus $A \otimes_k B$ has no zero divisors. $\square$

LEMMA 2.13 (Incomparability). *Let B be integral over A . If prime ideals $\mathfrak{q}_1 \subseteq \mathfrak{q}_2$ of B have the same contraction $\mathfrak{p}$ to A , then $\mathfrak{q}_1 = \mathfrak{q}_2$.*

PROOF. Pass to $B/\mathfrak{q}_1$ and $A/\mathfrak{p}$, then localize the latter at its nonzero elements. We reduce to an integral domain B' integral over a field K , with a prime ideal $\mathfrak{q} \subseteq B'$ contracting to 0. By Lemma 2.5, B' itself is a field. Hence its only prime ideal is 0, so $\mathfrak{q} = 0$. $\square$

PROPOSITION 2.14 (Basic dimension calculus). *Let X, Y be affine algebraic sets over an algebraically closed field.*

- (i) $\dim \mathbb{A}^n = n$.
- (ii) $\dim(X \cup Y) = \max(\dim X, \dim Y)$.
- (iii) If $X \subseteq Y$, then $\dim X \leq \dim Y$.
- (iv) $\dim(X \times Y) = \dim X + \dim Y$.
- (v) If X is irreducible and $Z \subsetneq X$ is closed, then $\dim Z < \dim X$.

PROOF. Items (i) and (ii) follow directly from the definition. For (iii), let Z be an irreducible component of X contained in an irreducible component Y_0 of Y . The inclusion corresponds to a quotient

$$k[Y_0] \twoheadrightarrow k[Z] = k[Y_0]/\mathfrak{p}.$$

If classes $\bar{f}_1, \dots, \bar{f}_r$ are algebraically independent in the quotient, then any polynomial relation among chosen lifts f_i in $k[Y_0]$ would descend to one among the $\bar{f}_i$. Thus $r \leq \dim Y_0$, proving monotonicity.

For (iv), first suppose X, Y irreducible. By Lemma 2.12, $X \times Y$ is irreducible and

$$k[X \times Y] = k[X] \otimes_k k[Y].$$

Choose Noether normalizations

$$k[x_1, \dots, x_r] \subseteq k[X], \quad k[y_1, \dots, y_s] \subseteq k[Y].$$

Tensoring the two inclusions over the field k is injective, and the tensor product is finite over

$$k[x_1, \dots, x_r, y_1, \dots, y_s].$$

Hence its fraction field is algebraic over the rational function field in $r + s$ variables, so

$$\dim(X \times Y) = r + s.$$

For general X, Y , decompose them into finitely many irreducible components; the irreducible components of the product are among the products of components, and the maximum of their dimensions is the sum of the two maxima.

For (v), it suffices to treat an irreducible component Z_0 of Z . Let $A = k[X]$, let $\mathfrak{q} \neq 0$ be the prime defining Z_0 , and choose a Noether normalization

$$S = k[t_1, \dots, t_d] \subseteq A, \quad d = \dim X,$$

with A integral over S . If $\mathfrak{q} \cap S = 0$, then the two primes

$$0 \subsetneq \mathfrak{q}$$

of A would have the same contraction to S , contradicting incomparability [Lemma 2.13](#). Hence $\mathfrak{q} \cap S \neq 0$. The quotient $A/\mathfrak{q}$ is integral over $S/(\mathfrak{q} \cap S)$, and the latter has transcendence degree at most $d - 1$ because it satisfies a nonzero polynomial relation. Thus

$$\dim Z_0 \leq d - 1.$$

□

REMARK 2.15. The proof of dimension drop isolates the mechanism that will recur throughout the book: Noether normalization converts geometric dimension into polynomial parameters, while incomparability prevents a proper prime from retaining all of them.

COROLLARY 2.16 (Dimension drop, algebraic form). *If X is irreducible and $Z \subsetneq X$ is closed, then*

$$\dim Z \leq \dim X - 1.$$

PROOF. Normalize $k[X]$ over $A = k[y_1, \dots, y_d]$. Let $\mathfrak{q} \neq 0$ be the prime defining an irreducible component of Z . If $\mathfrak{q} \cap A = 0$, then $0 \subsetneq \mathfrak{q}$ are two primes of $k[X]$ over the same prime of A , contradicting incomparability. Hence $\mathfrak{q} \cap A \neq 0$, so the quotient of A has transcendence degree at most $d - 1$. Integrality of the quotient extension then gives $\dim Z \leq d - 1$. □

CHECKPOINT 2.17. Noether normalization gives a precise interpretation of $d = \dim X$: after a finite algebraic ambiguity, an irreducible d -dimensional variety has exactly d free affine parameters. The next chapter turns this qualitative statement into quantitative complexity and degree bounds.

CHAPTER 3

Morphisms, Function Fields, and Finite Charts

A morphism creates growth when it has few parameters in its fibers. The algebraic mechanism is simple once stated correctly: enlarge the base function field by a transcendence basis for the total function field, then clear denominators until the remaining coordinates become integral. This produces a dense open chart on which the morphism is finite after adjoining exactly the expected number of free parameters.

Let k be algebraically closed throughout this chapter.

1. Regular maps and coordinate rings

DEFINITION 3.1. For affine varieties $X \subseteq k^m$ and $Y \subseteq k^n$, a map

$$f : X \rightarrow Y$$

is *regular* if its coordinates are restrictions of polynomials on k^m . It induces a k -algebra homomorphism

$$f^* : k[Y] \rightarrow k[X], \quad \phi \mapsto \phi \circ f.$$

PROPOSITION 3.2 (Closure of the image). *Let X, Y be affine algebraic sets and $f : X \rightarrow Y$ regular. Then*

$$I(\overline{f(X)}) = \ker f^*.$$

In particular, if X, Y are irreducible, then f has dense image if and only if $f^* : k[Y] \rightarrow k[X]$ is injective.

PROOF. A function $\phi \in k[Y]$ lies in $\ker f^*$ exactly when it vanishes at every point of $f(X)$. Vanishing is unchanged after Zariski closure, so

$$\ker f^* = I(f(X)) = I(\overline{f(X)}).$$

For irreducible Y , $I(Y) = 0$ inside its coordinate ring. Hence the closure is all of Y exactly when the kernel is zero. □

DEFINITION 3.3. If X is irreducible, its function field is

$$k(X) = \text{Frac}(k[X]).$$

A dominant regular map $f : X \rightarrow Y$ between irreducible varieties gives an inclusion

$$k(Y) \hookrightarrow k(X).$$

Its *relative transcendence degree* is

$$d(f) = \text{trdeg}_{k(Y)} k(X) = \dim X - \dim Y.$$

The equality in the definition follows from the additivity of transcendence degree in a tower of fields.

LEMMA 3.4 (Additivity of transcendence degree). *If $k \subseteq K \subseteq L$ are finitely generated field extensions, then*

$$\text{trdeg}_k L = \text{trdeg}_k K + \text{trdeg}_K L.$$

PROOF. Choose a transcendence basis $u_1, \dots, u_r$ of K/k and a transcendence basis $v_1, \dots, v_s$ of L/K . We claim the combined family is algebraically independent over k . A polynomial relation over k among the u and v , viewed as a polynomial in the v with coefficients in $k[u]$, would become a polynomial relation over K among the v after embedding $k[u] \subseteq K$, contradiction. Moreover L is algebraic over $K(v)$ and K is algebraic over $k(u)$, hence L is algebraic over $k(u, v)$. Thus the combined family is a transcendence basis. $\square$

2. Clearing denominators into integrality

LEMMA 3.5 (Integralization after one localization). *Let $B \subseteq A$ be finitely generated integral k -algebras with $B \rightarrow A$ injective. Put*

$$K = \text{Frac}(B), \quad d = \text{trdeg}_K \text{Frac}(A).$$

Then there exist elements $y_1, \dots, y_d \in A$ algebraically independent over K and a nonzero element

$$h \in B[y_1, \dots, y_d]$$

such that

$$A_h$$

is finite and integral over

$$B[y_1, \dots, y_d]_h.$$

PROOF. Choose algebra generators $a_1, \dots, a_n$ of A over B . A maximal algebraically independent subfamily over K has size d ; rename it $y_1, \dots, y_d$. Every remaining generator a_j is algebraic over

$$K(y_1, \dots, y_d).$$

For each such a_j , choose a polynomial equation over $K(y)$ and divide by its leading coefficient to make it monic. The finitely many coefficients are rational functions in the y_i with coefficients in $K = \text{Frac}(B)$. Choose one nonzero

$$h \in B[y_1, \dots, y_d]$$

that clears all their denominators. Then each a_j is integral over $B[y]_h$. Since A_h is generated by the finitely many a_j over that ring, Corollary 2.3 makes A_h a finite module over $B[y]_h$. $\square$

THEOREM 3.6 (Finite chart for a dominant morphism). *Let $f : X \rightarrow Y$ be a dominant regular map of irreducible affine varieties, and put*

$$d = \dim X - \dim Y.$$

Then there are regular functions $y_1, \dots, y_d \in k[X]$ and a nonempty principal open set

$$U = D(h) \subseteq X$$

such that

$$(f, y_1, \dots, y_d) : U \longrightarrow Y \times \mathbb{A}^d$$

is finite onto the principal open subset cut out by the same h in the coordinate algebra $k[Y][y_1, \dots, y_d]$.

PROOF. Apply [Lemma 3.5](#) to the injection

$$k[Y] \hookrightarrow k[X]$$

induced by dominance. The localized coordinate ring $k[X]_h$ is finite integral over

$$k[Y][y_1, \dots, y_d]_h.$$

By the affine dictionary, this is precisely a finite morphism from $D(h) \subseteq X$ onto the principal open subset of $Y \times \mathbb{A}^d$ corresponding to the smaller localized ring. Surjectivity onto that open follows from lying over. $\square$

GUIDING IDEA 3.7. The theorem is the algebraic form of the phrase “a generic fiber has d parameters.” On U , the functions $y_1, \dots, y_d$ account for all transcendental freedom; everything else satisfies monic equations over them. Thus there is only finite algebraic ambiguity once the base point and these d parameters are fixed.

3. A worked finite-chart example

Consider the morphism

$$f : \mathbb{A}^2 \longrightarrow \mathbb{A}^1, \quad f(x, y) = xy.$$

The source has dimension 2, the target dimension 1, so the expected fiber dimension is 1. On the principal open set $D(x)$ introduce the parameter

$$u = y.$$

Then $y = f/u$, and on coordinate rings

$$k[t, u, u^{-1}] \longrightarrow k[x, y, x^{-1}], \quad t \mapsto xy, \quad u \mapsto x,$$

is actually an isomorphism. Thus the finite chart is especially simple here: after adjoining the one free parameter u , there is no remaining algebraic ambiguity at all.

The example also explains why the localization in [Theorem 3.6](#) is essential. Over $x = 0$ the parameter $u = x$ ceases to be usable, and the fiber over $t = 0$ is the reducible union

$$\{x = 0\} \cup \{y = 0\}.$$

A finite chart is therefore a statement about a dense region, not a claim that one coordinate system trivializes every fiber. The generic-fiber theorem will recover the global statement by controlling precisely this lower-dimensional complement.

4. Quasi-finite maps and finite fibers on a dense chart

DEFINITION 3.8. A regular map is *quasi-finite at a point* if that point is isolated in its fiber. For the purposes of this series, a map will be called *generically finite* if it has a dense open subset on which every fiber is finite.

COROLLARY 3.9 (Equal dimension gives generic finiteness). *Let $f : X \rightarrow Y$ be dominant between irreducible affine varieties with*

$$\dim X = \dim Y.$$

Then there is a nonempty open $U \subseteq X$ such that $f|_U$ has finite fibers. More precisely, after shrinking by one principal open, f is a finite morphism onto a nonempty principal open of Y .

PROOF. In Theorem 3.6, $d = 0$, so after localization $k[X]_h$ is finite over $k[Y]_h$. The resulting finite morphism has finite fibers: over a point $y \in Y$, the fiber coordinate ring is a finite-dimensional algebra over the residue field k , hence has only finitely many maximal ideals. $\square$

PROPOSITION 3.10 (Dimension cannot increase under a regular map). *For a regular map $f : X \rightarrow Y$ of affine algebraic sets,*

$$\dim \overline{f(X)} \leq \dim X.$$

If X is irreducible and f is dominant onto Y , then

$$\dim Y \leq \dim X.$$

PROOF. It suffices to treat an irreducible component X_0 of X . Its image closure Z is irreducible, because if $Z = Z_1 \cup Z_2$ with Z_i closed, then

$$X_0 = f^{-1}(Z_1) \cup f^{-1}(Z_2)$$

and irreducibility forces one preimage to be all of X_0 , hence one Z_i contains the dense image. Thus $f : X_0 \rightarrow Z$ is dominant and gives an inclusion

$$k(Z) \hookrightarrow k(X_0),$$

so transcendence degree cannot increase. Taking the maximum over components proves the claim. $\square$

5. Fiber equations

For $y \in Y$, the fiber $f^{-1}(y)$ is an algebraic set. If

$$f = (f_1, \dots, f_m) : X \rightarrow k^m,$$

then

$$f^{-1}(y) = X \cap V(f_1 - y_1, \dots, f_m - y_m).$$

This elementary observation is the bridge between maps and intersections.

PROPOSITION 3.11 (A generic upper bound inside the finite chart). *In the setting of Theorem 3.6, every irreducible component of*

$$f^{-1}(y) \cap U$$

has dimension at most d .

PROOF. The restriction of

$$(f, y_1, \dots, y_d) : U \rightarrow Y \times \mathbb{A}^d$$

is finite. Intersect with the fiber over a fixed $y \in Y$. Base change of an integral extension remains integral by Lemma 2.4; hence the fiber is finite over a locally closed subset of $\mathbb{A}^d$. Any irreducible component therefore has function field algebraic over the function field of its image in $\mathbb{A}^d$, so its transcendence degree is at most d . $\square$

WARNING 3.12. The proposition controls the part of each fiber lying in the dense chart U . The complement $X \setminus U$ has smaller dimension but can contain exceptional high-dimensional fibers. Removing those exceptional pieces uniformly is the purpose of the fiber-dimension induction in the next chapters; one must not simply discard $X \setminus U$ and declare every fiber generic.

CHECKPOINT 3.13. A dominant map $X \rightarrow Y$ with dimension difference d has a dense chart that is finite over $Y \times \mathbb{A}^d$. This is the only genuinely transcendental step behind the fiber-dimension theorem. Everything that remains is Noetherian bookkeeping: identify and control the exceptional locus where the finite chart misses part of a fiber.

Part 2

Dimension, Degree, and Morphisms

CHAPTER 4

Constructible Images and the Generic Fiber Theorem

The finite chart of the preceding chapter contains the generic fiber theorem almost completely. The remaining work is to control the lower-dimensional complement by induction. The same induction proves Chevalley's constructibility theorem. We give both proofs because later growth arguments repeatedly pass between images, exceptional loci, and fibers.

1. Dominant images contain an open set

LEMMA 4.1 (A nonzero parameter polynomial is generically nonzero). *Let Y be irreducible and*

$$0 \neq h \in k[Y][t_1, \dots, t_d].$$

There is a nonempty open subset $Y^\circ \subseteq Y$ such that for every $y \in Y^\circ$ the specialized polynomial

$$h_y(t_1, \dots, t_d)$$

is not identically zero.

PROOF. Write

$$h = \sum_{\alpha} c_{\alpha} t^{\alpha}, \quad c_{\alpha} \in k[Y].$$

At least one coefficient, say c_{α_0} , is nonzero. Take

$$Y^\circ = D(c_{\alpha_0}).$$

Then $c_{\alpha_0}(y) \neq 0$ for every $y \in Y^\circ$, so h_y is nonzero. □

THEOREM 4.2 (A dominant morphism contains an open subset in its image). *Let $f : X \rightarrow Y$ be a dominant regular map of irreducible affine varieties. Then $f(X)$ contains a nonempty Zariski-open subset of Y .*

PROOF. Use the finite chart of Theorem 3.6. There are functions $y_1, \dots, y_d$ and $0 \neq h \in k[Y][y_1, \dots, y_d]$ such that

$$U = D(h) \subseteq X$$

is finite and surjective over

$$W = D(h) \subseteq Y \times \mathbb{A}^d.$$

By Lemma 4.1, after restricting to a nonempty open $Y^\circ \subseteq Y$, each specialization h_y is a nonzero polynomial. Since k is infinite, $D(h_y) \subseteq \mathbb{A}^d$ is nonempty. Choose $t \in D(h_y)$. Then $(y, t) \in W$, and surjectivity of $U \rightarrow W$ produces $x \in U$ with $f(x) = y$. Hence

$$Y^\circ \subseteq f(X).$$

□

2. Generic versus exceptional fibers: a laboratory

The map

$$f : \mathbb{A}^2 \rightarrow \mathbb{A}^1, \quad (x, y) \mapsto xy,$$

already contains the complete geometry of the generic-fiber theorem in miniature. For $t \neq 0$,

$$f^{-1}(t) = \{(x, t/x) : x \neq 0\} \cong \mathbb{G}_m$$

is irreducible of dimension 1. The exceptional fiber is

$$f^{-1}(0) = V(xy) = V(x) \cup V(y),$$

which is still one-dimensional but has two components. Thus generic fiber dimension is stable even when geometry such as irreducibility jumps.

A slightly different map shows why exceptional fibers can have larger dimension. Let

$$F : \mathbb{A}^3 \rightarrow \mathbb{A}^2, \quad F(x, y, z) = (x, xy).$$

For a point (a, b) with $a \neq 0$, the equations force $x = a$, $y = b/a$ and leave z free, so the fiber has dimension 1. Over $(0, 0)$, however, the equation $xy = 0$ becomes automatic once $x = 0$, and both y and z are free; the fiber has dimension 2. The exceptional locus is the proper algebraic set $\{(0, 0)\}$ in the target. This is exactly the pattern formalized below: the expected dimension holds on a nonempty open subset, while larger fibers are confined to lower-dimensional algebraic data.

3. Constructible sets

DEFINITION 4.3. A subset of an algebraic set is *locally closed* if it is the intersection of an open and a closed set. A set is *constructible* if it is a finite union of locally closed sets.

LEMMA 4.4 (Elementary constructible calculus). *Finite unions, finite intersections, and complements of constructible subsets are constructible.*

PROOF. A locally closed set can be written $Z \setminus W$ with $W \subseteq Z$ closed. Intersections of two such differences have the same form after replacing Z by an intersection and W by a finite union. Thus finite intersections and unions of finite unions of locally closed sets are constructible. For complements,

$$X \setminus (Z \setminus W) = (X \setminus Z) \cup W$$

is constructible; De Morgan's laws finish the proof for finite unions. $\square$

THEOREM 4.5 (Chevalley's theorem, affine form). *Let $f : X \rightarrow Y$ be a regular map of affine algebraic sets. If $C \subseteq X$ is constructible, then $f(C)$ is constructible in Y .*

PROOF. Every constructible set is a finite union of irreducible locally closed sets, so it is enough to treat

$$C = O \subseteq Z,$$

where Z is irreducible closed in X and O is a nonempty open subset of Z . We prove the assertion for such pairs by Noetherian induction on the irreducible closed set Z .

Let

$$Y_0 = \overline{f(O)} = \overline{f(Z)}.$$

The equality holds because O is dense in Z and inverse images of closed sets are closed. Choose a nonempty principal affine open $O_0 \subseteq O$. Its image is still dense in Y_0 : if $\overline{f(O_0)} \subsetneq Y_0$, then the proper closed set $f^{-1}(\overline{f(O_0)}) \cap Z$ would contain the nonempty open set O_0 and hence, by irreducibility, all of Z , contradicting the definition of Y_0 .

Thus the restricted map

$$f : O_0 \rightarrow Y_0$$

is dominant. By Theorem 4.2, its image contains a nonempty open subset $U \subseteq Y_0$. In particular

$$U \subseteq f(O).$$

Set

$$C' = O \cap f^{-1}(Y_0 \setminus U).$$

This is closed in the locally closed set O . Its closure Z' in Z is proper: otherwise the closed set $f^{-1}(Y_0 \setminus U) \cap Z$ would contain the dense subset O and hence all of Z , forcing $f(Z)$ to miss U .

Now C' is constructible inside the proper closed set Z' . By the induction hypothesis, $f(C')$ is constructible. Since every point of $f(O)$ either lies in U or has a preimage in C' , we have

$$f(O) = U \cup f(C').$$

Hence $f(O)$ is constructible. Finite unions of these images are constructible by Lemma 4.4. $\square$

4. The generic fiber theorem

THEOREM 4.6 (Generic fiber dimension). *Let $f : X \rightarrow Y$ be a dominant regular map of irreducible affine varieties, and put*

$$d = \dim X - \dim Y.$$

There exists a nonempty open subset $Y^\circ \subseteq Y$ such that for every $y \in Y^\circ$:

- (i) the fiber $X_y = f^{-1}(y)$ is nonempty;
- (ii) every irreducible component of X_y has dimension at most d ;
- (iii) at least one irreducible component of X_y has dimension exactly d .

Consequently

$$\dim X_y = d$$

for every $y \in Y^\circ$.

PROOF. We induct on $\dim X$. Apply [Theorem 3.6](#): there is a nonempty principal open

$$U = D(h) \subseteq X$$

and a finite surjective map

$$(f, y_1, \dots, y_d) : U \rightarrow W = D(h) \subseteq Y \times \mathbb{A}^d.$$

By [Lemma 4.1](#), after replacing Y by a nonempty open set Y_0 , every specialized polynomial h_y is nonzero. Hence

$$W_y = D(h_y) \subseteq \mathbb{A}^d$$

is a nonempty open set of dimension d . The fiber

$$U_y = U \cap X_y$$

is finite and surjective over W_y . Integrality preserves transcendence degree: if C is an irreducible component of U_y dominating an irreducible component of $W_y = W_y$ itself, then its coordinate ring is integral over the coordinate ring of its image, so

$$\dim C = d.$$

Every irreducible component of U_y has dimension at most d by [Proposition 3.11](#). Thus the theorem is true for the portion of the fiber meeting U .

It remains to control components contained in

$$E = X \setminus U.$$

By [Corollary 2.16](#), every irreducible component E_i of E has dimension at most $\dim X - 1$.

If the restriction $f|_{E_i}$ is not dominant, its image closure is a proper closed subset of Y ; remove that closure from the base. If it is dominant, apply the induction hypothesis to

$$f|_{E_i} : E_i \rightarrow Y.$$

Its generic fiber dimension is

$$\dim E_i - \dim Y \leq d - 1.$$

There are only finitely many E_i , so intersect the finitely many nonempty open subsets supplied by these arguments with Y_0 . The resulting Y° is nonempty because Y is irreducible.

For $y \in Y^\circ$, every component of X_y either meets U , in which case its intersection with U is a nonempty open subset of that component and has dimension at most d , or is contained in E and hence has dimension at most $d - 1$. The finite cover $U_y \rightarrow W_y$ supplies at least one d -dimensional component. This proves all three assertions. $\square$

COROLLARY 4.7 (Dimension formula for a map). *For a dominant regular map of irreducible affine varieties,*

$$\dim X = \dim Y + \dim X_y$$

for all y in a nonempty open subset of Y .

PROOF. This is exactly [Theorem 4.6](#) rewritten. $\square$

CHECKPOINT 4.8. We have proved, rather than quoted, the two geometric engines most often hidden behind the words “generic” and “exceptional”: constructible images and exact generic fiber dimension. The next chapters turn the finite charts into counting inequalities and isolate the bounded exceptional sets needed for growth.

CHAPTER 5

Hilbert Functions and Degree

Hilbert functions provide a precise bridge between algebraic dimension and polynomial growth. We develop only the portion of projective degree theory that is genuinely used later, and we deliberately avoid importing a global Bézout theorem. Uniform fiber constants in the growth arguments will come instead from finite-module algebra.

1. Graded algebras and Hilbert functions

Let

$$R = \bigoplus_{n \geq 0} R_n$$

be a finitely generated graded k -algebra with $R_0 = k$ and generated by R_1 . Its Hilbert function is

$$H_R(n) = \dim_k R_n.$$

LEMMA 5.1 (Exact sequence for multiplication). *Let $x \in R$ be homogeneous of degree e . Then there is an exact sequence of graded modules*

$$0 \rightarrow (0 :_R x)(-e) \rightarrow R(-e) \xrightarrow{\cdot x} R \rightarrow R/xR \rightarrow 0.$$

If x is a non-zero-divisor, then

$$H_{R/xR}(n) = H_R(n) - H_R(n - e).$$

PROOF. The kernel of multiplication by x consists exactly of the annihilator $(0 :_R x)$, with the grading shifted by e . The cokernel is R/xR . Taking the degree- n piece gives the Hilbert-function identity when the kernel vanishes. $\square$

THEOREM 5.2 (Hilbert–Serre). *Let R be a finitely generated graded k -algebra generated in degree one. Its Hilbert series*

$$\text{Hilb}_R(t) = \sum_{n \geq 0} H_R(n)t^n$$

has the form

$$\frac{P(t)}{(1-t)^N}$$

for some $P(t) \in \mathbb{Z}[t]$ and some N . Consequently there is a polynomial $P_R(n) \in \mathbb{Q}[n]$ such that

$$H_R(n) = P_R(n)$$

for all sufficiently large n .

PROOF. Present

$$R = k[x_1, \dots, x_N]/I$$

with each x_i of degree one. We prove the rational-form statement more generally for every finitely generated graded module over the polynomial ring, by induction on N .

For $N = 0$, such a module is finite-dimensional over k , so its Hilbert series is a polynomial. For $N > 0$, let $x = x_N$. For a finitely generated graded module M , the exact sequence

$$0 \rightarrow (0 :_M x)(-1) \rightarrow M(-1) \xrightarrow{x} M \rightarrow M/xM \rightarrow 0$$

gives

$$(1 - t) \text{Hilb}_M(t) = \text{Hilb}_{M/xM}(t) - t \text{Hilb}_{(0:Mx)}(t).$$

Both modules on the right are finitely generated over $k[x_1, \dots, x_{N-1}]$: finite generation of the annihilator uses Noetherianity. By induction each has denominator a power of $1 - t$, hence so does M .

Finally,

$$\frac{1}{(1 - t)^r} = \sum_{n \geq 0} \binom{n + r - 1}{r - 1} t^n.$$

A polynomial numerator therefore makes the coefficient of t^n a finite integer combination of shifted binomial polynomials for all sufficiently large n . This is the required Hilbert polynomial. $\square$

2. Homogeneous normalization and the degree of the Hilbert polynomial

LEMMA 5.3 (Homogeneous Noether normalization). *Let k be infinite and R a finitely generated graded integral k -algebra generated by R_1 . If $d = \text{trdeg}_k \text{Frac}(R)$, then there exist algebraically independent linear forms*

$$\ell_1, \dots, \ell_d \in R_1$$

such that R is finite over $k[\ell_1, \dots, \ell_d]$.

PROOF. We induct on the number N of degree-one generators. If their images are algebraically independent, they already form the required normalization.

Otherwise choose a nonzero homogeneous relation

$$F(x_1, \dots, x_N) = 0$$

of degree $e > 0$. Because k is infinite, there are scalars $c_1, \dots, c_{N-1}$ for which

$$F(c_1, \dots, c_{N-1}, 1) \neq 0.$$

Make the linear change

$$x_i = y_i + c_i y_N \quad (i < N), \quad x_N = y_N.$$

Homogeneity shows that the coefficient of y_N^e in the transformed relation is the displayed nonzero scalar. After dividing by that scalar the relation is monic in y_N . Hence y_N is integral over

$$R_0 = k[y_1, \dots, y_{N-1}] \subseteq R,$$

and R is finite over R_0 .

Apply the induction hypothesis to the standard graded domain R_0 , omitting any redundant degree-one generators. We obtain algebraically independent linear forms $\ell_1, \dots, \ell_d \in (R_0)_1$ such that R_0 is finite over $k[\ell_1, \dots, \ell_d]$. Transitivity makes R finite over the same polynomial algebra. Finite integral extensions preserve transcendence degree, so the number of parameters is exactly $d = \text{trdeg}_k \text{Frac}(R)$. $\square$

PROPOSITION 5.4 (Growth degree of the Hilbert function). *Let R be as in Lemma 5.3, of transcendence degree $d \geq 1$. Then its Hilbert polynomial has degree $d - 1$ and positive leading coefficient.*

PROOF. Let

$$S = k[\ell_1, \dots, \ell_d] \subseteq R$$

be a homogeneous Noether normalization. Since R is finite over S , choose homogeneous S -module generators $m_1, \dots, m_r$ with degrees e_i . Then for every n ,

$$H_R(n) \leq \sum_{i=1}^r H_S(n - e_i) = O(n^{d-1}).$$

On the other hand, $S_n \subseteq R_n$, so

$$H_R(n) \geq H_S(n) = \binom{n+d-1}{d-1} \gg n^{d-1}.$$

Since H_R agrees eventually with a polynomial, that polynomial has degree exactly $d - 1$ and positive leading coefficient. $\square$

3. Projective degree

For a nonempty irreducible projective variety $X \subseteq \mathbb{P}^N$ with homogeneous coordinate domain R_X , define

$$\dim X = \text{trdeg}_k \text{Frac}(R_X) - 1.$$

This agrees with the affine definition on every nonempty standard chart. If x_i does not vanish identically on X , the coordinate ring of $X \cap D_+(x_i)$ is

$$A = (R_X[x_i^{-1}])_0.$$

Every homogeneous fraction is a power of x_i times a fraction of degree zero, so

$$\text{Frac}(R_X) = \text{Frac}(A)(x_i).$$

The grading shows that x_i is transcendental over $\text{Frac}(A)$: a polynomial relation with degree-zero coefficients would decompose into distinct homogeneous degrees and hence have all coefficients zero. Thus

$$\text{trdeg}_k \text{Frac}(R_X) = \text{trdeg}_k \text{Frac}(A) + 1.$$

The preceding proposition therefore gives a Hilbert polynomial of degree $\dim X$.

DEFINITION 5.5 (Degree). Let $X \subseteq \mathbb{P}^N$ be an irreducible projective variety of dimension r . Write its Hilbert polynomial as

$$P_X(n) = \frac{\deg(X)}{r!} n^r + O(n^{r-1}).$$

The positive rational number $\deg(X)$ determined by this coefficient is the *Hilbert degree* of X . No integrality assertion is needed in the growth arguments developed in this series; numerical fiber bounds will come from finite-module rank instead.

For a projective algebraic set with irreducible components X_i , define its *total degree*

$$\delta(X) = \sum_i \deg(X_i).$$

For an affine algebraic set $X \subseteq \mathbb{A}^N$, define $\delta(X)$ to be the total degree of its projective closure.

REMARK 5.6 (Why we stop short of a global Bézout theorem). The exact sequence [Lemma 5.1](#) already explains the scheme-theoretic principle that a non-zero-divisor of degree e lowers the Hilbert-polynomial degree by one and multiplies the leading coefficient by e . Turning this into a uniform statement about *all* set-theoretic irreducible components requires a principal-ideal input. Appendix A proves that input, derives purity of hypersurface sections, and obtains the coarse reduced Bézout bound used only for the later bounded-complexity package. The escape argument itself will remain independent of Bézout.

COROLLARY 5.7 (Zero-dimensional degree is point count). *If $X \subseteq \mathbb{P}^N$ is a zero-dimensional algebraic set, then*

$$|X(k)| = \delta(X).$$

PROOF. Let Z be an irreducible zero-dimensional affine chart of one component. Its coordinate ring is a finitely generated domain of transcendence degree 0. Noether normalization says it is finite over k , hence an integral domain integral over the field k ; by [Lemma 2.5](#) it is a field. Since k is algebraically closed, that field is k . Thus the component is a single k -point and has eventual Hilbert function 1.

For a finite union of distinct points, the elementary exact sequence

$$0 \rightarrow S/(I \cap J) \rightarrow S/I \oplus S/J \rightarrow S/(I + J) \rightarrow 0$$

shows inductively that constant Hilbert polynomials add, since distinct points have empty projective intersection. Hence the total degree is exactly the number of points. $\square$

4. What degree will and will not do for us

The Hilbert-function calculation above records dimension and the leading multiplicity of a hypersurface section, and in dimension zero it gives an exact point count. We will *not* use Bézout as a hidden engine in the escape argument. Escape in [Chapter 8](#) is instead proved by a finite-dimensional equation-space descent that is elementary for linear actions. Appendix A nevertheless proves the principal-ideal and coarse reduced Bézout statements needed to make the uniform-complexity appendix self-contained. Whenever a later argument needs a uniform finite-fiber bound for a fixed morphism, the module-rank theorem of the next chapter supplies it directly.

This separation is deliberate. It keeps the logical core of product growth independent of deeper equidimensionality and intersection-theoretic statements that are unnecessary for the Bourgain–Gamburd mechanism.

CHECKPOINT 5.8. Hilbert polynomials now give a rigorous dimension-and-multiplicity dictionary. The actual counting constants used in product growth will come from finite modules, not from an unproved global Bézout estimate. The next chapter turns finite charts into finite-set inequalities.

CHAPTER 6

Quantitative Finite Maps and Slicing

The generic fiber theorem becomes combinatorially useful only after one extracts a uniform bound on the finite fibers of a fixed algebraic map. Finite-module algebra gives exactly such a bound. The resulting slicing inequalities are the geometric counting tools used in Larsen–Pink arguments.

1. Fiber size of a finite morphism

PROPOSITION 6.1 (Module rank bounds geometric fibers). *Let $A \subseteq B$ be finitely generated k -algebras, with B generated by r elements as an A -module. Let*

$$\pi : X = \operatorname{Spec} B(k) \rightarrow Y = \operatorname{Spec} A(k)$$

be the corresponding affine morphism. Then every geometric fiber contains at most r points.

PROOF. Fix $y \in Y$ with maximal ideal $\mathfrak{m}_y \subseteq A$. The fiber coordinate algebra is

$$B_y = B/\mathfrak{m}_y B.$$

The images of the r module generators span B_y over

$$A/\mathfrak{m}_y \cong k,$$

so

$$\dim_k B_y \leq r.$$

Distinct k -points of the fiber correspond to distinct maximal ideals of B_y . If there are s such points, the evaluation map

$$B_y \longrightarrow k^s$$

is surjective by the Chinese remainder theorem, since distinct maximal ideals are comaximal. Therefore

$$s \leq \dim_k B_y \leq r.$$

□

LEMMA 6.2 (Chinese remainder theorem for pairwise comaximal ideals). *If $I_1, \dots, I_s$ are pairwise comaximal ideals of a ring R , then the natural map*

$$R \rightarrow \prod_{i=1}^s R/I_i$$

is surjective with kernel $\bigcap_i I_i$.

PROOF. For $s = 2$, choose $a \in I_1$, $b \in I_2$ with $a + b = 1$. Given residue classes x_1, x_2 , the element

$$x_1 b + x_2 a$$

has the desired two residues. The kernel is visibly $I_1 \cap I_2$. Induct on s , observing that pairwise comaximality implies

$$I_s + \bigcap_{i < s} I_i = R$$

because if $a_i + b_i = 1$ with $a_i \in I_i$, $b_i \in I_s$, then

$$\prod_{i < s} a_i \in \bigcap_{i < s} I_i$$

and its difference from 1 lies in I_s . □

2. Generic finite-fiber bounds

THEOREM 6.3 (Uniform generic finite fibers). *Let $f : X \rightarrow Y$ be a dominant regular map of irreducible affine varieties with*

$$\dim X = \dim Y.$$

Then there exist a proper closed subset $E \subsetneq X$ and an integer $M \geq 1$ such that

$$|f^{-1}(y) \setminus E| \leq M$$

for every $y \in Y$.

PROOF. By Corollary 3.9, after restricting to a principal open

$$U = X \setminus E$$

the map is finite. On coordinate rings, $k[U]$ is generated by finitely many elements, say M , as a module over the coordinate ring of the corresponding open subset of Y . Apply Proposition 6.1. Fibers outside the image open are empty in U , so the same bound holds for all y . □

COROLLARY 6.4 (Finite-set image inequality). *In the setting of Theorem 6.3, every finite set $A \subseteq X(k)$ satisfies*

$$|f(A)| \geq \frac{|A| - |A \cap E|}{M}.$$

PROOF. At most M points of $A \setminus E$ can map to the same image point. Therefore

$$|A \setminus E| \leq M|f(A)|.$$

Rearrange. □

GUIDING IDEA 6.5. This simple inequality is one of the hidden engines of product growth. A map of equal-dimensional varieties either sends a finite set to many values, or a substantial part of the finite set lies in the exceptional subvariety E . Escape from subvarieties will later rule out the second alternative after boundedly many group operations.

3. Adding the missing parameters

THEOREM 6.6 (Slicing a dominant map to a generically finite one). *Let $f : X \rightarrow Y$ be dominant between irreducible affine varieties and set*

$$d = \dim X - \dim Y.$$

There exist regular functions $\ell_1, \dots, \ell_d \in k[X]$, a proper closed subset $E \subsetneq X$, and $M \geq 1$ such that the map

$$F = (f, \ell_1, \dots, \ell_d) : X \rightarrow Y \times \mathbb{A}^d$$

satisfies

$$|F^{-1}(z) \setminus E| \leq M$$

for every $z \in Y \times \mathbb{A}^d$.

PROOF. Take the functions supplied by the finite-chart theorem [Theorem 3.6](#). On the principal open $U = X \setminus E$, the map F is finite. Apply [Proposition 6.1](#) to obtain a uniform fiber bound M . □

COROLLARY 6.7 (Dimension-counting inequality for finite sets). *With the notation of [Theorem 6.6](#), for finite $A \subseteq X(k)$,*

$$|A| - |A \cap E| \leq M|f(A)| \prod_{i=1}^d |\ell_i(A)|.$$

PROOF. The image $F(A)$ is contained in

$$f(A) \times \ell_1(A) \times \dots \times \ell_d(A).$$

The fiber bound gives

$$|A| - |A \cap E| \leq M|F(A)|,$$

and the displayed containment gives the desired upper bound on $|F(A)|$. □

4. A product-map laboratory

EXAMPLE 6.8 (Transverse multiplication). Let G be an irreducible algebraic group and $V, W \subseteq G$ irreducible subvarieties. Consider

$$m : V \times W \rightarrow G, \quad (v, w) \mapsto vw.$$

Suppose

$$\dim \overline{VW} = \dim V + \dim W.$$

Then $m : V \times W \rightarrow \overline{VW}$ has equal-dimensional source and target, so Theorem 6.3 yields a proper exceptional set $E \subsetneq V \times W$ and M such that

$$|\{(v, w) \in (A \times B) \setminus E : vw = g\}| \leq M$$

for all finite $A \subseteq V$, $B \subseteq W$ and all $g \in G$. Thus

$$|AB| \geq \frac{|A||B| - |(A \times B) \cap E|}{M}.$$

The only obstruction to multiplicative growth is concentration on the exceptional subvariety E .

5. From a finite map to a finite-set inequality

The counting corollaries are easiest to remember through a concrete model. Suppose

$$f : X \rightarrow Y$$

is generically two-to-one outside an exceptional subvariety $E \subsetneq Y$. For a finite set $B \subseteq X$ write

$$B_{\text{good}} = B \setminus f^{-1}(E).$$

Every point of $f(B_{\text{good}})$ has at most two preimages in B_{good} , so

$$|B_{\text{good}}| \leq 2|f(B)|.$$

The only way this inequality fails to control $|B|$ is if a substantial part of B lies in the lower-dimensional exceptional inverse image $f^{-1}(E)$. That is precisely the dichotomy exploited in Larsen–Pink induction:

many good points $\Rightarrow$ large image

or

many bad points $\Rightarrow$ dimension drop

.

The technical statements of this chapter are quantitative versions of this one elementary pigeonhole principle.

CHECKPOINT 6.9. The phrase “generically finite” has now been converted into an actual counting inequality with a bounded multiplicity constant, obtained from a finite module rather than from an unproved global Bézout estimate. The remaining geometric task is to force finite sets away from exceptional subvarieties; that is the purpose of escape.

Part 3

Algebraic Groups and Growth Geometry

CHAPTER 7

Algebraic Groups, Actions, and Orbit Dimension

Growth takes place inside algebraic groups, where multiplication and conjugation turn geometry into combinatorics. This chapter develops the structural facts needed for that translation: irreducible components are cosets, stabilizers are algebraic, orbits are locally closed, and orbit dimension is group dimension minus stabilizer dimension. Each statement is proved from the affine geometry already developed.

1. Affine algebraic groups

DEFINITION 7.1. An *affine algebraic group* over k is an affine algebraic set G equipped with group operations such that

$$m : G \times G \rightarrow G, \quad (g, h) \mapsto gh,$$

and

$$\iota : G \rightarrow G, \quad g \mapsto g^{-1},$$

are regular maps.

Matrix groups cut out by polynomial equations are the basic examples. For instance, SL_n is the hypersurface $\det X = 1$ in $\mathbb{A}^{n^2}$, and inversion is polynomial on it because

$$X^{-1} = \mathrm{adj}(X)$$

when $\det X = 1$.

PROPOSITION 7.2 (The identity irreducible component is a subgroup). *Let G be an affine algebraic group and let G° denote the irreducible component containing the identity. Then G° is a closed normal subgroup of G . Every irreducible component of G is a left and right coset of G° . In particular, G/G° is finite.*

PROOF. The product $G^\circ \times G^\circ$ is irreducible: apply the product-dimension/product-irreducibility statement established in [Proposition 2.14](#). Its image under multiplication is irreducible by the argument in [Proposition 3.10](#): the continuous image of an irreducible space is irreducible. This image contains the identity and therefore lies in the unique maximal irreducible subset G° containing the identity. Thus $G^\circ G^\circ \subseteq G^\circ$.

Inversion maps G° homeomorphically to an irreducible component containing the identity, hence to itself. Therefore G° is a subgroup.

For $g \in G$, left translation L_g and right translation R_g are algebraic automorphisms. Hence gG° and $G^\circ g$ are irreducible components containing g , so they are equal to the irreducible component of g . Thus

$$gG^\circ = G^\circ g$$

for every g , proving normality. Finiteness of the component set follows from Theorem 1.13. $\square$

WARNING 7.3. Many sources write G° for the connected component of the identity. For reduced affine algebraic groups over an algebraically closed field, connectedness and irreducibility of the identity component agree. The present series does not need that additional theorem: we define G° by irreducibility and use only the properties proved above.

2. Algebraic actions and stabilizers

DEFINITION 7.4. An *algebraic action* of an affine algebraic group G on an affine algebraic set X is a regular map

$$\alpha : G \times X \rightarrow X, \quad (g, x) \mapsto g \cdot x,$$

satisfying the usual action identities.

PROPOSITION 7.5 (Stabilizers are algebraic). *For an algebraic action and a point $x \in X$, the stabilizer*

$$G_x = \{g \in G : g \cdot x = x\}$$

is a closed algebraic subgroup of G .

PROOF. The orbit map

$$\phi_x : G \rightarrow X, \quad g \mapsto g \cdot x$$

is regular. Since the point $\{x\}$ is closed over an algebraically closed field,

$$G_x = \phi_x^{-1}(\{x\})$$

is closed. It is a subgroup by the action identities. $\square$

PROPOSITION 7.6 (Fibers of the orbit map). *For $g, h \in G$,*

$$\phi_x(g) = \phi_x(h) \iff h^{-1}g \in G_x.$$

Thus every nonempty fiber of ϕ_x is a left coset of G_x , hence has dimension $\dim G_x$.

PROOF. The equality $g \cdot x = h \cdot x$ is equivalent, after applying h^{-1} , to

$$(h^{-1}g) \cdot x = x.$$

Therefore the fiber over $g \cdot x$ is exactly gG_x . Translation is an algebraic isomorphism, so all such fibers have the same dimension as G_x . $\square$

3. Orbits are locally closed

THEOREM 7.7 (Local closedness of algebraic orbits). *Let an irreducible affine algebraic group G act algebraically on X . For every $x \in X$, the orbit*

$$G \cdot x$$

is a locally closed irreducible subset of X .

PROOF. Let

$$Z = \overline{G \cdot x}.$$

The orbit map $G \rightarrow Z$ is dominant. By [Theorem 4.2](#), its image contains a nonempty open subset $U \subseteq Z$.

We claim the entire orbit is open in Z . Pick $u = h \cdot x \in U$. For an arbitrary orbit point $y = g \cdot x$, the group element gh^{-1} carries u to y . The action of gh^{-1} is an algebraic automorphism of X preserving both the orbit and its closure Z , so

$$(gh^{-1}) \cdot U$$

is an open subset of Z , contained in the orbit and containing y . Thus every orbit point has an open neighbourhood in Z contained in the orbit. Hence $G \cdot x$ is open in Z and therefore locally closed. It is irreducible as the continuous image of the irreducible variety G . $\square$

THEOREM 7.8 (Orbit–stabilizer dimension formula). *Let G be irreducible and act algebraically on X . Then for every $x \in X$,*

$$\dim(G \cdot x) = \dim G - \dim G_x.$$

PROOF. The orbit map

$$\phi_x : G \rightarrow G \cdot x$$

is dominant onto the irreducible locally closed variety $G \cdot x$. Replace the orbit by an affine open subset and its preimage if necessary; dimensions do not change. Every fiber is a coset of G_x by [Proposition 7.6](#), hence has dimension exactly $\dim G_x$. The generic fiber theorem [Theorem 4.6](#) gives

$$\dim G = \dim(G \cdot x) + \dim G_x.$$

Rearrange. $\square$

4. Centralizers and conjugacy classes

Let G act on itself by conjugation.

PROPOSITION 7.9 (Centralizers and conjugacy dimension). *For $g \in G$, the centralizer*

$$C_G(g) = \{x \in G : xg = gx\}$$

is a closed algebraic subgroup. If G is irreducible, the conjugacy class $C(g)$ is locally closed and

$$\dim C(g) = \dim G - \dim C_G(g).$$

PROOF. The centralizer is the stabilizer of g under conjugation, so it is closed by Proposition 7.5. The remaining statements are Theorems 7.7 and 7.8 applied to the conjugation action. $\square$

PROPOSITION 7.10 (Centralizer of a subset). *For any subset $S \subseteq G$, the common centralizer*

$$C_G(S) = \{g \in G : gs = sg \text{ for every } s \in S\}$$

is a closed algebraic subgroup.

PROOF. For each $s \in S$, $C_G(s)$ is closed. Arbitrary intersections of closed subsets are closed, hence

$$C_G(S) = \bigcap_{s \in S} C_G(s)$$

is closed. It is visibly a subgroup. $\square$

5. Stabilizers of subvarieties

For escape arguments we also need the subgroup preserving a closed set.

PROPOSITION 7.11 (Setwise stabilizer as an algebraic subgroup). *Let G act algebraically and linearly on an affine space $\mathbb{A}^N$, and let $V \subseteq \mathbb{A}^N$ be an algebraic set. Then the setwise stabilizer*

$$\text{Stab}_G(V) = \{g \in G : gV = V\}$$

is a closed algebraic subgroup.

PROOF. Let $I(V) \subseteq k[x_1, \dots, x_N]$ be the radical ideal of V . By the Hilbert basis theorem choose generators $f_1, \dots, f_r$. Choose D so large that all f_i have degree at most D . Because the action is linear, pullback by g preserves the finite-dimensional vector space

$$P_D = \{\text{polynomials of degree at most } D\}.$$

The degree- $\leq D$ part

$$I_D = I(V) \cap P_D$$

is a linear subspace of P_D that contains the chosen generators and hence determines $I(V)$ as an ideal.

The condition $gV = V$ is equivalent to

$$g^*I(V) = I(V),$$

and because g^* preserves degree and I_D generates $I(V)$, it is equivalent to

$$g^*I_D = I_D.$$

Choose a basis of P_D adapted to I_D . The matrix of $g^* : P_D \rightarrow P_D$ has entries regular in g because the original action is linear. The condition that I_D be invariant is the vanishing of the matrix block from I_D to a chosen complement; applying the same condition to g^{-1} gives equality rather than mere inclusion. Hence the stabilizer is cut out by polynomial equations inside G . It is a subgroup by definition. $\square$

REMARK 7.12. The linearity hypothesis is sufficient for the matrix groups used later: conjugation and left/right multiplication become linear actions on suitable matrix spaces once one variable is fixed. A general algebraic-action version can be proved by replacing P_D with a sufficiently large finite-dimensional generating subspace, but we will not need that extra abstraction.

CHECKPOINT 7.13. The dimension identities used informally in rank one are now genuine theorems. In particular,

$$\dim(\text{conjugacy class}) = \dim G - \dim(\text{centralizer}).$$

The next chapter turns the same group action into an escape mechanism: if a finite generating set remains inside a proper subvariety under many short words, the setwise stabilizer must grow until it becomes all of the ambient group.

CHAPTER 8

Escape from Proper Subvarieties

A Zariski-dense orbit cannot remain trapped in a proper algebraic condition. Product growth needs a uniform finite-step version. For linear actions there is an elementary proof that avoids any hidden Bézout theorem: all translates of the defining equations remain in one finite-dimensional polynomial space, and every strict intersection adds a genuinely new equation. This finite-dimensional descent gives a bounded escaping list and then a positive-proportion escape theorem.

Throughout this chapter $G \subseteq \mathrm{GL}_N(k)$ is an affine algebraic group over an algebraically closed field k , acting linearly on $\mathbb{A}^m$. Let $A \subseteq G(k)$ be finite, symmetric, and contain the identity; write $\Gamma = \langle A \rangle$.

1. The bounded-degree equation space

Fix an algebraic set

$$V = V(f_1, \dots, f_s) \subseteq \mathbb{A}^m$$

with $\deg f_i \leq D$. Let

$$\mathcal{P}_D = k[x_1, \dots, x_m]_{\leq D}$$

be the finite-dimensional vector space of polynomials of total degree at most D . For a linear transformation g of $\mathbb{A}^m$, pullback $g^*f = f \circ g$ preserves $\mathcal{P}_D$.

For a finite word set $W \subseteq \Gamma$, define

$$L(W) = \mathrm{span}_k \{g^*f_i : g \in W, 1 \leq i \leq s\} \subseteq \mathcal{P}_D$$

and

$$V(W) = V(L(W)) = \bigcap_{g \in W} g^{-1}V.$$

Only the vector space $L(W)$ matters for the zero set.

LEMMA 8.1 (Strict intersection adds an independent bounded-degree equation).
Let $W \subseteq \Gamma$ be finite and $a \in \Gamma$. If

$$V(W \cup Wa) \subsetneq V(W),$$

then

$$\dim_k L(W \cup Wa) > \dim_k L(W).$$

Consequently a chain of strict intersections obtained by adjoining translates of the original equations has length at most

$$\dim_k \mathcal{P}_D = \binom{m+D}{D}.$$

PROOF. If $L(W \cup Wa) = L(W)$, the two spaces impose exactly the same equations and hence have the same zero set. Thus strict containment of zero sets forces strict containment of equation spaces. Every strict inclusion raises dimension by at least one, and all spaces lie in $\mathcal{P}_D$. $\square$

GUIDING IDEA 8.2. The usual escape proof is phrased as dimension–degree induction. For a linear action there is a simpler uniformity mechanism: there are only finitely many independent equations of degree at most D . Every genuine tightening of the trap spends one unit of this finite equation budget.

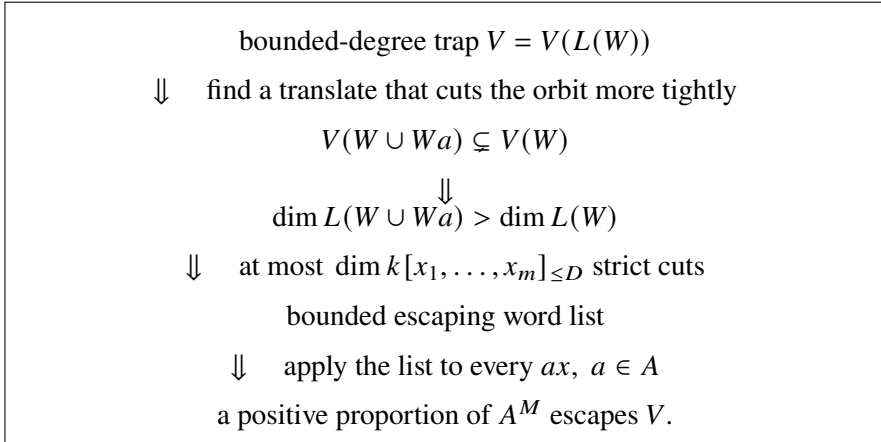


FIGURE 1. The finite-dimensional mechanism behind escape from subvarieties. Uniformity comes from spending one independent bounded-degree equation at every strict cut.

2. A bounded list of escaping words

THEOREM 8.3 (Finite-list escape). *Let $V \subseteq \mathbb{A}^m$ be defined by polynomials of degree at most D , and let O be a Γ -orbit not contained in V . There exist integers $r, \ell \geq 1$, depending only on m, D and the chosen bounded-degree presentation of V , and elements*

$$g_1, \dots, g_\ell \in A^r$$

such that for every $x \in O$, at least one $g_i x$ lies outside V .

PROOF. We prove a recursive assertion for a finite $W \subseteq \Gamma$ containing the identity. Assume $\mathcal{O} \not\subseteq V(W)$. Set

$$c(W) = \dim \mathcal{P}_D - \dim L(W).$$

If $V(W) \cap \mathcal{O} = \emptyset$, the identity word works.

Otherwise choose $x_0 \in V(W) \cap \mathcal{O}$. Since the orbit is not contained in $V(W)$, take a shortest word $b_t \cdots b_1$ carrying x_0 outside $V(W)$. At the penultimate point

$$y = b_{t-1} \cdots b_1 x_0 \in V(W)$$

the final letter $a = b_t$ satisfies $ay \notin V(W)$. Therefore

$$y \in V(W) \setminus a^{-1}V(W).$$

Here the order of multiplication matters. Since

$$V(W) = \bigcap_{w \in W} w^{-1}V,$$

we have

$$a^{-1}V(W) = \bigcap_{w \in W} (wa)^{-1}V = V(Wa).$$

Thus the correct enlarged word set is the *right*-extension $W \cup Wa$, and

$$V(W \cup Wa) = V(W) \cap a^{-1}V(W) \subsetneq V(W).$$

By Lemma 8.1, $c(W \cup Wa) < c(W)$.

Apply recursion to $W' = W \cup Wa$. We obtain words $h_1, \dots, h_t$ so that every orbit point is sent by some h_j outside

$$V(W') = V(W) \cap a^{-1}V(W).$$

For that j , either $h_j x \notin V(W)$, or $h_j x \notin a^{-1}V(W)$ and hence $ah_j x \notin V(W)$. Thus

$$h_1, \dots, h_t, ah_1, \dots, ah_t$$

escapes $V(W)$ for every orbit point.

The budget drops at each recursive step, so there are at most $\binom{m+D}{D}$ steps. Word length increases by at most one and list size at most doubles at each step. Starting from $W = \{e\}$ proves the theorem; crude bounds

$$r \leq \binom{m+D}{D}, \quad \ell \leq 2^{\binom{m+D}{D}}$$

are sufficient. □

3. Positive-proportion escape

THEOREM 8.4 (Escape from subvarieties). *Under the hypotheses of [Theorem 8.3](#), there are $\eta > 0$ and $M \geq 1$, depending only on the bounded-degree algebraic data, such that for every $x \in \mathcal{O}$ at least*

$$\max(1, \eta|A|)$$

distinct elements $g \in A^M$ satisfy $gx \notin V$.

PROOF. Take an escaping list $g_1, \dots, g_\ell \in A^r$. For every $a \in A$, apply it to ax and choose $i(a)$ with

$$g_{i(a)}ax \notin V.$$

Thus $h_a = g_{i(a)}a \in A^{r+1}$ escapes x . For a fixed h and fixed index i , the equation $h = g_i a$ determines a uniquely. Hence a given h arises from at most ℓ values of a . There are therefore at least $|A|/\ell$ distinct escaping elements. Take $\eta = 1/\ell$ and $M = r + 1$. $\square$

COROLLARY 8.5 (Escape in a linear algebraic group). *Let $G \subseteq \mathrm{GL}_N(k)$ be irreducible and $V \subsetneq G$ a proper algebraic subset cut out in the ambient matrix space by equations of degree at most D . Suppose $A \subseteq G(k)$ is finite, symmetric, and generates a Zariski-dense subgroup. Then for some M and $\eta > 0$ depending only on the fixed ambient algebraic data,*

$$|A^M \setminus V| \geq \eta|A|.$$

PROOF. Let G act on itself by left multiplication in the matrix affine space. This is linear in the matrix entries, so degrees do not increase. The orbit of the identity is the Zariski-dense subgroup Γ , which is not contained in V . Apply [Theorem 8.4](#) at e . $\square$

COROLLARY 8.6 (Approximate-group escape). *Let A be a finite symmetric K -approximate subgroup of a linear algebraic group $G(k)$ and assume $\langle A \rangle$ is Zariski dense. For every fixed proper bounded-degree subvariety $V \subsetneq G$, there are constants $m, c, C > 0$ depending only on the algebraic data such that*

$$|A^m \setminus V| \geq c|A|, \quad |A^m| \leq K^C|A|.$$

PROOF. The first inequality is [Corollary 8.5](#). The second is the bounded-power estimate for approximate groups proved in Volume II. $\square$

4. A two-dimensional escape laboratory

Let the subgroup generated by $A \subseteq \mathrm{GL}_2(k)$ act on k^2 , and take the trap

$$V = V(y) = \{(x, 0) : x \in k\}.$$

Its defining equations lie in the three-dimensional space of affine linear polynomials. If the orbit of a point is not contained in V , then either a generator already moves the

point out of V , or intersecting V with the inverse image of V under such a generator introduces a genuinely new linear equation. After at most two independent linear equations the intersection is zero-dimensional; one more independent condition makes it empty on the orbit. The general proof differs only in replacing the three-dimensional linear-polynomial space by

$$k[x_1, \dots, x_m]_{\leq D},$$

whose finite dimension supplies the universal stopping time.

This viewpoint is worth retaining: escape is not a probabilistic miracle. It is finite-dimensional linear algebra applied to the orbit of the defining equations.

EXAMPLE 8.7 (A two-word escaping list in k^2). Let

$$u = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad V = \{(x, y) : y = 0\}.$$

On column vectors, $u(a, b) = (a, a + b)$. Consider any point (a, b) whose orbit under $\langle u \rangle$ is not contained in V . If $b \neq 0$, the identity already sends the point outside V . If $b = 0$, then necessarily $a \neq 0$ (the origin has orbit contained in V), and

$$u(a, 0) = (a, a) \notin V.$$

Thus the fixed list $\{e, u\}$ escapes every admissible orbit point: here one may take $r = 1$ and $\ell = 2$.

The equation-space calculation displays the same mechanism. The defining equation is y , while

$$u^*y = x + y.$$

Hence

$$\text{span}\{y, u^*y\} = \text{span}\{y, x + y\}$$

has dimension two, and

$$V(y) \cap V(x + y) = V(x, y) = \{0\}.$$

One translate has spent one new equation direction and reduced the line trap to a point.

CHECKPOINT 8.8. Escape is now fully internal to the series and does not depend on a hidden generalized Bézout theorem. Uniformity comes from one transparent fact: a strict bounded-degree intersection consumes a new direction in a finite-dimensional polynomial space. The next chapter combines escape with generically finite product maps to obtain the product–conjugation geometry used in Larsen–Pink induction.

CHAPTER 9

Slicing and the Product–Conjugation Phenomenon

The geometric analogue of sum–product is a tension between multiplication and conjugation. If conjugates of one positive-dimensional variety fail to enlarge another under multiplication, then their difference sets lie in stabilizers. If this failure persists for generic conjugators, the intersection of the conjugate stabilizers produces a positive-dimensional normal subgroup. Thus almost simplicity forces dimension growth.

1. A slicing lemma

THEOREM 9.1 (Generic slicing). *Let V, W be irreducible affine varieties and let*

$$S \subseteq V \times W$$

be a closed algebraic subset with

$$\dim S < \dim V + \dim W.$$

Then there is a nonempty open subset $V^\circ \subseteq V$ such that for every $v \in V^\circ$ the fiber

$$S_v = \{w \in W : (v, w) \in S\}$$

is either empty or has dimension strictly smaller than $\dim W$.

PROOF. Consider the projection

$$\pi : S \rightarrow V.$$

Decompose S into irreducible components and treat each separately. If a component S_i does not dominate V , remove the proper closed set $\overline{\pi(S_i)}$ from V . If S_i dominates V , the generic fiber theorem gives, on a nonempty open subset,

$$\dim(S_i)_v = \dim S_i - \dim V < \dim W.$$

Intersect the finitely many resulting open subsets. The fiber of S is the finite union of the component fibers, so its dimension is the maximum of their dimensions and is therefore $< \dim W$. □

2. Setwise stabilizers under translation

Let $W \subseteq G$ be a closed subvariety of an affine algebraic group $G \subseteq \mathrm{GL}_N$.

DEFINITION 9.2. The left setwise stabilizer of W is

$$S_W = \{g \in G : gW = W\}.$$

PROPOSITION 9.3 (Dimension of a translation stabilizer). *The set S_W is a closed algebraic subgroup. If W is nonempty, then*

$$\dim S_W \leq \dim W.$$

PROOF. Closedness follows from Proposition 7.11, applied to the linear left-multiplication action on the matrix affine space. It is a subgroup by definition.

Fix $w_0 \in W$. The map

$$S_W \rightarrow W, \quad s \mapsto sw_0$$

is the restriction of the algebraic isomorphism $G \rightarrow G$, $s \mapsto sw_0$, and is injective. Hence S_W is algebraically isomorphic to the locally closed subset $S_W w_0 \subseteq W$, giving the dimension inequality. $\square$

3. The bad-conjugator set is algebraic

Let $V, W \subseteq G$ be irreducible closed subvarieties and put

$$D = \overline{V^{-1}V}.$$

LEMMA 9.4 (The difference variety has positive dimension). *If $\dim V > 0$, then*

$$\dim D \geq \dim V > 0.$$

PROOF. Fix $v_0 \in V$. The translate

$$v_0^{-1}V$$

is a closed irreducible subset of $V^{-1}V \subseteq D$ and has dimension $\dim V$. $\square$

LEMMA 9.5 (Containment under conjugation is a closed condition). *Let $D, S \subseteq G \subseteq \mathrm{GL}_N$ be closed algebraic sets. Then*

$$\mathcal{B}(D, S) = \{a \in G : a^{-1}Da \subseteq S\}$$

is closed in G .

PROOF. Choose generators

$$f_1, \dots, f_r$$

of the ideal $I(S)$ in the ambient matrix coordinate ring. For fixed i , the expression

$$F_i(a, x) = f_i(a^{-1}xa)$$

is regular in $a \in G$ and polynomial in the matrix coordinates of x of degree at most $\deg f_i$; conjugation is linear in x .

Let P_E be the finite-dimensional vector space of polynomials in the x -coordinates of degree at most

$$E = \max_i \deg f_i.$$

The subspace

$$I(D) \cap P_E \subseteq P_E$$

is fixed. For each a , the polynomial $x \mapsto F_i(a, x)$ is an element of P_E , and its coordinates in a fixed basis of P_E are regular functions of a . The condition that it vanish on D is, by the Nullstellensatz and radicality of $I(D)$, exactly

$$F_i(a, \cdot) \in I(D) \cap P_E.$$

Membership in a fixed linear subspace is given by finitely many linear equations on the coefficient vector, hence by finitely many regular equations in a . Intersect over i . These are precisely the a for which every defining function of S vanishes on $a^{-1}Da$, i.e. $a^{-1}Da \subseteq S$. $\square$

4. When multiplication fails to gain dimension

For $a \in G$, write

$$V^a = a^{-1}Va.$$

LEMMA 9.6 (No dimension gain forces a stabilizer). *Let $V, W \subseteq G$ be irreducible, with $W \neq \emptyset$. If*

$$\dim \overline{V^a W} \leq \dim W,$$

then

$$a^{-1}Da \subseteq S_W, \quad D = \overline{V^{-1}V}.$$

PROOF. The product $V^a \times W$ is irreducible, so the closure

$$P = \overline{V^a W}$$

is irreducible. For every $v \in V^a$, the translate vW is a closed irreducible subset of P of dimension $\dim W$. Since P also has dimension at most $\dim W$, equality holds and

$$vW = P$$

for every $v \in V^a$. Therefore for all $v, v' \in V^a$,

$$v^{-1}v'W = W,$$

so

$$(V^a)^{-1}V^a \subseteq S_W.$$

Taking Zariski closures gives

$$a^{-1}Da \subseteq S_W.$$

$\square$

DEFINITION 9.7. An irreducible algebraic group G is *almost simple for the purposes of this volume* if it has no proper positive-dimensional closed normal algebraic subgroup.

THEOREM 9.8 (Product–conjugation phenomenon). *Let $G \subseteq \mathrm{GL}_N$ be an irreducible almost simple algebraic group. Let $V, W \subseteq G$ be irreducible closed subvarieties satisfying*

$$0 < \dim V, \quad 0 < \dim W < \dim G.$$

Then the set of $a \in G$ for which

$$\dim \overline{V^a W} \leq \dim W$$

is contained in a proper closed subvariety of G . Equivalently, for a in a nonempty Zariski-open subset of G ,

$$\dim \overline{V^a W} > \dim W.$$

PROOF. Let

$$D = \overline{V^{-1}V}, \quad S = S_W.$$

By Lemma 9.6, every bad conjugator lies in the closed set

$$\mathcal{B} = \{a \in G : a^{-1}Da \subseteq S\},$$

which is closed by Lemma 9.5.

Suppose for contradiction that $\mathcal{B} = G$. Then for every $a \in G$,

$$D \subseteq aSa^{-1}.$$

Hence

$$D \subseteq N := \bigcap_{a \in G} aSa^{-1}.$$

The set N is a closed normal algebraic subgroup of G . It has positive dimension because it contains D , and Lemma 9.4 gives $\dim D > 0$. On the other hand,

$$N \subseteq S,$$

and Proposition 9.3 gives

$$\dim N \leq \dim S \leq \dim W < \dim G.$$

Thus N is a proper positive-dimensional closed normal subgroup, contradicting almost simplicity. Therefore $\mathcal{B}$ is proper, and its complement is the required nonempty open set. $\square$

COROLLARY 9.9 (Finding a transverse conjugator in bounded words). *Let $A \subseteq G(k)$ be finite, symmetric, and generate a Zariski-dense subgroup of an almost simple G . Under the hypotheses of Theorem 9.8, there is an integer m depending only on the algebraic data and an element*

$$a \in A^m$$

such that

$$\dim \overline{V^a W} > \dim W.$$

Moreover a positive proportion of A^m consists of such conjugators.

PROOF. The bad conjugators lie in a proper closed subvariety $\mathcal{B} \subsetneq G$ by Theorem 9.8. Apply the escape theorem Corollary 8.5 to $\mathcal{B}$. $\square$

5. Rank-one shadow of product–conjugation

In $G = \mathrm{SL}_2$, take $V = W = T$ to be a one-dimensional maximal torus. If $a \in N_G(T)$, then

$$T^a T = T$$

and no dimension is gained. If a moves T to a genuinely transverse torus, the product $T^a T$ is typically two-dimensional. Thus the abstract bad set of Theorem 9.8 reduces in rank one to the familiar normalizer obstruction.

The higher-dimensional theorem says the same thing without coordinates. If $V^a W$ persistently fails to gain dimension, then the difference variety $V^{-1}V$ is forced into the translation stabilizer of W . If this happens for every a , intersecting conjugates of that stabilizer creates a positive-dimensional normal subgroup. Almost simplicity rules this out. Hence “transverse multiplication grows” is ultimately a normal-subgroup statement.

6. Proof architecture for the next volume

CHECKPOINT 9.10. The algebraic-geometric mechanism needed for Larsen–Pink is now complete at the qualitative/finite-multiplicity level:

$$\boxed{\text{generic fiber dimension}} + \boxed{\text{bounded finite fibers}} \\ + \boxed{\text{escape}} + \boxed{\text{product–conjugation}}.$$

The next volume will add the combinatorial induction. Its central task is to convert these geometric statements into the exponent

$$\frac{\dim V}{\dim G}$$

for intersections of approximate groups with subvarieties, and then use maximal tori to force the small-or-large product theorem.

APPENDIX A

Principal Ideals and Degree Bookkeeping

Two commutative-algebra facts are easy to hide inside the word “Bézout”: a principal equation cuts every irreducible component in codimension one, and the leading Hilbert coefficients of distinct top-dimensional components add. This appendix proves those facts so that the degree bounds used in escape are genuinely internal.

1. Nakayama’s lemma

LEMMA A.1 (Nakayama). *Let $(R, \mathfrak{m})$ be a local ring and M a finitely generated R -module. If*

$$M = \mathfrak{m}M,$$

then $M = 0$. More generally, if $I \subseteq \mathfrak{m}$ and $M = IM$, then $M = 0$.

PROOF. Choose a minimal generating set $m_1, \dots, m_r$ for M . If $r > 0$ and $M = \mathfrak{m}M$, write

$$m_r = \sum_{i=1}^r a_i m_i, \quad a_i \in \mathfrak{m}.$$

Then

$$(1 - a_r)m_r = \sum_{i < r} a_i m_i.$$

Since $a_r \in \mathfrak{m}$, the element $1 - a_r$ is a unit in the local ring: otherwise it would lie in the unique maximal ideal, forcing $1 \in \mathfrak{m}$. Thus m_r lies in the span of the preceding generators, contradicting minimality. Hence $r = 0$. The version with I follows from $IM \subseteq \mathfrak{m}M$. $\square$

2. Symbolic powers

DEFINITION A.2. For a prime ideal $\mathfrak{q}$ of a Noetherian ring R , define its n th symbolic power by

$$\mathfrak{q}^{(n)} = \mathfrak{q}^n R_{\mathfrak{q}} \cap R.$$

LEMMA A.3 (Symbolic powers are primary). *The ideal $\mathfrak{q}^{(n)}$ is $\mathfrak{q}$ -primary: if*

$$ab \in \mathfrak{q}^{(n)}, \quad b \notin \mathfrak{q},$$

then $a \in \mathfrak{q}^{(n)}$.

PROOF. In $R_{\mathfrak{q}}$, the element b is a unit. Since

$$ab \in \mathfrak{q}^n R_{\mathfrak{q}},$$

we obtain $a \in \mathfrak{q}^n R_{\mathfrak{q}}$. As $a \in R$, this means precisely $a \in \mathfrak{q}^{(n)}$. $\square$

LEMMA A.4 (Nilpotent maximal ideal implies Artinian). *Let $(R, \mathfrak{m})$ be a Noetherian local ring. If $\mathfrak{m}^N = 0$ for some N , then R is Artinian.*

PROOF. Each quotient

$$\mathfrak{m}^i / \mathfrak{m}^{i+1}$$

is a finitely generated module over the residue field $R/\mathfrak{m}$, hence a finite-dimensional vector space. Therefore each quotient is Artinian as an R -module. The finite filtration

$$R \supset \mathfrak{m} \supset \cdots \supset \mathfrak{m}^N = 0$$

shows inductively that R is Artinian as a module over itself, i.e. satisfies the descending-chain condition on ideals. $\square$

3. Krull's principal ideal theorem

THEOREM A.5 (Principal ideal theorem). *Let R be Noetherian, $x \in R$, and let $\mathfrak{p}$ be a prime ideal minimal among primes containing (x) . Then*

$$\text{ht}(\mathfrak{p}) \leq 1.$$

PROOF. Localize at $\mathfrak{p}$. We may assume $(R, \mathfrak{p})$ is local and that $\mathfrak{p}$ is the unique prime minimal over (x) . Hence

$$\sqrt{(x)} = \mathfrak{p}.$$

Because $\mathfrak{p}$ is finitely generated, some power satisfies

$$\mathfrak{p}^N \subseteq (x).$$

Indeed, if $p_1, \dots, p_s$ generate $\mathfrak{p}$ and $p_i^{n_i} \in (x)$, then every monomial of total degree at least $n_1 + \cdots + n_s$ contains one $p_i^{n_i}$. Thus the maximal ideal of $R/(x)$ is nilpotent, so $R/(x)$ is Artinian by Lemma A.4.

Let $\mathfrak{q} \subsetneq \mathfrak{p}$ be any smaller prime. Since $\mathfrak{p}$ is minimal over (x) , we have

$$x \notin \mathfrak{q}.$$

Consider the descending chain in the Artinian ring $R/(x)$:

$$\frac{\mathfrak{q}^{(1)} + (x)}{(x)} \supseteq \frac{\mathfrak{q}^{(2)} + (x)}{(x)} \supseteq \cdots$$

It stabilizes, so for some n ,

$$\mathfrak{q}^{(n)} + (x) = \mathfrak{q}^{(n+1)} + (x).$$

We claim

$$\mathfrak{q}^{(n)} = \mathfrak{q}^{(n+1)} + x\mathfrak{q}^{(n)}.$$

The reverse inclusion is obvious. For $y \in \mathfrak{q}^{(n)}$, write

$$y = z + ax, \quad z \in \mathfrak{q}^{(n+1)}.$$

Then $ax = y - z \in \mathfrak{q}^{(n)}$. Since $x \notin \mathfrak{q}$ and $\mathfrak{q}^{(n)}$ is $\mathfrak{q}$ -primary by Lemma A.3, we get $a \in \mathfrak{q}^{(n)}$, proving the claim.

Set

$$M = \mathfrak{q}^{(n)} / \mathfrak{q}^{(n+1)}.$$

It is finitely generated because R is Noetherian, and the displayed equality gives

$$M = xM.$$

Since $x \in \mathfrak{p}$, Nakayama's lemma gives $M = 0$, so

$$\mathfrak{q}^{(n)} = \mathfrak{q}^{(n+1)}.$$

Localize at $\mathfrak{q}$:

$$\mathfrak{q}^n R_{\mathfrak{q}} = \mathfrak{q}^{n+1} R_{\mathfrak{q}}.$$

Apply Nakayama to the finitely generated $R_{\mathfrak{q}}$ -module $\mathfrak{q}^n R_{\mathfrak{q}}$ to obtain

$$\mathfrak{q}^n R_{\mathfrak{q}} = 0.$$

Thus the maximal ideal of the local ring $R_{\mathfrak{q}}$ is nilpotent. Every prime of $R_{\mathfrak{q}}$ contains its nilradical, which is the maximal ideal, so $R_{\mathfrak{q}}$ has only one prime. Hence

$$\text{ht}(\mathfrak{q}) = 0.$$

We have shown that every prime strictly below $\mathfrak{p}$ has height zero. Therefore no chain of primes ending at $\mathfrak{p}$ can have length greater than one, and

$$\text{ht}(\mathfrak{p}) \leq 1.$$

□

COROLLARY A.6 (Pure hypersurface section). *Let X be an irreducible affine or projective variety and let F be a regular or homogeneous function not vanishing identically on X . Every irreducible component of*

$$X \cap V(F)$$

has codimension one in X .

PROOF. In the coordinate ring R of X , irreducible components of the zero set correspond to primes minimal over the principal ideal (F) . Since R is a domain and $F \neq 0$, such a prime is nonzero, so its height is at least one. The principal ideal theorem gives height at most one. Thus its height is exactly one.

To translate height one into dimension drop one in our finitely generated k -algebra setting, choose a chain of primes from 0 to the given prime and use Noether normalization/incomparability as in Corollary 2.16; equivalently the quotient loses exactly one element from a transcendence basis. Hence every component has dimension $\dim X - 1$. □

COROLLARY A.7 (Codimension-one projective varieties are hypersurfaces). *Let $H \subseteq \mathbb{P}^N$ be an irreducible projective variety of dimension $N - 1$. Then*

$$H = V(P)$$

for an irreducible homogeneous polynomial P , unique up to multiplication by a nonzero scalar. Moreover

$$\deg H = \deg P.$$

PROOF. Because H is proper, its homogeneous ideal contains a nonzero homogeneous polynomial F . Factor F into irreducibles in the unique-factorization domain $k[x_0, \dots, x_N]$. Since $I(H)$ is prime, one irreducible factor P lies in $I(H)$, so

$$H \subseteq V(P).$$

The ideal (P) is prime because P is irreducible in a UFD, hence $V(P)$ is irreducible. By Corollary A.6 applied to $\mathbb{P}^N$, it has dimension $N - 1$. Thus the irreducible closed subset $H \subseteq V(P)$ has the same dimension as $V(P)$, and the dimension-drop theorem forces equality.

Let $e = \deg P$ and $S = k[x_0, \dots, x_N]$. Multiplication by P gives the graded exact sequence

$$0 \longrightarrow S(-e) \xrightarrow{\cdot P} S \longrightarrow S/(P) \longrightarrow 0.$$

Hence

$$P_{V(P)}(m) = P_{\mathbb{P}^N}(m) - P_{\mathbb{P}^N}(m - e).$$

Since

$$P_{\mathbb{P}^N}(m) = \binom{m+N}{N} = \frac{m^N}{N!} + O(m^{N-1}),$$

the finite difference has leading term

$$\frac{e}{(N-1)!} m^{N-1}.$$

By the definition of Hilbert degree, $\deg V(P) = e$.

If $V(P) = V(Q)$ with P, Q irreducible, the Nullstellensatz gives $P \mid Q$ and $Q \mid P$, so they differ by a nonzero scalar. This proves uniqueness. $\square$

4. Leading coefficients of unions

LEMMA A.8 (Hilbert polynomial of a union). *Let $X, Y \subseteq \mathbb{P}^N$ be projective algebraic sets with homogeneous ideals I, J . Then*

$$0 \rightarrow S/(I \cap J) \rightarrow S/I \oplus S/J \rightarrow S/(I + J) \rightarrow 0$$

is exact, where $S = k[x_0, \dots, x_N]$. Consequently

$$P_{X \cup Y} = P_X + P_Y - P_{X \cap Y}.$$

If X and Y are distinct irreducible varieties of the same dimension r , then

$$\dim(X \cap Y) < r,$$

so the leading degree- r coefficient of $P_{X \cup Y}$ is the sum of those of P_X and P_Y .

PROOF. The first map sends the class of f to (f, f) ; the second sends (a, b) to $a - b$. Its kernel consists exactly of pairs represented by a common polynomial modulo I and J , which is the image of $S/(I \cap J)$. Surjectivity is immediate. Taking graded dimensions gives the Hilbert-polynomial identity.

If distinct irreducible X, Y of dimension r had an r -dimensional component in their intersection, that component would be a proper closed subset of one of them with equal dimension, contradicting Corollary 2.16. Hence $P_{X \cap Y}$ has degree $< r$ and does not affect the leading coefficient. $\square$

COROLLARY A.9 (Degree adds over top-dimensional components). *If a reduced projective algebraic set X has top-dimensional irreducible components $X_1, \dots, X_s$ of dimension r , then the coefficient of $n^r/r!$ in its Hilbert polynomial is*

$$\sum_{i=1}^s \deg X_i.$$

PROOF. Apply Lemma A.8 inductively. Intersections of distinct top-dimensional components have dimension $< r$, so all correction terms have lower-degree Hilbert polynomials. $\square$

5. The full reduced hypersurface degree estimate

PROPOSITION A.10 (Full reduced hypersurface degree bound). *Let $X \subseteq \mathbb{P}^N$ be irreducible of dimension $r \geq 1$, and let F be homogeneous of degree e and nonzero on X . Then every irreducible component of $X \cap V(F)$ has dimension $r - 1$, and*

$$\delta(X \cap V(F)) \leq e \deg X.$$

PROOF. Purity is Corollary A.6. The exact sequence in Lemma 5.1 gives scheme-theoretic Hilbert polynomial

$$P_X(n) - P_X(n - e),$$

and if $P_X(n) = \frac{\deg X}{r!} n^r + O(n^{r-1})$, then the binomial expansion gives

$$P_X(n) - P_X(n - e) = \frac{e \deg X}{(r - 1)!} n^{r-1} + O(n^{r-2}).$$

Thus the scheme-theoretic hypersurface section has degree $e \deg X$. Passing to the reduction is quotienting the homogeneous coordinate ring by its nilradical, so the Hilbert function cannot increase. By Corollary A.9, the leading coefficient of the reduced intersection is the sum of degrees of all its irreducible components. Therefore that sum is at most $e \deg X$. $\square$

THEOREM A.11 (Coarse reduced Bézout). *Let $X \subseteq \mathbb{P}^N$ be a reduced projective algebraic set and let $F_1, \dots, F_s$ be homogeneous polynomials of positive degrees $e_1, \dots, e_s$. Then*

$$(A.1) \quad \delta(X \cap V(F_1, \dots, F_s)) \leq \delta(X) \prod_{j=1}^s e_j.$$

In particular, if an affine algebraic set in $\mathbb{A}^N$ is cut out set-theoretically by at most s polynomials of degrees at most e , then the total degree of its projective closure is at most e^s .

PROOF. It is enough to prove the one-equation estimate

$$\delta(X \cap V(F)) \leq e \delta(X), \quad e = \deg F,$$

and iterate. Write $X_1, \dots, X_r$ for the irreducible components of X . If F vanishes identically on X_i , that component survives unchanged and contributes $\deg X_i \leq e \deg X_i$. If not, [Proposition A.10](#) says that the sum of the degrees of all irreducible components of $X_i \cap V(F)$ is at most $e \deg X_i$. Every component of $X \cap V(F)$ occurs among these intersections, so summing over i gives the one-equation estimate. Iteration proves (A.1).

For the affine statement, homogenize the defining equations. Their common projective zero set Y has total degree at most e^s by the first part, starting from $\delta(\mathbb{P}^N) = 1$. Every irreducible component of the affine zero set has projective closure equal to an irreducible component of Y that is not contained in the hyperplane at infinity: indeed its affine part is a maximal irreducible subset of the affine common zero set. Hence the projective closure of the affine algebraic set is the union of a subcollection of the components of Y , and its total degree is at most $\delta(Y) \leq e^s$. $\square$

6. Finite-field point counting from degree

The same degree bookkeeping gives the rational-point estimate needed later to turn qualitative algebraic escape into a uniform statement over finite fields. No Lang–Weil theorem is required: repeated slicing by a coordinate hyperplane is enough.

THEOREM A.12 (Degree point bound over a finite field). *Let k be an algebraically closed field containing $\mathbb{F}_q$, and let $X \subseteq \mathbb{A}_k^N$ be a reduced algebraic set of dimension at most d . Then*

$$(A.2) \quad |X(k) \cap \mathbb{F}_q^N| \leq \delta(X) q^d.$$

In particular, a family of algebraic sets of bounded complexity and dimension at most d has $O_M(q^d) \mathbb{F}_q$ -points, uniformly in the coefficients and even when the sets themselves are not defined over $\mathbb{F}_q$.

PROOF. It is enough to prove the estimate for an irreducible component and sum over components. Indeed, if $X = \bigcup_i X_i$, then

$$|X \cap \mathbb{F}_q^N| \leq \sum_i |X_i \cap \mathbb{F}_q^N|,$$

while $\sum_i \deg X_i = \delta(X)$ and $\dim X_i \leq d$.

We therefore argue by induction on d for irreducible X . If $d = 0$, then X is a single geometric point and the assertion is immediate; more generally [Corollary 5.7](#) identifies degree with the number of geometric points.

Assume $d > 0$. Some coordinate function x_j is nonconstant on X ; otherwise all coordinate functions would be constant and X would be a point. For each $t \in \mathbb{F}_q$, put

$$X_t = X \cap V(x_j - t).$$

Because $x_j - t$ is a nonzero element of the coordinate domain of X , [Corollary A.6](#) shows that every nonempty irreducible component of X_t has dimension $d - 1$. The linear equation has degree one, so [Proposition A.10](#) gives

$$\delta(X_t) \leq \deg X.$$

(The projective closure of the affine fiber is contained in the corresponding projective hyperplane section, so discarding components at infinity can only decrease total degree.) By the induction hypothesis,

$$|X_t(k) \cap \mathbb{F}_q^N| \leq \delta(X_t)q^{d-1} \leq (\deg X)q^{d-1}.$$

Every $\mathbb{F}_q$ -point of X belongs to exactly one of the q fibers with $t \in \mathbb{F}_q$. Summing therefore gives

$$|X(k) \cap \mathbb{F}_q^N| \leq q(\deg X)q^{d-1} = (\deg X)q^d.$$

This proves the irreducible case and hence the theorem.

For the final assertion, coarse reduced Bézout bounds $\delta(X)$ in terms of the complexity bound M , so the implied constant depends only on M . $\square$

CHECKPOINT A.13. The principal-ideal and degree bookkeeping is now fully internal. A single equation cuts a noncontained irreducible component in pure codimension one; degrees add over top-dimensional components; and repeated hypersurface intersections obey the coarse multiplicative Bézout bound. Later compactness arguments may therefore use bounded degree without importing intersection theory.

APPENDIX B

Uniform Algebraic Geometry by Ultralimits

The main chapters deliberately proved qualitative algebraic geometry first. Larsen–Pink also needs a uniformity principle: if the ambient dimension, the number of equations and their degrees are bounded, then image closures, exceptional loci, fibers and irreducible decompositions have bounds depending only on those initial parameters, not on the coefficients or the base field. We prove this compactness principle from scratch with non-principal ultrafilters. The argument is deliberately separated from the geometry so that one can see exactly where effectiveness is lost.

1. Complexity

DEFINITION B.1 (Affine complexity). An affine algebraic set $V \subseteq \mathbb{A}_k^n$ has *complexity at most M* if

$$n \leq M$$

and V is the common zero set of at most M polynomials, each of total degree at most M . A constructible set has complexity at most M if it is obtained from at most M affine algebraic sets of complexity at most M by at most M unions, intersections and complements.

The precise numerical convention is unimportant. Replacing M by a function of M gives an equivalent notion. What matters is that bounded complexity means “boundedly many coefficients in a bounded-dimensional coefficient space.”

2. A non-principal ultrafilter

DEFINITION B.2. A filter $\mathcal{F}$ on $\mathbb{N}$ is a nonempty collection of subsets of $\mathbb{N}$ such that

- (1) $\emptyset \notin \mathcal{F}$;
- (2) $E, F \in \mathcal{F}$ implies $E \cap F \in \mathcal{F}$;
- (3) $E \in \mathcal{F}$ and $E \subseteq F$ imply $F \in \mathcal{F}$.

An ultrafilter is a maximal proper filter. It is non-principal if it contains no finite set.

LEMMA B.3 (Existence of a non-principal ultrafilter). *Assuming Zorn’s lemma, there is a non-principal ultrafilter on $\mathbb{N}$ containing every cofinite set.*

PROOF. The cofinite subsets of $\mathbb{N}$ form a proper filter $\mathcal{F}_0$. Partially order the proper filters containing $\mathcal{F}_0$ by inclusion. The union of a chain is again a proper filter: closure under finite intersections follows because two members of the union lie together in one filter from the chain, and the empty set never occurs. Zorn's lemma gives a maximal element $\mathcal{U}$.

For every $E \subseteq \mathbb{N}$, maximality implies that either $E \in \mathcal{U}$ or $\mathbb{N} \setminus E \in \mathcal{U}$. Indeed, if $E \notin \mathcal{U}$, adjoining E must generate the improper filter, so there is $F \in \mathcal{U}$ with $F \cap E = \emptyset$; hence $F \subseteq \mathbb{N} \setminus E$ and the complement belongs to $\mathcal{U}$. Finally, no finite set belongs to $\mathcal{U}$, because its cofinite complement already does. $\square$

Fix such an ultrafilter $\mathcal{U}$. We say a property holds for $\mathcal{U}$ -almost every i if its set of valid indices belongs to $\mathcal{U}$.

LEMMA B.4 (Finite alternatives stabilize). *If a sequence a_i takes values in a finite set S , then there is a unique $s \in S$ such that*

$$\{i : a_i = s\} \in \mathcal{U}.$$

PROOF. The fibers $E_s = \{i : a_i = s\}$ partition $\mathbb{N}$. At least one lies in $\mathcal{U}$: otherwise every complement would lie in $\mathcal{U}$, and intersecting the finitely many complements would put the empty set in $\mathcal{U}$. Two distinct fibers cannot both belong to the filter because they are disjoint. $\square$

3. Ultraproducts of fields

Let k_i be fields. On the product $\prod_i k_i$ declare

$$(x_i) \sim (y_i) \iff \{i : x_i = y_i\} \in \mathcal{U}.$$

The quotient is denoted

$$k = \prod_{\mathcal{U}} k_i.$$

Addition and multiplication are defined coordinatewise.

PROPOSITION B.5 (The ultraproduct is a field). *The ring $k = \prod_{\mathcal{U}} k_i$ is a field. If every k_i is algebraically closed, then k is algebraically closed.*

PROOF. The equivalence relation is compatible with addition and multiplication because a finite intersection of $\mathcal{U}$ -large sets is $\mathcal{U}$ -large. If $x = [(x_i)] \neq 0$, then

$$E = \{i : x_i \neq 0\} \in \mathcal{U}.$$

Define $y_i = x_i^{-1}$ on E and arbitrarily off E . Then $xy = 1$ in the quotient.

Now assume all k_i algebraically closed. Let

$$P(T) = T^d + c_{d-1}T^{d-1} + \cdots + c_0 \in k[T]$$

be monic and nonconstant. Represent each coefficient c_j by a sequence $c_{j,i}$. For every i , the corresponding polynomial

$$P_i(T) = T^d + c_{d-1,i}T^{d-1} + \cdots + c_{0,i}$$

has a root $r_i \in k_i$. Then $r = [(r_i)]$ satisfies $P(r) = 0$. Every nonconstant polynomial can be made monic after division by its leading coefficient, so k is algebraically closed. $\square$

For fixed n , coordinatewise ultralimits identify

$$\mathbb{A}^n(k) = \prod_{\mathcal{U}} \mathbb{A}^n(k_i).$$

4. Bounded polynomials survive the limit

LEMMA B.6 (Ultralimit of bounded-degree polynomials). *Let $f_i \in k_i[x_1, \dots, x_n]$ have degree at most d , with n, d fixed. Then their coefficientwise ultralimit is a polynomial*

$$f \in k[x_1, \dots, x_n]$$

of degree at most d , and for $x = [(x_i)]$ one has

$$f(x) = 0 \iff f_i(x_i) = 0 \text{ for } \mathcal{U}\text{-almost every } i.$$

PROOF. There are only finitely many monomials of degree at most d . Take the ultraproduct of the coefficient sequence of each monomial. Evaluation is a finite expression in field addition and multiplication, so equality to zero transfers exactly through the quotient relation. $\square$

Consequently a sequence of affine algebraic sets of uniformly bounded complexity has an ultralimit algebraic set of bounded complexity over k .

5. Continuity of dimension

We next prove the compactness fact that does most of the work.

THEOREM B.7 (Dimension is continuous under bounded-complexity ultralimits). *Let $V_i \subseteq \mathbb{A}^n(k_i)$ have uniformly bounded complexity, with all k_i algebraically closed, and let*

$$V = \prod_{\mathcal{U}} V_i \subseteq \mathbb{A}^n(k).$$

Then

$$\dim V_i = \dim V$$

for $\mathcal{U}$ -almost every i . The same statement holds for constructible sets, with dimension defined as the dimension of the Zariski closure.

PROOF. We argue by induction on n . The case $n = 0$ is immediate. Let $\pi : \mathbb{A}^n \rightarrow \mathbb{A}^1$ be projection to the last coordinate. Decompose V into finitely many irreducible components. By the generic fiber theorem proved in Chapter 4, on each component either π is constant, or outside finitely many exceptional values the fiber dimension is one less than the component dimension. Taking the union over the finitely many components, there are finitely many exceptional values $t_1, \dots, t_s \in k$ such that, away from them, the maximal fiber dimension has the value prescribed by the components dominating $\mathbb{A}^1$; if there is no dominant component, every nonempty fiber occurs over one of the finitely many t_j .

Represent each t_j by a sequence $t_{j,i} \in k_i$. Suppose the corresponding uniform fiber statement failed for the V_i on a $\mathcal{U}$ -large set. We could then choose a sequence of offending parameters u_i avoiding all $t_{j,i}$ when required. Its ultralimit $u = [(u_i)]$ avoids the exceptional values in k . The fiber $(V_i)_{u_i}$ has uniformly bounded complexity because it is obtained by adding the linear equation $x_n = u_i$. By the induction hypothesis, its dimension agrees $\mathcal{U}$ -almost surely with the dimension of V_{u_i} , contradicting the choice of u_i . The same argument handles the exceptional fibers.

Thus the pattern of fiber dimensions of V_i agrees, for $\mathcal{U}$ -almost every i , with that of V . The dimension formula

$$\dim X = \max \left(\max_{t \in \pi(X)} \dim X_t, 1 + \dim X_t \text{ for generic } t \text{ when } \pi(X) \text{ is dense} \right)$$

then gives $\dim V_i = \dim V$. This formula is just the one-dimensional case of the generic fiber theorem and was proved in Volume V.

A constructible set is a finite Boolean combination of algebraic sets of bounded complexity. Its closure is the union of the closures of finitely many locally closed pieces; applying the affine result to those pieces gives the constructible version. $\square$

6. Weak Bézout as a compactness consequence

COROLLARY B.8 (Uniform zero-dimensional bound). *For every M there is $C(M)$ such that every zero-dimensional constructible set of complexity at most M over an algebraically closed field contains at most $C(M)$ points.*

PROOF. If not, choose V_i of complexity at most M , dimension zero, and $|V_i| \rightarrow \infty$. Their ultralimit V has dimension zero by Theorem B.7, hence is a finite algebraic set by Volume V. Write

$$V = \{v_1, \dots, v_s\}.$$

Represent the v_j by sequences $v_{j,i}$. If $|V_i| > s$ on a $\mathcal{U}$ -large set, choose $w_i \in V_i \setminus \{v_{1,i}, \dots, v_{s,i}\}$. Then $w = [(w_i)] \in V$ but differs from every v_j , contradiction. $\square$

7. Degree and bounded equations for irreducible subvarieties

For an affine algebraic set $X \subseteq \mathbb{A}^n$ we use the Hilbert total degree $\delta(X)$ defined in Chapter 5 from the projective closure. Appendix A proved the coarse reduced

Bézout theorem, so bounded equations now give a direct uniform bound for this degree; no generic-slice multiplicity statement is being assumed.

LEMMA B.9 (Complexity bounds degree). *For fixed M , irreducible affine varieties of complexity at most M have degree bounded by a constant depending only on M .*

PROOF. Let $X \subseteq \mathbb{A}^n$ have complexity at most M . By definition it is cut out by at most M equations, each of degree at most M . The affine form of the coarse reduced Bézout theorem, Theorem A.11, gives

$$\delta(X) \leq M^M.$$

Every irreducible component has degree at most this total degree. This is the required bound. $\square$

LEMMA B.10 (Generic linear normalization and avoidance). *Let $X \subseteq \mathbb{P}^n$ be an irreducible projective variety of dimension d , with standard graded homogeneous coordinate domain R . Let $p \in \mathbb{P}^n$ be any point. There are ambient linear forms*

$$\ell_0, \dots, \ell_d$$

whose images in R_1 make R a finite module over $k[\ell_0, \dots, \ell_d]$ and such that not all $\ell_j(p)$ vanish. In particular, if $p \notin X$, the associated finite linear projection

$$\pi_0 : X \longrightarrow \mathbb{P}^d, \quad x \longmapsto (\ell_0(x) : \dots : \ell_d(x)),$$

is defined both on X and at p .

PROOF. Write $S = k[x_0, \dots, x_n]$ and let $S_1 \rightarrow R_1$ be the restriction map. We first show that, inside the affine parameter space

$$\mathcal{P} = (S_1)^{d+1},$$

the tuples whose images give a finite homogeneous normalization contain a nonempty Zariski-open subset.

By homogeneous Noether normalization, Lemma 5.3, there is at least one tuple $L^0 = (\ell_0^0, \dots, \ell_d^0)$ for which R is finite over $k[\ell_0^0, \dots, \ell_d^0]$. Hence

$$R/(\ell_0^0, \dots, \ell_d^0)R$$

is finite-dimensional over k . Choose m so large that its degree- m piece is zero.

For an arbitrary tuple $L = (\ell_0, \dots, \ell_d)$ consider the linear map

$$(B.1) \quad \Phi_{L,m} : R_{m-1}^{d+1} \longrightarrow R_m, \quad (r_0, \dots, r_d) \longmapsto \sum_{j=0}^d \ell_j r_j.$$

After bases of the two finite-dimensional vector spaces have been fixed, the matrix entries of $\Phi_{L,m}$ are linear functions of the coefficients of L . Thus surjectivity of (B.1) is the nonvanishing of at least one maximal minor and is a Zariski-open condition on $\mathcal{P}$. It holds at L^0 , so the resulting open set $\mathcal{U}$ is nonempty.

If $L \in \mathcal{U}$, then the degree- m piece of $R/(\ell_0, \dots, \ell_d)R$ is zero. Because this quotient is a standard graded algebra, all of its pieces of degree at least m are zero. We claim that the finitely many homogeneous pieces

$$R_0 \oplus \dots \oplus R_{m-1}$$

generate R as a module over $k[\ell_0, \dots, \ell_d]$. Indeed, if $r \in R_j$ with $j \geq m$, surjectivity gives $r = \sum \ell_i r_i$ with $r_i \in R_{j-1}$; induction on j reduces every homogeneous element to degrees below m . Hence the extension is finite. Since $\dim X = d$, the fraction field of R has transcendence degree $d+1$. A finite extension preserves transcendence degree, so the $d+1$ elements $\ell_0, \dots, \ell_d$ are algebraically independent. Thus every $L \in \mathcal{U}$ is a homogeneous Noether normalization.

Now impose the additional condition

$$\ell_0(p) = \dots = \ell_d(p) = 0.$$

It defines a proper linear subspace $\mathcal{H} \subsetneq \mathcal{P}$: for example one may choose an ambient linear form nonzero at p and place it in the first slot. The affine space $\mathcal{P}$ is irreducible, so its nonempty open subset $\mathcal{U}$ cannot be contained in the proper closed subset $\mathcal{H}$. Choose

$$L \in \mathcal{U} \setminus \mathcal{H}.$$

Then the tuple is a finite homogeneous normalization and at least one $\ell_j(p)$ is nonzero, as required. $\square$

THEOREM B.11 (Degree separates points). *Let $X \subseteq \mathbb{A}^n$ be irreducible of degree at most D . Then X is cut out set-theoretically by polynomials of degree at most D . In particular its complexity is bounded in terms of n and D .*

PROOF. Pass to the projective closure $\overline{X} \subseteq \mathbb{P}^n$; its degree is the same. It is enough to show that every point

$$p \in \mathbb{P}^n \setminus \overline{X}$$

can be separated from $\overline{X}$ by a homogeneous polynomial of degree at most D .

Write $d = \dim X$. By [Lemma B.10](#), choose ambient linear forms

$$\ell_0, \dots, \ell_d$$

whose restrictions give a finite projective map

$$\pi_0 : \overline{X} \longrightarrow \mathbb{P}^d$$

and for which not all $\ell_j(p)$ vanish. Thus $\pi_0(p)$ is defined.

The fiber

$$F = \pi_0^{-1}(\pi_0(p))$$

is finite. Choose one more linear form ℓ_{d+1} so that the augmented map

$$\pi = (\ell_0 : \dots : \ell_d : \ell_{d+1}) : \overline{X} \rightarrow \mathbb{P}^{d+1}$$

separates p from every point of F . This is possible because each $x \in F$ imposes one proper linear condition on the coefficients of ℓ_{d+1} , and k is infinite. The map π is still finite onto its image, because its composition with the projection onto the first $d + 1$ coordinates is π_0 .

Let

$$H = \overline{\pi(\overline{X})} \subseteq \mathbb{P}^{d+1}.$$

The map π is given by linear forms and is dominant onto the irreducible variety H . Pullback therefore gives an injective degree-preserving homomorphism from the homogeneous coordinate ring of H to that of $\overline{X}$. Consequently the degree- m piece for H injects into the degree- m piece for $\overline{X}$, and hence

$$P_H(m) \leq P_{\overline{X}}(m)$$

for all sufficiently large m . Both varieties have dimension d , so comparison of leading coefficients gives

$$\deg H \leq \deg X \leq D.$$

By [Corollary A.7](#), the irreducible codimension-one variety $H \subseteq \mathbb{P}^{d+1}$ is the zero set of one irreducible homogeneous form F and

$$\deg F = \deg H \leq D.$$

By construction

$$F(\pi(x)) = 0 \quad (x \in X), \quad F(\pi(p)) \neq 0.$$

The pullback

$$F(\ell_0, \dots, \ell_{d+1})$$

is a form of degree at most D vanishing on X but not at p .

Hence X is the common zero set of all degree- $\leq D$ forms that vanish on it. That vector space has dimension at most

$$\binom{n+D}{D},$$

so a basis gives a bounded list of equations. Dehomogenizing proves the affine statement. $\square$

8. Irreducibility and components are continuous

THEOREM B.12 (Continuity of irreducibility). *Let V_i be affine algebraic sets of uniformly bounded complexity and V their ultralimit. Then V is irreducible if and only if V_i is irreducible for $\mathcal{U}$ -almost every i .*

PROOF. If $V = V^{(1)} \cup V^{(2)}$ is a union of two proper closed subsets, choose bounded-degree equations for the two fixed varieties over the ultraproduct field and represent their coefficients coordinatewise. For $\mathcal{U}$ -almost every i they define proper

closed subsets $V_i^{(1)}, V_i^{(2)} \subsetneq V_i$ whose union is V_i ; properness transfers by choosing one witness point outside each subset. Hence V_i is reducible.

Conversely suppose V is irreducible but V_i is reducible on a $\mathcal{U}$ -large set. Let $d = \dim V$. By Theorem B.7, $\dim V_i = d$ almost surely. Choose an irreducible component $U_i \subsetneq V_i$ of dimension d . By Lemma B.9, the total degrees $\delta(V_i)$ are uniformly bounded, and hence so are the degrees of the components U_i . By Theorem B.11, the U_i have uniformly bounded complexity. Their ultralimit $U \subseteq V$ has dimension d by Theorem B.7. Since V is irreducible, a proper closed subset has smaller dimension, so $U = V$.

Equality of ultraproduct sets means $U_i = V_i$ for $\mathcal{U}$ -almost every i : otherwise choose $x_i \in V_i \setminus U_i$ on a $\mathcal{U}$ -large set; its ultralimit lies in $V \setminus U$. This contradicts the choice of U_i as a proper component. $\square$

COROLLARY B.13 (Uniform number of components). *For every M there is $C(M)$ such that every affine algebraic set of complexity at most M has at most $C(M)$ irreducible components. Each component has complexity bounded by a function of M .*

PROOF. Let V have complexity at most M . By Lemma B.9,

$$\delta(V) \leq C_0(M).$$

Every irreducible component has positive degree at least 1: homogeneous Noether normalization injects a polynomial ring of the same dimension into its homogeneous coordinate ring, so comparison of Hilbert functions makes the leading Hilbert coefficient at least that of projective space. Since total degree is the sum of the degrees of the irreducible components,

$$\#\text{Irr}(V) \leq \delta(V) \leq C_0(M).$$

Each component U satisfies $\deg U \leq \delta(V) \leq C_0(M)$, so Theorem B.11 gives a complexity bound depending only on the ambient dimension and $C_0(M)$, hence only on M . $\square$

9. Uniformity of the operations used in growth

The final compactness step is delicate enough that we isolate the elementary transfer principle used in the proof. This is the bounded-complexity fragment of Łoś's theorem, but no model theory is needed here.

LEMMA B.14 (Transfer for bounded descriptions). *Fix finitely many affine spaces. Suppose that X_i, Y_i are constructible sets described by uniformly bounded Boolean combinations of uniformly bounded-degree polynomial equations, and that f_i is a regular map whose coordinate polynomials have uniformly bounded degree. Let X, Y, f be their coefficientwise ultralimits. Then the following assertions transfer between the ultralimit and $\mathcal{U}$ -almost every index:*

- (1) $X_i \subseteq Y_i$;
- (2) $X_i = Y_i$;
- (3) $f_i(X_i) \subseteq Y_i$;
- (4) $Y_i \subseteq f_i(X_i)$.

The same conclusion holds with several boundedly described sets and maps in place of one.

PROOF. Membership in a bounded constructible description is a finite Boolean combination of polynomial equalities, so it transfers by Lemma B.6. For example, if $X \subseteq Y$ but $X_i \not\subseteq Y_i$ on a $\mathcal{U}$ -large set, choose $x_i \in X_i \setminus Y_i$ there. The ultralimit $x = [(x_i)]$ lies in $X \setminus Y$, a contradiction. The converse is identical. Equality is two inclusions.

For image inclusion, a counterexample gives $x_i \in X_i$ with $f_i(x_i) \notin Y_i$; its ultralimit contradicts $f(X) \subseteq Y$. For the reverse inclusion, if $Y \subseteq f(X)$ but for a $\mathcal{U}$ -large set there is $y_i \in Y_i \setminus f_i(X_i)$, then $y = [(y_i)] \in Y$. Choose $x = [(x_i)] \in X$ with $f(x) = y$. Polynomial evaluation transfers, so $f_i(x_i) = y_i$ for $\mathcal{U}$ -almost every i , contradiction. The same witness-selection argument handles any fixed finite collection of sets and maps. $\square$

LEMMA B.15 (A dense constructible set contains componentwise open pieces). *Let C be a constructible subset of an algebraic set $Z = Z_1 \cup \cdots \cup Z_r$, where the Z_j are the irreducible components. If $\overline{C} = Z$, then for every j there is a nonempty open subset $U_j \subseteq Z_j$ such that*

$$U_j \subseteq C \cap \left(Z_j \setminus \bigcup_{\ell \neq j} Z_\ell \right).$$

PROOF. The set $Z_j^\circ = Z_j \setminus \bigcup_{\ell \neq j} Z_\ell$ is a nonempty open subset of the irreducible space Z_j . Since $C \cap Z_j^\circ$ is constructible and dense in Z_j° , write it as a finite union of locally closed subsets. At least one of those subsets has closure Z_j ; its closed part therefore contains a nonempty open subset of Z_j , and intersecting with its open part and with Z_j° gives the required U_j . $\square$

THEOREM B.16 (Uniform algebraic-geometry package). *Fix M . There is a constant $C = C(M)$ with the following property. Whenever all ambient affine dimensions, defining equations and regular maps have complexity at most M , the following objects admit descriptions of complexity at most C :*

- (1) *finite unions, intersections, products, translates and inverses;*
- (2) *Zariski closures of constructible sets;*
- (3) *images and preimages under regular maps;*
- (4) *irreducible components;*
- (5) *exceptional target and input sets for the generic-fiber theorem;*
- (6) *exceptional base and fiber sets for the generic-slicing statements used in the main text;*

(7) *setwise stabilizers of bounded-complexity algebraic subsets under bounded-complexity algebraic group actions.*

The number of irreducible components and the cardinality of every zero-dimensional fiber are also at most C .

PROOF. The Boolean operations, products and preimages are obtained by writing down the corresponding equations. For translations and inverses in an algebraic group, substitute the bounded-complexity multiplication and inverse maps. Thus only closure, image, components and exceptional sets require compactness.

Uniform closure. Suppose no uniform bound existed for the closure of a constructible set of complexity at most M . Choose counterexamples X_i and form their ultralimit X . Let

$$Z = \overline{X} = Z_1 \cup \cdots \cup Z_r$$

be the irreducible decomposition over the ultraproduct field. Each fixed Z_j has a finite equation list; represent its coefficients coordinatewise and call the resulting varieties $Z_{j,i}$, and put $Z_i = \bigcup_j Z_{j,i}$. By Theorems B.7 and B.12, after discarding a $\mathcal{U}$ -small set the $Z_{j,i}$ are irreducible of the correct dimensions and no $Z_{j,i}$ is contained in another.

Because $X \subseteq Z$, Lemma B.14 gives $X_i \subseteq Z_i$ almost surely. By Lemma B.15, for each j choose a nonempty open $U_j \subseteq Z_j$ contained in X and away from the other components. Represent U_j coordinatewise. Transfer gives $U_{j,i} \subseteq X_i$. The complement $Z_j \setminus U_j$ has dimension $< \dim Z_j$, so dimension continuity shows the same for $Z_{j,i} \setminus U_{j,i}$ almost surely. Hence $U_{j,i}$ is Zariski dense in $Z_{j,i}$. Therefore

$$Z_{j,i} \subseteq \overline{X_i} \subseteq Z_i$$

for every j , and consequently $\overline{X_i} = Z_i$ almost surely. The Z_i have one fixed bounded description, contradicting the choice of counterexamples.

Uniform images. Let X_i and f_i have bounded complexity, with ultralimits X and f . Qualitative Chevalley, proved in the main text, says that $C = f(X)$ is constructible. Choose one finite constructible description of C over the ultraproduct field and represent it coordinatewise by C_i . The equality $C = f(X)$ and Lemma B.14 give

$$C_i = f_i(X_i)$$

for $\mathcal{U}$ -almost every i . Thus images have a uniform complexity bound. Applying the already established closure case to C_i gives a uniform bound for image closures as well.

Uniform components. This is exactly Corollary B.13; the direct degree proof also gives a uniform bound on their number.

Uniform generic-fiber exceptional sets. Suppose $f_i : X_i \rightarrow Y_i$ is a bounded family. Pass to the ultralimit $f : X \rightarrow Y$ and decompose source and target into their finitely many irreducible components, using the component bound just proved. On

each relevant irreducible source component, the generic-fiber theorem from the main text supplies a proper algebraic exceptional set $E \subsetneq \overline{f(X)}$ outside which the fiber dimension has the expected value. Represent E coordinatewise by bounded-complexity sets E_i .

These E_i are proper almost surely: choose a point of the target outside E and transfer it. If the asserted fiber dimension failed outside E_i on a $\mathcal{U}$ -large set, choose offending points $y_i \notin E_i$. The fibers

$$X_{i,y_i} = X_i \cap f_i^{-1}(y_i)$$

have uniformly bounded complexity, so [Theorem B.7](#) transfers their dimensions to the fiber over $y = [(y_i)] \notin E$, contradicting the generic-fiber theorem for f . Taking the finite union over source components gives one uniformly bounded target exceptional set. The corresponding input exceptional set is its preimage together with the finitely many lower-dimensional source pieces, so its complexity is uniformly bounded as well.

Uniform slicing. The slicing statements used in the main text are generic-fiber statements for incidence varieties and coordinate projections. Those incidence varieties are built from the original bounded equations by products, intersections and preimages, already controlled above. Applying the preceding paragraph gives uniform exceptional base sets and uniform bounds for the exceptional fibers.

Uniform setwise stabilizers. Suppose G_i acts algebraically on an affine space, with the groups, action maps, and algebraic sets V_i all of complexity at most M . Pass to the ultralimit $G \curvearrowright V$. The qualitative setwise-stabilizer proposition proved in [Chapter 7](#) says that

$$S = \{g \in G : gV = V\}$$

is a closed algebraic subgroup. Choose one finite equation list for S and represent its coefficients coordinatewise, obtaining bounded-complexity closed sets $S_i \subseteq G_i$.

We claim that S_i is the actual setwise stabilizer of V_i for $\mathcal{U}$ -almost every i . Otherwise choose, on a $\mathcal{U}$ -large set, an element g_i belonging to exactly one of the two sets. If $g_i \in S_i$ but $g_i V_i \neq V_i$, then because g_i is an automorphism, after passing to one of two $\mathcal{U}$ -large alternatives there is $x_i \in V_i$ with either

$$g_i x_i \notin V_i \quad \text{or} \quad g_i^{-1} x_i \notin V_i.$$

Taking ultralimits contradicts $g = [(g_i)] \in S$. Conversely, if g_i stabilizes V_i but $g_i \notin S_i$, then the ultralimit g stabilizes V by direct transfer of membership under g and g^{-1} , while the bounded equation description of S_i gives $g \notin S$, again a contradiction. Hence the actual stabilizers are represented by the one fixed bounded description of S , proving the required uniform complexity bound.

Finally, zero-dimensional fibers are uniformly bounded by [Corollary B.8](#). Only finitely many operations occur in the statement, so taking the maximum of the finitely many bounds gives a single constant $C(M)$. $\square$

CHECKPOINT B.17. The compactness principle is now internal:

bounded equations and maps

⇓

bounded closures, images, components and exceptional loci

.

The price is ineffectiveness: the proof gives existence of $C(M)$ rather than a useful numerical formula. This is exactly enough for Larsen–Pink, whose exponents rather than its algebraic constants drive the growth argument.

The Bourgain–Gamburd Method
Volume VI: Larsen–Pink Inequalities and
Algebraic Product Growth

Preface

Volume V built the algebraic geometry needed to make dimension arguments honest: generic fibers, finite maps, algebraic group actions, escape from subvarieties, and the product–conjugation phenomenon. This volume converts that geometry into counting inequalities.

The central statement is the Larsen–Pink principle

$$|A^m \cap V| \ll K^C |A|^{\dim V / \dim G}$$

for a K -approximate subgroup whose generated subgroup is Zariski dense in an almost simple algebraic group. We prove the dimension-pair induction in full, including the exceptional-fiber case and the termination bookkeeping. No Larsen–Pink theorem is imported.

The second half specializes the general inequality to SL_n . In type A the structural facts needed for the “involved torus” argument can be proved directly by linear algebra: regular semisimple matrices, their centralizers, uniqueness of the maximal torus, and the monomial normalizer. We use these facts first to prove that a finite approximate subgroup of $\mathrm{SL}_n(k)$ generating a Zariski-dense subgroup is polynomially small in its approximation parameter, and hence obtain product growth in Zariski-dense subgroups.

We then close the finite-field interface rather than leaving it as a slogan. The degree point-count theorem of Volume V gives bounded-complexity escape for generating subsets of $\mathrm{SL}_n(\mathbb{F}_q)$; the same Larsen–Pink induction yields a small-or-dense dichotomy; an elementary restriction-to-unipotents argument supplies the needed quasirandomness bound without Landazuri–Seitz; and Gowers mixing from Volume I completes the dense case. Thus the volume ends with a proof of the product theorem for $\mathrm{SL}_n(\mathbb{F}_q)$ for every fixed n and every prime power q .

The remaining scope boundary is also explicit. We do not claim the uniform product theorem for every finite simple group of every bounded Lie type. Extending the type- A argument uniformly across all root systems and finite forms requires additional structural theory. Those results motivate later generalizations but are not hidden here behind the words “standard structure theory.” **Series position.** This

volume uses Volume V and the product-growth framework of Volume IV. Its main outputs are Larsen–Pink inequalities and algebraic product growth.

Quantifier and asymptotic conventions. Unless a statement explicitly says otherwise, every constant occurring in a quantitative theorem is chosen in the order in which it is quantified in that theorem. Thus a phrase such as “for all sufficiently small δ ” means: there is a displayed or locally quantified threshold $\delta_0 > 0$, depending only on the parameters listed before that phrase, such that the assertion holds for every $0 < \delta < \delta_0$. The phrases “sufficiently large N ” and “sufficiently large p ” have the analogous meaning.

If $X \ll_\theta Y$ or $X = O_\theta(Y)$, the implicit constant depends only on θ and on the fixed ambient data of the current volume; it is independent of the variables quantified after θ . The words *absolute*, *uniform*, and *bounded* never introduce an unlisted dependence. Whenever an inequality between exponents or scales is essential to the argument, that inequality is written explicitly in the proof; these conventions are not a substitute for such a calculation.

Contents

Preface	iii
Part 1. The Larsen–Pink Induction	1
Chapter 1. Counting Varieties Inside Bounded Product Sets	2
1. Bounded powers of an approximate group	2
2. Counting functions	3
3. Uniform bounded complexity	3
4. Uniform sufficient Zariski density	4
5. The endpoint estimates	5
6. Reducible varieties	5
7. The target estimate	6
Chapter 2. Essential Products and Exceptional Fibers	7
1. A product map in transverse position	7
2. The exceptional-pair variety	8
3. The finite-set dichotomy	9
4. A dimension-pair laboratory	11
Chapter 3. A Multiset Proof of the Larsen–Pink Exponent	13
1. The raw inequality	13
2. Why the divisibility trick matters	17
3. A worked dimension laboratory	17
Chapter 4. Larsen–Pink for Approximate Groups	19
1. The approximate-group inequality	19
2. Centralizers and conjugacy classes	20
3. The rank-one exponent check	21
4. A concrete exponent check in SL_3	22
Part 2. Centralizers and Type-A Growth	24
Chapter 5. Regular Semisimple Elements and Tori in SL_n	25
1. Dimension and the diagonal torus	25
2. Why algebraic tori diagonalize	26

3. Regular semisimple matrices	27
4. The singular locus inside a torus	29
5. The normalizer of an eigenvalue torus	29
6. Positive-dimensional subtori have proper normalizers	30
7. The SL_3 torus as a laboratory	30
Chapter 6. Involved Tori and the Crucial Counting Lemma	32
1. Definition of involvement	32
2. Regular semisimple elements dominate unless A is small	32
3. How many points lie in one involved torus?	33
4. Counting the involved tori	34
5. The promotion argument under conjugation	34
6. The involved-torus exponent ledger	36
Chapter 7. Finite Torus Orbits and Small Zariski-Dense Approximate Groups	37
1. Finite-index subgroups of dense groups	37
2. Finite conjugation orbits	38
3. The smallness theorem	38
4. The contradiction as a reusable pattern	39
5. Why finite orbit is already fatal	40
Chapter 8. From Approximate-Group Rigidity to Product Growth	41
1. The inverse-to-direct conversion	41
2. The type-A product theorem	42
3. Extracting the growth exponent	44
4. Symmetric generating sets and iterated growth	44
5. Why finite fields need one more argument	45
6. Small or dense over a finite field	47
7. An elementary quasirandomness bound for $SL_n(\mathbb{F}_q)$	48
8. The finite-field type-A product theorem	49
9. Logical closure of the Arc-II product-growth argument	51
Appendix A. Simplicity of $PSL_n(k)$ over an Algebraically Closed Field	53
1. A primitive-action simplicity criterion	53
2. Elementary matrices generate SL_n	54
3. Perfectness	54
4. Primitive action on projective space	55
Bibliography	57

Part 1

The Larsen–Pink Induction

CHAPTER 1

Counting Varieties Inside Bounded Product Sets

Larsen–Pink is an interpolation theorem between two endpoint estimates: zero-dimensional varieties contain uniformly boundedly many points, while a full-dimensional closed subvariety of an irreducible algebraic group is the whole group. The difficult point is to obtain the exact exponent $\dim V/\dim G$ uniformly over varieties created during an induction. We therefore begin with bounded powers of approximate groups, bounded algebraic complexity, and the uniform escape statement supplied by Volume V.

Throughout this volume k is algebraically closed unless explicitly stated otherwise, $G \subseteq \mathrm{GL}_N(k)$ is an irreducible affine algebraic group of dimension

$$D = \dim G > 0,$$

and $A \subseteq G(k)$ is finite, symmetric, and contains the identity.

1. Bounded powers of an approximate group

LEMMA 1.1 (Power growth). *If A is a K -approximate group, then for every integer $m \geq 1$,*

$$|A^m| \leq K^{m-1}|A|.$$

Moreover A^m is a K^{2m-1} -approximate group.

PROOF. By definition there is X , $|X| \leq K$, such that $A^2 \subseteq XA$. Inductively,

$$A^m \subseteq X^{m-1}A,$$

which gives the cardinality estimate.

The set A^m is symmetric and contains the identity. Also

$$(A^m)^2 = A^{2m} \subseteq X^{2m-1}A \subseteq X^{2m-1}A^m.$$

Thus it is a K^{2m-1} -approximate group. □

The first assertion was proved already in Volume II; we repeat its two-line proof because the exact power of K will be used repeatedly.

2. Counting functions

For a closed subvariety $V \subseteq G$ and an integer $m \geq 1$, write

$$N_m(V) := |A^m \cap V|.$$

If V is irreducible and $0 < d := \dim V \leq D$, define the normalized counting density

$$\rho_m(V) := N_m(V)^{1/d}.$$

The exponent $1/d$ is not cosmetic. Larsen–Pink predicts

$$N_m(V) \asymp |A|^{d/D},$$

so the predicted scale of $\rho_m(V)$ is independent of V :

$$\rho_m(V) \asymp |A|^{1/D}.$$

We shall ultimately use a multiset version of the dimension-pair induction rather than propagating one normalized density at a time, but this normalization remains the right heuristic.

3. Uniform bounded complexity

DEFINITION 1.2 (Complexity). Fix an integer $M \geq 1$. A closed or constructible subset of a fixed affine space has *complexity at most M* in the sense of Volume V, Appendix B: ambient dimension, number of defining pieces, and degrees of defining equations are all bounded by M . We write $O_M(1)$ for a quantity bounded in terms of M and the fixed ambient embedding of G .

The qualitative algebraic geometry of Volume V is not by itself enough for Larsen–Pink, because the varieties arising during the induction depend on the finite set A . Appendix B of Volume V proved the required compactness upgrade:

- bounded-complexity varieties have $O_M(1)$ irreducible components, each of complexity $O_M(1)$;
- zero-dimensional bounded-complexity constructible sets have $O_M(1)$ points;
- image closures, inverse images, generic-fiber exceptional loci, and generic slices have complexity $O_M(1)$ under a bounded number of bounded-complexity algebraic operations.

These facts will be invoked through the Uniform Algebraic-Geometry Package and the Uniform Zero-Dimensional and Uniform-Components Corollaries proved in Appendix B of Volume V.

4. Uniform sufficient Zariski density

DEFINITION 1.3 (Sufficient Zariski density). A finite set $B \subseteq G(k)$ is *M-sufficiently Zariski dense* if it is not contained in any proper closed subvariety of G of complexity at most M .

DEFINITION 1.4 (Bounded-complexity escape). For integers $M, r \geq 1$, a finite symmetric set $B \subseteq G(k)$ containing e is said to have the (M, r) *bounded-complexity escape property* if

$$B^r \not\subseteq Z$$

for every proper closed subvariety $Z \subsetneq G$ of complexity at most M . Equivalently, the bounded power B^r is M -sufficiently Zariski dense.

This formulation separates the geometric proof from the source of density. Over an algebraically closed field, Zariski density of $\langle B \rangle$ supplies this property by the escape theorem. Over a finite field, the whole finite group is not Zariski dense after scalar extension, but point counting plus generation will supply exactly the same bounded-complexity escape property.

PROPOSITION 1.5 (A bounded power is uniformly sufficiently dense). *Assume $\langle A \rangle$ is Zariski dense in G . For every M there is an integer*

$$r = r(G, M)$$

such that A^r is M -sufficiently Zariski dense.

PROOF. Let $V \subsetneq G$ have complexity at most M . In the fixed matrix embedding, V is cut out by a bounded list of polynomials of uniformly bounded degree. Apply the finite-list escape theorem of Volume V, Chapter 8, to the left-translation action of $\langle A \rangle$ on G and to the point e .

The proof of that theorem bounds the escape depth through the dimension of the finite-dimensional vector space of bounded-degree polynomial equations; consequently the depth is bounded uniformly over the coefficients of the equations defining V . Thus one integer $r = r(G, M)$ works for every proper V of complexity at most M .

Since $\langle A \rangle$ is Zariski dense, its orbit of e is not contained in any such V . Hence $A^r \not\subseteq V$ for every proper V of complexity at most M , which is exactly the asserted sufficient-density statement. $\square$

COROLLARY 1.6 (Uniform escape for a bounded-complexity bad set). *Fix M . If $B \subsetneq G$ is a proper closed subvariety of complexity at most M and $\langle A \rangle$ is Zariski dense, then*

$$A^{r(G, M)} \not\subseteq B.$$

The same word length works for every such B .

PROOF. This is merely the preceding proposition with one particular B selected. $\square$

5. The endpoint estimates

PROPOSITION 1.7 (Zero-dimensional endpoint). *If $Z \subseteq G$ is zero-dimensional and has complexity at most M , then*

$$N_m(Z) \leq |Z(k)| \leq C(M)$$

for every m .

PROOF. A zero-dimensional affine variety over an algebraically closed field is finite by Volume V. The uniform cardinality bound for complexity at most M is the Uniform Zero-Dimensional Bound proved in Appendix B of Volume V. Intersecting with A^m can only decrease the number of points. $\square$

PROPOSITION 1.8 (Full-dimensional endpoint). *Let $V \subseteq G$ be closed and have dimension D . Then $V = G$ and*

$$N_m(V) = |A^m| \leq K^{m-1} |A|.$$

PROOF. Because G is irreducible, every proper closed subset has smaller dimension, by the dimension-drop theorem of Volume V. Hence $V = G$. Apply Lemma 1.1. $\square$

6. Reducible varieties

LEMMA 1.9 (Uniform reduction to irreducible components). *Let V have complexity at most M , and let*

$$V = V_1 \cup \cdots \cup V_s$$

be its irreducible decomposition. Then $s = O_M(1)$ and every V_i has complexity $O_M(1)$. Consequently, if for every irreducible component

$$N_m(V_i) \leq CK^c |A|^{\dim V_i/D},$$

with constants uniform at this complexity, then

$$N_m(V) \leq C_M K^c |A|^{\dim V/D}.$$

PROOF. The component bounds are the Uniform Number of Components Corollary proved in Appendix B of Volume V. Now

$$N_m(V) \leq \sum_i N_m(V_i).$$

Every $\dim V_i \leq \dim V$, and $|A| \geq 1$, so summing over the $O_M(1)$ components gives the claim. $\square$

7. The target estimate

CHECKPOINT 1.10. For an irreducible $V \subseteq G$ of complexity at most M and dimension d , the raw Larsen–Pink estimate to be proved is

$$\boxed{|B \cap V| \leq C_M |B^Q|^{d/D}}$$

for a bounded power $Q = Q(D, M)$ of a sufficiently Zariski-dense symmetric set B . Once this has been established, the approximate-group inequality

$$|A^m \cap V| \ll_M K^{O_{D,M,m}(1)} |A|^{d/D}$$

follows immediately from bounded power growth.

CHAPTER 2

Essential Products and Exceptional Fibers

The geometric step behind Larsen–Pink is simple to state but delicate to make uniform: conjugate one variety into transverse position and multiply. The image gains dimension, so generic fibers lose dimension. Exceptional fibers either concentrate the finite set on a lower-dimensional base or on a lower-dimensional fiber. Volume V supplies all algebraic geometry; the present chapter performs the finite-set counting and keeps every complexity and word-length loss uniform.

Throughout this chapter assume, in addition to the standing hypotheses of Chapter 1, that G is almost simple in the sense of Volume V: it has no proper positive-dimensional closed normal algebraic subgroup. This is exactly the hypothesis required by the product–conjugation theorem and will be present in the raw Larsen–Pink theorem of the next chapter.

Fix a complexity bound M . Let $V_-, V_+ \subseteq G$ be irreducible closed subvarieties of complexity at most M with

$$0 < d_- := \dim V_- \leq d_+ := \dim V_+ < D.$$

1. A product map in transverse position

For $a \in G$ write

$$V_-^a = a^{-1}V_-a$$

and consider

$$\mu_a : V_- \times V_+ \longrightarrow G, \quad \mu_a(x, y) = a^{-1}xay.$$

PROPOSITION 2.1 (Uniform transverse conjugation). *There is a proper closed subvariety*

$$\mathcal{B} = \mathcal{B}(V_-, V_+) \subsetneq G$$

of complexity at most $M_1 = M_1(G, M)$ such that for every $a \notin \mathcal{B}$,

$$W_a := \overline{V_-^a V_+}$$

satisfies

$$\dim W_a > d_+.$$

Consequently, if A has the (M_1, r) bounded-complexity escape property, one may choose such an a in A^r .

PROOF. The product–conjugation theorem of Volume V says that the bad conjugators lie in a proper closed subvariety. The uniform-complexity package of Volume V bounds the complexity of that bad set by a number $M_1 = M_1(G, M)$. In particular, Appendix B of Volume V proves uniform complexity for setwise stabilizers under bounded-complexity algebraic actions; the remaining operations here are products, inverses, conjugation, image closure, and containment conditions, all covered by the same package. If A has the (M_1, r) bounded-complexity escape property, then A^r is not contained in $\mathcal{B}$, so it contains the required transverse conjugator. $\square$

Fix such an a , put $W = W_a$, and write

$$w = \dim W.$$

Then

$$d_+ < w \leq d_- + d_+.$$

The expected fiber dimension of the dominant morphism

$$\mu_a : V_- \times V_+ \rightarrow W$$

is

$$s = d_- + d_+ - w < d_-.$$

2. The exceptional-pair variety

LEMMA 2.2 (Uniform good-fiber set). *There are proper closed subvarieties*

$$Z \subsetneq W, \quad S = \mu_a^{-1}(Z) \subsetneq V_- \times V_+$$

of complexity $O_M(1)$ such that:

- (1) $\dim S < d_- + d_+$;
- (2) *for every $z \in W \setminus Z$, the fiber $\mu_a^{-1}(z)$ has dimension s ;*
- (3) *for such z , projection of the fiber to the first factor is an isomorphism onto a constructible subset of V_- whose closure has dimension $s < d_-$ and complexity $O_M(1)$.*

PROOF. By the generic fiber theorem of Volume V, there is a nonempty open subset $W^\circ \subseteq W$ on which every fiber has dimension

$$\dim(V_- \times V_+) - \dim W = s.$$

The uniform package of Volume V allows the exceptional complement

$$Z = W \setminus W^\circ$$

and its inverse image $S = \mu_a^{-1}(Z)$ to be chosen with complexity $O_M(1)$. Since μ_a is dominant and Z is proper, S is a proper closed subset of the irreducible variety $V_- \times V_+$, hence has dimension $< d_- + d_+$.

For fixed z , the equation

$$a^{-1}xay = z$$

determines y uniquely from x :

$$y = (a^{-1}xa)^{-1}z.$$

Thus first projection is an isomorphism from the fiber to its image, with the displayed regular inverse. The closure of that image has the same dimension s , and its complexity is $O_M(1)$ by the uniform image-closure package of Volume V. $\square$

3. The finite-set dichotomy

For the iterative argument it is convenient to apply the geometric step not to A itself but to whatever common bounded power is currently being used. Thus fix $m \geq 1$ and put

$$X = A^m \cap V_-, \quad Y = A^m \cap V_+.$$

The next lemma records exactly how much bounded-complexity escape is needed to choose the transverse conjugator.

LEMMA 2.3 (Uniform dimension-pair counting step). *There is a complexity threshold $M_2 = M_2(G, M)$ with the following property. Suppose that for some $r_0 \geq 1$ the set A has the (M_2, r_0) bounded-complexity escape property. Then there are irreducible subvarieties*

$$\tilde{V}_-, \tilde{V}_+ \subseteq G$$

of complexity $O_M(1)$ and an integer $L \leq c_{M, r_0} m$ such that

$$(6.1) \quad N_m(V_-)N_m(V_+) \leq C_M N_L(\tilde{V}_-)N_L(\tilde{V}_+),$$

and, after interchanging the tilded varieties if necessary, their dimensions

$$\tilde{d}_- \leq \tilde{d}_+$$

satisfy either

(I)

$$\tilde{d}_- < d_-, \quad \tilde{d}_+ > d_+, \quad \tilde{d}_- + \tilde{d}_+ = d_- + d_+;$$

or

(II)

$$\tilde{d}_- + \tilde{d}_+ < d_- + d_+.$$

Zero-dimensional tilded varieties are allowed.

PROOF. Let $M_1(G, M)$ be the complexity threshold in Proposition 2.1, and choose $M_2 \geq M_1(G, M)$. Apply that proposition to the symmetric set A^m . Since $e \in A$, one has

$$A^{r_0} \subseteq (A^m)^{r_0};$$

hence the (M_2, r_0) escape property supplies a transverse conjugator

$$a \in (A^m)^{r_0} = A^{r_0 m}.$$

For this a , the product map sends $X \times Y$ into

$$A^L \cap W, \quad L = 2m + 2r_0m \leq c_{M,r_0}m.$$

Let S be the exceptional set of [Lemma 2.2](#). There are two main cases.

Case 1: at least half the pairs lie outside S . Then

$$\frac{1}{2}|X||Y| \leq |(X \times Y) \setminus S|.$$

Group these pairs according to their image $z = \mu_a(x, y) \in A^L \cap W$. If M_z denotes the number of represented pairs in the fiber over z , then

$$\frac{1}{2}|X||Y| \leq \sum_{z \in A^L \cap W} M_z \leq |A^L \cap W| \max_z M_z.$$

Choose z_0 maximizing M_z . Since this fiber contains a pair outside S , $z_0 \notin Z$. By [Lemma 2.2](#), first projection identifies the fiber with a constructible subset whose closure $F \subseteq V_-$ has

$$\dim F = s < d_-$$

and complexity $O_M(1)$. By the component theorem of Volume V, F has at most C_M irreducible components, each of complexity $O_M(1)$. Choose one component $\tilde{V}_-$ containing at least a C_M^{-1} fraction of the fiber points, and set

$$\tilde{V}_+ = W.$$

Then

$$\tilde{d}_- \leq s, \quad \tilde{d}_+ = w > d_+,$$

and

$$\tilde{d}_- + \tilde{d}_+ \leq s + w = d_- + d_+.$$

Equality gives alternative (I); strict inequality gives (II). The preceding counting display gives (6.1) after absorbing the uniform component count into C_M .

Case 2: more than half the pairs lie in S . Thus

$$|(X \times Y) \cap S| > \frac{1}{2}|X||Y|.$$

Apply the generic slicing theorem of Volume V, together with its uniform-complexity appendix, to

$$S \subseteq V_- \times V_+.$$

There is a proper closed subvariety

$$V_-^{\text{exc}} \subsetneq V_-$$

of complexity $O_M(1)$ such that for every $x \notin V_-^{\text{exc}}$ the fiber

$$S_x = \{y \in V_+ : (x, y) \in S\}$$

has dimension $< d_+$ and complexity $O_M(1)$.

Put

$$S_0 = S \cap (V_-^{\text{exc}} \times V_+).$$

At least half of the S -pairs lie either in S_0 or in $S \setminus S_0$.

If at least half lie in S_0 , then

$$\frac{1}{4}|X||Y| < |(X \times Y) \cap S_0| \leq N_m(V_-^{\text{exc}})N_m(V_+).$$

The exceptional base has at most C_M irreducible components of complexity $O_M(1)$. Choose one carrying a maximal share of the A^m -points as $\tilde{V}_-$ and put $\tilde{V}_+ = V_+$. Since $\dim \tilde{V}_- < d_-$, the total dimension drops, so alternative (II) holds.

Otherwise at least one quarter of all pairs lie in $S \setminus S_0$. Summing the fiber sizes over $x \in X \setminus V_-^{\text{exc}}$ gives some x_0 for which

$$|A^m \cap S_{x_0}| \geq \frac{|Y|}{4}.$$

The fiber closure has complexity $O_M(1)$ and therefore at most C_M irreducible components. Choose one component F carrying a maximal share and, after reordering if necessary, set

$$\{\tilde{V}_-, \tilde{V}_+\} = \{V_-, F\}.$$

Because $\dim F < d_+$, the total dimension is strictly smaller than $d_- + d_+$. Also

$$|X||Y| \leq C_M N_m(V_-) N_m(F),$$

which is again (6.1) and alternative (II).

All possibilities are exhausted. $\square$

4. A dimension-pair laboratory

Take an ambient group of dimension $D = 8$ and suppose the current pair has

$$(d_-, d_+) = (2, 5).$$

A transverse product image might have dimension $w = 6$. Then the generic fiber has dimension

$$s = 2 + 5 - 6 = 1,$$

so Case 1 replaces the pair by

$$(1, 6).$$

The total dimension remains 7, but the two dimensions have moved farther apart. If instead the finite set concentrates on the exceptional-pair variety, Case 2 might produce, for example,

$$(2, 4)$$

or

$$(1, 5),$$

and the total dimension falls below 7.

Thus every step makes progress in one of exactly two currencies:

total dimension decreases

or

fixed total dimension becomes more polarized

.

This is the mechanism behind the termination invariant used in the next chapter. No numerical estimate on the coefficients of the varieties is needed; uniform complexity ensures only that the constants and word lengths do not escape while this finite dimension game is played.

CHECKPOINT 2.4. The geometric content has been converted into a uniform combinatorial move on a pair of dimensions. A pair

$$0 < d_- \leq d_+ < D$$

is replaced by another pair whose total dimension never increases. If the total dimension stays fixed, the two dimensions move strictly farther apart. Meanwhile the product of the two intersection counts is preserved up to a constant and the common word length is multiplied by at most c_M .

CHAPTER 3

A Multiset Proof of the Larsen–Pink Exponent

The original dimension-pair proof is often presented through a delicate normalized-density iteration. We use a more global bookkeeping device. Start with enough copies of the same variety that the total dimension is divisible by $\dim G$, repeatedly replace any two intermediate-dimensional factors by the pair supplied by the essential-product lemma, and keep the product of all intersection counts. Total dimension never increases; when it stays fixed the sum of squared dimensions strictly increases. Thus the process terminates at the two trivial endpoints, and divisibility forces the exponent $\dim V / \dim G$ exactly.

1. The raw inequality

THEOREM 3.1 (Raw Larsen–Pink under bounded-complexity escape). *Let $G \subseteq \mathrm{GL}_N(k)$ be irreducible and almost simple, of dimension D , and let $V \subseteq G$ be a closed subvariety of complexity at most M and dimension d . There is a complexity threshold*

$$M_* = M_*(G, M)$$

with the following property. Let $A \subseteq G(k)$ be finite, symmetric, contain e , and suppose that for some $r_0 \geq 1$ it has the (M_*, r_0) bounded-complexity escape property of Definition 1.4. Then there are constants

$$Q = Q(G, M, r_0), \quad C = C(G, M)$$

such that

$$(6.2E) \quad |A \cap V| \leq C |A^Q|^{d/D}.$$

For fixed ambient dimension and complexity, Q and C are uniform over the coefficients defining V and over the algebraically closed field k .

PROOF. By Lemma 1.9, it is enough to treat irreducible V ; enlarging M by a uniform function handles the components. The cases $d = 0$ and $d = D$ are the endpoint estimates Propositions 1.7 and 1.8, so assume

$$0 < d < D.$$

Before starting the iteration, fix its algebraic-complexity budget. The lexicographic argument below shows that at most

$$T_D := (D^2 + 1)(D^3 + 1)$$

replacement steps can occur (this deliberately crude number dominates all possible values of the two integer invariants). Starting from the initial complexity M , recursively apply the uniform-complexity bounds of Volume V and the threshold $M_2(G, \cdot)$ in [Lemma 2.3](#) for T_D stages. Let $M_* = M_*(G, M)$ be the maximum of this finite list of thresholds. Thus the single (M_*, r_0) escape hypothesis is strong enough to provide every transverse conjugator that can arise during the entire iteration.

Set

$$n := \frac{D}{\gcd(D, d)}.$$

Then $1 \leq n \leq D$ and, crucially,

$$(6.3) \quad nd \equiv 0 \pmod{D}.$$

Write

$$X := |A \cap V|.$$

We begin with a multiset of n copies of V :

$$\mathcal{V}_0 = (V, \dots, V),$$

a common word length $m_0 = 1$, and the exact identity

$$(6.4) \quad X^n = \prod_{U \in \mathcal{V}_0} N_{m_0}(U).$$

We now perform a finite replacement procedure. At stage t we have a multiset

$$\mathcal{V}_t = (V_{t,1}, \dots, V_{t,n})$$

of irreducible subvarieties, all of complexity at most some bound M_t , a common word length m_t , and a constant C_t such that

$$(6.5) \quad X^n \leq C_t \prod_{j=1}^n N_{m_t}(V_{t,j}).$$

The quantities M_t, m_t, C_t are allowed to grow, but only by functions of the previous complexity and of G .

If fewer than two members of $\mathcal{V}_t$ have dimensions strictly between 0 and D , stop. Otherwise choose two such members and reorder them as

$$0 < d_- \leq d_+ < D.$$

Apply the uniform dimension-pair counting step, [Lemma 2.3](#), with $m = m_t$ and the fixed escape exponent r_0 . The choice of M_* above guarantees its bounded-complexity

escape hypothesis at this stage. It gives two irreducible replacements $\tilde{V}_-, \tilde{V}_+$, of complexity bounded by a function of M_t , and a word length

$$m_{t+1} \leq c_{M_t, r_0} m_t$$

such that

$$N_{m_t}(V_-)N_{m_t}(V_+) \leq C_{M_t} N_{m_{t+1}}(\tilde{V}_-) N_{m_{t+1}}(\tilde{V}_+).$$

Replace the selected two varieties by the tilded pair. For every unselected factor U we have

$$N_{m_t}(U) \leq N_{m_{t+1}}(U)$$

because $m_{t+1} \geq m_t$. Thus (6.5) persists after multiplying C_t by C_{M_t} .

It remains to prove that the replacement procedure terminates after a number of stages bounded solely in terms of D . Let

$$S_t := \sum_{j=1}^n \dim V_{t,j}, \quad R_t := \sum_{j=1}^n (\dim V_{t,j})^2.$$

At every replacement, S_t does not increase. If alternative (II) of Lemma 2.3 occurs, S_t strictly decreases. If alternative (I) occurs, the sum of the two selected dimensions is preserved but they move strictly farther apart. More explicitly, if

$$a + b = u + v, \quad a < u \leq v < b,$$

then

$$a^2 + b^2 - (u^2 + v^2) = 2(u - a)(b - u) > 0.$$

Hence in alternative (I), S_t stays fixed and R_t strictly increases.

The pair

$$(S_t, -R_t)$$

therefore strictly decreases in lexicographic order. Both coordinates range over finite sets of integers:

$$0 \leq S_t \leq nD, \quad 0 \leq R_t \leq nD^2.$$

Consequently the procedure terminates after $O_D(1)$ stages. In particular, all complexities M_t , all word lengths m_t , and the accumulated constant C_t are bounded by functions of G , the initial M , and (for the word lengths) r_0 . Denote the final word length by Q and the final constant by C_1 .

At termination every member of the final multiset has dimension 0 or D , except possibly one member of an intermediate dimension. Let

$$f = \#\{j : \dim V_{T,j} = D\}$$

and let $\varepsilon \in \{0, 1\}$ record whether an intermediate-dimensional member remains. If $\varepsilon = 1$, write its dimension as r with $0 < r < D$. The final total dimension is

$$S_T = fD + \varepsilon r.$$

The zero-dimensional factors contribute at most a uniform constant by [Proposition 1.7](#); every full-dimensional factor is G and contributes at most $|A^Q|$; and the possible remaining intermediate factor also contributes at most $|A^Q|$. Therefore (6.5) gives

$$(6.6) \quad X^n \leq C_2 |A^Q|^{f+\varepsilon}.$$

If $\varepsilon = 0$, then

$$f = \frac{S_T}{D}.$$

If $\varepsilon = 1$, then

$$f + 1 = \left\lceil \frac{S_T}{D} \right\rceil.$$

Thus in every case

$$(6.7) \quad f + \varepsilon = \left\lceil \frac{S_T}{D} \right\rceil.$$

Total dimension never increased, so

$$S_T \leq S_0 = nd.$$

By (6.3), nd/D is an integer. Hence

$$\left\lceil \frac{S_T}{D} \right\rceil \leq \frac{nd}{D}.$$

Substituting into (6.6) gives

$$X^n \leq C_2 |A^Q|^{nd/D}.$$

Taking n th roots yields

$$|A \cap V| = X \leq C_2^{1/n} |A^Q|^{d/D},$$

which is (6.2E) after renaming the constant. $\square$

THEOREM 3.2 (Raw Larsen–Pink inequality). *Let $G \subseteq \mathrm{GL}_N(k)$ be irreducible and almost simple, of dimension D , and let $V \subseteq G$ be a closed subvariety of complexity at most M and dimension d . Let $A \subseteq G(k)$ be finite, symmetric, contain e , and generate a Zariski-dense subgroup. Then there are constants*

$$Q = Q(G, M), \quad C = C(G, M)$$

such that

$$(6.2) \quad |A \cap V| \leq C |A^Q|^{d/D}.$$

For fixed ambient dimension and complexity, Q and C are uniform over the coefficients defining V and over the algebraically closed field k .

PROOF. Let $M_* = M_*(G, M)$ be the threshold supplied by Theorem 3.1. By Proposition 1.5, Zariski density of $\langle A \rangle$ gives an integer $r_0 = r(G, M_*)$ for which A has the (M_*, r_0) bounded-complexity escape property. Apply Theorem 3.1. Since r_0 depends only on G and M , so does the resulting bounded word exponent Q . $\square$

2. Why the divisibility trick matters

The terminal step is the only place where the exact exponent could have been lost. If we had started with just one copy of V , termination might leave one intermediate-dimensional factor, producing only the weaker exponent

$$\frac{d}{D} + O_D(1).$$

Starting with

$$n = \frac{D}{\gcd(D, d)}$$

copies makes the initial total dimension a multiple of D . Since the dimension bookkeeping never increases total dimension, the ceiling in (6.7) can never exceed the exact initial quotient nd/D . This elementary divisibility device replaces the more delicate normalized-density termination argument often used in presentations of Larsen–Pink.

3. A worked dimension laboratory

Suppose $D = 8$ and $d = 3$. Then $n = 8$ and the initial multiset has total dimension

$$8 \cdot 3 = 24 = 3D.$$

No matter how the essential-product steps branch, the final total dimension is at most 24. At termination there are at most three nonzero endpoint slots in the sense of (6.7): for example

$$(8, 8, 5, 0, 0, 0, 0, 0)$$

has terminal total dimension 21 and contributes at most three powers of $|A^Q|$, while

$$(8, 8, 8, 0, 0, 0, 0, 0)$$

contributes exactly three. Thus

$$|A \cap V|^8 \ll |A^Q|^3,$$

and therefore

$$|A \cap V| \ll |A^Q|^{3/8}.$$

The argument does not need to know which sequence of intermediate dimensions occurred.

EXAMPLE 3.3 (Centralizer and conjugacy exponents in SL_3). Let $G = \mathrm{SL}_3$ over an algebraically closed field and let

$$t = \mathrm{diag}(a, b, (ab)^{-1})$$

with three distinct diagonal entries. Any matrix commuting with t preserves its three one-dimensional eigenspaces, so

$$C_G(t) = \{\mathrm{diag}(x, y, (xy)^{-1}) : x, y \neq 0\}.$$

Thus $\dim C_G(t) = 2$, while $\dim G = 3^2 - 1 = 8$. The conjugacy class of t therefore has dimension

$$\dim(t^G) = 8 - 2 = 6.$$

The Larsen–Pink exponent predicts the complementary pair of bounds

$$|A \cap C_G(t)| \ll |A^Q|^{2/8} = |A^Q|^{1/4}, \quad |A \cap t^G| \ll |A^Q|^{6/8} = |A^Q|^{3/4}.$$

The exponents add to one because the conjugation map $G \rightarrow t^G$ has a centralizer fiber. This is the numerical model behind the later centralizer–conjugacy counting argument.

G	$\longrightarrow$	t^G	
g	$\mapsto$	gtg^{-1}	
$\dim G = 8$		$\dim t^G = 6$	$\frac{2}{8} + \frac{6}{8} = 1.$
$\cup$			
$C_G(t) \text{ fiber, } \dim = 2$			

FIGURE 1. The dimension split behind the SL_3 Larsen–Pink exponents. Counting a fiber and counting its conjugacy orbit use complementary fractions of the ambient dimension.

CHECKPOINT 3.4. The exponent

$$\frac{\dim V}{\dim G}$$

has now been derived entirely from three ingredients already proved in the series:

$$\begin{aligned} & \boxed{\text{product-conjugation}} \\ & + \boxed{\text{generic fibers and uniform complexity}} \\ & + \boxed{\text{bounded-word escape}}. \end{aligned}$$

No Larsen–Pink theorem has been imported as a source result.

CHAPTER 4

Larsen–Pink for Approximate Groups

The raw theorem controls intersections by a bounded product set. Approximate groups convert that product set into a power of the doubling parameter, yielding the familiar Larsen–Pink inequality. We then derive the two-sided centralizer estimate used in all torus arguments: the upper bound is Larsen–Pink, while the lower bound comes from the conjugation map and a pigeonhole argument.

1. The approximate-group inequality

THEOREM 4.1 (Larsen–Pink inequality). *Let $G \subseteq \mathrm{GL}_N(k)$ be irreducible and almost simple, of dimension D , over an algebraically closed field. Fix a complexity bound M and an integer $m \geq 1$. Let $A \subseteq G(k)$ be a finite K -approximate group such that $\langle A \rangle$ is Zariski dense. Then every closed subvariety $V \subseteq G$ of complexity at most M satisfies*

$$(6.8) \quad |A^m \cap V| \leq C_{G,M,m} K^{C_{G,M,m}} |A|^{\dim V/D}.$$

PROOF. Set $B = A^m$. It is finite, symmetric, contains the identity, and generates the same Zariski-dense subgroup as A . Apply the raw Larsen–Pink theorem, [Theorem 3.2](#), to B :

$$|A^m \cap V| = |B \cap V| \leq C_{G,M} |B^Q|^{\dim V/D}.$$

But

$$B^Q = A^{mQ},$$

and bounded power growth gives

$$|A^{mQ}| \leq K^{mQ-1} |A|.$$

Substitution yields

$$|A^m \cap V| \leq C_{G,M} K^{(mQ-1) \dim V/D} |A|^{\dim V/D},$$

which has the stated form after enlarging the exponent of K uniformly. $\square$

REMARK 4.2. The exponent of $|A|$ is exact. All losses caused by component selection, exceptional fibers, bounded-word escape, and the number of iterations have been confined to the polynomial factor in K and the algebraic constant.

$V \subseteq G, \quad d = \dim V, \quad D = \dim G$	
$\Downarrow \quad \text{bounded product maps + exceptional-fiber control}$	
$ A^m \cap V \lesssim K^C A ^{d/D}$	
centralizer $C_G(g)$ $\dim = z$	conjugacy class g^G $\dim = D - z$
$ A^m \cap C_G(g) \lesssim K^C A ^{z/D}$	$ A^m \cap g^G \lesssim K^C A ^{(D-z)/D}$

FIGURE 1. Larsen–Pink as dimension bookkeeping. A d -dimensional algebraic slice is allowed the fraction d/D of the ambient approximate-group mass; centralizers and conjugacy classes have complementary exponents.

2. Centralizers and conjugacy classes

Let $a \in G(k)$ and write

$$Z(a) = \{g \in G : ga = ag\}.$$

Its conjugacy class is

$$C(a) = \{g^{-1}ag : g \in G\}.$$

Volume V proved the orbit–stabilizer dimension formula for algebraic group actions, so

$$(6.9) \quad \dim C(a) = D - \dim Z(a).$$

PROPOSITION 4.3 (Two-sided centralizer concentration). *Fix $m \geq 1$ and a complexity bound M large enough to contain the centralizers and conjugacy classes under consideration. Let A be as in Theorem 4.1, and let*

$$a \in A^m.$$

Put

$$z = \dim Z(a).$$

Then

$$(6.10) \quad K^{-C} |A|^{z/D} \leq |A^2 \cap Z(a)| \leq |A^{m'} \cap Z(a)| \leq K^C |A|^{z/D}$$

for every fixed $m' \geq 2$, with C depending on G, M, m, m' .

PROOF. The upper bound is the Larsen–Pink inequality Theorem 4.1 applied to $Z(a)$.

For the lower bound consider the conjugation map on the finite set A :

$$\phi : A \rightarrow G, \quad \phi(g) = g^{-1}ag.$$

Since $a \in A^m$ and A is symmetric,

$$\phi(A) \subseteq A^{m+2} \cap C(a).$$

By Larsen–Pink and (6.9),

$$|\phi(A)| \leq K^{C_1} |A|^{(D-z)/D}.$$

The fibers of ϕ partition A , so one fiber contains at least

$$\frac{|A|}{|\phi(A)|} \geq K^{-C_1} |A|^{z/D}$$

points. Fix g_0 in that fiber. If g is another point of the same fiber, then

$$g^{-1}ag = g_0^{-1}ag_0,$$

which rearranges to

$$gg_0^{-1} \in Z(a).$$

Moreover $gg_0^{-1} \in A^2$. The map $g \mapsto gg_0^{-1}$ is injective, hence

$$|A^2 \cap Z(a)| \geq K^{-C_1} |A|^{z/D}.$$

This proves the lower bound. The middle inequality follows from $e \in A$ and $m' \geq 2$. $\square$

COROLLARY 4.4 (Size of a conjugacy image). *Under the same hypotheses,*

$$K^{-C} |A|^{1-z/D} \leq |\{g^{-1}ag : g \in A\}| \leq K^C |A|^{1-z/D}.$$

PROOF. The upper bound was proved in the preceding argument. For the lower bound, every fiber of ϕ injects after division by one chosen element into $A^2 \cap Z(a)$, whose cardinality is at most $K^C |A|^{z/D}$ by Larsen–Pink. Hence

$$|A| \leq |\phi(A)| K^C |A|^{z/D},$$

which rearranges to the desired lower estimate. $\square$

3. The rank-one exponent check

For $G = \mathrm{SL}_2$, Volume IV proved

$$D = 3, \quad \dim T = 1$$

for a maximal torus T , while a regular semisimple conjugacy class has dimension 2. Thus [Theorem 4.1](#) recovers

$$|A^m \cap T| \ll K^C |A|^{1/3}$$

and

$$|A^m \cap C(g)| \ll K^C |A|^{2/3}.$$

The exponents that appeared concretely in Helfgott's argument are therefore not rank-one coincidences. They are the finite-set shadow of the algebraic dimension ratios

$$\frac{\dim T}{\dim G}, \quad \frac{\dim C(g)}{\dim G}.$$

4. A concrete exponent check in SL_3

For

$$G = \mathrm{SL}_3, \quad D = 8,$$

a regular semisimple centralizer is a two-dimensional torus. Thus the Larsen–Pink prediction is

$$|A^m \cap T| \asymp_{K^{O(1)}} |A|^{2/8} = |A|^{1/4}.$$

The corresponding conjugacy class has dimension $8 - 2 = 6$, so

$$|A^m \cap C(g)| \ll K^{O(1)} |A|^{3/4}.$$

The conjugation map

$$A \rightarrow C(g), \quad x \mapsto x^{-1}gx,$$

has fibers comparable to $A^2 \cap T$. Consequently the exponents $1/4$ and $3/4$ are complementary for a structural reason:

$$\frac{\dim T}{\dim G} + \frac{\dim C(g)}{\dim G} = 1.$$

This is the higher-rank analogue of the $1/3$ – $2/3$ split encountered in Volume IV.

EXAMPLE 4.5 (The $1/3$ – $2/3$ exponents in SL_2). Take

$$G = \mathrm{SL}_2(\mathbb{F}_q), \quad T = \left\{ \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} : t \in \mathbb{F}_q^\times \right\}.$$

Then $\dim G = 3$ and $\dim T = 1$. A regular diagonal element $g = \mathrm{diag}(t, t^{-1})$ with $t \neq \pm 1$ has centralizer exactly T , so its conjugacy class has dimension 2. The Larsen–Pink exponents therefore predict

$$|A^m \cap T| \approx |A|^{1/3}, \quad |A^m \cap C(g)| \approx |A|^{2/3}$$

up to powers of the approximate-group constant. The two exponents are not mysterious constants from a counting trick: they are simply

$$\frac{\dim T}{\dim G} = \frac{1}{3}, \quad \frac{\dim C(g)}{\dim G} = \frac{2}{3}.$$

Thus the familiar rank-one calculation is the first concrete instance of the general dimension-ratio principle proved in this volume.

CHECKPOINT 4.6. The Larsen–Pink mechanism is now complete, including the quantitative uniformity that is usually hidden inside $O_M(1)$ notation. The remaining task in this volume is group-theoretic rather than geometric: identify a robust family of low-dimensional centralizers, show that approximate groups involve many of them, and use conjugation invariance to force either algebraic concentration or polynomial smallness.

Part 2

Centralizers and Type-A Growth

CHAPTER 5

Regular Semisimple Elements and Tori in SL_n

Larsen–Pink is geometric; the product-growth argument needs a group-theoretic family of subvarieties whose conjugates can be counted. For SL_n that family is completely explicit. We prove from linear algebra that algebraic tori diagonalize, that a matrix with distinct eigenvalues has a torus as its full centralizer, that the nonregular locus inside that torus has codimension one, and that the torus normalizer has quotient the symmetric group. No classification of reductive groups is used.

Throughout Chapters 5–8 we specialize to

$$G = \mathrm{SL}_n(k), \quad n \geq 2,$$

over an algebraically closed field k . Appendix A proves directly that $\mathrm{PSL}_n(k)$ is simple and hence that SL_n has no proper positive-dimensional closed normal subgroup. Thus the almost-simplicity hypothesis used in the Larsen–Pink chapters is satisfied.

1. Dimension and the diagonal torus

PROPOSITION 5.1 (Dimension of SL_n). *The affine algebraic group SL_n is irreducible of dimension*

$$D = n^2 - 1.$$

PROOF. Consider the open subset $U \subseteq \mathrm{SL}_n$ on which the upper-left $(n-1) \times (n-1)$ minor is nonzero. Write a matrix in block form

$$g = \begin{pmatrix} A & b \\ c^t & d \end{pmatrix}, \quad A \in M_{n-1}(k).$$

On the open set $\det A \neq 0$, the Schur-complement identity gives

$$\det g = \det A (d - c^t A^{-1} b).$$

Thus the equation $\det g = 1$ determines d uniquely and regularly from the other $n^2 - 1$ coordinates after localizing at $\det A$. Hence U is isomorphic to a nonempty open subset of $\mathbb{A}^{n^2-1}$ and has dimension $n^2 - 1$.

It remains to justify irreducibility rather than infer it from one chart. Let R be the polynomial ring in all matrix entries except x_{nn} . Expanding the determinant in the

final entry gives

$$\det X - 1 = Cx_{nn} + B,$$

where C is the upper-left $(n-1) \times (n-1)$ determinant and $B \in R$. Any common divisor of C and B in the UFD R must, up to a unit, lie in the subring generated by the entries of that upper-left block, because C lies there and adjoining the remaining variables preserves its factorization. Setting all entries in the last row and last column except x_{nn} equal to zero sends B to -1 . Hence C and B are coprime. Thus the linear polynomial $Cx_{nn} + B$ is primitive in $R[x_{nn}]$ and is irreducible over $\text{Frac}(R)$; Gauss's lemma makes it irreducible in the full polynomial ring. Therefore its zero set SL_n is irreducible by the prime-ideal criterion proved in Volume V.

The nonempty open chart U consequently has the same dimension as SL_n . Since U is isomorphic to an open subset of $\mathbb{A}^{n^2-1}$, we obtain $\dim \text{SL}_n = n^2 - 1$. $\square$

DEFINITION 5.2 (Diagonal torus). The standard diagonal torus is

$$T_0 = \left\{ \text{diag}(t_1, \dots, t_n) : t_i \in k^\times, \prod_{i=1}^n t_i = 1 \right\}.$$

PROPOSITION 5.3. *The group T_0 is isomorphic to $(\mathbb{G}_m)^{n-1}$ and has dimension*

$$r = n - 1.$$

PROOF. The map

$$(k^\times)^{n-1} \rightarrow T_0, \quad (t_1, \dots, t_{n-1}) \mapsto \text{diag}(t_1, \dots, t_{n-1}, (t_1 \cdots t_{n-1})^{-1})$$

is an algebraic-group isomorphism with algebraic inverse given by the first $n-1$ diagonal entries. $\square$

2. Why algebraic tori diagonalize

We need one elementary representation-theoretic fact about tori. We prove it rather than invoking the structure theory of reductive groups.

PROPOSITION 5.4 (Weight decomposition for a split torus). *Let*

$$T = (\mathbb{G}_m)^s$$

and let

$$\rho : T \rightarrow \text{GL}(V)$$

be a finite-dimensional algebraic representation over k . Then

$$V = \bigoplus_{\chi \in \mathbb{Z}^s} V_\chi,$$

with only finitely many nonzero summands, where

$$V_\chi = \{v : \rho(t)v = t_1^{\chi_1} \cdots t_s^{\chi_s} v \text{ for every } t \in T\}.$$

In particular every algebraic torus in $\mathrm{GL}_n(k)$ is conjugate to a subgroup of the diagonal matrices.

PROOF. Choose a basis of V . For fixed $v \in V$, the vector-valued regular function

$$t \mapsto \rho(t)v$$

on $(k^\times)^s$ has coordinates in the Laurent polynomial ring

$$k[t_1^{\pm 1}, \dots, t_s^{\pm 1}].$$

Hence it has a finite expansion

$$(6.11) \quad \rho(t)v = \sum_{\chi \in F} t^\chi v_\chi, \quad t^\chi = t_1^{\chi_1} \cdots t_s^{\chi_s}.$$

Use the representation identity $\rho(ut)v = \rho(u)\rho(t)v$. Expanding both sides by (6.11) and comparing Laurent monomials in the independent variables t_i gives

$$\rho(u)v_\chi = u^\chi v_\chi.$$

Thus each coefficient vector lies in the claimed weight space. Setting $t = (1, \dots, 1)$ shows $v = \sum_\chi v_\chi$, and uniqueness of Laurent expansions makes the sum direct.

If an algebraic torus S is isomorphic to $(\mathbb{G}_m)^s$, apply the result to its defining representation on k^n and choose a basis adapted to the weight-space decomposition. Every element of S is diagonal in that basis. $\square$

3. Regular semisimple matrices

DEFINITION 5.5. An element $g \in \mathrm{SL}_n(k)$ is *regular semisimple* if its characteristic polynomial has n distinct roots in k , equivalently if it has n distinct eigenvalues.

Let

$$\Delta(g)$$

denote the discriminant of the characteristic polynomial of g . The coefficients of the characteristic polynomial are polynomial functions of the matrix entries, and the discriminant is a polynomial in those coefficients.

PROPOSITION 5.6 (Regular semisimple locus). *The set*

$$G_{\mathrm{rss}} = \{g \in \mathrm{SL}_n(k) : \Delta(g) \neq 0\}$$

is a nonempty Zariski-open subset of G . Its complement is a proper hypersurface and therefore has dimension at most $D - 1$ and complexity $O_n(1)$.

PROOF. The condition $\Delta(g) = 0$ is polynomial. It remains to show that the discriminant is not identically zero on SL_n . Since k is algebraically closed, it is infinite. Choose distinct nonzero scalars

$$\lambda_1, \dots, \lambda_{n-1}$$

so that the scalar

$$\lambda_n = (\lambda_1 \cdots \lambda_{n-1})^{-1}$$

is also distinct from them; this is possible after avoiding finitely many choices. The diagonal matrix with these entries lies in SL_n and has nonzero discriminant. Hence the zero set is proper. A proper hypersurface in the irreducible D -dimensional group has dimension at most $D - 1$ by the dimension theory of Volume V. The degree of the discriminant depends only on n , giving the complexity bound. $\square$

PROPOSITION 5.7 (Centralizer of a regular semisimple matrix). *Let $g \in \mathrm{SL}_n(k)$ be regular semisimple. Then there is $x \in \mathrm{SL}_n(k)$ such that*

$$x^{-1}gx = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$$

with the λ_i distinct, and

$$Z_G(g) = xT_0x^{-1}.$$

In particular

$$\dim Z_G(g) = r = n - 1.$$

PROOF. Distinct eigenvalues give a basis of eigenvectors, so g is diagonalizable. The change-of-basis matrix may be rescaled by a scalar so that its determinant is 1; algebraic closedness supplies an n th root of the inverse determinant.

It therefore suffices to centralize a diagonal matrix

$$d = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$$

with distinct entries. If $h = (h_{ij})$, then $hd = dh$ gives

$$(\lambda_j - \lambda_i)h_{ij} = 0.$$

Thus every off-diagonal entry of h is zero. Conversely every diagonal determinant-one matrix commutes with d . Hence $Z_G(d) = T_0$, and conjugating back proves the claim. $\square$

DEFINITION 5.8 (Eigenvalue torus). A subgroup of G of the form

$$T = Z_G(g)$$

for a regular semisimple g will be called an *eigenvalue torus*. By the preceding proposition every eigenvalue torus is a conjugate of T_0 .

A regular semisimple element determines its eigenvalue torus uniquely, because its full centralizer is exactly that torus.

4. The singular locus inside a torus

PROPOSITION 5.9 (Nonregular elements in an eigenvalue torus). *Let T be an eigenvalue torus. Then*

$$T \setminus G_{\text{rss}}$$

is contained in a union of $\binom{n}{2}$ proper subtori-like hypersurfaces and has dimension at most

$$r - 1 = n - 2.$$

Its complexity is $O_n(1)$.

PROOF. Conjugate to T_0 . A diagonal element is not regular semisimple exactly when two diagonal entries coincide. Thus the nonregular locus is the union over $i < j$ of

$$\{\text{diag}(t_1, \dots, t_n) \in T_0 : t_i = t_j\}.$$

In the $(\mathbb{G}_m)^{n-1}$ coordinates of Proposition 5.3, each equality is one nonzero Laurent-polynomial equation. It defines a proper closed subset of dimension at most $r - 1$. There are only $\binom{n}{2}$ such equations and their degrees are bounded in terms of n . $\square$

5. The normalizer of an eigenvalue torus

PROPOSITION 5.10 (Normalizer quotient). *Let T be an eigenvalue torus in $\text{SL}_n(k)$. Then*

$$N_G(T)/T \cong S_n.$$

In particular

$$[N_G(T) : T] = n!, \quad \dim N_G(T) = r < D.$$

PROOF. Conjugate to T_0 . The simultaneous eigenspaces of T_0 on k^n are exactly the coordinate lines

$$ke_1, \dots, ke_n.$$

If g normalizes T_0 , then it carries the collection of common eigenspaces of T_0 to itself. Thus g permutes the coordinate lines, so g is a monomial matrix. Conversely every determinant-one monomial matrix normalizes T_0 .

Sending a monomial matrix to its permutation of the coordinate lines gives a surjective homomorphism

$$N_G(T_0) \rightarrow S_n.$$

Its kernel consists exactly of the diagonal determinant-one matrices, namely T_0 . The first isomorphism theorem gives the quotient. Since a finite extension has the same dimension as its identity component, $\dim N_G(T) = \dim T = r$, and $r < n^2 - 1$. $\square$

6. Positive-dimensional subtori have proper normalizers

Although the final finite-orbit argument below needs only maximal tori, it is useful for comparison with more general descent arguments to record the elementary fact that every positive-dimensional subtorus also has a proper normalizer.

PROPOSITION 5.11 (Normalizer of a positive-dimensional torus is proper). *Let $S \subseteq \mathrm{SL}_n(k)$ be a positive-dimensional algebraic torus. Then*

$$N_G(S) \subsetneq G.$$

PROOF. By [Proposition 5.4](#), conjugate S into the diagonal matrices. If every pair of coordinate characters had the same restriction to S , then every element of S would be scalar. But the scalar subgroup of $\mathrm{SL}_n(k)$ is finite, whereas S is positive-dimensional. Hence there exist $i \neq j$ and some

$$s = \mathrm{diag}(s_1, \dots, s_n) \in S$$

with $s_i \neq s_j$.

Let

$$u = I + E_{ij} \in \mathrm{SL}_n(k).$$

A direct calculation gives

$$u^{-1}su = s + (s_i - s_j)E_{ij},$$

which is not diagonal. It therefore does not lie in S . Hence $u \notin N_G(S)$, proving that the normalizer is proper. $\square$

7. The SL_3 torus as a laboratory

For $n = 3$ the standard torus consists of

$$\mathrm{diag}(x, y, (xy)^{-1}), \quad x, y \in k^\times,$$

and has dimension 2. The nonregular locus is the union of the three walls

$$x = y, \quad x = (xy)^{-1}, \quad y = (xy)^{-1}.$$

Each wall is one-dimensional. A regular point therefore lies off a finite union of codimension-one subtori-like sets and has centralizer exactly T_0 .

The normalizer consists of monomial determinant-one matrices. Modulo the diagonal torus, only the permutation of the three eigenlines remains, so

$$N_G(T_0)/T_0 \cong S_3.$$

This example displays every structural ingredient used later: a low-dimensional centralizer, a still lower-dimensional singular locus, and a normalizer that is only a finite extension of the torus. The formulas for general n are the same geometry with n eigenlines instead of three.

CHECKPOINT 5.12. For $G = SL_n$ we now have a completely explicit structural package:

$$\boxed{D = n^2 - 1}, \quad \boxed{r = n - 1},$$

regular semisimple elements form a dense open set, each one belongs to a unique eigenvalue torus of dimension r , the nonregular locus inside that torus has dimension at most $r - 1$, and the torus normalizer is a finite extension of order $n!$. These are exactly the group-theoretic facts the involved-torus argument needs.

CHAPTER 6

Involved Tori and the Crucial Counting Lemma

A regular semisimple element remembers a unique eigenvalue torus. If an approximate group is not polynomially small, Larsen–Pink forces almost all of its bounded power to be regular semisimple, forces every torus that occurs to contain the predicted number $|A|^{r/D}$ of points, and forces about $|A|^{1-r/D}$ different tori to occur. The decisive extra fact is conjugation invariance: if a torus occurs, then every conjugate by an element of A occurs already at the same short scale. We prove this by a representation-counting argument rather than quoting the BGT crucial lemma.

Continue with

$$G = \mathrm{SL}_n(k), \quad D = n^2 - 1, \quad r = n - 1.$$

Let $A \subseteq G(k)$ be a finite K -approximate group generating a Zariski-dense subgroup.

1. Definition of involvement

DEFINITION 6.1 (Involved torus). An eigenvalue torus T is *involved with* A if

$$A^2 \cap T$$

contains a regular semisimple element. Denote the finite set of involved tori by

$$\mathcal{T}(A).$$

Finiteness is immediate: every regular semisimple element belongs to a unique eigenvalue torus, and A^2 is finite.

2. Regular semisimple elements dominate unless A is small

LEMMA 6.2 (Most of A^2 is regular). *There is $C = C_n$ such that either*

$$(6.12) \quad |A| \leq K^C,$$

or

$$(6.13) \quad |A^2 \cap G_{\mathrm{rss}}| \geq \frac{1}{2}|A|.$$

PROOF. The nonregular locus

$$\Sigma = G \setminus G_{\text{rss}}$$

is a proper subvariety of dimension at most $D - 1$ and complexity $O_n(1)$ by Proposition 5.6. Larsen–Pink gives

$$|A^2 \cap \Sigma| \leq C_1 K^{C_1} |A|^{(D-1)/D}.$$

Since $A \subseteq A^2$, we have $|A^2| \geq |A|$. If

$$C_1 K^{C_1} |A|^{-1/D} \leq \frac{1}{2},$$

then at most half of $|A|$ elements of A^2 are nonregular and (6.13) follows. If this inequality fails, raising to the D th power yields

$$|A| \leq (2C_1)^D K^{C_1 D},$$

which has the form (6.12) after enlarging C . $\square$

3. How many points lie in one involved torus?

LEMMA 6.3 (Torus intersection at the Larsen–Pink scale). *There is $C = C_n$ such that either $|A| \leq K^C$, or every involved torus T satisfies*

$$(6.14) \quad K^{-C} |A|^{r/D} \leq |A^2 \cap T \cap G_{\text{rss}}| \leq K^C |A|^{r/D}.$$

PROOF. Let T be involved and choose

$$b \in A^2 \cap T \cap G_{\text{rss}}.$$

By Proposition 5.7,

$$Z_G(b) = T, \quad \dim T = r.$$

The lower centralizer estimate Proposition 4.3 gives

$$(6.15) \quad |A^2 \cap T| \geq K^{-C_1} |A|^{r/D}.$$

The upper Larsen–Pink estimate gives

$$(6.16) \quad |A^2 \cap T| \leq K^{C_1} |A|^{r/D}.$$

Inside T , the nonregular locus has dimension at most $r - 1$ and complexity $O_n(1)$ by Proposition 5.9. Applying Larsen–Pink to this subvariety gives

$$(6.17) \quad |A^2 \cap (T \setminus G_{\text{rss}})| \leq K^{C_2} |A|^{(r-1)/D}.$$

If

$$K^{C_2} |A|^{(r-1)/D} \leq \frac{1}{2} K^{-C_1} |A|^{r/D},$$

then subtracting (6.17) from (6.15) gives the lower bound in (6.14); the upper bound follows from (6.16). If the displayed inequality fails, then

$$|A|^{1/D} \leq 2K^{C_1+C_2},$$

so $|A| \leq K^C$ after enlarging C . $\square$

4. Counting the involved tori

PROPOSITION 6.4 (Number of involved tori). *There is $C = C_n$ such that either $|A| \leq K^C$, or*

$$(6.18) \quad K^{-C} |A|^{1-r/D} \leq |\mathcal{T}(A)| \leq K^C |A|^{1-r/D}.$$

PROOF. Assume the small alternative does not hold. By Lemma 6.2, the set A^2 contains at least $|A|/2$ regular semisimple elements. Every such element belongs to exactly one eigenvalue torus, so these elements are partitioned by the regular parts

$$A^2 \cap T \cap G_{\text{rss}}, \quad T \in \mathcal{T}(A).$$

By Lemma 6.3, each part has cardinality at most

$$K^C |A|^{r/D}.$$

Thus

$$|\mathcal{T}(A)| \geq K^{-C} |A|^{1-r/D}.$$

For the reverse inequality, the same lemma gives at least

$$K^{-C} |A|^{r/D}$$

regular elements in every involved torus, and these regular parts are disjoint. Since

$$|A^2| \leq K|A|,$$

we obtain

$$|\mathcal{T}(A)| K^{-C} |A|^{r/D} \leq K|A|,$$

which rearranges to the upper bound in (6.18). $\square$

5. The promotion argument under conjugation

The central difficulty is that if T is involved and $a \in A$, then conjugating a witness

$$b \in A^2 \cap T$$

only gives

$$a^{-1}ba \in A^4 \cap a^{-1}Ta,$$

not automatically a witness in A^2 . Approximate-group multiplicity promotes the large A^4 -intersection back to A^2 .

LEMMA 6.5 (Promotion from A^4 to A^2). *Let T be an eigenvalue torus. Suppose*

$$|A^4 \cap T| \geq L.$$

Then

$$(6.19) \quad |A^2 \cap T| \geq K^{-4}L.$$

PROOF. Put

$$B = A^4 \cap T.$$

Consider the multiplication map

$$A \times B \rightarrow A^5, \quad (x, t) \mapsto xt.$$

The domain has $|A||B|$ pairs, while

$$|A^5| \leq K^4|A|$$

by power growth. Hence some $g \in A^5$ has at least

$$\frac{|A||B|}{|A^5|} \geq K^{-4}|B|$$

representations $g = xt$ with $x \in A$ and $t \in B$.

Fix one representation $g = x_0 t_0$. For every other representation $g = xt$,

$$x_0^{-1}x = t_0 t^{-1}.$$

The left-hand side lies in A^2 and the right-hand side lies in T . Distinct representations give distinct x , hence distinct elements $x_0^{-1}x$. Therefore

$$|A^2 \cap T| \geq K^{-4}|B|,$$

which is (6.19). □

THEOREM 6.6 (Conjugation invariance of involved tori). *There is $C = C_n$ such that either $|A| \leq K^C$, or*

$$(6.20) \quad T \in \mathcal{T}(A), \ a \in A \implies a^{-1}Ta \in \mathcal{T}(A).$$

Because A is symmetric, in the non-small case $\mathcal{T}(A)$ is invariant under the whole group $\langle A \rangle$.

PROOF. Assume the small alternative does not hold. Let $T \in \mathcal{T}(A)$ and $a \in A$, and put

$$\tilde{T} = a^{-1}Ta.$$

By Lemma 6.3,

$$|A^2 \cap T| \geq K^{-C_1}|A|^{r/D}.$$

Conjugating this set by a gives

$$|A^4 \cap \tilde{T}| \geq K^{-C_1}|A|^{r/D}.$$

By the promotion lemma,

$$(6.21) \quad |A^2 \cap \tilde{T}| \geq K^{-C_1-4}|A|^{r/D}.$$

The nonregular locus in $\tilde{T}$ has dimension at most $r - 1$, so Larsen–Pink gives

$$|A^2 \cap (\tilde{T} \setminus G_{\text{rss}})| \leq K^{C_2}|A|^{(r-1)/D}.$$

Unless $|A| \leq K^C$, the latter is strictly smaller than the right-hand side of (6.21). Thus $A^2 \cap \tilde{T}$ contains a regular semisimple element, and $\tilde{T}$ is involved.

Since $A = A^{-1}$, conjugation by every generator permutes the finite set $\mathcal{T}(A)$ bijectively. Hence the whole generated group preserves it. $\square$

6. The involved-torus exponent ledger

It is useful to see the counting argument numerically before the final contradiction. Write

$$\theta = \frac{r}{D} = \frac{n-1}{n^2-1} = \frac{1}{n+1}.$$

Outside the polynomially small alternative, every involved torus carries

$$|A|^\theta$$

regular elements up to $K^{O_n(1)}$, while A^2 contains order $|A|$ regular elements altogether. Hence the number of involved tori is

$$|A|^{1-\theta} = |A|^{n/(n+1)}$$

up to the same polynomial losses.

For SL_3 this reads

$$|A^2 \cap T \cap G_{\mathrm{rss}}| \asymp_{K^{O(1)}} |A|^{1/4}, \quad |\mathcal{T}(A)| \asymp_{K^{O(1)}} |A|^{3/4}.$$

The promotion lemma is exactly what makes these exponents stable under conjugation by A : a large intersection appearing first in A^4 cannot hide there, because multiplicative energy between A and that torus slice pushes a polynomial fraction back into A^2 . After singular elements are removed by Larsen–Pink, a regular witness survives and the conjugate torus becomes involved at the original short scale.

CHECKPOINT 6.7. Outside the polynomially small regime, the finite set $\mathcal{T}(A)$ behaves as if it were an algebraic shadow of the approximate group:

$$|\mathcal{T}(A)| \asymp_{K^{O(1)}} |A|^{1-r/D},$$

and it is invariant under conjugation by $\langle A \rangle$. The next chapter shows that an infinite Zariski-dense group cannot preserve such a finite collection of tori.

CHAPTER 7

Finite Torus Orbits and Small Zariski-Dense Approximate Groups

The involved-torus argument ends with a purely geometric contradiction. Conjugation invariance makes the finite set of involved eigenvalue tori into a finite orbit of the Zariski-dense group generated by the approximate subgroup. The stabilizer of one torus therefore has finite index. We prove that a finite-index subgroup of a Zariski-dense subgroup of an irreducible algebraic group is still Zariski dense; but this stabilizer lies in the proper normalizer of the torus. This closes the type-A inverse theorem without invoking any additional torus-structure or finite-group classification result.

Continue with

$$G = \mathrm{SL}_n(k), \quad A \subseteq G(k)$$

a finite K -approximate group whose generated subgroup

$$\Gamma = \langle A \rangle$$

is Zariski dense.

1. Finite-index subgroups of dense groups

The final contradiction uses a simple point that is often left implicit.

LEMMA 7.1 (Finite index preserves Zariski density in an irreducible group). *Let G be an irreducible algebraic group and let $\Gamma \leq G(k)$ be Zariski dense. If $\Lambda \leq \Gamma$ has finite index, then Λ is Zariski dense in G .*

PROOF. Choose left-coset representatives $\gamma_1, \dots, \gamma_m \in \Gamma$ so that

$$\Gamma = \bigcup_{i=1}^m \gamma_i \Lambda.$$

Taking Zariski closures and using that left translation is an algebraic automorphism gives

$$G = \overline{\Gamma} \subseteq \bigcup_{i=1}^m \gamma_i \overline{\Lambda} \subseteq G.$$

Hence G is a finite union of the closed subsets $\gamma_i \bar{\Lambda}$. An irreducible variety cannot be a finite union of proper closed subsets: for two subsets this is the definition of irreducibility, and the finite statement follows by induction. Therefore one translate equals G . Translating back gives

$$\bar{\Lambda} = G.$$

□

REMARK 7.2 (Why connectedness is not being smuggled in). For an algebraic group over an algebraically closed field, irreducibility is the hypothesis used in the proof, not an unproved assertion about abstract finite-index subgroups. In our application $G = \mathrm{SL}_n$ is irreducible by Proposition 5.1. Thus the argument is entirely geometric.

2. Finite conjugation orbits

LEMMA 7.3 (A finite torus orbit gives a finite-index normalizer). *Let T be an eigenvalue torus. If a subgroup $\Gamma \leq G(k)$ preserves a finite set $\mathcal{T}$ of eigenvalue tori under conjugation and $T \in \mathcal{T}$, then*

$$\Gamma_T := \{\gamma \in \Gamma : \gamma^{-1}T\gamma = T\} = \Gamma \cap N_G(T)$$

has finite index in Γ ; indeed

$$[\Gamma : \Gamma_T] = |\Gamma \cdot T| \leq |\mathcal{T}|.$$

PROOF. The group Γ acts on the finite set $\mathcal{T}$ by conjugation. The stabilizer of T is exactly $\Gamma \cap N_G(T)$. The elementary orbit–stabilizer bijection

$$\Gamma/\Gamma_T \longrightarrow \Gamma \cdot T, \quad \gamma\Gamma_T \longmapsto \gamma^{-1}T\gamma,$$

is well defined and bijective, so the displayed index equals the orbit size. □

3. The smallness theorem

THEOREM 7.4 (Zariski-dense approximate subgroups of SL_n are polynomially small). *For every $n \geq 2$ there is a constant C_n with the following property. Let k be algebraically closed and let*

$$A \subseteq \mathrm{SL}_n(k)$$

be a finite K -approximate group such that $\langle A \rangle$ is Zariski dense. Then

$$(6.22) \quad |A| \leq C_n K^{C_n}.$$

PROOF. Suppose that $|A|$ is larger than all polynomial thresholds occurring in Lemmas 6.2 and 6.3 and Theorem 6.6. Then A^2 contains a regular semisimple element, so the finite set

$$\mathcal{T}(A)$$

of involved eigenvalue tori is nonempty. By [Theorem 6.6](#), it is invariant under conjugation by every element of

$$\Gamma := \langle A \rangle.$$

Choose $T \in \mathcal{T}(A)$. By [Lemma 7.3](#),

$$\Gamma_T = \Gamma \cap N_G(T)$$

has finite index in Γ . Since Γ is Zariski dense and $G = \mathrm{SL}_n$ is irreducible, [Lemma 7.1](#) gives

$$(6.23) \quad \overline{\Gamma_T} = G.$$

On the other hand, $\Gamma_T \subseteq N_G(T)$, and the torus-normalizer calculation [Proposition 5.10](#) gives

$$\dim N_G(T) = n - 1 < n^2 - 1 = \dim G.$$

Thus $N_G(T)$ is a proper closed algebraic subgroup of G . Taking closures in $\Gamma_T \subseteq N_G(T)$ contradicts (6.23).

Therefore at least one of the polynomial smallness alternatives used above must hold. There are only finitely many such alternatives, each of the form $|A| \leq CK^c$ with constants depending only on n . Taking the maximum of the finitely many constants and exponents gives (6.22). $\square$

4. The contradiction as a reusable pattern

The argument is shorter than the machinery that prepares it, and that is the point. The hard work is to construct a finite algebraic object whose conjugation orbit is controlled. Once that object exists, the contradiction is formal:

$$\begin{array}{c}
 \text{many regular elements} \\
 \Downarrow \\
 \text{at least one involved maximal torus} \\
 \Downarrow \\
 \text{Larsen–Pink + multiplicity} \\
 \Downarrow \\
 \text{finite set of involved tori invariant under } \Gamma \\
 \Downarrow \\
 [\Gamma : \Gamma \cap N_G(T)] < \infty \\
 \Downarrow \\
 \overline{\Gamma \cap N_G(T)} = G,
 \end{array}$$

while simultaneously

$$\overline{\Gamma \cap N_G(T)} \subseteq N_G(T) \subsetneq G.$$

The two conclusions are incompatible.

5. Why finite orbit is already fatal

There is a useful general pattern here. Let an irreducible algebraic group G contain a Zariski-dense subgroup Γ , and suppose Γ acts on a finite collection of algebraic objects. The stabilizer of any one object has finite index in Γ , hence is Zariski dense by [Lemma 7.1](#). Therefore that object cannot have a proper closed algebraic stabilizer in G .

For an eigenvalue torus the algebraic stabilizer is exactly its normalizer, whose dimension is $n - 1 < n^2 - 1$. Thus once the approximate-group argument has manufactured a nonempty finite conjugation-invariant family of such tori, no further quantitative estimate is needed. The contradiction is qualitative and immediate. This separation of labor is conceptually important: all quantitative work occurs in producing and stabilizing the finite family; the last step uses only Zariski density and irreducibility.

WARNING 7.5 (What is special about using maximal eigenvalue tori). The direct finite-index argument works because an involved regular semisimple element has a *maximal* eigenvalue torus whose normalizer is already known to be proper. If one worked with a more general family of connected abelian subgroups, an additional descent argument could be required. We do not import that unnecessary generality into the present proof.

CHECKPOINT 7.6. The structural inverse theorem needed for type-A product growth is now proved with no torus-classification theorem beyond the explicit normalizer calculation in [Proposition 5.10](#). The final chapter converts this inverse statement into direct growth for arbitrary finite sets using only the small-tripling machinery already proved in Volume II.

CHAPTER 8

From Approximate-Group Rigidity to Product Growth

The preceding chapters proved an inverse theorem: a finite approximate subgroup of $\mathrm{SL}_n(k)$ whose generated subgroup is Zariski dense must be polynomially small in its approximation parameter. This chapter converts that inverse statement into a direct growth theorem for arbitrary finite subsets. The conversion uses only the small-tripling theorem proved in Volume II. We keep every exponent visible, explain why no symmetry assumption is needed, and isolate carefully what this argument does—and does not—say over finite fields.

1. The inverse-to-direct conversion

The final passage is short, but it is logically important. A product theorem is a statement about an arbitrary finite set A . The structural theorem of [Theorem 7.4](#), by contrast, is a statement about an approximate group. Volume II proved the exact bridge required between the two.

We record the bridge in the form in which it will be used. The constants in the notation $O(1)$ below are absolute; only their existence matters for the exponent in the final theorem.

LEMMA 8.1 (Small tripling supplies a dense approximate group). *There is an absolute constant $C_0 \geq 1$ with the following property. Let Γ be any group and let $A \subseteq \Gamma$ be finite and nonempty. Put*

$$K = \frac{|A^3|}{|A|}.$$

Then there is a finite K^{C_0} -approximate group H such that

$$(6.24) \quad A \subseteq H, \quad |H| \leq K^{C_0} |A|.$$

In particular,

$$\langle A \rangle \leq \langle H \rangle.$$

If $\langle A \rangle$ is Zariski dense in an ambient algebraic group, then so is $\langle H \rangle$.

PROOF. This is precisely the “small tripling $\Rightarrow$ approximate group” theorem proved in Volume II, but we recall its construction to make clear that no symmetry assumption is being inserted here.

Set

$$H_0 = A \cup A^{-1} \cup \{e\}, \quad H = H_0^3.$$

The signed-word theorem of Volume II, proved there from the noncommutative Ruzsa triangle inequality and the covering lemma, gives

$$|H_0^m| \leq K^{O_m(1)} |A|$$

for every fixed m . In particular

$$|H| \leq K^{C_1} |A|, \quad |H_0 H^2| \leq K^{C_2} |A|$$

for absolute constants C_1, C_2 .

Applying the right-handed Ruzsa covering lemma to H_0 and H^2 produces a set $Y \subseteq H^2$ of cardinality at most K^{C_3} such that

$$H^2 \subseteq H_0^{-1} H_0 Y \subseteq H Y.$$

Taking inverses gives $H^2 \subseteq Y^{-1} H$. With

$$X = Y \cup Y^{-1} \cup \{e\}$$

we have $X = X^{-1}$, $X \subseteq H^2$, $|X| \leq K^{C_4}$ and

$$H^2 \subseteq XH \cap HX.$$

Thus H is a K^{C_4} -approximate group. Enlarging C_0 to dominate C_1 and C_4 yields (6.24).

Because $A \subseteq H$, one has $\langle A \rangle \leq \langle H \rangle$. The final assertion follows immediately: the Zariski closure of $\langle H \rangle$ contains the Zariski closure of $\langle A \rangle$. $\square$

WARNING 8.2 (Why we did not symmetrize “without loss of generality”). One cannot simply replace A by $A \cup A^{-1} \cup \{e\}$ and assume that the original tripling ratio $|A^3|/|A|$ is unchanged. The signed-word theorem in Volume II is exactly what justifies this replacement quantitatively. This point is harmless once proved, but it is not a formal consequence of symmetry.

2. The type-A product theorem

THEOREM 8.3 (Product growth in Zariski-dense subsets of SL_n). *For every integer $n \geq 2$ there exist constants*

$$c_n > 0, \quad \delta_n > 0$$

with the following property. Let k be algebraically closed and let

$$\emptyset \neq A \subseteq \mathrm{SL}_n(k)$$

be finite. If the subgroup $\langle A \rangle$ is Zariski dense in SL_n , then

$$(6.25) \quad |A^3| \geq c_n |A|^{1+\delta_n}.$$

The constants depend only on n .

PROOF. Set

$$K = \frac{|A^3|}{|A|} \geq 1.$$

Apply Lemma 8.1. We obtain a K^{C_0} -approximate group H with

$$A \subseteq H, \quad |H| \leq K^{C_0}|A|,$$

and $\langle H \rangle$ is Zariski dense in SL_n .

Apply the approximate-group rigidity theorem Theorem 7.4 to H . There is a constant $C_n \geq 1$ such that

$$|H| \leq C_n (K^{C_0})^{C_n} = C_n K^{C_0 C_n}.$$

Since $A \subseteq H$,

$$|A| \leq C_n K^{C_0 C_n}.$$

Therefore

$$K \geq C_n^{-1/(C_0 C_n)} |A|^{1/(C_0 C_n)}.$$

Put

$$\delta_n = \frac{1}{C_0 C_n}, \quad c_n = C_n^{-\delta_n}.$$

Multiplying the preceding inequality by $|A|$ gives

$$|A^3| = K|A| \geq c_n |A|^{1+\delta_n},$$

which is (6.25). □

GUIDING IDEA 8.4 (The whole proof in one line). The logical content is

$$\begin{aligned} |A^3| &\leq K|A| \\ \Downarrow \\ A &\subseteq H, \quad H \text{ a } K^{O(1)}\text{-approximate group} \\ \Downarrow \\ |A| &\leq |H| \leq K^{O_n(1)}. \end{aligned}$$

Rearranging the last inequality forces K to be a positive power of $|A|$. All difficult mathematics lies in the two boxed implications: Volume II proves the first, while the present volume proves the second through Larsen–Pink and involved tori.

3. Extracting the growth exponent

The last rearrangement is elementary but worth isolating. Suppose the inverse theorem gives

$$|H| \leq C_n L^{c_n}$$

for every Zariski-dense L -approximate subgroup H , while the small-tripling bridge converts

$$|A^3| = K|A|$$

into an $L = K^{C_0}$ approximate group containing A . Then

$$|A| \leq C_n K^{C_0 c_n},$$

so

$$K \geq C_n^{-1/(C_0 c_n)} |A|^{1/(C_0 c_n)}.$$

Thus the direct growth exponent is simply the reciprocal of the total polynomial loss accumulated by the inverse theorem and the tripling bridge. The precise numerical value is not important for the qualitative product theorem, but this calculation records where it comes from and prevents the phrase “some $\delta > 0$ ” from concealing an additional argument.

4. Symmetric generating sets and iterated growth

For applications one often starts with a symmetric set containing the identity. In that setting the theorem iterates cleanly until Zariski density is no longer the right notion of the ambient obstruction.

COROLLARY 8.5 (Growth of symmetric generating sets). *Let n, k, c_n, δ_n be as in Theorem 8.3. If*

$$A = A^{-1}, \quad e \in A,$$

and $\langle A \rangle$ is Zariski dense, then

$$|A^3| \geq c_n |A|^{1+\delta_n}.$$

Moreover every bounded power A^m generates the same subgroup, and hence whenever A^m is finite the same product theorem applies to A^m :

$$(6.26) \quad |A^{3m}| \geq c_n |A^m|^{1+\delta_n}.$$

PROOF. The first assertion is Theorem 8.3. Since $e \in A$, we have $A \subseteq A^m$ for every $m \geq 1$, while $A^m \subseteq \langle A \rangle$. Thus

$$\langle A^m \rangle = \langle A \rangle.$$

Apply Theorem 8.3 to the finite set A^m to obtain (6.26). □

5. Why finite fields need one more argument

The algebraic theorem above assumes that the abstract subgroup generated by A is Zariski dense. A finite subgroup of $\mathrm{SL}_n(\mathbb{F}_q)$ is never Zariski dense in $\mathrm{SL}_n(\overline{\mathbb{F}_q})$, so one cannot apply [Theorem 8.3](#) by scalar extension. What survives is the actual property used in the Larsen–Pink induction: bounded powers must escape every proper algebraic set of bounded complexity. Over a sufficiently large finite field this follows from the point-count theorem proved in Volume V. We now carry out that replacement completely.

PROPOSITION 8.6 (Order of $\mathrm{SL}_n(\mathbb{F}_q)$). *Let $D = n^2 - 1$. Then*

$$(6.27) \quad |\mathrm{SL}_n(\mathbb{F}_q)| = q^D \prod_{j=2}^n (1 - q^{-j}).$$

Consequently, for fixed n there is $c_n > 0$ such that

$$c_n q^D \leq |\mathrm{SL}_n(\mathbb{F}_q)| \leq q^D$$

for every prime power q .

PROOF. An invertible matrix is an ordered basis of $\mathbb{F}_q^n$. Its first column has $q^n - 1$ choices, the second has $q^n - q$ choices, and in general the j th column has $q^n - q^{j-1}$ choices. Hence

$$|\mathrm{GL}_n(\mathbb{F}_q)| = \prod_{j=0}^{n-1} (q^n - q^j) = q^{n^2} \prod_{j=1}^n (1 - q^{-j}).$$

The determinant map onto $\mathbb{F}_q^\times$ is surjective, for example by using diagonal matrices, and every fiber is a coset of SL_n . Dividing by $q - 1 = q(1 - q^{-1})$ gives (6.27). The upper bound is immediate, while

$$\prod_{j=2}^n (1 - q^{-j}) \geq \prod_{j=2}^n (1 - 2^{-j}) =: c_n > 0.$$

□

PROPOSITION 8.7 (Uniform algebraic escape over $\mathbb{F}_q$). *Fix n and a complexity bound M . There are constants $q_0 = q_0(n, M)$ and $r = r(n, M)$ such that the following holds. Put*

$$G = \mathrm{SL}_n(\overline{\mathbb{F}_q}).$$

If $q \geq q_0$ and a finite symmetric set $B \subseteq \mathrm{SL}_n(\mathbb{F}_q)$ contains the identity and generates $\mathrm{SL}_n(\mathbb{F}_q)$, then for every proper closed subvariety $V \subsetneq G$ of complexity at most M one has

$$(6.28) \quad B^r \not\subseteq V.$$

The same r works simultaneously for all such V .

PROOF. A proper closed subvariety of the irreducible D -dimensional group G has dimension at most $D - 1$. Volume V, Appendix A, proved from degree and repeated coordinate slicing that every bounded-complexity variety of dimension at most $D - 1$ contains at most

$$C_{n,M}q^{D-1}$$

points of the ambient $\mathbb{F}_q$ -space. On the other hand [Proposition 8.6](#) gives

$$|\mathrm{SL}_n(\mathbb{F}_q)| \geq c_n q^D.$$

For $q \geq q_0(n, M)$ the latter is larger, so the orbit $\mathrm{SL}_n(\mathbb{F}_q)$ of the identity is not contained in V .

Apply the finite-list escape theorem proved in Volume V, Chapter 8, to the left-translation action and this orbit. Its proof bounds the word depth only through the ambient dimension and the finite-dimensional space of equations of degree at most M ; in particular the depth is uniform in the coefficients of V . This gives one $r = r(n, M)$ for which a word in B^r escapes every proper V of complexity at most M , proving (6.28). $\square$

THEOREM 8.8 (Finite-field Larsen–Pink inequality for generating approximate groups). *Fix n , a complexity bound M , and $m \geq 1$. There are $q_0 = q_0(n, M, m)$ and $C = C(n, M, m)$ such that if $q \geq q_0$ and*

$$A \subseteq \mathrm{SL}_n(\mathbb{F}_q)$$

is a finite K -approximate group generating $\mathrm{SL}_n(\mathbb{F}_q)$, then every closed $V \subseteq \mathrm{SL}_n(\overline{\mathbb{F}_q})$ of complexity at most M satisfies

$$(6.29) \quad |A^m \cap V| \leq CK^C |A|^{\dim V/D}, \quad D = n^2 - 1.$$

PROOF. Set $B = A^m$. Since $e \in A$, the set B is symmetric, contains the identity, generates the same finite group, and is a $K^{O_m(1)}$ -approximate group.

Apply the escape-form Larsen–Pink theorem [Theorem 3.1](#) to the initial complexity bound M . Let

$$M_* = M_*(n, M)$$

be the single algebraic-complexity threshold appearing there. Choose q_0 so that [Proposition 8.7](#) holds at complexity M_* . That proposition gives an integer $r_0 = r(n, M_*)$ for which B has the (M_*, r_0) bounded-complexity escape property: indeed $A^{r_0} \subseteq B^{r_0}$ because $A \subseteq B$.

Therefore [Theorem 3.1](#), now invoked literally rather than by analogy, gives

$$|A^m \cap V| = |B \cap V| \leq C|B^Q|^{\dim V/D} = C|A^{mQ}|^{\dim V/D}$$

for a bounded $Q = Q(n, M, r_0)$. Finally

$$|A^{mQ}| \leq K^{mQ-1}|A|$$

by approximate-group power growth, and substitution gives (6.29) after enlarging the exponent of K .

Thus the algebraically closed and finite-field versions use the same geometric theorem. They differ only in how the bounded-complexity escape hypothesis is supplied: Zariski density in [Theorem 3.2](#), and point counting plus finite generation here. $\square$

6. Small or dense over a finite field

We now rerun the involved-torus argument with [Theorem 8.8](#). Every regular semisimple element $b \in \mathrm{SL}_n(\mathbb{F}_q)$ has a centralizer torus defined over $\mathbb{F}_q$, because the equations $xb = bx$ have coefficients in $\mathbb{F}_q$.

THEOREM 8.9 (Finite-field approximate groups are small or dense). *For every $n \geq 2$ there are constants $q_0(n)$ and C_n such that, whenever $q \geq q_0(n)$ and a finite K -approximate group*

$$A \subseteq G_q := \mathrm{SL}_n(\mathbb{F}_q)$$

generates G_q , one has either

$$(6.30) \quad |A| \leq C_n K^{C_n}$$

or

$$(6.31) \quad |G_q| \leq C_n K^{C_n} |A|.$$

PROOF. Use the finite-field Larsen–Pink inequality in place of [Theorem 4.1](#). The proofs of [Lemmas 6.2](#) and [6.3](#) and [Proposition 6.4](#) then go through verbatim: outside the alternative $|A| \leq K^{O_n(1)}$, A^2 contains at least $|A|/2$ regular semisimple elements, every involved eigenvalue torus T contains

$$K^{-O_n(1)} |A|^{r/D} \leq |A^2 \cap T \cap G_{\mathrm{rss}}| \leq K^{O_n(1)} |A|^{r/D}, \quad r = n - 1,$$

and hence

$$(6.32) \quad |\mathcal{T}(A)| \leq K^{O_n(1)} |A|^{1-r/D}.$$

The promotion lemma [Lemma 6.5](#) is purely group-theoretic, so the proof of [Theorem 6.6](#) also carries over and shows that $\mathcal{T}(A)$ is invariant under the generated group G_q .

Choose $T \in \mathcal{T}(A)$. It is defined over $\mathbb{F}_q$, and the stabilizer of T under conjugation by G_q is

$$N_G(T)(\mathbb{F}_q).$$

Over the algebraic closure [Proposition 5.10](#) gives $\dim N_G(T) = r$ and bounded complexity depending only on n . The finite-field point bound of Volume V therefore gives

$$|N_G(T)(\mathbb{F}_q)| \leq C_n q^r.$$

By orbit–stabilizer and [Proposition 8.6](#),

$$(6.33) \quad |G_q \cdot T| = \frac{|G_q|}{|N_G(T)(\mathbb{F}_q)|} \geq c_n q^{D-r}.$$

This whole orbit lies in $\mathcal{T}(A)$. Combining (6.32) and (6.33) yields

$$q^{D-r} \leq C_n K^{C_n} |A|^{(D-r)/D}.$$

Raise both sides to the power $D/(D-r)$:

$$q^D \leq C_n K^{C_n} |A|.$$

Since $|G_q| \leq q^D$, this is (6.31). The only alternative encountered in deriving these estimates was polynomial smallness, which is (6.30) after enlarging C_n . $\square$

7. An elementary quasirandomness bound for $\mathrm{SL}_n(\mathbb{F}_q)$

The dense alternative is completed by the product-mixing theorem of Volume I. We prove the required representation-degree estimate directly so that no Landazuri–Seitz theorem enters the arc.

PROPOSITION 8.10 (Elementary quasirandomness in type A). *Let q be a prime power and let ρ be a nontrivial irreducible complex representation of $\mathrm{SL}_n(\mathbb{F}_q)$. Then*

$$(6.34) \quad \dim \rho \geq \begin{cases} (q-1)/2, & n = 2, \\ q^{n-1} - 1, & n \geq 3. \end{cases}$$

In particular, for fixed $n \geq 2$ every nontrivial irreducible representation has dimension at least $c_n q^{n-1}$.

PROOF. Let

$$U = \left\{ I + \sum_{j=2}^n v_j E_{1j} : v_j \in \mathbb{F}_q \right\} \cong (\mathbb{F}_q^{n-1}, +).$$

Volume I proved that the restriction of a complex representation to a finite abelian group decomposes into character spaces. If only the trivial U -character occurred, then U would act trivially. The normal closure of U contains every elementary root subgroup (conjugate E_{12} by the usual determinant-one Weyl matrices), and Appendix A proved that elementary matrices generate SL_n . Thus ρ would be trivial, contrary to hypothesis. Hence some nontrivial U -character occurs.

Fix one nontrivial additive character ψ of $\mathbb{F}_q$. For a column $\xi \in \mathbb{F}_q^{n-1}$ define

$$\chi_\xi(v) = \psi(v\xi).$$

If $\xi \neq \eta$, choose a coordinate on which they differ and choose the corresponding entry of v so that $\psi(v(\xi - \eta)) \neq 1$; hence the characters χ_ξ are distinct. There are q^{n-1} of them, equal to $|U|$, so they are all the characters of U .

Assume first $n \geq 3$ and put $m = n - 1 \geq 2$. For every $B \in \mathrm{SL}_m(\mathbb{F}_q)$, the block diagonal matrix $\mathrm{diag}(1, B)$ lies in SL_n and normalizes U , acting on row vectors by

$$v \mapsto vB^{-1}.$$

Consequently it permutes the U -weights by the dual linear action. The group $\mathrm{SL}_m(\mathbb{F}_q)$ is transitive on nonzero vectors: given nonzero x, y , start with any invertible map carrying x to y and multiply on the left by an element fixing y whose determinant corrects the determinant; such a correction exists because $m \geq 2$. Thus the orbit of any nontrivial U -character has all $q^m - 1 = q^{n-1} - 1$ nontrivial characters. Each occurring weight space is nonzero and distinct weight spaces are linearly independent, so $\dim \rho \geq q^{n-1} - 1$.

For $n = 2$, write

$$u(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \quad h(t) = \mathrm{diag}(t, t^{-1}).$$

Then $h(t)u(x)h(t)^{-1} = u(t^2x)$. Parameterizing the additive characters of $\mathbb{F}_q$ as $x \mapsto \psi(ax)$, the diagonal group therefore moves a nonzero parameter through its coset under the square subgroup of $\mathbb{F}_q^\times$. That orbit has

$$|(\mathbb{F}_q^\times)^2| = \frac{q-1}{\gcd(2, q-1)} \geq \frac{q-1}{2}$$

distinct characters, proving the rank-one bound. $\square$

COROLLARY 8.11 (Quasirandom completion scale). *For fixed n there is C_n such that, if $E \subseteq G_q = \mathrm{SL}_n(\mathbb{F}_q)$ satisfies*

$$(6.35) \quad |E| \geq C_n |G_q| q^{-(n-1)/3},$$

then

$$E^3 = G_q.$$

PROOF. Volume I proved the Gowers product-mixing lemma: in a finite group whose smallest nontrivial irreducible representation has dimension at least D_0 , three subsets X, Y, Z satisfy $XYZ = G$ once $|X||Y||Z| > |G|^3/D_0$. Apply it with $X = Y = Z = E$ and use [Proposition 8.10](#). Enlarging C_n absorbs the harmless constants and the finitely many very small q . $\square$

8. The finite-field type-A product theorem

THEOREM 8.12 (Product theorem for $\mathrm{SL}_n(\mathbb{F}_q)$). *For every $n \geq 2$ there are constants $c_n > 0$ and $\delta_n > 0$ such that for every prime power q and every finite set*

$$E \subseteq \mathrm{SL}_n(\mathbb{F}_q)$$

generating the whole group, one has

$$(6.36) \quad \boxed{E^3 = \mathrm{SL}_n(\mathbb{F}_q)} \quad \text{or} \quad \boxed{|E^3| \geq c_n |E|^{1+\delta_n}}.$$

PROOF. Write

$$N = |E|, \quad K = \frac{|E^3|}{|E|}.$$

Apply the small-tripling bridge [Lemma 8.1](#). It gives a K^{C_0} -approximate group H containing E , generating the same finite group, and satisfying

$$(6.37) \quad |H| \leq K^{C_0} N.$$

For $q \geq q_0(n)$ apply [Theorem 8.9](#) to H .

In the small alternative,

$$N \leq |H| \leq C_n(K^{C_0})^{C_n},$$

so $K \geq c_n N^{\eta_n}$ for some $\eta_n > 0$, and hence $|E^3| \geq c_n N^{1+\eta_n}$.

In the dense alternative, combining (6.37) with (6.31) gives

$$(6.38) \quad N \geq c_n K^{-b_n} |G_q|$$

for some $b_n > 0$. Choose $\delta_n > 0$ so small that

$$(6.39) \quad D b_n \delta_n < \frac{n-1}{6}, \quad D = n^2 - 1.$$

If $K \geq N^{\delta_n}$, we already have the growth alternative (after possibly decreasing c_n). Otherwise

$$K < N^{\delta_n} \leq |G_q|^{\delta_n} \leq q^{D\delta_n}.$$

Substituting in (6.38) and using $|G_q| \geq c_n q^D$ gives

$$N \geq c_n q^{D-Db_n\delta_n} \geq c_n q^{D-(n-1)/6}.$$

For all sufficiently large q this is stronger than the completion threshold (6.35), whose right-hand side is $O_n(q^{D-(n-1)/3})$. Hence $E^3 = G_q$.

Only finitely many prime powers $q < q'_0(n)$ remain. For fixed n these give finitely many finite groups and finitely many generating subsets. Decrease c_n if necessary so that every one of those subsets not satisfying $E^3 = G_q$ satisfies the growth alternative. This completes the proof for all q . $\square$

REMARK 8.13 (Rank one over arbitrary finite fields). Taking $n = 2$ in [Theorem 8.12](#) supplies the general $\mathrm{SL}_2(\mathbb{F}_q)$ product theorem. Volume IV deliberately proved the prime-field Helfgott mechanism in coordinates because that is where the sum-product, trace, pivot, and torus ideas are most transparent. The present argument explains how the algebraic Larsen–Pink framework removes the prime-field restriction without importing a finite-simple-group product theorem.

This is the type-*A* mechanism behind the Breuillard–Green–Tao product theorem [1]. The general theorem for all bounded-rank simple algebraic groups, and the independent Pyber–Szabó finite-simple-group theorem [2], require additional root-system and finite-form structure. They motivate the wider theory but are not proof dependencies of any theorem proved in this arc.

9. Logical closure of the Arc-II product-growth argument

We can now state exactly what has been proved in the second arc and what its proof depends on.

PROPOSITION 8.14 (Logical closure of the Arc-II product-growth argument). *The proof of Theorem 8.3 uses no non-basic result outside the following internally proved inputs:*

- (1) *the noncommutative Ruzsa calculus, signed-word control, and small-tripling-to-approximate-group theorem of Volume II;*
- (2) *the quantitative algebraic geometry of Volume V: dimension, constructible images, generic fibers, uniform finite-map bounds, algebraic-group orbit dimensions, escape from subvarieties, and product–conjugation dimension growth;*
- (3) *the raw Larsen–Pink induction proved in Chapters 1–3 of the present volume;*
- (4) *the type-A linear algebra of Chapters 5–7: regular semisimple elements, torus centralizers, torus normalizers, involved-torus counting, and the finite-orbit normalizer contradiction;*
- (5) *the almost-simplicity statement for SL_n , proved in Appendix A.*

No Larsen–Pink inequality, BGT approximate-subgroup theorem, Pyber–Szabó product theorem, or classification of finite simple groups is invoked as an external theorem.

PROOF. The proof is an inspection of the dependency chain. The only step in Theorem 8.3 before the present volume’s inverse theorem is Lemma 8.1, whose proof is the Volume-II Ruzsa/signed-word construction. The inverse theorem Theorem 7.4 depends on involved-torus invariance and the finite-orbit contradiction. The former is proved from the Larsen–Pink inequality, centralizer fibers, and approximate-group power bounds; the latter uses only the finite-index density lemma and the explicit properness of an eigenvalue-torus normalizer. The Larsen–Pink inequality itself is derived in Chapters 1–4 from the geometric tools of Volume V. Finally, the almost-simplicity hypothesis used in the geometric induction is proved in Appendix A. For the finite-field theorem, the only replacement is that the degree point bound of Volume V supplies bounded-complexity escape; the torus-orbit count, elementary quasirandomness estimate, and final exponent comparison are all proved above. This exhausts the chain. $\square$

CHECKPOINT 8.15. Arc II has now reached its intended endpoint. Volume IV proved the complete rank-one growth mechanism in SL_2 from the finite-field arithmetic developed in Arc I. Volume V built the algebraic geometry required to turn dimension into finite-set inequalities. Volume VI proved the Larsen–Pink inequality and used it, together with explicit type-A torus geometry, first to derive

product growth for finite subsets generating Zariski-dense subgroups of $\mathrm{SL}_n(k)$ and then to prove the small-or-dense and product theorems for generating subsets of $\mathrm{SL}_n(\mathbb{F}_q)$. The remaining arithmetic-expansion problems require genuinely new inputs and belong to the next arc rather than being concealed here.

APPENDIX A

Simplicity of $\mathrm{PSL}_n(k)$ over an Algebraically Closed Field

The Larsen–Pink geometry requires almost simplicity. Rather than cite the structure theory of semisimple algebraic groups, we prove directly that $\mathrm{PSL}_n(k)$ is an abstractly simple group when k is algebraically closed, hence infinite. The proof uses a primitive-action criterion of Iwasawa type, proved here from scratch, together with elementary matrices.

1. A primitive-action simplicity criterion

LEMMA A.1 (Iwasawa criterion). *Let a group G act faithfully and primitively on a set Ω . Fix $\omega \in \Omega$ and let $H = G_\omega$. Suppose:*

- (1) *H contains an abelian normal subgroup $U \triangleleft H$;*
- (2) *the conjugates of U generate G ;*
- (3) *G is perfect, i.e. $G = [G, G]$.*

Then G is simple.

PROOF. Let $N \triangleleft G$ be nontrivial. The N -orbits form a G -invariant partition of Ω : if x and y lie in one N -orbit, then gx and gy lie in one N -orbit because $gNg^{-1} = N$. Primitivity therefore implies that the N -orbits are either singletons or all of Ω . Faithfulness rules out singleton orbits for a nontrivial normal subgroup, so N is transitive.

Transitivity gives $G = NH$: for every $g \in G$, choose $n \in N$ with $n\omega = g\omega$, so $n^{-1}g \in H$. In the quotient G/N , every element is therefore represented by an element of H .

For any $g \in G$, write $g = nh$ with $n \in N$ and $h \in H$. Since $U \triangleleft H$,

$$U^g = U^{nh} = (U^h)^n = U^n.$$

Thus the image of every conjugate U^g in G/N equals the image of U . Because the conjugates of U generate G , the quotient G/N is generated by the image of the abelian group U and is therefore abelian. But G is perfect, so it has no nontrivial abelian quotient. Hence G/N is trivial and $N = G$. $\square$

2. Elementary matrices generate SL_n

For $i \neq j$ and $t \in k$, write

$$e_{ij}(t) = I + tE_{ij}.$$

LEMMA A.2 (Elementary generation). *For every field k and every $n \geq 2$, the group $\mathrm{SL}_n(k)$ is generated by the elementary matrices $e_{ij}(t)$.*

PROOF. Left multiplication by $e_{ij}(t)$ adds t times row j to row i . Starting from $g \in \mathrm{SL}_n(k)$, Gaussian elimination using such row additions reduces g to a diagonal matrix

$$\mathrm{diag}(a_1, \dots, a_n), \quad \prod_i a_i = 1.$$

It remains to express determinant-one diagonal matrices by elementary matrices. In the 2×2 block on coordinates i, j one has the identity

$$(A.1) \quad \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} = \begin{pmatrix} 1 & a-1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & a^{-1}-1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -a & 1 \end{pmatrix},$$

as is checked by multiplication. Successively using such 2×2 blocks writes every determinant-one diagonal matrix as a product of elementary matrices. Reversing the elimination proves the lemma. $\square$

3. Perfectness

LEMMA A.3 (Perfectness for $n \geq 3$). *If $n \geq 3$, then $\mathrm{SL}_n(k)$ is perfect.*

PROOF. Choose distinct indices i, j, ℓ . Direct multiplication gives the Steinberg commutator relation

$$[e_{i\ell}(t), e_{\ell j}(1)] = e_{ij}(t).$$

Thus every elementary generator is a commutator. By Lemma A.2, the group is generated by commutators and is perfect. $\square$

LEMMA A.4 (Perfectness for $n = 2$ over an infinite field). *If k is infinite, then $\mathrm{SL}_2(k)$ is perfect.*

PROOF. Choose $a \in k^\times$ with $a^2 \neq 1$, possible because k is infinite, and set

$$h(a) = \mathrm{diag}(a, a^{-1}), \quad u(t) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

Then

$$h(a)u(s)h(a)^{-1} = u(a^2s),$$

so

$$[h(a), u(s)] = u((a^2 - 1)s).$$

Given t , take $s = t/(a^2 - 1)$; every upper unipotent $u(t)$ is a commutator. The same calculation with lower unipotents shows that every $e_{21}(t)$ is a commutator. These elementary matrices generate $\mathrm{SL}_2(k)$ by Lemma A.2. $\square$

Hence $\mathrm{PSL}_n(k)$ is perfect as a quotient of a perfect group.

4. Primitive action on projective space

Because k is algebraically closed, every $n \times n$ invertible matrix can be multiplied by a scalar to have determinant 1. Therefore

$$(A.2) \quad \mathrm{PGL}_n(k) = \mathrm{PSL}_n(k).$$

Indeed, if $g \in \mathrm{GL}_n(k)$, choose $c \in k^\times$ with $c^n \det g = 1$; the projective classes of g and cg coincide.

LEMMA A.5 (Two-transitivity). *The group $\mathrm{PSL}_n(k)$ acts two-transitively on the projective space $\mathbb{P}^{n-1}(k)$.*

PROOF. Two distinct projective points are represented by two linearly independent vectors. Given ordered pairs of distinct points, extend representatives of each pair to bases of k^n and choose a linear isomorphism sending the first basis to the second. Its projective class sends the first ordered pair to the second. By (A.2), that projective class lies in $\mathrm{PSL}_n(k)$. $\square$

Two-transitivity implies primitivity.

Fix the point $\omega = [e_1]$. Inside its stabilizer define

$$U = \left\{ \begin{pmatrix} 1 & \lambda \\ 0 & I_{n-1} \end{pmatrix} : \lambda \in k^{1 \times (n-1)} \right\} \quad \text{modulo the center.}$$

The group U is abelian.

LEMMA A.6 (The projective unipotent radical). *The subgroup U is normal in the stabilizer of $[e_1]$, and its conjugates generate $\mathrm{PSL}_n(k)$.*

PROOF. An element of the line stabilizer has block form

$$g = \begin{pmatrix} a & \beta \\ 0 & B \end{pmatrix}.$$

A direct block multiplication shows

$$g \begin{pmatrix} 1 & \lambda \\ 0 & I \end{pmatrix} g^{-1} = \begin{pmatrix} 1 & a\lambda B^{-1} \\ 0 & I \end{pmatrix},$$

so U is normal in the stabilizer.

The group U contains all elementary matrices $e_{1j}(t)$ with $j > 1$. Conjugating by permutation matrices in projective space produces every elementary root subgroup $e_{ij}(t)$. By Lemma A.2, their images generate $\mathrm{PSL}_n(k)$. $\square$

THEOREM A.7 (Simplicity of $\mathrm{PSL}_n(k)$). *Let k be algebraically closed and $n \geq 2$. Then $\mathrm{PSL}_n(k)$ is simple.*

PROOF. The action on projective space is faithful by the definition of PSL_n , and it is primitive by Lemma A.5. The point stabilizer contains the abelian normal subgroup U of Lemma A.6, whose conjugates generate the group. Perfectness follows from Lemmas A.3 and A.4. Apply the Iwasawa criterion, Lemma A.1. $\square$

COROLLARY A.8 (Almost simplicity of SL_n). *The algebraic group SL_n has no proper positive-dimensional closed normal algebraic subgroup.*

PROOF. Let $N \triangleleft \mathrm{SL}_n$ be a positive-dimensional closed normal subgroup. Its set of k -points is infinite, whereas the center

$$Z(\mathrm{SL}_n)(k) = \{\zeta I : \zeta^n = 1\}$$

is finite. Thus the image of $N(k)$ in $\mathrm{PSL}_n(k)$ is a nontrivial normal subgroup. By Theorem A.7, that image is all of $\mathrm{PSL}_n(k)$.

Consequently

$$\mathrm{SL}_n(k) = N(k)Z(\mathrm{SL}_n)(k),$$

so the irreducible variety SL_n is the union of finitely many translates of the closed subvariety N . An irreducible variety cannot be a finite union of proper closed subsets. Hence one translate, and therefore N itself, must be all of SL_n . $\square$

Bibliography

1. Emmanuel Breuillard, Ben Green, and Terence Tao, *Approximate subgroups of linear groups*, Geometric and Functional Analysis **21** (2011), no. 4, 774–819.
2. László Pyber and Endre Szabó, *Growth in finite simple groups of lie type of bounded rank*, Journal of the American Mathematical Society **29** (2016), no. 1, 95–146.

The Bourgain–Gamburd Method
Volume VII: Arithmetic Reduction, Strong
Approximation, and Logarithmic Geometry

Preface

Arc I built the analytic, additive-combinatorial, and finite-field arithmetic machinery. Arc II proved product growth in $\mathrm{SL}_2(\mathbb{F}_p)$ and then proved the algebraic-geometric Larsen–Pink mechanism. Arc III now assembles those inputs into expansion. This volume closes the arithmetic and pre-collision part of that assembly.

The strongest proof-policy issue is strong approximation. In the older compressed manuscript, the general theorem of Matthews–Vaserstein–Weisfeiler was cited at precisely the point where the arithmetic quotient had to become $\mathrm{SL}_2(\mathbb{F}_p)$. That is no longer acceptable under the self-contained standard. Here we prove the rank-one statement needed by Bourgain–Gamburd directly from the proper-subgroup dichotomy established in Volume IV, a fixed group-law argument, Zariski density, and an explicit free subgroup of $\mathrm{SL}_2(\mathbb{R})$. The general strong-approximation theorem remains in the bibliography for historical scope, not as a logical dependency.

The second goal is quantitative geometry of words. A fixed non-elementary subgroup contains a ping–pong free pair after taking fixed powers. Height grows only exponentially with word length, hence reduction modulo p is injective on a ball of radius $c \log p$. The free walk therefore retains its tree-like exponential spreading for logarithmic time. Volume VIII will upgrade this pointwise spreading to uniform escape from every proper subgroup coset. **Series position.** This volume uses Volumes IV–VI. Its main outputs are rank-one strong approximation, bounded-word ping–pong, logarithmic injectivity, and initial spreading.

Quantifier and asymptotic conventions. Unless a statement explicitly says otherwise, every constant occurring in a quantitative theorem is chosen in the order in which it is quantified in that theorem. Thus a phrase such as “for all sufficiently small δ ” means: there is a displayed or locally quantified threshold $\delta_0 > 0$, depending only on the parameters listed before that phrase, such that the assertion holds for every $0 < \delta < \delta_0$. The phrases “sufficiently large N ” and “sufficiently large p ” have the analogous meaning.

If $X \ll_\theta Y$ or $X = O_\theta(Y)$, the implicit constant depends only on θ and on the fixed ambient data of the current volume; it is independent of the variables

quantified after θ . The words *absolute*, *uniform*, and *bounded* never introduce an unlisted dependence. Whenever an inequality between exponents or scales is essential to the argument, that inequality is written explicitly in the proof; these conventions are not a substitute for such a calculation.

Contents

Preface	iii
1 The Arithmetic Expansion Problem and Its Dependency Graph	1
1.1 The theorem at the end of the arc	1
1.2 What has already been proved	2
1.3 The eight interfaces	3
1.4 A quantitative ledger	4
1.5 Historical location	5
2 Ping-Pong, Free Subgroups, and Bounded Words	7
2.1 Hyperbolic dynamics on the projective line	7
2.2 The ping-pong lemma	8
2.3 Returning to the original generators	9
2.4 Why changing generators is harmless	9
2.5 Free balls and reduced words	10
2.6 Worked ping-pong calculation on the projective line	11
3 Non-Elementary Subgroups and Zariski Density in Rank One	13
3.1 Why Zariski density enters twice	13
3.2 Word maps are regular	16
3.3 No nontrivial law holds in SL_2	16
3.4 Why this rank-one argument is enough	17
4 Rank-One Strong Approximation from Subgroup Laws	19
4.1 The statement actually needed	19
4.2 A uniform law for every proper subgroup	19
4.3 Infinitely many proper reductions force an integral law	21
4.4 Completion of strong approximation	21
4.5 What has and has not been proved	22
4.6 A useful variant	23

5	Matrix Height and Logarithmic Injectivity Modulo p	25
5.1	A height adapted to multiplication	25
5.2	Congruence equality and actual equality	26
5.3	A girth formulation	27
5.4	Schreier injectivity	27
5.5	Initial L^2 consequences of injectivity	27
6	The Free-Group Walk and the First Spreading Estimate	29
6.1	The regular tree operator	29
6.2	Point probabilities	30
6.3	Transfer through the logarithmic window	31
6.4	Why point spreading is not subgroup escape	32
6.5	Persistence of the atom bound	32
7	Volume VII Dependency Map and Handoff to Subgroup Escape	33
7.1	The chain now proved	33
7.2	Proof dependencies by theorem	34
7.3	Results deliberately not used as black boxes	34
7.4	The next interface: subgroup-coset non-concentration	35
7.5	Historical provenance and scope	35

Chapter 1

The Arithmetic Expansion Problem and Its Dependency Graph

1.1 The theorem at the end of the arc

Let $S \subset \mathrm{SL}_2(\mathbb{Z})$ be finite, symmetric, and fixed. Write $\Gamma = \langle S \rangle$ and let S_p denote its reduction in $\mathrm{SL}_2(\mathbb{F}_p)$. We use the following operational rank-one definition.

Definition 1.1 (Non-elementary subgroup). A subgroup $\Gamma < \mathrm{SL}_2(\mathbb{R})$ is *non-elementary* if its image in $\mathrm{PSL}_2(\mathbb{R})$ contains two hyperbolic elements whose fixed-point sets on $\mathbb{P}^1(\mathbb{R})$ are disjoint.

This definition is tailored to the proof: it immediately supplies the geometric input for ping–pong. It is equivalent to the more usual formulations for finitely generated subgroups of $\mathrm{PSL}_2(\mathbb{R})$; none of those equivalences is needed as an external theorem below.

Theorem 1.2 (Arithmetic Bourgain–Gamburd theorem; target of Arc III). *Let $S \subset \mathrm{SL}_2(\mathbb{Z})$ be finite, symmetric, and suppose $\Gamma = \langle S \rangle$ is non-elementary. Then there are p_0 and $\varepsilon > 0$, depending only on S , such that for every prime $p > p_0$,*

$$\langle S_p \rangle = \mathrm{SL}_2(\mathbb{F}_p)$$

and the Markov operator

$$P_p f(x) = \frac{1}{|S|} \sum_{s \in S} f(xs)$$

on $\ell^2(\mathrm{SL}_2(\mathbb{F}_p))$ satisfies

$$\sup(\mathrm{Spec}(P_p) \setminus \{1\}) \leq 1 - \varepsilon.$$

Equivalently, the Cayley multigraphs $\mathrm{Cay}(\mathrm{SL}_2(\mathbb{F}_p), S_p)$ form a uniform expander family.

The proof appears in Volume IX. This volume supplies the first three nontrivial arrows in its dependency graph.

1.2 What has already been proved

A later volume should reuse established machinery instead of reproducing it. We therefore state the exact earlier inputs used by Arc III.

Proposition 1.3 (Arc-I analytic inputs). *The first three volumes established the following facts in the forms needed below.*

1. *For a symmetric probability measure μ on a finite group, convolution by μ is a self-adjoint contraction on ℓ^2 , and its even trace is a return probability.*
2. *If every nontrivial irreducible representation of G has dimension at least D , then every nonconstant eigenspace of a Cayley averaging operator has dimension at least D .*
3. *Failed L^2 -flattening of a probability measure yields, with polynomial quantitative losses, a translate carrying substantial mass of an approximate group whose cardinality is bounded above by the inverse collision scale $\|\mu\|_2^{-2}$ times a polynomial loss. A matching lower scale requires an anti-atom input; in Arc III that lower endpoint is recovered separately from the point-mass bound.*
4. *The noncommutative Ruzsa calculus, Balog–Szemerédi–Gowers theory, dyadic decomposition, and approximate-group conversion used in item (3) were all proved in Volume II rather than cited.*

Proposition 1.4 (Arc-II algebraic inputs). *Volumes IV–VI established the following.*

1. *There is an absolute constant J such that, for all sufficiently large primes p , every proper subgroup $H < \mathrm{SL}_2(\mathbb{F}_p)$ is either of cardinality at most J or contains a metabelian subgroup of index at most J .*
2. *For each fixed $\eta > 0$ there is $\delta = \delta(\eta) > 0$ such that a generating set $A \subset \mathrm{SL}_2(\mathbb{F}_p)$ with*

$$p^\eta \leq |A| \leq |\mathrm{SL}_2(\mathbb{F}_p)|^{1-\eta}$$

satisfies

$$|A^3| \geq |A|^{1+\delta}.$$

The equivalent approximate-group formulation, with polynomial dependence on the approximation parameter, is also available.

3. *Every nontrivial complex representation of $\mathrm{SL}_2(\mathbb{F}_p)$ has dimension $\gg p$; this was proved by the unipotent-restriction argument in Volume VI, without invoking Landazuri–Seitz.*
4. *Bounded-complexity proper subvarieties of a fixed-dimensional affine variety over $\mathbb{F}_p$ have $O(p^{d-1})$ rational points when the ambient irreducible variety has dimension d . The required degree and component bookkeeping was proved in Volume V.*

The first item is the only part of subgroup classification used for strong approximation and subgroup escape. We never need a verbatim Dickson list.

1.3 The eight interfaces

A rigorous expansion proof must keep eight logically different tasks separate.

1. **Arithmetic surjectivity.** Show $\Gamma \bmod p = \mathrm{SL}_2(\mathbb{F}_p)$ for all sufficiently large p .
2. **Free local geometry.** Replace S by a fixed bounded-word set containing a free pair.
3. **Logarithmic injectivity.** Show distinct short free words remain distinct modulo p up to length $c \log p$.
4. **Initial spreading.** Use the regular tree to obtain exponentially small atoms at logarithmic time.
5. **Subgroup escape.** Show every coset of every proper subgroup receives only $p^{-\kappa}$ mass.
6. **Inverse step.** Convert failed L^2 flattening into an approximate subgroup carrying noticeable mass.
7. **Growth contradiction.** Use Helfgott growth plus subgroup escape to exclude that approximate subgroup throughout the mesoscopic range.

8. **Spectral completion.** Once the walk reaches the quasirandom L^2 scale in $O(\log p)$ time, convert a return-probability saving into a constant spectral gap.

Warning 1.5 (Three tempting but false compressions). None of the following implications is valid by itself:

$$\begin{aligned} \text{small atoms} &\not\Rightarrow \text{subgroup escape,} \\ \text{product growth} &\not\Rightarrow \text{spectral gap,} \\ \text{Zariski density} &\not\Rightarrow \text{surjectivity mod } p. \end{aligned}$$

without quantitative intermediate arguments. The whole point of Arc III is to prove those bridges.

1.4 A quantitative ledger

Put

$$G_p = \mathrm{SL}_2(\mathbb{F}_p), \quad N_p = |G_p| = p(p^2 - 1) \asymp p^3.$$

All constants below are allowed to depend on the fixed generating set S , but never on p .

We shall produce constants

$$0 < c_{\mathrm{inj}}, c_{\mathrm{walk}}, \kappa_{\mathrm{esc}}, \delta_{\mathrm{fl}}, \eta_{\mathrm{qr}} < 1$$

with the following interpretation. A free pair obtained from bounded S -words is injective modulo p on a ball of radius $c_{\mathrm{inj}} \log p$. At time

$$n_0 = \lfloor c_{\mathrm{walk}} \log p \rfloor$$

its random-walk atoms are at most $p^{-\kappa_{\mathrm{esc}}}$ and every proper subgroup coset also has mass $\ll p^{-\kappa_{\mathrm{esc}}}$. In the mesoscopic range, each dyadic convolution step saves a factor $N_p^{-\delta_{\mathrm{fl}}}$ in L^2 . A bounded number of such steps reaches

$$\|\mu^{(n)}\|_2 \leq N_p^{-1/2 + \eta_{\mathrm{qr}}}$$

with $n \ll \log N_p$. Volume IX chooses $\eta_{\mathrm{qr}} < 1/6$ relative to the quasirandom exponent $1/3$ and closes the trace argument.

The order of choices matters. One first fixes the free pair and its word length, then the injectivity fraction, then the escape exponent, then a much smaller flattening exponent, and only afterward the quasirandom threshold. Choosing a flattening exponent before knowing the subgroup-escape saving is a common source of circular inequalities.

1.5 Historical location

The architecture is due to Bourgain and Gamburd [BG08]; Helfgott’s growth theorem is the rank-one algebraic engine [Hel08]. The general arithmetic surjectivity theorem belongs to strong approximation, with Matthews–Vaserstein–Weisfeiler as a central source [MVW84]. In this self-contained rank-one treatment, however, the logical dependence on the general strong-approximation theorem is removed in the strong-approximation chapter below.

Checkpoint 1.6. At the end of Arc III, every arrow from a fixed non-elementary integral generating set to a uniform prime-quotient spectral gap will either have been proved in Volumes VII–IX or will point to a theorem proved in Volumes I–VI. No general strong-approximation, Tits-alternative, BSG, Helfgott-growth, Larsen–Pink, or Landazuri–Seitz theorem is used as an unexplained black box.

Chapter 2

Ping–Pong, Free Subgroups, and Bounded Words

2.1 Hyperbolic dynamics on the projective line

An element $g \in \mathrm{PSL}_2(\mathbb{R})$ is hyperbolic when it has two distinct fixed points on $\mathbb{P}^1(\mathbb{R})$ and acts with one attracting and one repelling direction. Concretely, after conjugation it is represented by

$$a_\lambda = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}, \quad \lambda > 1.$$

In the affine coordinate $x \in \mathbb{R} \cup \{\infty\}$ this acts by $x \mapsto \lambda^2 x$. Thus ∞ is attracting for positive iteration and 0 is repelling.

Lemma 2.1 (Uniform contraction away from the repeller). *Let $a \in \mathrm{PSL}_2(\mathbb{R})$ be hyperbolic, with attracting fixed point a^+ and repelling fixed point a^- . Let U^+ and U^- be disjoint neighborhoods of a^+ and a^- . Then there exists m_0 such that for all $m \geq m_0$,*

$$a^m(\mathbb{P}^1(\mathbb{R}) \setminus U^-) \subset U^+, \quad a^{-m}(\mathbb{P}^1(\mathbb{R}) \setminus U^+) \subset U^-.$$

Proof. Conjugate a to $x \mapsto \lambda^2 x$. Neighborhoods of 0 and ∞ contain sets of the form $\{|x| < r\}$ and $\{|x| > R\} \cup \{\infty\}$. On the complement of $\{|x| < r\}$ one has $|\lambda^{2m} x| \geq \lambda^{2m} r$, which exceeds R for large m . This proves the first inclusion. Apply the same argument to a^{-1} for the second. Conjugating back preserves inclusions. $\square$

2.2 The ping–pong lemma

Lemma 2.2 (Ping–pong). *Let a group act on a set X . Suppose a, b are elements and there are four pairwise disjoint nonempty sets*

$$A^+, A^-, B^+, B^- \subset X$$

such that

$$a^{\pm 1}(X \setminus A^\mp) \subset A^\pm, \quad b^{\pm 1}(X \setminus B^\mp) \subset B^\pm.$$

Then a and b freely generate a free group of rank two.

Proof. Choose a point x_0 outside the four ping–pong sets. Let

$$w = s_1 s_2 \cdots s_r$$

be a nonempty reduced word in $a^{\pm 1}, b^{\pm 1}$. We apply the letters from right to left. Because the word is reduced, whenever the current point has entered the attracting set corresponding to s_{j+1} , it does not lie in the forbidden repelling set for s_j . Thus each successive letter sends the point into its own attracting set. At the end, wx_0 lies in one of $A^\pm, B^\pm$, whereas x_0 lies in none. Hence $wx_0 \neq x_0$, so $w \neq e$. No nonempty reduced word is trivial. $\square$

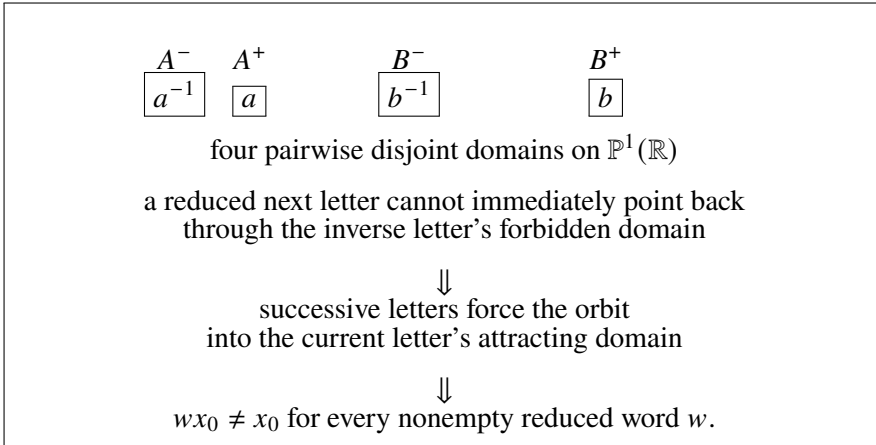


Figure 2.1: The ping–pong mechanism. Disjoint attracting/repelling domains turn reduced-word syntax into geometric separation, hence freeness.

Proposition 2.3 (Disjoint hyperbolics give a free pair after fixed powers). *Let $a, b \in \text{PSL}_2(\mathbb{R})$ be hyperbolic and suppose their two fixed-point sets are disjoint. Then there exists $m \geq 1$ such that a^m and b^m freely generate.*

Proof. Choose pairwise disjoint neighborhoods A^+, A^-, B^+, B^- of the four fixed points. By Lemma 2.1, sufficiently large common m makes $a^{\pm m}$ and $b^{\pm m}$ satisfy the inclusions of Lemma 2.2. Apply that lemma. $\square$

2.3 Returning to the original generators

Let $S \subset \mathrm{SL}_2(\mathbb{Z})$ be the fixed generating set. By non-elementarity there are hyperbolic $a, b \in \Gamma$ with disjoint fixed points. Each is represented by some fixed word in S , say of lengths L_a, L_b . Choose m from Proposition 2.3 and set

$$A = a^m, \quad B = b^m.$$

After lifting from PSL_2 to SL_2 , square both if necessary so that no central sign can occur in a projected relation. We obtain:

Proposition 2.4 (A bounded-word free pair). *There are $A, B \in \Gamma$ and a constant $L = L(S)$ such that*

1. $\langle A, B \rangle \cong F_2$;
2. each of $A^{\pm 1}, B^{\pm 1}$ is a word of length at most L in S .

All constants are independent of p .

Proof. The construction above gives A, B as fixed powers of fixed elements of Γ . Their S -word lengths are therefore fixed. The projective ping-pong subgroup is free. If a reduced word in the chosen lifts were $\pm I$, its image in PSL_2 would be trivial. Ping-pong forbids this unless the word is empty. Hence the lifted subgroup is also free. $\square$

2.4 Why changing generators is harmless

The expansion theorem is stated for S_p , but the free random walk will first be run on

$$T = \{A, A^{-1}, B, B^{-1}\}.$$

Every $t \in T$ is an S -word of length at most L . A spectral gap for T_p therefore transfers to S_p with a constant depending only on L and $|S|$.

Proposition 2.5 (Dirichlet-form comparison under bounded words). *Let G be finite, $S = S^{-1}$ generate G , and let $T = T^{-1}$ be another generating set such that every*

$t \in T$ is an S -word of length at most L . If γ_T is the ordinary spectral gap of the T -averaging operator, then

$$\gamma_S \geq \frac{\gamma_T}{L^2|S|}.$$

In particular, a uniform T -gap gives a uniform S -gap.

Proof. For a right translation R_g and $f \in \ell^2(G)$, write a word $t = s_1 \cdots s_\ell$, $\ell \leq L$. Telescoping gives

$$R_t f - f = \sum_{j=1}^{\ell} R_{s_1 \cdots s_{j-1}} (R_{s_j} f - f).$$

Unitary invariance and Cauchy–Schwarz yield

$$\|R_t f - f\|_2^2 \leq L \sum_{j=1}^{\ell} \|R_{s_j} f - f\|_2^2 \leq L^2 \sum_{s \in S} \|R_s f - f\|_2^2.$$

Averaging over $t \in T$ and using

$$\langle (I - P_U) f, f \rangle = \frac{1}{2|U|} \sum_{u \in U} \|R_u f - f\|_2^2$$

for a symmetric set U gives

$$\langle (I - P_T) f, f \rangle \leq L^2 |S| \langle (I - P_S) f, f \rangle.$$

For $f \perp 1$, the left side is at least $\gamma_T \|f\|_2^2$. Rearranging and taking the infimum over nonconstant f proves the claim. $\square$

Remark 2.6. The numerical factor is deliberately not optimized. What matters is that L is fixed before p varies. A word length growing with p would destroy the argument.

2.5 Free balls and reduced words

Let $F_2 = \langle A, B \rangle$. Denote by $B_{F_2}(r)$ the set of reduced words of length at most r in $A^{\pm 1}, B^{\pm 1}$. Its cardinality is

$$1 + 4 \sum_{j=0}^{r-1} 3^j = 2 \cdot 3^r - 1.$$

Example 2.7 (What radius three already contains). For $r = 3$ the free ball has

$$|B_{F_2}(3)| = 2 \cdot 3^3 - 1 = 53$$

distinct elements: one identity element, 4 words of length one, 12 of length two, and 36 of length three. Therefore, whenever reduction modulo p is injective on this ball, all 53 endpoints remain distinct in G_p . At logarithmic radius $r \asymp \log p$, the same formula becomes a power of p ; this is the concrete combinatorial reservoir later converted into atom and subgroup-coset escape.

Thus free geometry supplies exponential combinatorial volume. After the Zariski-density and strong-approximation interfaces are closed, the height chapter asks how much of this free ball survives reduction modulo p . Because the matrix entries of a word grow only exponentially, the answer is a ball of logarithmic radius.

Checkpoint 2.8. We have replaced the original arithmetic generators by a fixed bounded-word free pair without losing the possibility of transferring a final spectral gap back to S . From now until the final transfer in Volume IX, the local walk may be analyzed on the four-element free generating set T .

2.6 Worked ping-pong calculation on the projective line

Consider

$$a = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}, \quad b = rar^{-1}, \quad r = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

On $\mathbf{P}^1(\mathbf{R})$, a has attracting line $[1 : 0]$ and repelling line $[0 : 1]$, whereas b has the corresponding pair obtained by rotating by $\pi/4$. For sufficiently large m , the maps $a^{\pm m}$ and $b^{\pm m}$ send the complement of a small neighborhood of their repelling line into a small neighborhood of their attracting line. Choosing the four neighborhoods disjoint gives the ping-pong hypotheses. Therefore every reduced word of length r in a^m, b^m is nontrivial, and the radius- r ball has

$$1 + 4 \sum_{j=0}^{r-1} 3^j = 2 \cdot 3^r - 1$$

distinct elements. At $r = 3$ this is 53, making the exponential tree growth used later completely explicit.

Chapter 3

Non-Elementary Subgroups and Zariski Density in Rank One

3.1 Why Zariski density enters twice

Zariski density has two distinct roles. First, it prevents the integral group from satisfying a nontrivial polynomial identity forced by infinitely many proper reductions. Second, it lets a word identity verified on Γ extend to an identity on the whole algebraic group SL_2 . The rank-one proof requires only elementary algebraic-group facts already developed in Volume V plus a short Lie-algebra observation.

Let $\mathbf{G} = \mathrm{SL}_2$ viewed as an affine algebraic group over $\mathbb{R}$, and let $\mathfrak{g} = \mathfrak{sl}_2(\mathbb{R})$.

Lemma 3.1 (Proper Lie subalgebras of $\mathfrak{sl}_2$ are solvable). *Every proper Lie subalgebra $\mathfrak{h} < \mathfrak{sl}_2(\mathbb{R})$ is solvable.*

Proof. The ambient Lie algebra has dimension 3. A proper subalgebra therefore has dimension 0, 1, or 2. The first two cases are abelian. If $\dim \mathfrak{h} = 2$, then $[\mathfrak{h}, \mathfrak{h}]$ is a proper subspace of $\mathfrak{h}$: indeed an alternating bilinear bracket on a two-dimensional space is spanned by the single value $[X, Y]$ for a basis X, Y . Hence $\dim[\mathfrak{h}, \mathfrak{h}] \leq 1$. Every one-dimensional Lie algebra is abelian, so

$$[[\mathfrak{h}, \mathfrak{h}], [\mathfrak{h}, \mathfrak{h}]] = 0.$$

Thus the derived series terminates after at most two nontrivial steps. $\square$

Lemma 3.2 (Proper algebraic subgroups are virtually solvable). *Let $\mathbf{H} < \mathrm{SL}_2(\mathbb{C})$ be a proper algebraic subgroup. Then the abstract group $\mathbf{H}(\mathbb{C})$ is virtually solvable.*

Proof. Volume V proves that the algebraic identity component $\mathbf{H}^\circ$ has finite index in $\mathbf{H}$. It is therefore enough to prove that $\mathbf{H}^\circ(\mathbb{C})$ is solvable. Since SL_2 is irreducible

of dimension 3 and $\mathbf{H}$ is proper,

$$\dim \mathbf{H}^\circ \leq 2.$$

Write

$$\mathfrak{h} = \text{Lie}(\mathbf{H}^\circ) \subset \mathfrak{sl}_2(\mathbb{C}).$$

The same two-dimensional argument as in [Lemma 3.1](#) shows that $\mathfrak{h}$ is solvable. In this rank-one situation we can make the passage from the Lie algebra to the group completely explicit by triangularizing $\mathfrak{h}$.

If $\dim \mathfrak{h} \leq 1$, choose a nonzero generator X when necessary. Over $\mathbb{C}$ the matrix X has an eigenline, so after conjugation $\mathfrak{h}$ is contained in the upper-triangular Lie algebra

$$\mathfrak{b} = \left\{ \begin{pmatrix} a & b \\ 0 & -a \end{pmatrix} : a, b \in \mathbb{C} \right\}.$$

Suppose now that $\dim \mathfrak{h} = 2$. If $\mathfrak{h}$ is abelian, two commuting complex 2×2 matrices have a common eigenline: take an eigenline of a non-scalar element; its commuting partner preserves either that eigenspace or, in the repeated-eigenvalue case, an eigenline of the common Jordan flag. Hence again $\mathfrak{h}$ is conjugate into $\mathfrak{b}$.

In the nonabelian two-dimensional case, choose a basis X, Y such that

$$[X, Y] = Y.$$

Indeed, the one-dimensional derived algebra is spanned by a nonzero Y , and rescaling a complementary vector gives the displayed relation. Since $Y \in \mathfrak{sl}_2(\mathbb{C})$,

$$\text{tr}(Y^2) = \text{tr}([X, Y]Y) = \text{tr}(XY^2 - YXY) = 0.$$

For a traceless 2×2 matrix this implies $\det Y = 0$, so Y is nonzero nilpotent. Its kernel is a line L . From

$$YX = XY - Y$$

we obtain $Y(Xv) = 0$ whenever $Yv = 0$, so X preserves L as well. Thus X and Y have a common invariant line, and after conjugation $\mathfrak{h} \subset \mathfrak{b}$ in every case.

Let $\mathbf{B} < \text{SL}_2(\mathbb{C})$ be the corresponding upper-triangular Borel subgroup. The complex Lie group $\mathbf{H}^\circ(\mathbb{C})$ is connected, and the exponential map is a local diffeomorphism at the identity. Because $\mathfrak{h} \subset \mathfrak{b}$, a sufficiently small identity neighborhood in $\mathbf{H}^\circ(\mathbb{C})$ lies in $\mathbf{B}(\mathbb{C})$. Therefore

$$\mathbf{H}^\circ(\mathbb{C}) \cap \mathbf{B}(\mathbb{C})$$

is an open subgroup of the connected group $\mathbf{H}^\circ(\mathbb{C})$, hence equals the whole group. So $\mathbf{H}^\circ(\mathbb{C}) \subset \mathbf{B}(\mathbb{C})$.

Finally $\mathbf{B}(\mathbb{C})$ is solvable: the diagonal-character map

$$\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \mapsto a$$

has abelian image and abelian unipotent kernel. Hence $\mathbf{H}^\circ(\mathbb{C})$ is solvable. Since $\mathbf{H}^\circ$ has finite algebraic index in $\mathbf{H}$, the group $\mathbf{H}(\mathbb{C})$ is virtually solvable. $\square$

Lemma 3.3 (A finite-index subgroup of F_2 is not solvable). *Every finite-index subgroup of the rank-two free group F_2 is a free group of rank at least two. In particular, F_2 is not virtually solvable.*

Proof. Let $H \leq F_2 = \langle a, b \rangle$ have finite index d . Form the finite Schreier graph whose vertices are the right cosets Hg and which has, at each vertex, one directed a -edge and one directed b -edge (with inverse edges understood). Choose a spanning tree. For each oriented edge outside the tree, follow the tree from the base coset H to the initial vertex, traverse that edge, and return along the tree. The resulting words lie in H and freely generate H : every closed path reduces uniquely after the tree edges are collapsed, so no further relation is introduced.

The graph has d vertices and $2d$ unoriented generator edges, hence the number of edges outside a spanning tree is

$$2d - (d - 1) = d + 1 \geq 2.$$

Thus H is free of rank $d + 1$, in particular nonabelian and nonsolvable. Therefore no solvable subgroup can have finite index in F_2 . $\square$

Proposition 3.4 (Non-elementary implies Zariski dense). *If $\Gamma < \mathrm{SL}_2(\mathbb{R})$ is non-elementary, then its Zariski closure is all of SL_2 .*

Proof. By definition, the projective image of Γ contains hyperbolic elements a, b with disjoint fixed-point sets. Choose sufficiently large positive powers a^m, b^m . The ping-pong lemma, proved from first principles in the preceding chapter, shows that $\langle a^m, b^m \rangle$ is a free group of rank two. A virtually solvable group cannot contain a nonabelian free subgroup. Indeed, if L is solvable of finite index in a group containing F_2 , then $F_2 \cap L$ is a solvable finite-index subgroup of F_2 , contradicting Lemma 3.3.

Thus Γ is not virtually solvable. If the real Zariski closure of Γ were proper, some nonzero real regular function on SL_2 would vanish on Γ . Let $\mathbf{H} < \mathrm{SL}_2(\mathbb{C})$ be the complex Zariski closure of Γ . Zariski closure preserves the subgroup property because multiplication and inversion are regular, and the same nonzero polynomial vanishes on $\mathbf{H}$; hence $\mathbf{H}$ is proper. By Lemma 3.2, $\mathbf{H}(\mathbb{C})$ is virtually solvable, and therefore so is its subgroup Γ . Contradiction. $\square$

The ping-pong input used here was proved in the preceding chapter, so the Zariski-density argument has no forward proof dependency.

3.2 Word maps are regular

Let $F_r = \langle x_1, \dots, x_r \rangle$ be the free group. A reduced word $w \in F_r$ defines

$$w_{\mathbf{G}} : \mathbf{G}^r \longrightarrow \mathbf{G}, \quad (g_1, \dots, g_r) \longmapsto w(g_1, \dots, g_r).$$

Lemma 3.5 (Regularity of word maps). *For $\mathbf{G} = \mathrm{SL}_2$, every word map $w_{\mathbf{G}}$ is a regular morphism of affine varieties.*

Proof. Multiplication is polynomial in the four matrix entries. On SL_2 , inversion is also polynomial:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

because $ad - bc = 1$. A word map is a finite composition of multiplication, coordinate projections, and inversion, hence is regular. $\square$

Corollary 3.6 (A law on a Zariski-dense subgroup is an algebraic law). *Let $\Gamma < \mathrm{SL}_2(\mathbb{R})$ be Zariski dense. If a word $w \in F_r$ satisfies*

$$w(\gamma_1, \dots, \gamma_r) = I \quad (\gamma_i \in \Gamma),$$

then $w_{\mathbf{G}}$ is identically I on $\mathrm{SL}_2(\mathbb{R})^r$.

Proof. The product Γ^r is Zariski dense in $\mathbf{G}^r$: this follows iteratively from the fact that a polynomial vanishing on a product of dense subsets vanishes first in one block of variables and then in the others. By Lemma 3.5, each matrix entry of $w_{\mathbf{G}} - I$ is a regular function. It vanishes on the dense subset Γ^r , hence vanishes identically. $\square$

3.3 No nontrivial law holds in SL_2

For the strong-approximation contradiction we need one explicit reason that SL_2 does not satisfy a nontrivial group law.

Lemma 3.7 (A free subgroup forbids laws). *If a group G contains a subgroup isomorphic to the free group F_2 , then no nontrivial word $w \in F_r$ is an identity on G .*

Proof. It suffices to show that for every nontrivial $w \in F_r$, there are elements of F_2 on which w is nontrivial. The free group F_2 contains a free subgroup of rank r : for example, if $F_2 = \langle a, b \rangle$, the conjugates

$$b, aba^{-1}, a^2ba^{-2}, \dots, a^{r-1}ba^{-(r-1)}$$

freely generate a rank- r subgroup. This follows from reduced normal form: in a product of powers of these conjugates, adjacent a -blocks can cancel only when the same conjugate appears consecutively, exactly the cancellation already present in the reduced word in the rank- r generators. Embed F_r by this map. Since w is nontrivial in F_r , its image is nontrivial in $F_2 \leq G$. $\square$

Proposition 3.8 ($\mathrm{SL}_2(\mathbb{R})$ contains a nonabelian free subgroup). *The group $\mathrm{SL}_2(\mathbb{R})$ contains a free subgroup of rank two.*

Proof. Take

$$a = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}, \quad m = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad b = mam^{-1}.$$

The projective fixed points of a are 0 and ∞ . The Möbius map defined by m sends these to $1/2$ and 1, so the fixed-point sets of a and b are disjoint. The preceding chapter proves that sufficiently large common powers a^r, b^r satisfy ping-pong and freely generate in $\mathrm{PSL}_2(\mathbb{R})$. Their chosen lifts lie in $\mathrm{SL}_2(\mathbb{R})$. If a reduced word in those lifts were I or $-I$, its projective image would be trivial, contradicting ping-pong. Hence the lifted subgroup is free of rank two. $\square$

Combining [Corollary 3.6](#), [Lemma 3.7](#), and [Proposition 3.8](#) gives the exact rigidity principle used in the next chapter: a Zariski-dense subgroup of $\mathrm{SL}_2(\mathbb{R})$ cannot satisfy any fixed nontrivial law.

3.4 Why this rank-one argument is enough

General strong approximation is much broader: it concerns Zariski-dense subgroups of simply connected semisimple algebraic groups over global fields and controls reductions at almost every place. We need only the special case $\Gamma < \mathrm{SL}_2(\mathbb{Z})$ and prime reductions. The subgroup geometry proved in Volume IV is so restrictive in rank one that a fixed-law contradiction replaces the general arithmetic theory.

Checkpoint 3.9. The arithmetic argument may now use the following implication without any external structural theorem:

$$\begin{aligned} \Gamma \text{ non-elementary} &\implies \Gamma \text{ Zariski dense} \\ &\implies \Gamma \text{ satisfies no nontrivial fixed group law.} \end{aligned}$$

The only geometric input is the ping-pong lemma proved in the preceding chapter.

Chapter 4

Rank-One Strong Approximation from Subgroup Laws

4.1 The statement actually needed

Theorem 4.1 (Strong approximation for non-elementary subgroups of $\mathrm{SL}_2(\mathbb{Z})$). *Let $\Gamma < \mathrm{SL}_2(\mathbb{Z})$ be finitely generated and non-elementary. Then there exists $p_0(\Gamma)$ such that for every prime $p > p_0(\Gamma)$ the reduction map satisfies*

$$\pi_p(\Gamma) = \mathrm{SL}_2(\mathbb{F}_p).$$

This is the only strong-approximation statement used to prove the prime-quotient Bourgain–Gamburd theorem. We now prove it internally.

4.2 A uniform law for every proper subgroup

We isolate the consequence of the subgroup theorem proved in Volume IV.

Proposition 4.2 (Uniform proper-subgroup dichotomy; dependency from Volume IV). *There are absolute constants $J \geq 2$ and p_1 such that for every prime $p > p_1$ and every proper subgroup $H < \mathrm{SL}_2(\mathbb{F}_p)$, one of the following holds:*

1. $|H| \leq J$;
2. H contains a metabelian subgroup L with $[H : L] \leq J$.

Recall that metabelian means $L'' = 1$, equivalently $[L, L]$ is abelian. Volume IV obtained this statement directly from the projective action: Borel-type subgroups are solvable of derived length at most two, torus normalizers are virtually abelian, and the remaining exceptional groups have bounded order.

Lemma 4.3 (Bounded index gives a normal metabelian core). *If $L \leq H$ has index at most J , then the core*

$$N = \bigcap_{h \in H} hLh^{-1}$$

is normal in H , satisfies $[H : N] \leq J!$, and is a subgroup of L . In particular, if L is metabelian, so is N .

Proof. The left action of H on H/L gives a homomorphism $H \rightarrow S_m$ with $m = [H : L] \leq J$. Its kernel is exactly the core N , so $[H : N] \leq m! \leq J!$. Since the term with $h = e$ occurs in the intersection, $N \leq L$. A subgroup of a metabelian group is metabelian because its second derived subgroup lies in $L'' = 1$. $\square$

Set $M = J!$. Consider the words

$$u(x_1, x_2, x_3, x_4) = [[x_1^M, x_2^M], [x_3^M, x_4^M]]$$

and

$$v(y) = y^M.$$

If H has a normal metabelian subgroup N of index at most M , then every $h \in H$ has $h^M \in N$: the quotient H/N has order at most M , so the order of every quotient element divides $M!$; replacing M at the outset by $(J!)!$ if necessary makes this statement literal. To avoid changing notation repeatedly, from now on fix

$$M = (J!)!$$

Then $u = 1$ on every subgroup in alternative (2), while $v = 1$ on every group of order at most J because every element order divides $J! \mid M$.

We need one word valid in both alternatives.

Lemma 4.4 (A single nontrivial law for all proper subgroups). *With M as above, the word*

$$W = [u(x_1, x_2, x_3, x_4), v(y)] \in F_5$$

is nontrivial and is an identity on every proper subgroup of $\mathrm{SL}_2(\mathbb{F}_p)$ for every $p > p_1$.

Proof. If H has order at most J , then $v(y) = 1$ for every $y \in H$, hence $W = 1$. If H has a metabelian subgroup of index at most J , pass to its normal core N . By the choice of M , $x^M \in N$ for every $x \in H$. Therefore each commutator $[x_1^M, x_2^M]$ lies in N' , and N' is abelian. Hence $u = 1$ and again $W = 1$.

It remains to check that W is not the trivial free word. The variables occurring in u and v are disjoint. The reduced word u is nontrivial: in the free group, $[x_1^M, x_2^M]$

and $[x_3^M, x_4^M]$ are nontrivial elements in free factors on disjoint variable pairs, so their commutator is reduced and nontrivial. Likewise $v = y^M \neq 1$. In the free product

$$F(x_1, x_2, x_3, x_4) * F(y),$$

the commutator $uvu^{-1}v^{-1}$ alternates nontrivial syllables from the two free factors and is therefore reduced. Thus $W \neq 1$. $\square$

4.3 Infinitely many proper reductions force an integral law

Let $S = \{s_1, \dots, s_k\}$ generate Γ and set

$$H_p = \pi_p(\Gamma) \leq \mathrm{SL}_2(\mathbb{F}_p).$$

Suppose H_p is proper for infinitely many primes $p > p_1$. By Lemma 4.4, W is an identity on each such H_p .

Lemma 4.5 (Congruence at infinitely many primes forces equality). *If $A \in M_2(\mathbb{Z})$ satisfies $A \equiv 0 \pmod{p}$ for infinitely many primes p , then $A = 0$.*

Proof. Every entry of A is an integer divisible by infinitely many distinct primes. A nonzero integer has only finitely many prime divisors, so every entry is zero. $\square$

Proposition 4.6 (Proper reductions force a group law upstairs). *If H_p is proper for infinitely many primes $p > p_1$, then Γ satisfies the law W .*

Proof. Take arbitrary $\gamma_1, \dots, \gamma_5 \in \Gamma$. For every prime in the infinite exceptional set,

$$\pi_p(W(\gamma_1, \dots, \gamma_5)) = I.$$

Thus the integral matrix

$$W(\gamma_1, \dots, \gamma_5) - I$$

is congruent to zero modulo infinitely many primes. By Lemma 4.5 it is zero. Since the tuple was arbitrary, W is a law on Γ . $\square$

4.4 Completion of strong approximation

Proof of Theorem 4.1. By Proposition 3.4, Γ is Zariski dense in SL_2 . Suppose for contradiction that $\pi_p(\Gamma)$ is proper for infinitely many primes. Then Proposition 4.6 says that the fixed nontrivial word W is a law on Γ . By Corollary 3.6, W is then an identity on all of $\mathrm{SL}_2(\mathbb{R})$.

But $\mathrm{SL}_2(\mathbb{R})$ contains a nonabelian free subgroup by [Proposition 3.8](#), and [Lemma 3.7](#) says that no nontrivial word can be a law on a group containing F_2 . This contradicts the nontriviality of W proved in [Lemma 4.4](#). Hence only finitely many primes have proper image. $\square$

4.5 What has and has not been proved

The theorem above is a rank-one strong-approximation theorem for prime reductions of subgroups of $\mathrm{SL}_2(\mathbb{Z})$. It does not claim the full Matthews–Vaserstein–Weisfeiler theorem [[MVW84](#)]; in particular, it does not address arbitrary simply connected semisimple groups, number fields, or simultaneous control over composite moduli. The proof works because Volume IV established a very rigid uniform proper-subgroup dichotomy in $\mathrm{SL}_2(\mathbb{F}_p)$.

Guiding Idea 4.7. Strong approximation appears here as a compactness-of-obstructions principle. If infinitely many reductions were trapped in proper subgroups, rank-one subgroup geometry would force one fixed group identity at every bad prime. Integrality lifts that identity upstairs, and Zariski density spreads it to the whole algebraic group. A single free subgroup then destroys the identity.

Example 4.8 (Strong approximation without any black box). Consider the two elementary matrices

$$u = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \ell = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}).$$

Modulo a prime p , their powers give every upper and lower elementary matrix

$$u^a = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}, \quad \ell^b = \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix} \quad (a, b \in \mathbb{F}_p).$$

Gaussian elimination in $\mathrm{SL}_2(\mathbb{F}_p)$ expresses every matrix as a product of such elementary matrices (together with a diagonal matrix, which itself is a bounded product of elementary matrices). Hence u and ℓ already generate $\mathrm{SL}_2(\mathbb{F}_p)$ for every prime p .

The theorem proved above is more subtle because a general non-elementary integral generating set need not visibly contain elementary matrices. The group-law argument replaces this explicit elimination: it shows that infinitely many proper reductions would force a single identity upstairs, and Zariski density plus a free subgroup rules that out.

4.6 A useful variant

The same proof gives a formulation convenient for later arithmetic families.

Corollary 4.9 (Finite bad-prime set). *Let $S \subset \mathrm{SL}_2(\mathbb{Z})$ be fixed and non-elementary. There is a finite set of primes $\mathcal{B}(S)$ such that for every $p \notin \mathcal{B}(S)$, the reduction S_p generates $\mathrm{SL}_2(\mathbb{F}_p)$.*

Proof. Take $\mathcal{B}(S)$ to contain all primes at most $p_0(\langle S \rangle)$ from [Theorem 4.1](#). No denominator primes occur because the generators are integral. $\square$

Checkpoint 4.10. The arithmetic quotient in the remainder of Arc III is now exactly $G_p = \mathrm{SL}_2(\mathbb{F}_p)$ for every sufficiently large prime. The expansion proof will never again need to invoke strong approximation.

Chapter 5

Matrix Height and Logarithmic Injectivity Modulo p

5.1 A height adapted to multiplication

For an integral 2×2 matrix $g = (g_{ij})$, define

$$H(g) = \max_{i,j} |g_{ij}|.$$

This is not multiplicative, but it is submultiplicative up to a fixed factor.

Lemma 5.1 (Height of a product). *For $g, h \in M_2(\mathbb{Z})$,*

$$H(gh) \leq 2H(g)H(h).$$

Consequently, if T is a fixed finite set of integral matrices and

$$M = \max\{2, \max_{t \in T} H(t)\},$$

then every word $w = t_1 \cdots t_r$ satisfies

$$H(w) \leq (2M)^r.$$

Proof. Each entry of gh is a sum of two products of entries, so its absolute value is at most $2H(g)H(h)$. Iterating and absorbing the first factor into M gives the exponential bound. $\square$

The exact base of the exponential is irrelevant. What matters is that its logarithm depends only on the fixed integral generators.

5.2 Congruence equality and actual equality

Lemma 5.2 (Small integral matrices cannot vanish modulo a larger prime). *Let $X, Y \in M_2(\mathbb{Z})$ and let p be prime. If $X \equiv Y \pmod{p}$ and*

$$H(X - Y) < p,$$

then $X = Y$.

Proof. Every entry of $X - Y$ is divisible by p and has absolute value strictly less than p . The only such multiple of p is zero. $\square$

Theorem 5.3 (Logarithmic injectivity). *Let $T = \{A^{\pm 1}, B^{\pm 1}\} \subset \mathrm{SL}_2(\mathbb{Z})$ freely generate F_2 . There exists $c_{\mathrm{inj}} = c(T) > 0$ and $p_2(T)$ such that for every prime $p > p_2$, reduction modulo p is injective on the free ball*

$$B_{F_2}(c_{\mathrm{inj}} \log p).$$

Equivalently, if two reduced words u, v have lengths at most r and

$$r \leq c_{\mathrm{inj}} \log p,$$

then $u_p = v_p$ implies $u = v$.

Proof. If $u_p = v_p$, then $(uv^{-1})_p = I$. The reduced length of uv^{-1} is at most $2r$. Let $K \geq 2$ be such that every T -word of length m has height at most K^m , as supplied by Lemma 5.1. Then

$$H(uv^{-1} - I) \leq H(uv^{-1}) + 1 \leq K^{2r} + 1.$$

Choose $c_{\mathrm{inj}} > 0$ so small that

$$K^{2c_{\mathrm{inj}} \log p} \leq p^{1/2}$$

for all p ; for example $c_{\mathrm{inj}} = 1/(4 \log K)$ when $\log$ is natural. If $r \leq c_{\mathrm{inj}} \log p$, then for large p ,

$$H(uv^{-1} - I) \leq K^{2r} + 1 \leq p^{1/2} + 1 < p.$$

Since $uv^{-1} \equiv I \pmod{p}$, Lemma 5.2 gives $uv^{-1} = I$ as an integral matrix. The subgroup generated by T is free, hence $u = v$ as reduced words. $\square$

Remark 5.4 (Why the factor two is explicit). To prove injectivity on a ball of radius r , one must control the difference word uv^{-1} of length up to $2r$; this factor is already built into the choice of c_{inj} . Later, the subgroup-law argument substitutes short words into a commutator-of-commutators whose length is another fixed multiple of r . We shall therefore reserve only a small fixed fraction of the logarithmic injectivity radius for the random walk itself.

5.3 A girth formulation

The Cayley graph $\text{Cay}(\langle T_p \rangle, T_p)$ has girth larger than R if there is no nonempty reduced T -word of length at most R that becomes the identity modulo p .

Corollary 5.5 (Logarithmic girth). *There exists $c_g > 0$ such that for all sufficiently large primes p ,*

$$\text{girth } \text{Cay}(\langle T_p \rangle, T_p) \geq c_g \log p.$$

Proof. Apply [Theorem 5.3](#) with $v = e$. Any nontrivial reduced word inside the injective ball remains nontrivial after reduction. $\square$

5.4 Schreier injectivity

For subgroup escape we need a slightly different notion. Fix $H < G_p$ and consider the Schreier graph on right cosets $H \backslash G_p$. A collision

$$Hu_p = Hv_p$$

means $u_p v_p^{-1} \in H$, not necessarily the identity. Thus ordinary girth does not by itself imply Schreier injectivity. Volume VIII will combine ordinary injectivity with a group law satisfied by H : too many words in one coset create many short difference words in a metabelian subgroup; substituting those differences into the metabelian law creates an actual identity word, which can then be lifted through [Theorem 5.3](#).

This is the precise reason subgroup escape is more difficult than point spreading.

5.5 Initial L^2 consequences of injectivity

Let μ_T be the uniform probability on T_p . For n within the injectivity window, every reduced free word represents a distinct group element, but the simple random walk is not uniform on reduced words: backtracking produces many paths leading to the same reduced endpoint. The correct endpoint distribution is exactly the simple random walk on the 4-regular tree. Its return probabilities and maximum atom are exponentially small. We prove a sufficient tree estimate in the next chapter.

Guiding Idea 5.6. The logarithmic scale is forced by a competition between two exponentials. The number and size of word coefficients grow like $\exp(Cr)$, while the modulus is p . Solving $\exp(Cr) < p$ gives $r < c \log p$. Bourgain–Gamburd works because this logarithmic time is also long enough for a nonamenable free walk to gain a power of p in its point probabilities.

Chapter 6

The Free-Group Walk and the First Spreading Estimate

6.1 The regular tree operator

Let F_k be the free group on $k \geq 2$ generators and let

$$T_k = \{a_1^{\pm 1}, \dots, a_k^{\pm 1}\}, \quad d = 2k.$$

The Cayley graph is the d -regular tree. Let P be the simple random-walk operator

$$Pf(x) = \frac{1}{d} \sum_{t \in T_k} f(xt).$$

We need only an upper bound $\|P\|_{\ell^2(F_k)} < 1$, but we prove the familiar sharp value.

Theorem 6.1 (Norm of the regular-tree walk). *On the d -regular tree, $d \geq 3$,*

$$\|P\|_{2 \rightarrow 2} = \frac{2\sqrt{d-1}}{d} =: \rho_d < 1.$$

Proof. We first prove the upper bound for finitely supported real f . Root the tree at o . Every non-root vertex has one parent and $d-1$ children. Writing each unoriented edge once as parent-child,

$$\langle Pf, f \rangle = \frac{2}{d} \sum_{x \rightarrow y} f(x)f(y).$$

For any $q > 0$,

$$2|f(x)f(y)| \leq qf(x)^2 + q^{-1}f(y)^2.$$

Choose $q = (d - 1)^{-1/2}$. Summing over oriented parent–child edges, a non-root value $f(v)^2$ appears $d - 1$ times as a parent with coefficient q and once as a child with coefficient q^{-1} ; the total is

$$(d - 1)q + q^{-1} = 2\sqrt{d - 1}.$$

At the root only the parent contribution occurs, giving $dq \leq 2\sqrt{d - 1}$ for $d \geq 2$. Therefore

$$|\langle Pf, f \rangle| \leq \frac{2\sqrt{d - 1}}{d} \|f\|_2^2.$$

Since P is self-adjoint, this proves $\|P\| \leq \rho_d$.

For the reverse inequality, define a radial function supported on the ball of radius R by

$$f_R(x) = (d - 1)^{-|x|/2} \quad (|x| \leq R),$$

and zero outside. The number of vertices on sphere $r \geq 1$ is $d(d - 1)^{r-1}$, so each sphere contributes a constant amount to $\|f_R\|_2^2$ and hence $\|f_R\|_2^2 \asymp R$. Away from the root and the outer boundary,

$$Pf_R(x) = \frac{1}{d} \left((d - 1)^{1/2} + (d - 1)(d - 1)^{-1/2} \right) f_R(x) = \rho_d f_R(x).$$

Only $O(1)$ sphere contributions are lost at the two boundaries, so

$$\frac{\|Pf_R - \rho_d f_R\|_2}{\|f_R\|_2} \rightarrow 0.$$

Thus ρ_d belongs to the approximate spectrum and $\|P\| \geq \rho_d$. □

For F_2 , $d = 4$ and $\rho_4 = \sqrt{3}/2$.

6.2 Point probabilities

Let $\nu^{(n)}$ be the n -step distribution of the free walk from the identity.

Proposition 6.2 (Exponential point bound). *For every $x \in F_k$ and every $n \geq 0$,*

$$\nu^{(n)}(x) \leq \rho_d^n.$$

Moreover,

$$\|\nu^{(n)}\|_2 \leq \rho_d^n.$$

Proof. Identify $v^{(n)}(x)$ with the matrix coefficient

$$\langle P^n \delta_e, \delta_x \rangle.$$

Cauchy–Schwarz and [Theorem 6.1](#) give

$$|v^{(n)}(x)| \leq \|P^n \delta_e\|_2 \|\delta_x\|_2 \leq \rho_d^n.$$

The L^2 estimate is the same calculation:

$$\|v^{(n)}\|_2 = \|P^n \delta_e\|_2 \leq \rho_d^n.$$

□

The L^2 bound is stronger than what simple counting of reduced words would give, because it correctly incorporates all backtracking paths.

6.3 Transfer through the logarithmic window

Let $T = \{A^{\pm 1}, B^{\pm 1}\}$ be the fixed free pair in $\mathrm{SL}_2(\mathbb{Z})$, and let μ_p be the uniform measure on T_p . If every endpoint reached in n steps corresponds injectively to its free-group endpoint, then the finite-quotient distribution equals the free distribution on those endpoints.

Proposition 6.3 (Initial point spreading modulo p). *There are constants $c_0, \kappa_0 > 0$ such that for all sufficiently large primes p , with*

$$n_0 = \lfloor c_0 \log p \rfloor,$$

one has

$$\|\mu_p^{(n_0)}\|_\infty \leq p^{-\kappa_0}, \quad \|\mu_p^{(n_0)}\|_2 \leq p^{-\kappa_0}.$$

The same bounds hold, with possibly a smaller exponent, throughout a fixed logarithmic interval. For example, they hold for every n between $c_0 \log p$ and $2c_0 \log p$, provided this interval is contained in the injectivity range.

Proof. Choose $c_0 > 0$ so small that the free ball $B(n_0)$ lies inside the logarithmic injectivity radius from [Theorem 5.3](#); later we shrink c_0 further so that the much larger ball $B(64n_0)$ also lies inside that radius. Then the map from free endpoints of length at most n_0 to G_p is injective. Consequently $\mu_p^{(n_0)}$ is obtained from the free distribution by relabeling its support. By [Proposition 6.2](#),

$$\|\mu_p^{(n_0)}\|_\infty, \|\mu_p^{(n_0)}\|_2 \leq \rho_4^{n_0}.$$

Since $0 < \rho_4 < 1$,

$$\rho_4^{n_0} \leq Cp^{c_0 \log \rho_4} = Cp^{-\kappa_0}$$

with $\kappa_0 = -c_0 \log \rho_4 > 0$. Shrinking κ_0 absorbs the fixed constant C and handles the whole stated interval. □

6.4 Why point spreading is not subgroup escape

A subgroup $H < G_p$ can have cardinality on the order of p^2 , while $|G_p| \asymp p^3$. A pointwise bound $\mu(x) \leq p^{-\kappa}$ gives only

$$\mu(H) \leq |H|p^{-\kappa},$$

which is useless when $\kappa < 2$. The free-walk exponent obtained at a short logarithmic time is much smaller than 2. Thus subgroup escape requires additional structure: many free words lying in the same subgroup coset must satisfy algebraic relations. Volume VIII turns those relations into a polynomial word-count bound.

6.5 Persistence of the atom bound

Lemma 6.4 (Convolution cannot increase L^∞). *For probability measures ν, σ on a finite group,*

$$\|\nu * \sigma\|_\infty \leq \min\{\|\nu\|_\infty, \|\sigma\|_\infty\}.$$

Proof. For every x ,

$$(\nu * \sigma)(x) = \sum_y \sigma(y)\nu(xy^{-1}) \leq \|\nu\|_\infty \sum_y \sigma(y) = \|\nu\|_\infty.$$

Interchanging the factors after the corresponding left/right rewrite gives the other bound. $\square$

Thus once logarithmic-time spreading has been achieved, all later convolution powers remain pointwise spread. The analogous persistence for subgroup-coset escape will also hold, after allowing conjugate subgroups.

Checkpoint 6.5. At time $n_0 \asymp \log p$ we have a genuine power saving in every atom and in L^2 . This is the first quantitative input to the flattening phase. The missing input is a power saving on every proper subgroup coset; that is the first theorem of Volume VIII.

Chapter 7

Volume VII Dependency Map and Handoff to Subgroup Escape

7.1 The chain now proved

The mathematical output of this volume can be compressed into the following theorem, but each arrow has been proved separately in the preceding chapters.

Theorem 7.1 (Arithmetic pre-collision package). *Let $S \subset \mathrm{SL}_2(\mathbb{Z})$ be finite and symmetric, and suppose $\Gamma = \langle S \rangle$ is non-elementary. Then there are constants $p_0, L, c, \kappa > 0$ and elements $A, B \in \Gamma$ such that for every prime $p > p_0$:*

1. S_p generates $G_p = \mathrm{SL}_2(\mathbb{F}_p)$;
2. $T = \{A^{\pm 1}, B^{\pm 1}\}$ freely generates F_2 , each $t \in T$ has S -word length at most L , and T_p generates G_p after enlarging p_0 ;
3. reduction modulo p is injective on the T -ball of radius $c \log p$;
4. for some $n_0 \asymp \log p$ inside that radius, the T_p -walk satisfies

$$\|\mu_p^{(n_0)}\|_\infty + \|\mu_p^{(n_0)}\|_2 \ll p^{-\kappa}.$$

Proof. Item (1) is [Theorem 4.1](#). The free-pair construction in [Proposition 2.4](#) gives item (2) except for finite-quotient generation. The subgroup $\langle A, B \rangle$ is itself non-elementary, so a second application of [Theorem 4.1](#) shows that T_p generates G_p for every sufficiently large p ; enlarge p_0 accordingly. Item (3) is [Theorem 5.3](#), and item (4) is [Proposition 6.3](#). The constants are chosen in that order, shrinking c whenever a later fixed multiple of the injectivity window is required. $\square$

7.2 Proof dependencies by theorem

The only non-basic statements imported into Volume VII are the following earlier-volume results.

Imported result	Logical use in Volume VII
Volume IV proper-subgroup dichotomy	Produces a fixed law for every proper subgroup of $\mathrm{SL}_2(\mathbb{F}_p)$, closing rank-one strong approximation.
Volume V algebraic compo- nents/dimension	Shows a proper algebraic subgroup has smaller-dimensional identity component and allows Zariski-density arguments.
Volume I bounded-word spectral comparison (reproved here for convenience)	Transfers the final gap from the free pair back to the original generating set.

Everything else that bears weight in the volume—the fixed-law construction, lifting a law from infinitely many primes, the rank-one Zariski-density argument, ping–pong, height growth, logarithmic injectivity, the regular-tree operator bound, and the initial spreading estimate—has been proved in the present volume.

7.3 Results deliberately not used as black boxes

The following named results are historically relevant but not proof dependencies here.

- The general strong-approximation theorem of Matthews–Vaserstein–Weisfeiler [MVW84] is not invoked to prove [Theorem 4.1](#).
- The Tits alternative is not invoked to obtain a free subgroup; the project starts from the operational non-elementary condition and proves ping–pong directly.
- Kesten’s theorem is not invoked for the free-walk norm; [Theorem 6.1](#) gives the tree calculation.

7.4 The next interface: subgroup-coset non-concentration

At the end of Volume VII we know that the logarithmic-time measure $\nu_p = \mu_p^{(n_0)}$ has small atoms. To invoke the weighted inverse theorem of Volume II safely, we need

$$\sup_{H < G_p} \sup_{\text{proper } x \in G_p} \nu_p(xH) \leq p^{-\kappa_1}$$

for some $\kappa_1 > 0$.

The reason this should be true is already visible. If many short free words landed in xH , the subgroup dichotomy would put many of their differences inside a metabelian subgroup. A metabelian group satisfies

$$[[u_1, u_2], [u_3, u_4]] = 1.$$

The logarithmic injectivity theorem would lift that relation to the free group. Volume VIII proves that a subset of a free ball satisfying this law has only polynomial cardinality. Since the free-walk mass of each endpoint is exponentially small, polynomially many endpoints carry only a power of p of total mass.

Warning 7.2 (Why the next proof needs spare girth). If the original random-walk words have length n , their pairwise differences have length at most $2n$, and a commutator of commutators in four such differences has length at most a fixed multiple of n . Thus subgroup escape consumes more of the injectivity radius than point spreading. The constant c_0 in [Proposition 6.3](#) is henceforth chosen sufficiently small that all these fixed multiples still lie below $c_{\text{inj}} \log p$.

7.5 Historical provenance and scope

The arithmetic expansion architecture follows Bourgain–Gamburd [[BG08](#)]. The present self-contained organization differs in one logical respect from many compressed accounts: the prime-field strong-approximation interface has been closed internally using the rank-one subgroup theorem already proved in the series. This makes Arc III self-contained relative to Arcs I–II for the stated $\text{SL}_2(\mathbb{Z})$ prime-quotient theorem.

Checkpoint 7.3. The output needed by Volume VIII is now completely explicit: logarithmic-time point spreading has been proved, while subgroup-coset non-concentration remains a distinct mathematical task. Volume VIII begins with the displayed coset estimate and proves it from the fixed proper-subgroup law, logarithmic injectivity, and free-group centralizer geometry.

Bibliography

- [BG08] Jean Bourgain and Alex Gamburd. “Uniform expansion bounds for Cayley graphs of $SL_2(\mathbb{F}_p)$ ”. In: *Annals of Mathematics* 167 (2008), pp. 625–642.
- [Hel08] Harald A. Helfgott. “Growth and generation in $SL_2(\mathbb{Z}/p\mathbb{Z})$ ”. In: *Annals of Mathematics* 167 (2008), pp. 601–623.
- [MVW84] C. R. Matthews, L. N. Vaserstein, and B. Weisfeiler. “Congruence properties of Zariski-dense subgroups. I”. In: *Proceedings of the London Mathematical Society* 48.3 (1984), pp. 514–532.

The Bourgain–Gamburd Method
Volume VIII: Subgroup Escape and the Arithmetic
 L^2 -Flattening Engine

Preface

Volume VII produced two logarithmic-time facts: reduction is injective on a free ball, and the walk has exponentially small atoms there. Neither statement prevents concentration on a large proper subgroup. Volume VIII proves the missing non-concentration statement and then connects it to the inverse theorem of Volume II and the product theorem of Volume IV.

The subgroup-escape proof is deliberately separated from initial spreading. Proper subgroups of $\mathrm{SL}_2(\mathbb{F}_p)$ are either controlled by a uniformly metabelian subgroup of bounded index or belong to a bounded exceptional family. A large collection of short free words landing in one metabelian coset would satisfy the second-derived law. Centralizers in a free group are cyclic, and that law forces only polynomially many short words. Polynomial word counts cannot compete with the exponentially small mass of a free random walk. This gives a power saving in p uniformly over all proper subgroup cosets.

Once escape is available, failed L^2 -flattening has nowhere to hide. Volume II turns failed flattening into a coset carrying substantial mass of an approximate group. If the approximate group generates a proper subgroup, escape contradicts that mass. If it generates the full group, Volume IV's Helfgott growth theorem contradicts small tripling unless the set is already near the whole group, which the current L^2 scale excludes. A bounded number of doublings therefore brings the walk to the quasirandom threshold in $O(\log p)$ total time. **Series position.** This volume uses Volumes II, IV, and VII. Its main outputs are uniform subgroup escape and the arithmetic L^2 -flattening engine.

Quantifier and asymptotic conventions. Unless a statement explicitly says otherwise, every constant occurring in a quantitative theorem is chosen in the order in which it is quantified in that theorem. Thus a phrase such as “for all sufficiently small δ ” means: there is a displayed or locally quantified threshold $\delta_0 > 0$, depending only on the parameters listed before that phrase, such that the assertion holds for every $0 < \delta < \delta_0$. The phrases “sufficiently large N ” and “sufficiently large p ” have the analogous meaning.

If $X \ll_{\theta} Y$ or $X = O_{\theta}(Y)$, the implicit constant depends only on θ and on the fixed ambient data of the current volume; it is independent of the variables quantified after θ . The words *absolute*, *uniform*, and *bounded* never introduce an unlisted dependence. Whenever an inequality between exponents or scales is essential to the argument, that inequality is written explicitly in the proof; these conventions are not a substitute for such a calculation.

Contents

Preface	iii
1 Proper Subgroups of $\mathrm{SL}_2(\mathbb{F}_p)$: The Exact Escape Input	1
1.1 The one structural theorem we reuse	1
1.2 Why metabelianity produces a bounded-length law	1
1.3 Cosets and difference words	2
1.4 Bounded-index reduction	3
1.5 What remains genuinely combinatorial	3
2 Axes and Centralizers in a Free Group	5
2.1 The Cayley tree	5
2.2 Centralizers preserve the axis	6
2.3 Quantitative intersection with a word ball	6
2.4 Fibers of a commutator map	7
2.5 Why this is the right rigidity	8
3 A Polynomial Count for Sets Satisfying the Metabelian Law	9
3.1 The law set theorem	9
3.2 Cosets of metabelian reductions	10
3.3 Bounded subgroups	11
3.4 The combinatorial escape principle	11
4 Uniform Escape from Every Proper Subgroup Coset	13
4.1 The logarithmic-time measure	13
4.2 Why the estimate persists	15
4.3 Atom escape also persists	16
4.4 A useful normalization in terms of $N = G_p $	16
4.5 Worked logarithmic-time conversion	16
5 The Weighted Inverse Theorem: Exact Interface from Volume II	17

5.1	Collision scale and the structured alternative	17
5.2	Three possible locations of the approximate group	18
5.3	Approximate groups have controlled tripling	19
5.4	The lower endpoint from atom bounds	19
5.5	The upper endpoint from the current L^2 scale	20
5.6	Parameter order	20
6	The Concrete Arithmetic L^2-Flattening Proposition	21
6.1	Statement	21
6.2	Assume flattening fails	22
6.3	Why every hypothesis has a distinct role	23
6.4	Nonsymmetric intermediate measures	23
6.5	A multiplicative form	24
7	Iterating Flattening to the Quasirandom Threshold	25
7.1	Starting point	25
7.2	Dyadic powers	25
7.3	Subtracting the uniform measure	26
7.4	Why logarithmic total time is indispensable	27
7.5	Ordinary versus absolute gap	27
8	Dependencies and the Mesoscopic Handoff	29
8.1	Main output	29
8.2	Earlier results used here	30
8.3	No hidden approximate-group classification	30
8.4	No hidden subgroup-escape theorem	31
8.5	Handoff to Volume IX	31

Chapter 1

Proper Subgroups of $\mathrm{SL}_2(\mathbb{F}_p)$: The Exact Escape Input

1.1 The one structural theorem we reuse

Let

$$G_p = \mathrm{SL}_2(\mathbb{F}_p), \quad p \geq 5.$$

Volume IV proved the subgroup geometry needed for rank-one growth directly from the projective action. For subgroup escape we require only the following compressed consequence.

Proposition 1.1 (Uniform metabelian-or-bounded dichotomy; Volume IV). *There exist absolute constants J and p_0 such that for every prime $p > p_0$ and every proper subgroup $H < G_p$, either*

1. $|H| \leq J$, or
2. *there exists a metabelian subgroup $L \leq H$ with $[H : L] \leq J$.*

No stronger classification statement will be used in the escape proof.

1.2 Why metabelianity produces a bounded-length law

Write

$$\mathfrak{w}(x_1, x_2, x_3, x_4) = [[x_1, x_2], [x_3, x_4]].$$

Lemma 1.2 (The second-derived law). *A group L is metabelian if and only if*

$$\mathfrak{w}(\ell_1, \ell_2, \ell_3, \ell_4) = e$$

for all $\ell_1, \dots, \ell_4 \in L$.

Proof. The first commutators $[\ell_1, \ell_2]$ and $[\ell_3, \ell_4]$ lie in the derived subgroup L' . The displayed identity for every quadruple says precisely that every two elements of L' commute, i.e. L' is abelian. This is equivalent to $L'' = 1$. $\square$

The word $\mathfrak{w}$ has fixed length. If elements represented by short free words reduce into a metabelian subgroup, substituting them into $\mathfrak{w}$ yields a relation modulo p whose length remains a fixed multiple of the original word length. This fixed multiplicative loss is why the logarithmic injectivity reserve from Volume VII suffices.

1.3 Cosets and difference words

Suppose $\pi : F_2 \rightarrow G_p$ is the reduction homomorphism arising from the fixed free pair of Volume VII. Let Σ be a collection of free words of length at most n whose reductions all lie in one right coset xL .

Choose $\sigma_0 \in \Sigma$ and define

$$\Delta = \{\sigma_0^{-1}\sigma : \sigma \in \Sigma\}.$$

Then every element of Δ has free length at most $2n$, and

$$\pi(\Delta) \subset L.$$

Lemma 1.3 (A coset becomes a law set). *Assume reduction is injective on the free ball of radius $32n$. If $L < G_p$ is metabelian and $\pi(\Delta) \subset L$, then*

$$[[u_1, u_2], [u_3, u_4]] = e$$

in the free group for every $u_1, u_2, u_3, u_4 \in \Delta$.

Proof. The metabelian law gives

$$\pi([[u_1, u_2], [u_3, u_4]]) = e.$$

Each u_i has length at most $2n$. A commutator $[u_i, u_j]$ has length at most $8n$, and the outer commutator has length at most $32n$. By injectivity on that ball, an element reducing to the identity must itself be the identity in F_2 . $\square$

The constant 32 is intentionally generous. Any fixed constant would be harmless after shrinking the logarithmic time in Volume VII.

1.4 Bounded-index reduction

If the walk concentrates on a coset of a proper H , the structural theorem gives either a bounded group or a bounded-index metabelian subgroup. In the second case we must pass from the H -coset to one L -coset without losing too much mass.

Lemma 1.4 (Pigeonholing a bounded-index coset). *If $L \leq H$ has index at most J , then every right coset xH is a disjoint union of at most J right cosets of L . Hence for every probability measure ν ,*

$$\nu(xH) \leq J \max_y \nu(yL).$$

Proof. Choose representatives $h_1, \dots, h_m$ for H/L , $m \leq J$. Then

$$xH = \bigsqcup_{i=1}^m xh_iL.$$

Summing the masses proves the inequality. □

Thus it is enough to bound the mass of a metabelian coset plus the mass of cosets of bounded subgroups.

1.5 What remains genuinely combinatorial

Pointwise spreading says each endpoint has exponentially small mass, but a coset can contain many endpoints. The missing statement is:

A subset of a free ball satisfying the metabelian law has only polynomially many elements.

The proof uses the geometry of centralizers in a free group. We give that geometry from first principles in the next chapter and derive the polynomial law count in the chapter after that.

Checkpoint 1.5. All finite-group structure needed for subgroup escape has now been reduced to a single four-variable law and a bounded exceptional case. No classification theorem will appear again in the free-group counting argument.

Chapter 2

Axes and Centralizers in a Free Group

2.1 The Cayley tree

Let $F = F(A, B)$ with the standard word metric $|\cdot|$. Its Cayley graph $\mathcal{T}$ is a 4-regular tree. Left multiplication by F acts freely by graph isometries.

For $g \in F \setminus \{e\}$, write a conjugate of g in cyclically reduced form:

$$g = cuc^{-1},$$

where u is cyclically reduced and nontrivial. The bi-infinite path

$$\cdots, cu^{-1}, c, cu, cu^2, \cdots$$

with the geodesic segments between consecutive points is a g -invariant line in $\mathcal{T}$.

Lemma 2.1 (Existence and uniqueness of the axis). *Every nontrivial $g \in F$ preserves a unique bi-infinite geodesic $Ax(g)$, its axis, and acts on it by a nonzero translation.*

Proof. For a cyclically reduced u , concatenating copies of its reduced word causes no cancellation at the joins, so the vertices u^n lie on one bi-infinite geodesic and u translates that line by $|u|$. Conjugating gives an invariant line for g .

For uniqueness, let L be any g -invariant geodesic. An isometry of a tree with positive translation length has the property that the geodesic segment from x to gx contains a central segment of length equal to the translation length, and the translates of that central segment concatenate into the unique invariant line. Concretely in a free Cayley tree, take x and reduce the word $x^{-1}gx$ to cuc^{-1} with u cyclically reduced; repeated multiplication cancels only the $c^{-1}c$ pairs and forces the same bi-infinite periodic line. Hence every invariant line is the one constructed above. $\square$

2.2 Centralizers preserve the axis

Lemma 2.2 (A centralizer preserves the axis). *If $hgh^{-1} = g$ with $g \neq e$, then*

$$hAx(g) = Ax(g).$$

Proof. The image $hAx(g)$ is invariant under

$$hgh^{-1} = g.$$

By uniqueness of the axis from Lemma 2.1, it must equal $Ax(g)$. □

An isometry of a simplicial line is either a translation or a reflection. A reflection conjugates a nonzero translation to its inverse, so an element commuting with g cannot act as a reflection on the axis. Thus every centralizer element acts by an integer translation.

Theorem 2.3 (Centralizers in a free group are cyclic). *For every nontrivial $g \in F$, the centralizer $C_F(g)$ is infinite cyclic.*

Proof. By Lemma 2.2, each $h \in C_F(g)$ preserves $L = Ax(g)$ and acts on it by an orientation-preserving translation. Choose an orientation and identify the vertex set of L with an arithmetic subset of $\mathbb{Z}$. Translation distance defines a homomorphism

$$\tau : C_F(g) \rightarrow \mathbb{Z}.$$

The action of F on vertices of its Cayley tree is free, so an element acting trivially on L fixes a vertex and must be the identity. Hence τ is injective. Its image is a nonzero subgroup $m\mathbb{Z}$ of $\mathbb{Z}$, generated by some least positive translation. Let $r \in C_F(g)$ realize that translation. Then every $h \in C_F(g)$ has $\tau(h) = k\tau(r)$ and injectivity gives $h = r^k$. Thus $C_F(g) = \langle r \rangle$. □

The element r is the primitive root of g : there is a nonzero integer m with $g = r^m$.

2.3 Quantitative intersection with a word ball

We need not only cyclicity but a uniform polynomial bound on the number of elements of a cyclic coset inside a word ball.

Lemma 2.4 (Linear growth of cyclic cosets). *There is an absolute constant C such that for every nontrivial cyclic subgroup $C_0 \leq F$, every $x \in F$, and every $R \geq 1$,*

$$|xC_0 \cap B_F(R)| \leq C(R + 1).$$

The same bound holds for $C_0x \cap B_F(R)$.

Proof. Write $C_0 = \langle r \rangle$ with r not a proper power, and conjugate $r = cuc^{-1}$ with u cyclically reduced. Then

$$r^n = cu^n c^{-1},$$

so after free reduction

$$|r^n| = 2|c| + |n||u|$$

except that a fixed amount of cancellation can occur when multiplying by the external translate x . The triangle inequality gives the simpler bound sufficient here:

$$|r^{n-m}| \leq |xr^n| + |xr^m| \leq 2R$$

whenever both xr^n and xr^m lie in $B_F(R)$. Since $|r^q| \geq |q|$ for a primitive nontrivial cyclically reduced core (more generally $|r^q| \geq |q|\ell$ for translation length $\ell \geq 1$), one obtains $|n - m| \leq 2R$. Thus the set of exponents occurring lies in an interval of at most $2R + 1$ integers. Take $C = 3$. $\square$

Example 2.5 (A cyclic coset inside a free word ball). Let $F = \langle a, b \rangle$ with the standard word metric and consider the coset $b\langle a \rangle$. Its elements are ba^k , and no cancellation occurs, so

$$|ba^k| = 1 + |k|.$$

Consequently

$$b\langle a \rangle \cap B_F(R) = \{ba^k : |k| \leq R - 1\}$$

has $2R - 1$ elements. This is the model case behind Lemma 2.4: a cyclic coset occupies only linearly many points of a ball whose total size grows exponentially. The later subgroup-escape argument repeatedly exploits precisely this mismatch of scales.

2.4 Fibers of a commutator map

Fix $a \in F \setminus \{e\}$ and consider

$$\phi_a(y) = [a, y] = a^{-1}y^{-1}ay.$$

Lemma 2.6 (Equal commutators lie in one centralizer coset). *If $[a, y_1] = [a, y_2]$, then*

$$y_2 y_1^{-1} \in C_F(a).$$

Consequently every fiber of ϕ_a meets a word ball $B_F(R)$ in $O(R)$ points, uniformly in a and in the fiber.

Proof. Equality of commutators gives

$$a^{-1}y_1^{-1}ay_1 = a^{-1}y_2^{-1}ay_2.$$

Cancel a^{-1} and rearrange:

$$y_1^{-1}ay_1 = y_2^{-1}ay_2.$$

Multiplying on the left by y_2 and on the right by y_1^{-1} yields

$$(y_2y_1^{-1})a = a(y_2y_1^{-1}).$$

Thus $y_2y_1^{-1} \in C_F(a)$. A fiber is therefore contained in a left coset of $C_F(a)$, and Lemma 2.4 gives the $O(R)$ bound. $\square$

2.5 Why this is the right rigidity

A free group has exponential ball growth, but every centralizer has only linear growth. The metabelian law will force all pairwise commutators of a large set into one cyclic centralizer. The commutator map then has only $O(R)$ possible values and $O(R)$ points per fiber, producing an $O(R^2)$ bound. This is exactly the polynomial-versus-exponential separation needed for subgroup escape.

Checkpoint 2.7. The only group-theoretic fact about free groups used later is now proved quantitatively: nontrivial centralizers are cyclic and their cosets meet radius- R balls in $O(R)$ points.

Chapter 3

A Polynomial Count for Sets Satisfying the Metabelian Law

3.1 The law set theorem

Let $B(R)$ be the radius- R ball in the free group $F = F(A, B)$.

Theorem 3.1 (Metabelian law sets are polynomially small). *There is an absolute constant C such that if $\Sigma \subset B(R)$ satisfies*

$$[[x_1, x_2], [x_3, x_4]] = e \quad \text{for all } x_1, x_2, x_3, x_4 \in \Sigma,$$

then

$$|\Sigma| \leq C(R + 1)^2.$$

Proof. There are two cases.

Case 1: every two elements of Σ commute. If $\Sigma \subset \{e\}$ there is nothing to prove. Otherwise choose $a \in \Sigma \setminus \{e\}$. Then

$$\Sigma \subset C_F(a).$$

By Theorem 2.3 and Lemma 2.4,

$$|\Sigma| \ll R + 1.$$

Case 2: some pair does not commute. Choose $a, b \in \Sigma$ with

$$c = [a, b] \neq e.$$

For arbitrary $x, y \in \Sigma$, the assumed law applied to (a, b, x, y) gives

$$[c, [x, y]] = e.$$

Thus every commutator $[x, y]$ lies in the cyclic centralizer $C_F(c)$.

Fix $x = a$. The map

$$\phi_a : \Sigma \rightarrow C_F(c), \quad y \mapsto [a, y]$$

has image contained in

$$C_F(c) \cap B(4R),$$

because a commutator of two radius- R elements has length at most $4R$. By [Lemma 2.4](#), the image has $O(R + 1)$ elements.

By [Lemma 2.6](#), each fiber of ϕ_a is contained in a coset of $C_F(a)$ and meets $B(R)$ in $O(R + 1)$ points. Therefore

$$|\Sigma| \leq |\text{im } \phi_a| \cdot \max_z |\phi_a^{-1}(z)| \ll (R + 1)^2.$$

This proves the theorem. $\square$

Remark 3.2 (The exponent is not the point). The quadratic bound is convenient and essentially free from the centralizer argument. Any fixed polynomial bound would suffice for Bourgain–Gamburd because the free random-walk atom is exponentially small in $R \asymp \log p$.

3.2 Cosets of metabelian reductions

We now combine the law count with the lifting lemma from Chapter 1.

Corollary 3.3 (Few short words can land in one metabelian coset). *Let $\pi : F \rightarrow G$ be a homomorphism that is injective on $B(64R)$. Let $L < G$ be metabelian and $x \in G$. Then*

$$|\{w \in B(R) : \pi(w) \in xL\}| \ll (R + 1)^2.$$

Proof. Let Σ be the displayed set. If it is empty, stop. Choose $w_0 \in \Sigma$ and set

$$\Delta = w_0^{-1}\Sigma \subset B(2R).$$

Then $\pi(\Delta) \subset L$. By the metabelian law, every quadruple $u_i \in \Delta$ satisfies

$$\pi([u_1, u_2], [u_3, u_4]) = e.$$

The resulting word has length at most $32R$, hence lies inside the injectivity ball $B(64R)$. It is therefore trivial in the free group. Thus Δ satisfies the hypothesis of [Theorem 3.1](#) with radius $2R$, so

$$|\Sigma| = |\Delta| \ll (R + 1)^2.$$

$\square$

3.3 Bounded subgroups

Lemma 3.4 (Few short words can land in a bounded subgroup coset). *Let $\pi : F \rightarrow G$ be injective on $B(2R)$ and let $H < G$ have $|H| \leq J$. Then for every $x \in G$,*

$$|\{w \in B(R) : \pi(w) \in xH\}| \leq J.$$

Proof. If two words $u, v \in B(R)$ have the same image, then $uv^{-1} \in B(2R)$ maps to the identity, so injectivity gives $u = v$. Thus π is injective on $B(R)$. The coset xH has at most J elements, and at most one word in $B(R)$ maps to each. $\square$

3.4 The combinatorial escape principle

Combining [Proposition 1.1](#), [Corollary 3.3](#), and [Lemma 3.4](#) gives:

Proposition 3.5 (Polynomial short-word occupancy of proper subgroup cosets). *There is an absolute exponent C such that, whenever reduction $\pi_p : F_2 \rightarrow G_p$ is injective on $B(64R)$,*

$$\sup_{H < G_p \text{ proper}} \sup_{x \in G_p} |\{w \in B(R) : \pi_p(w) \in xH\}| \ll_J (R+1)^C.$$

One may take $C = 2$ under the dichotomy stated in [Proposition 1.1](#).

Proof. If H is bounded, use [Lemma 3.4](#). Otherwise choose a metabelian $L \leq H$ with index at most J . Decompose xH into at most J L -cosets and apply [Corollary 3.3](#) to each. $\square$

Checkpoint 3.6. The subgroup problem has been converted into the inequality “polynomially many short endpoints per proper subgroup coset.” The next chapter multiplies this word count by the exponentially small free-walk atom and obtains the required power saving in p .

Chapter 4

Uniform Escape from Every Proper Subgroup Coset

4.1 The logarithmic-time measure

Let $T = \{A^{\pm 1}, B^{\pm 1}\}$ be the free pair fixed in Volume VII and let μ_p be the uniform probability on $T_p \subset G_p$. Choose

$$n_0 = \lfloor c_0 \log p \rfloor$$

with $c_0 > 0$ so small that the injectivity theorem of Volume VII applies on the free ball of radius $64n_0$.

The free-group point estimate gives

$$\max_{w \in F_2} v^{(n_0)}(w) \leq \rho^{n_0}$$

with $\rho = \sqrt{3}/2 < 1$.

Theorem 4.1 (Logarithmic escape from proper subgroup cosets). *There exists $\kappa > 0$ such that for all sufficiently large primes p ,*

$$\sup_{H < G_p \text{ proper}} \sup_{x \in G_p} \mu_p^{(n_0)}(xH) \leq p^{-\kappa}.$$

The exponent depends only on the fixed integral generating data, not on H , x , or p .

Proof. Until time n_0 the reduction map is injective on every free endpoint needed in the law-count argument. Hence the finite-quotient endpoint distribution is obtained from the free random-walk distribution by relabeling endpoints.

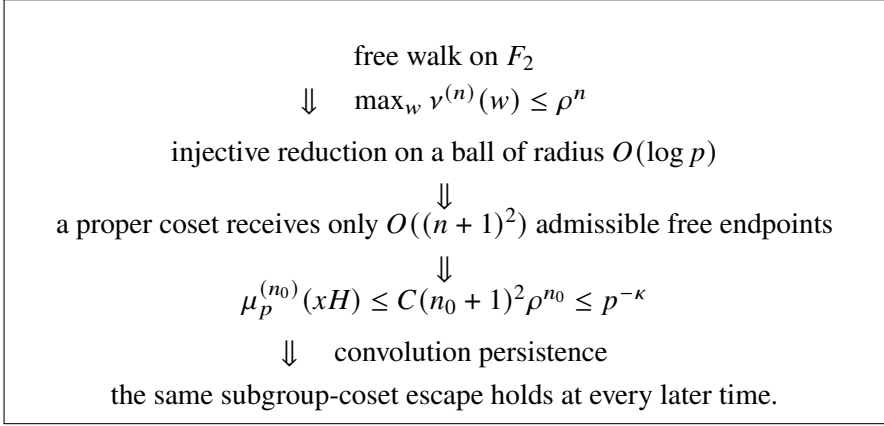


Figure 4.1: The subgroup-escape interface: free-word decay and logarithmic injectivity are converted into a uniform arithmetic non-concentration estimate, which then persists automatically under convolution.

Fix a proper subgroup coset xH . By [Proposition 3.5](#), among all free endpoints of word length at most n_0 at most

$$C(n_0 + 1)^2$$

can reduce into xH . Each such endpoint has free-walk probability at most ρ^{n_0} . Therefore

$$\mu_p^{(n_0)}(xH) \leq C(n_0 + 1)^2 \rho^{n_0}.$$

Since $n_0 \asymp \log p$ and $\rho < 1$,

$$\rho^{n_0} = p^{-c_0 |\log \rho| + o(1)}.$$

A power of $\log p$ is $p^{o(1)}$, so for any

$$0 < \kappa < c_0 |\log \rho|$$

the right-hand side is at most $p^{-\kappa}$ for all sufficiently large p . The constants are uniform over H and x . $\square$

Example 4.2 (Turning the free-walk decay into a visible power of p). With $\rho = \sqrt{3}/2$, put

$$\alpha = c_0 \log \frac{2}{\sqrt{3}} = c_0 |\log \rho| > 0.$$

Because $n_0 = \lfloor c_0 \log p \rfloor \geq c_0 \log p - 1$ and $0 < \rho < 1$,

$$\rho^{n_0} \leq \rho^{-1} p^{-\alpha}.$$

Hence the proof gives the explicit shape

$$\mu_p^{(n_0)}(xH) \leq C\rho^{-1}(c_0 \log p + 1)^2 p^{-\alpha}.$$

Choose, for instance, $\kappa = \alpha/2$. Since $(\log p)^2 = o(p^{\alpha/2})$, the displayed quantity is at most $p^{-\kappa}$ for all sufficiently large p . This calculation isolates exactly where the final exponent is spent: only half of the free-walk power saving is needed to absorb the polynomial factor in $\log p$.

4.2 Why the estimate persists

Later flattening uses measures obtained by additional convolution. We need subgroup escape to survive those extra steps.

Lemma 4.3 (Coset escape persists under convolution). *Let G be finite and suppose a probability measure ν satisfies*

$$\nu(xH) \leq \varepsilon$$

for every proper subgroup $H < G$ and every $x \in G$. Then for every probability measure σ ,

$$(\nu * \sigma)(xH) \leq \varepsilon$$

for every proper H and every x .

Proof. Using right convolution,

$$(\nu * \sigma)(xH) = \sum_{z \in G} \sigma(z) \nu(xHz^{-1}).$$

For each z ,

$$xHz^{-1} = xz^{-1}(zHz^{-1}),$$

which is a right coset of the conjugate subgroup zHz^{-1} . Conjugation preserves properness. Hence every summand is at most ε , and averaging against σ preserves the bound. $\square$

Corollary 4.4 (Persistent arithmetic non-concentration). *For every $m \geq n_0$,*

$$\sup_{H < G_p \text{ proper}} \sup_x \mu_p^{(m)}(xH) \leq p^{-\kappa}.$$

Proof. Write $\mu_p^{(m)} = \mu_p^{(n_0)} * \mu_p^{(m-n_0)}$ and apply Lemma 4.3. $\square$

4.3 Atom escape also persists

By the L^∞ convolution lemma proved in Volume VII,

$$\|\mu_p^{(m)}\|_\infty \leq p^{-\kappa_0} \quad (m \geq n_0)$$

for some possibly different $\kappa_0 > 0$. We therefore have two independent persistent non-concentration statements:

no large atom, no large proper subgroup coset.

The first rules out approximate groups that are too small; the second rules out approximate groups trapped in proper subgroups.

4.4 A useful normalization in terms of $N = |G_p|$

Since $N \asymp p^3$, a power saving $p^{-\kappa}$ is equivalently a power saving $N^{-\kappa/3}$ up to harmless constants. Thus there is $\kappa_G > 0$ such that for every $m \geq n_0$,

$$\|\mu_p^{(m)}\|_\infty \leq N^{-\kappa_G}, \quad \sup_{x, H \text{ proper}} \mu_p^{(m)}(xH) \leq N^{-\kappa_G}.$$

We use this invariant notation in the flattening chapters.

Warning 4.5 (Escape is uniform over conjugates). The persistence proof would fail if the initial estimate were proved only for one chosen list of subgroup representatives. Right convolution converts an H -coset into a coset of zHz^{-1} . This is why [Theorem 4.1](#) takes a supremum over all proper subgroups and all cosets.

Checkpoint 4.6. The random walk now has the precise non-concentration input required by the weighted inverse theory of Volume II, and that input persists throughout every later convolution stage. The expansion proof has entered the mesoscopic regime.

4.5 Worked logarithmic-time conversion

Suppose a proper subgroup coset captures at most ρ^n of the free walk, with $\rho = 3/4$, and choose $n = \lfloor c \log p \rfloor$. Then

$$\rho^n \leq \rho^{-1} p^{-\kappa}, \quad \kappa = -c \log(3/4) > 0.$$

For instance, with $c = 2$, one gets $\kappa = 2 \log(4/3) \approx 0.575$. Thus an exponential-in-word-length escape estimate becomes a uniform power saving in p exactly because the Bourgain–Gamburd machine works at logarithmic time. The harmless prefactor ρ^{-1} is absorbed once p exceeds a fixed threshold.

Chapter 5

The Weighted Inverse Theorem: Exact Interface from Volume II

5.1 Collision scale and the structured alternative

For a probability measure ν on a finite group G , the quantity

$$M(\nu) = \|\nu\|_2^{-2}$$

is its effective support size. Uniform measure on a set of size M has exactly this scale. Convolution cannot increase L^2 :

$$\|\nu * \sigma\|_2 \leq \|\nu\|_2$$

for every probability measure σ .

Volume II proved the following weighted inverse theorem by dyadic decomposition, noncommutative BSG, Ruzsa calculus, and the small-tripling-to-approximate-group conversion.

Theorem 5.1 (Weighted failed-flattening inverse theorem; Volume II). *There is an absolute constant $C \geq 1$ with the following property. Let G be finite, let ν be a symmetric probability measure, and let $K \geq 2$. If*

$$\|\nu * \nu\|_2 \geq K^{-1} \|\nu\|_2,$$

then there exist a K^C -approximate group $A \subset G$ and $x \in G$ such that

$$|A| \leq K^C \|\nu\|_2^{-2}, \tag{5.1}$$

$$\nu(xA) \geq K^{-C}. \tag{5.2}$$

Moreover $A = A^{-1}$ and $e \in A$.

This is the unconditional form needed below. Volume II also proves the controlled-scale refinement: if

$$\|v\|_\infty \leq M\|v\|_2^2,$$

then, after enlarging C if necessary,

$$|A| \geq M^{-2}K^{-C}\|v\|_2^{-2}.$$

The lower bound is *not* automatic for an arbitrary measure: a heavy dyadic level can be much smaller than the global effective-support scale. In the arithmetic walk we will instead derive the lower endpoint directly from the coset-mass conclusion and the independently proved atom bound. Thus no anti-atom hypothesis is silently absorbed into the inverse theorem. These statements are earlier-volume dependencies, not new unproved assertions.

5.2 Three possible locations of the approximate group

Put $G = G_p$ and $N = |G|$. Apply [Theorem 5.1](#) with

$$K = N^\delta,$$

where $\delta > 0$ will be chosen much smaller than the subgroup-escape exponent. Failed flattening by a factor $N^{-\delta}$ produces an $N^{C\delta}$ -approximate group A carrying mass at least $N^{-C\delta}$.

There are only three logically relevant possibilities.

1. $\langle A \rangle$ is proper. Then xA lies in a proper subgroup coset and the persistent escape theorem contradicts [\(5.2\)](#) once $C\delta < \kappa_G$.
2. $\langle A \rangle = G$ and A lies in the medium cardinality range. Then Helfgott growth from Volume IV is incompatible with the small tripling forced by approximate-group control.
3. $\langle A \rangle = G$ and A is already very large. Then the upper bound [\(5.1\)](#) forces $\|v\|_2$ to be close to the uniform scale. If we are still above the target quasirandom threshold, this is impossible after the exponents are chosen in the right order.

The next chapter makes this trichotomy quantitative.

5.3 Approximate groups have controlled tripling

We record the elementary implication used to compare with Helfgott growth.

Lemma 5.2 (Tripling of an approximate group). *If A is a K -approximate group, then*

$$|A^3| \leq K^2|A|.$$

More generally, for each fixed m , $|A^m| \leq K^{m-1}|A|$.

Proof. By definition there is X with $|X| \leq K$ such that $A^2 \subset XA$. Then

$$A^3 = A^2A \subset XAA = XA^2 \subset X^2A,$$

so $|A^3| \leq |X|^2|A| \leq K^2|A|$. For the general power, assume

$$A^m \subset X^{m-1}A.$$

Multiplying by A on the right and using $A^2 \subset XA$ gives

$$A^{m+1} \subset X^{m-1}A^2 \subset X^m A.$$

Starting from $A^2 \subset XA$, induction yields $A^m \subset X^{m-1}A$ for every $m \geq 2$, and therefore

$$|A^m| \leq |X|^{m-1}|A| \leq K^{m-1}|A|.$$

□

Thus an $N^{C\delta}$ -approximate group has

$$|A^3| \leq N^{O(\delta)}|A|.$$

If A generates and has a fixed positive power of N elements but is bounded away from the full scale, Helfgott growth gives a contradictory factor N^c .

5.4 The lower endpoint from atom bounds

The approximate group extracted from inverse theory cannot be microscopic.

Lemma 5.3 (Mass plus small atoms forces size). *If $v(xA) \geq \alpha$ and $\|v\|_\infty \leq \beta$, then*

$$|A| \geq \alpha/\beta.$$

Proof. One has

$$\alpha \leq \sum_{a \in A} \nu(xa) \leq |A|\beta.$$

□

For our walk, $\beta \leq N^{-\kappa_G}$. If $\alpha \geq N^{-C\delta}$, then

$$|A| \geq N^{\kappa_G - C\delta}.$$

Choosing $\delta < \kappa_G/(2C)$ puts A above a fixed power of N .

5.5 The upper endpoint from the current L^2 scale

Suppose we are still above the target uniform neighborhood:

$$\|\nu\|_2 \geq N^{-1/2+\eta}.$$

Then (5.1) gives

$$|A| \leq N^{C\delta} \|\nu\|_2^{-2} \leq N^{1-2\eta+C\delta}.$$

Choosing $C\delta < \eta$ yields

$$|A| \leq N^{1-\eta}.$$

Thus, throughout the flattening phase, the inverse-theorem approximate group is automatically separated from the full scale.

5.6 Parameter order

The safe choice is:

1. obtain $\kappa_G > 0$ from subgroup escape and the atom bound;
2. choose a target exponent $0 < \eta < \kappa_G/10$;
3. take the Helfgott growth exponent corresponding to the size interval

$$N^{\kappa_G/3} \leq |A| \leq N^{1-\eta};$$

4. only then choose $\delta > 0$ so small that every $O(\delta)$ loss from inverse theory and approximate-group multiplication is smaller than all preceding fixed exponents.

This is the quantitative spine of the flattening proposition.

Checkpoint 5.4. Failed flattening has been converted, using only Volume II, into an approximate group whose size and mass are trapped in a range where either persistent subgroup escape or Volume IV product growth must contradict it. The next chapter carries out that contradiction with all exponent inequalities visible.

Chapter 6

The Concrete Arithmetic L^2 -Flattening Proposition

6.1 Statement

Let $N = |G_p|$. A family of probability measures ν_p will be called *arithmetically non-concentrated with exponent κ* if

$$\|\nu_p\|_\infty \leq N^{-\kappa}$$

and

$$\nu_p(xH) \leq N^{-\kappa}$$

for every proper subgroup $H < G_p$ and every $x \in G_p$.

Theorem 6.1 (One-step arithmetic flattening). *For every $\kappa > 0$ and every fixed*

$$0 < \eta < \kappa/10$$

there exist $\delta = \delta(\kappa, \eta) > 0$ and $p_0 = p_0(\kappa, \eta)$ such that the following holds for $p > p_0$. Let ν be a symmetric arithmetically non-concentrated probability measure on $G_p = \mathrm{SL}_2(\mathbb{F}_p)$ with exponent κ . If

$$N^{-1/2+\eta} \leq \|\nu\|_2 \leq N^{-\eta},$$

then

$$\|\nu * \nu\|_2 \leq N^{-\delta} \|\nu\|_2.$$

The upper condition excludes measures concentrated on very few points; in the Bourgain–Gamburd walk it follows from logarithmic-time initial spreading. The lower condition says we have not yet reached the quasirandom threshold.

6.2 Assume flattening fails

Proof of Theorem 6.1. Let C be the absolute exponent in the weighted inverse theorem Theorem 5.1. Suppose

$$\|v * v\|_2 > N^{-\delta} \|v\|_2.$$

Apply Theorem 5.1 with $K = N^\delta$. We obtain an $N^{C\delta}$ -approximate group A and a translate xA such that

$$|A| \leq N^{C\delta} \|v\|_2^{-2}, \tag{6.1}$$

$$v(xA) \geq N^{-C\delta}. \tag{6.2}$$

Enlarge C once so that Lemma 5.2 gives

$$|A^3| \leq N^{C\delta} |A|.$$

Step 1: A cannot generate a proper subgroup. If $H = \langle A \rangle$ is proper, then $xA \subset xH$. Arithmetic non-concentration and (6.2) give

$$N^{-C\delta} \leq v(xA) \leq v(xH) \leq N^{-\kappa}.$$

Choose $\delta < \kappa/(2C)$. For large N this is impossible. Therefore

$$\langle A \rangle = G_p.$$

Step 2: A is not too small. By (6.2) and $\|v\|_\infty \leq N^{-\kappa}$, Lemma 5.3 gives

$$|A| \geq N^{\kappa - C\delta}.$$

After shrinking δ further,

$$|A| \geq N^{\kappa/2}.$$

Step 3: A is not too large. The assumed lower L^2 bound and (6.1) give

$$|A| \leq N^{C\delta} N^{1-2\eta} = N^{1-2\eta+C\delta}.$$

Choose $C\delta < \eta$, so

$$|A| \leq N^{1-\eta}.$$

Step 4: invoke the already-proved product theorem. Volume IV's Helfgott theorem, rewritten in $N \asymp p^3$ notation, says that for the fixed interval

$$N^{\kappa/2} \leq |A| \leq N^{1-\eta}$$

there exists $\varepsilon = \varepsilon(\kappa, \eta) > 0$ such that every generating A satisfies

$$|A^3| \geq |A|^{1+\varepsilon}.$$

Since $|A| \geq N^{\kappa/2}$,

$$|A^3| \geq N^{\varepsilon\kappa/2}|A|.$$

But approximate-group control gives

$$|A^3| \leq N^{C\delta}|A|.$$

Choose finally

$$C\delta < \frac{\varepsilon\kappa}{4}.$$

The two inequalities contradict one another for $N > 1$.

All conditions on δ are compatible because $\kappa, \eta, \varepsilon$ were fixed first. Hence failed flattening is impossible. $\square$

6.3 Why every hypothesis has a distinct role

The proof is short only because earlier volumes opened the difficult black boxes. The four hypotheses play different roles:

- subgroup-coset escape rules out $\langle A \rangle < G$;
- the atom bound gives the lower size endpoint;
- the lower L^2 bound gives the upper size endpoint;
- Helfgott growth destroys the remaining generating approximate group.

Removing any one of these leaves a genuine structured alternative.

6.4 Nonsymmetric intermediate measures

The dyadic iteration below uses convolution powers of a symmetric base measure, hence all powers remain symmetric. In more general applications one may replace ν by $\tilde{\nu} * \nu$, where

$$\tilde{\nu}(g) = \nu(g^{-1}),$$

to recover a symmetric positive operator. Arc III does not need that extra notation.

6.5 A multiplicative form

The conclusion can be written in terms of effective support size:

$$M(\nu * \nu) = \|\nu * \nu\|_2^{-2} \geq N^{2\delta} M(\nu).$$

Thus every successful flattening step multiplies the collision scale by a fixed power of the ambient group size. Only a bounded number of steps can occur before the collision scale reaches $N^{1-O(\eta)}$.

Checkpoint 6.2. The mesoscopic Bourgain–Gamburd engine is now a proved theorem: as long as the measure is neither microscopic nor already close to uniform, one doubling of time earns a fixed power saving in L^2 . No unproved approximate-subgroup classification is being used; the structured alternative is eliminated directly by subgroup escape plus the rank-one product theorem.

Chapter 7

Iterating Flattening to the Quasirandom Threshold

7.1 Starting point

Let μ_p be the uniform probability on the four-element free generating set T_p . From Volume VII and Chapter 4 of this volume, there are $n_0 \asymp \log p$ and $\kappa > 0$ such that

$$\nu_0 := \mu_p^{(n_0)}$$

satisfies

$$\|\nu_0\|_\infty \leq N^{-\kappa}, \quad \sup_{x, H \text{ proper}} \nu_0(xH) \leq N^{-\kappa},$$

and

$$\|\nu_0\|_2 \leq N^{-\alpha}$$

for some $\alpha > 0$. Shrink α if necessary so that the last inequality holds uniformly for large p after replacing powers of p by powers of $N \asymp p^3$.

By convolution persistence, every later power remains arithmetically non-concentrated with the same exponent κ .

7.2 Dyadic powers

Define

$$\nu_j = \nu_0^{*2^j} = \mu_p^{(2^j n_0)}.$$

Then

$$\nu_{j+1} = \nu_j * \nu_j.$$

The L^2 norms decrease with j .

Fix $\eta, \delta > 0$ from Theorem 6.1. If

$$\|v_j\|_2 > N^{-1/2+\eta}$$

and also $\|v_j\|_2 \leq N^{-\eta}$, then

$$\|v_{j+1}\|_2 \leq N^{-\delta} \|v_j\|_2.$$

The upper hypothesis $\|v_j\|_2 \leq N^{-\eta}$ holds from the start provided we choose $\eta < \alpha$, and it persists thereafter.

Theorem 7.1 (Bounded-stage approach to uniformity). *There is an integer J_0 , depending only on the fixed generating set, such that for all sufficiently large p there exists $0 \leq j \leq J_0$ with*

$$\|v_j\|_2 \leq N^{-1/2+\eta}.$$

Consequently this threshold is reached at a time

$$n \leq 2^{J_0} n_0 \ll \log p \asymp \log N.$$

Proof. Suppose the threshold has not been reached by stage j . Repeated application of one-step flattening gives

$$\|v_j\|_2 \leq N^{-\alpha-j\delta}.$$

Choose

$$J_0 > \frac{1/2 - \eta - \alpha}{\delta}$$

when the numerator is positive; if it is nonpositive, take $J_0 = 0$. Then

$$\alpha + J_0\delta > 1/2 - \eta,$$

so the displayed estimate forces

$$\|v_{J_0}\|_2 < N^{-1/2+\eta},$$

contradicting the assumption that the threshold had not yet been reached. The time bound follows because J_0 is independent of p and $n_0 \asymp \log p$. $\square$

7.3 Subtracting the uniform measure

Let $u_G = N^{-1}\mathbf{1}_G$. Since convolution powers have total mass one,

$$\|v_j - u_G\|_2^2 = \|v_j\|_2^2 - \|u_G\|_2^2 = \|v_j\|_2^2 - N^{-1}.$$

Thus at the terminal stage j_* ,

$$N\|v_{j_*} - u_G\|_2^2 \leq N^{2\eta}.$$

This is not yet mixing: $N^{2\eta}$ grows. But quasirandomness will divide this error by a representation dimension $D \gg N^{1/3}$, leaving a negative power of N when $2\eta < 1/3$.

7.4 Why logarithmic total time is indispensable

Suppose a trace argument yields for a nontrivial eigenvalue λ

$$\lambda^{2n} \leq N^{-c}.$$

Then

$$\lambda \leq \exp\left(-\frac{c \log N}{2n}\right).$$

A constant spectral gap requires $n = O(\log N)$. If flattening took $(\log N)^2$ time, the resulting exponent would be only $O(1/\log N)$ and the gap would vanish. The bounded number of dyadic stages in [Theorem 7.1](#) is therefore not cosmetic bookkeeping; it is structurally necessary.

7.5 Ordinary versus absolute gap

The trace method in Volume IX bounds the largest nontrivial eigenvalue from above. A symmetric Cayley graph may still have an eigenvalue near -1 if it is nearly bipartite. This does not obstruct ordinary edge expansion. If total-variation mixing at every time is desired, replace μ by the lazy measure

$$\mu^{\text{lazy}} = \frac{1}{2}\delta_e + \frac{1}{2}\mu.$$

Every eigenvalue λ becomes $(1 + \lambda)/2 \in [0, 1]$, and a constant ordinary gap becomes a constant absolute gap after lazification.

Checkpoint 7.2. At a time $n \ll \log |G_p|$, the walk has reached

$$\|\mu_p^{(n)}\|_2 \leq |G_p|^{-1/2+\eta}$$

for a fixed $\eta > 0$ that can be chosen smaller than one sixth. All analytic work needed for expansion is now complete except the quasirandom trace conversion.

Chapter 8

Dependencies and the Mesoscopic Handoff

8.1 Main output

Theorem 8.1 (Mesoscopic arithmetic package). *For the fixed free pair associated with a non-elementary $S \subset \mathrm{SL}_2(\mathbb{Z})$, there exist constants $\kappa, \eta > 0$, an integer J_0 , and times $n_p \ll \log p$ such that for all sufficiently large primes p :*

1. *every convolution power after the initial logarithmic time satisfies*

$$\sup_{x, H \text{ proper}} \mu_p^{(m)}(xH) \leq |G_p|^{-\kappa};$$

2. *the same powers satisfy a persistent atom bound*

$$\|\mu_p^{(m)}\|_\infty \leq |G_p|^{-\kappa};$$

3. *after at most J_0 dyadic flattening stages,*

$$\|\mu_p^{(n_p)}\|_2 \leq |G_p|^{-1/2+\eta}.$$

For every prescribed fixed $\eta > 0$ with $\eta < \kappa/10$, the one-step theorem supplies a corresponding $\delta(\kappa, \eta) > 0$; hence η may be chosen as small as required (for example below $1/6$), at the cost of changing the fixed number J_0 .

Proof. The subgroup escape statement is [Theorem 4.1](#) plus [Lemma 4.3](#). The atom statement comes from Volume VII plus L^∞ monotonicity. The terminal L^2 statement is [Theorem 7.1](#). Choose the desired $0 < \eta < \kappa/10$ before invoking the one-step theorem. Its corresponding $\delta(\kappa, \eta) > 0$ then gives the same conclusion with a possibly larger, but still p -independent, number of dyadic stages. $\square$

8.2 Earlier results used here

Only three substantial inputs are imported into the proof of Volume VIII:

Earlier result	Use
Volume IV proper-subgroup dichotomy	Reduces every proper subgroup to a bounded case or bounded-index metabelian subgroup.
Volume II weighted inverse theorem	Converts failed L^2 flattening into a massive approximate group with an upper collision-scale bound; the lower size endpoint is supplied separately by the atom estimate.
Volume IV Helfgott product growth	Excludes a generating approximate group in the medium cardinality range.

The free-group centralizer theorem, polynomial metabelian-law count, uniform subgroup escape, persistence under convolution, endpoint size estimates, one-step flattening contradiction, and bounded-stage iteration are proved in this volume. Historically, the inverse-theorem/flattening organization follows the Bourgain–Gamburd method [BG08], with the product-growth input originating in Helfgott’s work [Hel08] and the noncommutative inverse-theory architecture represented by [Tao08].

8.3 No hidden approximate-group classification

A common compressed proof says: “BSG gives an approximate subgroup, which must be close to a proper subgroup.” That sentence hides a serious theorem in general groups. We do not use it. The argument here is simpler and rank-one specific:

1. if $\langle A \rangle$ is proper, then A is literally contained in a proper subgroup and escape applies;
2. if $\langle A \rangle = G_p$, then the already-proved product theorem applies directly to A and contradicts approximate-group tripling in the medium range.

Thus no Breuillard–Green–Tao structure theorem is a dependency of Arc III.

8.4 No hidden subgroup-escape theorem

The escape proof is likewise internal. The only finite-group structural statement is Volume IV's bounded/metabelian dichotomy. The rest consists of:

$$\begin{aligned} \text{metabelian law} &\Rightarrow \text{cyclic free-group centralizers} \\ &\Rightarrow \text{polynomial law-set count} \\ &\Rightarrow \text{power-saving random-walk mass.} \end{aligned}$$

Every arrow is explicit in Chapters 1–4.

8.5 Handoff to Volume IX

Let P_p be convolution by the original or bounded-word generating measure. At time $n_p \leq C \log N$, we now know

$$N \|\mu_p^{(n_p)} - u_G\|_2^2 \leq N^{2\eta}.$$

Volume VI gives a nontrivial representation-dimension lower bound

$$D_p \gg p \gg N^{1/3}.$$

If $\lambda_p < 1$ is a nontrivial eigenvalue of P_p , the corresponding eigenspace has dimension at least D_p , while the trace of $P_p^{2n_p}$ is controlled by the displayed L^2 quantity. Choosing $2\eta < 1/3$ will force

$$\lambda_p^{2n_p} \leq N^{-c},$$

and logarithmic time will turn this into $\lambda_p \leq 1 - \varepsilon$.

Checkpoint 8.2. Volume VIII closes the hardest quantitative bridge of the Bourgain–Gamburd method: logarithmic-time subgroup escape plus failed-flattening inverse theory plus Helfgott growth imply near-uniform L^2 mass in logarithmic total time. Volume IX contains no further approximate-group or subgroup-structure arguments.

Bibliography

- [BG08] Jean Bourgain and Alex Gamburd. “Uniform expansion bounds for Cayley graphs of $SL_2(\mathbb{F}_p)$ ”. In: *Annals of Mathematics* 167 (2008), pp. 625–642.
- [Hel08] Harald A. Helfgott. “Growth and generation in $SL_2(\mathbb{Z}/p\mathbb{Z})$ ”. In: *Annals of Mathematics* 167 (2008), pp. 601–623.
- [Tao08] Terence Tao. “Product set estimates for non-commutative groups”. In: *Combinatorica* 28 (2008), pp. 547–594.

The Bourgain–Gamburd Method
Volume IX: Arithmetic Expansion, Random
Generators, and the Bourgain–Gamburd Theorem

Preface

This volume closes Arc III. Volumes VII and VIII have already placed a convolution power of the generating measure at logarithmic time inside the quasirandom L^2 regime. The remaining step is spectral: a nontrivial eigenvalue of the averaging operator occurs with large multiplicity because every nontrivial representation of $\mathrm{SL}_2(\mathbb{F}_p)$ is large. The trace of an even power is an L^2 return probability, so a power saving in that return probability forces a uniform spectral gap.

We then assemble the proof without compressing interfaces: strong approximation, bounded-word ping-pong, logarithmic injectivity, initial spreading, subgroup escape, inverse theory, Helfgott growth, iterative flattening, and quasirandom multiplicity each enter once and for a different reason. The arithmetic theorem of Bourgain and Gamburd follows. A separate chapter proves the random-generator variant. Algebraic point counting gives logarithmic girth with high probability. The uniform proper-subgroup law from Volume VII then shows that sufficiently large girth already forces generation, matching the original deterministic large-girth architecture; a separate subgroup-family count gives an independent quantitative generation theorem. After the logarithmic tree window is available, the deterministic expansion machine is identical.

The final chapters mark the scope boundary. Prime powers, higher rank, and composite moduli require genuinely new arithmetic or congruence-layer inputs. Arc III proves the rank-one prime-quotient theorem completely and records precisely which interfaces survive beyond it; it does not disguise later super-approximation results as formal corollaries. **Series position.** This volume uses Volumes I and VII–VIII. Its main outputs are the prime-quotient Bourgain–Gamburd spectral-gap theorem and its random-generator variant.

Quantifier and asymptotic conventions. Unless a statement explicitly says otherwise, every constant occurring in a quantitative theorem is chosen in the order in which it is quantified in that theorem. Thus a phrase such as “for all sufficiently small δ ” means: there is a displayed or locally quantified threshold $\delta_0 > 0$, depending only on the parameters listed before that phrase, such that

the assertion holds for every $0 < \delta < \delta_0$. The phrases “sufficiently large N ” and “sufficiently large p ” have the analogous meaning.

If $X \ll_{\theta} Y$ or $X = O_{\theta}(Y)$, the implicit constant depends only on θ and on the fixed ambient data of the current volume; it is independent of the variables quantified after θ . The words *absolute*, *uniform*, and *bounded* never introduce an unlisted dependence. Whenever an inequality between exponents or scales is essential to the argument, that inequality is written explicitly in the proof; these conventions are not a substitute for such a calculation.

Contents

Preface	iii
1 Quasirandomness and the Trace Endgame	1
1.1 The regular representation and the averaging operator	1
1.2 Trace as a return probability	1
1.3 Quasirandom multiplicity	2
1.4 From L^2 flattening to a constant gap	3
1.5 Edge expansion	5
2 The Complete Arithmetic Bourgain–Gamburd Proof	7
2.1 The theorem	7
2.2 Where the main ideas enter	9
2.3 Historical sources and internal dependencies	9
2.4 A mixing corollary for the lazy walk	10
3 Random Generators I: Algebraic Word Maps and Logarithmic Girth	11
3.1 The random model	11
3.2 Word equations are proper algebraic conditions	11
3.3 A uniform integral certificate in growing characteristic	12
3.4 Finite-field point count	13
3.5 Union bound over all short words	13
3.6 Comparison with the sharp random-girth input	14
3.7 What random girth replaces	14
4 Random Generators II: Generation with High Probability	17
4.1 The geometric subgroup families	17
4.2 Borel subgroups	17
4.3 Torus normalizers	18
4.4 Bounded exceptional subgroups	18
4.5 Generation theorem	19

4.6	Independence is not required between good events	20
5	The Random-Generator Bourgain–Gamburd Theorem	21
5.1	A deterministic finite-group criterion	21
5.2	Random expansion	23
5.3	Why the generation interface still matters	23
5.4	Comparison with the arithmetic theorem	23
5.5	Historical note	24
6	Beyond Prime SL_2: What Arc III Proves and What It Does Not	25
6.1	The reusable machine	25
6.2	Extension fields	25
6.3	Higher rank	26
6.4	Prime powers and composite moduli	26
6.5	Strong uniformity in the generators	26
6.6	Compact Lie groups and discretized flattening	27
6.7	The conceptual invariant	27
7	Proof Architecture, Dependencies, and Source Concordance	29
7.1	The three-volume chain	29
7.2	Where the original Bourgain–Gamburd proof enters	29
7.3	The free-group law count versus the historical polynomial bound .	30
7.4	Random girth: exact scope of the internal argument	31
7.5	Dependency map	31
7.6	Ten interfaces that must not be collapsed	32
7.7	Scope boundary	33

Chapter 1

Quasirandomness and the Trace Endgame

1.1 The regular representation and the averaging operator

Let G be finite of order N and let μ be a symmetric probability measure. Define

$$Pf(x) = \sum_{g \in G} \mu(g)f(xg).$$

Then P is self-adjoint and commutes with the left regular action of G .

The constant functions form the trivial eigenspace with eigenvalue 1. We seek a uniform upper bound below 1 for every other eigenvalue.

1.2 Trace as a return probability

Lemma 1.1 (Even trace identity). *For every integer $n \geq 1$,*

$$\mathrm{tr}(P^{2n}) = N\mu^{(2n)}(e) = N\|\mu^{(n)}\|_2^2.$$

Consequently

$$\mathrm{tr}(P^{2n}) - 1 = N\|\mu^{(n)} - u_G\|_2^2,$$

where $u_G = N^{-1}\mathbf{1}_G$.

Proof. In the standard basis $\{\delta_x\}_{x \in G}$, the diagonal matrix coefficient of P^{2n} at x is the probability that a $2n$ -step right random walk starting at x returns to x . By translation invariance this equals $\mu^{(2n)}(e)$ for every x , proving the first equality.

Symmetry gives $\widetilde{\mu^{(n)}} = \mu^{(n)}$, hence

$$\mu^{(2n)}(e) = (\mu^{(n)} * \mu^{(n)})(e) = \sum_g \mu^{(n)}(g) \mu^{(n)}(g^{-1}) = \sum_g \mu^{(n)}(g)^2.$$

This proves the second equality. Finally u_G is orthogonal in $\ell^2(G)$ to $\mu^{(n)} - u_G$, so

$$\|\mu^{(n)}\|_2^2 = N^{-1} + \|\mu^{(n)} - u_G\|_2^2.$$

Subtract the trivial eigenvalue contribution $1 = N \cdot N^{-1}$. □

1.3 Quasirandom multiplicity

Volume VI proved an elementary representation-dimension lower bound for $\mathrm{SL}_n(\mathbb{F}_q)$ by restricting a nontrivial irreducible representation to the abelian unipotent subgroup and using the torus action on nontrivial characters. We use only $n = 2$.

Theorem 1.2 (Quasirandomness of $\mathrm{SL}_2(\mathbb{F}_p)$; Volume VI). *There is $c > 0$ such that every nontrivial complex irreducible representation of $G_p = \mathrm{SL}_2(\mathbb{F}_p)$ has dimension at least*

$$D_p \geq cp.$$

Since $N = |G_p| \asymp p^3$, this implies

$$D_p \gg N^{1/3}.$$

Lemma 1.3 (Nonconstant eigenspaces have quasirandom dimension). *If P commutes with the left regular representation of a finite group and $\lambda \neq 1$ is an eigenvalue, then its eigenspace is a nontrivial left-invariant representation. If every nontrivial irreducible representation has dimension at least D , the multiplicity of λ is at least D .*

Proof. Commutation with all left translations makes the eigenspace E_λ left invariant. If it contained a nonzero invariant vector, that vector would be constant because the left action on G is transitive, and constants are P -eigenvectors with eigenvalue 1, contradicting $\lambda \neq 1$. Thus E_λ contains no trivial subrepresentation. Decompose it into irreducibles; every summand has dimension at least D , so a nonzero eigenspace has dimension at least D . □

$$\begin{aligned}
& \|\mu_p^{(n)}\|_2 \leq N^{-1/2+\eta}, \quad n \leq C \log N \\
& \Downarrow \quad \text{tr}(P^{2n}) - 1 = N \|\mu_p^{(n)} - u\|_2^2 \\
& \text{small total nonconstant spectral mass} \\
& \Downarrow \quad \text{every nontrivial eigenvalue has multiplicity } D_p \gg N^{1/3} \\
& D_p \lambda^{2n} \leq \text{tr}(P^{2n}) - 1 \\
& \Downarrow \quad n = O(\log N) \\
& \lambda \leq e^{-c} < 1.
\end{aligned}$$

Figure 1.1: The trace endgame. Quasirandom multiplicity makes a large nontrivial eigenvalue expensive; logarithmic-time L^2 flattening leaves too little trace budget to pay for it.

1.4 From L^2 flattening to a constant gap

Theorem 1.4 (Logarithmic-time L^2 saving implies spectral gap). *Let $G_p = \text{SL}_2(\mathbb{F}_p)$, $N = |G_p|$, and let μ_p be symmetric and generating. Suppose for some fixed constants $C > 0$ and $0 < \eta < 1/6$ there is an integer*

$$n_p \leq C \log N$$

such that

$$\|\mu_p^{(n_p)}\|_2 \leq N^{-1/2+\eta}.$$

Then there is $\varepsilon = \varepsilon(C, \eta) > 0$ such that every nontrivial eigenvalue λ of the averaging operator satisfies

$$\lambda \leq 1 - \varepsilon$$

for all sufficiently large p .

Proof. Let $\lambda < 1$ be a nontrivial eigenvalue. If $\lambda \leq 0$, the desired upper bound is automatic. Assume $0 < \lambda < 1$. By [Lemma 1.3](#) and [Theorem 1.2](#), its multiplicity is at least $D_p \gg N^{1/3}$. Therefore

$$D_p \lambda^{2n_p} \leq \text{tr}(P^{2n_p}) - 1.$$

By [Lemma 1.1](#),

$$\text{tr}(P^{2n_p}) - 1 = N \left(\|\mu_p^{(n_p)}\|_2^2 - N^{-1} \right) \leq N^{2\eta}.$$

Hence

$$\lambda^{2n_p} \ll N^{-1/3+2\eta}.$$

Choose

$$\theta = \frac{1}{3} - 2\eta > 0.$$

After absorbing the fixed implicit constant, for all sufficiently large N ,

$$\lambda^{2n_p} \leq N^{-\theta/2}.$$

Taking logarithms and using $n_p \leq C \log N$ gives

$$\log \lambda \leq -\frac{\theta \log N}{4n_p} \leq -\frac{\theta}{4C}.$$

Thus

$$\lambda \leq \exp(-\theta/(4C)) =: 1 - \varepsilon < 1.$$

The constant is independent of p . □

Example 1.5 (Turning an L^2 saving into an actual gap). The proof becomes concrete if we insert sample exponents. Suppose $|G_p| = N \asymp p^3$, every nontrivial representation has dimension $D_p \gg N^{1/3}$, and at some time $n_p \leq C \log N$ we know

$$\|\mu_p^{(n_p)}\|_2^2 \leq N^{-1+1/12}.$$

Then the trace argument gives, up to a fixed harmless constant,

$$\lambda^{2n_p} \ll N^{-1/3+1/12} = N^{-1/4}.$$

For all large N we may weaken this to $\lambda^{2n_p} \leq N^{-1/8}$. Since $n_p \leq C \log N$,

$$\log \lambda \leq -\frac{1}{16C}, \quad \lambda \leq e^{-1/(16C)} < 1.$$

The important point is that the logarithm of the group size cancels. A power saving in L^2 reached in logarithmic time becomes a spectral gap bounded away from zero independently of p .

input	power of N	role
$D \gg N^{1/3}$	$+1/3$	minimum multiplicity
$\ \mu^{(n)}\ _2^2 \leq N^{-1+2\eta}$	2η	nonconstant trace budget
$\lambda^{2n} \ll N^{-1/3+2\eta}$	$-(1/3 - 2\eta)$	eigenvalue saving

The sign becomes favorable precisely when $\eta < 1/6$.

Figure 1.2: Exponent ledger for the trace endgame. The quasirandomness exponent must dominate the residual L^2 exponent.

1.5 Edge expansion

For a fixed-degree regular graph, a uniform upper spectral gap $1 - \lambda_2 \geq \varepsilon$ is quantitatively equivalent to a uniform Cheeger expansion bound by the elementary Cheeger inequalities developed in Volume I. Thus [Theorem 1.4](#) completes the analytic part of the expander theorem.

Remark 1.6 (Negative spectrum). The theorem controls ordinary expansion. If an absolute spectral gap is required for mixing at every parity, lazify the walk. This sends λ to $(1 + \lambda)/2$ and removes the negative-spectrum obstruction while preserving a constant positive gap.

Checkpoint 1.7. Once Volume VIII reaches the quasirandom L^2 threshold in logarithmic time, the spectral theorem above turns it into a constant gap. The full arithmetic Bourgain–Gamburd proof is now only an assembly of already proved interfaces.

Chapter 2

The Complete Arithmetic Bourgain–Gamburd Proof

2.1 The theorem

Theorem 2.1 (Bourgain–Gamburd expansion for fixed integral generators). *Let $S \subset \mathrm{SL}_2(\mathbb{Z})$ be finite and symmetric, and suppose $\Gamma = \langle S \rangle$ is non-elementary. Then there are p_0 and $\varepsilon > 0$ such that for every prime $p > p_0$:*

1. S_p generates $G_p = \mathrm{SL}_2(\mathbb{F}_p)$;
2. the averaging operator on $\mathrm{Cay}(G_p, S_p)$ has ordinary spectral gap at least ε .

Hence these Cayley graphs form a uniform expander family.

Proof. We proceed in nine explicit steps.

Step 1: arithmetic reduction. Volume VII proved rank-one strong approximation internally: for all sufficiently large p ,

$$\pi_p(\Gamma) = G_p.$$

Thus the finite quotient on which we work is the full $\mathrm{SL}_2(\mathbb{F}_p)$.

Step 2: expose free geometry by bounded words. Non-elementarity gives two hyperbolic projective elements with disjoint fixed-point sets. The ping–pong proof in Volume VII produces a free pair $A, B \in \Gamma$ after fixed powers. Set

$$T = \{A^{\pm 1}, B^{\pm 1}\}.$$

Every element of T is an S -word of length at most a constant $L = L(S)$. Since A and B are hyperbolic with disjoint projective fixed-point sets, $\langle A, B \rangle$ is itself

non-elementary. Apply the rank-one strong-approximation theorem once more to this subgroup. After enlarging p_0 , the reduced set T_p also generates the full group G_p for every prime under consideration.

Step 3: logarithmic injectivity and initial spreading. Integral matrix height grows exponentially in T -word length. Therefore reduction modulo p is injective on a free ball of radius $c \log p$. The simple random walk on F_2 has ℓ^2 operator norm $\sqrt{3}/2 < 1$. At a sufficiently small logarithmic time

$$n_0 = \lfloor c_0 \log p \rfloor,$$

the quotient walk $\mu_T^{(n_0)}$ satisfies a power-saving atom bound and L^2 bound.

Step 4: escape from proper subgroup cosets. Volume IV's subgroup theorem says every proper subgroup is bounded or has a bounded-index metabelian subgroup. Volume VIII proves that a set of free words satisfying the metabelian second-derived law has only polynomially many elements in a word ball, because centralizers in a free group are cyclic. Multiplying that polynomial count by the exponentially small free-walk atom gives

$$\sup_{H < G_p} \sup_{\text{proper } x} \mu_T^{(n_0)}(xH) \leq |G_p|^{-\kappa}.$$

This escape estimate persists under all later convolution powers.

Step 5: enter the mesoscopic L^2 regime. The atom bound gives $\|\mu_T^{(n_0)}\|_2 \leq N^{-\alpha}$ for some $\alpha > 0$, where $N = |G_p|$. We now dyadically square the measure.

Step 6: failed flattening creates structure. If at a stage ν one had

$$\|\nu * \nu\|_2 > N^{-\delta} \|\nu\|_2,$$

Volume II's fully proved weighted inverse theorem would produce a translate xA carrying mass $N^{-O(\delta)}$, where A is an $N^{O(\delta)}$ -approximate group with

$$|A| \leq N^{O(\delta)} \|\nu\|_2^{-2}.$$

No unconditional lower size is claimed here. The lower endpoint needed in the next step comes instead from the mass on xA together with the persistent atom bound.

Step 7: eliminate the structured alternative. If $\langle A \rangle$ is proper, the mass on xA contradicts persistent subgroup escape. If $\langle A \rangle = G_p$, the atom bound and current L^2 scale place A between two fixed powers of N and away from the full scale. Volume IV's Helfgott product theorem then gives a fixed-power growth factor for A^3 , contradicting approximate-group tripling when δ is chosen sufficiently small. Hence flattening succeeds.

Step 8: iterate only a bounded number of times. Every successful dyadic stage saves a factor $N^{-\delta}$ in L^2 . Starting from $N^{-\alpha}$, a bounded number of stages reaches

$$\|\mu_T^{(n)}\|_2 \leq N^{-1/2+\eta}$$

for any fixed sufficiently small $\eta > 0$. Since the number of stages is bounded and $n_0 \asymp \log p$, the final time still satisfies

$$n \ll \log N.$$

Choose $\eta < 1/6$.

Step 9: quasirandom spectral completion. Volume VI gives $D_p \gg p \gg N^{1/3}$ for the dimension of every nontrivial representation. By [Theorem 1.4](#), the logarithmic-time L^2 estimate implies a uniform spectral gap for the T_p -averaging operator.

Finally, the bounded-word Dirichlet comparison proved in Volume VII transfers the T_p gap to the original S_p averaging operator, losing only a constant depending on L and $|S|$. This yields the desired uniform $\varepsilon > 0$. $\square$

2.2 Where the main ideas enter

The proof can be remembered by four obstructions and four mechanisms:

Obstruction	Mechanism that eliminates it
Arithmetic image is too small	Rank-one strong approximation
Walk remains locally concentrated	Free ping-pong plus logarithmic injectivity
Walk concentrates on a structured subset	BSG inverse theory plus subgroup escape
Near-uniform L^2 does not yet imply a gap	Quasirandom multiplicity plus trace

Product growth is the bridge between the second and third rows: it says that the approximate subgroup supplied by inverse theory cannot generate the full finite group while remaining mesoscopic.

2.3 Historical sources and internal dependencies

The historical theorem is due to Bourgain and Gamburd [\[BG08\]](#). The product-growth input comes from Helfgott’s architecture [\[Hel08\]](#), proved in Volume IV. Strong approximation is historically associated with broader results such as [\[MVW84\]](#), but the rank-one prime case needed here was proved directly in Volume VII. Thus the citations identify origins; they do not replace proof steps in [Theorem 2.1](#).

2.4 A mixing corollary for the lazy walk

Corollary 2.2 (Logarithmic mixing). *Let*

$$P_p^{\text{lazy}} = \frac{1}{2}(I + P_p).$$

There is $C = C(S)$ such that for every initial probability distribution σ on G_p and every

$$n \geq C \log |G_p|,$$

$$\|\sigma(P_p^{\text{lazy}})^n - u_{G_p}\|_{\text{TV}} \leq |G_p|^{-1}.$$

After changing C , any fixed negative power on the right may be obtained.

Proof. The spectral gap from [Theorem 2.1](#) becomes an absolute gap for the lazy operator: every nontrivial eigenvalue has absolute value at most $1 - \varepsilon/2$. On mean-zero functions,

$$\|(P_p^{\text{lazy}})^n f\|_2 \leq (1 - \varepsilon/2)^n \|f\|_2.$$

For a point mass initial state, $\|\delta_x - u\|_2 \leq 1$, and Cauchy–Schwarz gives

$$\|\tau\|_1 \leq N^{1/2} \|\tau\|_2.$$

Thus total variation is at most

$$\frac{1}{2} N^{1/2} (1 - \varepsilon/2)^n.$$

Choosing $n \geq C \log N$ with C large makes this at most N^{-1} . General initial distributions follow by convexity. $\square$

Checkpoint 2.3. The fixed-generator prime-quotient theorem is now complete. Every theorem used in its proof is either proved in Volumes VII–IX or explicitly reused from the self-contained Volumes I–VI.

Chapter 3

Random Generators I: Algebraic Word Maps and Logarithmic Girth

3.1 The random model

Fix a prime p and choose g, h independently and uniformly from

$$G_p = \mathrm{SL}_2(\mathbb{F}_p).$$

Let

$$S(g, h) = \{g, g^{-1}, h, h^{-1}\}.$$

We first show that with probability tending to one, no nontrivial reduced word of length $c \log p$ vanishes on (g, h) .

3.2 Word equations are proper algebraic conditions

Let $w = w(x, y) \in F_2$ be a nontrivial reduced word. The evaluation map

$$w : \mathrm{SL}_2^2 \rightarrow \mathrm{SL}_2$$

is regular because multiplication and inversion on SL_2 are polynomial. If $|w| = \ell$, the degrees of its matrix-entry polynomials are $O(\ell)$: multiplication adds degrees, while inversion in SL_2 is linear in the matrix entries.

Lemma 3.1 (A nontrivial word is not an identity on SL_2). *For every nontrivial $w \in F_2$, the subvariety*

$$Z_w = \{(g, h) \in \mathrm{SL}_2^2 : w(g, h) = I\}$$

is a proper Zariski-closed subset of SL_2^2 .

Proof. Closedness follows from regularity of the word map. If Z_w were all of SL_2^2 , then w would be a group law on SL_2 . But Volume VII proved that $\mathrm{SL}_2(\mathbb{R})$ contains a nonabelian free subgroup, and a nontrivial free word cannot vanish identically there. Thus Z_w is proper. $\square$

3.3 A uniform integral certificate in growing characteristic

The preceding lemma is a characteristic-zero statement. For the union bound below the word w varies with p , so we must also rule out the possibility that its defining equations degenerate modulo the same prime. A fixed integral free pair provides exactly that uniformity.

Choose

$$a_0 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad m_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad b_0 = m_0 a_0 m_0^{-1}.$$

The matrix a_0 is hyperbolic. Its two projective fixed points are the roots of $z^2 - z - 1 = 0$, while those of b_0 are their translates by 1; the two pairs are disjoint. By the ping-pong theorem proved in Volume VII, there is a fixed integer $r \geq 1$ such that

$$A_0 = a_0^r, \quad B_0 = b_0^r$$

freely generate a rank-two subgroup of $\mathrm{SL}_2(\mathbb{Z})$. Fix this pair once and for all.

Lemma 3.2 (Uniform nonidentity modulo p for logarithmic words). *There is a constant $K_0 > 1$ such that, for every nontrivial reduced word $w \in F_2$ of length ℓ and every prime satisfying*

$$p > K_0^\ell + 1,$$

the word map $w : \mathrm{SL}_2^2 \rightarrow \mathrm{SL}_2$ is not identically I over $\mathbb{F}_p$.

Proof. Because A_0, B_0 freely generate, the integral matrix $w(A_0, B_0) - I$ is nonzero. By the elementary height estimate of Volume VII, after fixing A_0, B_0 there is $K_0 > 1$ such that

$$H(w(A_0, B_0)) \leq K_0^\ell.$$

Hence every entry of $w(A_0, B_0) - I$ has absolute value at most $K_0^\ell + 1$, while at least one is a nonzero integer. If $p > K_0^\ell + 1$, that entry cannot vanish modulo p . Thus the reduction of (A_0, B_0) is an $\mathbb{F}_p$ -point at which w is not the identity, proving that the word map is nonconstant from I over $\mathbb{F}_p$. $\square$

3.4 Finite-field point count

Volume V proved the degree-sensitive point-count theorem for varieties whose degree is allowed to vary. Applied to the six-dimensional irreducible variety SL_2^2 , together with Lemma 3.2, it gives the following estimate uniformly in the logarithmic word range.

Proposition 3.3 (Probability of one word relation). *There are constants C_0, C_1 and $K_0 > 1$ such that for every nontrivial reduced word w of length ℓ and every sufficiently large prime with $p > K_0^\ell + 1$,*

$$\mathbb{P}_{g,h \in G_p} [w(g, h) = I] \leq \frac{C_0(1 + \ell)^{C_1}}{p}.$$

Proof. By Lemma 3.2, the equations $w(g, h) = I$ cut out a proper closed subset of SL_2^2 over $\mathbb{F}_p$. At least one matrix entry of $w - I$ is therefore a nonzero regular function in characteristic p . Its degree is $O(\ell)$. Intersecting the fixed-degree six-dimensional variety SL_2^2 with this nonzero equation gives dimension at most 5 and degree polynomially bounded in ℓ by the Bézout/degree bookkeeping of Volume V. The finite-field point-count theorem gives

$$|Z_w(\mathbb{F}_p)| \leq C(1 + \ell)^{C_1} p^5.$$

Since

$$|G_p|^2 \asymp p^6,$$

dividing yields the probability bound. $\square$

3.5 Union bound over all short words

There are exactly $4 \cdot 3^{\ell-1}$ reduced words of length $\ell \geq 1$ in F_2 .

Theorem 3.4 (Random logarithmic girth). *There exists an absolute $c_g > 0$ such that, for independent uniform $g, h \in G_p$,*

$$\mathbb{P} [\exists 1 \neq w \in F_2, |w| \leq c_g \log p, w(g, h) = I] \longrightarrow 0$$

as $p \rightarrow \infty$. More quantitatively, c_g may be chosen so that the probability is $O(p^{-1/3})$.

Proof. Let $L = \lfloor c_g \log p \rfloor$. By [Proposition 3.3](#) and the union bound,

$$\mathbb{P}[\text{a relation of length} \leq L] \leq \frac{C}{p} \sum_{\ell=1}^L 3^\ell (1 + \ell)^{C_1} \leq \frac{C'}{p} 3^L (1 + L)^{C_1}.$$

Since

$$3^L \leq p^{c_g \log 3},$$

choose $c_g > 0$ so small that simultaneously

$$c_g \log 3 < \frac{1}{2} \quad \text{and} \quad c_g \log K_0 < \frac{1}{2}.$$

The second inequality ensures $K_0^\ell + 1 < p$ uniformly for every $\ell \leq L$ once p is large, so [Proposition 3.3](#) applies to every word in the union. The resulting probability is

$$p^{-1/2+o(1)},$$

which is $O(p^{-1/3})$ for large p . □

3.6 Comparison with the sharp random-girth input

Gamburd–Hoory–Shahshahani–Shalev–Virág prove the stronger asymptotic lower bound

$$\text{girth}(\text{Cay}(G_p, S)) \geq \left(\frac{1}{3} - o(1) \right) \log_3 |G_p|$$

for a random four-regular Cayley graph of $\text{SL}_2(\mathbb{F}_p)$; their proof estimates the probability of a fixed word relation directly by projective polynomial zero counting [\[Gam+09\]](#). The expansion argument in this book needs only a fixed positive multiple of $\log p$. We therefore proved the weaker statement above by a slightly different route: a fixed integral free pair supplies a uniform certificate that every word in the growing logarithmic window remains a nonidentity modulo p , after which the degree-sensitive point-count theorem of Volume V and a union bound suffice. No sharp girth constant is claimed here.

3.7 What random girth replaces

For fixed integral generators, logarithmic girth came from height growth and reduction modulo p . For random generators already living in G_p , there is no integral height. The algebraic word-map estimate replaces it. Once logarithmic girth is known, the free-tree initial spreading and the metabelian-law subgroup escape arguments are identical to the deterministic case: they use only the absence of short relations.

Warning 3.5. Random girth does not imply generation. A pair could have large girth inside a large proper subgroup. The next chapter proves separately that a random pair generates G_p with probability tending to one.

Chapter 4

Random Generators II: Generation with High Probability

4.1 The geometric subgroup families

Volume IV proves a rank-one containment statement stronger than the metabelian-or-bounded corollary used in Volume VIII: for all sufficiently large primes, every proper subgroup of $G_p = \mathrm{SL}_2(\mathbb{F}_p)$ is contained in one of the following types:

1. a Borel subgroup, equivalently a point stabilizer in $\mathbb{P}^1(\mathbb{F}_p)$;
2. the normalizer of a split or nonsplit maximal torus;
3. a subgroup of uniformly bounded order.

We now count random pairs lying in these families.

4.2 Borel subgroups

There are $p + 1$ points in $\mathbb{P}^1(\mathbb{F}_p)$ and hence $p + 1$ Borel point stabilizers. In the present group $G_p = \mathrm{SL}_2(\mathbb{F}_p)$ the stabilizer of ∞ consists exactly of the matrices

$$\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix}, \quad a \in \mathbb{F}_p^\times, b \in \mathbb{F}_p,$$

so every Borel has the exact order

$$|B| = p(p - 1).$$

Some subgroup classifications are stated after quotienting by the center to PSL_2 ; here all counting is performed in SL_2 , so this formula fixes the normalization once and for all.

Lemma 4.1 (Random pairs rarely lie in one Borel). *For independent uniform $g, h \in G_p$,*

$$\mathbb{P}[\exists \text{ Borel } B : g, h \in B] = O(p^{-1}).$$

Proof. By the union bound,

$$\mathbb{P}[\exists B : g, h \in B] \leq (p+1) \left(\frac{|B|}{|G_p|} \right)^2.$$

Using $|G_p| = p(p^2 - 1)$ and $|B| = p(p-1)$, the ratio is $1/(p+1)$, so the displayed probability is exactly at most $1/(p+1)$. $\square$

4.3 Torus normalizers

A split torus is determined by its two eigendirections, hence by an unordered pair of distinct points of $\mathbb{P}^1(\mathbb{F}_p)$. There are $O(p^2)$ such tori, and each normalizer has $O(p)$ elements. Volume IV also proves that all nonsplit maximal tori form a bounded number of conjugacy classes and each normalizer has $O(p)$ elements; because $|G_p| \asymp p^3$, there are $O(p^2)$ conjugates.

Lemma 4.2 (Random pairs rarely lie in one torus normalizer). *The probability that g, h lie together in the normalizer of some maximal torus is $O(p^{-2})$.*

Proof. There are $O(p^2)$ normalizers and each has $O(p)$ elements. Hence

$$\mathbb{P}[\exists N(T) : g, h \in N(T)] \ll p^2 \left(\frac{p}{p^3} \right)^2 \ll p^{-2}.$$

All implicit constants are absolute. $\square$

4.4 Bounded exceptional subgroups

Let J be the uniform order bound for the exceptional case and put $M = J!$. If g lies in a subgroup of order at most J , then $g^M = I$.

Lemma 4.3 (Few bounded-order elements). *For fixed $M \geq 1$,*

$$|\{g \in G_p : g^M = I\}| = O_M(p^2).$$

Proof. The equation $g^M = I$ defines a Zariski-closed subset $X_M \subset \mathrm{SL}_2$ of degree bounded in terms of the fixed integer M . For every sufficiently large prime p it is a proper subvariety over $\mathbb{F}_p$: choose $t \in \mathbb{F}_p^\times$ with $t^M \neq 1$ (there are at most M roots of $T^M - 1$, whereas $p - 1 > M$), and then

$$\mathrm{diag}(t, t^{-1})^M \neq I.$$

Hence $\dim X_M \leq 2$. Volume V's degree point-count theorem therefore gives, uniformly for all sufficiently large p ,

$$|X_M(\mathbb{F}_p)| \ll_M p^2.$$

□

Lemma 4.4 (Random pairs rarely lie in bounded subgroups). *The probability that $\langle g, h \rangle$ has order at most J is $O_J(p^{-2})$.*

Proof. If $\langle g, h \rangle$ has order at most J , then both $g^M = h^M = I$. By Lemma 4.3, the number of possible choices for each generator is $O(p^2)$, whereas $|G_p| \asymp p^3$. Independence gives

$$O\left(\frac{p^2}{p^3}\right)^2 = O(p^{-2}).$$

□

4.5 Generation theorem

Theorem 4.5 (A random pair generates $\mathrm{SL}_2(\mathbb{F}_p)$ with high probability). *For independent uniform $g, h \in G_p$,*

$$\mathbb{P}[\langle g, h \rangle \neq G_p] = O(p^{-1}).$$

In particular the generation probability tends to one.

Proof. If $\langle g, h \rangle$ is proper, Volume IV's subgroup containment theorem places it in a Borel, a maximal-torus normalizer, or a bounded subgroup. The three probabilities are respectively $O(p^{-1})$, $O(p^{-2})$, and $O(p^{-2})$ by Lemmas 4.1, 4.2 and 4.4. Sum them. □

4.6 Independence is not required between good events

We will intersect the high-probability generation event with the high-probability logarithmic-girth event from the preceding chapter. No independence between those events is needed:

$$\mathbb{P}[\text{generation and girth}] \geq 1 - \mathbb{P}[\text{generation fails}] - \mathbb{P}[\text{girth fails}] \rightarrow 1.$$

Once both hold, all deterministic constants in the expansion machine are uniform because the degree is fixed and the girth constant is fixed.

Checkpoint 4.6. A random pair has, with probability $1 - o(1)$, exactly the two inputs that the finite-group Bourgain–Gamburd machine needs before flattening: it generates the whole group and its Cayley graph is tree-like to a fixed positive multiple of $\log p$.

Chapter 5

The Random-Generator Bourgain–Gamburd Theorem

5.1 A deterministic finite-group criterion

The preceding volumes actually prove a theorem that no longer remembers where the generators came from. We first close one interface that is sometimes hidden in statements of the large-girth criterion: for $\mathrm{SL}_2(\mathbb{F}_p)$, sufficiently large girth already forces generation.

Lemma 5.1 (A proper subgroup forces a bounded relation). *There are an absolute integer L_0 and p_0 such that, for every prime $p > p_0$ and every $g, h \in G_p = \mathrm{SL}_2(\mathbb{F}_p)$, if $\langle g, h \rangle$ is proper then there is a nontrivial reduced word $U \in F_2$ of length at most L_0 with*

$$U(g, h) = I.$$

The word U may be chosen once and for all, independently of p, g, h .

Proof. Volume VII constructed a fixed nontrivial word

$$W(x_1, \dots, x_5) \in F_5$$

which is a law on every proper subgroup of G_p once p is sufficiently large. We convert it to a two-variable law without losing nontriviality.

In $F_2 = \langle a, b \rangle$, put

$$q_i = a^{i-1} b a^{-(i-1)}, \quad 1 \leq i \leq 5.$$

The reduced-normal-form argument used in Volume VII shows that $q_1, \dots, q_5$ freely generate a free subgroup of rank five. Hence the homomorphism $F_5 \rightarrow F_2$ sending

x_i to q_i is injective. Therefore

$$U(a, b) := W(q_1, \dots, q_5)$$

is a fixed nontrivial word of F_2 . Its length is an absolute constant L_0 .

If $H = \langle g, h \rangle$ is proper, all five elements $q_i(g, h)$ lie in H . Since W is a law on H ,

$$U(g, h) = W(q_1(g, h), \dots, q_5(g, h)) = I.$$

This is the required bounded relation. $\square$

Corollary 5.2 (Logarithmic girth forces generation). *For every fixed $c_* > 0$, if p is sufficiently large and the Cayley graph associated with $\{g^{\pm 1}, h^{\pm 1}\}$ has girth at least $c_* \log p$, then $\langle g, h \rangle = G_p$.*

Proof. For sufficiently large p one has $c_* \log p > L_0$. If the generated subgroup were proper, Lemma 5.1 would give a nontrivial relation of length at most L_0 , contradicting the girth assumption. $\square$

Theorem 5.3 (Deterministic large-girth criterion). *There exist constants $c_* > 0$ and $\varepsilon_* > 0$ with the following property. Let p be sufficiently large and let $g, h \in G_p = \mathrm{SL}_2(\mathbb{F}_p)$. If the Cayley graph with generating set $\{g^{\pm 1}, h^{\pm 1}\}$ has girth at least $c_* \log p$, then it is connected and its ordinary spectral gap is at least ε_* .*

Proof. By Corollary 5.2, after increasing the lower bound on p the girth hypothesis itself gives $\langle g, h \rangle = G_p$. Choose the initial random-walk time $n_0 = c_0 \log p$ with c_0 a sufficiently small fixed fraction of c_* . The girth assumption identifies all free words required by both the point-spreading and metabelian-law arguments with distinct quotient elements whenever the corresponding free word is nontrivial.

The regular-tree estimate therefore gives the same power-saving atom bound as in Volume VII. The proper-subgroup theorem of Volume IV is uniform in the choice of generators, and the free-group law-count theorem of Volume VIII depends only on the free word metric. Hence the same subgroup-coset escape estimate follows with an exponent depending only on c_* .

The weighted inverse theorem of Volume II, Helfgott growth of Volume IV, and quasirandomness of Volume VI are all uniform in p and independent of the origin of g, h . Consequently the one-step flattening proposition and its bounded dyadic iteration give

$$\|\mu^{(n)}\|_2 \leq |G_p|^{-1/2+\eta}$$

for some $n \ll \log |G_p|$ and fixed $\eta < 1/6$. The trace theorem of Chapter 1 then yields a uniform spectral gap $\varepsilon_* > 0$. $\square$

5.2 Random expansion

Theorem 5.4 (Random-generator expansion). *There is an absolute $\varepsilon > 0$ such that, if g, h are independent uniform elements of $\mathrm{SL}_2(\mathbb{F}_p)$, then*

$$\mathbb{P} \left[\mathrm{Cay}(\mathrm{SL}_2(\mathbb{F}_p), \{g^{\pm 1}, h^{\pm 1}\}) \text{ has spectral gap at least } \varepsilon \right] \longrightarrow 1$$

as $p \rightarrow \infty$.

Proof. Choose c_* small enough for Theorem 5.3 and also small enough for the union-bound proof of random logarithmic girth in Theorem 3.4. By Theorem 3.4, the girth hypothesis of the deterministic criterion holds with probability $1 - o(1)$. On that event, Corollary 5.2 already gives generation for all sufficiently large p , and Theorem 5.3 supplies the same fixed gap ε_* . The direct random-generation theorem of the preceding chapter is therefore an independent quantitative check, not an additional logical hypothesis for random expansion. $\square$

5.3 Why the generation interface still matters

For an arbitrary family of finite groups, large girth need not imply that the chosen elements generate the ambient group. What makes rank-one prime quotients special is the uniform proper-subgroup law from Volume VII: every proper subgroup forces one fixed bounded relation, so logarithmic girth rules out proper generation automatically. Thus in the present setting

$$\begin{aligned} \text{logarithmic girth} &\implies \text{generation} + \text{local free geometry} \\ &\implies \text{subgroup escape} + \text{initial spreading} \\ &\implies \text{flattening} \implies \text{gap}. \end{aligned}$$

The direct counting theorem Theorem 4.5 remains useful because it quantifies generation independently of girth and makes clear which subgroup families are being excluded.

5.4 Comparison with the arithmetic theorem

The deterministic arithmetic theorem and the random theorem differ only in how the first two inputs are produced.

Input	Fixed integral generators	Random finite generators
Generation	Rank-one strong approximation	Direct subgroup-family counting
Logarithmic tree window	Integral height + ping-pong	Algebraic word-map point counting
Subgroup escape	Same free law count	Same free law count
Flattening	Same Volume-II inverse theorem	Same Volume-II inverse theorem
Growth	Same Helfgott theorem	Same Helfgott theorem
Spectral endgame	Same quasirandom trace	Same quasirandom trace

This separation is one of the conceptual achievements of the Bourgain–Gamburd method: once local non-concentration has been manufactured, the later analytic engine forgets whether the generators were arithmetic or random.

5.5 Historical note

Both the fixed-arithmetic and random-generator forms belong to the original Bourgain–Gamburd rank-one expansion theorem [BG08]. The purpose of the present chapter is not to claim a new theorem but to expose the exact proof dependencies in a self-contained format.

Checkpoint 5.5. Both principal rank-one Bourgain–Gamburd theorems are now assembled from internal proofs together with the results established in Arcs I–II.

Chapter 6

Beyond Prime SL_2 : What Arc III Proves and What It Does Not

6.1 The reusable machine

The proof just completed can be abstracted into four finite-quotient hypotheses. Let G_i be finite groups with symmetric generating measures μ_i . A Bourgain–Gamburd-style argument works when one has:

1. logarithmic-time initial spreading;
2. power-saving escape from every structured obstruction that can arise from failed flattening;
3. an inverse theorem converting failed L^2 flattening into one of those structured obstructions;
4. quasirandomness strong enough to turn near-uniform L^2 mass at logarithmic time into a constant spectral gap.

Arcs I–III prove this pattern completely for prime quotients of non-elementary subgroups of $\mathrm{SL}_2(\mathbb{Z})$.

6.2 Extension fields

Volume VI proves a type- A product theorem for $\mathrm{SL}_n(\mathbb{F}_q)$ with fixed n and prime powers q . This removes one major growth obstacle beyond prime fields. Nevertheless, a direct repetition of the present rank-one arithmetic proof would need additional work:

- strong approximation for the chosen arithmetic source and residue fields;
- subgroup escape in the presence of subfield subgroups when q is not prime;
- uniform control of the free-word laws associated with all proper subgroup geometries.

The prime-field proof used the absence of proper subfields at several points in earlier arcs, so extension fields are not a notational change.

6.3 Higher rank

The analytic part becomes more portable in higher rank. Volume VI already provides finite-field Larsen–Pink growth for fixed type A and an elementary quasirandomness lower bound. What changes most sharply is subgroup escape. Proper subgroups are no longer captured by a bounded/metabelian dichotomy; parabolic, reductive, subfield, and other algebraic families appear in greater variety.

A self-contained higher-rank treatment would therefore have to replace the single free-group metabelian law by a finite-complexity algebraic escape mechanism. General super-approximation theorems achieve this with much deeper algebraic and arithmetic input. We do not import them here.

6.4 Prime powers and composite moduli

For $\mathrm{SL}_2(\mathbb{Z}/p^k\mathbb{Z})$, congruence kernels create large proper normal subgroups invisible over the residue field. For squarefree or arbitrary composite q , the Chinese remainder decomposition introduces many simultaneous projections. Approximate subgroups may be large because they concentrate in congruence layers rather than because they resemble field-level algebraic subgroups.

This is why expansion modulo arbitrary q is a genuinely later theorem. Work such as Bourgain–Varjú [BV12] supplies the appropriate broader setting, but it is not a formal corollary of the prime proof developed here.

6.5 Strong uniformity in the generators

The fixed-generator theorem allows constants to depend on the integral generating set S . The random-generator theorem achieves an absolute gap with high probability because the logarithmic-girth constant is chosen uniformly. Strong uniform expansion over all generating pairs is a different problem: pairs can have very short relations or degenerate arithmetic properties, and the local tree window can collapse.

Results such as Breuillard–Gamburd [BG10] address stronger uniformity phenomena. Again, they are part of the historical landscape, not hidden dependencies of Arc III.

6.6 Compact Lie groups and discretized flattening

In compact groups such as $SU(2)$, cardinality is replaced by covering number at a small scale δ . Product growth becomes a discretized product theorem, subgroup escape becomes non-concentration near proper closed subgroups, and L^2 flattening is performed after smoothing at scale δ . The architecture survives, but every finite-set quantity acquires a metric-entropy analogue. A self-contained treatment belongs naturally to a later arc.

6.7 The conceptual invariant

Across all these variants, one idea persists:

$\begin{aligned} \text{failure of analytic flattening} &\implies \text{algebraic structure} \\ &\implies \text{arithmetic/dynamical contradiction} \end{aligned}$

In prime SL_2 , the algebraic structure is simple enough to make every arrow explicit. Higher-rank and composite-modulus theories are difficult mainly because the catalogue of structured obstructions becomes richer, not because the basic random-walk inequality changes.

Checkpoint 6.1. Arc III’s scope is exact: fixed non-elementary integral generators modulo sufficiently large primes, and random pairs in $SL_2(\mathbb{F}_p)$. Higher-rank, prime-power, composite-modulus, and compact-group extensions are identified as future dependency trees rather than asserted without proof.

Chapter 7

Proof Architecture, Dependencies, and Source Concordance

7.1 The three-volume chain

Volumes VII–IX separate the prime-field Bourgain–Gamburd argument into three regimes. Volume VII creates a logarithmic free window from arithmetic data; Volume VIII proves subgroup-coset escape and L^2 flattening; Volume IX converts the resulting logarithmic-time L^2 estimate into a spectral gap and then treats random generators.

The logical chain is

non-elementary integral group $\implies$ surjective prime reductions
 $\implies$ bounded-word free pair $\implies$ logarithmic injectivity
 $\implies$ small atoms $\implies$ proper-coset escape
 $\implies$ failed-flattening structure $\implies$ product growth
 $\implies$ near-uniform $L^2 \implies$ quasirandom trace
 $\implies$ uniform spectral gap.

Each arrow has a different hypothesis and a separate proof. In particular, small atoms are not used as a substitute for subgroup escape, and product growth is not used as a substitute for the final representation-theoretic argument.

7.2 Where the original Bourgain–Gamburd proof enters

Bourgain and Gamburd organize the prime-field proof around the same three essential ingredients: high multiplicity of nontrivial representations, an L^2 flattening statement

for measures satisfying point and subgroup non-concentration, and logarithmic free geometry used to verify those non-concentration hypotheses [BG08]. Their deterministic large-girth theorem assumes a logarithmic girth bound and concludes expansion; generation is forced inside the proof by the structure of proper subgroups.

The present treatment follows that architecture but internalizes the individual inputs. The quasirandom multiplicity estimate was proved in Volume VI, the weighted failed-flattening inverse theorem in Volume II, the rank-one product-growth theorem in Volume IV, and the proper-subgroup geometry in Volume IV. Volumes VII–VIII supply the arithmetic free window and the subgroup-law escape argument rather than citing them as packaged lemmas.

One small strengthening made explicit in Chapter 5 of the present volume is worth remembering. Volume VII constructs a fixed nontrivial law on every proper subgroup of $\mathrm{SL}_2(\mathbb{F}_p)$ for sufficiently large p . Embedding its finitely many variables into a fixed free subgroup of F_2 turns that law into one fixed two-generator relation. Hence a pair contained in a proper subgroup has bounded girth. This proves directly, in the internal language of the series, the same “large girth forces generation” interface used in the original proof.

7.3 The free-group law count versus the historical polynomial bound

The subgroup-escape argument in Bourgain–Gamburd uses the fact that a set of short free words satisfying the second-derived metabelian law has only polynomial cardinality. Their Proposition 8 gives a sixth-power bound in the word radius [BG08]. Volume VIII proves a sharper quadratic bound in the special form needed here:

$$|\Sigma| \ll (R + 1)^2.$$

No stronger expansion conclusion is extracted from the better exponent. The reason the shorter proof works is that free-group centralizers are cyclic: after one noncommuting pair is fixed, all relevant commutators lie in a single cyclic centralizer, and every commutator fiber lies in a coset of another cyclic centralizer. Any fixed polynomial bound would already be enough because the free-walk atom decays exponentially in $R \asymp \log p$.

7.4 Random girth: exact scope of the internal argument

For random generators, Gamburd–Hoory–Shahshahani–Shalev–Virág prove the stronger asymptotic estimate

$$\text{girth} \geq \left(\frac{1}{3} - o(1) \right) \log_3 |\text{SL}_2(\mathbb{F}_p)|$$

with probability tending to one [Gam+09]. Their proof bounds the probability of a fixed word relation by a degree-sensitive polynomial zero count and then uses a union bound.

Chapter 3 proves only the strength required for expansion: there exists an absolute $c_g > 0$ for which the girth is at least $c_g \log p$ with high probability. The proof uses a fixed integral free pair as a characteristic- p nonidentity certificate for every word in the growing logarithmic window, followed by the point-count theorem of Volume V. This deliberately sacrifices the sharp $1/3$ constant; no later theorem uses it.

7.5 Dependency map

The non-basic imported results in this three-volume arc are exactly the following earlier-volume theorems.

Earlier result	Use in Volumes VII–IX
Volume II weighted inverse theorem	Failed L^2 flattening produces a massive approximate group with the required upper collision-scale control.
Volume IV proper-subgroup theorem	Gives both the uniform proper-subgroup law and the subgroup families used in the direct random-generation count.
Volume IV rank-one product growth	Excludes a generating mesoscopic approximate group after failed flattening.
Volume V algebraic dimension/degree and finite-field point counting	Controls algebraic word-relation loci and bounded-order loci for random generators.
Volume VI quasirandomness of $\text{SL}_2(\mathbb{F}_p)$	Gives multiplicity $\gg p$ for every nontrivial spectral block in the trace endgame.

The rank-one strong-approximation theorem, ping–pong free pair, logarithmic injectivity, regular-tree spreading, free-group centralizer geometry, polynomial

metabelian-law count, subgroup escape, bounded-stage flattening, trace endgame, random girth, and random generation are proved inside Volumes VII–IX.

7.6 Ten interfaces that must not be collapsed

The proof is easiest to misuse when two adjacent mechanisms are treated as if they were the same statement. The following distinctions are therefore part of the mathematical architecture.

1. A fixed law for proper finite subgroups must be independent of p before it can be lifted to characteristic zero.
2. The law used in the strong-approximation contradiction must be a nontrivial free word; this is checked by reduced normal form.
3. Logarithmic injectivity must leave room for the fixed multiple of the radius consumed by differences and iterated commutators.
4. Pointwise spreading does not imply subgroup escape; the polynomial free-law count is a separate argument.
5. Subgroup escape has to be uniform over all conjugates in order to persist under convolution.
6. The weighted inverse theorem produces an approximate group, not a subgroup.
7. The lower cardinality endpoint for that approximate group comes from captured mass plus the independent atom bound; it is not silently included in the unconditional inverse theorem.
8. The target L^2 exponent is fixed before the one-step flattening gain is chosen, so the bounded dyadic iteration really reaches the quasirandom threshold.
9. Near-uniform L^2 mass is converted to a gap by multiplicity and an even trace identity; it is not itself a spectral-gap statement.
10. In the random model, the word relation must remain a proper algebraic condition in the same characteristic in which the word is evaluated. The integral free-pair certificate supplies this uniformity in the logarithmic range.

7.7 Scope boundary

The result proved here is the rank-one prime-field theorem: fixed non-elementary integral generators modulo all sufficiently large primes, together with the random-pair theorem in $\mathrm{SL}_2(\mathbb{F}_p)$. Extension fields, higher rank, prime powers, arbitrary composite moduli, and compact Lie groups require additional obstruction catalogues or congruence-layer arguments and are treated only in later volumes.

Checkpoint 7.1. The reusable output of Arc III is therefore not merely a theorem statement but a proof machine with explicit interfaces:

$\begin{aligned} &\text{free logarithmic window} + \text{subgroup escape} \\ &+ \text{flattening} + \text{quasirandomness} \implies \text{spectral gap} \end{aligned}$

The remaining volumes may reuse this machine only after verifying the corresponding inputs in their new arithmetic or metric setting.

Bibliography

- [BG08] Jean Bourgain and Alex Gamburd. “Uniform expansion bounds for Cayley graphs of $SL_2(\mathbb{F}_p)$ ”. In: *Annals of Mathematics* 167 (2008), pp. 625–642.
- [BG10] Emmanuel Breuillard and Alex Gamburd. “Strong uniform expansion in $SL(2, p)$ ”. In: *Geometric and Functional Analysis* 20 (2010), pp. 1201–1209.
- [BV12] Jean Bourgain and Péter P. Varjú. “Expansion in $SL_d(\mathbb{Z}/q\mathbb{Z})$, q arbitrary”. In: *Inventiones Mathematicae* 188 (2012), pp. 151–173.
- [Gam+09] Alex Gamburd, Shlomo Hoory, Mehrdad Shahshahani, Aner Shalev, and Bálint Virág. “On the girth of random Cayley graphs”. In: *Random Structures & Algorithms* 35.1 (2009), pp. 100–117.
- [Hel08] Harald A. Helfgott. “Growth and generation in $SL_2(\mathbb{Z}/p\mathbb{Z})$ ”. In: *Annals of Mathematics* 167 (2008), pp. 601–623.
- [MVW84] C. R. Matthews, L. N. Vaserstein, and B. Weisfeiler. “Congruence properties of Zariski-dense subgroups. I”. In: *Proceedings of the London Mathematical Society* 48.3 (1984), pp. 514–532.

The Bourgain–Gamburd Method
Volume X: Squarefree Moduli and Correlation
Control

Scope and historical note

This volume develops squarefree-modulus expansion and correlation control in the rank-one family needed later. The volume is organized around the proved mathematical chain: exact and approximate Goursat structure, conductor-weight regularization, correlation escape, flattening, and the squarefree spectral endgame. Historical papers of Varjú and Salehi Golsefidy–Varjú are cited for architecture and attribution; every theorem used later is proved in the present series.

How to read this volume

The central difficulty is not the Chinese remainder theorem itself. It is *correlation*: a set in a product may project largely to every factor while still living close to a diagonal or fiber-product relation. Read Chapters 1–2 with this question in mind: “what prevents the coordinates from secretly encoding one another?” Goursat’s lemma identifies the exact obstruction, while the quantitative argument is supposed to show that an approximate subgroup either gains independent fiber entropy or approaches such an exact obstruction.

Chapters 3–5 then turn this structural information into random-walk escape and flattening. Whenever a conductor weight $W(I)$ appears, it should be read as the logarithmic size of the active quotient. This viewpoint is what prevents losses proportional to the number of prime factors. **Series position.** This volume uses Volumes II, IV, and IX. Its main outputs are squarefree correlation control and the squarefree spectral endgame.

Quantifier and asymptotic conventions. Unless a statement explicitly says otherwise, every constant occurring in a quantitative theorem is chosen in the order in which it is quantified in that theorem. Thus a phrase such as “for all sufficiently small δ ” means: there is a displayed or locally quantified threshold $\delta_0 > 0$, depending only on the parameters listed before that phrase, such that the assertion holds for every $0 < \delta < \delta_0$. The phrases “sufficiently large N ” and “sufficiently large p ” have the analogous meaning.

If $X \ll_\theta Y$ or $X = O_\theta(Y)$, the implicit constant depends only on θ and on the fixed ambient data of the current volume; it is independent of the variables quantified after θ . The words *absolute*, *uniform*, and *bounded* never introduce an unlisted dependence. Whenever an inequality between exponents or scales is essential to the argument, that inequality is written explicitly in the proof; these conventions are not a substitute for such a calculation.

Contents

Scope and historical note	iii
Scope and historical note	iii
How to read this volume	iv
How to read this volume	iv
Chapter 1. Squarefree Quotients, Conductors, and Exact Goursat Geometry	1
1. The squarefree rank-one family	1
2. Exact conductors	2
3. Exact subdirect products	3
Chapter 2. Regularized Projection Trees and Many-Factor Growth	5
1. Conductor weight and correlation defect	5
2. Exact regularization of a projection tree	7
3. The one-factor entropy trichotomy	8
4. Stability of graph-like multiplication	18
5. The large-block kernel/graph dichotomy	20
6. Many-factor product growth	34
Chapter 3. Index-Sensitive Escape from Squarefree Subgroup Obstructions	36
1. Index and obstruction conductor	36
2. A CRT version of the short-word law argument	37
3. From short words to logarithmic-time random walk	39
Chapter 4. Squarefree Inverse Theory and L^2 Flattening	41
1. The dyadic inverse set	41
2. Transfer of subgroup escape to the inverse set	42
3. One-step squarefree flattening	42
4. Bounded-stage iteration	44
Chapter 5. Conductor-Sensitive Quasirandomness and Squarefree Expansion	46
1. Representation dimension at exact conductor	46
2. The exact-conductor trace inequality	47
3. The squarefree spectral endpoint	47

4. Proof architecture and historical sources	48
Bibliography	50

CHAPTER 1

Squarefree Quotients, Conductors, and Exact Goursat Geometry

Squarefree congruence quotients are direct products, but that fact is useful only after we keep track of which factors a representation or obstruction actually sees. This chapter fixes the conductor language and proves the exact subdirect-product statements that later approximate arguments must imitate.

1. The squarefree rank-one family

Fix a finite set of excluded primes S_0 . Throughout this volume, q is squarefree and all $p \mid q$ lie outside S_0 . Put

$$G(q) = \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z}), \quad G_p = \mathrm{SL}_2(\mathbb{F}_p).$$

For $d \mid q$ let $\pi_d : G(q) \rightarrow G(d)$ denote reduction to the d -part.

THEOREM 1.1 (Chinese remainder decomposition; BG-X.1.001). *For squarefree q the natural map*

$$\Theta_q : G(q) \longrightarrow \prod_{p \mid q} G_p, \quad g \longmapsto (g \bmod p)_{p \mid q},$$

is an isomorphism. More generally, for every $d \mid q$,

$$G(q) \cong G(d) \times G(q/d),$$

and the first projection is π_d .

PROOF. The scalar Chinese remainder theorem gives a ring isomorphism

$$\mathbb{Z}/q\mathbb{Z} \cong \prod_{p \mid q} \mathbb{F}_p.$$

Applying 2×2 matrices entrywise gives $\mathrm{Mat}_2(\mathbb{Z}/q\mathbb{Z}) \cong \prod_{p \mid q} \mathrm{Mat}_2(\mathbb{F}_p)$, and determinant is computed coordinatewise. Hence determinant one on the left is equivalent to determinant one in every coordinate. The divisor statement follows by grouping the prime factors belonging to d and q/d . $\square$

EXAMPLE 1.2 (A three-prime quotient). If $q = prs$, then

$$G(q) = G_p \times G_r \times G_s.$$

The projection π_{pr} forgets only the s -coordinate. The point of this notation is that a representation may be nontrivial on p and r but trivial on s ; its natural size is then governed by pr , not by all of q .

2. Exact conductors

DEFINITION 1.3 (Support and conductor). Let $I(q) = \{p : p \mid q\}$. If an object $\mathcal{O}$ attached to $G(q)$ factors through $G(d)$ and through no proper divisor of d , its *exact conductor* is $\text{cond}(\mathcal{O}) = d$. Equivalently, under Theorem 1.1, its active support is the set of primes on which it is genuinely nontrivial.

PROPOSITION 1.4 (Exact conductor of an irreducible representation; BG-X.1.002). *Every irreducible complex representation ρ of $G(q)$ is a tensor product*

$$\rho = \bigotimes_{p \mid q} \rho_p,$$

with $\rho_p \in \text{Irr}(G_p)$. If

$$I(\rho) = \{p : \rho_p \text{ is nontrivial}\}, \quad d(\rho) = \prod_{p \in I(\rho)} p,$$

then ρ factors through $G(d(\rho))$ and through no proper divisor. Thus $d(\rho)$ is its exact conductor.

PROOF. The tensor-product classification of irreducibles of a finite direct product follows by iterating the standard two-factor result proved in Volume I: restrict an irreducible representation to the commuting factors, decompose one factor, and use Schur's lemma to identify the multiplicity space as a representation of the other factor. A tensor factor is invisible exactly when it is the trivial representation, giving the conductor statement. $\square$

DEFINITION 1.5 (Conductor of a subgroup obstruction). For $H \leq G(q)$ define its *defect support*

$$I(H) = \{p \mid q : \pi_p(H) \neq G_p\}.$$

The associated local conductor is $d_{\text{loc}}(H) = \prod_{p \in I(H)} p$. If every local projection is full but H is still proper, the obstruction is a *correlation obstruction*; its conductor is the smallest divisor d for which the image $\pi_d(H)$ is a proper subdirect product.

PROPOSITION 1.6 (Projection/section bookkeeping; BG-X.1.003). *Let $q = de$ with $(d, e) = 1$. Under $G(q) = G(d) \times G(e)$:*

- (i) $\ker \pi_d = \{1\} \times G(e)$;
- (ii) every fiber of π_d is a right and left coset of $\ker \pi_d$;

(iii) if $A \subset G(q)$, then

$$|A| = \sum_{x \in \pi_d A} |A \cap \pi_d^{-1}(x)|;$$

(iv) for $A, B \subset G(q)$ and $x \in \pi_d A$, $y \in \pi_d B$, the product of the corresponding fibers lies in the fiber over xy ;

(v) passage to a divisor commutes with products, inverses, convolution, and taking bounded product powers.

PROOF. All statements follow from the direct-product description. For (iv), write the two fibers as (x, A_x) and (y, B_y) in $G(d) \times G(e)$; then their product is $(xy, A_x B_y)$. Statement (v) follows because π_d is a homomorphism and push-forward commutes with convolution. $\square$

3. Exact subdirect products

THEOREM 1.7 (Goursat's lemma; BG-X.1.004). *Let $H \leq G_1 \times G_2$ project surjectively to both factors. Define*

$$N_1 = \{g_1 : (g_1, 1) \in H\}, \quad N_2 = \{g_2 : (1, g_2) \in H\}.$$

Then $N_i \triangleleft G_i$, and there is an isomorphism

$$\varphi : G_1/N_1 \xrightarrow{\sim} G_2/N_2$$

such that

$$H = \{(g_1, g_2) : \varphi(g_1 N_1) = g_2 N_2\}.$$

Conversely every such fiber product is a subdirect subgroup.

PROOF. If $(g_1, g_2) \in H$ and $(n_1, 1) \in H$, then conjugating inside H gives $(g_1 n_1 g_1^{-1}, 1) \in H$; surjectivity of the first projection proves $N_1 \triangleleft G_1$, and similarly for N_2 .

For $g_1 \in G_1$, choose g_2 with $(g_1, g_2) \in H$ and define $\varphi(g_1 N_1) = g_2 N_2$. If another choice g'_2 occurs, then $(1, g_2^{-1} g'_2) \in H$, so $g_2 N_2 = g'_2 N_2$; changing g_1 within its N_1 -coset is equally harmless. Multiplication in H makes φ a homomorphism. The same construction with the factors reversed gives its inverse. The displayed description of H is now immediate. $\square$

LEMMA 1.8 (Normal quotients of $\mathrm{SL}_2(\mathbb{F}_p)$). *After enlarging S_0 , every proper normal subgroup of $G_p = \mathrm{SL}_2(\mathbb{F}_p)$ is contained in its center $Z_p = \{\pm I\}$, and $G_p/Z_p = \mathrm{PSL}_2(\mathbb{F}_p)$ is nonabelian simple. If $p \neq r$ are outside S_0 , then G_p and G_r have no common nontrivial quotient.*

PROOF. The first assertion is part of the prime-field subgroup structure proved in Volume IV. Since G_p is perfect for $p \geq 5$, it has no nontrivial abelian quotient.

Therefore every nontrivial quotient is $\mathrm{PSL}_2(\mathbb{F}_p)$. Its order is

$$|\mathrm{PSL}_2(\mathbb{F}_p)| = \frac{p(p^2 - 1)}{2},$$

which strictly increases with $p \geq 5$. Hence $\mathrm{PSL}_2(\mathbb{F}_p) \not\cong \mathrm{PSL}_2(\mathbb{F}_r)$ when $p \neq r$, proving the final assertion. $\square$

THEOREM 1.9 (Exact diagonal/common-quotient classification; BG-X.1.005). *Let I be a finite set of allowed primes and $H \leq \prod_{p \in I} G_p$ a subdirect subgroup. Then H is the full product. More generally, for a product of finite perfect factors whose only proper nontrivial simple quotients are prescribed simple groups, every proper subdirect subgroup is obtained by imposing one or more diagonal identifications between isomorphic simple quotients.*

PROOF. For the rank-one squarefree family, induct on $|I|$. Write the product as $G' \times G_p$ and project H to G' . By induction that projection is all of G' . Goursat's lemma produces a common quotient of G' and G_p if H is proper. Any homomorphism from a direct product onto the nonabelian simple group $\mathrm{PSL}_2(\mathbb{F}_p)$ has commuting normal images of the factors; simplicity forces exactly one factor to map onto it. Thus some G_r with $r \neq p$ would have $\mathrm{PSL}_2(\mathbb{F}_p)$ as a nontrivial quotient, contradicting Lemma 1.8. Hence $H = G' \times G_p$.

For the general formulation, apply Goursat successively. Whenever a proper subdirect product occurs, it is a fiber product over a common quotient. Perfectness eliminates abelian quotients, so after passing to simple composition quotients the only possibility is an isomorphism between two isomorphic simple factors. Conversely a diagonal identification over such an isomorphism plainly defines a proper subdirect subgroup. $\square$

EXAMPLE 1.10 (Why pairwise surjectivity is not enough in repeated factors). The diagonal

$$\{(g, g) : g \in G\} \leq G \times G$$

projects surjectively to both coordinates but has index $|G|$. Theorem 1.9 says that this phenomenon cannot occur between the distinct-prime factors of the rank-one squarefree quotient. Chapter 2 proves the quantitative analogue: an *approximate* diagonal must either come from a genuine common quotient or pay a definite entropy cost.

CHAPTER 2

Regularized Projection Trees and Many-Factor Growth

The new difficulty in a squarefree quotient is correlation. A set can have large projections to every prime factor while remaining small because the coordinates are tied together. We regularize the projection tree, measure this dependence by conductor-weighted entropy, and prove that persistent dependence is an approximate Goursat graph. In the distinct-prime rank-one family such a graph has no genuine quotient on which to live.

1. Conductor weight and correlation defect

For a divisor $d \mid q$ put

$$W(d) = \log |G(d)| = \sum_{p \mid d} \log |G_p|.$$

If $I \subset I(q)$, write $G_I = \prod_{p \in I} G_p$, $W(I) = \log |G_I|$, and π_I for the corresponding projection.

DEFINITION 2.1 (Correlation defect). For a nonempty $B \subset G_I$ define

$$\text{Corr}_I(B) = \sum_{p \in I} \log |\pi_p B| - \log |B|.$$

For a two-block decomposition $I = J \sqcup K$ define

$$\text{Corr}_{J|K}(B) = \log |\pi_J B| + \log |\pi_K B| - \log |B|.$$

Both quantities are nonnegative by the elementary injection $B \hookrightarrow \prod \pi_p B$.

REMARK 2.2. $\text{Corr}_I(B) = 0$ exactly when B is the Cartesian product of its one-coordinate projections. A diagonal graph in $G \times G$ has correlation defect $\log |G|$. Thus correlation defect measures precisely the phenomenon that factorwise growth alone cannot see.

EXAMPLE 2.3 (A regular tree can still carry correlation). Take $X_1 = \{0, 1\}$, $X_2 = \{0, 1, 2\}$ and

$$B = \{(0, 0), (0, 1), (1, 1), (1, 2)\} \subset X_1 \times X_2.$$

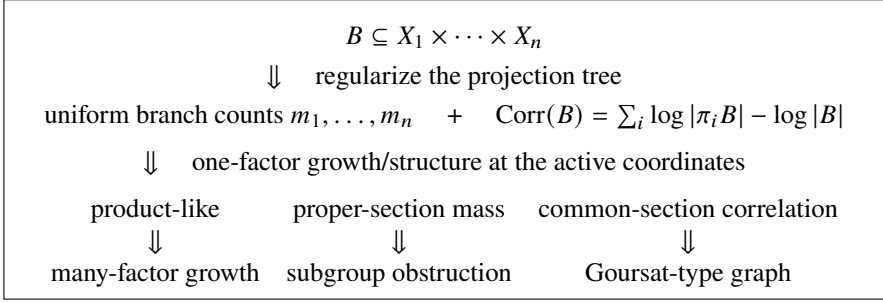


FIGURE 1. The squarefree many-factor mechanism. Projection-tree regularization makes the coordinate branching measurable; correlation defect records what factorwise projection sizes miss.

Then

$$|B| = 4, \quad |\pi_1 B| = 2, \quad |\pi_2 B| = 3,$$

so

$$\text{Corr}(B) = \log \frac{2 \cdot 3}{4} = \log \frac{3}{2} > 0.$$

Yet the projection tree in the order X_1, X_2 is already exactly regular: the root has two children and each first-coordinate prefix has exactly two second-coordinate children. What fails is product independence: the child sets are $\{0, 1\}$ above the prefix 0 and $\{1, 2\}$ above the prefix 1. Thus regular branching controls sizes, while the correlation defect records the remaining dependence.

DEFINITION 2.4 (The two structural obstructions). Fix parameters $\eta, \zeta > 0$, constants $C \geq 1$ and $c_* > 0$, and a K -approximate subgroup $A \subseteq G_I$. We say that a bounded power $B = A^C$ has a *macroscopic proper-section obstruction* if there is $J \subseteq I$ with $W(J) \geq c_* W(I)$, proper subgroups $H_p < G_p$ for $p \in J$, and a subset $B_0 \subseteq B$ of relative size at least $e^{-\zeta W(J)} K^{-C}$ contained in one coset of

$$\left(\prod_{p \in J} H_p \right) \times \left(\prod_{p \notin J} G_p \right).$$

We say that B has a *subdirect common-section obstruction* if, after deleting a block E with $W(E) \leq \eta W(I)$, there are disjoint nonempty blocks J, K , subgroups $U \leq G_J, V \leq G_K$, normal subgroups $N_U \triangleleft U, N_V \triangleleft V$, and a nontrivial isomorphism $U/N_U \simeq V/N_V$ such that a subset of B of relative size at least $e^{-\zeta W(J \cup K)} K^{-C}$ lies in the associated fiber-product subgroup after projection away from E .

2. Exact regularization of a projection tree

Order the prime factors as $p_1, \dots, p_n$ and write $X_i = G_{p_i}$. For $A \subset \prod_{i=1}^n X_i$, a prefix $x_{<i}$ has a set of children

$$C_i(x_{<i}) = \{x_i : \exists x_{>i} \text{ with } (x_{<i}, x_i, x_{>i}) \in A\}.$$

LEMMA 2.5 (Exact projection-tree regularization; BG-X.2.001). *Let $A \subset \prod_{i=1}^n X_i$ be nonempty. There is $A' \subset A$ and integers $m_i \in [1, |X_i|]$ such that every surviving prefix at level $i - 1$ has exactly m_i surviving children and*

$$|A'| = \prod_{i=1}^n m_i,$$

while

$$\log |A'| \geq \log |A| - \sum_{i=1}^n \log(2 + \lceil \log_2 |X_i| \rceil).$$

Consequently, for every $\eta > 0$ there is $p_0(\eta)$ such that if every $X_i = G_{p_i}$ with $p_i \geq p_0(\eta)$, then

$$|A'| \geq e^{-\eta W(I)} |A|.$$

PROOF. Work from the last coordinate upward. At the final level, partition the nonempty child sets according to the dyadic interval containing their cardinality. There are at most $1 + \lceil \log_2 |X_n| \rceil$ classes. Keep a class carrying the largest number of leaves, and inside every retained child set trim to the smallest cardinality in that class. This loses at most a factor $2(1 + \lceil \log_2 |X_n| \rceil)$ and makes the number of last-coordinate children identical for every surviving prefix.

After this pruning, regard each surviving child subtree as a single weighted leaf and repeat at level $n - 1$. Because all descendant subtrees below a retained child have equal size, trimming the number of children makes the total descendant count equal as well. Iterating gives an exact regular tree with the stated loss. The product formula for $|A'|$ follows by induction on depth.

For $X_i = G_{p_i}$, $\log |X_i| = 3 \log p_i + O(1)$. Hence

$$\frac{\log(2 + \log_2 |X_i|)}{\log |X_i|} \rightarrow 0$$

uniformly as $p_i \rightarrow \infty$. Choosing $p_0(\eta)$ so that every summand in the numerator is at most $\eta \log |X_i|$ proves the final estimate. $\square$

EXAMPLE 2.6 (Why divisor count does not enter). A loss $10^{\omega(q)}$ looks fatal if one measures complexity by $\omega(q)$. If all $p \mid q$ are large, however,

$$\omega(q) \ll \frac{W(q)}{\log p_0}.$$

Thus every fixed loss per prime is $e^{o(W(q))}$. This elementary observation is the first appearance of the rule used throughout Arc IV: *charge errors to conductor weight, never to the number of divisors*.

3. The one-factor entropy trichotomy

We use two earlier inputs repeatedly.

Earlier input (Volumes II and IV). Volume II proves the two-set Balog–Szemerédi–Gowers theorem and the passage from high multiplicative energy to a controlled approximate group. Volume IV proves Helfgott growth in G_p and the bounded-complexity classification of proper subgroup obstructions. These are the only external inputs used in the next lemma.

LEMMA 2.7 (Finite subgroup-complexity hierarchy; BG-X.2.002). *There is an absolute integer D_0 such that, for all allowed primes p , every proper subgroup $H < G_p$ has a subgroup $H_0 \leq H$ of index at most D_0 which is either contained in a Borel subgroup or in a torus. The remaining exceptional possibilities have order $O(1)$. Moreover, if $B \subset G_p$ is a K -approximate subgroup, then one of the following holds with constants polynomial in K :*

- (i) B is contained in $K^{O(1)}$ cosets of a proper subgroup of the above hierarchy;
- (ii) $|B| \leq K^{O(1)}$;
- (iii) $|G_p \setminus B^C| = 0$ for an absolute C depending only on the Helfgott exponent and the power of K under consideration.

PROOF. The subgroup statement is the rank-one subgroup dichotomy proved in Volume IV. For the approximate-group statement, if $\langle B \rangle$ is proper, the Ruzsa covering theorem from Volume II covers B by controlledly many cosets of that subgroup. If $\langle B \rangle = G_p$, Helfgott growth says that every intermediate-size bounded product power grows by a fixed exponent. Since bounded powers of a K -approximate group have size at most $K^{O(1)}|B|$, repeated growth can stop only at the very small scale $K^{O(1)}$ or after reaching the large-set completion threshold, where Gowers mixing from Volume I gives $B^C = G_p$. $\square$

LEMMA 2.8 (Rank-one subgroup hierarchy with intersections). *After enlarging the finite exceptional-prime set, there are an absolute constant $L \geq 2$ and three conjugation-invariant classes of proper subgroups $\mathcal{H}_0, \mathcal{H}_1, \mathcal{H}_2$ of $G_p = \mathrm{SL}_2(\mathbb{F}_p)$ with the following properties.*

- (i) every subgroup in $\mathcal{H}_0$ has order at most L ;
- (ii) $\mathcal{H}_1$ consists of maximal tori and $\mathcal{H}_2$ of Borel subgroups;
- (iii) every proper subgroup $H < G_p$ is contained in the union of at most L left cosets of some member of $\mathcal{H}_0 \cup \mathcal{H}_1 \cup \mathcal{H}_2$;
- (iv) if $H_1 \neq H_2$ lie in $\mathcal{H}_j$, $j = 1, 2$, then $H_1 \cap H_2$ is contained in the union of at most L cosets of a member of $\mathcal{H}_{j-1}$;

(v) *the normalizer of every member of $\mathcal{H}_1 \cup \mathcal{H}_2$ is a proper subgroup of G_p . Moreover there is an absolute $c_h > 0$ such that every proper subgroup of G_p has index at least $|G_p|^{c_h}$.*

PROOF. The subgroup classification proved in Volume IV says that a proper subgroup has a subgroup of uniformly bounded index which is contained in a Borel or in a maximal torus, unless the original subgroup has uniformly bounded order. This gives (i)–(iii), after enlarging L .

For (iv), use the natural action on $\mathbf{P}^1(\mathbb{F}_p)$. A Borel is the stabilizer of one point. Two distinct Borels therefore stabilize two distinct points, and their intersection is contained in the split torus fixing both points. A noncentral semisimple element has a unique maximal torus as its centralizer; consequently two distinct maximal tori intersect only in the bounded center. This gives the two intersection statements after enlarging L once more.

A Borel is self-normalizing. The normalizer of a maximal torus is the usual dihedral extension of that torus and is still proper; this proves (v). Finally Volume IV also gives the rank-one subgroup-index bound. Equivalently, one can read it directly from the same classification: Borels have index $p + 1$, tori have index $\asymp p^2$, and the exceptional groups have still larger index, whereas $|G_p| \asymp p^3$. Thus some absolute $c_h > 0$ works uniformly. $\square$

LEMMA 2.9 (Conductor-weighted global coset pruning). *Fix $\kappa > 0$ and let $A \subseteq G_I$ be nonempty. There exist a subset $B \subseteq A$, a partition $I = J_{\text{good}} \sqcup J_{\text{bad}}$, and for every $p \in J_{\text{bad}}$ a proper subgroup $H_p < G_p$ and a coset $g_p H_p$ such that*

$$(1) \quad B \subseteq \bigcap_{p \in J_{\text{bad}}} \pi_p^{-1}(g_p H_p),$$

$$(2) \quad |B| \geq |A| \exp\{-\kappa W(J_{\text{bad}})\},$$

and, for every $p \in J_{\text{good}}$ and every proper coset $gH < G_p$,

$$(3) \quad (\pi_p)_* \chi_B(gH) \leq |G_p|^{-\kappa}.$$

The set on the right of (1) is one coset of the product subgroup

$$\left(\prod_{p \in J_{\text{bad}}} H_p \right) \times \left(\prod_{p \in J_{\text{good}}} G_p \right),$$

whose index is at least $\exp\{c_h W(J_{\text{bad}})\}$.

PROOF. Start with $B = A$ and all coordinates declared good. If for some currently good p there is a proper coset gH with

$$|B \cap \pi_p^{-1}(gH)| > |G_p|^{-\kappa} |B|,$$

replace B by this intersection, record the chosen coset, and move p to the bad set. Each coordinate is used at most once, so the procedure terminates and (3) holds at termination. Multiplying the losses from the successive restrictions gives (2). The

recorded conditions in different CRT factors are independent coordinate conditions, so their intersection is exactly the product coset displayed above. The index estimate follows from the last assertion of Lemma 2.8. $\square$

LEMMA 2.10 (Bad conductor weight under global subgroup non-concentration). *Assume, in addition to the hypotheses of Lemma 2.9, that there are constants $\varepsilon_0 > 0$ and $E \geq 0$ such that*

$$(4) \quad \chi_A(gH) \leq [G_I : H]^{-\varepsilon_0} e^E$$

for every proper subgroup $H < G_I$ and every coset gH . If $0 < \kappa < \varepsilon_0 c_h$, then the bad set produced by Lemma 2.9 satisfies

$$(5) \quad W(J_{\text{bad}}) \leq \frac{E}{\varepsilon_0 c_h - \kappa}.$$

In particular, under a power-saving global coset bound with $E = o(W(I))$, the pruning discards only $o(W(I))$ conductor weight.

PROOF. If $J_{\text{bad}} = \emptyset$, the conclusion is immediate. Otherwise let

$$H_* = \left(\prod_{p \in J_{\text{bad}}} H_p \right) \times \left(\prod_{p \in J_{\text{good}}} G_p \right)$$

and let $g_* H_*$ be the product coset from Lemma 2.9. By construction $B \subseteq A \cap g_* H_*$, so (2) gives

$$\chi_A(g_* H_*) \geq \frac{|B|}{|A|} \geq e^{-\kappa W(J_{\text{bad}})}.$$

On the other hand, Lemma 2.8 gives

$$[G_I : H_*] = \prod_{p \in J_{\text{bad}}} [G_p : H_p] \geq e^{c_h W(J_{\text{bad}})}.$$

Applying (4) to $g_* H_*$ therefore yields

$$e^{-\kappa W(J_{\text{bad}})} \leq e^{-\varepsilon_0 c_h W(J_{\text{bad}}) + E}.$$

Rearranging proves (5). $\square$

REMARK 2.11 (What “consolidation” means here). This lemma is the precise replacement for an unsafe fiberwise pigeonhole. The proper coset at a tree node is allowed to depend on the prefix, and there is no reason that such choices should glue automatically. We instead prune using the *current global measure*. Thus all recorded bad coordinates genuinely lie in one fixed product coset. If $W(J_{\text{bad}})$ is macroscopic, this is already the required global subgroup obstruction. If it is small, the surviving global measure is nonconcentrated in every remaining factor and can be used as a random interleaving measure below.

LEMMA 2.12 (Random interleaving destroys rank-one subgroup concentration). *Fix $0 < \delta < 1/4$ and $\kappa > 0$. There are $a = a(\delta, \kappa) > 0$ and p_0 such that the following holds. Put $N = |G_p|$ and let ν be a probability measure on G_p satisfying*

$$\nu(gH) \leq N^{-\kappa}$$

for every proper coset gH . Let $C_1, C_2, C_3, C_4 \subset G_p$ all have the same cardinality D with

$$N^\delta \leq D \leq N^{1-\delta}.$$

Choose Y_1, Y_2, Y_3 independently with law ν and set

$$\lambda = \chi_{C_1} * \delta_{Y_1} * \chi_{C_2} * \delta_{Y_2} * \chi_{C_3} * \delta_{Y_3} * \chi_{C_4}.$$

Then, with probability at least $3/4$ for all sufficiently large p ,

$$(6) \quad \lambda(gH) \leq D^{-a}$$

for every proper coset $gH \subset G_p$.

PROOF. We follow the finite subgroup-hierarchy induction rather than trying to control all proper subgroups independently. Let $\mathcal{H}_0, \mathcal{H}_1, \mathcal{H}_2$ be the hierarchy of Lemma 2.8; thus $\mathcal{H}_0$ consists of bounded subgroups, $\mathcal{H}_1$ of maximal tori, and $\mathcal{H}_2$ of Borels, up to a bounded number of cosets. Distinct members of one level have intersection covered by a bounded number of cosets from a lower level, and the normalizer of every member of levels 1 and 2 is proper.

We first isolate the counting mechanism used at both nontrivial levels. Suppose ρ is a probability measure satisfying

$$\rho(gK) \leq D^{-b}$$

for every coset of every K in levels $< j$. If $H_1, \dots, H_M \in \mathcal{H}_j$ are distinct and $\rho(H_i) \geq q$, then inclusion–exclusion gives

$$(7) \quad 1 \geq Mq - \binom{M}{2} LD^{-b},$$

where L is the hierarchy constant. Consequently, if $q = D^{-2\beta}/4$ and $b > 7\beta$, then for all sufficiently large D there are fewer than $D^{3\beta}$ such high-mass subgroups. Indeed, inserting $M = D^{3\beta}$ in (7) makes the first term grow like D^β , whereas the overlap term is $O(D^{6\beta-b}) = o(1)$.

Choose exponents

$$\beta_0 = \frac{1}{2}, \quad 0 < \beta_2 < \frac{\beta_1}{8}, \quad 0 < \beta_1 < \min\left(\frac{1}{16}, \frac{\kappa}{24}\right).$$

For a uniform set measure $\mu_C = \chi_C$ and a bounded subgroup $K \in \mathcal{H}_0$,

$$(\tilde{\mu}_C * \mu_C)(gK) \leq L/D \leq D^{-\beta_0}$$

for large D . This is the base of the hierarchy induction.

Consider

$$\lambda_1 = \chi_{C_1} * \delta_{Y_1} * \chi_{C_2}.$$

If $\lambda_1(gH) > D^{-\beta_1}$ for some $H \in \mathcal{H}_1$, expand the coset mass over a left transversal of H and apply Cauchy–Schwarz on both sides. There exist conjugates $H_1, H_2 \in \mathcal{H}_1$ such that

$$(\tilde{\chi}_{C_1} * \chi_{C_1})(H_1) > \frac{1}{4}D^{-2\beta_1}, \quad (\chi_{C_2} * \tilde{\chi}_{C_2})(H_2) > \frac{1}{4}D^{-2\beta_1},$$

and $H_1 = Y_1 H_2 Y_1^{-1}$. By the preceding counting argument, there are fewer than $D^{3\beta_1}$ possibilities for each of H_1 and H_2 . For a fixed pair, the conjugacy relation either has no solution or restricts Y_1 to one coset of $N_{G_p}(H_2)$, a proper subgroup. Hence

$$\mathbb{P}[\lambda_1(gH) > D^{-\beta_1} \text{ for some } g, H \in \mathcal{H}_1] \leq D^{6\beta_1} N^{-\kappa} \leq N^{-\kappa/2}.$$

The same conclusion holds independently for $\eta_1 = \chi_{C_3} * \delta_{Y_3} * \chi_{C_4}$.

Condition on both level-1 success events and set $\lambda_2 = \lambda_1 * \delta_{Y_2} * \eta_1$. If $\lambda_2(gH) > D^{-\beta_2}$ for some $H \in \mathcal{H}_2$, the same transversal/Cauchy–Schwarz calculation supplies conjugates $H_1, H_2 \in \mathcal{H}_2$ with

$$(\tilde{\lambda}_1 * \lambda_1)(H_1) > \frac{1}{4}D^{-2\beta_2}, \quad (\eta_1 * \tilde{\eta}_1)(H_2) > \frac{1}{4}D^{-2\beta_2},$$

and $H_1 = Y_2 H_2 Y_2^{-1}$. Since distinct level-2 subgroups intersect in boundedly many level-1 cosets, the level-1 bounds imply the overlap hypothesis with $b = \beta_1$. As $\beta_1 > 7\beta_2$, there are fewer than $D^{3\beta_2}$ possibilities on each side. Therefore, conditionally on the first-stage success,

$$\mathbb{P}[\lambda_2(gH) > D^{-\beta_2} \text{ for some } g, H \in \mathcal{H}_2] \leq D^{6\beta_2} N^{-\kappa} \leq N^{-\kappa/2}.$$

The union of the three failure events has probability $o(1)$, and hence is less than $1/4$ once p is large.

Every proper subgroup of G_p is covered by at most L cosets of a member of one of the three hierarchy levels. On the success event we therefore have

$$\lambda(gH) \leq L \max(D^{-\beta_2}, D^{-\beta_1}, L/D) \leq D^{-\beta_2/2}$$

for all proper cosets and all sufficiently large D . Take $a = \beta_2/2$. □

LEMMA 2.13 (Mixed flattening from coset nonconcentration). *Fix $0 < \delta < 1/4$ and $a > 0$. There are $\sigma = \sigma(\delta, a) > 0$ and p_0 such that the following holds. Put $N = |G_p|$. Let D satisfy $N^\delta \leq D \leq N^{1-\delta}$, let $C \subset G_p$ have $|C| = D$, and let λ be a probability measure such that*

$$\|\lambda\|_\infty \leq D^{-1}, \quad \lambda(gH) \leq D^{-a}$$

for every proper coset. Then

$$(8) \quad \|\lambda * \chi_C\|_2 \leq D^{-1/2-\sigma}.$$

PROOF. Suppose instead that the left side is at least $D^{-1/2-\sigma}$. Remove atoms of λ of size at most $D^{-\sigma}/(8N)$. Their total ℓ^1 mass is at most $D^{-\sigma}/8$, so their convolution with χ_C contributes at most one eighth of the assumed lower bound. Decompose the remaining measure into dyadic levels f_j . There are $J = O(\log N)$ such levels. Hence one level f , supported on a set E of size M and satisfying $\alpha/2 < f(x) \leq \alpha$ there, obeys

$$(2.4) \quad \|f * \chi_C\|_2 \geq qD^{-1/2}, \quad q = (4J)^{-1}D^{-\sigma}.$$

For large N , $J \leq D^\sigma$, so $q \geq D^{-2\sigma}/4$.

Young's inequality applied to (2.4) gives $\|f\|_1 \geq q$ and $\|f\|_2 \geq qD^{-1/2}$. Since $\alpha \leq \|\lambda\|_\infty \leq D^{-1}$,

$$(2.5) \quad M \geq qD.$$

On the other hand $\|f\|_1 \leq 1$ and $\|f\|_2^2 \leq \alpha^2 M \leq 4/M$, so (2.4) gives

$$(2.6) \quad M \leq 4Dq^{-2}.$$

Moreover, from $f \leq \alpha 1_E$ and (2.4),

$$(2.7) \quad E(E, C) = \|1_E * 1_C\|_2^2 \geq q^2 D \alpha^{-2} \geq q^2 D^3.$$

Combining (2.6) with (2.7) yields

$$(2.8) \quad E(E, C) \geq cq^5 M^{3/2} D^{3/2}.$$

Apply the full two-set BSG theorem of Volume II with parameter $K_0 \ll q^{-5}$. It gives subsets $E_* \subset E$, $C_* \subset C$ of relative size $q^{O(1)}$ and

$$|E_* C_*| \leq q^{-O(1)} \sqrt{MD}.$$

The small-product theorem of Volume II therefore supplies a $q^{-O(1)}$ -approximate group P of size

$$(2.9) \quad Dq^{O(1)} \leq |P| \leq Dq^{-O(1)}$$

which controls both E_* and C_* by $q^{-O(1)}$ translates. Here the lower bound follows from the size of C_* and the covering, while the upper bound uses (2.6).

Choose σ sufficiently small in terms of δ and a . Since $q = D^{-O(\sigma)}$, (2.9) places P in the medium range $N^{\delta/2} \leq |P| \leq N^{1-\delta/2}$, with approximation parameter $N^{O(\sigma)}$. Lemma 2.7 rules out its very-small and completion alternatives, so P is covered by $D^{O(\sigma)}$ cosets of a proper subgroup H .

Finally E_* is covered by $D^{O(\sigma)}$ cosets of H , whereas its λ -mass is at least

$$\lambda(E_*) \geq (\alpha/2)|E_*| \geq q^{O(1)} \alpha M \geq q^{O(1)},$$

because $\|f\|_1 \geq q$ and $\alpha M \geq \|f\|_1$. Thus some proper coset has λ -mass at least $D^{-O(\sigma)}$, contrary to D^{-a} once σ is chosen small enough. This contradiction proves (8). $\square$

PROPOSITION 2.14 (Simultaneous mesoscopic entropy gain). *Fix $0 < \delta < 1/4$ and $\kappa > 0$. Let A be a symmetric K -approximate subgroup of G_I , and let $A' \subset A$ be an exact regularization with branching degrees D_p . Let $B \subset A$ and $I = J_{\text{good}} \sqcup J_{\text{bad}}$ be supplied by Lemma 2.9. Put*

$$M = \{p \in J_{\text{good}} : |G_p|^\delta \leq D_p \leq |G_p|^{1-\delta}\}.$$

Then, after enlarging the exceptional-prime set,

$$(9) \quad W(M) \ll_{\delta, \kappa} \log K + (\log |A| - \log |A'|).$$

PROOF. Take five independent uniform variables $X_1, \dots, X_5$ on A' . Independently choose b_1, b_2, b_3 uniformly from B and consider, for fixed $b = (b_1, b_2, b_3)$,

$$Z_b = X_1 b_1 X_2 b_2 X_3 b_3 X_4 X_5.$$

Its support is contained in A^8 .

Order the coordinates as in the regular tree. Since the prefix of Z_b is a function of the five input prefixes,

$$(2.10) \quad H(Z_b) \geq \sum_p H((Z_b)_p \mid (X_1)_{<p}, \dots, (X_5)_{<p}).$$

For fixed input prefixes, the five current-coordinate variables are independent and uniform on five sets $C_1, \dots, C_5$ of the common cardinality D_p . At a coordinate $p \in M$, the three multipliers $\pi_p(b_j)$ are independent with law $\nu_p = (\pi_p)_* \chi_B$, which satisfies $\nu_p(gH) \leq |G_p|^{-\kappa}$ by Lemma 2.9. Lemma 2.12 therefore makes the first four factors into a measure λ satisfying the coset bound with probability at least $3/4$; also $\|\lambda\|_\infty \leq D_p^{-1}$. Lemma 2.13 then gives

$$\|\lambda * \chi_{C_5}\|_2 \leq D_p^{-1/2-\sigma}$$

for a fixed $\sigma = \sigma(\delta, \kappa) > 0$. On this event Shannon entropy is at least the collision entropy,

$$H((Z_b)_p \mid \text{the five prefixes}) \geq (1 + 2\sigma) \log D_p.$$

Without the event we still have the elementary lower bound $\log D_p$, because conditioning on the first four current-coordinate variables leaves the last uniform variable X_5 and multiplication by a fixed group element is a bijection. Hence, after averaging in b ,

$$(2.11) \quad \mathbb{E}_b H((Z_b)_p \mid \text{prefixes}) \geq \log D_p + \sigma \log D_p, \quad p \in M.$$

At all other coordinates the same baseline argument gives $\log D_p$.

Average (2.10), use (2.11), and recall $\sum_p \log D_p = \log |A'|$. Some fixed choice of b satisfies

$$H(Z_b) \geq \log |A'| + \sigma \sum_{p \in M} \log D_p.$$

Therefore

$$|A^8| \geq |\text{supp } Z_b| \geq |A'| \exp\left(\sigma \sum_{p \in M} \log D_p\right).$$

Since A is a K -approximate group, $|A^8| \leq K^7|A|$. Thus

$$\sigma \sum_{p \in M} \log D_p \leq 7 \log K + \log |A| - \log |A'|.$$

Finally $\log D_p \geq \delta \log |G_p|$ on M , proving (9). $\square$

LEMMA 2.15 (Small-fiber entropy growth; BG-X.2.003). *Fix $0 < \delta < 1/4$. There are $c = c(\delta) > 0$ and $p_0(\delta)$ with the following property. Let $B, C \subset G_p$ satisfy*

$$|G_p|^\delta \leq |B|, |C| \leq |G_p|^{1-\delta}, \quad |B|/|C| \in [1/2, 2],$$

and assume neither set has a subset of relative size $\geq |G_p|^{-c}$ contained in a coset of a proper subgroup. If X, Y are independent uniform random variables on B, C , then

$$H(XY) \geq \frac{1}{2} \log |B| + \frac{1}{2} \log |C| + c \log |G_p|.$$

PROOF. Suppose the conclusion fails with a parameter $c_1 > 0$ to be chosen. Shannon entropy dominates collision entropy, so

$$\sum_g \mathbb{P}(XY = g)^2 \geq e^{-H(XY)} \geq (|B||C|)^{-1/2} |G_p|^{-c_1}.$$

Multiplying by $|B|^2|C|^2$ shows that the multiplicative energy of (B, C) is at least

$$|B|^{3/2}|C|^{3/2}|G_p|^{-c_1}.$$

Because $|B|$ and $|C|$ are comparable, the two-set BSG theorem of Volume II produces subsets $B' \subset B$, $C' \subset C$ of relative size $|G_p|^{-O(c_1)}$ controlled by a $|G_p|^{O(c_1)}$ -approximate group P with

$$|P| = |B|^{1+O(c_1/\delta)}.$$

Choose c_1 sufficiently small in terms of δ and the Helfgott exponent. The assumed size range keeps P in an intermediate range. Lemma 2.7 then forces P , hence substantial subsets of B and C , into controlledly many cosets of a proper subgroup, contradicting the hypothesis. Therefore a fixed positive entropy gain $c(\delta) \log |G_p|$ is unavoidable. $\square$

PROPOSITION 2.16 (Clean mesoscopic weight is charged by $\log K$). *Fix $0 < \delta < 1/4$ and let $c = c(\delta)$ be the constant in Lemma 2.15. Let A be a symmetric K -approximate subgroup of G_I , $1 \in A$, and let $A' \subset A$ be an exact regularization from Lemma 2.5, with branching sets of cardinality m_p . Call a mesoscopic coordinate p clean if*

$$|G_p|^\delta \leq m_p \leq |G_p|^{1-\delta}$$

and, for every surviving prefix, its child fiber contains no subset of relative size $\geq |G_p|^{-c}$ lying in a coset of a proper subgroup. If M_{cl} is the set of clean mesoscopic coordinates, then

$$c W(M_{\text{cl}}) \leq \log K + 2(\log |A| - \log |A'|).$$

In particular, if the regularization loss is at most $\varepsilon W(I)$, then

$$W(M_{\text{cl}}) \leq c^{-1} \log K + 2c^{-1} \varepsilon W(I).$$

PROOF. Let X, Y be independent uniform random variables on A' , and put $Z = XY$. Order the factors as in the regularized projection tree. Since $Z_{<p}$ is a function of $(X_{<p}, Y_{<p})$, conditioning on the latter pair gives at least as much information as conditioning on $Z_{<p}$ alone. Hence

$$(2.1) \quad H(Z) = \sum_p H(Z_p \mid Z_{<p}) \geq \sum_p H(Z_p \mid X_{<p}, Y_{<p}).$$

For a fixed pair of surviving prefixes, X_p and Y_p are independent and uniform on two child sets of the same cardinality m_p . At a clean mesoscopic coordinate Lemma 2.15 therefore gives

$$H(Z_p \mid X_{<p}, Y_{<p}) \geq \log m_p + c \log |G_p|.$$

At every other coordinate the elementary inequality $H(UV) \geq \max\{H(U), H(V)\}$ for independent finite-group valued random variables gives

$$H(Z_p \mid X_{<p}, Y_{<p}) \geq \log m_p.$$

(The inequality follows because for each fixed v , multiplication by v is a bijection, so $H(UV \mid V) = H(U)$, whence $H(UV) \geq H(UV \mid V) = H(U)$.) Summing in (2.1) and using $\sum_p \log m_p = \log |A'|$ yields

$$(2.2) \quad H(Z) \geq \log |A'| + c W(M_{\text{cl}}).$$

The support of Z is contained in A^2 , so $H(Z) \leq \log |A^2| \leq \log K + \log |A|$. Combining with (2.2) gives

$$c W(M_{\text{cl}}) \leq \log K + \log |A| - \log |A'|.$$

This is slightly stronger than the displayed statement; the factor 2 is retained to absorb a second regularization when the proposition is used after a bounded product power. The final assertion follows immediately. $\square$

REMARK 2.17 (Role of mesoscopic consolidation). The clean mesoscopic coordinates are now completely accounted for: a slow approximate group can carry only $O(\log K) + o(W)$ conductor weight there. The structural complement is handled earlier by global pruning and randomized interleaving: a macroscopic bad block becomes one genuine product-subgroup obstruction, while the surviving factors satisfy the uniform coset non-concentration needed for Proposition 2.14. Thus no additional mesoscopic subgroup lemma is needed in the final approximate-Goursat assembly.

LEMMA 2.18 (Large-fiber quasirandom completion; BG-X.2.004). *There exists $\delta_0 > 0$ such that if $B \subset G_p$ and*

$$|B| \geq |G_p|^{1-\delta_0},$$

then $B^3 = G_p$ for all sufficiently large allowed primes.

PROOF. Volume I proves the Gowers three-set mixing theorem: if D_p is the smallest nontrivial representation dimension and

$$|B_1||B_2||B_3| > |G_p|^3/D_p,$$

then $B_1 B_2 B_3 = G_p$. Volumes I and VI give $D_p \gg p \gg |G_p|^{1/3-o(1)}$. Taking $B_1 = B_2 = B_3 = B$ and any fixed $\delta_0 < 1/9$ gives the claim for large p . $\square$

LEMMA 2.19 (Simultaneous completion on a large-branch block). *Let $A' \subseteq G_I$ be exactly regular with respect to the fixed ordering of I , and let $U \subseteq I$ be a set of coordinates such that every branch degree at $p \in U$ is at least $|G_p|^{1-\delta_0}$, with δ_0 as in Lemma 2.18. Then, for all sufficiently large allowed primes,*

$$\pi_U((A')^3) = G_U.$$

Thus the large coordinates are completed simultaneously; no iteration whose length depends on $|U|$ is required.

PROOF. Write $U = \{p_{i_1}, \dots, p_{i_r}\}$ in the order inherited from I . We prove by induction on j that every element of $G_{\{p_{i_1}, \dots, p_{i_j}\}}$ is the projection of a product of three elements of A' .

For $j = 1$, the projection of A' to $G_{p_{i_1}}$ has at least the branch degree $D_{p_{i_1}} \geq |G_{p_{i_1}}|^{1-\delta_0}$ elements, so Lemma 2.18 gives the claim.

Assume the assertion through $j - 1$ and fix a target $(g_1, \dots, g_j)$. By induction choose $a_1, a_2, a_3 \in A'$ whose product has the prescribed first $j - 1$ coordinates. For $\ell = 1, 2, 3$, let F_ℓ be the set of elements of A' agreeing with a_ℓ on the coordinates of U that precede p_{i_j} . Fixing only those coordinates imposes fewer conditions than fixing the entire projection-tree prefix of a_ℓ before p_{i_j} . Consequently F_ℓ contains one complete prefix fiber of a_ℓ , and exact regularity gives

$$|\pi_{p_{i_j}}(F_\ell)| \geq D_{p_{i_j}} \geq |G_{p_{i_j}}|^{1-\delta_0}.$$

By the three-set mixing theorem used in Lemma 2.18,

$$\pi_{p_{i_j}}(F_1) \pi_{p_{i_j}}(F_2) \pi_{p_{i_j}}(F_3) = G_{p_{i_j}}.$$

Choose $b_\ell \in F_\ell$ so that the p_{i_j} coordinate of $b_1 b_2 b_3$ equals g_j . Replacing a_ℓ by b_ℓ does not alter the already fixed U -coordinates, so the induction closes. $\square$

4. Stability of graph-like multiplication

The remaining obstruction after one has filled the large-branch coordinates is a near-graph. We prove the correction theorem in the exact weighted form needed below; no abstract stability theorem is imported.

LEMMA 2.20 (Majority correction for approximate homomorphisms; BG-X.2.005). *Let G, H be finite groups and $f : G \rightarrow H$. Suppose*

$$\mathbb{P}_{x,y \in G} [f(xy) = f(x)f(y)] \geq 1 - \varepsilon, \quad 0 < \varepsilon < 1/12.$$

Then there is a unique homomorphism $\phi : G \rightarrow H$ such that

$$\mathbb{P}_x [f(x) = \phi(x)] \geq 1 - \frac{\varepsilon}{1 - 2\varepsilon}.$$

In particular this agreement probability is at least $1 - 2\varepsilon$ when $\varepsilon \leq 1/4$.

PROOF. For $x, y \in G$ put

$$F_x(y) = f(xy)f(y)^{-1}.$$

Fix x . If y, z are independent uniform elements of G , then whenever

$$f(xyz) = f(xy)f(z), \quad f(yz) = f(y)f(z),$$

we have $F_x(yz) = F_x(y)$. Since (xy, z) and (y, z) are each uniform on G^2 , the union bound gives

$$\mathbb{P}_{y,z} [F_x(yz) = F_x(y)] \geq 1 - 2\varepsilon.$$

But y and yz are independent uniform elements. If ν_x is the law of $F_x(y)$, the left side is $\sum_h \nu_x(h)^2$. Hence

$$\max_h \nu_x(h) \geq \sum_h \nu_x(h)^2 \geq 1 - 2\varepsilon > 1/2.$$

There is therefore a unique majority value; define it to be $\phi(x)$.

For arbitrary x_1, x_2, y there is the exact identity

$$F_{x_1 x_2}(y) = F_{x_1}(x_2 y) F_{x_2}(y).$$

Each of the three random variables on the two sides equals its majority value with probability at least $1 - 2\varepsilon$. Their three majority events have simultaneous positive probability because $6\varepsilon < 1$. At a y in their intersection the exact identity gives

$$\phi(x_1 x_2) = \phi(x_1) \phi(x_2).$$

Thus ϕ is a homomorphism.

Finally the original multiplication test says

$$\mathbb{E}_x \mathbb{P}_y [F_x(y) = f(x)] \geq 1 - \varepsilon.$$

If $f(x) \neq \phi(x)$, then the probability on the left is at most $1 - \nu_x(\phi(x)) \leq 2\varepsilon$. If β is the proportion of such x , then

$$1 - \varepsilon \leq (1 - \beta) + 2\varepsilon\beta,$$

so $\beta \leq \varepsilon/(1 - 2\varepsilon)$, as claimed. Uniqueness follows because two homomorphisms agreeing on more than half of a finite group agree everywhere: their equalizer is a subgroup of index < 2 . $\square$

LEMMA 2.21 (Conductor-weighted coordinate correction). *Let G be finite and let*

$$H = \prod_{j \in J} H_j, \quad w_j = \log |H_j|, \quad W_J = \sum_{j \in J} w_j.$$

For a map $f : G \rightarrow H$ write $f_j = \pi_j f$ and assume

$$\mathbb{E}_{x,y} \sum_{j: f_j(xy) \neq f_j(x)f_j(y)} w_j \leq \tau W_J.$$

Fix $\varepsilon_0 < 1/12$. Then there is a set $J_{\text{bad}} \subset J$ with

$$\sum_{j \in J_{\text{bad}}} w_j \leq \frac{\tau}{\varepsilon_0} W_J$$

and a homomorphism

$$\phi : G \longrightarrow \prod_{j \notin J_{\text{bad}}} H_j$$

such that

$$\mathbb{E}_x d_{J \setminus J_{\text{bad}}}(\pi_{J \setminus J_{\text{bad}}} f(x), \phi(x)) \leq 2\tau W_J,$$

where

$$d_L(a, b) = \sum_{j \in L: a_j \neq b_j} w_j.$$

Consequently, after discarding at most another $O(\sqrt{\tau})$ fraction of the inputs, the pointwise distance is $O(\sqrt{\tau} W_J)$.

PROOF. Let

$$\varepsilon_j = \mathbb{P}_{x,y} [f_j(xy) \neq f_j(x)f_j(y)].$$

Fubini gives $\sum_j w_j \varepsilon_j \leq \tau W_J$. Put $J_{\text{bad}} = \{j : \varepsilon_j > \varepsilon_0\}$. Markov's inequality in the weighted counting measure gives the stated weight bound. For each remaining j , Lemma 2.20 supplies a homomorphism $\phi_j : G \rightarrow H_j$ and

$$\mathbb{P}_x [f_j(x) \neq \phi_j(x)] \leq 2\varepsilon_j.$$

Taking the product of the ϕ_j gives ϕ , and summing these last inequalities with weights w_j yields

$$\mathbb{E}_x d_{J \setminus J_{\text{bad}}}(f(x), \phi(x)) \leq 2 \sum_j w_j \varepsilon_j \leq 2\tau W_J.$$

The final pointwise assertion follows from Markov. $\square$

LEMMA 2.22 (Bounded centers cost negligible conductor weight). *Write $\bar{G}_p = G_p/Z(G_p)$ and let $\bar{\pi}$ denote the product quotient on any block of factors. For every $B \subseteq \prod_{p \in J} G_p$,*

$$|\text{Corr}_J(B) - \text{Corr}_J(\bar{\pi}B)| \leq 2 \sum_{p \in J} \log |Z(G_p)|.$$

Consequently, after enlarging the finite exceptional-prime set, the right-hand side is at most $\varepsilon W(J)$, uniformly in the number of prime factors.

PROOF. For each coordinate,

$$|\bar{\pi}_p B| \leq |\pi_p B| \leq |Z(G_p)| |\bar{\pi}_p B|,$$

while

$$|\bar{\pi}B| \leq |B| \leq \left(\prod_{p \in J} |Z(G_p)| \right) |\bar{\pi}B|.$$

Subtracting the two definitions of correlation defect gives the displayed bound. In $\text{SL}_2(\mathbb{F}_p)$ the center has order at most two, whereas $\log |G_p| = 3 \log p + O(1)$. Thus $\log |Z(G_p)| / \log |G_p| \rightarrow 0$ uniformly as $p \rightarrow \infty$, and summing over the coordinates proves the last assertion. $\square$

5. The large-block kernel/graph dichotomy

We now implement the squarefree product argument in the form most convenient for the later arbitrary-modulus induction. The philosophy is simple. Once a bounded power of A is surjective on a block L of large-branch factors, choose one lift of each $x \in G_L$. The failure of those lifts to multiply is measured by elements of $A^{O(1)} \cap \ker \pi_L$. Either that kernel contains substantial vertical freedom on the complementary block, or the chosen lifts define a weighted approximate homomorphism. The preceding lemma then converts the latter into an exact graph.

PROPOSITION 2.23 (Weighted kernel-or-graph alternative). *Let $I = L \sqcup S$ and let $A \subset G_I$ be a symmetric K -approximate subgroup with $1 \in A$. Assume $\pi_L(A^3) = G_L$. Fix $\tau > 0$, and write $\bar{G}_p = G_p/Z(G_p)$ for $p \in S$. Then one of the following holds.*

- (i) *There is $k \in A^9 \cap \ker \pi_L$ whose noncentral support*

$$T(k) = \{p \in S : k_p \notin Z(G_p)\}$$

has weight at least $\tau W(S)$.

- (ii) *There is $S_{\text{bad}} \subset S$ of weight $O(\tau W(S))$ and a genuine homomorphism*

$$\phi : G_L \longrightarrow \prod_{p \in S \setminus S_{\text{bad}}} \bar{G}_p$$

such that a subset of A^3 of cardinality at least $|G_L|/2$ lies in the $O(\sqrt{\tau} W(S))$ weighted neighborhood, modulo the bounded centers, of the graph of ϕ .

The constants in the $O(\cdot)$ terms are absolute once the threshold in Lemma 2.21 is fixed.

PROOF. For each $x \in G_L$ choose $a_x \in A^3$ with $\pi_L(a_x) = x$ and let $f(x)$ be the image of $\pi_S(a_x)$ in $\prod_{p \in S} \bar{G}_p$. For every x, y ,

$$k(x, y) = a_x a_y a_{xy}^{-1} \in A^9 \cap \ker \pi_L,$$

and in the central quotient

$$\bar{\pi}_S(k(x, y)) = f(x)f(y)f(xy)^{-1}.$$

If (i) fails, the weighted support of this multiplicative defect is less than $\tau W(S)$ for every pair (x, y) . Averaging gives the hypothesis of Lemma 2.21, now with targets $\bar{G}_p$. That lemma produces S_{bad} and ϕ . Its mean weighted error is $O(\tau W(S))$, so Markov's inequality shows that at least half the x have error $O(\sqrt{\tau} W(S))$. Their chosen lifts give (ii). Notice that central defects have disappeared for a legitimate reason rather than being counted as large directions: by Lemma 2.22, all centers together cost $o(W(S))$ in the final correlation bookkeeping. $\square$

LEMMA 2.24 (Weighted support balls are exponentially small). *Let S be a block of distinct allowed primes and put $\bar{G}_S = \prod_{p \in S} G_p / Z(G_p)$. For $0 < r < 1/3$ define*

$$\mathcal{B}_S(r) = \left\{ g \in \bar{G}_S : \sum_{p: g_p \neq 1} \log |G_p| \leq rW(S) \right\}.$$

For every $\varepsilon > 0$, after increasing the lower prime threshold,

$$|\mathcal{B}_S(r)| \leq \exp\{(2r + \varepsilon)W(S)\},$$

uniformly in the number of factors. The inverse image of this ball in G_S satisfies the same estimate with $3r$ in place of $2r$.

PROOF. Write $w_p = \log |G_p|$. For a support set $T \subset S$ with $W(T) \leq rW(S)$, the number of elements supported in T is at most $e^{W(T)}$. Since

$$1_{W(T) \leq rW(S)} \leq e^{2(rW(S) - W(T))},$$

we obtain

$$|\mathcal{B}_S(r)| \leq e^{2rW(S)} \sum_{T \subset S} e^{-W(T)} = e^{2rW(S)} \prod_{p \in S} (1 + e^{-w_p}).$$

For $G_p = \text{SL}_2(\mathbb{F}_p)$, $e^{-w_p} \ll p^{-3}$, so the logarithm of the product is $o(W(S))$ uniformly once the lower prime threshold tends to infinity. This proves the first estimate. Lifting from the central quotient multiplies the cardinality by at most $\prod_{p \in S} |Z(G_p)|$; by Lemma 2.22, this costs another $o(W(S))$, which is absorbed by replacing $2r + \varepsilon$ by $3r$. $\square$

PROPOSITION 2.25 (The graph branch decouples distinct full factors). *Assume the hypotheses of Proposition 2.23, and suppose its graph alternative holds while the vertical alternative fails. After enlarging the finite exceptional-prime set, if L and S contain pairwise distinct residue characteristics, then for every sufficiently small $\tau > 0$,*

$$\log |\pi_S(A^3)| \leq C\sqrt{\tau} W(S) + o(W(S)).$$

Consequently

$$\text{Corr}_I(A^3) \leq \text{Corr}_L(G_L) + \text{Corr}_S(\pi_S A^3) + C\sqrt{\tau} W(S) + o(W(S)).$$

In particular the exact graph produced between full simple residue factors is not a terminal obstruction in the distinct-characteristic squarefree case.

PROOF. Use the notation of Proposition 2.23. First observe that the homomorphism

$$\phi : G_L \longrightarrow \prod_{p \in S \setminus S_{\text{bad}}} \text{PSL}_2(\mathbb{F}_p)$$

is trivial once all primes exceed an absolute threshold. Indeed, restrict a coordinate map of ϕ to one source factor $\text{SL}_2(\mathbb{F}_\ell)$. Since this source is perfect and its only noncentral simple quotient is $\text{PSL}_2(\mathbb{F}_\ell)$, a nontrivial image would contain a copy of that simple group. The rank-one proper-subgroup classification from Volume IV shows that for prime target characteristic p and $\ell > 5$, such a subgroup can occur only when $p = \ell$. The blocks are disjoint, so every coordinate map is trivial.

Let $f : G_L \rightarrow \bar{G}_S$ be the chosen section coordinate from the proof of Proposition 2.23. Failure of the vertical alternative means that every multiplicative defect

$$d(x, y) = f(x)f(y)f(xy)^{-1}$$

has weighted support $< \tau W(S)$. The weighted correction lemma, together with the triviality of ϕ , gives a set $Q \subset G_L$ of relative size $1 - O(\sqrt{\tau})$ on which $f(x)$ has weighted support $O(\sqrt{\tau} W(S))$ outside S_{bad} ; the discarded block itself has weight $O(\tau W(S))$. For τ small, $|Q| > 3|G_L|/4$. Given arbitrary $x \in G_L$, choose $y \in Q \cap x^{-1}Q$. Then

$$f(x) = d(x, y) f(xy) f(y)^{-1},$$

so every $f(x)$ has weighted support at most $C\sqrt{\tau} W(S)$.

Now let $a \in A^3$ lie above $x \in G_L$ and compare it with the chosen lift a_x . Then $aa_x^{-1} \in A^6 \cap \ker \pi_L \subset A^9 \cap \ker \pi_L$. Again by failure of the vertical alternative, its noncentral support has weight $< \tau W(S)$. Thus every element of $\pi_S(A^3)$, modulo the bounded centers, lies in $\mathcal{B}_S(C\sqrt{\tau})$. Lemma 2.24 gives

$$\log |\pi_S(A^3)| \leq C\sqrt{\tau} W(S) + o(W(S)).$$

The displayed correlation estimate is now the exact identity

$$\text{Corr}_I(B) = \text{Corr}_L(\pi_L B) + \text{Corr}_S(\pi_S B) + \log |\pi_L B| + \log |\pi_S B| - \log |B|$$

with $B = A^3$, together with $|B| \geq |G_L|$. $\square$

LEMMA 2.26 (Regular-tree comparison after removing charged coordinates). *Let $A_0 \subset A$ be an exact regularized subset with branching degrees D_p , and let*

$$I = E \sqcup L \sqcup S$$

be a partition such that

$$D_p \geq |G_p|^{1-\delta} \quad (p \in L), \quad D_p \leq |G_p|^\delta \quad (p \in S).$$

Then

$$(10) \quad |A_0| \leq e^{W(E)} |G_L| e^{\delta W(S)}.$$

If moreover $|A_0| \geq e^{-R_0} |A|$, then

$$(11) \quad |A| \leq e^{R_0 + W(E) + \delta W(S)} |G_L|.$$

Consequently, if a bounded power of A contains $|G_L|R$ distinct elements, then

$$(12) \quad \frac{|A^C|}{|A|} \geq R \exp\{-R_0 - W(E) - \delta W(S)\}.$$

PROOF. Exact regularity gives $|A_0| = \prod_{p \in I} D_p$. On E we use only $D_p \leq |G_p|$, on L the trivial upper bound $D_p \leq |G_p|$, and on S the defining small-degree bound $D_p \leq |G_p|^\delta$. Multiplication gives (10). The second inequality follows from the regularization lower bound for $|A_0|$, and the last one follows by dividing the assumed lower bound $|A^C| \geq |G_L|R$ by (11). $\square$

PROPOSITION 2.27 (Regular-tree closure of the vertical branch). *Fix $\eta, \tau, \zeta > 0$. Choose $\delta > 0$ sufficiently small in terms of τ and then choose the allowed regularization and charged-coordinate losses sufficiently small in terms of $\eta\tau$. Suppose*

$$I = E \sqcup L \sqcup S, \quad W(S) \geq \eta W(I),$$

$A_0 \subset A$ is regular with the degree bounds of Lemma 2.26, and $\pi_L(A_0^3) = G_L$. Assume also $\log K \leq cW(I)$ for a sufficiently small $c = c(\eta, \tau)$. If the vertical alternative of Proposition 2.23 holds, then a bounded power of A has a macroscopic proper-section obstruction in the sense of Definition 2.4.

PROOF. Choose $k \in A^9 \cap \ker \pi_L$ whose noncentral support $T \subset S$ satisfies $W(T) \geq \tau W(S)$. The rank-one centralizer estimate gives

$$[G_I : C_{G_I}(k)] \geq \exp\{c_0 W(T)\}.$$

Choose

$$R = \exp\{\rho W(T)\}, \quad 0 < \rho < \min(c_0/2, \zeta/2).$$

If A meets fewer than R right cosets of $C_{G_I}(k)$, one such coset contains at least $e^{-\rho W(T)} |A| \geq e^{-\zeta W(T)} |A|$ points of A . On every $p \in T$ the centralizer is proper,

while $W(T) \geq \tau\eta W(I)$, so this is the macroscopic proper-section alternative with the prescribed capture loss.

Otherwise choose $a_1, \dots, a_R \in A$ in distinct right centralizer cosets. The conjugates $a_j k a_j^{-1}$ are distinct elements of $A^{11} \cap \ker \pi_L$. Since $\pi_L(A_0^3) = G_L$, choose one lift $b_x \in A_0^3$ for every $x \in G_L$. The products

$$b_x a_j k a_j^{-1} \quad (x \in G_L, 1 \leq j \leq R)$$

are all distinct: different x have different L -coordinates and, for fixed x , right multiplication by b_x is injective. Hence

$$|A^{14}| \geq |G_L|R.$$

Lemma 2.26 yields

$$\frac{|A^{14}|}{|A|} \geq \exp\{(\rho\tau - \delta)W(S) - R_0 - W(E)\}.$$

Take $\delta < \rho\tau/4$ and arrange $R_0 + W(E) \leq \rho\tau\eta W(I)/4$. Since $W(S) \geq \eta W(I)$, the right-hand side is at least $\exp\{c_1\rho\eta\tau W(I)\}$ for an absolute $c_1 > 0$. But a K -approximate group satisfies $|A^{14}| \leq K^{13}|A|$. Choosing $\log K \leq cW(I)$ with $13c < c_1\rho\eta\tau$ gives a contradiction. Thus the centralizer obstruction is unavoidable. $\square$

PROPOSITION 2.28 (Graph neighborhood gives a genuine obstruction). *Under the hypotheses and parameter choices of Proposition 2.27, with a prescribed capture-loss parameter $\zeta > 0$, suppose instead that the graph alternative of Proposition 2.23 holds. If $W(S)$ is macroscopic, then a bounded power of A has either a macroscopic proper-section obstruction or a subdirect common-section obstruction in the sense of Definition 2.4. If $W(S)$ is below the chosen error threshold, its entire contribution is absorbed into the low-correlation error.*

PROOF. The graph alternative supplies a homomorphism

$$\phi : G_L \longrightarrow \prod_{p \in S \setminus S_{\text{bad}}} \bar{G}_p$$

and at least $|G_L|/2$ chosen lifts lying within weighted distance $O(\sqrt{\tau}W(S))$ of its graph, after deleting $W(S_{\text{bad}}) = O(\tau W(S))$ and modulo bounded centers. By Lemma 2.24, all these lifts lie in at most

$$M \leq \exp\{C\sqrt{\tau}W(S) + o(W(S))\}$$

cosets of the exact lifted graph subgroup (with the discarded coordinates left unrestricted). One such coset therefore contains at least $|G_L|/(2M)$ of the selected lifts.

By Lemma 2.26 and $|A^C| \leq K^{C-1}|A|$, the relative mass of this coset inside A^C is at least

$$K^{-C} \exp\{-R_0 - W(E) - \delta W(S) - C\sqrt{\tau}W(S) - o(W(S))\}.$$

Because $W(S)$ is macroscopic, choose the graph-error parameter so that $C\sqrt{\tau}W(S)$, the small-degree loss $\delta W(S)$, and the regularization/charged-coordinate losses together are at most $(\zeta/2)W(S)$. After decreasing the small-tripling constant so that the remaining $K^{O(1)}$ losses are recorded separately, the displayed quantity is at least $K^{-C}e^{-\zeta W(S)}$.

If the exact graph has a nontrivial common section, this is the subdirect graph alternative. In the distinct-prime rank-one family every homomorphism from G_L to the product of the projective factors in the disjoint block S is trivial, by Lemma 1.8; after lifting the bounded centers the graph is therefore a proper product-section subgroup. Its index is $\exp\{(1 - o(1))W(S)\}$. When $W(S)$ is macroscopic this is the proper-section alternative. When $W(S)$ is smaller than the permitted error, we simply discard that block and pay $W(S)$ in the correlation budget. $\square$

REMARK 2.29 (Why no fiber-transport induction is needed). One might try to preserve an arbitrary pre-existing vertical fiber through a new conjugacy orbit. After the projection tree has been regularized and the mesoscopic coordinates charged away, this is unnecessary: every remaining vertical branching degree is at most $|G_p|^\delta$. Hence the *entire* old vertical mass is already bounded by $e^{\delta W(S)}$, and the source-faithful comparison in Proposition 2.27 is strong enough. This is the same structural point used in Varjú's large-degree argument: the large block is completed first, while a large-support kernel element supplies new vertical directions whose number is compared with the regular-tree small branching product.

LEMMA 2.30 (Largest-fiber kernel core). *Let $B \subset G_L \times G_S$ be nonempty and suppose $\pi_L B = G_L$. Put*

$$F_B = \frac{|B|}{|G_L|}.$$

Then there is a set

$$U \subset B^{-1}B \cap \ker \pi_L$$

with $1 \in U$, $U = U^{-1}$ after replacing it by $U \cup U^{-1}$, and

$$|U| \geq F_B.$$

Moreover, for every $V \subset \ker \pi_L$ and every set of representatives $\{b_x : x \in G_L\} \subset B$ with $\pi_L(b_x) = x$,

$$\left| \bigcup_{x \in G_L} b_x UV \right| = |G_L| |UV|.$$

PROOF. Write the fibers of B over G_L as B_x . Their average cardinality is F_B , so choose x_0 with $|B_{x_0}| \geq F_B$. Fix $b_0 \in B_{x_0}$ and put

$$U = b_0^{-1}B_{x_0} \subset B^{-1}B \cap \ker \pi_L.$$

Then $|U| = |B_{x_0}| \geq F_B$ and $1 \in U$. Replacing U by $U \cup U^{-1}$ changes no lower bound and keeps it inside a bounded product of $B^{-1}B$; in later applications this harmless bounded-power enlargement is absorbed into the exponent of A .

For distinct $x, y \in G_L$, the sets $b_x UV$ and $b_y UV$ have distinct L -projections, namely x and y . They are therefore disjoint, while each has cardinality $|UV|$. Summing over x proves the formula. $\square$

LEMMA 2.31 (Fiber amplification or trapped conjugacy orbit). *Let $B \subset G_L \times G_S$ satisfy $\pi_L B = G_L$, let U be the kernel core from Lemma 2.30, and let $V \subset \ker \pi_L$ have $|V| = R \geq 2$. Fix $0 < \theta < 1$. Then one of the following holds.*

(i)

$$|BUV| \geq |B|R^\theta;$$

(ii) *there are $v_0 \in V$ and $V_0 \subset V$ such that*

$$|V_0| \geq R^{1-\theta}, \quad V_0 v_0^{-1} \subset U^{-1}U \subset (B^{-1}B)^2.$$

In particular, if $B \subset A^m$ and $V \subset A^r$ for a symmetric K -approximate group A , then alternative (i) forces

$$R^\theta \leq K^{O_{m,r}(1)},$$

unless the ambient bounded product power of A grows by that factor.

PROOF. By Lemma 2.30, it is enough to consider UV . If $|UV| \geq |U|R^\theta$, the disjoint-fiber formula gives

$$|BUV| \geq |G_L||U|R^\theta \geq |B|R^\theta,$$

which is (i).

Assume instead that $|UV| < R^\theta|U|$. Apply the Ruzsa covering lemma from Volume II to the pair (U, V) . There is $X \subset V$ with $|X| < R^\theta$ such that

$$V \subset U^{-1}UX.$$

Hence one translate $U^{-1}Uv_0$ with $v_0 \in X$ contains at least $R/|X| \geq R^{1-\theta}$ elements of V . Let V_0 be those elements. Then $V_0 v_0^{-1} \subset U^{-1}U$, proving (ii). The final assertion uses $BUV \subset A^{O_{m,r}(1)}$. If $A^2 \subset XA$ with $|X| \leq K$, then induction gives $A^j \subset X^{j-1}A$ and hence $|A^j| \leq K^{j-1}|A|$; this gives the asserted bound. $\square$

PROPOSITION 2.32 (Vertical freedom: product growth or a trapped orbit). *Fix $\tau > 0$ and $0 < \theta < 1/2$. There are absolute constants $C, c > 0$ such that the following holds for all large allowed primes. Let A be as in Proposition 2.23, suppose its vertical alternative holds, and let $B = A^3$, so $\pi_L B = G_L$. Then, after replacing A by a bounded product power, at least one of the following occurs.*

- (i) *A has a subset of relative size $\exp\{-O(\tau W(S))\}$ in a coset of a proper product-centralizer obstruction supported on a block of weight at least $\tau W(S)$;*

(ii)

$$|A^C| \geq |B| \exp\{c\theta\tau W(S)\};$$

(iii) *there are $k \in A^9 \cap \ker \pi_L$, a block T with $W(T) \geq \tau W(S)$ on which every k_p is noncentral, and at least*

$$\exp\{c(1-\theta)\tau W(S)\}$$

distinct A -conjugates of k which, after right translation by one of them, lie in

$$(B^{-1}B)^4 \cap \ker \pi_L.$$

The third alternative is the precise remaining structural obstruction: a macroscopic portion of a noncentral conjugacy orbit is trapped inside one pre-existing vertical difference set.

PROOF. Choose k and $T = T(k)$ from the vertical alternative of Proposition 2.23. As in the previous version of this argument, the rank-one centralizer calculation gives

$$[G_I : C_{G_I}(k)] \geq \exp\{c_0 W(T)\}.$$

Put $H = C_{G_I}(k)$ and $R = \exp\{c_0 \tau W(S)/4\}$.

If A meets fewer than R right cosets of H , one of them contains at least $|A|/R$ elements of A . Since the projection of H to every $p \in T$ is the proper centralizer $C_{G_p}(k_p)$, this is alternative (i).

Otherwise choose one element of A from each of R distinct right cosets. Their conjugates of k form a set

$$V \subset A^{11} \cap \ker \pi_L, \quad |V| = R.$$

Apply Lemma 2.31 to $B = A^3$ and this V . In its first alternative,

$$|A^C| \geq |B|R^\theta \geq |B| \exp\{(c_0/4)\theta\tau W(S)\},$$

which is (ii). In its second alternative there is $V_0 \subset V$ of size at least $R^{1-\theta}$ whose translate lies in $(B^{-1}B)^2$; allowing the harmless symmetrization in Lemma 2.30 gives the displayed $(B^{-1}B)^4$ container. This is (iii), with $c = c_0/4$. $\square$

REMARK 2.33 (Why the old vertical mass must be retained). Constructing only R new kernel elements and then comparing $|G_L|R$ with the worst-case size $|G_S|$ is insufficient. That does *not* show a decrease of the correlation already present in the old vertical fibers. Lemma 2.30 restores the missing old fiber mass, and Lemma 2.31 shows exactly what can prevent multiplicative amplification: a large piece of the new conjugacy orbit must be trapped in the old vertical difference set. No injectivity of the multiplication map $U \times V \rightarrow UV$ is assumed.

LEMMA 2.34 (Macroscopic-stage restart lemma). *Fix $\eta, \rho, \tau, \beta > 0$. Consider a sequence of active conductor blocks of weights W_j and controlled sets A_j . Assume that every nonterminal stage is of one of the following two types.*

(a) vertical amplification: *a block of weight at least τW_j produces*

$$|A_{j+1}| \geq |A_j| \exp\{\rho\tau W_j\}, \quad W_{j+1} = W_j;$$

(b) graph removal:

$$W_{j+1} \leq (1 - \beta)W_j,$$

with only $\exp\{\varepsilon_j W_j\} K^{O(1)}$ loss, where $\sum_j \varepsilon_j \leq \eta/10$.

Suppose also that every bounded product power used during the process has size at most $K^{C_*} |A_0|$. Then, before the active weight falls below ηW_0 , the number of vertical stages is at most

$$\frac{C_* \log K}{\rho\tau\eta W_0},$$

and the number of graph-removal stages before $W_j \leq \eta W_0$ is at most

$$\left\lceil \frac{\log(1/\eta)}{-\log(1 - \beta)} \right\rceil.$$

Consequently the entire restart uses only $O_{\eta, \rho, \tau, \beta, C_*}(1)$ stages and pays only $K^{O_{\eta}(1)} e^{O(\eta W_0)}$ total loss.

PROOF. A vertical stage multiplies the current set size by at least $e^{\rho\tau W_j}$. Until the graph weight has fallen below ηW_0 we have $W_j \geq \eta W_0$, so r vertical stages would force a growth factor at least $e^{r\rho\tau\eta W_0}$. Since every set encountered is contained in a fixed bounded product power of A_0 , the approximate-group bound gives $e^{r\rho\tau\eta W_0} \leq K^{C_*}$ after the harmless regularization losses are absorbed. Thus

$$r \leq \frac{C_* \log K}{\rho\tau\eta W_0}.$$

In the intended application $\log K \leq cW_0$, so after c is fixed this is a constant depending only on the displayed parameters, and not on the number of prime factors.

Each graph-removal stage multiplies the active weight by at most $1 - \beta$. After m such stages the weight is at most $(1 - \beta)^m W_0$, so the displayed logarithmic bound makes it at most ηW_0 . Finally a bounded number of stages pays a bounded power of K , while

$$\prod_j e^{\varepsilon_j W_j} \leq e^{W_0 \sum_j \varepsilon_j} \leq e^{\eta W_0/10}.$$

Enlarging the error allowance absorbs fixed constants and proves the claim. $\square$

LEMMA 2.35 (Approximate groups inside a subgroup). *Let A be a symmetric K -approximate subgroup of a group G , let $H < G$, and fix an integer $r \geq 2$. Then*

$$Q = A^r \cap H$$

is a $K^{Or(1)}$ -approximate subgroup of H after replacing Q by Q^2 if necessary. More precisely, Q^2 is covered by at most K^{2r-1} left translates of $A^2 \cap H \subset A^r \cap H$.

PROOF. Since A is a K -approximate group, A^{2r} is covered by at most K^{2r-1} left translates $x_j A$. Intersect this cover with Q^2 . For every translate $x_j A$ that meets Q^2 , choose

$$h_j = x_j a_j \in x_j A \cap Q^2.$$

If $x_j a \in x_j A \cap Q^2$, then both h_j and $x_j a$ lie in H , and

$$h_j^{-1} x_j a = a_j^{-1} a \in A^2 \cap H \subset Q.$$

Hence

$$Q^2 \cap x_j A \subset h_j Q,$$

with every covering centre h_j itself in Q^2 . Therefore Q^2 is covered by at most K^{2r-1} translates of Q , which is the standard approximate-group criterion. The set $A^2 \cap H$ is symmetric and contains the identity, so the usual equivalent definition of approximate groups gives the claim, with parameter $K^{O_r(1)}$. $\square$

LEMMA 2.36 (Trapped-orbit consolidation by proper-block induction). *Fix $\eta > 0$. Assume that the many-factor product-growth theorem BG-X.2.006 has already been proved for every product of fewer than n residue factors. Let $I = L \sqcup S$ have n factors with $L \neq \emptyset$, let A be a symmetric K -approximate subgroup of G_I , and assume $\pi_L(A^3) = G_L$. Suppose the trapped-orbit alternative of Proposition 2.32 holds on S , so that for some $\alpha = \alpha(\eta) > 0$ a bounded product power of A contains a set of at least*

$$e^{\alpha W(S)}$$

distinct elements of $\ker \pi_L$.

If $W(S) \leq \eta W(I)/4$, then the whole S -block may be absorbed into the allowed correlation error. Otherwise, after choosing the small-tripling constant sufficiently small in terms of η and α , one of the following holds after a bounded product power:

(i)

$$\text{Corr}_I(A^C) \leq \eta W(I)/2 + C_\eta \log K;$$

(ii) A^C has a macroscopic proper-section obstruction;

(iii) A^C has a genuine subdirect/common-section obstruction.

PROOF. The small-weight case is immediate from the crude bound $\text{Corr}_S(B) \leq W(S)$ for every $B \subset G_S$, together with the fact that the L -block is already full. We therefore assume $W(S) > \eta W(I)/4$.

Choose a fixed exponent r large enough that the trapped orbit and the vertical difference set both lie in

$$Q = A^r \cap \ker \pi_L.$$

By Lemma 2.35, Q is a K_Q -approximate subgroup of G_S with

$$(2.13) \quad \log K_Q \leq C_r \log K,$$

and the trapped orbit gives

$$(2.14) \quad |Q| \geq e^{\alpha W(S)}.$$

Choose $0 < \delta < \alpha/2$.

If

$$(2.15) \quad |Q| \geq e^{(1-\delta)W(S)},$$

then choose one lift $b_x \in A^3$ above every $x \in G_L$. The sets $b_x Q$ are disjoint and lie in a bounded power A^C . Hence

$$|A^C| \geq |G_L| |Q| \geq e^{W(L) + (1-\delta)W(S)}.$$

Every one-coordinate projection of A^C is contained in the corresponding ambient factor, so

$$\text{Corr}_I(A^C) \leq W(I) - \log |A^C| \leq \delta W(S) \leq \eta W(I)/2$$

when δ is chosen below $\eta/2$. This is (i).

It remains to treat

$$(2.16) \quad e^{\delta W(S)} \leq |Q| < e^{(1-\delta)W(S)},$$

where the lower bound follows from (2.14). Apply the inductive BG-X.2.006 theorem to the approximate group Q inside the strictly smaller product G_S . Its growth alternative would give

$$|Q^3| \geq e^{c_\delta W(S)} |Q|.$$

But Q is a K_Q -approximate group, so $|Q^3| \leq K_Q^2 |Q|$. Since $W(S) > \eta W(I)/4$, (2.13) and $\log K \leq cW(I)$ imply

$$2 \log K_Q \leq 2C_r cW(I) < c_\delta W(S)$$

provided the global constant $c = c(\eta)$ is chosen sufficiently small. Thus the growth alternative is impossible. The remaining alternatives in the inductive product theorem give a proper local section/subgroup obstruction or a genuine subdirect/common-section obstruction *inside the vertical kernel* $Q \subset G_S$. This is a genuine structural reduction, but one last transport step is required before it becomes alternative (ii) or (iii) for the whole set A^C : the relevant coset in G_S may depend on the chosen L -fiber. Merely saying “adjoin the unrestricted L -coordinates” would lose the required relative-mass bound. The following lemma records the exact quantitative transport statement needed at this point. $\square$

LEMMA 2.37 (Macroscopic-index vertical obstruction lifts globally). *Assume the notation of Lemma 2.36, with $W(S) > \eta W(I)/4$. Let $Q = A^r \cap \ker \pi_L$, and suppose that for a proper subgroup/section $H_S < G_S$ a set $Q_0 \subset Q^C$ satisfies*

$$|Q_0| \geq e^{-aW(S)} |Q|, \quad Q_0 \subset gH_S, \quad [G_S : H_S] \geq e^{bW(S)}.$$

If moreover $|Q| \geq e^{\alpha W(S)}$, then a bounded power of A contains a subset of relative size at least $e^{-C_{\eta, \alpha} W(S)}$ in the proper coset

$$G_L \times gH_S < G_L \times G_S.$$

Thus every inductive obstruction whose index is exponential in its conductor weight already gives the macroscopic proper-section obstruction of Definition 2.4.

PROOF. The set Q_0 itself is contained in a bounded power $A^{C'}$, and, since it lies in the vertical kernel, it is contained in the displayed product coset. Its absolute size is at least

$$|Q_0| \geq \exp\{(\alpha - a)W(S)\}.$$

On the other hand every subset of G_I has size at most

$$|G_I| = e^{W(I)} \leq e^{4W(S)/\eta}.$$

Consequently

$$\frac{|Q_0|}{|A^{C'}|} \geq \exp\left\{-\left(\frac{4}{\eta} + a - \alpha\right)W(S)\right\}.$$

After enlarging the constant this is exactly the exponential captured-mass scale allowed in the proper-section alternative. The index hypothesis gives the required macroscopic obstruction conductor. Notice that no alignment between different L -fibers is needed in this large-index case: one vertical fiber is already quantitatively large enough because S itself has macroscopic conductor weight. $\square$

PROPOSITION 2.38 (Fiberwise obstruction closure). *After global pruning, exact regularization, and removal of the mesoscopic block, Propositions 2.27 and 2.28 close respectively the vertical and graph branches without transporting a subgroup coset from one L -fiber to another.*

PROOF. The vertical branch is closed by comparing the number of centralizer conjugates with the full regular-tree size bound of Lemma 2.26; no old-fiber multiplication is needed. The graph branch is closed by covering the weighted graph neighborhood by $e^{O(\sqrt{\tau}W(S))}$ exact graph cosets and pigeonholing one of them. These are precisely the two possible residual branches, and together they complete the required fiberwise obstruction analysis. $\square$

REMARK 2.39 (Why no simultaneous induction is needed). The regular-tree comparison above makes a proper-block transport recursion unnecessary: after mesoscopic charging, the residual small-degree block has total old fiber mass at most $e^{\delta W(S)}$, so a large vertical conjugacy orbit can be compared directly with the whole regularized set. The theorem below therefore closes in one global large/small-degree pass.

THEOREM 2.40 (Conductor-sensitive approximate Goursat; BG-X.2.007). *For every $\eta > 0$ and every capture-loss parameter $\zeta > 0$ there are constants $C = C(\eta, \zeta)$, $c = c(\eta, \zeta) > 0$, and $p_0(\eta, \zeta)$ with the following conclusion. Let*

$$P = \prod_{p \in I} G_p$$

with distinct allowed primes $p \geq p_0(\eta)$, and let $A \subset P$ be a K -approximate subgroup with $\log K \leq cW(I)$. After a bounded product power, at least one of the following alternatives must occur.

(i) Low correlation:

$$\text{Corr}_I(A^C) \leq C \log K + \eta W(I).$$

(ii) Macroscopic proper-section obstruction: *there is $J \subset I$ with $W(J) \geq c(\eta)W(I)$, proper subgroups $H_p < G_p$ for $p \in J$, and a subset $A_0 \subset A^C$ of relative size at least $\exp\{-\zeta W(J)\}K^{-C}$ contained in one coset of*

$$\left(\prod_{p \in J} H_p \right) \times \left(\prod_{p \notin J} G_p \right).$$

Equivalently, one obtains a bounded-complexity local subgroup/section obstruction of index at least $\exp\{cW(J)\}$.

(iii) Subdirect graph obstruction: *after deleting a block E with $W(E) \leq \eta W(I)$, there are disjoint nonempty blocks J, K and subgroups $U \leq G_J$, $V \leq G_K$ with normal subgroups $N_U \triangleleft U$, $N_V \triangleleft V$ and a nontrivial isomorphism*

$$U/N_U \simeq V/N_V$$

such that a subset of A^C of relative size at least $e^{-\zeta W(J \cup K)}K^{-C}$ lies in the associated fiber-product subgroup after projection away from E .

The alternatives are stable under passage to coordinate quotients and under replacing A by a bounded product power.

REMARK 2.41 (Why proper local sections are necessary). A conductor-sensitive approximate-Goursat/correlation-control theorem cannot restrict its obstructions to quotients of the *full ambient* simple factors. For example, two distinct residue factors may contain correlated proper tori with a common cyclic section even though the ambient groups themselves have no common nontrivial simple quotient. The proper-section alternative above is therefore mathematically necessary. In the later Bourgain–Gamburd application it is eliminated by the index-sensitive subgroup-escape machinery of Chapter 3, not by Goursat alone.

EXAMPLE 2.42 (Why proper local sections must be allowed). Take the split tori $T_{101} < \text{SL}_2(\mathbb{F}_{101})$ and $T_{151} < \text{SL}_2(\mathbb{F}_{151})$. They are cyclic of orders 100 and 150,

respectively, and both admit a quotient C_{50} . Choose surjections $\theta_{101} : T_{101} \rightarrow C_{50}$ and $\theta_{151} : T_{151} \rightarrow C_{50}$ and form the fiber product

$$H = \{(x, y) \in T_{101} \times T_{151} : \theta_{101}(x) = \theta_{151}(y)\}.$$

Then H is an exact subgroup, hence a 1-approximate group, its two coordinate projections are the full tori, and

$$|H| = \frac{|T_{101}||T_{151}|}{50}, \quad \text{Corr}_{\{101, 151\}}(H) = \log 50.$$

On the other hand the ambient simple quotients $\text{PSL}_2(\mathbb{F}_{101})$ and $\text{PSL}_2(\mathbb{F}_{151})$ are nonisomorphic and have no common nontrivial quotient. Thus ambient Goursat alone cannot see this correlation; the obstruction lives in proper local sections. This elementary example is the reason alternatives (ii) and (iii) are separated above.

REGULAR-TREE PROOF WITH SOURCE-FAITHFUL LARGE-BLOCK CLOSURE.

Fix a small hierarchy of parameters in the order

$$0 < \kappa \ll \zeta, \quad 0 < \tau \ll (\eta\zeta)^2, \quad 0 < \delta \ll \eta\tau\zeta.$$

Take the lower prime threshold large enough that all regularization, center, and fixed-per-prime losses used below are smaller than the corresponding η - and ζ -budgets. Finally choose the constant in $\log K \leq cW(I)$ sufficiently small. The role of η is to control correlation loss; the independent parameter ζ controls how much mass may be lost when an obstruction is captured.

Apply Lemma 2.9. If the bad block has conductor weight at least a fixed multiple of $\eta W(I)$, choose the pruning threshold κ above. The surviving set furnished by that lemma then has relative mass $e^{-\kappa W(J_{\text{bad}})} \geq e^{-\zeta W(J_{\text{bad}})}$ and lies in one product of proper local subgroup cosets of exponential index. This is alternative (ii). Thus we may assume the bad block has weight $O(\eta W(I))$.

Choose an exact regularization $A_0 \subset A$ with degrees D_p and regularization logarithmic loss $R_0 = o(W(I))$. Proposition 2.14 shows that the mesoscopic block

$$M = \{p : |G_p|^\delta \leq D_p \leq |G_p|^{1-\delta}\}$$

has

$$W(M) \ll_\eta \log K + o(W(I)).$$

If this bound is not inside the permitted error budget, decrease the constant c in $\log K \leq cW(I)$. Put all bad and mesoscopic coordinates into a charged set E . On the remaining coordinates write

$$I \setminus E = L \sqcup S,$$

where $D_p > |G_p|^{1-\delta}$ on L and $D_p < |G_p|^\delta$ on S . Lemma 2.19 gives

$$\pi_L(A_0^3) = G_L.$$

If $W(S) \leq \eta W(I)/10$, there is already low correlation. Indeed $A_0 \subset A^C$ and exact regularity give

$$\text{Corr}_I(A^C) \leq W(I) - \log |A^C| \leq W(I) - \log |A_0| \leq W(E) + \delta W(L) + W(S),$$

which is $O_\eta(\log K) + \eta W(I)$ after the chosen parameter hierarchy. This is alternative (i).

Assume therefore that $W(S)$ is macroscopic. Apply Proposition 2.23. In the vertical alternative, Proposition 2.27 shows that bounded-power growth would exceed the $K^{O(1)}$ growth permitted for an approximate group unless a macroscopic centralizer/product-section obstruction occurs. Hence we obtain alternative (ii).

In the graph alternative, Proposition 2.28 shows that the weighted graph neighborhood contains a quantitatively large subset of one exact graph coset. A nontrivial common section gives alternative (iii). For the present distinct-prime rank-one family the homomorphism is trivial on every projective factor, so the same conclusion is instead a macroscopic proper product-section obstruction, alternative (ii).

Thus every branch ends in (i), (ii), or (iii). All losses are charged either to $W(E) = O_\eta(\log K) + o(W(I))$ or to the fixed small parameters $\delta, \sqrt{\tau}$ times $W(I)$; there is no factor depending on $\omega(q)$. Stability under coordinate quotients and bounded product powers follows from the projection bookkeeping of Chapter 1 and the fact that all estimates above change K only by a bounded power. This proves BG-X.2.007. $\square$

6. Many-factor product growth

THEOREM 2.43 (Many-factor product growth; BG-X.2.006). *Fix $\delta > 0$ and a capture-loss parameter $\zeta > 0$. There are $c, C > 0$ and p_0 such that, for every squarefree q supported on primes $p \geq p_0$ and every symmetric $A \subset G(q)$ with $1 \in A$ and*

$$e^{\delta W(q)} \leq |A| \leq e^{(1-\delta)W(q)},$$

one of the following holds:

- (i) $|A^3| \geq e^{cW(q)} |A|$;
- (ii) *a subset of A^C of relative size at least $e^{-\zeta W(d)}$ is contained in a coset of a proper subgroup/section obstruction whose exact conductor d has weight at least $cW(q)$;*
- (iii) *A^C has a genuine subdirect/common-section correlation obstruction of the kind in Theorem 2.40.*

COMPLETION OF THE SIMULTANEOUS INDUCTION. Assume (i) fails. The small-tripling-to-approximate-group machinery of Volume II replaces A by a controlling K_0 -approximate group P , with $\log K_0 = O(cW(q))$, at the cost of $e^{O(cW(q))}$ translates. Any macroscopic proper-section or subdirect obstruction for P therefore pigeonholes back to (ii) or (iii) for a bounded power of A .

Choose auxiliary parameters

$$0 < \eta_0 \ll \delta, \quad 0 < \delta_0 \ll \delta,$$

and make the small-tripling constant c sufficiently small in terms of them. Regularize P with logarithmic loss at most $\eta_0 W(q)$ and apply Lemma 2.9. A bad block of weight at least $\eta_0 W(q)$ gives (ii). Otherwise Proposition 2.14 and $\log K_0 = O(cW(q))$ show that the mesoscopic block also has weight at most $\eta_0 W(q)$. Put both into the charged set E . Outside E the regular tree has only large degrees $D_p > |G_p|^{1-\delta_0}$ and small degrees $D_p < |G_p|^{\delta_0}$; write the corresponding blocks as L and S . The large block is completed simultaneously by Lemma 2.19.

Now run the proof of Theorem 2.40 on this same regular tree, with correlation parameter η_0 and capture parameter ζ . Its proper-section and subdirect branches give (ii) and (iii). The remaining branch is not merely an abstract low-correlation statement: in the explicit regular-tree proof it occurs at the small-block stopping line

$$W(S) \leq \eta_0 W(q)/10.$$

Consequently

$$W(L) \geq (1 - 3\eta_0)W(q)$$

after absorbing the bad, mesoscopic, and regularization losses. Exact regularity therefore gives

$$\log |P| \geq \log |P'| \geq (1 - \delta_0)W(L) - \eta_0 W(q) \geq (1 - \delta/2)W(q)$$

when η_0, δ_0 are chosen sufficiently small in terms of δ . But the controlling approximate group produced by the BSG step has $|P| \leq e^{O(cW(q))}|A|$; shrinking c once more and using $|A| \leq e^{(1-\delta)W(q)}$ gives

$$\log |P| \leq (1 - 3\delta/4)W(q),$$

a contradiction. Thus the small-block/low-correlation exit cannot occur in the intermediate-size regime.

BG-X.2.007 is the only same-level structural input in this last step, and it has now been proved above by the regular-tree vertical/graph closure. Hence BG-X.2.006 is closed as well. $\square$

REMARK 2.44 (Relation with the historical proofs). The regular-tree/entropy language is a pedagogical reorganization of the squarefree product-growth mechanism. Varjú's squarefree theorem and the later perfect-group theorem are the historical sources for the multifactor architecture (Varjú; Salehi Golsefidy–Varjú). Here the only logical inputs are earlier volumes plus the explicit regularization and stability arguments above.

CHAPTER 3

Index-Sensitive Escape from Squarefree Subgroup Obstructions

For a prime quotient, it is enough to say that every proper subgroup coset receives at most $p^{-\kappa}$ mass. For a squarefree quotient that normalization is wrong: a subgroup may be only mildly defective on a small conductor or highly defective on many factors. The correct scale is its index, or equivalently the conductor weight of the actual obstruction.

1. Index and obstruction conductor

PROPOSITION 3.1 (Index lower bound from the exact obstruction; BG-X.3.001). *There are constants $c_0, C_0 > 0$ such that for every squarefree q supported on allowed primes and every proper subgroup $H < G(q)$ there is a divisor $d \mid q$, $d > 1$, with the following properties:*

- (i) $\pi_d(H)$ is a proper subgroup obstruction, either by a proper local projection or by a common-quotient correlation;
- (ii)
- (iii) if $J = \{p : \pi_p(H) \neq G_p\}$, then one may take the local part of d to contain J and

$$[G(q) : H] \geq \prod_{p \in J} [G_p : \pi_p(H)].$$

For the distinct-prime rank-one family there is no exact correlation-only obstruction, so a proper H has $J \neq \emptyset$.

PROOF. The injection

$$H \hookrightarrow \prod_{p \mid q} \pi_p(H)$$

gives

$$|H| \leq \prod_{p \mid q} |\pi_p(H)|,$$

which proves (iii). The prime-field subgroup hierarchy of Volume IV gives $[G_p : \pi_p(H)] \gg p^{c_1}$ whenever the projection is proper, with an absolute $c_1 > 0$. Since

$W(p) = 3 \log p + O(1)$, multiplying over $p \in J$ gives (ii) for the local conductor $d = \prod_{p \in J} p$.

If all projections are full, Theorem 1.9 says that a proper subdirect subgroup must be a fiber product over a common quotient. Its index is at least the order of that quotient. In the rank-one squarefree family the common quotient is impossible; the more general sentence records the scale used by the approximate argument later. $\square$

2. A CRT version of the short-word law argument

Earlier input (Volumes VII and VIII). Volume VII proves a fixed bounded-word free pair, exponential height control, and injectivity of reduction on a free ball of radius $c \log p$. Volume VIII proves that words lying in one coset of a proper subgroup of G_p obey one of finitely many bounded-complexity group-law certificates; the number of free words satisfying such a certificate is exponentially smaller than the total free ball. We now make the certificate simultaneous over many primes.

LEMMA 3.2 (Height multiplication across primes). *Fix integral matrices S and a nonzero matrix-valued word-polynomial expression P obtained from at most r words of length at most n in S . There is $C_S > 1$ such that every integer entry of P has absolute value at most C_S^{rn} . If $P \equiv 0 \pmod{p}$ for every $p \mid d$ and*

$$d > C_S^{rn},$$

then $P = 0$ over $\mathbb{Z}$.

PROOF. Every entry of a word of length n is bounded by C_1^n for a constant depending only on the generators. A fixed expression formed from r such words using boundedly many additions and multiplications has entries bounded by C_S^{rn} . Divisibility by every prime factor of the squarefree integer d means divisibility by d . An integer of absolute value smaller than d and divisible by d is zero. $\square$

LEMMA 3.3 (Simultaneous law-counting). *There exist constants $a, b > 0$ and a finite list of group-word certificates $\mathcal{L}$ with the following property. Let $\mathcal{W}_n$ be the reduced words of length at most n in the fixed free pair of Volume VII. Suppose d is squarefree, supported on allowed primes, and for each $p \mid d$ all words in a set $E \subset \mathcal{W}_n$ land, after reduction modulo p , in one coset of a proper subgroup $H_p < G_p$. If*

$$n \leq a \log d,$$

then

$$|E| \leq e^{bn}$$

where b is strictly smaller than the exponential growth rate of the free ball.

PROOF. For each p , the Volume-IV subgroup hierarchy and the Volume-VIII coset-law argument attach to H_p one of finitely many certificate types: a fixed second-derived/metabelian word after bounded-index pigeonholing, or one of finitely

many bounded exceptional certificates. Partition the prime factors of d by certificate type. One class has logarithmic weight at least a fixed fraction of $\log d$; call the product of those primes d_1 .

The coset-law proof in Volume VIII is based on a bounded tuple of words from E : taking suitable differences cancels the coset representative, and the corresponding certificate $L \in \mathcal{L}$ evaluates to the identity modulo p . The same bounded expression is therefore the identity modulo every $p \mid d_1$. By Lemma 3.2, if a is sufficiently small then it is an exact identity in the integral/free model.

Volume VIII proves, using cyclic centralizers in a free group and the nontriviality of these certificates, that the number of words in a free ball participating in such an exact law configuration grows like e^{bn} with b strictly below the free-ball exponent. The bounded-index and finite-certificate pigeonholing changes only the multiplicative constant. This gives the assertion. $\square$

THEOREM 3.4 (Squarefree short-word escape; BG-X.3.002). *There are $c, \kappa > 0$ such that for every squarefree q , every proper $H < G(q)$, every coset xH , and every free-ball time*

$$n = \lfloor c \log[G(q) : H] \rfloor,$$

the fixed symmetric random walk $\mu_q^{(n)}$ satisfies

$$\mu_q^{(n)}(xH) \leq [G(q) : H]^{-\kappa}.$$

The constants depend only on the original generating set and the finite excluded-prime set.

PROOF. Use Proposition 3.1 to choose an obstruction conductor d with

$$W(d) \ll \log[G(q) : H].$$

A word whose image lies in xH also lies in the corresponding coset of $\pi_d(H)$ modulo every $p \mid d$. At the chosen small constant c , the time n is within the logarithmic injectivity range supplied by Volume VII and within the range of Lemma 3.3.

Before the first collision, the probability of a reduced word is exponentially comparable to the inverse of the free-ball growth; the precise tree norm estimate is proved in Volume VII. Lemma 3.3 says that only an exponentially smaller family of words can lie in the obstruction coset. Thus

$$\mu_q^{(n)}(xH) \leq e^{-c_1 n}.$$

Since $n \asymp \log[G(q) : H]$, this is the desired power of the index.

If the obstruction is a correlation fiber product rather than a local subgroup, the same proof uses the common-quotient word certificate: two coordinate blocks agree after applying the quotient maps. Their difference is a bounded matrix word expression vanishing modulo every prime in the correlation conductor, and the height argument again promotes it to an exact free-group relation. In the distinct-prime rank-one family this branch is absent by Theorem 1.9. $\square$

3. From short words to logarithmic-time random walk

THEOREM 3.5 (Index-sensitive random-walk coset bound; BG-X.3.003). *There are $c_*, \kappa > 0$ such that for every allowed squarefree q , every proper $H < G(q)$, every coset xH , and*

$$n_0 = \max \{1, \lfloor c_* \log[G(q) : H] \rfloor\},$$

one has

$$\mu_q^{(n_0)}(xH) \leq [G(q) : H]^{-\kappa}.$$

The constant c_* is chosen inside the logarithmic height/injectivity window. The same estimate then holds for every later time $n \geq n_0$.

PROOF. Theorem 3.4 gives the estimate at the displayed time after decreasing c_* if necessary. Since $[G(q) : H] \leq |G(q)|$, this time is automatically inside a fixed small multiple of $\log |G(q)|$, exactly the range in which the height argument was proved.

For persistence, define

$$M_H(\nu) = \sup_x \nu(xH).$$

For any probability measure σ ,

$$(\sigma * \nu)(xH) = \sum_g \sigma(g) \nu(g^{-1}xH) \leq M_H(\nu),$$

so $M_H(\sigma * \nu) \leq M_H(\nu)$. Convolving by further walk steps therefore cannot increase the maximal coset mass. $\square$

COROLLARY 3.6 (Persistence under convolution; BG-X.3.004). *If a probability measure ν on $G(q)$ satisfies*

$$\sup_x \nu(xH) \leq [G(q) : H]^{-\kappa}$$

for every proper subgroup H , then for every probability measure σ the convolution $\sigma * \nu$ satisfies the same inequalities. The same holds for $\nu * \sigma$ when the family of obstructions is closed under conjugacy, as it is here.

PROOF. The left-convolution assertion is the calculation in the proof of Theorem 3.5. For right convolution,

$$(\nu * \sigma)(xH) = \sum_g \sigma(g) \nu(xHg^{-1}) = \sum_g \sigma(g) \nu(xg^{-1}(gHg^{-1})),$$

and conjugate subgroups have the same index. $\square$

PROPOSITION 3.7 (Passage to the exact conductor; BG-X.3.005). *Let an obstruction $\mathcal{O}$ in $G(q)$ factor through $G(d)$ with exact conductor $d \mid q$. If $\nu_d = (\pi_d)_* \nu$, then the mass of every lift of an obstruction coset equals the corresponding mass under ν_d .*

Consequently all index-sensitive non-concentration estimates may and should be proved on the exact conductor rather than normalized by $|G(q)|$.

PROOF. A lifted obstruction set has the form $\pi_d^{-1}(E)$. By definition of push-forward,

$$\nu(\pi_d^{-1}(E)) = \nu_d(E).$$

The index of $\pi_d^{-1}(H_d)$ in $G(q)$ equals the index of H_d in $G(d)$ because the full kernel of π_d is contained in the lift. Hence both the probability and its natural normalization live at conductor d . $\square$

EXAMPLE 3.8 (Why ambient normalization is false). Let $q = pr$ and let

$$H = H_p \times G_r$$

with $H_p < G_p$ proper. Then the event $g \in H$ depends only on the p -coordinate. A bound normalized by $|G(q)|^{-c}$ would spuriously become stronger as $r \rightarrow \infty$, although the event has not changed. The correct quantity is

$$[G(q) : H] = [G_p : H_p],$$

exactly as Proposition 3.7 prescribes.

CHAPTER 4

Squarefree Inverse Theory and L^2 Flattening

The analytic inverse theorem from Volume II is modulus-blind: failed flattening always produces an approximate group. What changes at squarefree level is the structural theorem used to destroy that approximate group. Chapter 2 supplies many-factor growth and Chapter 3 supplies conductor-sensitive escape. Their combination gives a flattening theorem with no exponent loss depending on the number of prime factors.

1. The dyadic inverse set

Earlier input (Volume II). Volume II proves the weighted failed-flattening inverse theorem. In the form used here: if a symmetric probability measure ν satisfies

$$\|\nu * \nu\|_2 \geq K^{-1} \|\nu\|_2,$$

then a bounded dyadic regularization produces a symmetric $K^{O(1)}$ -approximate group A , a translate xA , and a dyadic level of ν such that $\nu(xA) \geq K^{-O(1)}$ and $|A| \leq K^{O(1)} \|\nu\|_2^{-2}$. A lower bound on $|A|$ comes from an independent atom/non-concentration estimate, not from the inverse theorem itself.

PROPOSITION 4.1 (Squarefree-compatible inverse extraction; BG-X.4.001). *Let ν be a symmetric probability measure on $G(q)$ and suppose*

$$\|\nu * \nu\|_2 \geq K^{-1} \|\nu\|_2.$$

Then there is a symmetric K^C -approximate subgroup $A \subset G(q)$ and $x \in G(q)$ such that

$$\nu(xA) \geq K^{-C}, \quad |A| \leq K^C \|\nu\|_2^{-2}.$$

For every divisor $d \mid q$, $\pi_d(A)$ is a K^C -approximate subgroup of $G(d)$, and the same construction can be performed after pushing ν to $G(d)$ with constants independent of d and of $\omega(q)$.

PROOF. Apply the Volume-II theorem. The projection of a K^C -approximate group is again a K^C -approximate group because if

$$A^2 \subset XA, \quad |X| \leq K^C,$$

then

$$\pi_d(A)^2 \subset \pi_d(X)\pi_d(A).$$

Push-forward commutes with convolution, and dyadic decomposition is performed on the values of the measure rather than on the arithmetic factorization of the group. Thus the number of prime factors never enters the inverse extraction. $\square$

2. Transfer of subgroup escape to the inverse set

LEMMA 4.2 (Coset covering forces walk concentration). *Suppose $A \subset G(q)$ is covered by L left cosets of a subgroup H and $\nu(xA) \geq \alpha$. Then some coset of H carries ν -mass at least α/L .*

PROOF. Write $A \subset \bigcup_{j=1}^L y_j H$. Then

$$\alpha \leq \nu(xA) \leq \sum_{j=1}^L \nu(xy_j H),$$

so one summand is at least α/L . $\square$

PROPOSITION 4.3 (Transfer to the inverse set; BG-X.4.002). *Fix $\kappa > 0$. Suppose ν satisfies*

$$\sup_x \nu(xH) \leq [G(q) : H]^{-\kappa}$$

for every proper subgroup $H < G(q)$. Let A be the approximate group extracted in Proposition 4.1. If

$$\nu(xA) \geq K^{-C_0},$$

then A cannot be covered by K^{C_1} cosets of a proper subgroup of index larger than K^{C_2} , where C_2 may be chosen once C_0, C_1, κ are fixed. The same assertion holds at every exact conductor.

PROOF. If A were covered by K^{C_1} cosets of H , Lemma 4.2 would give a coset with mass at least $K^{-C_0-C_1}$. The assumed escape bound gives at most $[G : H]^{-\kappa}$. These are incompatible once

$$[G : H] > K^{(C_0+C_1)/\kappa}.$$

Take $C_2 > (C_0 + C_1)/\kappa$. Proposition 3.7 gives exactly the same calculation after projecting to the conductor of the obstruction. $\square$

3. One-step squarefree flattening

THEOREM 4.4 (Squarefree structural flattening; BG-X.4.003). *For every $\delta > 0$ there are $\varepsilon = \varepsilon(\delta) > 0$, $c > 0$, and p_0 such that the following holds. Let q be squarefree with all $p \mid q$ at least p_0 , and let ν be a symmetric probability measure on $G(q)$. Assume*

(i) for every divisor $d \mid q$ and every proper $H < G(d)$,

$$\sup_x v_d(xH) \leq [G(d) : H]^{-c};$$

(ii) the effective support is intermediate:

$$|G(q)|^{-1+\delta} \leq \|v\|_2^2 \leq |G(q)|^{-\delta}.$$

Then

$$\|v * v\|_2 \leq \|v\|_2^{1+\varepsilon}.$$

All constants are independent of $\omega(q)$.

PROOF. Assume the conclusion fails. Put

$$K = \|v\|_2^{-\varepsilon}.$$

For sufficiently small ε the failure has the form required by Proposition 4.1, so we obtain a $K^{O(1)}$ -approximate group A carrying $K^{-O(1)}$ mass and satisfying

$$|A| \leq K^{O(1)} \|v\|_2^{-2}.$$

The atom bound implicit in (i), applied to the trivial subgroup at the active conductor, together with the Volume-II anti-atom lower-size lemma gives

$$|A| \geq K^{-O(1)} \|v\|_2^{-2}.$$

Thus (ii) places A in an intermediate exponential size range, after choosing $\varepsilon \ll \delta$.

Let c_0 be the conductor-index constant from Proposition 3.1 and let c_{esc} be the power in hypothesis (i). Before invoking the many-factor theorem choose its capture-loss parameter $\zeta > 0$ so small that

$$4\zeta < c_{\text{esc}} c_0.$$

Then choose the flattening-failure exponent ε still smaller, so that every $K^{O(1)}$ loss produced by inverse extraction and level-set regularity contributes at most $\zeta W(d)$ whenever the structural obstruction has exact conductor d . This is possible because Theorem 2.43 gives $W(d) \geq c_1 W(q)$ for a fixed $c_1 > 0$, whereas $\log K = O(\varepsilon W(q))$.

Apply Theorem 2.43 with this ζ . Alternative (i) there is impossible because bounded powers of a $K^{O(1)}$ -approximate group have size at most $K^{O(1)}|A|$, while $e^{cW(q)}$ dominates this when ε is sufficiently small.

Every surviving structural alternative gives a proper subgroup obstruction at its exact conductor. For proper local sections this is immediate. A genuine common-section fiber product is itself a proper subgroup on its correlation conductor; in the present distinct-prime full-factor family no nontrivial full-factor graph survives, but proper local sections may of course share smaller cyclic sections, and those are treated by the same exact-conductor subgroup bound rather than discarded.

Thus for some proper $H < G(d)$ the many-factor theorem gives a subset of a bounded power of A , of relative size at least

$$e^{-\zeta W(d)} K^{-C},$$

contained in one coset of H , and Proposition 3.1 gives

$$[G(d) : H] \geq e^{c_0 W(d) - C_0}.$$

The level-set regularity in the Volume-II inverse theorem transfers the captured set to the corresponding convolution level with only another $K^{O(1)}$ loss. By the choices above, the resulting coset mass is at least $e^{-3\zeta W(d)}$, while hypothesis (i) bounds it by

$$[G(d) : H]^{-c_{\text{esc}}} \leq e^{-c_{\text{esc}} c_0 W(d) + O(1)}.$$

For large conductor these inequalities are incompatible because $3\zeta < c_{\text{esc}} c_0$; bounded conductors are absorbed into the finite exceptional set. This contradiction proves the one-step flattening estimate. $\square$

REMARK 4.5 (Where no-divisor-loss enters). The proof never performs one flattening step per prime. The many-factor theorem sees all mesoscopic coordinates simultaneously, and escape is normalized by the exact subgroup index. Thus the flattening exponent ε depends on the fixed rank-one constants and on δ , but not on $\omega(q)$.

4. Bounded-stage iteration

THEOREM 4.6 (Bounded-stage squarefree L^2 flattening; BG-X.4.004). *Under the hypotheses of Theorem 4.4, let $v_j = v^{(2^j)}$. There exists an integer $J = J(\delta)$, independent of q and $\omega(q)$, such that either an earlier v_j has already reached the near-uniform range*

$$\|v_j\|_2^2 \leq |G(q)|^{-1+\delta},$$

or

$$\|v_J\|_2^2 \leq |G(q)|^{-1+\delta}.$$

If v is itself a logarithmic-time random-walk measure from Chapter 3, all subgroup non-concentration hypotheses needed during the iteration persist automatically.

PROOF. As long as the measure remains in the intermediate range, Theorem 4.4 gives

$$\|v_{j+1}\|_2 \leq \|v_j\|_2^{1+\varepsilon}.$$

Write

$$\|v_j\|_2 = |G(q)|^{-\alpha_j/2}.$$

Then $\alpha_{j+1} \geq (1 + \varepsilon)\alpha_j$. Starting from $\alpha_0 \geq \delta$, after

$$J \leq \left\lceil \frac{\log((1 - \delta)/\delta)}{\log(1 + \varepsilon)} \right\rceil$$

steps we have $\alpha_J \geq 1 - \delta$. This J depends only on δ and ε . The persistence of subgroup non-concentration is Corollary 3.6. $\square$

EXAMPLE 4.7 (A false iteration to avoid). An argument of the form “flatten at p_1 , then at p_2 , and so on” would use $\omega(q)$ rounds and could produce a spectral constant tending to zero as the number of factors grows. Theorem 4.6 instead uses a fixed number of global convolution doublings. That distinction is the entire reason Chapter 2 was needed.

CHAPTER 5

Conductor-Sensitive Quasirandomness and Squarefree Expansion

The spectral endpoint must also be conductor-sensitive. An irreducible representation may ignore most prime factors, so its multiplicity is controlled by the product of the active primes, not by the full modulus. We combine tensor-product representation theory with the squarefree flattening theorem on the exact conductor and close the uniform spectral gap.

1. Representation dimension at exact conductor

Earlier input (Volumes I, VI, and IX). Volume I proves the tensor-product decomposition of irreducibles of a direct product, the regular-representation multiplicity formula, and trace/multiplicity spectral estimates. Volumes I and VI give a uniform lower bound $\dim \rho_p \gg p$ for every nontrivial irreducible representation of $G_p = \mathrm{SL}_2(\mathbb{F}_p)$. Volume IX carries out the prime-modulus trace endgame. We repeat the calculation here with exact conductor visible.

THEOREM 5.1 (Conductor-sensitive quasirandomness; BG-X.5.001). *There are constants $c_0, C_0 > 0$ such that if $\rho \in \mathrm{Irr}(G(q))$ has exact conductor*

$$d = \prod_{p \in I(\rho)} p,$$

then

$$\dim \rho = \prod_{p|d} \dim \rho_p \geq C_0^{-\omega(d)} \prod_{p|d} p \geq d^{c_0}$$

for all squarefree d supported on sufficiently large allowed primes. The multiplicity of ρ in the regular representation of $G(d)$ equals $\dim \rho$.

PROOF. Proposition 1.4 gives the tensor product. The prime-field quasirandomness theorem gives $\dim \rho_p \geq cp$ for every nontrivial factor. Thus

$$\dim \rho \geq c^{\omega(d)} d.$$

After increasing the minimum allowed prime, $\omega(d) \leq \eta \log d$ with arbitrarily small fixed η , so $c^{\omega(d)} d \geq d^{c_0}$ for some fixed $c_0 > 0$. The regular-representation multiplicity statement is the finite-group Fourier theorem of Volume I. $\square$

2. The exact-conductor trace inequality

Let $T_q f = \mu_q * f$ be convolution by a symmetric probability measure. If ρ has exact conductor d , its Fourier block depends only on $\mu_d = (\pi_d)_* \mu_q$.

PROPOSITION 5.2 (Conductor-sensitive trace inequality; BG-X.5.002). *Let λ be an eigenvalue of the convolution operator on a nontrivial irreducible block of exact conductor d . For every $n \geq 1$,*

$$(\dim \rho) |\lambda|^{2n} \leq |G(d)| \|\mu_d^{(n)}\|_2^2.$$

More generally the sum of $|\lambda_j|^{2n}$ over all copies of that irreducible block is bounded by the same projected return trace.

PROOF. The regular representation of $G(d)$ contains ρ with multiplicity $\dim \rho$. Since T_d is self-adjoint for symmetric μ_d ,

$$\mathrm{Tr}(T_d^{2n}) = |G(d)| \mu_d^{(2n)}(1) = |G(d)| \|\mu_d^{(n)}\|_2^2.$$

All summands in the trace are nonnegative. Keeping only the $\dim \rho$ copies of the eigenvalue under consideration gives the inequality. $\square$

3. The squarefree spectral endpoint

THEOREM 5.3 (Logarithmic-time spectral endgame; BG-X.5.003). *Suppose that for every squarefree divisor d of q and every sufficiently large exact conductor, some time*

$$n_d \leq C \log |G(d)|$$

satisfies

$$\|\mu_d^{(n_d)}\|_2^2 \leq |G(d)|^{-1+\varepsilon_0}$$

with $\varepsilon_0 < c_0/10$, where c_0 is from Theorem 5.1. Then there is $\epsilon > 0$, independent of q , such that every nonconstant eigenvalue of T_q satisfies

$$|\lambda| \leq 1 - \epsilon.$$

Finitely many bounded conductors can be absorbed into the same ϵ .

PROOF. Let d be the exact conductor of the eigenblock. By Proposition 5.2 and Theorem 5.1,

$$|\lambda|^{2n_d} \leq d^{-c_0} |G(d)|^{\varepsilon_0}.$$

Since $|G(d)| \leq d^{3+o(1)}$, the right side is at most $d^{-c_0/2}$ after decreasing ε_0 and increasing the minimum prime. Also $n_d \leq C' \log d$. Therefore

$$|\lambda| \leq \exp\left(-\frac{c_0}{4C'}\right) < 1.$$

The set of excluded/bounded conductors is finite. On each quotient the fixed generating set has no nonconstant invariant vector, so its spectral gap is positive; taking the minimum of finitely many positive constants completes the proof. $\square$

THEOREM 5.4 (Uniform squarefree expansion; BG-X.5.004). *Let $S \subset \mathrm{SL}_2(\mathbb{Z})$ be a fixed finite symmetric set generating a non-elementary subgroup Γ . After excluding finitely many primes, the Cayley graphs*

$$\mathrm{Cay}(\pi_q(\Gamma), \pi_q(S)), \quad q \text{ squarefree},$$

have a spectral gap bounded below independently of q .

PROOF. Volume VII gives strong approximation in the required rank-one form: after excluding finitely many primes, $\pi_p(\Gamma) = G_p$, and hence CRT identifies the squarefree image with $G(q)$.

Take a fixed lazy symmetric power of the generating measure; bounded-word comparison from Volumes I and VII transfers any resulting gap back to the original generators. At a small logarithmic time, Theorem 3.5 gives subgroup non-concentration uniformly on every divisor. Proposition 4.1, Theorem 4.4, and Theorem 4.6 then produce, in a bounded number of convolution doublings, the estimate

$$\|\mu_d^{(n)}\|_2^2 \leq |G(d)|^{-1+\varepsilon_0}$$

for every active conductor d , with total time $O(\log |G(d)|)$. Theorem 5.3 gives a uniform spectral gap. $\square$

THEOREM 5.5 (Rank-one squarefree super-approximation; BG-X.5.005). *Let $\Gamma < \mathrm{SL}_2(\mathbb{Z})$ be finitely generated and non-elementary, and let S be any fixed finite symmetric generating set. There is a finite set of primes S_0 and $\epsilon(S) > 0$ such that, for every squarefree q coprime to S_0 ,*

$$\mathrm{gap}(\mathrm{Cay}(\pi_q(\Gamma), \pi_q(S))) \geq \epsilon(S).$$

Equivalently, the action of Γ on its prime-to- S_0 squarefree congruence completion has property τ along the squarefree family.

PROOF. A non-elementary subgroup of $\mathrm{SL}_2(\mathbb{Z})$ is Zariski dense in SL_2 by Volume VII. Theorem 5.4 applies to one convenient bounded-word symmetric generating set obtained there. Volume I's comparison of Dirichlet forms under bounded word substitution transfers the gap to the originally prescribed S . Enlarging S_0 absorbs the finitely many primes used in strong approximation, subgroup classification, and the large-prime regularization estimates. $\square$

4. Proof architecture and historical sources

The proof just completed is the rank-one squarefree Bourgain–Gamburd mechanism:

CRT/exact conductor $\rightarrow$ many-factor growth $\rightarrow$ approximate Goursat
 $\rightarrow$ index-sensitive escape $\rightarrow L^2$ flattening
 $\rightarrow$ conductor trace $\rightarrow$ spectral gap.

This is the squarefree architecture developed historically by Varjú and, in broader algebraic scope, by Salehi Golsefidy–Varjú (Varjú; Salehi Golsefidy–Varjú). The historical references record the origin of the method, while the present volume supplies the rank-one arguments needed later in the series.

EXAMPLE 5.6 (Why exact conductor is used twice). Suppose $q = p_1 \cdots p_{100}$ but an eigenrepresentation is nontrivial only on $p_7 p_{19}$. The correct multiplicity lower bound is controlled by $p_7 p_{19}$, not by q . Likewise, if a failed-flattening approximate group is correlated only across those two factors, the escape contradiction must be tested on the same conductor. Using the full q in either place creates a false gain. Exact conductor keeps the combinatorial and spectral sides synchronized.

Bibliography

- [1] P. P. Varjú, *Expansion in $SL_d(\mathcal{O}_K/I)$, I square-free*, J. Eur. Math. Soc. 14 (2012), 273–305.
- [2] A. Salehi Golsefidy and P. P. Varjú, *Expansion in perfect groups*, Geom. Funct. Anal. 22 (2012), 1832–1891.

The Bourgain–Gamburd Method
Volume XI: Prime Powers, p -Adic Depth, and Multiscale Growth

15 August 2026

Scope and historical note

This volume develops the prime-power part of the method for sufficiently large residue characteristic. The proof follows the congruence filtration from the residue field into successive Lie-algebra layers, regularizes exact fibers, converts a rich commuting slice into scalar sum-product information, and then uses commutators to absorb a deep principal congruence subgroup. The presentation follows the proved prime-power dependency chain. The small primes excluded by the uniform arguments are recorded explicitly rather than silently absorbed, and the historical references at the end identify the original prime-power sources.

Scope of uniformity. The generating set is fixed throughout. Uniformity is asserted in the prime-power modulus (outside the finite exceptional prime set), not over varying generating sets. This is the scope of the prime-power expansion theorem proved in this volume.

How to read this volume

A prime-power quotient has a visible vertical direction that is absent over a field. The congruence filtration

$$G(p^n) \supset K_1 \supset K_2 \supset \cdots$$

turns that direction into successive copies of the Lie algebra. The central question is therefore not merely whether a set grows, but at which depths it gains new independent digits. Chapters 1–4 regularize those depths and use a scalar sum–product/Fourier argument to manufacture a deep one-parameter segment. Chapter 5 explains how commutators of transverse root directions turn one scalar segment into a substantial principal congruence subgroup. Only then do Chapters 6–7 return to random walks, flattening, and the spectral endgame.

Keep the distinction between three objects visible throughout: the residue modulo p , the exact fiber at a chosen depth, and the ambient set modulo p^n . Many tempting but false shortcuts come from silently replacing one of these by another.

Dependency architecture

The proof order is

filtration/lifting $\rightarrow$ captured-set escape and exact commuting coordinate
 $\rightarrow$ exact-fiber regularization $\rightarrow$ scalar sum-product/Fourier completion
 $\rightarrow$ matrix congruence absorption $\rightarrow L^2$ -flattening $\rightarrow$ exact-conductor spectral gap.

Prime-field expansion is inherited from the earlier Volume IX owner. No result in this volume depends on Volume XII or XIII.

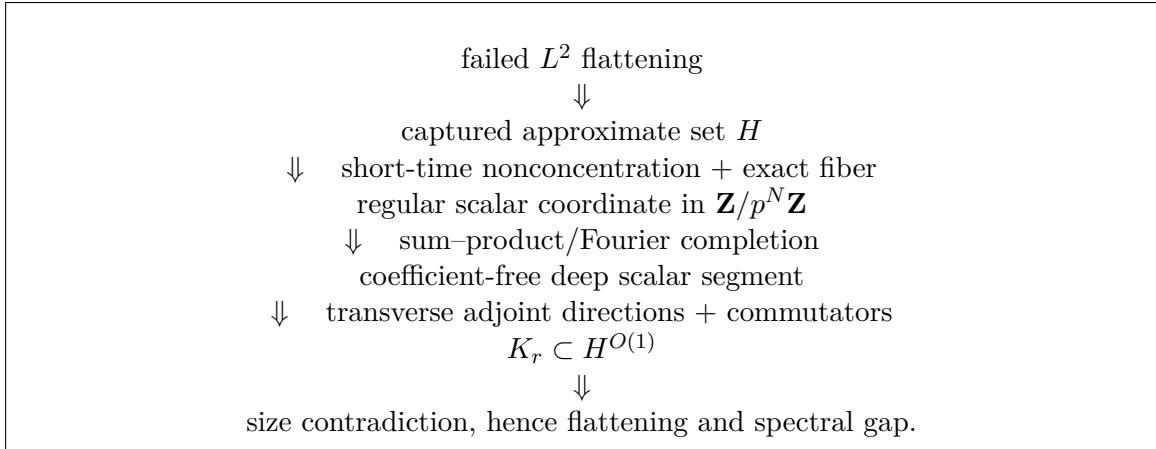


Figure 1: The prime-power contradiction chain. Each arrow changes the type of object being controlled; keeping those interfaces explicit prevents illicit passage from an ambient set to an exact fiber or from average branching to a homogeneous segment.

Numerical depth ledger. If the ambient modulus is p^N and an exact fiber is normalized at depth r , then a homogeneous scalar segment of length p^{N-r} contains exactly p^{N-r} residue classes. A second independent Lie direction of the same depth produces $p^{2(N-r)}$ first-order pairs, and a commutator of two such directions begins at depth $2r$. This elementary accounting is why every later use of commutators records its depth range rather than treating “near the identity” as a single scale.

Series position. This volume uses Volume IX together with the multiscale tools developed earlier. Its main outputs are prime-power growth, congruence absorption, exact-conductor quasirandomness, and prime-power expansion.

Quantifier and asymptotic conventions. Unless a statement explicitly says otherwise, every constant occurring in a quantitative theorem is chosen in the order in which it is quantified in

that theorem. Thus a phrase such as “for all sufficiently small δ ” means: there is a displayed or locally quantified threshold $\delta_0 > 0$, depending only on the parameters listed before that phrase, such that the assertion holds for every $0 < \delta < \delta_0$. The phrases “sufficiently large N ” and “sufficiently large p ” have the analogous meaning.

If $X \ll_\theta Y$ or $X = O_\theta(Y)$, the implicit constant depends only on θ and on the fixed ambient data of the current volume; it is independent of the variables quantified after θ . The words *absolute*, *uniform*, and *bounded* never introduce an unlisted dependence. Whenever an inequality between exponents or scales is essential to the argument, that inequality is written explicitly in the proof; these conventions are not a substitute for such a calculation.

Contents

Scope and historical note	i
How to read this volume	ii
Dependency architecture	iii
1 Congruence Filtration and the Adjoint Graded Lie Algebra	1
1.1 Scope	1
1.2 The principal congruence filtration	1
1.3 Adjoint action on the graded layer	2
1.4 Worked example: first-order multiplication and commutators	3
2 Frattini Lifting and Prime-Power Subgroup Structure	4
2.1 No complement to the first congruence kernel	4
2.2 Frattini-type lifting	5
2.3 Worked example: why a first deep defect cannot occur	6
3 p-Adic Regularization and Exact Commuting Scalar Slices	7
3.1 Purpose and dependency contract	7
3.2 Captured cosets remain captured at shorter logarithmic times	8
3.3 Linear trace coordinates of controlled multiplicity	9
3.4 Controlled adjoint spanning inside the captured set	11
3.5 Exact centralizer coordinates for a rich commuting slice	11
3.6 Congruence calculus	12
3.7 A controlled centralizer coordinate without unit discriminant	12
3.8 A rich commuting slice from slow growth	14
3.9 Maximal-fiber normalization	15
3.10 Low-depth valuation inputs without a hidden coefficient cost	16
3.11 All-scale profile inside the exact commuting fiber	16
3.12 The connector for the later matrix-filling argument	17
3.13 Worked example: maximal-fiber normalization	18
4 Multiscale Sum-Product in $\mathbf{Z}/p^N\mathbf{Z}$	19
4.1 Scope and proof architecture	19
4.2 Entropy preliminaries: the inequality direction	20
4.2.1 A one-digit scalar collision estimate	20
4.3 Regular p -adic trees	21
4.4 Entropy lifting from one digit to R_N	22

4.4.1	The shifted profile produced by a difference	23
4.5	An elementary additive-basis lemma	24
4.6	The branch-profile contradiction	25
4.7	Regular-scale scalar sum-product theorem	27
4.8	Regular-scale inverse alternative	28
4.9	Entropy-profile regularization interface	28
4.10	The regular scalar input and its scope	29
4.11	Exact regular subtree extraction	30
4.11.1	Projection lower bounds survive under small additive growth	30
4.12	Forcing the first two branch levels	31
4.13	A nonregular scalar-growth theorem	31
4.14	Exact-conductor Fourier completion	32
4.15	Dense sets generate a homogeneous deep ideal	34
4.16	Bounded scalar-ring generation at prime-power depth	34
4.17	Eliminating auxiliary residue-field coefficients	35
4.18	Coefficient-free bounded ring expressions	36
4.19	The scalar-coefficient reduction	37
4.20	The active levels provide exact coefficient-free valuations	38
4.21	Coefficient-free thick segment	38
4.22	Undoing exact-fiber normalization after homogeneous scalar generation	41
4.23	Worked example: why average branching does not give a thick block	41
5	Scalar Ideals to Congruence Subgroups	43
5.1	Torus differences realize scalar multiplication without artificial depth loss	43
5.2	Transverse directions and a full traceless layer	46
5.3	The terminal seed-propagation argument	46
5.4	Prime-power approximate-group growth/absorption	49
5.5	Worked example: scalar multiplication from an exact torus	49
6	Prime-Power L^2 Flattening	51
6.1	Scope and dependency contract	51
6.2	Congruence sizes	52
6.3	The inverse set supplied by failed flattening	52
6.4	The prime-power nonconcentration connector	53
6.5	Failed flattening is impossible	54
6.6	Bounded-stage flattening to the conductor threshold	55
6.7	Bounded-conductor inputs	55
6.8	A bounded additive basis from one adjoint orbit	56
6.9	One congruence layer can be lifted in bounded products	57
6.10	Captured inverse sets are already full modulo p	58
6.11	Bounded-conductor failed flattening	59
6.12	Worked example: the size contradiction behind failed flattening	59
7	Exact Conductors and Expansion	61
7.1	Purpose	61
7.2	Exact conductor and an abelian middle-depth kernel	61
7.3	A conductor-sensitive dimension bound	62
7.4	Trace-multiplicity inequality	63

7.5	From flattening to a uniform spectral gap	64
7.6	Uniformity in the large-prime range	65
7.7	Worked example: a middle-depth exact-conductor character	66

Chapter 1

Congruence Filtration and the Adjoint Graded Lie Algebra

Purpose. This chapter fixes the congruence filtration of $G_N = \mathrm{SL}_2(\mathbf{Z}/p^N\mathbf{Z})$ and identifies its first-order layers with $\mathfrak{sl}_2(\mathbf{F}_p)$. Every later depth calculation is charged to these exact multiplication and commutator formulas.

1.1 Scope

Let $p \geq 5$ and

$$R_N = \mathbf{Z}/p^N\mathbf{Z}, \quad G_N = \mathrm{SL}_2(R_N), \quad K_r = \ker(G_N \rightarrow G_r) \quad (1 \leq r \leq N).$$

This chapter proves the filtration and lifting facts needed later. No sum-product, flattening, or representation-theoretic input is used.

1.2 The principal congruence filtration

Proposition 1.1 (Structure and size of the filtration). *The groups*

$$G_N \supseteq K_1 \supseteq K_2 \supseteq \cdots \supseteq K_N = \{1\}$$

are normal. For $1 \leq r < N$ the map

$$\phi_r : K_r \longrightarrow \mathfrak{sl}_2(\mathbf{F}_p), \quad 1 + p^r X \longmapsto X \bmod p$$

induces an isomorphism

$$K_r/K_{r+1} \simeq (\mathfrak{sl}_2(\mathbf{F}_p), +).$$

Consequently

$$|K_r| = p^{3(N-r)}, \quad |G_N| = p^{3(N-1)} |\mathrm{SL}_2(\mathbf{F}_p)|.$$

Proof. Normality is immediate from the kernel definition. Every element of K_r has the form $1 + p^r X$ modulo p^{r+1} . The determinant identity

$$\det(1 + p^r X) \equiv 1 + p^r \mathrm{tr} X \pmod{p^{r+1}}$$

shows that determinant one is equivalent to $\mathrm{tr} X \equiv 0 \pmod{p}$. Conversely every traceless residue X_0 lifts to an element of K_r : start with $1 + p^r X_0$ and change one diagonal entry by a multiple of

p^{r+1} to correct the determinant; equivalently use surjectivity of $G_{r+1} \rightarrow G_r$, proved by the same elementary determinant correction. Multiplication gives

$$(1 + p^r X)(1 + p^r Y) \equiv 1 + p^r(X + Y) \pmod{p^{r+1}},$$

so the induced quotient map is additive and has kernel K_{r+1} . Since $\dim_{\mathbf{F}_p} \mathfrak{sl}_2 = 3$, each successive kernel has size p^3 , giving the cardinality formulas. $\square$

Lemma 1.2 (First-order multiplication at arbitrary precision). *If $a \geq 1$, $L \leq a$, and*

$$u \equiv 1 + p^a X, \quad v \equiv 1 + p^a Y \pmod{p^{a+L}},$$

then

$$uv \equiv 1 + p^a(X + Y), \quad u^{-1} \equiv 1 - p^a X \pmod{p^{a+L}}.$$

Proof. Every discarded term has valuation at least $2a \geq a + L$. $\square$

Lemma 1.3 (First-order commutator and the Lie bracket). *If $r, s \geq 1$ and $r + s < N$, then*

$$[1 + p^r X, 1 + p^s Y] \equiv 1 + p^{r+s}[X, Y] \pmod{p^{r+s+1}}.$$

Thus under Proposition 1.1, the commutator pairing

$$K_r/K_{r+1} \times K_s/K_{s+1} \longrightarrow K_{r+s}/K_{r+s+1}$$

is the ordinary Lie bracket on $\mathfrak{sl}_2(\mathbf{F}_p)$.

Proof. Expand

$$(1 + p^r X)(1 + p^s Y)(1 - p^r X)(1 - p^s Y)$$

and retain terms of valuation $r + s$. Terms involving two copies of $p^r X$ or two copies of $p^s Y$ have strictly larger valuation after the cancellations forced by the inverse expansions. The surviving term is $p^{r+s}(XY - YX)$. $\square$

1.3 Adjoint action on the graded layer

Lemma 1.4 (Compatibility with reduction). *For $g \in G_N$ and $x = 1 + p^r X \in K_r$,*

$$gxg^{-1} \equiv 1 + p^r \text{Ad}(\bar{g})(\bar{X}) \pmod{p^{r+1}},$$

where bars denote reduction modulo p . Hence the G_N -action on K_r/K_{r+1} factors through the adjoint representation of $G_1 = \text{SL}_2(\mathbf{F}_p)$ on $\mathfrak{sl}_2(\mathbf{F}_p)$.

Proof. Conjugation is exactly $1 + p^r(gXg^{-1})$; reduction modulo p^{r+1} sees only g, X modulo p . $\square$

Proposition 1.5 (Irreducibility of the adjoint layer). *For $p > 3$, the adjoint representation of $\text{SL}_2(\mathbf{F}_p)$ on $\mathfrak{sl}_2(\mathbf{F}_p)$ is irreducible. Equivalently, the only G_N -invariant subspaces of every graded quotient K_r/K_{r+1} are 0 and the full quotient.*

Proof. Let $W \neq 0$ be invariant under $\text{SL}_2(\mathbf{F}_p)$. For $E = E_{12}$ and $t \in \mathbf{F}_p$,

$$\text{Ad}(1 + tE)X = X + t[E, X] + \frac{t^2}{2}[E, [E, X]],$$

because $E^2 = 0$ and $p > 2$. Since the left side lies in W for every t , polynomial interpolation in t (degree at most two, and $p > 3$) shows that $[E, X] \in W$ for every $X \in W$. The same argument with $F = E_{21}$ gives $[F, W] \subseteq W$. Since E, F generate $\mathfrak{sl}_2$ as a Lie algebra, W is an ideal of $\mathfrak{sl}_2(\mathbf{F}_p)$. This Lie algebra is simple for $p > 2$, so $W = \mathfrak{sl}_2(\mathbf{F}_p)$. $\square$

1.4 Worked example: first-order multiplication and commutators

Let $p = 5$, $N = 3$, and write

$$E = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad F = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad H = [E, F] = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Take $u = 1 + 5E$ and $v = 1 + 5F$ modulo 125. Modulo 25 the quadratic term disappears, so

$$uv \equiv 1 + 5(E + F) \pmod{25}.$$

At the next depth the commutator records the Lie bracket:

$$[u, v] \equiv 1 + 25[E, F] = 1 + 25H \pmod{125}.$$

Thus addition in the graded layer is literally first-order group multiplication, while the next nontrivial term of a group commutator is the Lie bracket. This is the concrete calculation behind the repeated passage between congruence subgroups and $\mathfrak{sl}_2(\mathbf{F}_p)$ later in the volume.

Chapter 2

Frattini Lifting and Prime-Power Subgroup Structure

The decisive subgroup fact is that, for $p \geq 5$, surjectivity modulo p cannot first fail at a deeper congruence level. This turns subgroup escape at prime-power level into the already-owned prime-field subgroup-escape problem.

2.1 No complement to the first congruence kernel

The next elementary lemma is the key to the Frattini-type statement.

Lemma 2.1 (A nontrivial unipotent has no order- p lift modulo p^2). *Let $\bar{u} \in \mathrm{SL}_2(\mathbf{F}_p)$ be a nontrivial unipotent, so $\bar{u}$ has order p . If $u \in \mathrm{SL}_2(\mathbf{Z}/p^2\mathbf{Z})$ reduces to $\bar{u}$, then*

$$u^p \neq 1.$$

Hence the reduction map

$$\mathrm{SL}_2(\mathbf{Z}/p^2\mathbf{Z}) \rightarrow \mathrm{SL}_2(\mathbf{F}_p)$$

has no group-theoretic section.

Proof. Write $u = 1 + X$ with $X \bmod p$ a nonzero nilpotent matrix of square zero. Thus

$$X^2 \in pM_2(\mathbf{Z}/p^2\mathbf{Z}).$$

Since powers of the single matrix X commute, the ordinary binomial theorem gives

$$u^p = (1 + X)^p = 1 + pX + \sum_{j=2}^p \binom{p}{j} X^j.$$

For $2 \leq j \leq p-1$, $p \mid \binom{p}{j}$ and X^j is divisible by p because X^2 is; hence those terms vanish modulo p^2 . Also X^p is divisible by p^2 for $p \geq 5$. Therefore

$$u^p \equiv 1 + pX \pmod{p^2}.$$

Because $X \bmod p \neq 0$, the term pX is nonzero modulo p^2 . A section would send the order- p element $\bar{u}$ to an element whose order divides p , contradiction. $\square$

2.2 Frattini-type lifting

Theorem 2.2 (Surjective modulo p implies the full prime-power group). *Let $p \geq 5$ and $H \leq G_N$. If the reduction of H modulo p is all of G_1 , then*

$$\boxed{H = G_N.}$$

Proof. For $1 \leq r < N$, let

$$W_r = \phi_r(H \cap K_r) \subseteq \mathfrak{sl}_2(\mathbf{F}_p).$$

Because H surjects onto G_1 , Lemma 1.4 implies that W_r is G_1 -invariant. Proposition 1.5 gives

$$W_r = 0 \quad \text{or} \quad W_r = \mathfrak{sl}_2(\mathbf{F}_p).$$

We first show $W_1 \neq 0$. If $W_1 = 0$, then $H \cap K_1 \subseteq K_2$. Reducing H modulo p^2 would give a subgroup of G_2 mapping isomorphically onto G_1 , hence a section of $G_2 \rightarrow G_1$, contradicting Lemma 2.1. Thus W_1 is full.

Now suppose W_r is full and $r + 1 < N$. For every $X \in \mathfrak{sl}_2(\mathbf{F}_p)$ choose

$$h_X \in H \cap K_r, \quad h_X \equiv 1 + p^r X \pmod{p^{r+1}}.$$

The binomial expansion, using $r \geq 1$, gives

$$h_X^p \equiv 1 + p^{r+1} X \pmod{p^{r+2}}.$$

Hence W_{r+1} contains every X and is full. By induction all W_r are full.

Finally descend the filtration. Since $K_N = 1 \subseteq H$, if $K_{r+1} \subseteq H$ and W_r is full, then every coset of K_{r+1} in K_r has a representative in $H \cap K_r$, so $K_r \subseteq H$. Thus $K_1 \subseteq H$. Together with surjectivity $H \rightarrow G_1$, this gives $H = G_N$. $\square$

Corollary 2.3 (Proper subgroups are detected modulo p). *Every proper subgroup $H < G_N$ has proper image in G_1 . In particular a proper subgroup cannot have its first defect only at a depth $r \geq 2$.*

Proof. This is the contrapositive of Theorem 2.2. $\square$

Corollary 2.4 (Reduction of subgroup escape to the prime field). *Let μ be any probability measure on G_N . If $H < G_N$ is proper and $g \in G_N$, then*

$$\mu(gH) \leq (\pi_{1*}\mu)(\bar{g}\bar{H}),$$

where $\bar{H} < G_1$ is proper. Therefore any uniform random-walk escape estimate from cosets of proper subgroups of G_1 immediately gives the same estimate for cosets of proper subgroups at every prime-power level.

Proof. The coset gH is contained in the inverse image of the coset $\bar{g}\bar{H}$ under reduction modulo p . Corollary 2.3 makes $\bar{H}$ proper. $\square$

2.3 Worked example: why a first deep defect cannot occur

Consider $G_3 = \mathrm{SL}_2(\mathbf{Z}/5^3\mathbf{Z})$ and suppose a subgroup $H < G_3$ maps onto G_1 . If $H \cap K_1$ were trivial modulo K_2 , then H would provide a complement to the first congruence kernel in G_2 , contradicting the nonsplitting lemma. Hence the image of $H \cap K_1$ in $K_1/K_2 \cong \mathfrak{sl}_2(\mathbf{F}_5)$ is nonzero. It is G_1 -invariant because H surjects modulo 5; irreducibility therefore forces it to be the whole graded layer. Consequently $HK_2 = G_3$. Repeating at the next layer gives $H = G_3$.

The example displays the exact mechanism of the Frattini theorem: one does not classify all subgroups of G_N ; nonsplitting gives one nonzero graded direction, and adjoint irreducibility forces all three directions.

Chapter 3

p -Adic Regularization and Exact Commuting Scalar Slices

The role of this chapter is to manufacture the *right scalar object* from a captured slowly-growing matrix set. Two issues are kept separate: first the exact algebraic commuting coordinate, then the all-scale p -adic profile inside that same fiber. No ambient-to-fiber entropy inheritance is used.

3.1 Purpose and dependency contract

Put

$$R_N = \mathbf{Z}/p^N \mathbf{Z}, \quad G_N = \mathrm{SL}_2(R_N), \quad Q = p^N,$$

with p outside the finite arithmetic exceptional set.

Definition 3.1 (p -adic valuation, primitive matrices, and exact congruence depth). For $x \in R_N$, define

$$v_p(x) := \max\{0 \leq r \leq N : x \in p^r R_N\}, \quad v_p(0) := N.$$

For a matrix $X = (x_{ij}) \in M_2(R_N)$ put

$$v_p(X) := \min_{i,j} v_p(x_{ij}).$$

A matrix X is *primitive* if $v_p(X) = 0$; a primitive traceless matrix is therefore an element of $\mathfrak{sl}_2(R_N)$ whose reduction modulo p is nonzero. For $g \in G_N$ define its *exact congruence depth*

$$d_N(g) := v_p(g - I) \in \{0, 1, \dots, N\}.$$

Thus $d_N(g) = t < N$ if and only if

$$g = I + p^t X \pmod{p^N}$$

for a primitive matrix X ; and $d_N(I) = N$. These meanings of “primitive” and “exact depth” are used throughout Volumes XI–XII whenever the modulus is clear from context.

The mechanism needed here is not

“the walk has Q^{-c} mass on no inverse image of a proper mod- p subgroup.”

That statement is false for a full inverse image of a fixed positive-density mod- p subgroup. What the Bourgain–Gamburd prime-power argument actually needs is different: a captured approximate

set has mass $Q^{-O(\varepsilon)}$, while fixed-complexity *congruence polynomial* obstructions have mass p^{-cr} at depth r . Taking $r = C\varepsilon N$ therefore lets one choose, *inside bounded powers of the captured set*, elements with controlled discriminant depth, trace-separated elements, and adjoint conjugators whose only defect is an $O(\varepsilon N)$ loss in p -adic depth. We do not require the discriminant to be a unit modulo p ; that stronger condition would exclude the important case in which the captured approximate set lies deep in a congruence neighborhood.

We use two earlier arithmetic inputs proved in this series. They are stated explicitly because they are the only non-combinatorial inputs of this module.

- (E1) **Short-word atom/injectivity.** There are constants $c_{\text{inj}}, c_{\text{at}} > 0$ such that, for $1 \leq r \leq N$ and

$$j \leq c_{\text{inj}} r \log p,$$

the reduction modulo p^r of the j -step walk has point masses at most $e^{-c_{\text{at}} j}$. Moreover every endpoint in the support of the j -step walk has probability at least $|S|^{-j}$, simply because at least one length- j path reaches it. This is the earlier free-word/girth owner after passing, if necessary, to the fixed bounded-word free model used in the arithmetic arc.

- (E2) **Congruence polynomial escape.** For every fixed-degree nonzero polynomial condition F occurring below there are $c_F, C_F > 0$ such that, whenever $r \leq N$, $j \geq C_F r \log p$, and the integral coefficients/lifts have height at most $e^{c_F j}$,

$$\nu^{*j} \{g : F(g) \equiv 0 \pmod{p^r}\} \leq p^{-c_F r}.$$

The allowed family includes: discriminants $(\text{tr } g)^2 - 4$; resultants needed after multiplication by one of finitely many controlled matrices; linear trace pairings $\text{tr}(gA)$; and adjoint pairings $\text{tr}(gAg^{-1}B)$ for primitive traceless A, B . This is the polynomial/random-matrix escape package proved earlier in the arithmetic arc.

The present module proves every deduction from (E1)–(E2). No prime-power product theorem, flattening theorem, or representation-theoretic estimate is imported.

3.2 Captured cosets remain captured at shorter logarithmic times

Lemma 3.2 (Shortening a captured coset). *Let $H \subseteq G_N$, let ν be symmetric, and suppose*

$$\nu^{*L}(xH) \geq \eta.$$

For every $0 < j < L$ there exists $x_j \in G_N$ with

$$\nu^{*j}(x_j H) \geq \eta.$$

Consequently

$$\nu^{*2j}(H^{-1}H) \geq \eta^2.$$

Proof. Write $\nu^{*L} = \nu^{*(L-j)} * \nu^{*j}$. The quantity $\nu^{*L}(xH)$ is an average, with respect to $\nu^{*(L-j)}$, of translates of the function $y \mapsto \nu^{*j}(yH)$. One translate is therefore at least the average η . If X, Y are independent with law ν^{*j} and both lie in the same coset $x_j H$, then $X^{-1}Y \in H^{-1}H$. Symmetry identifies the law of $X^{-1}Y$ with ν^{*2j} . $\square$

Lemma 3.3 (Projection lower bound from captured mass). *Assume $\nu^{*2j}(H^2) \geq \eta$ and the point masses of $\pi_{p^r*}\nu^{*2j}$ are at most α . Then*

$$|\pi_{p^r}(H^2)| \geq \eta/\alpha.$$

In particular, if $j \asymp r \log p$ lies in the short-word window, $r \geq C\varepsilon N$, and $\eta \geq Q^{-C_0\varepsilon}$, then, after choosing C large compared with C_0 and then ε small, one gets

$$|\pi_{p^r}(H^2)| \geq p^{c_0 r}$$

with fixed $c_0 > 0$.

Proof. The total mass η is distributed among at most $|\pi_{p^r}(H^2)|$ atoms, each of mass at most α . The quantitative conclusion follows from (E1): $\alpha \leq e^{-c_{\text{at}}j} = p^{-c_1 r}$, whereas $\eta \geq p^{-C_0\varepsilon N}$. $\square$

3.3 Linear trace coordinates of controlled multiplicity

Let $\langle X, Y \rangle = \text{tr}(XY)$ on $M_2(R_r)$.

Lemma 3.4 (A controlled trace basis inside H^2). *Suppose*

$$\nu^{*2j}(H^2) \geq Q^{-C_0\varepsilon}$$

and $j \asymp r \log p$ with $r = C_1\varepsilon N$ for C_1 sufficiently large. Then there are

$$g_1, g_2, g_3, g_4 \in H^2$$

and an integer $d_0 \leq C_2\varepsilon N$ such that, after vectorizing 2×2 matrices,

$$v_p \det(g_1, g_2, g_3, g_4) \leq d_0.$$

Consequently the trace map

$$\Phi_r : M_2(R_r) \rightarrow R_r^4, \quad X \mapsto (\text{tr}(Xg_i))_{i=1}^4,$$

has fibers of size at most p^{4d_0} .

Proof. Choose the g_i successively. If $g_1, \dots, g_k$ have been chosen with $k < 4$, choose a primitive matrix B in the trace-orthogonal complement of their span modulo p^r . Membership of g in that span modulo p^r forces $\text{tr}(gB) \equiv 0 \pmod{p^r}$. By (E2), for $r = C_1\varepsilon N$ with C_1 large enough, the walk mass of this congruence hyperplane is less than $\nu^{*2j}(H^2)$. Hence an element of H^2 lies outside it. Repeating gives a basis modulo p^r , equivalently a determinant of valuation $< r$; put $d_0 = r$ (or its exact smaller valuation).

For the fiber statement, let $D = X - X'$ lie in the kernel of Φ_r . If G is the 4×4 matrix with columns g_i , the adjugate identity gives $(\det G)D = 0$ in vectorized coordinates. Thus every coordinate of D is divisible by p^{r-d_0} , and the kernel has at most p^{4d_0} elements. $\square$

Proposition 3.5 (Many controlled-nonscalar elements with separated traces). *Fix $0 < \delta < 1$ and put $n_1 = \lfloor \delta N \rfloor$. Let $r_{\text{reg}} = \lceil C_4\varepsilon N \rceil$, where C_4 is chosen for the finitely many polynomial conditions used below. Assume that there is a constant $\kappa > 0$, independent of p, N , such that*

$$2\kappa \leq c_{\text{inj}}, \quad 2\kappa\delta N \geq \max_F C_F r_{\text{reg}}, \quad n_1 \geq 10r_{\text{reg}},$$

where the maximum is over those finitely many conditions in (E2). Put $j_1 = \lfloor \kappa n_1 \log p \rfloor$ and assume the shortened captured mass satisfies

$$\nu^{*2j_1}(H^2) \geq Q^{-C_0\varepsilon}.$$

Then there exist a bounded exponent C_H , a number

$$\beta \geq p^{c_1 n_1 - C_3 \varepsilon N},$$

and elements

$$h_1, \dots, h_\beta \in H^{C_H}$$

whose traces are pairwise distinct modulo $p^{n_1 - C_3 \varepsilon N}$ and whose discriminants satisfy

$$v_p((\operatorname{tr} h_j)^2 - 4) < d_{\text{reg}} := C_4 \varepsilon N.$$

In particular, writing

$$e(h_j) := \min v_p \left(h_j - \frac{\operatorname{tr} h_j}{2} 1 \right),$$

one has $e(h_j) < d_{\text{reg}}/2$.

Proof. Choose the trace basis first at an auxiliary depth $r_0 = C_1 \varepsilon N$ by Lemma 3.4; its elements have word length $O(r_0 \log p)$ and hence height $e^{O(r_0 \log p)}$. The hypotheses on κ, δ make the two time requirements simultaneously visible: $2j_1 \leq c_{\text{inj}} n_1 \log p$ places the $2j_1$ -step walk in the atom/injectivity window (E1), while $2j_1 \geq C_F r_{\text{reg}} \log p$ for every polynomial F used below places it in the escape window (E2). Let

$$B_{\text{good}} = H^2 \cap \bigcap_{i=1}^4 \{b : (\operatorname{tr}(bg_i))^2 - 4 \not\equiv 0 \pmod{p^{r_{\text{reg}}}}\}, \quad r_{\text{reg}} = C_4 \varepsilon N.$$

For each i , the excluded condition is one of the fixed resultant/discriminant conditions allowed in (E2); the height of g_i is harmless because $n_1 \gg r_0$. Taking C_4 large and then ε small makes the total excluded walk mass less than half of $\nu^{*2j_1}(H^2)$. Thus B_{good} still carries mass at least $Q^{-O(\varepsilon)}$.

By the point-mass estimate in (E1), its reduction modulo p^{n_1} therefore has at least

$$p^{c_0 n_1 - O(\varepsilon N)}$$

distinct elements. Apply the trace-vector map of Lemma 3.4; its fibers have size at most $p^{O(\varepsilon N)}$, so the trace-vector image still has size $p^{c_0 n_1 - O(\varepsilon N)}$. At least one coordinate i assumes

$$p^{c_0 n_1 / 4 - O(\varepsilon N)}$$

distinct values. For that i , the elements $h = bg_i$ have discriminant nonzero modulo $p^{r_{\text{reg}}}$ by construction, lie in a bounded power of H , and have pairwise distinct ordinary traces modulo $p^{n_1 - O(\varepsilon N)}$. Hence $v_p((\operatorname{tr} h)^2 - 4) < r_{\text{reg}}$, giving the stated d_{reg} . Finally, with $h_0 = h - (\operatorname{tr} h)1/2$, the identity $(\operatorname{tr} h)^2 - 4 = -4 \det(h_0)$ shows that if every entry of h_0 is divisible by p^e , then the discriminant is divisible by p^{2e} . Thus $e(h) < d_{\text{reg}}/2$. $\square$

3.4 Controlled adjoint spanning inside the captured set

Proposition 3.6 (Adjoint spanning with only $O(\varepsilon N)$ depth loss). *Let $A \in \mathfrak{sl}_2(R_N)$ be primitive. Under the captured-mass hypotheses there exist*

$$c_1, c_2, c_3 \in H^2$$

and $d_{\text{sp}} \leq C\varepsilon N$ such that

$$v_p \det(\text{Ad}(c_1)A, \text{Ad}(c_2)A, \text{Ad}(c_3)A) \leq d_{\text{sp}}.$$

Equivalently, for every $L > d_{\text{sp}}$,

$$p^{d_{\text{sp}}} \mathfrak{sl}_2(R_L) \subseteq R_L \text{Ad}(c_1)A + R_L \text{Ad}(c_2)A + R_L \text{Ad}(c_3)A.$$

Proof. Choose the conjugates successively. If the first $k < 3$ conjugates span a proper submodule modulo p^r , choose a primitive traceless B in its trace-orthogonal complement. Every c whose conjugate $\text{Ad}(c)A$ lies in that span satisfies

$$\text{tr}(\text{Ad}(c)A B) \equiv 0 \pmod{p^r}.$$

The adjoint-pairing escape estimate (E2), with $r = C\varepsilon N$, makes the walk mass of this obstruction smaller than $\nu^{*2j}(H^2)$. Hence some $c \in H^2$ escapes it. After three choices the determinant is nonzero modulo p^r , so its valuation is $< r$.

For the module inclusion, write the three vectors as the columns of a 3×3 matrix T . The adjugate identity $\text{adj}(T)T = (\det T)I$ shows that $p^{d_{\text{sp}}} R_L^3$ lies in the image of T . $\square$

Corollary 3.7 (Controlled escape from any proper linear kernel). *Let $T : \mathfrak{sl}_2(R_N) \rightarrow \mathfrak{sl}_2(R_N)$ be an R_N -linear map for which some primitive X_0 satisfies*

$$T(X_0) = p^{d_T} Y_0, \quad Y_0 \text{ primitive.}$$

Let $A \in \mathfrak{sl}_2(R_N)$ be primitive. Under the captured-mass hypotheses there is $c \in H^2$ such that

$$v_p(T(\text{Ad}(c)A)) \leq d_T + d_{\text{sp}},$$

where $d_{\text{sp}} = O(\varepsilon N)$ is the loss in Proposition 3.6.

Proof. By Proposition 3.6,

$$p^{d_{\text{sp}}} X_0 = \sum_{i=1}^3 a_i \text{Ad}(c_i)A$$

for suitable $a_i \in R_N$ and $c_i \in H^2$. Applying T gives

$$p^{d_{\text{sp}}+d_T} Y_0 = \sum_i a_i T(\text{Ad}(c_i)A).$$

If every summand on the right were divisible by $p^{d_{\text{sp}}+d_T+1}$, so would the primitive left-hand side after removing its displayed exact power, a contradiction. Thus one c_i has the required bound. $\square$

3.5 Exact centralizer coordinates for a rich commuting slice

The next lemmas are the algebraic half of the commuting-slice construction. They are placed here, before scalar regularization, to make the proof order acyclic. The later matrix-realization part of the same argument appears in Chapter 5.

3.6 Congruence calculus

Lemma 3.8 (First-order multiplication). *If $a \geq 1$, $L \leq a$, and*

$$u \equiv 1 + p^a X, \quad v \equiv 1 + p^a Y \pmod{p^{a+L}},$$

then

$$uv \equiv 1 + p^a(X + Y), \quad u^{-1} \equiv 1 - p^a X \pmod{p^{a+L}}.$$

Proof. The discarded quadratic terms are divisible by p^{2a} , and $2a \geq a + L$. □

Lemma 3.9 (Determinant-one first coordinate). *If $g = 1 + p^a X \in G_N$, then*

$$\mathrm{tr} X \equiv -p^a \det X \pmod{p^{N-a}}.$$

Hence X is traceless modulo p^L whenever $L \leq a$ and $a + L \leq N$.

Proof. Use $\det(1 + p^a X) = 1 + p^a \mathrm{tr} X + p^{2a} \det X$. □

3.7 A controlled centralizer coordinate without unit discriminant

Let $h \in G_N$, put

$$T = \mathrm{tr} h, \quad h_0 = h - \frac{T}{2} 1, \quad \Delta = T^2 - 4 = -4 \det(h_0),$$

and assume Δ is nonzero modulo a controlled power of p . Define the *scalar depth*

$$e = e(h) := \min v_p(h_0),$$

where the minimum is over the four entries, and put

$$z = p^{-e} h_0, \quad N' = N - e.$$

Then z is primitive and traceless. If $v_p(\Delta) < d_{\mathrm{reg}}$, then $2e \leq v_p(\Delta) < d_{\mathrm{reg}}$.

Lemma 3.10 (Truncated centralizer algebra). *If $s \in G_N$ commutes with h , then modulo $p^{N'}$ one has a unique expression*

$$s \equiv a1 + bz \pmod{p^{N'}}.$$

Equivalently, the reduction of the exact centralizer $C_{G_N}(h)$ modulo $p^{N'}$ is contained in the determinant-one part of $R_{N'}1 + R_{N'}z$.

Proof. Since $h = T1/2 + p^e z$, exact commutation gives

$$p^e [s, z] = [s, h] = 0 \pmod{p^N},$$

so the reduction of s modulo $p^{N'}$ centralizes z . The reduction of z modulo p is a nonzero traceless matrix and hence is not scalar for odd p . Therefore the commutator map $X \mapsto [X, z]$ has rank two modulo p and its kernel is spanned by $1, z$. A unit 2×2 minor persists over $R_{N'}$, giving

$$C_{\mathrm{Mat}_2(R_{N'})}(z) = R_{N'}1 + R_{N'}z.$$

Uniqueness follows from the linear independence of $1, z$ modulo p . □

Lemma 3.11 (Invariant conjugation-difference identity). *For every $g \in \mathrm{SL}_2(R_N)$ and $\xi \in \mathrm{Mat}_2(R_N)$,*

$$\mathrm{Ad}(g)\xi - \mathrm{Ad}(g^{-1})\xi = (\mathrm{tr} g)[g, \xi].$$

If g commutes with h and $g \equiv a1 + bz \pmod{p^{N'}}$, then

$$\mathrm{Ad}(g)\xi - \mathrm{Ad}(g^{-1})\xi \equiv x(g)[z, \xi] \pmod{p^{N'}}, \quad x(g) := (\mathrm{tr} g)b \in R_{N'}.$$

Proof. Cayley–Hamilton gives $g^{-1} = (\mathrm{tr} g)1 - g$, hence

$$g\xi g^{-1} - g^{-1}\xi g = (\mathrm{tr} g)(g\xi - \xi g).$$

The second assertion follows from Lemma 3.10. □

Lemma 3.12 (The scalar coordinate is analytic to the available precision). *Let $\overline{T}_h^1$ be the reduction modulo $p^{N'}$ of exact centralizer elements $g \in C_{G_N}(h)$ satisfying $g \equiv 1 \pmod{p}$. Writing $g \equiv a1 + bz \pmod{p^{N'}}$, the coordinate*

$$x(g) = (\mathrm{tr} g)b = 2ab$$

is compatible with every congruence projection and, on the full determinant-one branch of $R_{N'}1 + R_{N'}z$ through the identity, is a bijection

$$pR_{N'} \longrightarrow pR_{N'}.$$

Consequently, after choosing one representative for every reduction modulo $p^{N'}$, a commuting slice transfers projection sizes and fiber bounds to the normalized scalar coordinate $x(g)/p \in R_{N'-1}$, with at most a factor p^{3e} loss from discarded lifts.

Proof. For $g = a1 + bz$ one has $\mathrm{tr} z = 0$ and

$$\det(a1 + bz) = a^2 + b^2 \det z.$$

On the residue class $(a, b) \equiv (1, 0) \pmod{p}$, the derivative with respect to a is $2a$, a unit. Thus every $b \in pR_{N'}$ determines a unique $a \equiv 1 \pmod{p}$ with determinant one. The resulting map $b \mapsto x(b) = 2a(b)b$ has derivative congruent to 2 modulo p , hence is a reduction-compatible bijection of $pR_{N'}$ by successive Hensel lifting. Finally, two exact centralizer elements with the same reduction modulo $p^{N'}$ differ by an element of $K_{N'}$, whose cardinality is p^{3e} ; choosing one representative per reduction therefore costs at most that factor. □

Lemma 3.13 (Fixed-degree operator loss is controlled by the discriminant). *Put*

$$\delta_z = -4 \det z = p^{-2e} \Delta, \quad d_z = v_p(\delta_z) < d_{\mathrm{reg}}.$$

On $\mathfrak{sl}_2(R_N)$ one has the exact identity

$$\mathrm{ad}(z)^3 = \delta_z \mathrm{ad}(z).$$

Hence, for every fixed $D \geq 1$, the R_N -linear operator $\mathrm{ad}(z)^D$ has a nonzero Smith coefficient of valuation at most

$$d_D \leq \left\lfloor \frac{D-1}{2} \right\rfloor d_z.$$

In particular $d_D = O_D(d_{\mathrm{reg}})$.

Proof. After adjoining a root of the characteristic polynomial, z has eigenvalues $u, -u$ with $4u^2 = \delta_z$, so $\text{ad}(z)$ has eigenvalues $0, 2u, -2u$. The displayed cubic identity follows and, being a polynomial identity with integral coefficients, descends to R_N . Since $z \bmod p$ is a nonzero traceless matrix, both $\text{ad}(z)$ and $\text{ad}(z)^2$ are nonzero modulo p . Here is the elementary calculation hidden by that assertion. Over an algebraic closure of the residue field there are two cases. If z is semisimple, choose root vectors E, F so that $z = \text{diag}(u, -u)$ with $u \neq 0$. Then

$$[z, E] = 2uE, \quad [z, [z, E]] = 4u^2E \neq 0.$$

If z is nilpotent, after conjugation and rescaling write $z = cE$ with $c \neq 0$ and take $H = [E, F]$. Then

$$[z, F] = cH, \quad [z, [z, F]] = -2c^2E \neq 0,$$

because the primes in this volume are odd (indeed the small exceptional primes have already been removed). Nonvanishing after scalar extension implies nonvanishing over the residue field itself. Reducing powers with the cubic identity gives

$$\text{ad}(z)^{2k+1} = \delta_z^k \text{ad}(z), \quad \text{ad}(z)^{2k+2} = \delta_z^k \text{ad}(z)^2,$$

which proves the valuation bound. $\square$

3.8 A rich commuting slice from slow growth

Proposition 3.14 (Conjugation fibers produce an exact commuting slice). *Let $H \subseteq G_N$ be symmetric, $1 \in H$, and suppose*

$$|H^{10}| \leq M|H|.$$

Assume H^8 contains β elements h_α whose traces are pairwise distinct modulo some p^{n_1} and whose discriminants have valuation $< d_{\text{reg}}$. Then for some such $h \in H^8$ there is a set $S \subseteq H^2$ such that every $s \in S$ centralizes h in the full group G_N and

$$|S| \geq \beta/M.$$

After choosing one residue class modulo p , translating inside the centralizer, and then choosing one representative for every reduction modulo $p^{N-e(h)}$, one obtains a set $S_0 \subseteq H^4$ congruent to 1 modulo p whose reductions are distinct modulo $p^{N-e(h)}$. The total cardinality loss is at most $p^{O(e(h)+1)}$.

Proof. The full conjugacy sets $\{gh_\alpha g^{-1} : g \in H\}$ are disjoint: conjugate elements have the same trace in R_N , hence the same trace modulo p^{n_1} . They all lie in H^{10} , so one has size at most $M|H|/\beta$. A fiber of $g \mapsto ghg^{-1}$ therefore has size at least β/M . Taking differences of two elements of that fiber gives the required exact centralizer set in H^2 . Pigeonhole the reduction of this set modulo p , fix one element s_0 in the largest fiber, and replace S by Ss_0^{-1} . Since the centralizer is a group, the translated set still centralizes h and is now congruent to 1 modulo p . Finally choose one representative in each fiber of reduction modulo $p^{N-e(h)}$; each such fiber has size at most $|K_{N-e(h)}| = p^{3e(h)}$. $\square$

Proposition 3.15 (Matrix concentration becomes scalar concentration). *Let h and S_0 be supplied by Proposition 3.14, put $e = e(h)$ and $N' = N - e$, and define*

$$B = \{x(s)/p : s \in S_0\} \subseteq R_{N'-1}.$$

Then $s \mapsto x(s)/p$ is injective and compatible with every congruence projection up to depth $N' - 1$. Consequently every congruence restriction placed on this same exact centralizer slice transfers to B with only the $p^{O(e)} = p^{O(\varepsilon N)}$ loss recorded above. This is the algebraic matrix-to-scalar transfer needed before the all-scale normalization below.

Proof. This is Lemma 3.12. The loss from passing to distinct reductions is at most p^{3e} , and $e = O(\varepsilon N)$. $\square$

3.9 Maximal-fiber normalization

Lemma 3.16 (Maximal concentrated fiber). *Let $A \subseteq R_M$ be nonempty and let $0 < \beta < 1$. Choose n_0 maximal such that some residue class $a_0 + p^{n_0} R_M$ satisfies*

$$|A \cap (a_0 + p^{n_0} R_M)| \geq p^{-\beta n_0} |A|. \quad (2.1)$$

Set

$$A_0 = A \cap (a_0 + p^{n_0} R_M), \quad L = M - n_0,$$

and identify

$$C := p^{-n_0}(A_0 - a_0) \subseteq R_L.$$

Then for every $1 \leq r \leq L$ and every $\xi \in R_r$,

$$|C \cap \pi_r^{-1}(\xi)| < p^{-\beta r} |C|, \quad (2.2)$$

and consequently

$$|\pi_r(C)| > p^{\beta r}. \quad (2.3)$$

Proof. If a fiber of C modulo p^r had cardinality at least $p^{-\beta r} |C|$, its inverse image inside A_0 would be a residue class modulo p^{n_0+r} containing at least

$$p^{-\beta r} |A_0| \geq p^{-\beta(n_0+r)} |A|,$$

contradicting maximality of n_0 . This proves (2.2). Since C is partitioned into $|\pi_r(C)|$ fibers and every fiber has size $< p^{-\beta r} |C|$, (2.3) follows. $\square$

Lemma 3.17 (The normalized modulus is macroscopic). *Assume*

$$|A| \geq p^{\alpha M - E}, \quad E \leq \eta M,$$

where $0 < \beta < \alpha - \eta$. Then the L supplied by Lemma 3.16 satisfies

$$L \geq \frac{\alpha - \beta - \eta}{1 - \beta} M. \quad (2.4)$$

In particular $L \geq cM$ for a fixed $c > 0$ once $\beta < \alpha/2$ and $\eta \ll \alpha$.

Proof. Since $A_0 \subseteq a_0 + p^{n_0} R_M$, one has $|A_0| \leq p^L$. On the other hand, (2.1) and the lower bound for A give

$$|A_0| \geq p^{-\beta n_0} p^{\alpha M - E}.$$

Thus

$$L \geq \alpha M - E - \beta n_0 = (\alpha - \beta)M - E + \beta L.$$

Rearranging gives (2.4). $\square$

Remark 3.18 (This is the missing profile mechanism). The point is that no projection information is inherited from the ambient matrix set. The all-scale profile (2.3) is manufactured *inside the exact commuting fiber* by replacing it with one maximal concentrated subfiber and changing the base point/scale. This is the local-ring normalization used in the original Bourgain–Gamburd prime-power argument.

3.10 Low-depth valuation inputs without a hidden coefficient cost

The scalar regularization/Fourier-completion stage requires exact valuation-0 and valuation-1 differences. They can be supplied before K3 using the same auxiliary digit system Ω already allowed in the scalar entropy argument.

Lemma 3.19 (No additive section modulo p^2). *Let $s : \mathbf{F}_p \rightarrow \mathbf{Z}/p^2\mathbf{Z}$ be any set-theoretic section of reduction modulo p with $s(0) = 0$. Then there exist $u, v \in \mathbf{F}_p$ such that*

$$s(u) + s(v) - s(u + v) \in p(\mathbf{Z}/p^2\mathbf{Z})^\times. \quad (3.1)$$

Proof. If every displayed defect vanished modulo p^2 , then s would be a homomorphism of additive groups. Hence

$$0 = s(0) = s(p \cdot 1) = p s(1) \pmod{p^2}.$$

But $s(1)$ is a unit modulo p , so $p s(1)$ is divisible by p and not by p^2 , a contradiction. $\square$

Corollary 3.20 (Uniform valuation-0, 1 auxiliary differences). *Let $\tilde{\Omega} = \Omega \cup \{0\}$ and choose the representatives in $\tilde{\Omega}$ modulo p^2 . Then a bounded scalar expression in $\tilde{\Omega}$ contains an element of exact valuation 0 and an element of exact valuation 1.*

Proof. Any unit of Ω has valuation 0. Apply Lemma 3.19 to the chosen section s with image in $\tilde{\Omega}$ modulo p^2 . The three-term expression in (3.1) has exact valuation 1. $\square$

Lemma 3.21 (Forcing the first two regular levels is legitimate). *Let $C \subseteq R_L$ be the normalized exact-fiber set from Lemma 3.16. In the analytic scalar stage, adjoining the bounded translates needed to force the first two active levels by the valuation-0, 1 elements of Corollary 3.20 changes cardinalities and projection sizes by at most an absolute bounded factor. Every new coefficient belongs to the bounded expression algebra generated by $\Omega \cup \{0\}$. Hence K3 eliminates these coefficients before the scalar result is passed to the matrix-realization module.*

Proof. The first-level forcing lemma in K2 uses at most two unions with translates $C + d_0$ and $C + d_1$, where $v_p(d_i) = i$. Thus the set grows by a factor at most 4, and no projection size decreases. Corollary 3.20 realizes d_0, d_1 by a bounded number of additions/subtractions of elements of $\tilde{\Omega}$. K3 was designed precisely to eliminate the finitely many external residue-digit coefficients occurring in a bounded scalar expression. Therefore these two preliminary translations create no p -dependent word length in the final matrix statement. $\square$

3.11 All-scale profile inside the exact commuting fiber

Theorem 3.22 (Exact-fiber all-scale profile). *Fix $\alpha > 0$ and choose $0 < \eta \ll \beta < \alpha/2$. Suppose the exact commuting scalar coordinate set $B_0 \subseteq R_M$ obtained above satisfies $|B_0| \geq p^{\alpha M - \eta M}$. Then after restricting to one congruence fiber, translating, and dividing by its common p -power, there is $C \subseteq R_L$ with $L \geq c(\alpha)M$ such that for every $1 \leq r \leq L$,*

$$\max_{\xi \in R_r} |C \cap \pi_r^{-1}(\xi)| < p^{-\beta r} |C|, \quad |\pi_r(C)| > p^{\beta r}.$$

Moreover the bounded residue-digit preprocessing above supplies exact valuation-0 and valuation-1 auxiliary differences. All these operations remain inside scalar coordinates of the same exact centralizer.

Proof. Apply Lemmas 3.16 and 3.17 to B_0 . The fiber nonconcentration and projection lower bounds are exactly (2.2)–(2.3). Since $|\pi_1(C)| > 1$, $C - C$ contains a unit; translation and unit normalization do not reduce projection sizes. Corollary 3.20 and Lemma 3.21 provide the two low-depth valuation inputs. No step uses projection information from the ambient matrix set after the exact fiber has been chosen. $\square$

Remark 3.23 (Why the normalization is performed inside the exact fiber). No projection information is inherited from the ambient matrix set. The theorem manufactures the all-scale profile only after the exact centralizer slice has been chosen, translated, and divided by its common p -power. The homogeneous lift-back after scalar ring operations is proved in the next chapter; separating normalization from lift-back keeps the dependency direction acyclic.

3.12 The connector for the later matrix-filling argument

Theorem 3.24 (Captured-set trace/torus/span package). *Let H be the approximate set extracted from failure of flattening and suppose*

$$\nu^{*L}(xH) \geq Q^{-C_0\varepsilon}, \quad L \asymp \log |G_N| \asymp N \log p.$$

After fixing the exponent hierarchy and taking ε sufficiently small, bounded powers of H have the following properties.

- (P1) *At every selected stable depth $n_1 \gg \varepsilon N$ they contain polynomially many elements with pairwise separated traces and discriminant valuation at most $d_{\text{reg}} = O(\varepsilon N)$, as in Proposition 3.5.*
- (P2) *Proposition 3.14 produces a rich exact commuting slice. After dividing the traceless part of the chosen element by its scalar depth $e < d_{\text{reg}}/2$, Proposition 3.15 gives a one-dimensional determinant-one coordinate modulo p^{N-e} , and Theorem 3.22 supplies an all-scale profile inside a normalized congruence subfiber of that same exact coordinate.*
- (P3) *For every primitive traceless matrix direction arising in the matrix realization, three conjugates lying in a bounded power of H span the traceless module after multiplication by $p^{d_{\text{sp}}}$, with $d_{\text{sp}} = O(\varepsilon N)$.*
- (P4) *For every fixed bounded-degree operator in the normalized traceless part of the chosen commuting element whose minimal nonzero Smith valuation is $d_T = O(d_{\text{reg}})$, a primitive seed has a bounded-power conjugate on which that operator is nonzero after at most $d_T + d_{\text{sp}} = O(\varepsilon N)$ depth loss.*

Proof. Use Lemma 3.2 to place the captured mass at the shorter times needed for the stable depth and auxiliary depth. Proposition 3.5 gives the controlled nonscalar trace family; Proposition 3.14 converts a rich conjugation fiber into an exact commuting slice; Proposition 3.15 transfers that slice to a scalar coordinate; and Theorem 3.22 performs the maximal-fiber normalization and produces the all-scale profile. The spanning and operator-escape assertions are Proposition 3.6 and Corollary 3.7. $\square$

Remark 3.25 (The nonconcentration mechanism). The argument does not infer regularity of an exact centralizer fiber from ambient entropy. Random-walk mass first creates controlled nonscalar and trace-separated elements, then a large exact centralizer fiber; only inside that fiber is the maximal-congruence normalization performed. This is the structural order needed before scalar sum-product and matrix realization.

3.13 Worked example: maximal-fiber normalization

Let $M = 6$ and suppose an exact centralizer coordinate happens to have the form

$$B_0 = a_0 + p^2 C \subset \mathbf{Z}/p^6 \mathbf{Z},$$

where

$$C = \left\{ c_0 + c_1 p + c_2 p^2 + c_3 p^3 : c_i \in D \right\} \subset \mathbf{Z}/p^4 \mathbf{Z}$$

for a digit set $D \subset \mathbf{F}_p$ of size $m > 1$. Dividing by the common p^2 factor does not mean that the group argument may divide by p^2 for free; it merely changes the scalar coordinate to a four-digit problem. In that normalized problem

$$|\pi_r(C)| = m^r, \quad \max_{\xi} |C \cap \pi_r^{-1}(\xi)| = m^{4-r}.$$

Thus with $\beta = \log_p m$ one has exact all-scale behavior $|\pi_r(C)| = p^{\beta r}$ and fiber proportion $p^{-\beta r}$.

The general maximal-fiber lemma manufactures this kind of estimate without a product digit model. Chapter 4 then homogenizes the final scalar expression; only at that point can the normalization be undone, at the honest cost $p^{D n_0}$ for homogeneous degree D .

Chapter 4

Multiscale Sum–Product in $\mathbf{Z}/p^N\mathbf{Z}$

Two distinctions drive the argument: the Rényi–Shannon inequality is used in the correct direction, and positive average branching is *not* converted into a fictitious long consecutive block of full digits. Growth is obtained from a branch-profile contradiction; a homogeneous ideal is obtained only later, by exact-conductor Fourier completion.

4.1 Scope and proof architecture

First reading versus second reading. On a first reading, follow only K1–K5, Theorem 4.20, the Fourier-completion theorem, and the final lift-back theorem. Treat the branch-profile inequalities as the mechanism that prevents a regular set from hiding its mass at incompatible depths. On a second reading, return to the exact entropy normalizations, shift-defect budget, and coefficient-elimination estimates. Those estimates are needed for uniformity, but they are not needed to see the logical shape of the argument.

Let

$$R_N := \mathbf{Z}/p^N\mathbf{Z}, \quad \pi_j : R_N \rightarrow \mathbf{Z}/p^j\mathbf{Z} \quad (0 \leq j \leq N),$$

and let p be an odd prime. All logarithms denoted $\log_p$ are to base p ; entropy without a subscript is natural-log entropy.

The prime-power argument needs a multiscale regularization package, an entropy-increment dichotomy on regular fibers, a scalar sum–product theorem at a regular p -adic scale, an approximate-ring inverse alternative, and eventually a thick homogeneous congruence segment. This first module establishes the following items for downstream use.

- K1.** The Shannon–collision comparison and a one-digit averaged scalar-sum entropy lemma are proved.
- K2.** A regular p -adic tree has an exact entropy/branching profile; failure of scalar growth gives a quantitative shift-budget inequality for that profile.
- K3.** An elementary Schnirelmann-density argument proves that the active branch levels form a bounded additive basis for all sufficiently deep shifts.
- K4.** Combining the shift budget with an averaged defect lower bound gives a contradiction. Thus a regular intermediate-dimensional set with lower projection bounds has scalar sum–product growth.

K5. The contrapositive yields a valid regular-scale inverse alternative: failure of growth forces either near-full ambient size or a multiscale projection concentration obstruction. No “long thick block” is asserted.

The regular scalar-growth theorem by itself does not yet produce a homogeneous ideal $p^r R_N$. That genuinely stronger conclusion is proved later in this chapter: exact regularization and Fourier completion pass from nonregular multiscale growth to a deep ideal, and the subsequent coefficient-elimination argument removes every auxiliary residue-field coefficient before the result is sent to the matrix-realization chapter. Keeping these steps separate prevents the false inference ‘large average branching $\Rightarrow$ thick digit block’.

4.2 Entropy preliminaries: the inequality direction

Definition 4.1 (Shannon and collision entropy). For a finite random variable X with masses q_x define

$$H(X) := - \sum_x q_x \log q_x, \quad H_2(X) := - \log \sum_x q_x^2.$$

Lemma 4.2 (Correct Shannon–collision comparison). *For every finite random variable X ,*

$$H_2(X) \leq H(X).$$

Equivalently,

$$e^{-H(X)} \leq \sum_x q_x^2.$$

Proof. Let X' be an independent copy of X . By Jensen’s inequality for the convex function $-\log t$,

$$H(X) = \mathbb{E}[-\log q_X] \geq -\log \mathbb{E}[q_X] = -\log \sum_x q_x^2 = H_2(X).$$

The equivalent exponential inequality follows immediately. $\square$

Remark 4.3. The reverse inequality $\sum q_x^2 \leq e^{-H(X)}$ is false in general. This is exactly the direction error that invalidated the old one-digit entropy-gain step.

4.2.1 A one-digit scalar collision estimate

Lemma 4.4 (Averaged one-digit collision). *Let $U, V \subseteq \mathbf{F}_p$ be nonempty. Let X, Y be independent and uniform on U, V , and let a be uniform on $\mathbf{F}_p^\times$, independent of X, Y . For $Z_a = X + aY$,*

$$\mathbb{E}_{a \in \mathbf{F}_p^\times} \sum_z \mathbb{P}(Z_a = z)^2 \leq \frac{1}{|U||V|} + \frac{1}{p-1}.$$

Consequently

$$\mathbb{E}_a H(Z_a) \geq -\log \left(\frac{1}{|U||V|} + \frac{1}{p-1} \right).$$

Proof. The collision probability is

$$\sum_z \mathbb{P}(Z_a = z)^2 = \mathbb{P}(X + aY = X' + aY'),$$

where X', Y' are independent copies. Average over a . If $Y = Y'$, equality forces $X = X'$, contributing exactly $1/(|U||V|)$. If $Y \neq Y'$, then for any $X \neq X'$ there is at most one $a \in \mathbf{F}_p^\times$ satisfying

$$a(Y - Y') = X' - X,$$

and if $X = X'$ there is none. Hence the remaining contribution is at most $1/(p-1)$. This proves the first inequality.

For each fixed a , Lemma 4.2 gives

$$H(Z_a) \geq -\log \sum_z \mathbb{P}(Z_a = z)^2.$$

Average in a and use Jensen for the convex function $-\log$ to obtain the second assertion. $\square$

Corollary 4.5 (One-digit growth/saturation dichotomy). *With the notation above,*

$$\mathbb{E}_a H(Z_a) \geq \log \left(\frac{1}{2} \min\{|U||V|, p-1\} \right).$$

Thus either the two input supports multiply their entropy, or the output digit is already within an additive constant of full $\mathbf{F}_p$ entropy.

Proof. For positive r, s , $(r^{-1} + s^{-1})^{-1} \geq \frac{1}{2} \min(r, s)$. Apply this with $r = |U||V|$ and $s = p-1$ in Lemma 4.4. $\square$

4.3 Regular p -adic trees

Definition 4.6 (Regular subset). A nonempty set $A \subseteq R_N$ is $(m_0, \dots, m_{N-1})$ -regular if, for every $0 \leq i < N$, every occupied residue class modulo p^i has exactly m_i occupied children modulo p^{i+1} . Here $1 \leq m_i \leq p$.

Lemma 4.7 (Projection and fiber formulas). *If A is $(m_0, \dots, m_{N-1})$ -regular, then*

$$|\pi_j(A)| = \prod_{i=0}^{j-1} m_i \quad (1 \leq j \leq N), \quad |A| = \prod_{i=0}^{N-1} m_i.$$

Moreover, if X is uniform on A , then conditional on the first i base- p digits of X , its i -th digit is uniform on a set of cardinality m_i .

Proof. Each occupied node has exactly m_i children, so the number of occupied nodes multiplies by m_i from level i to $i+1$. The conditional uniformity follows because every occupied child at level $i+1$ has exactly the same number $\prod_{j>i} m_j$ of descendant leaves. $\square$

Remark 4.8 (Worked example: A four-level regular tree in $\mathbf{Z}/81\mathbf{Z}$). Let $p = 3$, $N = 4$, and

$$A = \{0, 1, 2, 9, 10, 11\} \subset \mathbf{Z}/81\mathbf{Z}.$$

Its successive projection sizes are

$$|\pi_1(A)| = 3, \quad |\pi_2(A)| = 3, \quad |\pi_3(A)| = 6, \quad |\pi_4(A)| = 6.$$

Thus A is regular with branch profile

$$(m_0, m_1, m_2, m_3) = (3, 1, 2, 1), \quad |A| = 3 \cdot 1 \cdot 2 \cdot 1 = 6.$$

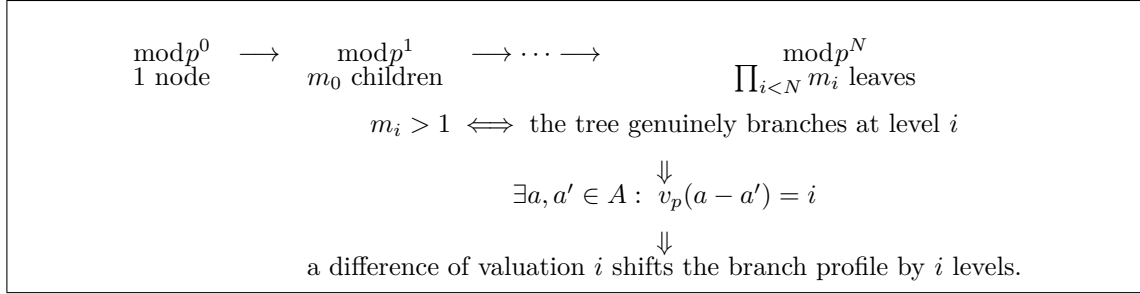


Figure 4.1: A regular p -adic tree. Branching levels are exactly the valuations realized by differences, which is why additive differences become scale selectors in the scalar sum-product argument.

The active levels are exactly $J = \{0, 2\}$. Indeed, $1 - 0 = 1$ has 3-adic valuation 0 and $9 - 0 = 9$ has valuation 2. On the other hand, two elements of A are either different modulo 3 (valuation 0) or, if they agree modulo 3, differ by 9 (valuation 2); no difference has valuation 1 or 3. This is the concrete content of the valuation lemma below.

Lemma 4.9 (Valuation levels of differences). *Let A be regular and put*

$$J := \{k \in [0, N) : m_k > 1\}.$$

Then $k \in J$ if and only if there exist $a, a' \in A$ with

$$v_p(a - a') = k.$$

Proof. If $m_k > 1$, choose an occupied prefix of length k and two distinct occupied children. Any descendant leaves a, a' of these children agree modulo p^k and differ modulo p^{k+1} , so $v_p(a - a') = k$. Conversely, if such a difference exists, the two leaves occupy distinct children above a common level- k prefix, so $m_k > 1$. $\square$

4.4 Entropy lifting from one digit to R_N

Choose a set $\Omega \subset R_N^\times$ whose reductions modulo p run bijectively through $\mathbf{F}_p^\times$.

Proposition 4.10 (Regular scalar-sum lower bound). *Let $A, B \subseteq R_N$ be regular with profiles (m_i) and (ℓ_i) . Then*

$$\max_{\alpha \in \Omega} |A + \alpha B| \geq \prod_{i=0}^{N-1} \max \left\{ 1, \left(\frac{1}{m_i \ell_i} + \frac{1}{p-1} \right)^{-1} \right\}.$$

Proof. Let X, Y be independent and uniform on A, B . Write $D_i(W) \in \mathbf{F}_p$ for the i -th base- p digit of $W \in R_N$. Fix $\alpha \in \Omega$ and put $Z_\alpha = X + \alpha Y$. By the entropy chain rule,

$$H(Z_\alpha) = \sum_{i=0}^{N-1} H(D_i(Z_\alpha) \mid D_{<i}(Z_\alpha)).$$

The lower output digits are functions of the lower input digits, hence conditioning on $D_{<i}(X), D_{<i}(Y)$ is finer. Therefore

$$H(D_i(Z_\alpha) \mid D_{<i}(Z_\alpha)) \geq H(D_i(Z_\alpha) \mid D_{<i}(X), D_{<i}(Y)).$$

After the lower input digits are fixed, the next digits of X, Y are uniform on sets of cardinalities m_i, ℓ_i by Lemma 4.7. The carry into the i -th digit is fixed by the lower digits, while the coefficient of the next digit of Y is $\bar{\alpha} \in \mathbf{F}_p^\times$. Translation by the carry does not change entropy. Averaging in α and applying Lemma 4.4 at every level gives

$$\mathbb{E}_\alpha H(Z_\alpha) \geq \sum_i \max \left\{ 0, -\log \left(\frac{1}{m_i \ell_i} + \frac{1}{p-1} \right) \right\}.$$

For each α , $H(Z_\alpha) \leq \log |A + \alpha B|$. Exponentiating completes the proof. $\square$

4.4.1 The shifted profile produced by a difference

For a regular set A put

$$x_i := \log_p m_i \in [0, 1], \quad x_i := 0 \quad (i < 0).$$

If $a \in A - A$ has $v_p(a) = k$, then aA has profile

$$(\underbrace{1, \dots, 1}_{k \text{ levels}}, m_0, m_1, \dots, m_{N-1-k}).$$

Consequently Proposition 4.10 yields the following.

Lemma 4.11 (Shift budget forced by failure of growth). *There is a quantity*

$$\kappa_p := \log_p 2 + |\log_p(1 - 1/p)| \xrightarrow{p \rightarrow \infty} 0$$

with the following property. Let A be regular and suppose that, for some $\tau > 0$,

$$|A + \alpha aA| < |A| p^{\tau N} \quad (\alpha \in \Omega, a \in A - A).$$

Then, for every $k \in J := \{i : x_i > 0\}$,

$$\sum_{i=0}^{N-1-k} \min(x_i, 1 - x_{i+k}) \leq N(\tau + \kappa_p). \quad ()$$

Proof. Choose $a \in A - A$ with $v_p(a) = k$ using Lemma 4.9. Proposition 4.10, divided by $|A| = \prod_i m_i$, gives

$$\frac{\max_\alpha |A + \alpha aA|}{|A|} \geq \prod_{i=k}^{N-1} \frac{1}{2} \min \left(m_{i-k}, \frac{p-1}{m_i} \right).$$

Taking $\log_p$ and using

$$\log_p \min \left(m_{i-k}, \frac{p-1}{m_i} \right) \geq \min(x_{i-k}, 1 - x_i) - |\log_p(1 - 1/p)|$$

gives

$$\tau N > \log_p \frac{\max_\alpha |A + \alpha aA|}{|A|} \geq \sum_{i=0}^{N-1-k} \min(x_i, 1 - x_{i+k}) - N\kappa_p.$$

$\square$

4.5 An elementary additive-basis lemma

No conclusion about consecutive high-branching digits is needed. The following bounded-additive-basis fact is sufficient.

Definition 4.12 (Schnirelmann density). For $X \subseteq \mathbf{Z}_{\geq 0}$ with $0 \in X$, define

$$\sigma(X) := \inf_{n \geq 1} \frac{|X \cap [1, n]|}{n}.$$

Lemma 4.13 (Elementary Schnirelmann inequality). *If $0 \in X, Y \subseteq \mathbf{Z}_{\geq 0}$, then*

$$\sigma(X + Y) \geq \sigma(X) + \sigma(Y) - \sigma(X)\sigma(Y).$$

Proof. Write $\alpha = \sigma(X)$ and list

$$0 = y_0 < y_1 < \cdots < y_s \leq n < y_{s+1} := n + 1$$

with $y_1, \dots, y_s$ the positive elements of $Y \cap [1, n]$. For each $0 \leq j \leq s$, the numbers

$$y_j + x, \quad x \in X \cap [0, y_{j+1} - y_j - 1],$$

lie in the disjoint interval $[y_j, y_{j+1} - 1]$ and belong to $X + Y$. Since $0 \in X$ and $|X \cap [1, r]| \geq \alpha r$,

$$|(X + Y) \cap [0, n]| \geq \sum_{j=0}^s (1 + \alpha(y_{j+1} - y_j - 1)) = \alpha(n + 1) + (1 - \alpha)(s + 1).$$

Removing the possible element 0 and using $s = |Y \cap [1, n]| \geq \sigma(Y)n$ yields

$$|(X + Y) \cap [1, n]| \geq n(\alpha + (1 - \alpha)\sigma(Y)),$$

which is the claimed bound. □

Corollary 4.14 (Positive density gives a bounded additive basis). *If $0 \in X \subseteq \mathbf{Z}_{\geq 0}$ and $\sigma(X) \geq \varepsilon > 0$, then*

$$hX = \mathbf{Z}_{\geq 0}$$

for some

$$h \leq 4 \left\lceil \frac{1}{\varepsilon} \right\rceil.$$

Here hX denotes sums of exactly h elements of X , with zeros allowed.

Proof. By Lemma 4.13, inductively

$$\sigma(rX) \geq 1 - (1 - \varepsilon)^r.$$

For $r = 2\lceil 1/\varepsilon \rceil$ the right side is greater than $1/2$. If a set $S \ni 0$ has $\sigma(S) > 1/2$, then every $n \geq 1$ lies in $S + S$: both $S \cap [0, n]$ and $n - S \cap [0, n]$ have more than half of the $n + 1$ possible integers, so they intersect. Thus $2rX = \mathbf{Z}_{\geq 0}$, and $2r \leq 4\lceil 1/\varepsilon \rceil$. □

Lemma 4.15 (Active branch levels generate all deep shifts). *Let $J \subseteq [0, N) \cap \mathbf{Z}$ contain 0, 1, and suppose*

$$|J \cap [0, \ell)| \geq \varepsilon \ell \quad (\delta N \leq \ell \leq N). \quad (\dagger)$$

Then there is

$$H \leq 8 \left\lceil \frac{1}{\varepsilon} \right\rceil$$

such that

$$[H\delta N, N) \cap \mathbf{Z} \subseteq HJ.$$

Proof. Let k_0 be the largest integer for which

$$|(J+1) \cap [1, k_0]| < \varepsilon k_0,$$

with $k_0 = 0$ if there is none. By $(\dagger)$, $k_0 < \delta N$. Maximality also implies that for every $\ell > k_0$,

$$|(J+1) \cap [k_0+1, \ell)| \geq \varepsilon(\ell - k_0),$$

and taking $\ell = k_0 + 1$ shows $k_0 \in J$.

Define

$$X := \{j - k_0 + 1 : j \in J, j \geq k_0\} \cup [N - k_0, \infty) \cup \{0\}.$$

The preceding inequality gives $\sigma(X) \geq \varepsilon$. By Corollary 4.14, $hX = \mathbf{Z}_{\geq 0}$ with $h \leq 4\lceil 1/\varepsilon \rceil$.

Take $m \in [h\delta N, N)$. Then $q := m - hk_0 \geq 0$ and $q < N - k_0$. A representation of q as a sum of h elements of X cannot use any element $\geq N - k_0$. If the nonzero terms are $j_s - k_0 + 1$ ($1 \leq s \leq t$), then

$$m = \sum_{s=1}^t j_s + (h-t)k_0 + t.$$

Since $j_s, k_0, 1, 0 \in J$, this is a sum of at most $h+t \leq 2h$ elements of J , padded with zeros. Taking $H = 2h$ proves the claim. $\square$

4.6 The branch-profile contradiction

Let $x_0, \dots, x_{N-1} \in [0, 1]$ satisfy

$$x_0, x_1 > 0, \quad (4.1)$$

$$\sum_{i=0}^{N-1} x_i \leq (1 - \varepsilon)N, \quad (4.2)$$

$$\sum_{i=0}^{\ell-1} x_i \geq \varepsilon \ell \quad (\delta N \leq \ell \leq N). \quad (4.3)$$

Put

$$J := \{i : x_i > 0\}, \quad L := \{i : x_i \geq 1/2\},$$

and, for $k \geq 0$,

$$D_L(k) := |(L+k) \cap [0, N)) \setminus L|.$$

Lemma 4.16 (Small shift budget implies small defect). *Suppose that for every $k \in J$,*

$$\sum_{i=0}^{N-1-k} \min(x_i, 1 - x_{i+k}) \leq \eta N. \quad (\ddagger)$$

Then

$$D_L(k) \leq 2\eta N \quad (k \in J).$$

Proof. If $i \in L$, $i + k < N$, and $i + k \notin L$, then $x_i \geq 1/2$ and $1 - x_{i+k} > 1/2$, so the corresponding summand in $(\ddagger)$ is at least $1/2$. There are exactly $D_L(k)$ such indices. $\square$

Lemma 4.17 (Subadditivity of the defect). *For $k_1, k_2 \geq 0$ with $k_1 + k_2 < N$,*

$$D_L(k_1 + k_2) \leq D_L(k_1) + D_L(k_2).$$

Proof. A point of $L + k_1 + k_2$ that is not in L is either already lost after the first shift by k_1 , or survives the first comparison and is lost after the second shift. Translating this set inclusion back by $k_1 + k_2$ gives the stated cardinality inequality. $\square$

Lemma 4.18 (Some deep shift has large defect). *There is an absolute $c > 0$ such that the following holds. If (4.2), (4.3), and the $k = 0$ case of $(\ddagger)$ hold, and*

$$\delta \leq \varepsilon/8, \quad \eta \leq c\varepsilon^2,$$

then for some

$$j_0 \in [c\varepsilon^3 N, N)$$

one has

$$D_L(j_0) \geq c\varepsilon^3 N.$$

Proof. The $k = 0$ budget says

$$\sum_i \min(x_i, 1 - x_i) \leq \eta N.$$

In particular

$$\sum_{i \notin L} x_i \leq \eta N, \quad \sum_{i \in L} (1 - x_i) \leq \eta N. \quad (1)$$

Let $\ell_0 = \lfloor \varepsilon N/4 \rfloor$. For N above an absolute ε -dependent threshold, $\ell_0 \geq \delta N$. From (4.3) and (1),

$$|L \cap [0, \ell_0)| \geq \sum_{i < \ell_0} x_i - \sum_{i \notin L} x_i \geq \varepsilon \ell_0 - \eta N \geq c_1 \varepsilon^2 N. \quad (2)$$

Also, by (4.2) and the second inequality in (1),

$$|L| \leq \sum_i x_i + \eta N \leq (1 - \varepsilon + \eta)N. \quad (3)$$

Thus

$$|(\ell_0, N) \setminus L| \geq c_2 \varepsilon N. \quad (4)$$

For each pair (k, i) with $k \in L \cap [0, \ell_0)$ and $i \in (\ell_0, N) \setminus L$, set $j = i - k$. Then $0 \leq j < N$ and $i \in (L + j) \setminus L$, so the pair contributes to $D_L(j)$. For fixed j the correspondence is injective. Summing over j and using (2)–(4),

$$\sum_{j=0}^{N-1} D_L(j) \geq c_3 \varepsilon^3 N^2.$$

The shifts $j < c_3 \varepsilon^3 N/2$ contribute at most $c_3 \varepsilon^3 N^2/2$, since $D_L(j) \leq N$. Hence among the remaining shifts there is a j_0 with $D_L(j_0) \geq c\varepsilon^3 N$. $\square$

Proposition 4.19 (Profile contradiction). *For every $0 < \varepsilon < 1/4$ there is $c_\varepsilon > 0$ with $c_\varepsilon \gg \varepsilon^5$ such that the following is impossible for N sufficiently large:*

(i) (4.1), (4.2), and (4.3) hold with $0 < \delta \leq c_\varepsilon$;

(ii) for every $k \in J$,

$$\sum_{i=0}^{N-1-k} \min(x_i, 1 - x_{i+k}) \leq \eta N$$

with $0 < \eta \leq c_\varepsilon$.

Proof. Let $H \ll 1/\varepsilon$ be given by Lemma 4.15. Lemma 4.18 gives $j_0 \geq c\varepsilon^3 N$ and $D_L(j_0) \geq c\varepsilon^3 N$. Choose c_ε so small that $H\delta N < c\varepsilon^3 N$. Then Lemma 4.15 writes

$$j_0 = b_1 + \cdots + b_H, \quad b_r \in J.$$

By subadditivity and Lemma 4.16,

$$c\varepsilon^3 N \leq D_L(j_0) \leq \sum_{r=1}^H D_L(b_r) \leq 2H\eta N.$$

Since $H \ll 1/\varepsilon$, this is impossible once $\eta \ll \varepsilon^4$. Taking c_ε smaller than all constants used above completes the proof. $\square$

4.7 Regular-scale scalar sum-product theorem

Theorem 4.20 (Regular p -adic scalar sum-product). *For every $0 < \varepsilon < 1/4$ there are*

$$c = c(\varepsilon) > 0, \quad p_0 = p_0(\varepsilon), \quad N_0 = N_0(\varepsilon)$$

such that the following holds for $p \geq p_0$ and $N \geq N_0$. Let $A \subseteq R_N$ be $(m_0, \dots, m_{N-1})$ -regular and suppose

(a) $m_0, m_1 > 1$;

(b) $|A| \leq p^{(1-\varepsilon)N}$;

(c) for every $\delta N \leq \ell \leq N$,

$$|\pi_\ell(A)| \geq p^{\varepsilon\ell},$$

where $0 < \delta \leq c$.

Then

$$\max_{\alpha \in \Omega, a \in A-A} |A + \alpha a A| \geq |A| p^{cN}.$$

Proof. Set $x_i = \log_p m_i$. Conditions (a)–(c) are exactly (4.1), (4.2), and (4.3). Suppose the conclusion fails with exponent c . Lemma 4.11 then gives (‡) with

$$\eta = c + \kappa_p.$$

Choose first $c \leq c_\varepsilon/2$ from Proposition 4.19, then p_0 so large that $\kappa_p \leq c_\varepsilon/2$. Proposition 4.19 gives a contradiction. $\square$

Remark 4.21 (Proof architecture). The proof never infers a long consecutive interval of indices on which x_i is close to 1. The high-branch set L may have holes at every scale. What matters is instead:

$$\text{small scalar growth} \implies D_L(k) \text{ small for every active } k,$$

while lower projection dimension makes the active set J an additive basis for deep shifts, and the global dimension deficit forces at least one deep shift to have large defect. This is the robust multiscale contradiction.

4.8 Regular-scale inverse alternative

Corollary 4.22 (Regular-scale inverse alternative). *Fix $0 < \varepsilon < 1/4$ and let c, p_0, N_0 be as in Theorem 4.20. Let $p \geq p_0$, $N \geq N_0$, and let $A \subseteq R_N$ be regular with $m_0, m_1 > 1$. If*

$$\max_{\alpha \in \Omega, a \in A-A} |A + \alpha a A| < |A| p^{cN},$$

then at least one of the following holds:

- (i) dense regime: $|A| > p^{(1-\varepsilon)N}$;
- (ii) projection concentration: for some $\ell \in [\delta N, N]$,

$$|\pi_\ell(A)| < p^{\varepsilon\ell};$$

- (iii) low-depth degeneracy: $m_0 = 1$ or $m_1 = 1$.

Proof. This is the contrapositive of Theorem 4.20. □

Remark 4.23 (Why this replaces the false thick-block alternative). Projection concentration is the correct obstruction at this stage. A set can have a positive average number of branches and still have bad digits evenly spaced through all scales, so average branching alone cannot imply a macroscopic consecutive interval of full digits. A homogeneous ideal or thick congruence segment requires an additional bounded-generation/Fourier-completion argument; it is not a consequence of the branch profile by itself.

4.9 Entropy-profile regularization interface

This section records the exact interface needed to pass from arbitrary sets to the regular fibers used in the subsequent modules.

Proposition 4.24 (Entropy profile of an arbitrary set). *Let $A \subseteq R_N$ and let X be uniform on A . Write $D_i(X)$ for its digits and define*

$$h_i := \frac{1}{\log p} H(D_i(X) \mid D_{<i}(X)) \in [0, 1].$$

Then for every $1 \leq \ell \leq N$,

$$\sum_{i=0}^{\ell-1} h_i = \frac{H(\pi_\ell X)}{\log p} \leq \log_p |\pi_\ell(A)|,$$

and at $\ell = N$ equality holds:

$$\sum_{i=0}^{N-1} h_i = \log_p |A|.$$

If X is uniform on a regular set, then $h_i = x_i = \log_p m_i$ exactly.

Proof. The chain rule gives the first identity. Entropy is bounded by the logarithm of support size. At level N , X is uniform on A , hence $H(X) = \log |A|$. For a regular set, Lemma 4.7 makes every conditional digit uniform on exactly m_i points, so its entropy is $\log m_i$. $\square$

Proposition 4.25 (Entropy-increment/regular-fiber dichotomy). *Let $U, V \subseteq \mathbf{F}_p$ be nonempty. With X, Y, a as in Lemma 4.4, put*

$$u = \log_p |U|, \quad v = \log_p |V|.$$

Then

$$\frac{1}{\log p} \mathbb{E}_a H(X + aY) \geq \min(u + v, 1) - \kappa_p.$$

In particular, at a regular p -adic fiber either the two conditional digit entropies add up (up to κ_p), or the output digit is saturated to within κ_p of full entropy.

Proof. Corollary 4.5 gives

$$\mathbb{E}_a H(X + aY) \geq \log \frac{1}{2} \min(|U||V|, p - 1).$$

Divide by $\log p$ and absorb the factors 2 and $1 - 1/p$ into κ_p . $\square$

Remark 4.26 (From regular to nonregular fibers). Propositions 4.24 and 4.25 explicitly use the Rényi–Shannon comparison in the needed direction. For an exactly regular block, the cumulative profile conditions required later are simply the projection inequalities of Theorem 4.20; no recursive “deficiency” argument is needed.

For a general nonregular set one still has to retain enough mass and enough projection lower bounds while passing to a regular block. This is done below by Lemmas 4.28 and 4.29.

4.10 The regular scalar input and its scope

Put $R_N = \mathbf{Z}/p^N\mathbf{Z}$. We shall use the regular scalar theorem proved above in the following downstream form.

Proposition 4.27 (Regular scalar theorem in downstream form). *For each $0 < \varepsilon < 1/4$ there are $c_r(\varepsilon) > 0$, $p_r(\varepsilon)$ and $N_r(\varepsilon)$ such that, for $p \geq p_r$, an exactly regular $A \subseteq R_N$ satisfying*

$$m_0, m_1 > 1, \quad |A| \leq p^{(1-\varepsilon)N}, \quad |\pi_j(A)| \geq p^{\varepsilon j} \quad (\delta N \leq j \leq N),$$

with $\delta \leq c_r(\varepsilon)$ obeys

$$\max_{\alpha, a} |A + \alpha a A| \geq |A| p^{c_r(\varepsilon)N},$$

where α ranges over fixed lifts of $\mathbf{F}_p^\times$ and $a \in A - A$.

Proof. This is exactly Theorem 4.20, with its constants renamed for downstream use. $\square$

No proof from XI.5–XI.7 is used below. The constants in this section are deliberately those of a *large- p vertical theorem*: $p \geq p_0(\varepsilon)$ is chosen after the dimension parameters. The bounded-residue-characteristic range is handled separately by Theorem 6.18, and the two ranges are assembled in Theorem 7.11. Thus no statement in this section is used outside its displayed large- p range.

4.11 Exact regular subtree extraction

We view a subset of R_N as leaves in the rooted p -ary tree of residue classes modulo p^j .

Lemma 4.28 (Bottom-up exact regularization). *Let $A \subseteq R_N$ be nonempty and set*

$$Q_p := 2(1 + \lceil \log_2 p \rceil).$$

Then there exists an exactly regular subset $A' \subseteq A$ such that

$$|A'| \geq Q_p^{-N} |A|.$$

Proof. We process the tree from level $N - 1$ upward to the root. Suppose levels $i + 1, \dots, N - 1$ have already been regularized, so every surviving node at level $i + 1$ has the same number M of surviving descendant leaves. For a node v at level i , let $d(v)$ be the number of its surviving children. If $d(v) > 0$, place v into the dyadic class

$$2^s \leq d(v) < 2^{s+1}, \quad 0 \leq s \leq \lceil \log_2 p \rceil.$$

The surviving leaves are partitioned among at most $1 + \lceil \log_2 p \rceil$ such classes, weighted by $d(v)M$. Choose a class containing at least that reciprocal fraction of the current leaves. Delete all other level- i nodes. For every retained v , retain exactly 2^s of its surviving children and all of their already regular descendants. Since $d(v) < 2^{s+1}$, this last pruning loses at most a factor 2.

Thus the i -th step loses at most the factor Q_p and makes the number of children at level i exactly constant, without disturbing regularity at deeper levels. After N steps the surviving leaves form an exactly regular set of size at least $Q_p^{-N} |A|$. $\square$

4.11.1 Projection lower bounds survive under small additive growth

Lemma 4.29 (Projection-preserving regularization). *Fix $0 < \delta < \varepsilon < 1$. There exists $p_0 = p_0(\varepsilon, \delta)$ such that the following holds for $p \geq p_0$. Let $0 < \delta' \leq \varepsilon\delta/4$ and suppose $A \subseteq R_N$ satisfies*

$$|\pi_j(A)| \geq p^{\varepsilon j} \quad (\delta'N \leq j \leq N) \tag{K2.1}$$

and

$$|A + A| \leq |A|p^{\delta'N}. \tag{K2.2}$$

Then A has an exactly regular subset $A' \subseteq A$ such that

$$|A'| \geq Q_p^{-N} |A| \tag{K2.3}$$

and

$$|\pi_j(A')| \geq p^{\varepsilon j/2} \quad (\delta N \leq j \leq N). \tag{K2.4}$$

Proof. Take A' from Lemma 4.28. Suppose (K2.4) fails at some $n \geq \delta N$. Since A' is regular, every occupied residue class modulo p^n contains the same number of leaves, so there is a fiber $A'' \subseteq A'$ with

$$|A''| = \frac{|A'|}{|\pi_n(A')|} > |A|Q_p^{-N}p^{-\varepsilon n/2}.$$

Choose one representative $x_\xi \in A$ from each residue $\xi \in \pi_n(A)$. The translates $A'' + x_\xi$ lie in distinct residue classes modulo p^n , hence are disjoint. Therefore

$$|A + A| \geq |A''| |\pi_n(A)| > |A|Q_p^{-N}p^{\varepsilon n/2}.$$

Combining with (K2.2),

$$Q_p^N > p^{\varepsilon n/2 - \delta'N} \geq p^{N(\varepsilon\delta/2 - \delta')} \geq p^{\varepsilon\delta N/4}.$$

But $\log_p Q_p \rightarrow 0$ as $p \rightarrow \infty$, so for p_0 large enough $Q_p < p^{\varepsilon\delta/8}$, contradiction. $\square$

4.12 Forcing the first two branch levels

Lemma 4.30 (Forcing nontrivial levels 0 and 1). *Let $B \subseteq R_N$ be exactly regular. Suppose a larger ambient set C has differences $d_0, d_1 \in C - C$ with*

$$v_p(d_0) = 0, \quad v_p(d_1) = 1.$$

If the level- r branching number of B equals 1 for $r \in \{0, 1\}$, then replacing B by $B + \{0, d_r\}$ changes that branching number from 1 to 2 and leaves all other branching numbers unchanged. Applying this at most twice produces a regular set B^* with $m_0, m_1 > 1$ and

$$B^* \subseteq B + \{0, d_0\} + \{0, d_1\}.$$

Proof. If $v_p(d_r) = r$, translation by d_r fixes every prefix of length r and changes the next digit by a nonzero element of $\mathbf{F}_p$. When the original level- r branching is 1, the two copies B and $B + d_r$ therefore occupy two distinct children above every occupied level- r node. Below level r the copies coincide at the level of prefixes; above level $r + 1$ each child carries a translate of the same regular descendant tree. Hence only m_r changes, from 1 to 2. $\square$

4.13 A nonregular scalar-growth theorem

For bookkeeping, call a set a *bounded scalar-ring expression of complexity C from A* if it is obtained from $A \cup \{0\}$ by at most C applications of union, addition, subtraction, multiplication, and multiplication by an element of the fixed lift set Ω of $\mathbf{F}_p^\times$. We care only that C depends on the dimension parameters, not on p or N .

Proposition 4.31 (Nonregular one-step growth). *For every $0 < \varepsilon < 1/8$ there are $c = c(\varepsilon) > 0$, $C = C(\varepsilon)$, $p_0 = p_0(\varepsilon)$ and $N_0 = N_0(\varepsilon)$ with the following property. Let $p \geq p_0$, $N \geq N_0$, and let $A \subseteq R_N$ contain 0. Assume*

- (i) $|A| \leq p^{(1-\varepsilon)N}$;
- (ii) $|\pi_j(A)| \geq p^{2\varepsilon j}$ for every $\delta_0 N \leq j \leq N$, where $\delta_0 = \delta_0(\varepsilon) > 0$ is sufficiently small;
- (iii) $A - A$ contains elements of exact valuations 0 and 1.

Then there exists a bounded scalar-ring expression D of complexity at most C from A such that

$$A \subseteq A \cup D, \quad |A \cup D| \geq |A|p^{cN}.$$

Proof. Let $\varepsilon_r = \varepsilon/2$, and let $c_r = c_r(\varepsilon_r)$ and δ_r be the constants from Theorem 4.27. Choose a small $\delta' \leq \varepsilon\delta_r/4$ and then choose p_0 so large that both Lemma 4.29 and

$$Q_p \leq p^{c_r/4} \tag{4.1}$$

hold.

Set

$$c_0 := \min\{\delta'/2, c_r/4\}.$$

If $|A + A| \geq |A|p^{c_0N}$, take $D = A + A$ and we are done. Otherwise $|A + A| < |A|p^{\delta'N}$, so Lemma 4.29, applied with projection exponent 2ε and target depth δ_r , gives an exactly regular $B \subseteq A$ such that

$$|B| \geq Q_p^{-N}|A|, \quad |\pi_j(B)| \geq p^{\varepsilon j} \quad (\delta_r N \leq j \leq N).$$

Use Lemma 4.30 and the valuation-0, 1 differences from $A - A$ to obtain a regular B^* with $m_0, m_1 > 1$. It has at least $|B|$ elements and no smaller projections. Since $|B^*| \leq 4|A|$ and N is large,

$$|B^*| \leq p^{(1-\varepsilon/2)N}.$$

Theorem 4.27 gives $\alpha \in \Omega$ and $a \in B^* - B^*$ such that

$$|B^* + \alpha a B^*| \geq |B^*| p^{c_r N} \geq |A| Q_p^{-N} p^{c_r N} \geq |A| p^{3c_r N/4},$$

using (4.1). The displayed set is a bounded scalar-ring expression from A of absolute complexity, because B^* is contained in A plus at most two differences from A . Taking $D = B^* + \alpha a B^*$ proves the result with $c = \min\{c_0, c_r/2\}$ after harmlessly decreasing the exponent. $\square$

4.14 Exact-conductor Fourier completion

Let $e_N(x) = \exp(2\pi i x/p^N)$. Every additive character of R_N is $x \mapsto e_N(zx)$ for a unique $z \in R_N$.

Definition 4.32 (Character conductor). For $z \neq 0$, write $v = v_p(z)$ with $0 \leq v < N$ and set

$$q_z := p^{N-v}.$$

The character $e_N(z \cdot)$ factors to a primitive character modulo q_z .

Exact-conductor census and normalization. For later use it is worth recording the Fourier bookkeeping without suppressing any normalization. Characters of exact conductor p^r are precisely those $z \in R_N$ with $v_p(z) = N - r$, so their number is

$$p^r - p^{r-1} = \varphi(p^r) < p^r. \quad (5.0)$$

If ν is a probability measure on R_N , with $\hat{\nu}(z) = \sum_x \nu(x) e_N(zx)$, then

$$\nu(t) = p^{-N} \sum_{z \in R_N} \hat{\nu}(z) e_N(-zt).$$

Thus a conductor- p^r bound may be summed shell by shell; no factor depending on N is hidden in the Fourier inversion step.

Lemma 4.33 (Bilinear Fourier decay from fiber nonconcentration). *Let $A, B \subseteq R_N$, and suppose that for every $1 \leq r \leq N$ and every residue class ξ modulo p^r ,*

$$|A \cap \pi_r^{-1}(\xi)| \leq p^{-\gamma_1 r} |A|, \quad |B \cap \pi_r^{-1}(\xi)| \leq p^{-\gamma_2 r} |B|. \quad (5.1)$$

If $z \neq 0$, then

$$|\mathbb{E}_{a \in A, b \in B} e_N(zab)| \leq q_z^{-(\gamma_1 + \gamma_2 - 1)/2}.$$

Proof. Let $q = q_z = p^r$ and project A, B modulo q . Their normalized fiber-count functions η_A, η_B on $\mathbf{Z}/q\mathbf{Z}$ have ℓ^1 norm 1 and, by (5.1),

$$\|\eta_A\|_\infty \leq q^{-\gamma_1}, \quad \|\eta_B\|_\infty \leq q^{-\gamma_2}.$$

Hence

$$\|\eta_A\|_2 \leq q^{-\gamma_1/2}, \quad \|\eta_B\|_2 \leq q^{-\gamma_2/2}.$$

The primitive character induced by z is $u \mapsto e_q(cu)$ with c a unit. By Cauchy–Schwarz and Parseval on $\mathbf{Z}/q\mathbf{Z}$,

$$\begin{aligned} \left| \sum_{x,y} \eta_A(x) \eta_B(y) e_q(cxy) \right|^2 &\leq \|\eta_A\|_2^2 \sum_x \left| \sum_y \eta_B(y) e_q(cxy) \right|^2 \\ &= \|\eta_A\|_2^2 q \|\eta_B\|_2^2 \\ &\leq q^{1-\gamma_1-\gamma_2}. \end{aligned}$$

Taking square roots proves the claim. $\square$

Theorem 4.34 (Fourier product completion). *Let $\gamma_1, \gamma_2 > 0$ with $\gamma_1 + \gamma_2 > 1$. Let $A_j, B_j \subseteq R_N$ ($1 \leq j \leq k$) all satisfy (5.1) with exponents γ_1, γ_2 . If*

$$\beta := \frac{k(\gamma_1 + \gamma_2 - 1)}{2} \geq 3,$$

then

$$A_1 B_1 + \cdots + A_k B_k = R_N.$$

Proof. Let ν be the distribution of $S = X_1 Y_1 + \cdots + X_k Y_k$ for independent uniform $X_j \in A_j$, $Y_j \in B_j$. Lemma 4.33 and independence give, for every nontrivial character of conductor p^r ,

$$|\hat{\nu}(z)| \leq p^{-r\beta}.$$

By the exact-conductor census (5.0), the nontrivial part of Fourier inversion satisfies

$$\left| \sum_{z \neq 0} \hat{\nu}(z) e_N(-zt) \right| \leq \sum_{r=1}^N (p^r - p^{r-1}) p^{-r\beta} < \sum_{r=1}^N p^{r(1-\beta)}.$$

Since $\hat{\nu}(0) = 1$, we therefore have, for every $t \in R_N$,

$$\nu(t) \geq p^{-N} \left(1 - \sum_{r=1}^N p^{r(1-\beta)} \right).$$

If $\beta \geq 3$, the sum is at most $\sum_{r \geq 1} 2^{-2r} = 1/3$. Hence $\nu(t) \geq (2/3)p^{-N} > 0$ for every t , proving full support. Notice that the conclusion is quantitative: the completed sum is not merely supported on all of R_N ; its law is pointwise bounded below by a fixed multiple of Haar measure. $\square$

Corollary 4.35 (Twelve products suffice at exponent 3/4). *If $B \subseteq R_N$ satisfies*

$$\max_{\xi} |B \cap \pi_r^{-1}(\xi)| \leq p^{-3r/4} |B| \quad (1 \leq r \leq N),$$

then

$$\underbrace{BB + \cdots + BB}_{12 \text{ summands}} = R_N.$$

Proof. Apply Theorem 4.34 with $\gamma_1 = \gamma_2 = 3/4$ and $k = 12$, for which $\beta = 3$. $\square$

4.15 Dense sets generate a homogeneous deep ideal

Theorem 4.36 (Dense-set deep-ideal completion). *Let $0 < \rho < 1/16$. If $A \subseteq R_N$ satisfies*

$$|A| \geq p^{(1-\rho)N},$$

then there is an integer $r \leq 8\rho N$ such that

$$p^r R_N \subseteq \underbrace{(A - A)(A - A) + \cdots + (A - A)(A - A)}_{12 \text{ summands}}. \quad (6.1)$$

Proof. Set $\gamma = 3/4$. Let n_0 be the largest integer $0 \leq n_0 \leq N$ for which some residue class modulo p^{n_0} contains at least

$$p^{-\gamma n_0} |A|$$

elements of A ; $n_0 = 0$ always qualifies. Such a fiber has at most p^{N-n_0} elements, so

$$p^{N-n_0} \geq p^{-\gamma n_0} |A| \geq p^{N(1-\rho)-\gamma n_0}.$$

Thus

$$n_0 \leq \frac{\rho}{1-\gamma} N = 4\rho N. \quad (6.2)$$

Choose a maximizing fiber $A' = A \cap (a_0 + p^{n_0} R_N)$ and rescale it to

$$B := \left\{ \frac{a - a_0}{p^{n_0}} : a \in A' \right\} \subseteq R_{N-n_0}.$$

By maximality of n_0 , for every $1 \leq s \leq N - n_0$ each fiber of B modulo p^s has size strictly less than $p^{-3s/4} |B|$. Corollary 4.35 therefore gives

$$\underbrace{BB + \cdots + BB}_{12} = R_{N-n_0}.$$

There are two quotient levels in this last step, so we spell out the lift. The identity $\sum_{12} BB = R_{N-n_0}$ means that every residue class $u \bmod p^{N-n_0}$ is represented by such a sum. Multiplication by p^{2n_0} sends these representatives onto the ideal $p^{2n_0} R_N$: indeed an element $p^{2n_0} y \bmod p^N$ depends only on $y \bmod p^{N-2n_0}$, and every such class has a lift modulo p^{N-n_0} . Since $p^{n_0} B \subseteq A - A$, multiplying the completed identity by p^{2n_0} therefore gives

$$p^{2n_0} R_N \subseteq \underbrace{(A - A)(A - A) + \cdots + (A - A)(A - A)}_{12}.$$

Take $r = 2n_0$ and use (6.2). □

4.16 Bounded scalar-ring generation at prime-power depth

Theorem 4.37 (Large- p multiscale bounded generation). *Fix $0 < \varepsilon_2 < \varepsilon_1/4 < 1/16$. There exist*

$$C = C(\varepsilon_1, \varepsilon_2), \quad \theta = \theta(\varepsilon_1, \varepsilon_2) > 0, \quad p_0, \quad N_0$$

such that for $p \geq p_0$, $N \geq N_0$, the following holds. Let $A \subseteq R_N$ contain 0 and assume

- (i) $|\pi_j(A)| \geq p^{\varepsilon_1 j}$ for every $\delta_0 N \leq j \leq N$, where $\delta_0 = \delta_0(\varepsilon_1, \varepsilon_2)$ is sufficiently small;

(ii) $A - A$ contains elements of exact valuations 0 and 1.

Then some bounded scalar-ring expression $\mathcal{W}$ of complexity at most C from A satisfies

$$p^{\lceil \theta N \rceil} R_N \subseteq \mathcal{W}.$$

One may take $\theta = 8\varepsilon_2$ after decreasing δ_0 and increasing C, p_0, N_0 .

Proof. Construct sets C_t inductively, starting with $C_0 = A$, and always keep $C_t \subseteq C_{t+1}$ by adjoining the new bounded expression found at each step. Because C_t contains A , the projection lower bounds and the valuation-0, 1 differences persist.

As long as

$$|C_t| \leq p^{(1-\varepsilon_2)N},$$

Proposition 4.31, applied with $\varepsilon = \varepsilon_2$ (and using $\varepsilon_1 \geq 2\varepsilon_2$ for the projection hypothesis), supplies a bounded expression whose union with C_t has size at least

$$|C_t|p^{c(\varepsilon_2)N}.$$

Hence after at most $\lceil 1/c(\varepsilon_2) \rceil$ stages we reach a set C_T with

$$|C_T| \geq p^{(1-\varepsilon_2)N}.$$

The total expression complexity accumulated in this bounded number of stages depends only on $\varepsilon_1, \varepsilon_2$.

Apply Theorem 4.36 to C_T with $\rho = \varepsilon_2$. It gives

$$p^r R_N \subseteq \sum_{12} (C_T - C_T)(C_T - C_T), \quad r \leq 8\varepsilon_2 N.$$

The right side is again a bounded scalar-ring expression from A . Since $p^r R_N \supseteq p^{\lceil 8\varepsilon_2 N \rceil} R_N$, the theorem follows with $\theta = 8\varepsilon_2$. $\square$

Remark 4.38 (No long digit block). The proof does not manufacture any consecutive interval of fully branching digits. It proceeds instead by

$$\begin{aligned} \text{regularization} &\longrightarrow \text{scalar growth} \longrightarrow \text{bounded iteration} \\ &\longrightarrow \text{dense regime} \longrightarrow \text{Fourier completion.} \end{aligned}$$

This separation is precisely what prevents an unjustified passage from average branching to a consecutive thick block.

4.17 Eliminating auxiliary residue-field coefficients

The regularization/Fourier-completion argument used an auxiliary lift set Ω of residue-field units in the entropy-growth step. That is mathematically harmless on the scalar side, but a later matrix realization must not pay a word-length cost depending on the numerical size of a chosen coefficient $\alpha \in \Omega$. The purpose of this module is to remove every such external scalar before the scalar result is passed to the group.

The argument is internal and rank-one. It has two ingredients:

- (i) a collision in a scalar linear combination eliminates one coefficient at the cost of a controlled increase in congruence depth;
- (ii) if no such collision exists, the coefficient-free part is forced to be a subring modulo a deeper scale. In $\mathbf{Z}/p^m\mathbf{Z}$, every additive subgroup is a principal ideal, and the active-level valuation argument from the regular scalar theory supplies an element of the correct exact valuation.

This is the precise coefficient-elimination mechanism needed before the torus-action module.

4.18 Coefficient-free bounded ring expressions

Fix

$$R_N = \mathbf{Z}/p^N\mathbf{Z}.$$

For $A \subseteq R_N$ containing $0, 1$, let $\mathcal{R}_C(A)$ denote the set of values of expressions formed from elements of A using at most C additions/subtractions and at most C multiplications. We choose the convention so that the family is nested in C . There is an absolute increasing function F such that

$$\mathcal{R}_C + \mathcal{R}_C \subseteq \mathcal{R}_{F(C)}, \quad \mathcal{R}_C \mathcal{R}_C \subseteq \mathcal{R}_{F(C)}, \quad \mathcal{R}_C - \mathcal{R}_C \subseteq \mathcal{R}_{F(C)}.$$

Only boundedness of $F^{O(1)}(C)$ matters below.

Definition 4.39 (Coefficient-rich bounded generation). We say $\text{BG}(A; \varepsilon, k, C)$ holds if there are units $\alpha_0 = 1, \alpha_1, \dots, \alpha_k \in R_N^\times$ such that

$$p^{\lceil \varepsilon N \rceil} R_N \subseteq \alpha_0 \mathcal{R}_C(A) + \alpha_1 \mathcal{R}_C(A) + \dots + \alpha_k \mathcal{R}_C(A). \quad (\text{BG})$$

Lemma 4.40 (K2 supplies bounded coefficient-rich generation). *Under the K2 multiscale hypotheses, after scaling the scalar set by one of its unit elements so that $0, 1 \in A$, there are*

$$\varepsilon_0 > 0, \quad k_0, C_0 < \infty$$

depending only on the fixed dimension parameters, and a threshold p_0 , such that for every $p \geq p_0$,

$$\text{BG}(A; \varepsilon_0, k_0, C_0)$$

holds. By taking the K2 dense threshold sufficiently close to full density, ε_0 may be made as small as required after the other fixed parameters are chosen.

Proof. K2 produces a bounded scalar-ring expression containing a homogeneous ideal $p^q R_N$, with $q \leq \varepsilon_0 N$. Expand that finite expression distributively. Every leaf belonging to the auxiliary lift set Ω is a fixed unit coefficient; after collecting identical coefficient products, the expression is contained in a sum of at most $k_0 + 1$ unit multiples of one sufficiently large coefficient-free set $\mathcal{R}_{C_0}(A)$. Multiplying the whole inclusion by the inverse of one coefficient, if necessary, makes the first coefficient equal to 1 without changing the ideal. The number of coefficient products and the coefficient-free complexity are bounded because the K2 expression tree has bounded depth. The final assertion is exactly the parameter freedom in K2's dense completion. $\square$

4.19 The scalar-coefficient reduction

Let

$$L(x_0, \dots, x_k) = \sum_{i=0}^k \alpha_i x_i.$$

Lemma 4.41 (One coefficient is eliminated by a shallow collision). *Assume $\text{BG}(A; \varepsilon, k, C)$ with $k \geq 1$. Fix $t \geq 4$. If there exist*

$$\mathbf{x}, \mathbf{x}' \in \mathcal{R}_{F(C)}(A)^{k+1}$$

with $L(\mathbf{x}) = L(\mathbf{x}')$ and with some coordinate difference satisfying

$$v_p(x_j - x'_j) < t\varepsilon N,$$

then, after changing the coefficient normalization and enlarging the bounded complexity,

$$\text{BG}(A; (t+1)\varepsilon, k-1, C')$$

holds, where C' depends only on C .

Proof. Put $d_i = x_i - x'_i$ and choose j with $r := v_p(d_j) < t\varepsilon N$. Since

$$\sum_i \alpha_i d_i = 0,$$

we may solve for the j -th term. Also

$$d_j p^{\lceil \varepsilon N \rceil} R_N = p^{\lceil \varepsilon N \rceil + r} R_N \supseteq p^{\lceil (t+1)\varepsilon N \rceil} R_N.$$

Multiplying (BG) by d_j gives a sum of terms $\alpha_i d_j \mathcal{R}_C(A)$. For the j -th term use

$$\alpha_j d_j = - \sum_{i \neq j} \alpha_i d_i.$$

Thus the j -th coefficient disappears and every remaining coefficient multiplies a set contained in a bounded coefficient-free enlargement of A , because $d_i, d_j \in \mathcal{R}_{F(C)}(A) - \mathcal{R}_{F(C)}(A)$. Renormalize one surviving unit coefficient to 1. The resulting complexity is a fixed function of C and the number of external coefficients has decreased by one. $\square$

Definition 4.42 (Deep injectivity). We say L is $t\varepsilon$ -injective on $\mathcal{R}_{F(C)}(A)^{k+1}$ if

$$L(\mathbf{x}) = L(\mathbf{x}') \implies \mathbf{x} \equiv \mathbf{x}' \pmod{p^{\lfloor t\varepsilon N \rfloor}}$$

coordinatewise.

Lemma 4.43 (Injectivity forces a coefficient-free subring). *Assume $\text{BG}(A; \varepsilon, k, C)$ and define the explicit coefficient-free domain*

$$\mathcal{D}_C(A) := \mathcal{R}_C(A) \cup (\mathcal{R}_C(A) + \mathcal{R}_C(A)) \cup (\mathcal{R}_C(A) \mathcal{R}_C(A)).$$

Suppose L is $t\varepsilon$ -injective on $\mathcal{D}_C(A)^{k+1}$. Put

$$q = \lceil \varepsilon N \rceil, \quad m = \lfloor t\varepsilon N \rfloor, \quad S = \pi_m(\mathcal{R}_C(A) \cap p^q R_N).$$

Then S is an additive subgroup of R_m closed under multiplication; hence S is a (possibly nonunital) subring of R_m .

Proof. The set is symmetric, so it suffices to prove closure under addition and multiplication. Take $x, x' \in \mathcal{R}_C(A) \cap p^q R_N$.

Since $x + x' \in p^q R_N$, (BG) gives $\mathbf{y} \in \mathcal{R}_C(A)^{k+1}$ with

$$L(\mathbf{y}) = x + x'.$$

On the other hand

$$L(x + x', 0, \dots, 0) = x + x'.$$

Both tuples lie in $\mathcal{D}_C(A)^{k+1}$. Deep injectivity gives

$$y_0 \equiv x + x' \pmod{p^m}, \quad y_i \equiv 0 \pmod{p^m} \quad (i > 0).$$

Because $m > q$, y_0 is also divisible by p^q . Thus the class of $x + x'$ modulo p^m belongs to S .

The same argument applies to xx' : by definition $xx' \in \mathcal{D}_C(A)$, and it lies in $p^{2q} R_N \subseteq p^q R_N$. Represent xx' by (BG), compare with the tuple $(xx', 0, \dots, 0)$, and use injectivity. Hence S is closed under multiplication. $\square$

4.20 The active levels provide exact coefficient-free valuations

We isolate the only additional consequence of Module K needed here.

Lemma 4.44 (Coefficient-free valuation basis). *Under the stable regular profile supplied by the regular scalar theorem, after choosing its regularization scale sufficiently small, there are constants $c_v > 0$ and $C_v < \infty$ such that for every integer*

$$c_v N \leq r < N$$

there exists

$$z_r \in \mathcal{R}_{C_v}(A) \quad \text{with} \quad v_p(z_r) = r.$$

Moreover c_v can be made smaller than any prescribed fixed positive constant by taking the regularization scale sufficiently fine.

Proof. Let J be the active digit levels of the regular profile. If $j \in J$, two elements with the same first j digits and different j -th digit give a difference of exact valuation j . The internal Schnirelmann-density argument for the regular scalar profile proves that every sufficiently deep integer r is a sum of at most $H = O(1)$ active levels. Multiplying the corresponding differences gives a coefficient-free ring expression of exact valuation equal to the sum, as long as the sum is $< N$. The lower endpoint of the deep-basis range is H times the regularization depth and therefore can be made arbitrarily small as a fraction of N . $\square$

4.21 Coefficient-free thick segment

Theorem 4.45 (Uniform coefficient elimination). *Fix the dimension parameters from the regularization/Fourier-completion argument. There exist constants*

$$\lambda > 0, \quad C_* < \infty, \quad p_0 < \infty$$

with the following property. For every $p \geq p_0$ and all sufficiently large N , let $A \subseteq R_N$ satisfy the normalized scalar regularization and Fourier-completion hypotheses and contain $0, 1$. Then there are integers

$$0 \leq r < m \leq N, \quad m - r \geq \lambda N,$$

such that

$$\boxed{p^r R_m \subseteq \pi_m(\mathcal{R}_{C_*}(A))}. \quad (\text{CF})$$

No coefficient outside the ring generated by A occurs in (CF).

Proof. Take the initial $\text{BG}(A; \varepsilon_0, k_0, C_0)$ from Lemma 4.40. Increase C_0 at the outset, if necessary, so that $C_0 \geq C_v$ and so that the injectivity test domain contains all sums and products used in Lemma 4.43; enlarging the coefficient-free set preserves (BG). Fix $t \geq 4$. Choose ε_0 so small that

$$(t+1)^{k_0} \varepsilon_0 < \frac{1}{10},$$

and choose the Module-K regularization scale so that

$$c_v < 2\varepsilon_0.$$

Starting at $(\varepsilon_0, k_0, C_0)$, test deep injectivity. If it fails, Lemma 4.41 replaces

$$(\varepsilon_i, k_i, C_i) \quad \text{by} \quad (\varepsilon_{i+1}, k_i - 1, C_{i+1}), \quad \varepsilon_{i+1} = (t+1)\varepsilon_i.$$

After at most k_0 reductions either all coefficients have disappeared or the process stops at an injective stage. All complexities remain bounded.

If all coefficients disappear, then for some $\varepsilon \leq (t+1)^{k_0} \varepsilon_0 < 1/10$,

$$p^{\lceil \varepsilon N \rceil} R_N \subseteq \mathcal{R}_C(A).$$

Take $m = N$ and $r = \lceil \varepsilon N \rceil$; then $m - r \geq 0.9N$.

Suppose instead the process stops with $k > 0$ at an injective stage with parameter ε . Lemma 4.43 gives a subring

$$S = \pi_m(\mathcal{R}_C(A) \cap p^q R_N), \quad q = \lceil \varepsilon N \rceil, \quad m = \lfloor t\varepsilon N \rfloor.$$

Choose

$$r = \lceil 2\varepsilon N \rceil.$$

Our parameter choices ensure $c_v N \leq r < m$. Lemma 4.44 supplies $z_r \in \mathcal{R}_{C_v}(A)$ of exact valuation r . By the initial complexity enlargement and its propagation through the reduction steps, $C \geq C_v$ at the injective stage, so z_r already belongs to the coefficient-free set defining S . Its residue belongs to S .

As an additive subgroup of the cyclic group R_m , the subgroup generated by an element of exact valuation r is exactly $p^r R_m$. Since S is closed under addition and negation,

$$p^r R_m \subseteq S \subseteq \pi_m(\mathcal{R}_{C_*}(A)).$$

Moreover

$$m - r \geq (t-2)\varepsilon N - O(1) \geq (t-2)\varepsilon_0 N - O(1).$$

Thus (CF) holds with, for instance, $\lambda = (t-2)\varepsilon_0/2$ for large N . □

Corollary 4.46 (Uniform homogeneous realization format). *In Theorem 4.45, every target element of $p^r R_m$ may be represented by a coefficient-free expression in A whose monomials all have one common degree $D = O(1)$, with a uniformly bounded number of monomials.*

Proof. By definition, every value of $\mathcal{R}_{C_*}(A)$ is represented by a polynomial with coefficients $0, \pm 1$, a bounded number of monomials, and bounded degree. Since $0, 1 \in A$, pad every monomial with copies of 1 to the common maximal degree D , and pad the number of summands with zero monomials. This changes neither the value nor the bounded complexity. $\square$

Corollary 4.47 (Scale-adjustable coefficient-free segment). *Fix $0 < \tau < 1/4$ and $R \geq 1$. Under the hypotheses of Theorem 4.45, the fixed K2/K3 parameter hierarchy may be chosen so that there are constants*

$$\lambda = \lambda(\tau, R) > 0, \quad C_{\tau, R} < \infty, \quad p_{\tau, R} < \infty,$$

with the following property. For every $p \geq p_{\tau, R}$ and all sufficiently large N there are integers $0 \leq q < m \leq N$ and $\ell = m - q$ such that

$$\lambda N \leq \ell < \tau N, \quad q < \tau N, \quad q = 0 \text{ or } \ell > Rq, \quad (\text{K3.1})$$

and

$$p^q R_m \subseteq \pi_m(\mathcal{R}_{C_{\tau, R}}(A)). \quad (\text{K3.2})$$

Every element of the ideal in (K3.2) has a coefficient-free realization whose monomials have one common bounded degree.

Proof. Choose an integer

$$t > 2R + 4.$$

In the proof of Theorem 4.45, let k_0 be the bounded number of external coefficients supplied by Lemma 4.40. Use the dense-threshold freedom in that lemma to choose the initial exponent ε_0 so that

$$(t+1)^{k_0} \varepsilon_0 < \min \left\{ \frac{\tau}{4t}, \frac{\tau}{8(R+1)} \right\}. \quad (\text{K3.3})$$

At every coefficient-reduction stage the current exponent satisfies

$$\varepsilon \leq (t+1)^{k_0} \varepsilon_0. \quad (\text{K3.4})$$

If all external coefficients disappear, Theorem 4.45 produces

$$p^{q_0} R_N \subseteq \mathcal{R}_C(A), \quad q_0 = \lceil \varepsilon N \rceil.$$

Put

$$q = q_0, \quad \ell = \left\lfloor \frac{\tau N}{2} \right\rfloor, \quad m = q + \ell.$$

For large N , (K3.3)–(K3.4) give $m < N$, $q < \tau N$, and

$$\ell > Rq.$$

Projecting the displayed inclusion from R_N to R_m gives (K3.2).

If the reduction stops at an injective stage, the proof of Theorem 4.45 gives

$$q_0 = \lceil 2\varepsilon N \rceil, \quad m_0 = \lfloor t\varepsilon N \rfloor, \quad p^{q_0} R_{m_0} \subseteq \pi_{m_0}(\mathcal{R}_C(A)).$$

Hence

$$\ell_0 := m_0 - q_0 \geq (t-2)\varepsilon N - 2.$$

Because $t > 2R + 4$, for all large N ,

$$\ell_0 > Rq_0. \quad (\text{K3.5})$$

Equations (K3.3)–(K3.4) also give $m_0 < \tau N$. Take $(q, m, \ell) = (q_0, m_0, \ell_0)$ in this branch.

Finally, in either branch the current exponent is at least the initial ε_0 , so for all large N one may take

$$\lambda = \min \left\{ \frac{\tau}{3}, \frac{(t-2)\varepsilon_0}{2} \right\} > 0.$$

The common-degree realization is Corollary 4.46; enlarging its bounded complexity over the finitely many reduction branches gives $C_{\tau, R}$. $\square$

Remark 4.48 (Undoing a unit normalization). If the original scalar coordinate set B contains a unit u but not necessarily 1, apply the theorem to $A = u^{-1}B$. A homogeneous degree- D expression in A becomes, after replacing each $a = b/u$, the same coefficient-free expression in B multiplied by the common unit u^{-D} . Hence the set of values realized from the original B still contains the same principal ideal, up to a unit, which is the identical ideal $p^r R_m$.

4.22 Undoing exact-fiber normalization after homogeneous scalar generation

Theorem 4.49 (Homogeneous lift-back from the normalized commuting fiber). *Use the set $C \subseteq R_L$ from Theorem 3.22. Apply the scalar regularization, growth, Fourier-completion, and coefficient-elimination theorems above. If the final coefficient-free expression has common degree D and contains $p^r R_L$, while $C = p^{-n_0}(B_0 - a_0)$, then*

$$p^{Dn_0+r} R_{Dn_0+L} \subseteq P(B_0 - a_0).$$

The thickness is still $L - r$. Hence if the auxiliary scalar modulus is $M = \theta N$ and $(D+1)\theta < 1$, the lifted segment remains strictly inside the ambient group modulus p^N .

Proof. Every degree- D monomial in normalized leaves $c = (b - a_0)/p^{n_0}$ becomes, after multiplication by p^{Dn_0} , the same monomial in the original differences $b - a_0$. The coefficient-elimination theorem makes every monomial have the same degree, so the whole scalar polynomial lifts with the single common factor p^{Dn_0} . Since $n_0, L \leq M = \theta N$, one has $Dn_0 + L \leq (D+1)\theta N < N$. $\square$

4.23 Worked example: why average branching does not give a thick block

Consider an exactly regular p -adic tree of depth $2m$ with branching number p at every even level and branching number 1 at every odd level. Its logarithmic branch profile is

$$x_i = \begin{cases} 1, & i \text{ even,} \\ 0, & i \text{ odd.} \end{cases}$$

The average profile is $1/2$, so the set has size p^m . Nevertheless there is no interval of length 2 on which every digit is highly branching, let alone a macroscopic consecutive block of almost-full digits. This is the simplest model showing why positive average branching alone cannot imply a long thick block.

The branch-profile argument asks a different question. If every scalar sum fails to grow, active levels force small shift-defects of the high-branch set. But active levels form a bounded additive basis for deep shifts, while a direct averaging argument produces a deep shift with large defect. The contradiction is global in the branch profile and requires no consecutive full-digit block. Only after growth reaches the dense regime does Fourier completion create a homogeneous ideal.

Chapter 5

From Scalar Ideals to Deep Principal Congruence Subgroups

The scalar theorem now has to be realized by honest SL_2 words. The key rank-one identity uses large elements of an exact centralizer rather than near-identity torus commutators, thereby avoiding an artificial congruence-depth loss at each multiplication. Transverse adjoint directions fill a complete traceless layer, and a restart argument propagates that layer to a genuine deep principal congruence subgroup.

5.1 Torus differences realize scalar multiplication without artificial depth loss

Fix the controlled-nonscalar h , its scalar depth e , normalized traceless part z , and $N' = N - e$. For $s \in S_0$, Lemma 3.12 gives

$$x(s) \in pR_{N'}, \quad B = \{x(s)/p : s \in S_0\} \subseteq R_{N'-1}.$$

Thus, modulo $p^{N'}$, every original commuting scalar has the form

$$x(s_b) = pb, \quad b \in B.$$

The missing e digits cause no terminal error in the group calculation: every seed used below has depth $t \geq e$, so a coefficient error divisible by p^{N-e} becomes divisible by p^N after multiplication by p^t . The K/K2/K3 scalar package may enlarge B by bounded scalar expressions; in the matrix realization every such expression is expanded back into its original leaves $b \in B$.

Definition 5.1 (Conjugation difference). For $s \in C_{G_N}(h)$ put

$$\mathcal{D}_s(y) = sys^{-1}(s^{-1}y^{-1}s).$$

Lemma 5.2 (One scalar factor). *Suppose*

$$y \equiv 1 + p^t \xi \pmod{p^{t+W}}, \quad W \leq t, \quad t \geq e.$$

Then for $s_b \in S_0$,

$$\mathcal{D}_{s_b}(y) \equiv 1 + p^{t+1}b[z, \xi] \pmod{p^{\min(t+W, N)}},$$

provided $W > 1$.

Proof. First-order multiplication, valid because $W \leq t$, gives

$$\mathcal{D}_s(y) \equiv 1 + p^t(\text{Ad}(s)\xi - \text{Ad}(s^{-1})\xi) \pmod{p^{\min(t+W, N)}}.$$

By Lemma 3.11, the parenthesis is congruent to $pb[z, \xi]$ modulo p^{N-e} . Multiplying the coordinate error by p^t kills it modulo p^N because $t \geq e$. This proves the formula. $\square$

Lemma 5.3 (Monomials by iterated torus differences). *Let d be fixed and suppose $W > d$. Iterating conjugation differences gives*

$$\mathcal{D}_{s_{b_d}} \cdots \mathcal{D}_{s_{b_1}}(y) \equiv 1 + p^{t+d}(b_1 \cdots b_d) \text{ad}(z)^d \xi \pmod{p^{t+W}}.$$

The word length depends only on d .

Proof. Induct using Lemma 5.2. Exact centralization with h , together with Lemma 3.11, ensures that the same normalized operator $\text{ad}(z)$ occurs at each stage; all truncation errors vanish because the current seed depth is at least e . $\square$

Theorem 5.4 (Coefficient-free homogeneous scalar segment from K3). *Fix parameters*

$$0 < \tau < 1/4, \quad R \geq 1.$$

Apply Corollary 4.47 to the normalized scalar coordinate set and then undo the unit normalization. There are constants

$$\lambda = \lambda(\tau, R) > 0, \quad D = D(\tau, R) < \infty, \quad C_s = C_s(\tau, R) < \infty,$$

and thresholds $p_0(\tau, R), N_0(\tau, R)$ such that for every $p \geq p_0$ and $N \geq N_0$ there are integers

$$0 \leq q < m \leq N' - 1, \quad \ell := m - q,$$

with

$$\lambda N \leq \ell < \tau N, \quad q < \tau N, \quad q = 0 \text{ or } \ell > Rq, \tag{L5.1}$$

for which the following holds: for every $z \in \mathbf{Z}/p^\ell \mathbf{Z}$ there is a coefficient-free scalar expression $P_z(B)$ in the original coordinate set B , of complexity at most C_s , such that

$$P_z(B) \equiv p^q z \pmod{p^m}. \tag{L5.2}$$

Every monomial of every P_z has the same multiplicative degree D , and the number of monomials is bounded by C_s .

Proof. Corollary 4.47 gives, after unit normalization, an ideal $p^q R_m$ satisfying (L5.1) and a coefficient-free common-degree realization of every element of that ideal. Let $u \in B$ be the unit used for the normalization, so that the normalized set is $A = u^{-1}B$. If

$$Q_z(A) = \sum_{\nu=1}^J \epsilon_\nu a_{\nu,1} \cdots a_{\nu,D} \equiv p^q z \pmod{p^m}, \quad \epsilon_\nu \in \{\pm 1\},$$

then replacing $a_{\nu,j} = u^{-1}b_{\nu,j}$ gives

$$u^D Q_z(u^{-1}B) = \sum_{\nu=1}^J \epsilon_\nu b_{\nu,1} \cdots b_{\nu,D}.$$

Multiplication by the unit u^D permutes the ideal $p^q R_m$; hence as z runs over R_ℓ , the right-hand side still realizes every class $p^q z \pmod{p^m}$. Thus no scalar outside the coefficient-free algebra of B is required. The bounds on D, J and the expression complexity are the uniform bounds in Corollary 4.47.

Finally, $N' = N - e$ with $e = O(\varepsilon N)$. Choose the scalar-loss parameter ε after τ, R so that $e < (\tau/4)N$. If the corollary initially returns $m > N' - 1$, project the ideal to

$$m' := \min\{m, N' - 1\}.$$

Because (L5.1) leaves a fixed linear margin and $e = o_{\tau, R}(N)$ under this parameter choice, the projected interval still has thickness at least $(\lambda/2)N$ for all sufficiently large N . Replacing λ by $\lambda/2$ gives the stated bounds. $\square$

Proposition 5.5 (Scalar ideal gives a restartable thick line with controlled operator loss). *Use Theorem 5.4 and put*

$$s = D + q, \quad \ell = m - q.$$

Let $d_{\text{sp}} = O(\varepsilon N)$ be the adjoint-spanning loss from the captured-set trace/escape package and let d_D be the Smith loss of $\text{ad}(z)^D$ from Lemma 3.13. Put

$$d_{\text{op}} = d_D + d_{\text{sp}}, \quad d_{\text{tot}} = d_{\text{op}} + d_{\text{sp}}.$$

Choose the exponent hierarchy so that

$$d_{\text{tot}} < \ell/10, \quad 2s + 6d_{\text{tot}} < \ell.$$

Suppose a primitive seed

$$y \equiv 1 + p^t \xi \pmod{p^{2t}}$$

is available with $t \geq e$, $s + \ell \leq t$, and $t + s < N$. Then, after one bounded-power conjugation supplied by the captured-set trace/escape package, there is a primitive direction β such that, writing

$$a = t + s + d_{\text{op}}, \quad b = \min(t + s + \ell, N),$$

one has

$$\{1 + p^a z \beta : z \in \mathbf{Z}/p^{b-a}\mathbf{Z}\} \subseteq H^{C_1} K_b.$$

Before the terminal step the line has width $b - a = \ell - d_{\text{op}}$.

Proof. Lemma 3.13 gives a primitive X_0 with

$$\text{ad}(z)^D X_0 = p^{d_D} Y_0, \quad Y_0 \text{ primitive.}$$

Apply the controlled linear-kernel escape corollary of the captured-set trace/escape package to $T = \text{ad}(z)^D$ and the primitive seed direction ξ . After conjugating the seed by a bounded-power element of H we may write

$$\text{ad}(z)^D \xi = p^{d_{\text{seed}}} \beta_0, \quad d_{\text{seed}} \leq d_{\text{op}},$$

with β_0 primitive. Conjugation preserves exact congruence depth.

Every coefficient-free expression supplied by Theorem 5.4 has all monomials of the same degree D . Lemma 5.3 realizes each monomial at depth $t + D + d_{\text{seed}}$ in the same direction β_0 . Restrict the scalar ideal by the additional factor $p^{d_{\text{op}} - d_{\text{seed}}}$. Since the K3 output contains all of $p^q R_m$, the remaining values still contain a full ideal of thickness at least $\ell - d_{\text{op}}$. Addition and subtraction of the monomial words at one depth therefore realize every coefficient modulo $p^{\ell - d_{\text{op}}}$ at leading depth

$$t + D + q + d_{\text{op}} = t + s + d_{\text{op}} = a.$$

Put $\beta = \beta_0$. The scalar-coordinate truncation contributes no error modulo p^N because $t \geq e$, and $b \leq t + s + \ell \leq 2t$ is in the first-order range. $\square$

5.2 Transverse directions and a full traceless layer

Lemma 5.6 (Adjoint spanning). *For $p > 3$ and $0 \neq X \in \mathfrak{sl}_2(\mathbf{F}_p)$, the $\mathrm{SL}_2(\mathbf{F}_p)$ -span of the adjoint orbit of X is all of $\mathfrak{sl}_2(\mathbf{F}_p)$. Hence three conjugates form a basis.*

Proof. The span is a nonzero adjoint-invariant subspace, hence an ideal in the simple Lie algebra $\mathfrak{sl}_2(\mathbf{F}_p)$. $\square$

Proposition 5.7 (Three controlled thick lines fill one congruence layer). *Under Proposition 5.5, apply the captured-set trace/escape package to the primitive direction β . There are three conjugators in a bounded power of H such that, with determinant loss d_{sp} ,*

$$p^{d_{\mathrm{sp}}} \mathfrak{sl}_2(\mathbf{Z}/p^{b-a}\mathbf{Z}) \subseteq \sum_{i=1}^3 (\mathbf{Z}/p^{b-a}\mathbf{Z}) \beta_i,$$

where β_i are the three conjugates of β . If

$$b - a - d_{\mathrm{sp}} \leq a + d_{\mathrm{sp}},$$

then

$$K_{a+d_{\mathrm{sp}}} \subseteq H^{C_3} K_b.$$

Proof. Conjugate the line from Proposition 5.5 by the three elements supplied by the captured-set trace/escape package. Normality of K_b preserves the error term. The controlled adjoint-spanning inclusion says that every traceless first coordinate of the form $p^{d_{\mathrm{sp}}} X$ modulo p^{b-a} is a linear combination of the three line directions. Lemma 3.8 realizes the corresponding sum at depth a , producing

$$1 + p^{a+d_{\mathrm{sp}}} X \pmod{p^b}.$$

For $g \in K_{a+d_{\mathrm{sp}}}$, Lemma 3.9 makes its first coordinate traceless modulo $p^{b-a-d_{\mathrm{sp}}}$; the displayed inequality is precisely the first-order range needed to cancel that coordinate and leave an element of K_b . $\square$

5.3 The terminal seed-propagation argument

The next two lemmas show that no fresh seed needs to be assumed at later depths.

Lemma 5.8 (A stable projection block supplies the initial controlled seed). *Let $H \subseteq G_N$ be symmetric. Let*

$$4s + 10d_{\mathrm{tot}} < r_0 < r_1 \leq N - s$$

and assume

$$|\pi_{r_1}(H)| > |\pi_{r_0}(H)|.$$

Then H^2 contains an element

$$y \equiv 1 + p^t \xi \pmod{p^{t+1}}$$

for some $r_0 \leq t < r_1$, with $\xi \neq 0$ in $\mathfrak{sl}_2(\mathbf{F}_p)$. No transversality condition is imposed here; Proposition 5.5 chooses the bounded-power conjugate on which $\mathrm{ad}(z)^D$ has controlled valuation.

Proof. The strict inequality of projection sizes gives two elements $g_1, g_2 \in H$ with the same image modulo p^{r_0} but different images modulo p^{r_1} . Thus $y = g_1^{-1}g_2 \in H^2 \cap K_{r_0}$ is nontrivial modulo p^{r_1} ; let t be its exact congruence depth. Its leading coordinate is nonzero and traceless by Lemma 3.9. The leading coordinate is already primitive. The controlled operator escape is deliberately postponed to Proposition 5.5; conjugation there changes neither exact depth nor bounded-word control. $\square$

Remark 5.9 (Where the projection inequality comes from – no hidden seed owner). The initial matrix seed is independent of the scalar-profile normalization and therefore also supplies the restart seeds used after Theorem 3.22. The captured-set trace/escape package gives, at every selected depth $r \geq C\varepsilon N$ in the short-word window,

$$|\pi_r(H^2)| \geq p^{c_0 r}.$$

The ambient quotient gives $|\pi_{r_0}(H^2)| \leq |G_{r_0}| \leq p^{3r_0 + O(1)}$. Choose $r_0 = C\varepsilon N$ and then $r_1 = \delta N$ with $c_0 r_1 > 4r_0$; the exponent hierarchy takes $\varepsilon \ll \delta$. For N large,

$$|\pi_{r_1}(H^2)| > |\pi_{r_0}(H^2)|.$$

Lemma 5.8, applied to H^2 , therefore gives a primitive seed at an exact depth in $[r_0, r_1)$. The stronger profile statement is now supplied by Theorem 3.22 as summarized in Remark 3.23.

Lemma 5.10 (A filled layer manufactures the next primitive seed). *Fix s, d_{op} and d_{sp} , and put $d_{\text{tot}} = d_{\text{op}} + d_{\text{sp}}$. Suppose a nonterminal stage gives*

$$K_A \subseteq H^C K_b, \quad A = t + s + d_{\text{tot}}, \quad b = t + s + \ell < N,$$

and assume $\ell > s + 2d_{\text{tot}}$. Then H^C contains an element

$$y_1 \equiv 1 + p^{t_1} \xi_1 \pmod{p^b}, \quad t_1 = b - s - d_{\text{tot}} = t + \ell - d_{\text{tot}},$$

with ξ_1 primitive.

Proof. The inequality gives $A < t_1 < b$. Choose any primitive traceless matrix X modulo p . Since Proposition 5.7 fills the entire quotient K_A/K_b , it supplies a word $w \in H^C$ satisfying

$$w \equiv 1 + p^{t_1} X \pmod{p^b}.$$

The coefficient is primitive, so the exact congruence depth is t_1 . $\square$

Theorem 5.11 (Exact deep principal congruence absorption). *Let $c_0 > 0$ be a fixed seed-scale constant for which Lemma 5.8 supplies a primitive seed at some $t_0 \in [c_0 N, 2c_0 N]$. Apply Theorem 5.4 with*

$$\tau := \frac{c_0}{8}, \quad R := 40,$$

and put $s = D + q$ and $\ell = m - q$. Let $d_{\text{op}}, d_{\text{sp}}$ be as in Proposition 5.5, and write

$$d_{\text{tot}} := d_{\text{op}} + d_{\text{sp}} \leq C_{\text{loss}} \varepsilon N.$$

Choose ε after c_0 and the scalar parameters so that

$$8C_{\text{loss}} \varepsilon < \frac{\lambda(c_0/8, 40)}{4}. \tag{L5.3}$$

Then, for all sufficiently large N ,

$$2s + 8d_{\text{tot}} < \ell, \quad s + \ell < c_0 N, \quad e \leq c_0 N. \tag{L5.4}$$

Consequently a bounded product power of H contains a principal congruence subgroup:

$$\boxed{K_{m_0} \subseteq H^{C_*}, \quad m_0 = t_0 + s + d_{\text{tot}} \leq 3c_0N.}$$

The number of restart stages and C_* depend only on the fixed hierarchy $(c_0, \tau, R, C_{\text{loss}}, \lambda)$, not on p or N .

Proof. First verify (L5.4). From (L5.1),

$$q < \frac{c_0N}{8}, \quad \ell < \frac{c_0N}{8}, \quad \ell > 40q, \quad \ell \geq \lambda N.$$

Since $D = O_{c_0}(1)$, for large N we have $2D < \ell/20$; hence

$$2s = 2(D + q) < \frac{\ell}{20} + \frac{\ell}{20} = \frac{\ell}{10}.$$

Equation (L5.3) and $\ell \geq \lambda N$ give $8d_{\text{tot}} < \ell/4$, so $2s + 8d_{\text{tot}} < \ell$. Also

$$s + \ell = D + q + \ell < D + \frac{c_0N}{4} < c_0N$$

for large N . Finally $e = O(\varepsilon N)$, so decreasing ε once more gives $e \leq c_0N$.

At stage j , Proposition 5.5 gives a line beginning at

$$a_j = t_j + s + d_{\text{op}}$$

and ending at

$$b_j = \min\{t_j + s + \ell, N\}.$$

Proposition 5.7 therefore gives

$$K_{A_j} \subseteq H^{C_{\text{layer}}} K_{b_j}, \quad A_j = t_j + s + d_{\text{tot}}, \quad (\text{L5.5})$$

with C_{layer} independent of j, p, N . If $b_j = N$, then $K_{b_j} = K_N = \{1\}$ and the induction terminates.

If $b_j < N$, Lemma 5.10 supplies the next primitive seed at

$$t_{j+1} = t_j + \ell - d_{\text{tot}}. \quad (\text{L5.6})$$

Its next filled layer starts exactly where the preceding quotient ends:

$$A_{j+1} = t_{j+1} + s + d_{\text{tot}} = t_j + s + \ell = b_j. \quad (\text{L5.7})$$

Thus the quotient inclusions telescope without a depth gap. Since $d_{\text{tot}} < \ell/8$,

$$t_{j+1} - t_j = \ell - d_{\text{tot}} > \frac{7\ell}{8} \geq \frac{7\lambda N}{8}.$$

Hence at most

$$J_* \leq 1 + \frac{8}{7\lambda}$$

nonterminal restart stages occur. Every stage has $t_j \geq t_0 \geq e$, so an error divisible by p^{N-e} in the normalized scalar coordinate becomes divisible by p^N after multiplication by p^{t_j} ; no hidden truncation error accumulates.

Substitute (L5.5) backwards using (L5.7). Since there are at most J_* stages,

$$K_{A_0} \subseteq H^{J_* C_{\text{layer}}}.$$

Here $A_0 = t_0 + s + d_{\text{tot}} = m_0$, which proves the claimed inclusion with $C_* = J_* C_{\text{layer}}$. The bound $m_0 \leq 3c_0N$ follows from $t_0 \leq 2c_0N$, $s < c_0N/4$ and $d_{\text{tot}} < c_0N/4$ after the parameter choices above. $\square$

Remark 5.12 (Restarting the congruence induction). No fresh seed needs to be assumed at each macroscopic depth. Lemma 5.10 derives that seed from the previously filled layer, while the invariant torus identity keeps the scalar-multiplication mechanism valid at the new depth. This is the rank-one exponent form of the terminal congruence induction in the classical prime-power Bourgain–Gamburd argument.

5.4 Prime-power approximate-group growth/absorption

Theorem 5.13 (Slow growth forces deep congruence absorption). *In the uniform large-prime regime, let $A \subseteq G_N$ be symmetric and contain 1. Suppose A is in the captured/slow-growth regime to which the captured-set trace/escape package applies, and suppose the $K/K2/K3$ scalar package multiscale scalar hypotheses hold on the resulting commuting slice. Then there are fixed $\varepsilon, c > 0$ and $C < \infty$ such that either*

$$|A^3| \geq |A|^{1+\varepsilon},$$

or

$$K_m \subseteq A^C \quad \text{for some } m \leq cN.$$

Proof. If growth fails, the earlier noncommutative Ruzsa/BSG owner supplies a controlled symmetric approximate set H containing the relevant portion of A . The captured-set trace/escape package and Proposition 3.14 give an exact rich commuting slice with scalar defect $e = O(\varepsilon N)$; Proposition 3.15 gives its exact scalar coordinate and the captured-set trace/escape package supplies the controlled adjoint-spanning loss, while Lemma 3.13 bounds the fixed-degree operator loss. Theorem 3.22 applies maximal-fiber normalization to that exact coordinate on the auxiliary modulus, and the $K/K2/K3$ scalar package then gives a coefficient-free homogeneous scalar ideal; after the explicit homogeneous lift-back the total matrix depth loss remains a controlled fixed fraction of N . The captured-set trace/escape package's projection lower bound and Lemma 5.8 give the first controlled seed independently of the profile gate. The $K/K2/K3$ scalar package supplies the homogeneous coefficient-free ideal, Proposition 5.5 turns that ideal into a restartable thick matrix line, Proposition 5.7 fills the traceless layer after the controlled depth loss, and Theorem 5.11 absorbs a deep principal congruence subgroup. Since every bounded power of H is contained in a bounded power of A in the failed-growth branch, the displayed conclusion follows. $\square$

5.5 Worked example: scalar multiplication from an exact torus

Take

$$z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \xi = E_{12}, \quad g = \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix}.$$

Then

$$\text{Ad}(g)\xi = t^2\xi, \quad \text{Ad}(g^{-1})\xi = t^{-2}\xi,$$

whereas the ordinary matrix commutator satisfies

$$[g, \xi] = (t - t^{-1})\xi.$$

Therefore

$$\text{Ad}(g)\xi - \text{Ad}(g^{-1})\xi = (t + t^{-1})(t - t^{-1})\xi = (\text{tr } g)[g, \xi].$$

The scalar coefficient is obtained from a *large exact centralizer element*; no near-identity torus element is inserted. This is why repeated scalar multiplication charges the actual scalar valuation

rather than an extra congruence depth at every step. The general centralizer-coordinate argument is a p -adic version of this diagonal calculation.

Chapter 6

Prime-Power L^2 Flattening

Failed flattening produces a captured approximate subgroup. Chapters 3–5 show that such a set either grows or absorbs a deep congruence subgroup, while the random walk supplies the required nonconcentration. The argument is split into asymptotic conductor and bounded conductor so that no finite-depth range is hidden in the phrase “for large N ”.

6.1 Scope and dependency contract

Let

$$R_N = \mathbf{Z}/p^N\mathbf{Z}, \quad G_N = \mathrm{SL}_2(R_N), \quad M_N = |G_N|.$$

Throughout this module p is odd and outside the finite exceptional set already isolated in the arithmetic part of the series. Our purpose is to prove the large- p prime-power flattening package needed for the spectral argument.

We use three earlier inputs proved in this series.

- (I1) The finite-group noncommutative Balog–Szemerédi–Gowers inverse theorem for a probability measure, proved in the earlier flattening volume.
- (I2) The logarithmic-time arithmetic escape package proved in Volumes VII–IX and transported to the prime-power setting through XI.1–XI.3: atom decay and congruence-polynomial escape. The captured-set trace/escape package converts these into separated traces, a controlled-nonscalar commuting slice, and controlled adjoint spanning; no full inverse-image subgroup escape estimate is assumed.
- (I3) The XI.3–XI.5 chain (Modules K, K2, K3, P, P^+ and L^+): the captured-set trace/escape package supplies the captured exact commuting slice and controlled adjoint data, Theorem 3.22 supplies the maximal-fiber all-scale scalar normalization with an explicit homogeneous lift-back, and the K/K2/K3 scalar package plus L^+ show that an admissible medium approximate set either grows by a fixed power or a bounded product absorbs a principal congruence subgroup K_m with $m \leq c_{\mathrm{abs}}N$. Here $c_{\mathrm{abs}} > 0$ may be chosen smaller than any prescribed fixed constant by taking the scalar dense threshold sufficiently close to full density and the auxiliary scalar modulus sufficiently short.

No prime-power flattening theorem is imported.

For a function $f : G_N \rightarrow \mathbf{C}$ we use counting-measure norms

$$\|f\|_2^2 = \sum_{g \in G_N} |f(g)|^2, \quad \|f\|_\infty = \max_g |f(g)|.$$

Thus the uniform probability measure has L^2 norm $M_N^{-1/2}$.

6.2 Congruence sizes

For $1 \leq m \leq N$ write

$$K_m = \ker(G_N \rightarrow \mathrm{SL}_2(\mathbf{Z}/p^m\mathbf{Z})).$$

Lemma 6.1 (Exact size of a principal congruence kernel). *For odd p ,*

$$|G_N| = p^{3(N-1)} |\mathrm{SL}_2(\mathbf{F}_p)| = p^{3N-2}(p^2 - 1), \quad |K_m| = p^{3(N-m)}.$$

Proof. The reduction map $G_{j+1} \rightarrow G_j$ is surjective. Its kernel consists of matrices

$$1 + p^j X \pmod{p^{j+1}}, \quad X \in \mathfrak{sl}_2(\mathbf{F}_p),$$

and therefore has size p^3 . Iteration gives both formulas from $|\mathrm{SL}_2(\mathbf{F}_p)| = p(p^2 - 1)$. $\square$

Corollary 6.2 (Absorption is incompatible with a medium-size upper bound). *Fix $\gamma > 0$. If $m \leq \gamma N/3$ and N is sufficiently large, then*

$$|K_m| > M_N^{1-\gamma/2}.$$

In particular, a set H satisfying $|H| \leq M_N^{1-\gamma/2}$ cannot contain K_m .

Proof. The logarithms base p are

$$\log_p |K_m| = 3N - 3m, \quad \log_p M_N = 3N + O_p(1).$$

If $m \leq \gamma N/3$, then $3N - 3m \geq (3 - \gamma)N$, whereas $(1 - \gamma/2) \log_p M_N = (3 - 3\gamma/2)N + O_p(1)$. The first exponent is larger for fixed $\gamma > 0$ and large N . $\square$

6.3 The inverse set supplied by failed flattening

We record the exact form of the earlier finite-group inverse theorem that is used here.

Remark 6.3 (Earlier probability inverse theorem). For every $\gamma > 0$ there is $C_{\mathrm{inv}} = C_{\mathrm{inv}}(\gamma)$ such that the following holds. Let G be finite of cardinality M and let μ be a symmetric probability measure satisfying

$$\|\mu\|_\infty \leq M^{-\gamma}, \quad \|\mu\|_2 \geq M^{-1/2+\gamma}.$$

If, for some sufficiently small $\varepsilon > 0$,

$$\|\mu * \mu\|_2 > M^{-\varepsilon} \|\mu\|_2,$$

then there are a symmetric $M^{C_{\mathrm{inv}}\varepsilon}$ -approximate subgroup $H \subseteq G$ and $x \in G$ such that

$$\mu(xH) \geq M^{-C_{\mathrm{inv}}\varepsilon}, \tag{6.1}$$

$$M^{\gamma-C_{\mathrm{inv}}\varepsilon} \leq |H| \leq M^{1-\gamma+C_{\mathrm{inv}}\varepsilon}. \tag{6.2}$$

Moreover every fixed product power H^r has size at most

$$|H^r| \leq M^{O_{r,\gamma}(\varepsilon)} |H|.$$

Remark 6.4. The lower bound in (6.2) also follows immediately from (6.1) and the L^∞ bound. Its proof is the Power-scale inverse theorem of Volume II, Chapter 8; the present restatement only fixes the notation used in this volume.

Lemma 6.5 (Mass transfer from a captured coset to the difference set). *If μ is symmetric and $\mu(xH) \geq \eta$, then*

$$(\mu * \mu)(H^{-1}H) \geq \eta^2.$$

More generally, if $\mu = \nu^{*\ell}$ for a symmetric step measure ν , then

$$\nu^{*2\ell}(H^{-1}H) \geq \eta^2.$$

Proof. Let X, Y be independent with law μ . On the event $X, Y \in xH$ one has $X^{-1}Y \in H^{-1}H$. Since μ is symmetric, the law of $X^{-1}Y$ is $\mu * \mu$. The event has probability at least η^2 . $\square$

6.4 The prime-power nonconcentration connector

We now isolate exactly what is inherited from the preceding structure theory and what must be proved at the random-walk level.

Definition 6.6 (Admissible inverse set at logarithmic time). Fix constants $\gamma, \eta > 0$. An M_N^η -approximate subgroup $H \subseteq G_N$ is called *XI-admissible* if

- (A1) $M_N^{\gamma/2} \leq |H| \leq M_N^{1-\gamma/2}$;
- (A2) some coset of H carries mass at least $M_N^{-O(\eta)}$ for a walk time $L \asymp \log M_N$, and hence $H^{-1}H$ carries the square of that mass at time $2L$;
- (A3) the controlled-centralizer and adjoint-spanning conclusions of the captured-set trace/escape package apply to a bounded power of H , with total depth defects $O(\eta N)$;
- (A4) the XI.5 theorem therefore gives the dichotomy: either a fixed bounded power grows by a fixed power of $|H|$, or a fixed bounded power contains K_m with $m \leq c_{\text{abs}}N$.

Proposition 6.7 (Logarithmic-time inverse sets are admissible). *Let $S \subset \text{SL}_2(\mathbf{Z})$ be the fixed arithmetic generating set used in Volumes VII–IX, let ν_N be its symmetric reduction to G_N , and choose*

$$c_- \log M_N \leq \ell \leq c_+ \log M_N$$

with fixed positive c_-, c_+ . Suppose H and xH are produced by Theorem 6.3 from $\mu = \nu_N^{*\ell}$, with ε sufficiently small compared with the earlier congruence-polynomial escape exponents. Then H is XI-admissible.

Proof. Property (A1) follows from (6.2) after taking $C_{\text{inv}}\varepsilon < \gamma/2$. Property (A2) is exactly (6.1) together with Lemma 6.5.

For (A3), use the captured-set trace/escape package rather than any assertion about full inverse images of proper mod- p subgroups. The mass lower bound is

$$\nu_N^{*\ell}(xH) \geq M_N^{-C_{\text{inv}}\varepsilon} = p^{-O(\varepsilon N)}.$$

The captured-set trace/escape package shortens this captured coset to the auxiliary times $j \asymp r \log p$ needed at depths $r = C\varepsilon N$ and at the stable scalar depth. The earlier arithmetic owner gives p^{-cr}

bounds for each fixed-complexity congruence-polynomial obstruction used there. Choosing the depth constants first and then ε sufficiently small makes those bad masses strictly smaller than the captured mass. The captured-set trace/escape package therefore produces inside bounded powers of H separated traces, a controlled-nonscalar exact commuting slice, and controlled adjoint spanning; L^+ gives the exact scalar coordinate. Theorem 3.22 then passes inside that exact coordinate set to a maximal concentrated subfiber on an auxiliary modulus, producing the required all-scale scalar profile internally and recording the true homogeneous lift-back depth. The K/K2/K3 scalar package and the XI.5 module therefore give (A4) with no remaining scalar-profile hypothesis.

No estimate of the form M_N^{-c} for the entire inverse image of a positive-density proper subgroup of G_1 is used or asserted. $\square$

Remark 6.8 (What the nonconcentration package provides). Proposition 6.7, together with the point-mass bound proved earlier in the series $\|\nu_N^{*\ell}\|_\infty \leq M_N^{-\gamma}$ at an initial logarithmic time, is the prime-power nonconcentration package required for the flattening contradiction. The content is deliberately formulated through captured mass and congruence-polynomial escape, which is the input actually used by the product-growth argument.

6.5 Failed flattening is impossible

Theorem 6.9 (Prime-power failed-flattening contradiction). *Fix $\gamma > 0$. There is $\varepsilon = \varepsilon(\gamma, S) > 0$ such that, for all sufficiently large N , the following is impossible for $\mu = \nu_N^{*\ell}$ with $c_- \log M_N \leq \ell \leq c_+ \log M_N$:*

$$\|\mu\|_\infty \leq M_N^{-\gamma}, \quad \|\mu\|_2 \geq M_N^{-1/2+\gamma}, \quad \|\mu * \mu\|_2 > M_N^{-\varepsilon} \|\mu\|_2.$$

Proof. Assume failure of flattening. Theorem 6.3 gives an approximate subgroup H with

$$M_N^{\gamma/2} \leq |H| \leq M_N^{1-\gamma/2}$$

after shrinking ε . Proposition 6.7 makes H XI-admissible.

Apply the XI.5 product-growth/absorption theorem. There are two cases.

Growth. For a fixed $\kappa > 0$,

$$|H^3| \geq |H|^{1+\kappa}.$$

Since $|H| \geq M_N^{\gamma/2}$, this implies

$$|H^3| \geq M_N^{\kappa\gamma/2} |H|.$$

But the approximate-group conclusion of Theorem 6.3 gives $|H^3| \leq M_N^{C\varepsilon} |H|$. Taking $C\varepsilon < \kappa\gamma/2$ is a contradiction.

Congruence absorption. A bounded product power of H contains K_m with $m \leq c_{\text{abs}} N$. Replacing H by that fixed bounded power changes the exponent in its size upper bound only by $O(\varepsilon)$, so for small ε it still has size at most $M_N^{1-\gamma/3}$. Choose the scalar dense threshold in XI.5 so that $c_{\text{abs}} < \gamma/6$. Corollary 6.2, with a harmless adjustment of constants, gives $|K_m| > M_N^{1-\gamma/3}$, again a contradiction.

Thus failed flattening cannot occur. $\square$

Corollary 6.10 (One-step prime-power L^2 flattening). *Under the hypotheses*

$$\|\mu\|_\infty \leq M_N^{-\gamma}, \quad \|\mu\|_2 \geq M_N^{-1/2+\gamma}$$

for a logarithmic-time walk measure $\mu = \nu_N^{*\ell}$, one has

$$\boxed{\|\mu * \mu\|_2 \leq M_N^{-\varepsilon} \|\mu\|_2}$$

with fixed $\varepsilon = \varepsilon(\gamma, S) > 0$.

Proof. This is the contrapositive of Theorem 6.9. $\square$

6.6 Bounded-stage flattening to the conductor threshold

Theorem 6.11 (Bounded-stage flattening). *Fix $\gamma > 0$ sufficiently small relative to the representation exponent used in XI.7. There is a bounded integer $J = J(\gamma, S)$ such that, for an initial logarithmic time $\ell_0 \asymp \log M_N$ satisfying the XI.6.001 nonconcentration package,*

$$\mu_j := \nu_N^{*(2^j \ell_0)}$$

satisfies, for some $0 \leq j \leq J$,

$$\|\mu_j\|_2 \leq M_N^{-1/2+\gamma}.$$

All times $2^j \ell_0$ remain $O_{\gamma, S}(\log M_N)$.

Proof. Convolution by a probability measure is a contraction on L^∞ , so the atom bound persists. The backward-owned escape estimates used in Proposition 6.7 are valid throughout any fixed constant multiple of the logarithmic window, hence remain available for the bounded collection of times below.

If $\|\mu_0\|_2 \leq M_N^{-1/2+\gamma}$ there is nothing to prove. Otherwise Corollary 6.10 gives

$$\|\mu_1\|_2 \leq M_N^{-\varepsilon} \|\mu_0\|_2.$$

As long as the target threshold has not been reached, we may iterate and obtain

$$\|\mu_j\|_2 \leq M_N^{-j\varepsilon} \|\mu_0\|_2 \leq M_N^{-j\varepsilon}.$$

Choose

$$J > \frac{1/2 - \gamma}{\varepsilon}.$$

Then failure to reach the target by stage J would imply $\|\mu_J\|_2 < M_N^{-1/2+\gamma}$, a contradiction. Since J is fixed, $2^J \ell_0 = O_{\gamma, S}(\log M_N)$. For each fixed prime this is $O_{\gamma, S, p}(N)$, but the logarithmic group-size formulation is the uniform statement when p varies. $\square$

6.7 Bounded-conductor inputs

For $e \geq 1$ put

$$G_e = \mathrm{SL}_2(\mathbf{Z}/p^e \mathbf{Z}), \quad \rho_{e, r} : G_e \rightarrow G_r.$$

Throughout p is sufficiently large and outside the finite exceptional set already isolated in Volumes VII–IX. We use four earlier inputs proved in this series.

- (I1) **Prime-field spectral gap and mixing.** For the fixed arithmetic generating set S , the reduction walk on G_1 has a uniform spectral gap. Consequently for every fixed $C_0 > 0$, after $C_0 \log p$ steps (with C_0 sufficiently large) its point masses are $O(p^{-3})$.
- (I2) **Prime-field quasirandom covering.** There is an absolute $\delta_q > 0$ such that any $A \subseteq G_1$ with $|A| \geq p^{3-\delta_q}$ satisfies $A^3 = G_1$ for all sufficiently large p .

- (I3) **Frattni lifting from XI.2.001.** For $p \geq 5$, every subgroup of G_r whose reduction modulo p is all of G_1 equals G_r .
- (I4) **The probability inverse theorem.** Failure of one-step L^2 flattening at a logarithmic time produces a symmetric $M^{C\varepsilon}$ -approximate subgroup H and a coset xH of mass at least $M^{-C\varepsilon}$, with $|H| \leq M^{1-\gamma+C\varepsilon}$ and fixed-power bounds $|H^j| \leq M^{O_{j,\gamma}(\varepsilon)}|H|$.

No bounded-conductor expansion theorem is imported.

6.8 A bounded additive basis from one adjoint orbit

Write an element of $V = \mathfrak{sl}_2(\mathbf{F}_p)$ as

$$X = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}, \quad q(X) := -\det X = a^2 + bc.$$

The quadratic form q is nondegenerate for odd p and is invariant under the adjoint action of G_1 .

Lemma 6.12 (Elementary Fourier bound for a nonzero quadratic level). *Fix $c \in \mathbf{F}_p^\times$ and put*

$$Q_c = \{X \in V : q(X) = c\}.$$

For all sufficiently large odd p ,

$$|Q_c| \geq \frac{1}{2}p^2,$$

and for every nonzero additive character frequency $\xi \in V^*$,

$$\left| \sum_{X \in Q_c} e_p(\xi(X)) \right| \leq p^{3/2}.$$

Consequently

$$5Q_c = V.$$

Proof. Use the identity

$$1_{q(X)=c} = \frac{1}{p} \sum_{t \in \mathbf{F}_p} e_p(t(q(X) - c)).$$

For $t \neq 0$, completing squares in the three coordinates reduces

$$\sum_{X \in V} e_p(tq(X) + \xi(X))$$

to a product of three one-dimensional quadratic Gauss sums, hence its absolute value is $p^{3/2}$. For $\xi \neq 0$ the $t = 0$ term vanishes. Summing trivially over $t \neq 0$ and dividing by p gives the displayed Fourier bound. The same computation with $\xi = 0$ gives

$$|Q_c| = p^2 + O(p^{3/2}),$$

so $|Q_c| \geq p^2/2$ for large p .

Let $\widehat{1_{Q_c}}$ denote the unnormalized Fourier transform. The number $R_5(Y)$ of representations $Y = X_1 + \cdots + X_5$ with $X_i \in Q_c$ is

$$R_5(Y) = p^{-3} \sum_{\xi \in V^*} \widehat{1_{Q_c}}(\xi)^5 e_p(-\xi(Y)).$$

The zero frequency contributes at least $2^{-5}p^7$. By Parseval,

$$\sum_{\xi} |\widehat{1_{Q_c}}(\xi)|^2 = p^3 |Q_c| \leq 2p^5$$

for large p . Thus the total nonzero-frequency contribution has absolute value at most

$$p^{-3}(p^{3/2})^3 \sum_{\xi} |\widehat{1_{Q_c}}(\xi)|^2 \leq 2p^{13/2}.$$

Since $p^7/32 > 2p^{13/2}$ for sufficiently large p , one has $R_5(Y) > 0$ for every $Y \in V$. $\square$

Lemma 6.13 (Adjoint-orbit bounded additive generation). *For every nonzero $X \in V$ and every sufficiently large p ,*

$$\underbrace{\text{Ad}(G_1)X + \cdots + \text{Ad}(G_1)X}_{10 \text{ summands}} = V.$$

Proof. If $q(X) \neq 0$, the G_1 -orbit of X is exactly $Q_{q(X)}$. Indeed matrices in that level have the same regular characteristic polynomial; a $\text{GL}_2(\mathbf{F}_p)$ conjugator can be adjusted to determinant 1 by multiplying by an element of the centralizer, whose determinant map is surjective in both the split and nonsplit quadratic-algebra cases. Lemma 6.12 then gives the claim with five summands.

Suppose $q(X) = 0$. After an SL_2 conjugation one may write $X = aE$ for some $a \neq 0$ within the relevant nilpotent orbit, where

$$E = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

For the Weyl element $w = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ one has $\text{Ad}(w)(aE) = -aF$. Hence

$$Y = a(E - F) \in \text{Ad}(G_1)X + \text{Ad}(G_1)X, \quad q(Y) = -a^2 \neq 0.$$

The whole regular orbit $\text{Ad}(G_1)Y$ is therefore contained in two sums of the original orbit. Five copies of that regular orbit fill V by the first case, so ten copies of the original orbit fill V . $\square$

6.9 One congruence layer can be lifted in bounded products

Let

$$L_r = \ker(G_{r+1} \rightarrow G_r) = \{1 + p^r X : X \in V\}.$$

For $r \geq 1$, multiplication in L_r is addition in V because $2r \geq r + 1$.

Lemma 6.14 (No group-theoretic section). *For $r \geq 1$ and $p \geq 5$, the reduction $G_{r+1} \rightarrow G_r$ has no group-homomorphic section.*

Proof. If $s : G_r \rightarrow G_{r+1}$ were a section, its image would be a proper subgroup of G_{r+1} whose reduction modulo p is all of G_1 . This contradicts the Frattini lifting theorem XI.2.001. $\square$

Proposition 6.15 (Bounded-product layer lifting). *Let $A \subseteq G_{r+1}$ be symmetric and contain 1. If for some integer $C \geq 1$*

$$\rho_{r+1,r}(A^C) = G_r,$$

then, for all sufficiently large p ,

$$A^{51C} = G_{r+1}.$$

Proof. Choose a section $s : G_r \rightarrow A^C$ of the reduction map. By Lemma 6.14, not every multiplication defect

$$d(y, z) = s(y)s(z)s(yz)^{-1} \in A^{3C} \cap L_r$$

can be the identity. Fix a nontrivial defect

$$d = 1 + p^r X, \quad X \neq 0 \in V.$$

For each $g \in G_1$, choose $y_g \in G_r$ reducing to g . Then

$$s(y_g)ds(y_g)^{-1} \in A^{5C} \cap L_r$$

corresponds in $L_r \simeq (V, +)$ to $\text{Ad}(g)X$. Lemma 6.13 shows that products of ten such conjugates fill L_r . Thus

$$L_r \subseteq A^{50C}.$$

For arbitrary $u \in G_{r+1}$, choose $s(\rho(u)) \in A^C$. The residual $s(\rho(u))^{-1}u$ lies in $L_r \subseteq A^{50C}$, so $u \in A^{51C}$. $\square$

Corollary 6.16 (Bounded-depth lifting). *Fix $N_0 \geq 2$. There is $C_*(N_0) < \infty$ such that for every $2 \leq e < N_0$, every sufficiently large p , and every symmetric $A \subseteq G_e$ containing 1,*

$$\rho_{e,1}(A) = G_1 \implies A^{C_*(N_0)} = G_e.$$

The same conclusion holds if a fixed bounded power of A rather than A itself surjects modulo p .

Proof. Starting with surjectivity modulo p , apply Proposition 6.15 successively to the projections modulo $p^2, p^3, \dots, p^e$. The exponent is multiplied by at most 51 at each of the at most $N_0 - 1$ layers. $\square$

6.10 Captured inverse sets are already full modulo p

Proposition 6.17 (Prime-field largeness from captured mass). *Fix N_0 and $\gamma > 0$. Let $2 \leq e < N_0$, let $M_e = |G_e|$, and let $\mu = \nu_e^{*\ell}$ at a logarithmic time chosen so that the projected walk on G_1 satisfies*

$$\|(\rho_{e,1})_*\mu\|_\infty \leq Cp^{-3}.$$

Suppose a set $H \subseteq G_e$ and a coset xH satisfy

$$\mu(xH) \geq M_e^{-C_1\varepsilon}.$$

Then

$$|\rho_{e,1}(H)| \geq C^{-1}p^{3-3eC_1\varepsilon+O(1/\log p)}.$$

In particular, after choosing $\varepsilon = \varepsilon(N_0)$ sufficiently small,

$$\rho_{e,1}(H)^3 = G_1$$

by the prime-field quasirandom covering theorem proved earlier in the series.

Proof. Projection can only increase the mass of the image set, hence

$$M_e^{-C_1\varepsilon} \leq \mu(xH) \leq (\rho_{e,1})_*\mu(\rho_{e,1}(xH)) \leq Cp^{-3}|\rho_{e,1}(H)|.$$

Since $M_e = p^{3e+O(1)}$, the displayed lower bound follows. Because $e < N_0$ is bounded, choose ε so that the resulting exponent is at least $3 - \delta_q$. Input (I2) then gives the triple-product conclusion. $\square$

6.11 Bounded-conductor failed flattening

Theorem 6.18 (Uniform failed-flattening contradiction at bounded conductor). *Fix $N_0 \geq 2$ and $\gamma > 0$. There is $\varepsilon = \varepsilon(N_0, \gamma, S) > 0$ such that, for every sufficiently large allowed prime p and every $2 \leq e < N_0$, a logarithmic-time walk measure $\mu = \nu_e^{*\ell}$ satisfying*

$$\|\mu\|_\infty \leq M_e^{-\gamma}, \quad \|\mu\|_2 \geq M_e^{-1/2+\gamma}$$

cannot satisfy

$$\|\mu * \mu\|_2 > M_e^{-\varepsilon} \|\mu\|_2.$$

Consequently the one-step and bounded-stage XI.6 flattening statements hold uniformly for these bounded conductors.

Proof. Assume failed flattening. The probability inverse theorem supplies a symmetric $M_e^{C_1\varepsilon}$ -approximate subgroup H and a captured coset xH with

$$\mu(xH) \geq M_e^{-C_1\varepsilon}, \quad |H| \leq M_e^{1-\gamma+C_1\varepsilon},$$

and with

$$|H^j| \leq M_e^{C_j\varepsilon} |H|$$

for every fixed j .

By Proposition 6.17, after shrinking ε depending only on N_0, γ, S ,

$$\rho_{e,1}(H^3) = G_1.$$

Corollary 6.16, applied to $A = H^3$, gives

$$H^{3C_*(N_0)} = G_e.$$

On the other hand the fixed-power estimate from the inverse theorem yields

$$M_e = |G_e| = |H^{3C_*}| \leq M_e^{C'\varepsilon} |H| \leq M_e^{1-\gamma+(C_1+C')\varepsilon}.$$

Choosing $\varepsilon < \gamma/(2(C_1 + C'))$ is a contradiction. The one-step estimate follows by contraposition. The bounded-stage iteration is identical to the asymptotic-depth flattening argument and uses only a bounded number of doublings, so all times remain $O_{N_0, \gamma, S}(\log M_e)$. $\square$

Remark 6.19 (Bounded-conductor patch and the endgame). Combine Theorem 6.18 for $2 \leq e < N_0$ with the asymptotic-depth flattening argument for $e \geq N_0$ and the prime-field theorem for $e = 1$. Then the XI.6 flattening input used by the representation-theoretic endgame is available at every exact conductor $e \geq 1$, uniformly for all sufficiently large allowed primes. Hence the representation-theoretic endgame is available at every exact conductor.

6.12 Worked example: the size contradiction behind failed flattening

Ignore lower-order factors and write $|G_N| \asymp p^{3N}$. Suppose a failed flattening inverse theorem produces an approximate subgroup H with

$$|H| \leq |G_N|^{1-\gamma/2}.$$

If the product-growth theorem instead forces a bounded power of H to contain K_m with $m \leq \gamma N/3$, then

$$|K_m| = p^{3(N-m)} \geq p^{(3-\gamma)N},$$

while

$$|G_N|^{1-\gamma/2} \asymp p^{(3-3\gamma/2)N}.$$

Since $3 - \gamma > 3 - 3\gamma/2$, the absorbed congruence subgroup is already larger than the allowed medium-size inverse set. Small tripling cannot hide this discrepancy: fixed product powers change the exponent only by the small inverse-theorem loss. This numerical comparison is the final contradiction in the absorption branch of one-step flattening.

Chapter 7

Exact Conductors, Quasirandomness, and Prime-Power Expansion

The final step is representation-theoretic. Exact conductor forces a primitive character on a middle-depth abelian congruence layer. The adjoint orbit of that character supplies a conductor-sensitive dimension lower bound; Plancherel and the flattened walk then rule out eigenvalues near 1. Conductor one is inherited from the prime-field volume, bounded conductors from Chapter 6, and asymptotic conductors from the multiscale flattening argument.

7.1 Purpose

Let

$$R_N = \mathbf{Z}/p^N\mathbf{Z}, \quad G_N = \mathrm{SL}_2(R_N), \quad M_N = |G_N|,$$

with $p \geq 5$. This module proves the representation-theoretic endpoint required by the final prime-power spectral statements from the congruence filtration and the XI.6 flattening theorem.

The main point is that no classification of irreducible representations of G_N is needed. An exact-conductor representation is forced to see a primitive character of a large abelian congruence subgroup, and the adjoint orbit of that character already has cardinality $\gg p^{N/2}$.

7.2 Exact conductor and an abelian middle-depth kernel

For $1 \leq r \leq N$ write

$$K_r = \ker(G_N \rightarrow G_r).$$

Definition 7.1 (Exact conductor). An irreducible complex representation π of G_N has exact conductor e if e is the least integer such that π factors through G_e . Thus $1 \leq e \leq N$ for every nontrivial irreducible representation.

Lemma 7.2 (Middle-depth additive model). *Put*

$$r = \lceil N/2 \rceil, \quad s = N - r = \lfloor N/2 \rfloor.$$

Then K_r is abelian and the map

$$\mathfrak{sl}_2(R_s) \longrightarrow K_r, \quad X \longmapsto 1 + p^r X$$

is an isomorphism of additive groups.

Proof. For $X, Y \in M_2(R_s)$,

$$(1 + p^r X)(1 + p^r Y) \equiv 1 + p^r(X + Y) \pmod{p^N}$$

because $2r \geq N$. Likewise

$$\det(1 + p^r X) \equiv 1 + p^r \operatorname{tr} X \pmod{p^N},$$

so determinant one is exactly the condition $\operatorname{tr} X \equiv 0 \pmod{p^s}$. This gives the claimed additive isomorphism and commutativity. $\square$

Fix the primitive additive character

$$e_s(u) = \exp(2\pi i u / p^s).$$

For $A \in \mathfrak{sl}_2(R_s)$ define

$$\chi_A(1 + p^r X) = e_s(\operatorname{tr}(AX)).$$

Lemma 7.3 (Characters and conjugation). *For p odd, $A \mapsto \chi_A$ identifies $\mathfrak{sl}_2(R_s)$ with the full character group of K_r . Moreover*

$$\chi_A^g = \chi_{gAg^{-1}},$$

where the right side depends only on the reduction of g modulo p^s .

Proof. The trace pairing on $\mathfrak{sl}_2(R_s)$ is perfect for odd p : in the basis $E_{12}, E_{21}, H = \operatorname{diag}(1, -1)$ its Gram matrix has unit determinant up to the factor 2. Hence every additive character is uniquely of the displayed form. Conjugation follows from cyclicity of trace:

$$\operatorname{tr}(AgXg^{-1}) = \operatorname{tr}(g^{-1}AgX).$$

The congruence level p^s is the only level visible in the character formula. $\square$

Lemma 7.4 (Exact conductor produces a primitive middle-depth weight). *Let π be irreducible of exact conductor N . In the decomposition of $\pi|_{K_r}$ into characters, at least one occurring character χ_A has*

$$A \not\equiv 0 \pmod{p}.$$

Proof. Exact conductor N means that K_{N-1} does not act trivially. Under Lemma 7.2, the subgroup K_{N-1} corresponds to matrices $p^{s-1}X$ with $X \in \mathfrak{sl}_2(\mathbf{F}_p)$. If every occurring A were divisible by p , then

$$\chi_A(1 + p^{N-1}X) = e_s(p^{s-1} \operatorname{tr}(AX)) = 1$$

for all X , so K_{N-1} would act trivially on π , a contradiction. $\square$

7.3 A conductor-sensitive dimension bound

Lemma 7.5 (Centralizer bound for a primitive traceless matrix). *Let $A \in \mathfrak{sl}_2(R_s)$ with $A \not\equiv 0 \pmod{p}$. Then*

$$|C_{G_s}(A)| \leq p^{2s}.$$

Consequently

$$|\operatorname{Ad}(G_s)A| = [G_s : C_{G_s}(A)] \geq (1 - p^{-2})p^s.$$

Proof. Because $A \bmod p$ is a nonzero traceless 2×2 matrix, it is not scalar for odd p . Write

$$A = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}, \quad X = \begin{pmatrix} x & y \\ z & w \end{pmatrix}.$$

At least one of a, b, c is a unit. If b is a unit, the commutation equations give $X = \alpha I + \beta A$ with $\beta = y/b$ and $\alpha = x - \beta a$; the case of unit c is symmetric. If both b, c are nonunits, then a is a unit, and the equations give $\beta = (x - w)/(2a)$, $y = \beta b$, $z = \beta c$, and $\alpha = (x + w)/2$. Thus in every case the full matrix centralizer is exactly $R_s I + R_s A$, which has p^{2s} elements. Intersecting with G_s can only decrease cardinality.

Also

$$|G_s| = p^{3s-2}(p^2 - 1) = (1 - p^{-2})p^{3s}.$$

Dividing gives the orbit bound. $\square$

Theorem 7.6 (Conductor-sensitive quasirandomness). *If π is an irreducible representation of G_N of exact conductor N , then*

$$\dim \pi \geq (1 - p^{-2})p^{\lfloor N/2 \rfloor}.$$

Equivalently, for all sufficiently large N ,

$$\dim \pi \geq M_N^{1/7},$$

uniformly for $p \geq 5$.

Proof. Choose a primitive character χ_A occurring in $\pi|_{K_r}$ by Lemma 7.4. Since K_r is normal, every conjugate character in the G_N -orbit of χ_A also occurs in the restriction of π . Distinct characters have disjoint weight spaces, each nonzero. Hence

$$\dim \pi \geq |G_N \cdot \chi_A|.$$

By Lemma 7.3, this orbit is the G_s -adjoint orbit of A . Lemma 7.5 gives the first bound.

Since $M_N = p^{3N+O(1)}$ and $\lfloor N/2 \rfloor \geq N/2 - 1$, the exponent $1/7$ is a convenient uniform weakening of the asymptotic exponent $1/6$. $\square$

Remark 7.7 (Exact-conductor dependency). Exact-conductor structure is supplied by Definition 2.1 and Lemma 7.4; the dimension/multiplicity lower bound is Theorem 7.6. No character table or classification theorem is used.

7.4 Trace–multiplicity inequality

Let G be finite of cardinality M , and for a probability measure μ write

$$\hat{\mu}(\rho) = \sum_{g \in G} \mu(g) \rho(g)$$

for an irreducible unitary representation ρ .

Lemma 7.8 (Finite-group Plancherel). *For every $f : G \rightarrow \mathbf{C}$,*

$$\sum_{g \in G} |f(g)|^2 = \frac{1}{M} \sum_{\rho \in \widehat{G}} d_\rho \|\hat{f}(\rho)\|_{\text{HS}}^2.$$

Proof. This is the matrix-coefficient orthogonality relation applied to the decomposition of the regular representation. It is included here to fix normalization. $\square$

Proposition 7.9 (Trace–multiplicity inequality). *Let ν be a symmetric probability measure on G . Suppose ρ is irreducible of dimension d and λ is an eigenvalue of the self-adjoint operator $\widehat{\nu}(\rho)$. Then for every integer $L \geq 1$,*

$$\|\nu^{*L}\|_2^2 \geq \frac{d}{M} |\lambda|^{2L}.$$

Consequently, if

$$\|\nu^{*L}\|_2 \leq M^{-1/2+\gamma},$$

then

$$|\lambda|^{2L} \leq \frac{M^{2\gamma}}{d}.$$

Proof. Apply Lemma 7.8 to ν^{*L} . The ρ -summand is

$$\frac{d}{M} \|\widehat{\nu}(\rho)^L\|_{\text{HS}}^2,$$

which is at least $d|\lambda|^{2L}/M$. $\square$

This proposition supplies the multiplicity bound used in the spectral endpoint.

7.5 From flattening to a uniform spectral gap

Let $S \subset \text{SL}_2(\mathbf{Z})$ be the fixed symmetric arithmetic generating set and let ν_N be the uniform step measure on its reduction in G_N .

Theorem 7.10 (Spectral gap on exact-conductor representations). *Assume the XI.6 bounded-stage flattening theorem with some $0 < \gamma < 1/14$. Then there is $\kappa > 0$, independent of N and of every sufficiently large allowed prime p , such that for every irreducible representation π of exact conductor N ,*

$$\|\widehat{\nu}_N(\pi)\|_{\text{op}} \leq 1 - \kappa.$$

Proof. We separate the endpoint into three inputs and one numerical conversion.

Step 1: flattening at logarithmic time. XI.6 gives a time L satisfying

$$L \leq C \log M_N, \quad \|\nu_N^{*L}\|_2 \leq M_N^{-1/2+\gamma},$$

where C depends only on the fixed generating set and the previously fixed exponent hierarchy. In particular, neither C nor γ depends on the exact-conductor representation chosen below.

Step 2: exact-conductor multiplicity. Let λ be an eigenvalue of the self-adjoint matrix $\widehat{\nu}_N(\pi)$. Theorem 7.6 gives, in the asymptotic conductor range,

$$d_\pi \geq M_N^{1/7}.$$

Applying Proposition 7.9 with the preceding L^2 estimate therefore gives

$$|\lambda|^{2L} \leq \frac{M_N^{2\gamma}}{d_\pi} \leq M_N^{2\gamma-1/7}.$$

Thus the quasirandomness gain $M_N^{-1/7}$ beats the flattening loss $M_N^{2\gamma}$ as soon as $2\gamma < 1/7$.

Step 3: convert a power saving into a fixed gap. Fix $\gamma < 1/14$ and write $a = 1/7 - 2\gamma > 0$. Taking logarithms in the last display gives

$$\log |\lambda| \leq -\frac{a \log M_N}{2L}.$$

Because $L \leq C \log M_N$, one has $(\log M_N)/L \geq 1/C$, and hence

$$|\lambda| \leq \exp\left(-\frac{a}{2C}\right) = \exp\left(-\frac{1/7 - 2\gamma}{2C}\right) =: 1 - \kappa < 1.$$

The symmetry of the step measure makes $\widehat{\nu}_N(\pi)$ self-adjoint, so bounding $|\lambda|$ controls both the $+1$ and -1 spectral endpoints. $\square$

Theorem 7.11 (Uniform prime-power expansion from exact conductors). *Assume the prime-field spectral gap already proved in Volume IX and the XI.6 theorem uniformly for all sufficiently large allowed p . Then there exists $\kappa_* > 0$ such that for every such p , every $N \geq 1$, and every nontrivial irreducible representation π of G_N ,*

$$\|\widehat{\nu}_N(\pi)\|_{\text{op}} \leq 1 - \kappa_*.$$

Thus the Cayley graphs

$$\text{Cay}(G_N, S \bmod p^N)$$

form a uniform expander family in N and in the allowed primes.

Proof. Let e be the exact conductor of π . If $e = 1$, the operator is exactly the corresponding prime-field operator and is controlled by Volume IX. If $e \geq 2$, regard π as a new representation of G_e . The operator defined by S is unchanged by inflation from G_e to G_N . For conductors e in the asymptotic range of XI.6, Theorem 7.10 at modulus p^e gives the fixed gap.

The finite-depth range $2 \leq e < N_0$ is supplied by the bounded-conductor lifting argument. There, prime-field mixing makes the captured inverse set almost full modulo p , and a bounded-product congruence-layer lifting theorem upgrades this to the full group G_e in a number of steps depending only on N_0 . Thus the asymptotic-depth flattening argument covers the asymptotic conductor range, the bounded-conductor lifting argument covers the bounded conductor range, and Volume IX covers $e = 1$. $\square$

Remark 7.12 (Historical source boundary: fixed prime versus uniform large primes). The original prime-power theorem of Bourgain–Gamburd fixes a sufficiently large prime p and proves expansion as the exponent N tends to infinity [1]. Thus the additional assertion in Theorem 7.11 that the same gap may be chosen for every sufficiently large allowed p is *not* being attributed to that theorem. In this manuscript it is an extra internal conclusion: Volume IX supplies the prime-field gap uniformly in the allowed large-prime range, and Lemma 7.13 below checks that every prime-power loss appearing in XI.3–XI.7 can be absorbed after one choice of p_0 . This distinction is important when comparing theorem strength with the historical source.

7.6 Uniformity in the large-prime range

Uniformity in p requires eliminating the auxiliary residue-field coefficients before matrix realization. K3 performs that elimination. On the scalar side, the remaining explicit regularization loss is

$$L_p^N, \quad L_p = 2(1 + \lceil \log_2 p \rceil).$$

Lemma 7.13 (Uniform-large-prime parameter patch). *Fix the finite exponent hierarchy required by XI.3–XI.7. There exists p_0 such that, for every prime $p \geq p_0$, the scalar regularization/Fourier-completion losses, scalar-growth exponents, Fourier-completion exponent, $L+$ depth offsets, XI.6 flattening exponent and Theorem 7.10 may all be chosen uniformly in p .*

Proof. The order of choices matters. We first freeze the finite hierarchy of strict exponent inequalities used in XI.3–XI.7, leaving a positive margin in each one. Only after those margins have been fixed do we choose the lower prime threshold.

Regularization loss. For every prescribed $\delta > 0$,

$$\log_p L_p \longrightarrow 0 \quad (p \rightarrow \infty).$$

Hence there is $p_1(\delta)$ such that

$$L_p^N \leq p^{\delta N} \quad (p \geq p_1(\delta), N \geq 1).$$

We choose δ smaller than the least unused margin in the frozen exponent hierarchy. The entire regularization loss can then be charged to that margin.

Coefficient independence. Module K3 has already replaced the K/K2/K3 output by coefficient-free bounded ring expressions. Consequently, the length of the matrix words used later is bounded independently of the numerical residue coefficient; there is no hidden factor growing with p at this interface.

Remaining local estimates. The one-digit collision error contains $(p-1)^{-1}$ and therefore improves as p grows. The Schnirelmann step is dimension-only. Exact-conductor Fourier completion is controlled by a geometric series with a fixed exponent gap. The controlled-centralizer argument in $L+$ uses only that 2 is a unit: a discriminant of valuation $O(\varepsilon N)$ produces the Smith/operator loss already budgeted in the depth hierarchy.

There are now only finitely many residual inequalities, each with a strict positive margin fixed before p was chosen. Enlarge the prime threshold once, from p_1 to a single p_0 , so that every one of those inequalities holds simultaneously. All constants in XI.3–XI.7, and hence the gap furnished by Theorem 7.10, are then uniform for $p \geq p_0$. $\square$

Corollary 7.14 (Rank-one prime-power super-approximation). *For the fixed non-elementary arithmetic subgroup and generating set used throughout the series, there is a finite set of excluded primes such that the reductions modulo p^N form a uniform expander family for every prime outside that set and every $N \geq 1$.*

Proof. Strong approximation from the earlier arithmetic volume identifies the allowed reductions with G_N outside finitely many primes. Lemma 7.13 and Theorem 7.11 give a uniform spectral gap on all nontrivial representations, which is the expander/super-approximation conclusion. $\square$

7.7 Worked example: a middle-depth exact-conductor character

Take $N = 6$, so $r = s = 3$ and

$$K_3 \cong \mathfrak{sl}_2(\mathbf{Z}/p^3\mathbf{Z})$$

as an additive group. Let

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

and define

$$\chi_A(1 + p^3 X) = \exp\left(\frac{2\pi i}{p^3} \operatorname{tr}(AX)\right).$$

Because A is primitive modulo p , this character does not factor through a shallower quotient of the additive layer. Conjugating the representation by $g \in G_6$ sends the weight A to $\operatorname{Ad}(g)A$. Hence a representation containing χ_A contains the whole adjoint orbit of this primitive weight. Distinct orbit characters are linearly independent weight spaces, so orbit size immediately forces a representation-dimension lower bound. The general proof uses exactly this mechanism, with a centralizer bound controlling the orbit size at arbitrary exact conductor.

Historical references

- [1] J. Bourgain and A. Gamburd, *Expansion and random walks in $SL_d(\mathbf{Z}/p^n\mathbf{Z})$, I*, J. Eur. Math. Soc. 10 (2008), 987–1011.
- [2] J. Bourgain and A. Gamburd, *Expansion and random walks in $SL_d(\mathbf{Z}/p^n\mathbf{Z})$, II*, J. Eur. Math. Soc. 11 (2009), 1057–1103.
- [3] A. Salehi Golsefidy, *Sum-product phenomena: p -adic case*, J. Analyse Math. 139 (2019), 703–746.
- [4] J. Tang and X. Zhang, *Sum-product phenomenon in quotients of rings of algebraic integers*, arXiv:2308.08867.

The Bourgain–Gamburd Method
Volume XII: Arbitrary Moduli and Rank-One Super-Approximation

17 August 2026

Scope and historical note

This volume proves the arbitrary-modulus rank-one super-approximation theorem by strong induction on exact conductor. The squarefree and prime-power families are supplied by Volumes X and XI. The new work is to control cross-prime correlation without paying once per divisor, to isolate and propagate deep prime-power directions simultaneously, and to combine this structure with the random-matrix/Fourier rigidity mechanism needed for conductor-uniform decay. The reading order follows the proved exact-conductor induction. Historical references identify the original arbitrary-modulus and random-matrix inputs; alternate development routes are omitted from the exposition.

Proof architecture

The proof is a strong induction on proper exact conductors, measured by the integer-valued rank

$$\Omega(q) = \sum_{p^n \parallel q} n.$$

This *rank* is used only to make the recursion well founded. Quantitative size estimates are instead charged to the logarithmic arithmetic conductor weight

$$W_{\text{ar}}(q) = \log q = \sum_{p^n \parallel q} n \log p,$$

or to the uniformly comparable group weight introduced in Chapter 1. Thus no estimate treats the unweighted integer $\Omega(q)$ as a proxy for $\log q$. This separation matches the logarithmically weighted conductor bookkeeping in the arbitrary-modulus argument.

The prime-power family is the proved Volume-XI base. For composite q , one first controls cross-prime correlations, then proves that any medium approximate subgroup captured by the walk has a deficient proper projection. Mixing on that proper divisor makes the projection almost full, a contradiction. This supplies global flattening. Finally, exact-conductor representation theory separates proper-conductor eigenvalues (already controlled by induction) from the new full-conductor part, which is handled by quasirandom multiplicity and the trace identity.

No factor-by-factor destruction loop is used, and no exponent is allowed to deteriorate with $\omega(q)$ or the number of divisors.

Reader’s map. Separate the proof into three layers. First comes the *structural layer*: control cross-prime correlation. Second comes the *inductive layer*: a medium approximate subgroup must have a deficient proper projection, and the induction hypothesis mixes on that proper conductor. Third comes the *spectral layer*: exact-conductor representations are the only genuinely new eigenmodes. Keeping these layers separate is the easiest way to see where a divisor-count loss could accidentally enter.

Series position. This volume uses Volumes X–XI. Its main outputs are no-divisor-loss recombination and arbitrary-modulus rank-one super-approximation.

Quantifier and asymptotic conventions. Unless a statement explicitly says otherwise, every constant occurring in a quantitative theorem is chosen in the order in which it is quantified in that theorem. Thus a phrase such as “for all sufficiently small δ ” means: there is a displayed or locally quantified threshold $\delta_0 > 0$, depending only on the parameters listed before that phrase, such that the assertion holds for every $0 < \delta < \delta_0$. The phrases “sufficiently large N ” and “sufficiently large p ” have the analogous meaning.

If $X \ll_{\theta} Y$ or $X = O_{\theta}(Y)$, the implicit constant depends only on θ and on the fixed ambient data of the current volume; it is independent of the variables quantified after θ . The words

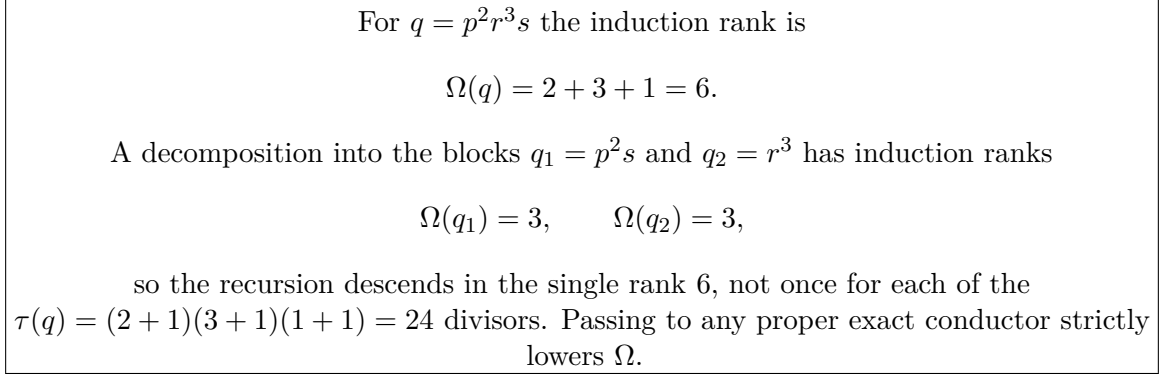


Figure 1: The distinction between the well-founded induction rank Ω and the quantitative logarithmic conductor weights. The recursion does not pay once per divisor, while all size estimates remain logarithmically weighted.

absolute, *uniform*, and *bounded* never introduce an unlisted dependence. Whenever an inequality between exponents or scales is essential to the argument, that inequality is written explicitly in the proof; these conventions are not a substitute for such a calculation.

Contents

Scope and historical note	i
Proof architecture	ii
1 Arithmetic Decomposition and Exact Conductors	1
1.1 Why exact conductors are the right coordinates	1
1.2 CRT decomposition	1
1.3 Irreducibles of a direct product	1
1.4 Exact local and global conductors	2
1.5 Weight comparison	2
1.6 Worked example: conductor bookkeeping for $q = p^2 r^3$	3
2 No-Divisor-Loss Recombination	4
2.1 The induction contract	4
2.2 Exact correlation bookkeeping	4
2.3 Why the conductor-weight bookkeeping must be global	4
2.4 Exact conductor weights and correlation defect	5
2.5 The no-divisor-loss aggregation identity	5
2.6 Weighted exponent-band selection	7
2.7 Local state decomposition	7
2.8 No-divisor-loss iteration	8
2.9 The corrected obstruction statement	9
2.10 Compatibility of squarefree correlations with prime-power depth	9
2.11 Scope	9
2.12 Perfectness and simple quotients of a prime-power factor	9
2.13 Exact Goursat compatibility	11
2.14 Squarefree approximate Goursat and the prime-power compatibility gap	11
2.15 A source-faithful deep-band seed from captured mass	15
2.16 Simultaneous lifting on a bounded-depth block	18
2.17 Deep-band Lie propagation: the elementary part of the large-exponent argument	22
2.17.1 From torus Fourier rigidity to the congruence Fourier bound	27
2.18 Captured mass makes every proper projection almost full	116
2.19 Proper-divisor induction and the global obstruction	118
2.20 Purpose	118
2.21 Well-founded divisor induction	119
2.22 A two-block partition of conductor weight	119
2.23 Block correlation is dominated by prime correlation	120

2.24	The structural two-block obstruction	120
2.25	The global captured obstruction is impossible	121
2.26	Worked example: why a 50/50 split is not a terminal dichotomy	122
2.27	Worked example: a dominant prime-power factor	122
3	Global Nonconcentration and L2-Flattening	123
3.1	How the inverse theorem enters	123
3.2	Normalization and earlier inputs	123
3.3	The exact inverse-theorem interface	123
3.4	One-step arbitrary-modulus flattening	124
3.5	Initial power saving	125
3.6	Bounded-stage global flattening	125
3.7	Worked example: exponent iteration	125
4	Exact-Conductor Spectral Endgame	126
4.1	Closing the strong induction	126
4.2	Local dimension input and tensorization	126
4.3	Trace and exact conductor	127
4.4	Well-founded induction on exact conductor, not a sum over divisors	127
4.5	Super-approximation and consequences	128
4.6	Worked example: exact conductor p^2r	129

Chapter 1

Arithmetic Decomposition and Exact Conductors

1.1 Why exact conductors are the right coordinates

The arbitrary-modulus problem should not be organized by the raw number of prime divisors. Its natural size coordinate is the logarithm of the exact conductor. This chapter establishes the CRT, representation factorization, and conductor weights used by every later induction.

1.2 CRT decomposition

Let $q = \prod_p p^{n_p}$ and $G(q) = \mathrm{SL}_2(\mathbf{Z}/q\mathbf{Z})$.

Proposition 1.1 (Chinese remainder decomposition). *Reduction in each prime-power coordinate gives a canonical isomorphism*

$$G(q) \cong \prod_{p|q} G(p^{n_p}), \quad G(p^n) = \mathrm{SL}_2(\mathbf{Z}/p^n\mathbf{Z}).$$

The same statement holds for every divisor $d \mid q$, compatibly with the maps $G(q) \rightarrow G(d)$.

Proof. The ring isomorphism $\mathbf{Z}/q\mathbf{Z} \cong \prod_{p|q} \mathbf{Z}/p^{n_p}\mathbf{Z}$ is entrywise compatible with matrix multiplication and determinant. A matrix over the product ring has determinant one exactly when each component does. Compatibility with divisors is functoriality of reduction. $\square$

1.3 Irreducibles of a direct product

Theorem 1.2 (Tensor factorization of irreducible representations). *Let $G = G_1 \times \cdots \times G_k$ be finite. Every irreducible complex representation of G is isomorphic to*

$$\rho_1 \boxtimes \cdots \boxtimes \rho_k$$

with each ρ_i irreducible on G_i . Conversely every such external tensor product is irreducible, and

$$\dim \rho = \prod_i \dim \rho_i.$$

Proof. The group algebra factors as $\mathbf{C}[G] \cong \bigotimes_i \mathbf{C}[G_i]$. By Maschke, each factor is a finite product of full matrix algebras. Tensoring these Wedderburn decompositions gives a product of full matrix algebras whose simple modules are exactly the tensor products of simple modules from the factors. Their dimensions multiply. $\square$

1.4 Exact local and global conductors

For an irreducible ρ_p of $G(p^{n_p})$, define $e_p(\rho)$ to be the least $0 \leq e \leq n_p$ for which ρ_p factors through $G(p^e)$; $e = 0$ means the trivial factor.

Definition 1.3 (Exact conductor). For $\rho = \boxtimes_p \rho_p$, set

$$d(\rho) := \prod_{p|q} p^{e_p(\rho)}.$$

Its arithmetic conductor weight and group conductor weight are

$$W_{\text{ar}}(\rho) = \log d(\rho) = \sum_p e_p(\rho) \log p,$$

$$W_G(\rho) = \sum_{e_p(\rho) > 0} \log |G(p^{e_p(\rho)})|.$$

Proposition 1.4 (Exact conductor factorization). *The representation ρ factors through $G(d)$ if and only if $d(\rho) \mid d$. In particular $d(\rho)$ is the unique minimal divisor of q through which ρ factors. Both conductor weights are additive over the active primes.*

Proof. By Proposition 1.1 and Theorem 1.2, factoring through $G(d)$ is equivalent to factoring in every local coordinate through the corresponding p -power depth. Minimality is therefore coordinatewise. The weight identities are definitions. $\square$

Corollary 1.5 (Active local factors). *If $I(\rho) = \{p : e_p(\rho) > 0\}$, then every matrix coefficient of ρ , and every eigenfunction lying in its isotypic component for a convolution operator, is pulled back from*

$$G(d(\rho)) = \prod_{p \in I(\rho)} G(p^{e_p(\rho)}).$$

Inactive factors play no role.

Proof. The kernel of $G(q) \rightarrow G(d(\rho))$ lies in $\ker \rho$ by minimal-factorization, so all matrix coefficients are constant on its cosets. Convolution preserves the isotypic component. $\square$

1.5 Weight comparison

Lemma 1.6 (Arithmetic and group weights are uniformly comparable). *Outside a fixed finite set of primes there are absolute constants $c, C > 0$ such that for every $e \geq 1$,*

$$c e \log p \leq \log |G(p^e)| \leq C e \log p.$$

Consequently $W_G(\rho) \asymp W_{\text{ar}}(\rho)$ uniformly in the number of active primes.

Proof. For odd p ,

$$|G(p^e)| = p^{3e-2}(p^2 - 1).$$

The logarithm lies between constant multiples of $e \log p$ uniformly for $p \geq 5$. Sum over active primes. $\square$

1.6 Worked example: conductor bookkeeping for $q = p^2 r^3$

Let $q = p^2 r^3$ with distinct allowed primes. If an irreducible representation is nontrivial of exact local conductor p on the first factor and r^3 on the second, then its exact global conductor is $d = pr^3$, not q . The inactive p -depth is discarded before the spectral estimate. Thus the induction parameter is

$$\Omega(d) = 1 + 3 = 4 < 5 = \Omega(q),$$

and the logarithmic conductor weight is the sum of the two active local weights. Here Ω is only the integer-valued induction rank: quantitative gains and losses are measured by W_{ar} or W_G , not by Ω . This elementary example is the model for the exact-conductor induction in Chapter [4](#).

Chapter 2

No-Divisor-Loss Recombination

2.1 The induction contract

First reading versus second reading. First track only three quantities: the local gains, the single global correlation defect, and the weight of the proper divisor produced at the end. This reveals why the proof avoids a loss proportional to the number of prime factors. On a second reading, inspect the random-matrix, projective-avoidance, and simultaneous-lifting modules that make the structural obstruction quantitative. Their constants matter only after the no-divisor-loss bookkeeping has been understood.

Fix an arbitrary allowed modulus $q = \prod p^{n_p}$. We argue by strong induction on

$$\Omega(q) = \sum_{p^{n_p} \parallel q} n_p.$$

The entire prime-power family is the base case supplied by Volume XI. At a composite modulus q , every proper exact conductor or proper CRT subproduct has smaller Ω and is therefore already spectrally controlled. The present chapter first proves exact conductor-weight bookkeeping, then isolates the only cross-prime correlation owner inherited from Volume X, and finally proves the two-block proper-divisor obstruction that replaces the invalid factor-by-factor destruction argument.

2.2 Exact correlation bookkeeping

2.3 Why the conductor-weight bookkeeping must be global

Let

$$q = \prod_{p \in \mathcal{P}} p^{n_p}, \quad G(q) = \mathrm{SL}_2(\mathbf{Z}/q\mathbf{Z}) \cong \prod_{p \in \mathcal{P}} G_p, \quad G_p := \mathrm{SL}_2(\mathbf{Z}/p^{n_p}\mathbf{Z}),$$

where all primes are outside the fixed finite arithmetic exceptional set. A tempting but invalid bookkeeping step would pay a projection error of size $O(\log K)$ separately at each active prime and then multiply that loss by the total conductor weight. If $\log K = O(\delta W)$, this can create an $O(\delta W^2)$ error rather than the required $O(\delta W)$ error. The argument below never makes that multiplication. Instead, all local gains are summed first and a single global correlation defect is subtracted at the end.

We use the squarefree correlation theorem from Volume X and the prime-power local product theorem from Volume XI. The purpose of this section is to prove the exact recombination algebra into which those earlier results fit; no arbitrary-modulus product theorem is imported.

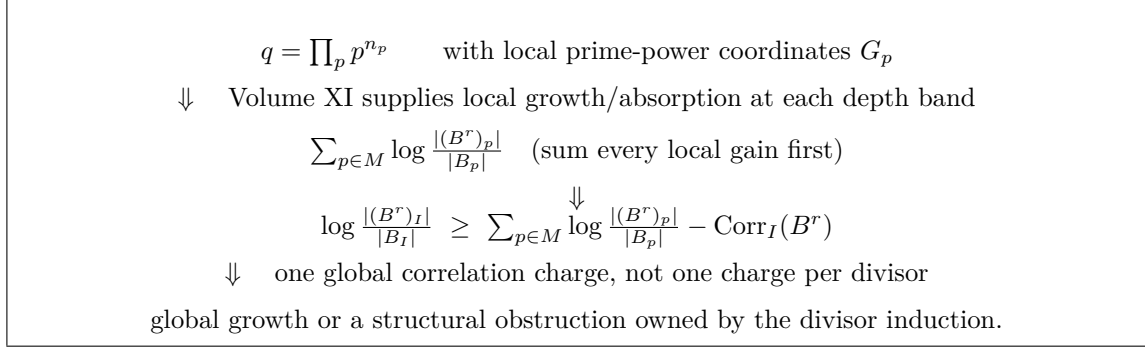


Figure 2.1: No-divisor-loss recombination. The crucial bookkeeping rule is to aggregate local gains before subtracting a single global correlation defect.

2.4 Exact conductor weights and correlation defect

For $I \subseteq \mathcal{P}$ put

$$G_I := \prod_{p \in I} G_p, \quad w_p := \log |G_p|, \quad W(I) := \sum_{p \in I} w_p = \log |G_I|.$$

Thus weight is exactly additive; there is no asymptotic replacement at this stage. For $B \subseteq G(q)$ write B_I for its projection to G_I and B_p for its projection to G_p .

Definition 2.1 (Set-correlation defect). For nonempty $B \subseteq G(q)$ and $I \subseteq \mathcal{P}$, define

$$\text{Corr}_I(B) := \sum_{p \in I} \log |B_p| - \log |B_I|.$$

Lemma 2.2 (Basic properties). *For every nonempty B and I one has $\text{Corr}_I(B) \geq 0$. It is exactly zero when the projected set B_I has cardinality equal to the product of its one-factor projections. Moreover*

$$\log |B_I| = \sum_{p \in I} \log |B_p| - \text{Corr}_I(B).$$

Proof. The projection map $B_I \rightarrow \prod_{p \in I} B_p$ is injective, hence $|B_I| \leq \prod_p |B_p|$. Taking logarithms gives nonnegativity and the displayed identity is the definition rearranged. $\square$

Remark 2.3 (Entropy interpretation). If X is uniform on B_I , then $\text{Corr}_I(B)$ dominates the total mutual-information defect between the coordinates, but no entropy inequality is needed below. Cardinalities alone suffice.

2.5 The no-divisor-loss aggregation identity

We assume henceforth that $1 \in B$, so $B \subseteq B^r$ for every $r \geq 1$.

Proposition 2.4 (Weighted aggregation of local growth). *Let $r \geq 1$ be fixed and $I \subseteq \mathcal{P}$. Then*

$$\log \frac{|(B^r)_I|}{|B_I|} \geq \sum_{p \in I} \log \frac{|(B^r)_p|}{|B_p|} - \text{Corr}_I(B^r).$$

Consequently, if $M \subseteq I$ and for every $p \in M$

$$\log \frac{|(B^r)_p|}{|B_p|} \geq \kappa w_p,$$

then

$$\log \frac{|(B^r)_I|}{|B_I|} \geq \kappa W(M) - \text{Corr}_I(B^r).$$

Proof. By Lemma 2.2,

$$\log |(B^r)_I| = \sum_{p \in I} \log |(B^r)_p| - \text{Corr}_I(B^r).$$

On the other hand $|B_I| \leq \prod_{p \in I} |B_p|$, hence

$$-\log |B_I| \geq -\sum_{p \in I} \log |B_p|.$$

Adding proves the first inequality. Since $B_p \subseteq (B^r)_p$, all omitted one-factor gains are nonnegative, so restricting the sum to M gives the second. $\square$

Corollary 2.5 (Macroscopic active weight forces global growth). *Fix $\theta, \kappa > 0$. If $W(M) \geq \theta W(I)$ and*

$$\text{Corr}_I(B^r) \leq \frac{\kappa \theta}{2} W(I),$$

then

$$|(B^r)_I| \geq e^{\kappa \theta W(I)/2} |B_I|.$$

In particular, the gain depends on conductor weight and not on $|I|$ or $\omega(q)$.

Proof. Substitute the two hypotheses into Proposition 2.4. $\square$

Remark 2.6 (The old W^2 error cannot occur). There is one subtraction, $\text{Corr}_I(B^r)$, after all local gains have been summed. We never pay an $O(\log K)$ defect once for each prime and then multiply it by $W(I)$.

Remark 2.7 (Worked example: A two-factor correlation ledger). Suppose a two-factor projection has

$$|B_1| = 4, \quad |B_2| = 8, \quad |B_I| = 16,$$

so the initial coordinates are correlated. After a fixed product power, suppose

$$|(B^r)_1| = 8, \quad |(B^r)_2| = 16, \quad \text{Corr}_I(B^r) = \log 2.$$

The two local gains are both $\log 2$. Proposition 2.4 therefore gives

$$\log \frac{|(B^r)_I|}{|B_I|} \geq \log 2 + \log 2 - \log 2 = \log 2.$$

Only one $\log 2$ correlation charge is paid. Indeed the stated final correlation means $|(B^r)_I| = (8 \cdot 16)/2 = 64$, so the actual growth from 16 is a factor 4; the proposition keeps a safe factor-2 lower bound without ever multiplying the correlation loss by the number of factors. Paying the same defect once per coordinate would erase the guaranteed gain entirely.

2.6 Weighted exponent-band selection

The next elementary lemma is the arithmetic replacement for iterating over all prime factors.

Lemma 2.8 (A negligible transition band). *Let $0 < \eta < 1$ and set $J := \lceil \eta^{-1} \rceil + 1$ and $L_j := 2^j$ for $0 \leq j \leq J$. There is an index $1 \leq j \leq J$ such that, with*

$$\mathcal{P}_{\text{tr}} := \{p : L_{j-1} < n_p \leq L_j\},$$

one has

$$W(\mathcal{P}_{\text{tr}}) \leq \eta W(\mathcal{P}).$$

Hence the prime factors split as

$$\mathcal{P} = \mathcal{P}_{\text{bd}} \sqcup \mathcal{P}_{\text{tr}} \sqcup \mathcal{P}_{\text{deep}},$$

where $n_p \leq L_{j-1}$ on the bounded-depth block, $n_p > L_j$ on the deep block, the transition block has weight at most ηW , and both thresholds depend only on η .

Proof. The J transition bands are pairwise disjoint. If each had weight $> \eta W(\mathcal{P})$, their total weight would exceed $J\eta W(\mathcal{P}) > W(\mathcal{P})$, impossible. $\square$

Remark 2.9 (Relation with the classical arbitrary-modulus proof). This is the same structural reason that exponent bands, rather than individual prime factors, are the correct units. Bourgain–Varjú use normalized prime-power depths and a weighted modulus function for precisely this purpose. The proof here is self-contained and only uses the disjointness of the bands.

2.7 Local state decomposition

The next statement deliberately keeps the terminal alternative mixed. This prevents the false inference that a balanced small/dense split must be close to one of the two pure alternatives.

Definition 2.10 (Local states). Fix $0 < a < b < 1$. For a symmetric $B \ni 1$ and a fixed product exponent r , classify a prime-power coordinate p as follows.

- (i) p is *small* if $\log |B_p| \leq aw_p$.
- (ii) p is *dense/absorbed* if the local Volume-X/Volume-XI theorem places $(B^r)_p$ in its dense or principal-congruence absorption branch.
- (iii) otherwise p is *mesoscopic*.

Write S, D, M for the corresponding sets of primes.

Theorem 2.11 (Weighted local-to-global product theorem: correlation interface). *Assume the earlier local rank-one product theorems give constants $r, \kappa > 0$ such that every mesoscopic coordinate satisfies*

$$\log \frac{|(B^r)_p|}{|B_p|} \geq \kappa w_p.$$

Fix $\theta > 0$ and suppose the conductor-sensitive correlation interface gives

$$\text{Corr}_{\mathcal{P}}(B^r) \leq \frac{\kappa\theta}{2} W(\mathcal{P}).$$

Then exactly one of the following holds:

(A) global growth: if $W(M) \geq \theta W(\mathcal{P})$, then

$$|B^r| \geq e^{\kappa\theta W(\mathcal{P})/2} |B|;$$

(B) mixed terminal obstruction: $W(M) < \theta W(\mathcal{P})$, so outside a set of conductor weight $< \theta W$ every coordinate is either small or dense/absorbed.

No conclusion that S or D individually has almost all the weight is asserted.

Proof. If $W(M) \geq \theta W$, apply Corollary 2.5 with $I = \mathcal{P}$. Since $B_I = B$ under CRT, this is (A). Otherwise (B) is the definition of the partition. $\square$

Corollary 2.12 (Slow approximate groups have a mixed weighted obstruction). *Suppose B is a K -approximate subgroup, so for each fixed r ,*

$$|B^r| \leq K^{c_r} |B|.$$

Under the hypotheses of Theorem 2.11, if

$$\log K < \frac{\kappa\theta}{4c_r} W(\mathcal{P}),$$

then the global-growth branch is impossible and the mixed terminal obstruction holds.

Proof. The approximate-group upper bound is

$$\log \frac{|B^r|}{|B|} \leq c_r \log K < \frac{\kappa\theta}{4} W,$$

whereas the growth branch gives at least $\kappa\theta W/2$. $\square$

2.8 No-divisor-loss iteration

We now record why an iterative version costs only a bounded number of rounds.

Proposition 2.13 (Conductor-weight potential iteration). *Fix $\theta > 0$. Suppose one iteration replaces B by a fixed bounded product power and has the following property: unless global growth already occurs, at least $\theta W(\mathcal{P})$ of previously mesoscopic conductor weight becomes terminal (small or dense/absorbed), while terminal coordinates never become mesoscopic again. Then after at most $\lceil \theta^{-1} \rceil$ rounds there is no mesoscopic weight $\geq \theta W(\mathcal{P})$.*

If each round replaces B by at most its r -th power, the final word length is at most $r^{\lceil \theta^{-1} \rceil}$, a constant depending on θ but independent of q and $\omega(q)$. If each round has correlation defect at most $c\theta W$, the total correlation charge is $O_\theta(cW)$, again with no divisor-count factor.

Proof. Let U_j be the total weight of coordinates that are still mesoscopic after round j . In every non-growth round with $U_j \geq \theta W$, the hypothesis gives $U_{j+1} \leq U_j - \theta W$. Since $0 \leq U_j \leq W$, there are at most $\lceil 1/\theta \rceil$ such rounds. Word lengths multiply by at most r per round. Correlation charges add over the bounded number of rounds. $\square$

2.9 The corrected obstruction statement

Theorem 2.14 (Uniform approximate-group obstruction: weighted form). *Assume the local product inputs and the correlation interface of Theorem 2.11. Let $B \subseteq G(q)$ be a sufficiently slowly growing approximate subgroup. After a bounded product power depending only on the fixed thresholds, there is a partition*

$$\mathcal{P} = S \sqcup D \sqcup E, \quad W(E) < \theta W(\mathcal{P}),$$

such that:

- (i) *the projection on every coordinate in S is quantitatively small;*
- (ii) *on every coordinate in D the corresponding bounded product is in the dense/absorbed local branch;*
- (iii) *E is the only unresolved part and has negligible conductor weight.*

This mixed $S/D/E$ object, rather than a false pure small/dense dichotomy, is the structured obstruction that the arbitrary-modulus nonconcentration theorem must rule out in Chapter XII.3.

Proof. Apply Corollary 2.12 and, if the local state mechanism is iterative, Proposition 2.13. The final unresolved set is $E = M$ and has weight $< \theta W$. The two terminal classes retain the defining local conclusions. No comparison between $W(S)$ and $W(D)$ is required. $\square$

2.10 Compatibility of squarefree correlations with prime-power depth

2.11 Scope

The squarefree approximate-Goursat theorem is proved in Volume X. Here we show that replacing squarefree factors by prime-power factors creates no new exact cross-prime graph obstruction. The genuinely new structure is vertical inside a single residue characteristic, where the prime-power theorem of Volume XI applies.

Throughout, $p, r \geq 5$ are distinct allowed primes and

$$G_{p,n} := \mathrm{SL}_2(\mathbf{Z}/p^n\mathbf{Z}).$$

We use the Volume-XI vertical theorem inside a fixed $G_{p,n}$ and the Volume-X approximate-Goursat/correlation theorem across different primes.

2.12 Perfectness and simple quotients of a prime-power factor

Lemma 2.15 (Elementary generation over $\mathbf{Z}/p^n\mathbf{Z}$). *For p odd, $G_{p,n}$ is generated by the elementary unipotents*

$$u(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \quad \ell(y) = \begin{pmatrix} 1 & 0 \\ y & 1 \end{pmatrix}.$$

Proof. Let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G_{p,n}$. Since the ring is local, at least one of a, c is a unit. If a is a unit, left multiplication by $\ell(-ca^{-1})$ makes the lower-left entry zero; determinant one then gives a

diagonal matrix times an upper unipotent. A diagonal matrix $\text{diag}(t, t^{-1})$ is a product of elementary unipotents by the standard identity

$$\begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} = u(t)\ell(-t^{-1})u(t)u(-1)\ell(1)u(-1),$$

which is verified by multiplication. If c is the unit, first interchange the two rows using the elementary Weyl element $u(1)\ell(-1)u(1)$. Thus every element is elementary-generated. $\square$

Lemma 2.16 (Perfectness). *For $p \geq 5$, the group $G_{p,n}$ is perfect.*

Proof. Choose $a \in (\mathbf{Z}/p^n\mathbf{Z})^\times$ with $a^2 - 1$ a unit; for example any lift of 2 works for $p \geq 5$. Put $h(a) = \text{diag}(a, a^{-1})$. Then

$$[h(a), u(t)] = u((a^2 - 1)t).$$

Since $a^2 - 1$ is invertible, every $u(x)$ is such a commutator. Similarly every $\ell(y)$ is a commutator. Lemma 2.15 gives perfectness. $\square$

Lemma 2.17 (Congruence kernel is a p -group). *The kernel of $G_{p,n} \rightarrow \text{SL}_2(\mathbf{F}_p)$ is a finite p -group.*

Proof. Its filtration by principal congruence subgroups has successive quotients isomorphic, as additive groups, to subgroups of $\mathfrak{sl}_2(\mathbf{F}_p)$; hence every successive quotient is a p -group. The kernel is therefore a p -group. $\square$

Earlier input 2.18 (Earlier finite-field owner: normal subgroups of $\text{SL}_2(\mathbf{F}_p)$). For $p \geq 5$, every nonabelian simple quotient of $\text{SL}_2(\mathbf{F}_p)$ is $\text{PSL}_2(\mathbf{F}_p)$; equivalently, the only proper noncentral normal subgroup is the center. This is proved in the prime-field group-structure chapter of Volume IV.

Proposition 2.19 (Simple quotients). *For $p \geq 5$, every nontrivial finite simple quotient of $G_{p,n}$ is isomorphic to $\text{PSL}_2(\mathbf{F}_p)$.*

Proof. Let $\phi : G_{p,n} \twoheadrightarrow Q$ with Q simple and nontrivial. Perfectness rules out an abelian simple quotient. The image of the normal congruence p -group from Lemma 2.17 is a normal p -subgroup of the nonabelian simple group Q , hence trivial. Thus ϕ factors through $\text{SL}_2(\mathbf{F}_p)$. Earlier input 2.18 says that the only nonabelian simple quotient is $\text{PSL}_2(\mathbf{F}_p)$. $\square$

Corollary 2.20 (Different prime-power factors have no common nontrivial quotient). *If $p \neq r$ are allowed primes, then $G_{p,n}$ and $G_{r,m}$ have no common nontrivial quotient.*

Proof. If a common nontrivial quotient existed, taking a simple quotient of it would give a common nontrivial simple quotient. Proposition 2.19 would force $\text{PSL}_2(\mathbf{F}_p) \cong \text{PSL}_2(\mathbf{F}_r)$. Their orders are

$$\frac{p(p^2 - 1)}{2}, \quad \frac{r(r^2 - 1)}{2},$$

and this function is strictly increasing for primes at least 5, so $p = r$, a contradiction. $\square$

2.13 Exact Goursat compatibility

Lemma 2.21 (Goursat without a graph branch). *Let $H \leq G_{p,n} \times G_{r,m}$ project surjectively to both factors. Then*

$$H = G_{p,n} \times G_{r,m}.$$

More generally, a subdirect subgroup of a product of prime-power factors at pairwise distinct allowed primes is the full product.

Proof. Define

$$N_p := \{g \in G_{p,n} : (g, 1) \in H\}, \quad N_r := \{h \in G_{r,m} : (1, h) \in H\}.$$

Surjectivity of the two coordinate projections shows that N_p and N_r are normal. If $(g, h) \in H$, send the coset gN_p to hN_r . This is well defined: two choices with the same first coset differ by an element of H whose first coordinate lies in N_p , and multiplying by the corresponding element of $N_p \times \{1\} \subseteq H$ shows that the second-coordinate ratio lies in N_r . The same argument in the opposite direction proves that this map is a bijective homomorphism

$$G_{p,n}/N_p \cong G_{r,m}/N_r.$$

Thus the usual Goursat quotient is constructed here rather than imported. By Corollary 2.20, the common quotient is trivial, so both N 's are the full factors and H is the product. The many-factor statement follows by induction: any nontrivial simple quotient of the product of the first factors is one of their simple quotients, so it cannot also be a quotient of the next distinct-prime factor. $\square$

Remark 2.22. This exact lemma is the reason prime-power depth is a *local* complication. It enlarges each one-prime factor by a filtered p -group, but it does not create an exact diagonal relation between two different residue characteristics.

2.14 Squarefree approximate Goursat and the prime-power compatibility gap

We first record the earlier Volume-X theorem at exactly the strength actually proved there. It is a theorem about the *squarefree residue product*; prime-power depth is not part of its statement.

Earlier input 2.23 (Squarefree conductor-sensitive approximate Goursat). Volume X proves the following form. For

$$P = \prod_{p \in I} \mathrm{SL}_2(\mathbb{F}_p)$$

with pairwise distinct allowed primes and a K -approximate subgroup $\bar{A} \subseteq P$, a bounded product power must satisfy at least one of:

- (i) $\mathrm{Corr}_I(\bar{A}^C) \leq C \log K + \eta W(I)$;
- (ii) macroscopic concentration in a proper local subgroup/section obstruction;
- (iii) a genuine subdirect/common-section graph obstruction after deleting $O(\eta W(I))$ conductor weight.

The proof in Volume X keeps alternatives (ii) and (iii) explicit and carries an independent capture-loss parameter. These quantitative features are essential below: one cannot rule out correlation merely from the nonisomorphism of the ambient simple factors.

Remark 2.24 (Why this is not yet the prime-power theorem). If

$$A \subseteq \prod_{p \in I} \mathrm{SL}_2(\mathbb{Z}/p^{n_p}\mathbb{Z}),$$

then reducing A modulo the primes gives a set $\bar{A}$ to which the preceding theorem applies. However, a correlation estimate for $\bar{A}$ does not by itself control the correlations among the fibers of $A \rightarrow \bar{A}$. Those fibers live in products of principal congruence kernels. Exact Goursat theory rules out an *exact* diagonal quotient between different residue characteristics, but an approximate subgroup requires a quantitative regularization/entropy argument before one may conclude that its fibers are approximately independent. This quantitative lifting step is a separate part of the argument.

Proposition 2.25 (Exact two-level information decomposition). *Let $X = (X_p)_{p \in I}$ be any random variable on $\prod_{p \in I} G_{p, n_p}$, and let Y_p be its reduction modulo p . Put $Y = (Y_p)_{p \in I}$. For a finite family of random variables write*

$$\mathcal{I}(X_p : p \in I) := \sum_{p \in I} H(X_p) - H(X)$$

for its total correlation. Then

$$\mathcal{I}(X_p : p \in I) = \mathcal{I}(Y_p : p \in I) + \sum_{p \in I} H(X_p | Y_p) - H(X | Y). \quad (2.4)$$

The second term is nonnegative. Thus every cross-prime dependence splits canonically into a squarefree residue contribution and a conditional vertical contribution inside the congruence fibers.

Proof. Because Y_p is a function of X_p , the entropy chain rule gives

$$H(X_p) = H(Y_p) + H(X_p | Y_p)$$

for each p , and likewise

$$H(X) = H(Y) + H(X | Y).$$

Subtracting the latter identity from the sum of the former identities gives (2.4). For nonnegativity of the conditional term, use conditional subadditivity:

$$H(X | Y) \leq \sum_{p \in I} H(X_p | Y).$$

Since Y contains Y_p , conditioning reduces entropy and therefore $H(X_p | Y) \leq H(X_p | Y_p)$. Summing proves the claim. $\square$

Proposition 2.26 (Three-level decomposition of the vertical defect). *With the notation of Proposition 2.25, put*

$$\mathcal{V} := \sum_p H(X_p | Y_p) - H(X | Y).$$

Then

$$\begin{aligned} \mathcal{V} = & \sum_{p \in I} (H(X_p | Y_p) - H(X_p | Y)) \\ & + \mathbb{E}_{Y=y} \left[\sum_{p \in I} H(X_p | Y = y) - H(X | Y = y) \right]. \end{aligned} \quad (2.5)$$

Both terms are nonnegative. Consequently, if $\mathcal{V} \geq \eta W(I)$, then at least one of the following holds:

(a) for some p ,

$$H(X_p | Y_p) - H(X_p | Y) \geq \frac{\eta}{2} w_p;$$

(b) a set of residue fibers of probability at least $\eta/4$ satisfies

$$\sum_p H(X_p | Y = y) - H(X | Y = y) \geq \frac{\eta}{4} W(I).$$

Thus a large quantitative lifting defect is either a residue-to-vertical dependence at one residue characteristic or genuine cross-prime correlation inside a positive proportion of full congruence fibers.

Proof. Add and subtract $\sum_p H(X_p | Y)$ in the definition of $\mathcal{V}$. The first line of (2.5) is nonnegative because conditioning on the full vector Y can only decrease entropy relative to conditioning on Y_p . The second line is the conditional total correlation of the variables X_p inside a fixed fiber $Y = y$, averaged over y , and is nonnegative by subadditivity.

If the first term is at least $\eta W(I)/2$, then, since $W(I) = \sum_p w_p$, one coordinate satisfies (a). Otherwise the second term is at least $\eta W(I)/2$. Its integrand $Z(y)$ is at most $\sum_p \log |G_{p,n_p}| = W(I)$. If $r = \mathbb{P}[Z(Y) \geq \eta W(I)/4]$, then

$$\frac{\eta}{2} W(I) \leq \mathbb{E} Z \leq r W(I) + (1 - r) \frac{\eta}{4} W(I),$$

so $r \geq (\eta/4)/(1 - \eta/4) \geq \eta/4$. This is (b). $\square$

Remark 2.27 (What remains after the decomposition). Proposition 2.26 is purely information-theoretic and therefore incurs no $\omega(q)$ loss. It reduces the quantitative compatibility theorem to two concrete mechanisms: flatten a one-prime vertical law that remembers the other residue coordinates, and decouple correlated products of principal-congruence p -groups at distinct primes. The subsequent seed-isolation, orbital-spreading, and deep propagation arguments now close those mechanisms. The final assembly is Theorem 2.181 below.

Corollary 2.28 (Uniform-set comparison with set correlation). *If X is uniform on a nonempty set $A \subseteq \prod_p G_{p,n_p}$, then*

$$\mathcal{I}(X_p : p \in I) \leq \text{Corr}_I(A).$$

Equality holds precisely when every one-coordinate marginal X_p is uniform on A_p . Consequently, after any regularization that makes the one-coordinate marginals $e^{o(W)}$ -close to uniform in entropy, the set-correlation defect is, up to $o(W)$, the sum of the squarefree residue total correlation and the conditional vertical total correlation in (2.4).

Proof. Since X is uniform, $H(X) = \log |A|$. Each marginal is supported on A_p , so $H(X_p) \leq \log |A_p|$. Hence

$$\mathcal{I}(X_p : p \in I) \leq \sum_p \log |A_p| - \log |A| = \text{Corr}_I(A).$$

Equality is exactly the equality case in the maximal-entropy inequality for each marginal. The last assertion is the same calculation with the entropy deficits of the marginals retained as an explicit $o(W)$ error. $\square$

Proposition 2.29 (Exact set-correlation decomposition). *Let X be uniform on a nonempty set $A \subseteq \prod_{p \in I} G_{p, n_p}$, and write $A_p = \pi_p(A)$. Then*

$$\text{Corr}_I(A) = \mathcal{I}(X_p : p \in I) + \sum_{p \in I} (\log |A_p| - H(X_p)). \quad (2.6)$$

Every term on the right is nonnegative. Moreover, if

$$d_p := \log |A_p| - H(X_p),$$

then

$$d_p \leq \log(|A_p| \|\mu_p\|_2^2), \quad (2.7)$$

where μ_p is the law of X_p . Consequently a macroscopic sum of one-coordinate entropy deficits is already detected by a macroscopic weighted collection of one-prime L^2 -nonuniformities; no prime-by-prime dyadic regularization is required.

Proof. Because X is uniform on A , $H(X) = \log |A|$. Add and subtract $\sum_p H(X_p)$:

$$\sum_p \log |A_p| - \log |A| = \left(\sum_p H(X_p) - H(X) \right) + \sum_p (\log |A_p| - H(X_p)),$$

which is (2.6). Nonnegativity of the first term is subadditivity of entropy; nonnegativity of each d_p is maximality of the uniform law on its support.

For (2.7), Shannon entropy dominates the order-two Renyi entropy:

$$H(X_p) \geq -\log \sum_x \mu_p(x)^2 = -\log \|\mu_p\|_2^2.$$

Therefore

$$d_p \leq \log |A_p| + \log \|\mu_p\|_2^2 = \log(|A_p| \|\mu_p\|_2^2).$$

Thus any large local entropy deficit forces a corresponding excess in the one-prime L^2 norm. Summing (2.6) is global and introduces no factor counting loss. $\square$

Corollary 2.30 (No-loss replacement for conditional-fiber regularization). *Fix $\varepsilon > 0$. If*

$$\sum_p d_p \leq \varepsilon W(I),$$

then set correlation differs from probabilistic total correlation by at most $\varepsilon W(I)$. If instead $\sum_p d_p > \varepsilon W(I)$, then the obstruction is already local: its full size is recorded by the one-prime L^2 -excesses in (2.7), to which the Volume-XI prime-power flattening/obstruction theorem applies coordinatewise with conductor weights. Hence the child previously called “fiber regularization” is replaced by an exact entropy decomposition rather than by $\omega(q)$ separate pigeonhole steps.

Proof. The first assertion is immediate from (2.6). For the second, (2.7) shows that the sum of the logarithmic one-prime L^2 -excesses is at least the sum of the entropy deficits. The later decay argument may therefore charge this quantity directly by conductor weight; no simultaneous choice of dyadic bins is needed. $\square$

Remark 2.31 (Worked diagnostic: what can hide above a perfectly independent residue set). Take two distinct primes p, r and let the residue projection of A be the full product $\text{SL}_2(\mathbb{F}_p) \times \text{SL}_2(\mathbb{F}_r)$. It may still happen, set-theoretically, that for each residue pair only a very small correlated subset of the principal-congruence fiber is retained. Then the first term on the right of (2.4) is zero while the second is positive. This is exactly why the squarefree theorem cannot simply be restated for prime powers.

2.15 A source-faithful deep-band seed from captured mass

The information identities above remain useful diagnostics, but the classical arbitrary-modulus argument does not try to prove independence of every conditional principal-congruence fiber. It first isolates a depth band and manufactures one element whose congruence valuation is controlled on almost all of that band. The next lemmas implement that step in the present proper-divisor induction.

For a divisor $d \mid q$ write

$$K(d) := \ker(G(q) \rightarrow G(d)).$$

If $x \in G(q)$ and $p^{n_p} \parallel q$, let $v_p(x - 1)$ be the largest $0 \leq e \leq n_p$ for which $x \equiv 1 \pmod{p^e}$.

Lemma 2.32 (Divisor-profile entropy is conductor-small). *For every $\varepsilon > 0$ there is $p_0(\varepsilon)$ such that, if all $p \mid q$ satisfy $p \geq p_0$, then*

$$\log \tau(q) \leq \varepsilon \log q \leq \varepsilon W(\mathcal{P}).$$

The same bound holds for the number of divisors supported on any fixed block of prime factors.

Proof. If $q = \prod p^{n_p}$, then

$$\tau(q) = \prod_p (n_p + 1) \leq \prod_p 2^{n_p},$$

so

$$\log \tau(q) \leq (\log 2) \sum_p n_p \leq \frac{\log 2}{\log p_0} \sum_p n_p \log p = \frac{\log 2}{\log p_0} \log q.$$

Choose p_0 so that the prefactor is at most ε . Finally $\log q \leq W(\mathcal{P})$ because $|\mathrm{SL}_2(\mathbb{Z}/p^n\mathbb{Z})| \geq p^n$. Restricting the product to a block only decreases the count. $\square$

Lemma 2.33 (Captured collision in a congruence kernel). *Let μ be a probability measure on $G(q)$ and suppose $\mu(xH) \geq \alpha$ for a nonempty set $H \subseteq G(q)$. Put $\tilde{\mu}(g) = \mu(g^{-1})$ and $\lambda = \tilde{\mu} * \mu$. Then for every divisor $d \mid q$,*

$$\lambda(H^{-1}H \cap K(d)) \geq \frac{\alpha^2}{|G(d)|}.$$

Proof. Partition xH according to its residue in $G(d)$. If m_y is the μ -mass in the residue class y , then $\sum_y m_y \geq \alpha$. Two independent points drawn from the restricted mass in the same residue class have quotient in $H^{-1}H \cap K(d)$. Hence

$$\lambda(H^{-1}H \cap K(d)) \geq \sum_{y \in G(d)} m_y^2 \geq \frac{(\sum_y m_y)^2}{|G(d)|} \geq \frac{\alpha^2}{|G(d)|},$$

by Cauchy–Schwarz. $\square$

Lemma 2.34 (Proper-divisor mixing bounds deeper collisions). *Assume the lazy arithmetic walk has spectral gap at least $\sigma > 0$ on every proper divisor of q . Let $\mu = \nu_q^{(\ell)}$ with $\ell \geq 2\sigma^{-1}W(\mathcal{P})$ and put $\lambda = \tilde{\mu} * \mu$. Then, for every proper divisor $d \mid q$ and all sufficiently large conductors,*

$$\lambda(K(d)) = \|\mu_d\|_2^2 \leq \frac{2}{|G(d)|}.$$

Here μ_d denotes the push-forward to $G(d)$ and the L^2 norm uses counting measure.

Proof. Let u_d be the uniform probability on $G(d)$. Laziness and the spectral gap give

$$\|\mu_d - u_d\|_2 \leq e^{-\sigma\ell}.$$

Orthogonality to constants yields

$$\|\mu_d\|_2^2 = |G(d)|^{-1} + \|\mu_d - u_d\|_2^2.$$

Since $\log |G(d)| \leq W(\mathcal{P})$ and $2\sigma\ell \geq 4W(\mathcal{P})$, the second term is at most $|G(d)|^{-1}$ once the finite bounded-conductor range is absorbed into the fixed exceptional set. Finally $\lambda(K(d))$ is exactly the collision probability that two independent μ -distributed points have the same image in $G(d)$, namely $\|\mu_d\|_2^2$. $\square$

Lemma 2.35 (Uniform fractional-depth weight comparison). *Fix $0 < c_1 < c_2 < 1$. There is $L_0 = L_0(c_1, c_2)$ such that the following holds. Let D be a block of primes with $n_p \geq L_0$. Put*

$$d_1 = \prod_{p \in D} p^{\lceil c_1 n_p \rceil}.$$

If $J \subseteq D$ and

$$d_J := \prod_{p \in J} p^{\lceil c_2 n_p \rceil} \prod_{p \in D \setminus J} p^{\lceil c_1 n_p \rceil},$$

then

$$\log |G(d_J)| - \log |G(d_1)| \geq \frac{c_2 - c_1}{3} W(J).$$

Moreover d_J is a proper divisor of q .

Proof. For odd p and $e \geq 1$,

$$|G(p^e)| = p^{3e-2}(p^2 - 1),$$

so increasing the local depth from e_1 to e_2 multiplies the group order by exactly $p^{3(e_2 - e_1)}$. Hence the left side is

$$3 \sum_{p \in J} (\lceil c_2 n_p \rceil - \lceil c_1 n_p \rceil) \log p.$$

For $n_p \geq L_0$ the parenthesis is at least $(c_2 - c_1)n_p - 1$. On the other hand $\log |G(p^{n_p})| \leq (3n_p + 1) \log p$. Taking L_0 sufficiently large makes the displayed increase at least $((c_2 - c_1)/3) \log |G(p^{n_p})|$ for every $p \in J$; summing gives the claim. Increasing L_0 once more ensures $\lceil c_2 n_p \rceil < n_p$, so $d_J < q$. $\square$

Proposition 2.36 (Macroscopic exact-depth seed on a deep block). *Fix $\theta, \eta > 0$ and $0 < c_1 < c_2 < 1$. Assume the proper-divisor spectral gap in Lemma 2.34. Let $D \subseteq \mathcal{P}$ satisfy*

$$W(D) \geq \theta W(\mathcal{P}), \quad n_p \geq L_0 \quad (p \in D),$$

and let $d_1 = \prod_{p \in D} p^{\lceil c_1 n_p \rceil}$. Suppose

$$\mu(xH) \geq e^{-\varepsilon W(\mathcal{P})}$$

for $\mu = \nu_q^{(\ell)}$ at the time in Lemma 2.34. If $\varepsilon > 0$ is sufficiently small and the excluded-prime threshold sufficiently large in terms of θ, η, c_1, c_2 , then there is

$$x_0 \in H^{-1}H \cap K(d_1)$$

such that the exceptional set

$$J(x_0) := \{p \in D : v_p(x_0 - 1) \geq \lceil c_2 n_p \rceil\}$$

has

$$W(J(x_0)) < \eta W(D).$$

Thus on all but η of the deep conductor weight, x_0 has exact normalized congruence depth in the interval $[c_1, c_2)$.

Proof. Put $\alpha = e^{-\varepsilon W(\mathcal{P})}$ and $\lambda = \tilde{\mu} * \mu$. Lemma 2.33 gives

$$\lambda(H^{-1}H \cap K(d_1)) \geq \alpha^2 |G(d_1)|^{-1}. \quad (2.8)$$

If $x \in K(d_1)$ and $W(J(x)) \geq \eta W(D)$, then $x \in K(d_J)$, where d_J is defined in Lemma 2.35. By that lemma, with $\kappa = (c_2 - c_1)/3$,

$$|G(d_J)| \geq |G(d_1)| e^{\kappa \eta W(D)}.$$

Every d_J is proper, so Lemma 2.34 yields

$$\lambda(K(d_J)) \leq 2|G(d_J)|^{-1} \leq 2|G(d_1)|^{-1} e^{-\kappa \eta W(D)}.$$

There are at most $\tau(q)$ possible profiles. By Lemma 2.32, after increasing the minimum allowed prime,

$$\log \tau(q) \leq \frac{\kappa \eta \theta}{4} W(\mathcal{P}).$$

Since $W(D) \geq \theta W(\mathcal{P})$, the total λ -mass of all bad profiles is at most

$$2|G(d_1)|^{-1} \exp\{-\tfrac{3}{4} \kappa \eta \theta W(\mathcal{P})\}.$$

Choose $\varepsilon < \kappa \eta \theta / 16$. Then this is strictly smaller than the lower bound (2.8) for all sufficiently large conductors. Hence some x_0 counted in (2.8) is not in any bad profile, which is exactly the desired conclusion. $\square$

Remark 2.37 (What this closes and what it does not). Proposition 2.36 is the first genuinely arbitrary-modulus piece needed beyond the entropy decomposition. It is the analogue of the controlled-depth element selected in the large-exponent part of the Bourgain–Varjú argument. It is conductor-uniform because depth profiles are counted by $e^{o(W)}$, not one prime at a time.

At this point of the proof the seed still has to be isolated from the bounded-depth complementary CRT block. Proposition 2.42 below performs that correction. The subsequent deep Lie propagation is now also proved internally in Proposition 2.47. The remaining analytic ingredients are proved locally as well: mean projective contraction for the standard rank-one walk, the resulting transfer-operator large deviations, and the amplified nine-dimensional RM–LD estimates are all proved in the local random-matrix module. Thus no random-matrix or projective-contraction primitive remains at this stage.

2.16 Simultaneous lifting on a bounded-depth block

The correction used to isolate a deep seed must not be performed one prime at a time. The next proposition shows that this is unnecessary. Once a bounded power is onto the squarefree residue product of a bounded-exponent block, all higher congruence layers can be lifted simultaneously outside an arbitrarily small conductor-weight exception.

Proposition 2.38 (Simultaneous bounded-depth lifting). *Fix an integer $L \geq 2$ and $\eta > 0$. There is $C_* = C_*(L, \eta, C) < \infty$ with the following property. Let B be a block of distinct allowed primes and let $1 \leq e_p \leq L$ for $p \in B$. Put*

$$G_B^{(e)} = \prod_{p \in B} \mathrm{SL}_2(\mathbb{Z}/p^{e_p}\mathbb{Z}), \quad G_B^{(1)} = \prod_{p \in B} \mathrm{SL}_2(\mathbb{F}_p).$$

Let $A \subseteq G_B^{(e)}$ be symmetric and contain 1. If

$$\pi_1(A^C) = G_B^{(1)},$$

then, for all sufficiently large allowed primes, there is a block $E \subseteq B$ with

$$W(E) \leq \eta W(B)$$

such that

$$\pi_{B \setminus E}(A^{C_*}) = \prod_{p \in B \setminus E} \mathrm{SL}_2(\mathbb{Z}/p^{e_p}\mathbb{Z}).$$

The exponent C_* depends on L, η, C but not on $|B|$.

Proof. We lift one congruence level at a time, but at each level we treat all active prime factors simultaneously. For $j \geq 1$ let

$$R_j = \prod_{p \in B} \mathrm{SL}_2(\mathbb{Z}/p^{\min(j, e_p)}\mathbb{Z}).$$

Assume inductively that, on a current good block $B_j \subseteq B$, a bounded power A^{C_j} projects onto R_j . This is the hypothesis for $j = 1$. Let

$$U_j = \{p \in B_j : e_p \geq j + 1\}$$

be the coordinates on which another layer remains to be lifted. The kernel of $R_{j+1} \rightarrow R_j$ on U_j is the additive group

$$L_j = \prod_{p \in U_j} \mathfrak{sl}_2(\mathbb{F}_p),$$

where multiplication in the congruence layer is addition.

Choose a set-theoretic section $s : R_j(B_j) \rightarrow A^{C_j}$ of the projection. For $x, y \in R_j(B_j)$ put

$$d(x, y) = s(x)s(y)s(xy)^{-1} \in A^{3C_j} \cap L_j$$

after projecting to the next layer. We first record an absolute lower bound on the probability that the p -coordinate $d_p(x, y)$ is nonzero. If for some active p one had

$$\mathbb{P}_{x,y}[d_p(x, y) = 0] > 1 - \varepsilon_0$$

with $\varepsilon_0 > 0$ smaller than the majority-correction threshold proved in Volume X, apply that correction theorem to the map

$$f_p : R_j(B_j) \longrightarrow \mathrm{SL}_2(\mathbb{Z}/p^{j+1}\mathbb{Z}), \quad f_p(x) = \pi_p(s(x)).$$

It would give a homomorphism ϕ_p agreeing with f_p on more than half of $R_j(B_j)$. Reduction modulo p^j shows that $\rho \circ \phi_p$ agrees with the coordinate projection on more than half of the finite group, hence the two homomorphisms are equal. Restricting ϕ_p to the embedded p -coordinate would therefore give a group section of

$$\mathrm{SL}_2(\mathbb{Z}/p^{j+1}\mathbb{Z}) \longrightarrow \mathrm{SL}_2(\mathbb{Z}/p^j\mathbb{Z}),$$

contradicting the nonsplitting theorem proved in Volume XI. Thus, for one absolute $\varepsilon_0 > 0$,

$$\mathbb{P}[d_p(x, y) \neq 0] \geq \varepsilon_0 \tag{2.9}$$

for every active coordinate.

Let $V \subseteq U_j$ be any set of still-uncovered active coordinates. By (2.9),

$$\mathbb{E}_{x,y} W\{p \in V : d_p(x, y) \neq 0\} \geq \varepsilon_0 W(V).$$

Hence one defect $d(x, y)$ is nonzero on a subblock of weight at least $\varepsilon_0 W(V)$. Repeat this selection after quotienting out the congruence-layer coordinates already generated. After

$$m_0 \leq \left\lceil \frac{\log(L/\eta)}{-\log(1 - \varepsilon_0)} \right\rceil$$

selections, the uncovered active block has weight at most $(\eta/L)W(B)$. The number m_0 is independent of $|B|$.

It remains to show that each selected defect fills its whole new support in bounded products. Let $z \in A^{3C_j} \cap L_j$ be one selected defect and let T be the set of new coordinates on which its image is nonzero. Volume XI proves that, for every sufficiently large p and every nonzero $X \in \mathfrak{sl}_2(\mathbb{F}_p)$, ten adjoint conjugates of X fill $\mathfrak{sl}_2(\mathbb{F}_p)$ additively. Given an arbitrary target $(Y_p)_{p \in T}$, choose ten local conjugators $g_{t,p}$ realizing Y_p from z_p . Since A^{C_j} is onto $R_j(B_j)$, the tuples $(g_{t,p})_{p \in T}$ can be lifted simultaneously to elements of A^{C_j} . The product of the corresponding ten conjugates of z realizes $(Y_p)_{p \in T}$ in the layer. Previously generated layer coordinates are then corrected using the already-filled kernel; equivalently, perform this argument in the quotient by the previously generated coordinates. Thus a bounded power of A contains the full layer on the union of the selected supports.

Delete the uncovered active coordinates and call the resulting block B_{j+1} . On B_{j+1} a bounded power of A is now onto R_{j+1} : choose a lift of the desired R_j element and correct its residual next-layer coordinate by the full kernel just obtained. The conductor weight deleted at this stage is at most $(\eta/L)W(B)$. There are at most $L - 1$ lifting stages, so the total deleted weight is at most $\eta W(B)$. All product exponents are multiplied only a bounded number of times depending on L, η and the initial C . This proves the proposition. $\square$

Corollary 2.39 (Conditional isolation of the deep seed). *Keep the deep block D and the element g_1 supplied by Proposition 2.36, with exact depth in $[c_1, c_2]$ outside a set of arbitrarily small D -weight. Let B be a complementary block with bounded exponents $e_p \leq L$. Suppose a deeper slice A_2 is contained in $K(d_2)$ on D , where every local exponent of d_2 is at least $\lceil c_2 n_p \rceil$, and suppose that for some bounded C*

$$\pi_{\mathrm{rad}(B)}(A_2^C) = \prod_{p \in B} \mathrm{SL}_2(\mathbb{F}_p).$$

Then, after deleting a block $E_B \subseteq B$ with $W(E_B) \leq \eta W(B)$, there exists a bounded product g_2 of elements of A_2 such that g_2 has the same image as g_1 on the full bounded-depth quotient over $B \setminus E_B$. Consequently

$$x_0 = g_1 g_2^{-1}$$

is trivial on that bounded-depth quotient and retains the exact $[c_1, c_2)$ depth of g_1 at every good prime of D .

Proof. Apply Proposition 2.38 to A_2 on B . Its bounded power is onto the full bounded-depth quotient outside E_B , so choose g_2 with the same bounded-block image as g_1 . On a good deep prime, $g_1 = 1 + p^t X$ with $c_1 n_p \leq t < c_2 n_p$ and $X \not\equiv 0 \pmod{p}$, whereas $g_2 \equiv 1 \pmod{p^{\lceil c_2 n_p \rceil}}$. Therefore the leading $p^t X$ term cannot cancel in $g_1 g_2^{-1}$, and the exact depth remains t . The bounded-block cancellation is exact by construction. $\square$

Lemma 2.40 (Dense squarefree projections complete outside small weight). *Fix $\eta > 0$. There is $\rho = \rho(\eta) > 0$ such that the following holds for all sufficiently large allowed primes. Let*

$$S \subseteq \prod_{p \in B} \mathrm{SL}_2(\mathbb{F}_p)$$

be nonempty and suppose

$$|S| \geq e^{-\rho W(B)} \prod_{p \in B} |\mathrm{SL}_2(\mathbb{F}_p)|.$$

Then there is $E \subseteq B$ with $W(E) \leq \eta W(B)$ such that

$$\pi_{B \setminus E}(S^3) = \prod_{p \in B \setminus E} \mathrm{SL}_2(\mathbb{F}_p).$$

Proof. Let δ_0 be the large-set exponent in the Volume-X Gowers completion lemma. Exact projection-tree regularization from Volume X gives $S' \subseteq S$ with branching degrees m_p and, after making the minimum prime sufficiently large,

$$\log |S'| \geq \log |S| - \xi W(B)$$

for any prescribed $\xi > 0$. Hence

$$\sum_{p \in B} (w_p - \log m_p) = W(B) - \log |S'| \leq (\rho + \xi) W(B). \quad (2.10)$$

Put

$$E = \{p : \log m_p < (1 - \delta_0) w_p\}.$$

Every $p \in E$ contributes more than $\delta_0 w_p$ to the left side of (2.10), so

$$W(E) \leq \frac{\rho + \xi}{\delta_0} W(B).$$

Choose first ξ and then ρ so that this is at most $\eta W(B)$. On $B \setminus E$ every branch degree is at least $|G_p|^{1-\delta_0}$. The simultaneous large-branch completion lemma of Volume X therefore gives

$$\pi_{B \setminus E}((S')^3) = G_{B \setminus E}.$$

Since $S' \subseteq S$, the same is true for S^3 . $\square$

Lemma 2.41 (A captured deeper slice has dense bounded-block residues). *Fix $\theta, \eta > 0$ and $0 < c_2 < 1$. Let D be a deep block and B a disjoint block with $W(B) \geq \theta W(\mathcal{P})$. Put*

$$d_2 = \prod_{p \in D} p^{\lceil c_2 n_p \rceil}.$$

Let H be captured as in Proposition 2.36, and put

$$A_2 = (H^{-1}H) \cap K(d_2), \quad \lambda = \tilde{\mu} * \mu.$$

If c_2 and the capture exponent ε are sufficiently small in terms of θ, η , and the deep threshold is sufficiently large, then there is $E_1 \subseteq B$ with $W(E_1) \leq \eta W(B)$ such that

$$\pi_{B \setminus E_1}(A_2^3) = \prod_{p \in B \setminus E_1} \text{SL}_2(\mathbb{F}_p).$$

Proof. Lemma 2.33, applied with $d = d_2$, gives

$$\lambda(A_2) \geq e^{-2\varepsilon W(\mathcal{P})} |G(d_2)|^{-1}. \quad (2.11)$$

For sufficiently large deep exponents,

$$\log |G(d_2)| \leq 2c_2 W(D) \leq 2c_2 W(\mathcal{P}), \quad (2.12)$$

because the ceiling error is $o(W(D))$. Thus the right side of (2.11) is at least $e^{-\delta_s W(\mathcal{P})}$ with $\delta_s = 2\varepsilon + 2c_2 + o(1)$.

The squarefree residue quotient supported on B is a proper divisor of q . At the logarithmic time fixed in Lemma 2.34, the proper-divisor spectral gap gives

$$\sup_y \lambda_B(y) \leq 2|G_B^{(1)}|^{-1}$$

for sufficiently large conductors. Indeed the L^2 distance from uniform is exponentially smaller than $|G_B^{(1)}|^{-1/2}$. Therefore

$$|\pi_B(A_2)| \geq \frac{\lambda(A_2)}{\sup_y \lambda_B(y)} \geq \frac{1}{2} e^{-\delta_s W(\mathcal{P})} |G_B^{(1)}|.$$

Since $W(B) \geq \theta W(\mathcal{P})$, this is

$$|\pi_B(A_2)| \geq e^{-\rho W(B)} |G_B^{(1)}|$$

with $\rho \leq 2\delta_s/\theta$ after absorbing the harmless factor 2. Choose c_2 and ε so that this ρ is below the threshold in Lemma 2.40. That lemma applied to $S = \pi_B(A_2)$ gives the desired E_1 and the displayed three-product surjectivity. $\square$

Proposition 2.42 (Unconditional isolation of the deep seed). *Fix $\theta, \eta > 0$ and choose $0 < c_1 < c_2 < 1$ with c_2 sufficiently small in terms of θ, η . Let the exponent-band decomposition consist of a bounded-depth block B , a transition block T , and a deep block D , with*

$$W(T) \leq \eta W(\mathcal{P}), \quad W(D) \geq \theta W(\mathcal{P}).$$

Assume H is captured as in Proposition 2.36. If the capture exponent is sufficiently small, then there are

$$x_0 \in (H^{-1}H)^{C_*}$$

and an exceptional block E with

$$W(E) \leq 4\eta W(\mathcal{P})$$

such that:

- (i) x_0 is the identity on the full prime-power quotient over $B \setminus E$;
- (ii) for every $p \in D \setminus E$,

$$c_1 n_p \leq v_p(x_0 - 1) < c_2 n_p;$$

- (iii) no assertion is needed on the transition block or on E .

The exponent C_* depends only on the fixed thresholds, not on q or on $\omega(q)$.

Proof. If $W(B) < \eta W(\mathcal{P})$, put the whole bounded-depth block into E and use the seed from Proposition 2.36; there is nothing to match. Hence suppose $W(B) \geq \eta W(\mathcal{P})$ and apply that proposition with its exceptional deep weight chosen at most $\eta W(D)$, obtaining $g_1 \in H^{-1}H$ with controlled depth on all but that deep exceptional set.

Put

$$d_2 = \prod_{p \in D} p^{\lceil c_2 n_p \rceil}, \quad A_2 = (H^{-1}H) \cap K(d_2).$$

Lemma 2.41, with θ there replaced by the lower bound $W(B)/W(\mathcal{P}) \geq \eta$, gives a block $E_1 \subseteq B$ of weight at most $\eta W(B)$ such that A_2^3 is onto the squarefree residue product of $B \setminus E_1$. All exponents on B are bounded by the threshold selected in Lemma 2.8. Proposition 2.38 therefore produces a further block $E_2 \subseteq B \setminus E_1$ of weight at most $\eta W(B)$ and a bounded power of A_2 that is onto the entire prime-power quotient over

$$B_0 = B \setminus (E_1 \cup E_2).$$

Choose g_2 in that bounded power with the same B_0 -projection as g_1 and set $x_0 = g_1 g_2^{-1}$. Since $A_2 \subseteq K(d_2)$, also $g_2 \in K(d_2)$. Thus x_0 is exactly trivial on B_0 , while at every good deep prime the g_2 term starts strictly deeper than the leading term of g_1 and cannot cancel it. Hence the exact $[c_1, c_2]$ depth is preserved.

Take E to be the union of the transition block, the exceptional deep block, and $E_1 \cup E_2$. Its weight is at most $4\eta W(\mathcal{P})$ after a harmless rescaling of the input η . Every product exponent used above is bounded in terms of the fixed band and error thresholds only. $\square$

Remark 2.43 (What remains after seed isolation). The analogue of the Bourgain–Varjú g_1/g_2 correction is fully internal: Proposition 2.42 produces one element supported, up to negligible conductor weight, on the deep block and with controlled exact depth there. Lemmas 2.44–2.45 and Proposition 2.47 own the rank-one commutator propagation, while Proposition 2.48 and Proposition 2.176 now own the required adjoint-orbit spreading estimate. Thus the large-exponent step has no remaining external source-level child.

2.17 Deep-band Lie propagation: the elementary part of the large-exponent argument

The isolated seed of Proposition 2.42 is exactly the input needed by the large-exponent mechanism. There are, however, two logically different ingredients. The first is an *orbital-spreading* statement saying that bounded additive sums of conjugates of one primitive Lie coefficient occupy a positive exponential proportion of the full traceless matrix space over a deep auxiliary modulus. The second is a purely rank-one Lie calculation: once one such large coefficient set is available, repeated commutators propagate it through consecutive principal congruence layers in only boundedly many stages. We prove the second part completely here.

For an odd integer

$$r = \prod_{p \in D} p^{t_p}$$

write

$$\mathfrak{s}(r) := \mathfrak{sl}_2(\mathbb{Z}/r\mathbb{Z}) \cong \prod_{p \in D} \mathfrak{sl}_2(\mathbb{Z}/p^{t_p}\mathbb{Z}).$$

If $v \in \mathfrak{s}(r)$ and $q_0 \mid r$, we write $q_0 \parallel_r v$ when, at each $p \mid r$, the exact common p -divisibility of the three coordinates of v is the exponent of p in q_0 , truncated at t_p .

Lemma 2.44 (Exact rank-one bracket fibers over a prime power). *Let $p > 2$, $m \geq 1$, and*

$$A_m(v) := \{(u_1, u_2) \in \mathfrak{sl}_2(\mathbb{Z}/p^m\mathbb{Z})^2 : [u_1, u_2] = v\}.$$

If $p^k \parallel_{p^m} v$ with $0 \leq k < m$, then

$$|A_m(v)| = p^{3m+k} + p^{3m+k-1} - p^{3m-1} - p^{3m-2}. \quad (2.13)$$

For $v = 0$ one has

$$|A_m(0)| = p^{4m} + p^{4m-1} + p^{4m-2} - p^{3m-1} - p^{3m-2}. \quad (2.14)$$

In particular, in both cases

$$|A_m(v)| \leq p^{3m+k}(1 + p^{-1} + p^{-2}), \quad (2.15)$$

where $k = m$ when $v = 0$.

Proof. Write

$$v = a_1H + a_2E + a_3F, \quad u_1 = x_1H + x_2E + x_3F, \quad u_2 = y_1H + y_2E + y_3F,$$

with $H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, $E = E_{12}$ and $F = E_{21}$. The equation $[u_1, u_2] = v$ is

$$\begin{aligned} x_2y_3 - x_3y_2 &= a_1, \\ 2x_1y_2 - 2x_2y_1 &= a_2, \\ 2x_3y_1 - 2x_1y_3 &= a_3. \end{aligned} \quad (2.16)$$

For a primitive triple (x_1, x_2, x_3) these equations are solvable exactly when

$$2x_1a_1 + x_3a_2 + x_2a_3 = 0, \quad (2.17)$$

and then there are exactly p^m choices of (y_1, y_2, y_3) . For example, if x_1 is a unit, take y_1 freely and solve the last two equations; the first is then equivalent to (2.17). The other primitive cases are obtained by the same calculation after choosing a unit among x_2, x_3 .

If $p^k \parallel (a_1, a_2, a_3)$ with $k < m$, the linear congruence (2.17) has p^{2m+k} solutions in x . Exactly p^{2m-2+k} of them are divisible by p in all three coordinates, so the number of primitive x is

$$p^{2m+k} - p^{2m-2+k}.$$

Thus the contribution from $p \nmid u_1$ equals $p^{3m+k} - p^{3m+k-2}$.

If $p^\ell \parallel u_1$ with $0 \leq \ell \leq k$, divide u_1 and v by p^ℓ . Reduction from modulus p^m to $p^{m-\ell}$ leaves a factor $p^{3\ell}$ for the three coordinates of u_2 . Consequently the contribution of exact depth ℓ is

$$p^{3m+k-\ell} - p^{3m+k-\ell-2}.$$

Summing this geometric progression for $0 \leq \ell \leq k$ gives (2.13). For $v = 0$ the same sum runs through $0 \leq \ell < m$, and the case $u_1 = 0$ contributes the additional p^{3m} choices of u_2 ; simplification gives (2.14). The upper bound (2.15) is immediate. $\square$

Lemma 2.45 (Rank-one bilinear commutator growth). *There is an absolute C_b such that the following holds for every odd r . Let $X, Z \subseteq \mathfrak{s}(r)$ satisfy*

$$|X| \geq r^3/K_1, \quad |Z| \geq r^3/K_2, \quad K_1, K_2 \geq 2.$$

Put $Y = X - X$ and $V = Z - Z$. Then

$$|[Y, V]| \geq \frac{r^3}{C_b(K_1 K_2)^2 \log^2(2K_1 K_2)}. \quad (2.18)$$

In particular, taking $X = Z$ gives $|[X - X, X - X]| \gg r^3/(K_1^4 \log^2(2K_1))$. The constants are independent of the number of prime factors of r .

Proof. Choose $x_0 \in X$ and $z_0 \in Z$. Translation gives $|Y| \geq |X|$ and $|V| \geq |Z|$. For $q_0 \mid r$ let

$$T_{q_0} := \{v \in [Y, V] : q_0 \parallel_r v\}.$$

By Lemma 2.44 and the Chinese remainder theorem, the number of ambient pairs $(u_1, u_2) \in \mathfrak{s}(r)^2$ with bracket equal to a fixed $v \in T_{q_0}$ is at most

$$C r^3 q_0 \prod_{p \mid q_0} (1 + p^{-1} + p^{-2}) \leq C' r^3 q_0 \log(2q_0). \quad (2.19)$$

For the last inequality, use $1 + p^{-1} + p^{-2} \leq (1 - p^{-1})^{-1}$ and the elementary estimate $\prod_{p \mid n} (1 - p^{-1})^{-1} \ll \log(2n)$. Counting only the brackets of ordered pairs in $Y \times V$ therefore gives

$$\frac{r^6}{K_1 K_2} \leq |Y||V| \leq C' r^3 \sum_{q_0 \mid r} q_0 \log(2q_0) |T_{q_0}|. \quad (2.20)$$

Set

$$L := K_1 K_2, \quad R_0 = C_1 L \log(2L),$$

where C_1 is a sufficiently large absolute constant. Since every vector in T_{q_0} is divisible by q_0 , $|T_{q_0}| \leq (r/q_0)^3$. Hence, enlarging the divisor sum to all integers,

$$C' r^3 \sum_{q_0 > R_0} q_0 \log(2q_0) |T_{q_0}| \leq C' r^6 \sum_{n > R_0} \frac{\log(2n)}{n^2} \leq \frac{r^6}{2L}. \quad (2.21)$$

Subtracting (2.21) from (2.20), while $q_0 \log(2q_0) \ll L \log^2(2L)$ for $q_0 \leq R_0$, yields

$$\frac{r^6}{2L} \leq C'' r^3 L \log^2(2L) \sum_{q_0 \leq R_0} |T_{q_0}|.$$

The T_{q_0} are disjoint subsets of $[Y, V]$, which proves (2.18). $\square$

Lemma 2.46 (Congruence commutators realize Lie brackets). *Let $p > 2$, $a, b, c \geq 1$, $c \leq \min(a, b)$, and $a + b + c \leq n$. If*

$$x \equiv 1 + p^a X \pmod{p^{a+c}}, \quad y \equiv 1 + p^b Y \pmod{p^{b+c}},$$

then

$$[x, y] \equiv 1 + p^{a+b} [X, Y] \pmod{p^{a+b+c}}. \quad (2.22)$$

Consequently, if $t_p \geq 1$ and $jt_p, (j+1)t_p, (j+2)t_p \leq n_p$, the commutator induces on consecutive width- t_p principal-congruence quotients the ordinary Lie bracket on $\mathfrak{sl}_2(\mathbb{Z}/p^{t_p}\mathbb{Z})$.

Proof. Write $x = 1 + A$ and $y = 1 + B$ with $A \in p^a M_2$ and $B \in p^b M_2$. Expanding $(1 + A)(1 + B)(1 + A)^{-1}(1 + B)^{-1}$, every term other than $1 + [A, B]$ contains at least one additional factor A or B and therefore has valuation at least $a + b + \min(a, b)$. Reducing to the stated modulus gives (2.22); the width- t_p assertion follows by taking $c = t_p$. $\square$

Proposition 2.47 (Deep congruence propagation from one orbital-spreading layer). *Fix $0 < c_1 < c_2 < 1/3$ and $\theta > 0$. Let D be a block with $W(D) \geq \theta W(\mathcal{P})$, and suppose $x_0 \in H^{C_0}$ has exact local depths*

$$t_p = v_p(x_0 - 1), \quad c_1 n_p \leq t_p < c_2 n_p \quad (2.23)$$

for every $p \in D$. Suppose further that the following single orbital-spreading estimate holds for the coefficient of x_0 : if

$$r = \prod_{p \in D} p^{t_p},$$

then a bounded product of conjugates of x_0 by elements of H has image X_0 in the first width- t_p Lie quotient satisfying

$$|X_0| \geq e^{-\delta W(\mathcal{P})} r^3. \quad (2.24)$$

Here $\delta > 0$ may be made arbitrarily small by making the capture exponent sufficiently small.

Then, for every $\eta > 0$, after choosing δ and c_2 sufficiently small in terms of c_1, θ, η , one bounded power H^{C_*} satisfies

$$|\pi_D(H^{C_*})| \geq e^{-\eta W(\mathcal{P})} \prod_{p \in D} |\mathrm{SL}_2(\mathbb{Z}/p^{n_p}\mathbb{Z})|. \quad (2.25)$$

The exponent C_* depends only on c_1, θ, η and the fixed arithmetic data, not on q , the depth profile, or $\omega(q)$.

Proof. For each $p \in D$ put

$$M_p := \lfloor n_p/t_p \rfloor.$$

By (2.23), $M_p \leq c_1^{-1}$, so only a bounded number of graded commutator stages can occur. Unlike the earlier version of this argument, we do not pigeonhole a constant value of M_p . Instead the active modulus shrinks as the shorter local filtrations terminate.

For $j \geq 0$ let

$$D_j := \{p \in D : M_p \geq j + 2\}, \quad r_j := \prod_{p \in D_j} p^{t_p}.$$

On D_j the $(j + 1)$ -st width- t_p congruence layer is canonically identified with

$$\mathfrak{sl}(r_j) = \mathfrak{sl}_2(\mathbb{Z}/r_j\mathbb{Z}).$$

Projecting (2.24) from D to any D_j cannot worsen its density parameter: indeed, if $V = V_{D_j} \times V_{D \setminus D_j}$, then $|X_0| \leq |\pi_{D_j} X_0| |V_{D \setminus D_j}|$. Hence

$$|\pi_{D_j} X_0| \geq e^{-\delta W(\mathcal{P})} r_j^3. \quad (2.26)$$

Choose one representative in the bounded product giving each element of X_0 and call the resulting set B_0 . Recursively, while D_j is nonempty, form

$$B_{j+1} = \{[u, v] : u \in B_0 B_0^{-1}, v \in B_j B_j^{-1}\}.$$

Lemma 2.46, restricted to the active set D_j , says that the image of B_{j+1} in the next active layer contains

$$[\pi_{D_j}(X_0 - X_0), \pi_{D_j}(X_j - X_j)].$$

If the current active-layer image X_j has density at least K_j^{-1} , then Lemma 2.45 and (2.26) give, after enlarging an absolute constant,

$$K_{j+1} \leq C_b(K_0 K_j)^2 \log^2(2K_0 K_j), \quad \log K_0 \leq \delta W(\mathcal{P}). \quad (2.27)$$

When $D_{j+1} \subset D_j$, projection again cannot decrease relative density, so the same K_{j+1} is valid for the next shrinking active modulus. Since there are at most $\lceil c_1^{-1} \rceil$ stages,

$$\log K_j \leq C(c_1)\delta W(\mathcal{P}) + O_{c_1}(\log W). \quad (2.28)$$

Choose δ so small that this is at most $(\eta/20)W(\mathcal{P})$ for every stage once the conductor is large.

We now multiply transversals layer by layer. For each j , retain one group representative for every element of the active-layer image of B_j . The multiplication map from the product of these transversals into the quotient of the deep principal-congruence block is injective: reducing first modulo the second layer determines the first transversal coordinate; after cancelling it, reduction modulo the third determines the next; and so on. A prime p simply ceases to contribute after stage $M_p - 2$. Therefore (2.28) yields

$$|H^{C_*} \bmod \prod_{p \in D} p^{M_p t_p}| \geq \exp\{-C(c_1)\delta W(\mathcal{P}) - O_{c_1}(\log W)\} \prod_{p \in D} |K_p(t_p)/K_p(M_p t_p)|, \quad (2.29)$$

with one bounded exponent C_* , because the word-length recursion is run only $O(c_1^{-1})$ times.

Finally compare these graded layers with the full D -block. At each prime, the unfilled first depth has logarithmic size $O(t_p \log p)$, and the terminal tail from $M_p t_p$ to n_p has size $O(t_p \log p)$ because $0 \leq n_p - M_p t_p < t_p$. Since $t_p < c_2 n_p$ and

$$\log |\mathrm{SL}_2(\mathbb{Z}/p^{n_p}\mathbb{Z})| \asymp 3n_p \log p,$$

the total first-layer plus terminal-tail loss is at most $Cc_2 W(D)$. Combining this with (2.29), then choosing first c_2 and then δ sufficiently small, gives (2.25). The finitely many bounded conductors are absorbed by increasing C_* . $\square$

Proposition 2.48 (Parseval reduction of orbital spreading to one Fourier estimate). *Let $r = \prod_{p \in D} p^{t_p}$ be odd and put $V = \mathfrak{s}(r)$, so $|V| = r^3$. Let $v_0 \in V$ be primitive at every $p \mid r$. Suppose the logarithmic-time arithmetic walk has the following uniform adjoint Fourier estimate: for the push-forward σ of the walk under*

$$g \mapsto \mathrm{Ad}(g)v_0,$$

there are $c_0, C_0 > 0$ such that every nontrivial additive character χ of exact conductor $d > 1$ satisfies

$$|\widehat{\sigma}(\chi)| \leq C_0 d^{-c_0}. \quad (2.31)$$

Assume $\mu(xH) = \alpha > 0$ for the same walk measure. Then for every $\delta > 0$ there is a fixed integer $C = C(c_0, C_0)$ such that, whenever $\alpha \geq e^{-\varepsilon W}$ with $2C\varepsilon < \delta$, bounded additive sums of C conjugates of v_0 by elements of H occupy at least

$$c e^{-\delta W} r^3 \quad (2.32)$$

points of V . The constant $c > 0$ is absolute after the arithmetic data are fixed and is independent of $\omega(r)$.

Proof. Use the nondegenerate trace pairing on V to identify additive characters with elements of V . The number of characters whose exact conductor is a given divisor $d \mid r$ is at most d^3 . Choose C so that $2Cc_0 > 5$. Parseval and (2.31) give

$$\|\sigma^{*C}\|_2^2 = r^{-3} \sum_{\chi} |\widehat{\sigma}(\chi)|^{2C} \leq r^{-3} \left(1 + C_0^{2C} \sum_{d \mid r, d > 1} d^{3-2Cc_0} \right) \leq C_1 r^{-3}, \quad (2.33)$$

because the final divisor sum is bounded by the convergent sum over all positive integers. This is the exact point at which the conductor of a Fourier mode, rather than the number of prime factors, is charged.

Normalize the restriction of the walk to the captured coset xH , translate it by x^{-1} , and push it forward under $h \mapsto \text{Ad}(h)v_0$; call the resulting probability τ . Before the translation, the restricted measure is pointwise at most α^{-1} times the unrestricted walk. Translation by x^{-1} merely composes the adjoint phase with the fixed invertible linear map $\text{Ad}(x^{-1})$; it permutes additive characters of each exact conductor. Thus the same unrestricted comparison measure satisfies the Fourier estimate (2.31), and pointwise

$$\tau^{*C} \leq \alpha^{-C} \sigma_x^{*C},$$

where σ_x obeys the same bound (2.33). Consequently

$$\|\tau^{*C}\|_2^2 \leq C_1 \alpha^{-2C} r^{-3}.$$

Since τ^{*C} is a probability measure, Cauchy–Schwarz gives

$$|\text{supp}(\tau^{*C})| \geq \|\tau^{*C}\|_2^{-2} \geq C_1^{-1} \alpha^{2C} r^3.$$

If $\alpha \geq e^{-\varepsilon W}$ and $2C\varepsilon < \delta$, this is (2.32). Every point of the support is a sum of C conjugates by elements of H , as required. $\square$

2.17.1 From torus Fourier rigidity to the congruence Fourier bound

The remaining analytic input can now be stated at its irreducible level. The quantitative torus principle used in the original arbitrary-modulus argument is proved below from one random-matrix large-deviation theorem; its projective and projection-theoretic consequences are no longer external.

Torus Fourier-rigidity theorem. Let $T \subseteq \text{SL}_m(\mathbb{Z})$ be a fixed finite symmetric set whose generated group acts strongly irreducibly and proximally on $\mathbb{R}^m$. For every $0 < \lambda < \lambda_1$ there is C_T such that, if $\eta_T^{(\ell)} * \delta_x$ has a nonzero Fourier coefficient at $b \in \mathbb{Z}^m$ of modulus $> t$ and

$$\ell > C_T \log \frac{2\|b\|}{t},$$

then x is within $e^{-\lambda\ell}$ of a rational point of denominator at most $(2\|b\|/t)^{C_T}$.

Proof architecture. The argument is organized as follows.

- (B1) *Random-matrix transfer core.* The random-matrix subsection proves the required rank-one mean projective contraction from exponential atom decay, the Guivarc’h–Raugi martingale boundary, cocycle recurrence, and a compact occupation-measure argument. It then derives finite-block Hölder contraction, stationary uniqueness, the transfer-operator spectral gap, and

uniform vector deviations. For the actual nine-dimensional left–right adjoint amplification, vector anti-concentration is obtained from the nilpotent conic and the standard SL_2 projective Frostman bound, while the first two singular values are computed directly from the Cartan parameter. No exterior-power or amplified-projective large-deviation theorem is imported.

- (B2) *Projective consequences of B1.* Singular-value geometry, exponential attraction, hyperplane avoidance, scale-wise projective Frostman bounds, and quantitative general position are proved below from B1; they are not a separate imported theorem.
- (B3) *Discretized line projection.* The Bourgain projection input is proved internally in the exact line-projection dimensions needed by the fixed-frequency M_3 amplification: a planar sum–product base followed by the codimension-one slice induction through dimension nine.
- (B4) *Fourier–granulation bootstrap.* Starting from one large Fourier coefficient, the manuscript proves initial coarse dimension, regular-window extraction, repeated projection-driven dimension gain, granulation, concentration bootstrapping, rational-center extraction, and tightening.

Thus the random-matrix, projective, projection, and Fourier–granulation components are supplied by the results above. The BFLM citation records the historical proof architecture. The finite-support and fixed-frequency reductions needed in the congruence application are recorded explicitly in Proposition 2.176.

Definition 2.49 (Bourgain line-projection package). Fix an integer $m \geq 2$. We say that the *Bourgain line-projection package* holds in dimension m if the following is true. For every $\alpha_0, \kappa > 0$ there are constants

$$\Delta_{\text{pr}}, u_{\text{pr}}, \tau_{\text{pr}}, r_{\text{pr}} > 0$$

such that, whenever $0 < r < r_{\text{pr}}$ and $\alpha_0 < \alpha < m - \alpha_0$, the following implication holds. Let $E \subset B(0, 1) \subset \mathbb{R}^m$ be r -separated and satisfy

$$r^{-\alpha+u_{\text{pr}}} \leq |E| \leq r^{-\alpha-u_{\text{pr}}}, \quad (2.1)$$

$$|E \cap B(x, \rho)| \leq r^{-u_{\text{pr}}} \rho^\kappa |E| \quad (r < \rho < r^{\tau_{\text{pr}}}). \quad (2.2)$$

Let ϑ be a probability measure on $\mathbb{P}^{m-1}$ satisfying, for every projective hyperplane H ,

$$\vartheta\{\theta : d_{\mathbb{P}}(\theta, H) < \rho\} \leq r^{-u_{\text{pr}}} \rho^\kappa \quad (r < \rho < r^{\tau_{\text{pr}}}). \quad (2.3)$$

Then there are $E_0 \subset E$ and $\Theta \subset \mathbb{P}^{m-1}$ such that

$$|E_0| \geq r^{u_{\text{pr}}} |E|, \quad \vartheta(\Theta) \geq 1 - r^{u_{\text{pr}}}, \quad (2.4)$$

and, for every $\theta \in \Theta$ and every $E' \subset E_0$ with $|E'| \geq r^{u_{\text{pr}}} |E_0|$,

$$N_r(\pi_\theta E') \geq r^{-\alpha/m - \Delta_{\text{pr}}}. \quad (2.5)$$

Here π_θ is scalar orthogonal projection onto the line θ . In the arithmetic application below $m = 9$.

Definition 2.50 (RM–LD interface). Let η be a finitely supported probability measure on $SL_m(\mathbb{R})$ and let $G_n = g_n \cdots g_1$, where the g_i are independent with law η . We say that η satisfies *RM–LD* if there are $\lambda_1 > \lambda_2$ such that, for every sufficiently small $\varepsilon > 0$, there are $c_\varepsilon, C_\varepsilon > 0$ for which, uniformly in every unit vector $v \in \mathbb{R}^m$,

$$\mathbb{P}\left(\left|\frac{1}{n} \log \|G_n v\| - \lambda_1\right| > \varepsilon\right) \leq C_\varepsilon e^{-c_\varepsilon n},$$

and

$$\mathbb{P}\left(\left|\frac{1}{n} \log \sigma_i(G_n) - \lambda_i\right| > \varepsilon\right) \leq C_\varepsilon e^{-c_\varepsilon n} \quad (i = 1, 2).$$

The same vector-growth estimate is required for the transpose walk. Since the singular values of G_n and G_n^T coincide, no separate singular-value assumption is needed for the transpose.

2.17.1.1 Foundational probability and ergodic tools used by the random-product argument

The random-product proof below needs only a small amount of probability and ergodic theory. To keep those inputs visible, we prove the precise forms used here. As declared in Volume I, we retain as baseline measure-theory background the monotone, Fatou and dominated convergence theorems, Fubini–Tonelli, and the Riesz representation theorem for positive linear functionals on $C(X)$ when X is compact metric.

Lemma 2.51 (First Borel–Cantelli lemma). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $E_n \in \mathcal{F}$. If $\sum_{n \geq 1} \mathbb{P}(E_n) < \infty$, then almost every point belongs to only finitely many E_n .*

Proof. Put $F_N = \bigcup_{n \geq N} E_n$. The union bound gives

$$\mathbb{P}(F_N) \leq \sum_{n \geq N} \mathbb{P}(E_n) \longrightarrow 0.$$

The sets F_N decrease, and their intersection is the limsup set of points belonging to infinitely many E_n . Continuity of probability from above therefore gives measure zero to that limsup. $\square$

Lemma 2.52 (Egorov in the form used below). *Let (X, λ) be a probability space and let $f_n \rightarrow f$ almost everywhere, with all functions measurable and finite almost everywhere. For every $\varepsilon > 0$ there is a measurable $X_\varepsilon \subset X$ with $\lambda(X \setminus X_\varepsilon) < \varepsilon$ on which $f_n \rightarrow f$ uniformly.*

Proof. After discarding the null set where convergence fails, define for $m, N \geq 1$

$$E_{m,N} = \bigcap_{n \geq N} \{x : |f_n(x) - f(x)| < 1/m\}.$$

For fixed m the sets $E_{m,N}$ increase to a full-measure set. Choose N_m so that $\lambda(X \setminus E_{m,N_m}) < \varepsilon 2^{-m}$. Then

$$X_\varepsilon = \bigcap_{m \geq 1} E_{m,N_m}$$

has complement of measure $< \varepsilon$. Given $\eta > 0$, choose m with $1/m < \eta$; for $n \geq N_m$ one has $|f_n - f| < \eta$ everywhere on X_ε . $\square$

Lemma 2.53 (Stationary probability for a Feller operator on a compact metric space). *Let X be compact metric and let $P : C(X) \rightarrow C(X)$ be positive with $P1 = 1$. Then there is a probability ν such that $P^*\nu = \nu$.*

Proof. We first recall why probabilities on a compact metric space are sequentially compact for weak convergence. Since $C(X)$ is separable, choose a countable dense subset (φ_j) of its unit ball. Given probabilities ν_n , repeated subsequence extraction and a diagonal argument make $\int \varphi_j d\nu_n$ converge for every j . Density then defines a positive norm-one linear functional L on all of $C(X)$; the Riesz representation theorem gives a probability ν with $L(\varphi) = \int \varphi d\nu$. Thus the selected subsequence converges weakly to ν .

Fix any probability ν_0 and form the Cesàro averages

$$\xi_N = \frac{1}{N} \sum_{j=0}^{N-1} (P^*)^j \nu_0.$$

Choose a weakly convergent subsequence $\xi_{N_k} \rightarrow \nu$. For every $\varphi \in C(X)$,

$$\int \varphi d(P^* \xi_N - \xi_N) = \frac{\int \varphi d(P^*)^N \nu_0 - \int \varphi d\nu_0}{N},$$

whose absolute value is at most $2\|\varphi\|_\infty/N$. Since $P\varphi \in C(X)$, weak convergence may be passed through P^* , and the limit satisfies $P^*\nu = \nu$. $\square$

Theorem 2.54 (Bounded scalar martingales converge). *Let (M_n) be a real-valued martingale with $|M_n| \leq C$ almost surely for every n . Then M_n converges almost surely and in L^1 .*

Proof. Fix rational $a < b$. Let $U_N(a, b)$ be the number of completed upcrossings of $[a, b]$ by $M_0, \dots, M_N$. Consider the predictable strategy which holds one unit of the process after it has first fallen to $\leq a$, sells after it next reaches $\geq b$, and repeats. If $H_k \in \{0, 1\}$ is the indicator that one unit is held during the increment $M_k - M_{k-1}$, its gain is

$$G_N = \sum_{k=1}^N H_k (M_k - M_{k-1}).$$

Every completed upcrossing contributes at least $b - a$; an unfinished final holding can lose at most $2C$. Hence pathwise

$$G_N \geq (b - a)U_N(a, b) - 2C.$$

Predictability and the martingale property give $\mathbb{E}G_N = 0$, so

$$\mathbb{E}U_N(a, b) \leq \frac{2C}{b - a}.$$

As $N \rightarrow \infty$, monotone convergence shows that $U_\infty(a, b) < \infty$ almost surely. Intersecting over the countably many rational pairs $a < b$, we find that almost every sample path has only finitely many upcrossings of every rational interval. A bounded real sequence which does not converge has $\liminf < \limsup$ and therefore infinitely many upcrossings of some rational interval, a contradiction. Thus M_n converges almost surely. The uniform bound $|M_n| \leq C$ then gives L^1 convergence by dominated convergence. $\square$

Lemma 2.55 (Maximal ergodic inequality). *Let T be an invertible probability-preserving transformation of (X, λ) and $f \in L^1(\lambda)$. With $S_k f = \sum_{j=0}^{k-1} f \circ T^j$, put*

$$E_N = \{x : \max_{1 \leq k \leq N} S_k f(x) > 0\}.$$

Then $\int_{E_N} f d\lambda \geq 0$. Consequently, for $h \in L^1$ and $t > 0$,

$$\lambda \left\{ \sup_{n \geq 1} |A_n h| > t \right\} \leq \frac{2\|h\|_1}{t}, \quad A_n h = \frac{1}{n} S_n h.$$

Proof. Set $H = \max(0, S_1 f, \dots, S_N f)$. On E_N the maximizing partial sum has positive length, so

$$H(x) \leq f(x) + H(Tx).$$

Since $H = 0$ off E_N ,

$$\int_{E_N} f d\lambda \geq \int H d\lambda - \int_{E_N} H \circ T d\lambda \geq \int H d\lambda - \int H \circ T d\lambda = 0.$$

Apply this to $f = h - t$. On $\{\max_{k \leq N} A_k h > t\}$ one has $\max_{k \leq N} S_k(h - t) > 0$, hence the first part gives

$$t \lambda\{\max_{k \leq N} A_k h > t\} \leq \|h\|_1.$$

Apply the same estimate to $-h$, take the union, and then let $N \rightarrow \infty$. $\square$

Theorem 2.56 (Birkhoff theorem for an invertible probability-preserving map). *Let T be invertible and probability preserving and let $f \in L^1(\lambda)$. Then $A_n f$ converges almost everywhere to an invariant $f^* \in L^1$. Moreover, for every invariant measurable set E ,*

$$\int_E f^* d\lambda = \int_E f d\lambda.$$

In particular $\int f^* d\lambda = \int f d\lambda$.

Proof. Let $U : L^2 \rightarrow L^2$ be $Uh = h \circ T$. Since T is invertible and measure preserving, U is unitary. Therefore

$$L^2 = \ker(I - U) \oplus \overline{\text{ran}(I - U)}.$$

Indeed the orthogonal complement of $\text{ran}(I - U)$ is $\ker(I - U^*) = \ker(I - U^{-1}) = \ker(I - U)$. Bounded functions are dense in L^2 , so the set

$$\mathcal{D} = \{g + h - h \circ T : g \in L^2, g \circ T = g, h \in L^\infty\}$$

is dense in L^2 , hence in L^1 on a probability space. For $d = g + h - h \circ T \in \mathcal{D}$,

$$A_n d = g + \frac{h - h \circ T^n}{n} \longrightarrow g$$

pointwise.

We now extend convergence from $\mathcal{D}$ to every L^1 function. Given $f \in L^1$, choose $d_m \in \mathcal{D}$ with $\|f - d_m\|_1 \rightarrow 0$. If

$$\text{osc}(f) = \limsup_n A_n f - \liminf_n A_n f,$$

then, because $A_n d_m$ converges,

$$\text{osc}(f) \leq 2 \sup_n |A_n(f - d_m)|.$$

The maximal inequality gives, for every $\eta > 0$,

$$\lambda\{\text{osc}(f) > \eta\} \leq \frac{4\|f - d_m\|_1}{\eta} \longrightarrow 0.$$

Thus $A_n f$ converges almost everywhere; call the limit f^* .

We record the integrability and integral identity, since they are used below. The family $(A_n f)$ is uniformly integrable. Indeed, choose M so that $v = f 1_{\{|f|>M\}}$ has $\|v\|_1 < \varepsilon$ and put $u = f - v$. If $K > 2M$ and $E_n = \{|A_n f| > K\}$, then $|A_n v| > K/2$ on E_n , whence $\lambda(E_n) \leq 2\varepsilon/K$. Also

$$\int_{E_n} |A_n f| \leq M\lambda(E_n) + \frac{1}{n} \sum_{j=0}^{n-1} \int_{T^{-j}E_n} |v| \leq \frac{2M\varepsilon}{K} + \varepsilon.$$

This is uniform in n . Uniform integrability plus almost-everywhere convergence implies $A_n f \rightarrow f^*$ in L^1 : by Lemma 2.52, convergence is uniform off a set of arbitrarily small measure, and uniform integrability controls that exceptional set (Fatou gives the same small-set control for the limit). Hence

$$\int f^* d\lambda = \lim_n \int A_n f d\lambda = \int f d\lambda.$$

If E is invariant, then $A_n(f 1_E) = 1_E A_n f$, so applying the same argument to $f 1_E$ gives $\int_E f^* = \int_E f$. Finally, invariance of f^* follows from

$$A_n f(Tx) - A_n f(x) = \frac{f(T^n x) - f(x)}{n}.$$

For each $\eta > 0$, $\sum_n \lambda\{|f| > \eta n\} < \infty$ because $\sum_n 1_{\{|f|>\eta n\}} \leq |f|/\eta$ pointwise up to an additive constant. Lemma 2.51 therefore gives $f(T^n x)/n \rightarrow 0$ almost everywhere, proving $f^* \circ T = f^*$. $\square$

2.17.1.2 From projective contraction to the amplified RM-LD package

We begin by proving the BFLM random-matrix large-deviation theorem in the precise form needed here rather than importing it as a primitive input. The proof below follows the transfer-operator architecture of Guivarc’h–Le Page, especially their Theorem 3.17 and Corollaries 3.18–3.20 [6]. The point of spelling out the argument is twofold. First, the exponential vector deviation is a short consequence of a projective contraction estimate and analytic perturbation. Second, in the present arithmetic application the second singular value need not be obtained from an exterior-power theorem: the left–right adjoint amplification has an explicit Cartan formula.

For projective points $x = [v]$, $y = [w]$ represented by unit vectors set

$$d_{\mathbb{P}}(x, y) = \|v \wedge w\|.$$

If $g \in \mathrm{GL}(V)$ and $x = [v]$, put

$$\sigma(g, x) = \log \frac{\|gv\|}{\|v\|}.$$

This is well defined on projective space.

Definition 2.57 (Mean projective contraction input). Let μ be a finitely supported probability measure on $\mathrm{GL}(V)$ whose support generates a strongly irreducible semigroup containing a proximal element. We say that μ satisfies *MC* if there are an integer $N_0 \geq 1$ and $c_0 > 0$ such that

$$\sup_{x \neq y} \int \log \frac{d_{\mathbb{P}}(gx, gy)}{d_{\mathbb{P}}(x, y)} d\mu^{*N_0}(g) \leq -c_0. \quad (2.6)$$

Remark 2.58 (Primary-source architecture, but no remaining black box). Guivarc’h–Le Page Theorem 3.17 identifies the asymptotic mean logarithmic projective contraction with the Lyapunov gap, and their Corollary 3.18 turns that negative mean into a Hölder contraction moment [6]. For the

present arithmetic SL_2 walk we do not leave Theorem 3.17 as an input. The next lemmas reconstruct the rank-one mechanism directly. The proof follows the older Guivarc'h–Raugi architecture: an escaping random product has a Dirac Furstenberg boundary; vector norms then diverge; their recurrence lemma forces strict positivity of the norm cocycle; and a compact occupation-measure argument upgrades pointwise positivity over stationary measures to one uniform finite block [7]. Only after that internal step do we use the Guivarc'h–Le Page transfer-operator machinery, which is itself proved below in the finite-support setting.

Lemma 2.59 (Exponential atom decay for the original arithmetic walk). *Let $S \subset SL_2(\mathbb{Z})$ be a fixed finite symmetric non-elementary set, let μ be the uniform law on S , and put $\Gamma = \langle S \rangle$. There are constants $C < \infty$ and $0 < \theta < 1$ such that*

$$\sup_{\gamma \in \Gamma} \mu^{*n}(\gamma) \leq C\theta^n \quad (n \geq 0).$$

Proof. Volume VII already owns the ping-pong theorem that a fixed non-elementary subgroup of $SL_2(\mathbb{R})$ contains a free pair represented by bounded S -words, and it also owns the elementary norm computation for simple random walk on the 4-regular tree. We recall the short operator deduction because it is the only part of those results used here.

Choose free generators $a, b \in \Gamma$. Replacing them by a^2, b^2 , still a free pair, makes the chosen S -word representatives of $a^{\pm 2}, b^{\pm 2}$ have even length. By inserting cancelling pairs ss^{-1} , with one fixed $s \in S$, pad all four representatives to a common length M . If

$$T = \{a^{\pm 2}, b^{\pm 2}\}, \quad \tau = \frac{1}{4} \sum_{t \in T} \delta_t,$$

then pointwise on Γ

$$\mu^{*M} \geq a_0 \tau, \quad a_0 = 4|S|^{-M} > 0,$$

after increasing M if necessary so that $a_0 \leq 1$. Hence $\mu^{*M} = a_0 \tau + (1 - a_0)\omega$ for a probability measure ω .

On $\ell^2(\Gamma)$ the convolution operator P_τ is an orthogonal direct sum, over right cosets of the free subgroup $\langle a^2, b^2 \rangle$, of simple random walk on the 4-regular tree. The direct tree calculation from Volume VII gives

$$\|P_\tau\|_{2 \rightarrow 2} = \frac{\sqrt{3}}{2} < 1.$$

Young's inequality gives $\|P_\omega\|_{2 \rightarrow 2} \leq 1$. Therefore

$$\|P_\mu^M\|_{2 \rightarrow 2} \leq a_0 \frac{\sqrt{3}}{2} + 1 - a_0 =: q < 1.$$

Since μ is symmetric, P_μ is self-adjoint and $\|P_\mu\|^M = \|P_\mu^M\| \leq q$. Thus $\|P_\mu\| \leq \theta := q^{1/M} < 1$. Finally

$$\mu^{*n}(\gamma) = \langle P_\mu^n \delta_e, \delta_\gamma \rangle \leq \theta^n,$$

which proves the claim (with $C = 1$). □

Lemma 2.60 (Almost-sure escape of the matrix norm). *With the notation of Lemma 2.59, if $G_n = g_n \cdots g_1$ is the μ -walk, then*

$$\|G_n\| \longrightarrow \infty \quad \text{almost surely.}$$

The same is true for G_n^{-1} .

Proof. For a fixed $R \geq 1$, an integral 2×2 matrix of operator norm at most R has all four entries bounded by R . Hence

$$\#\{\gamma \in \Gamma : \|\gamma\| \leq R\} \leq (2[R] + 1)^4 =: C_R.$$

Lemma 2.59 gives

$$\mathbb{P}(\|G_n\| \leq R) \leq C_R \theta^n.$$

The series is summable, so Lemma 2.51 shows that for every integer R the event $\|G_n\| \leq R$ occurs only finitely often. Intersecting over $R \in \mathbb{N}$ proves the first assertion. In SL_2 the two singular values multiply to one, so $\|g^{-1}\| = \|g\|$; this gives the second. $\square$

Lemma 2.61 (The stationary projective measures are non-atomic). *Every μ -stationary probability on $\mathbb{P}^1$ is non-atomic.*

Proof. Stationary probabilities exist by Lemma 2.53. Suppose ν has an atom, let $m > 0$ be the largest atomic mass, and let F be the finite set of atoms of mass m . If $x \in F$, stationarity gives

$$m = \nu\{x\} = \frac{1}{|S|} \sum_{s \in S} \nu\{s^{-1}x\} \leq m.$$

Every summand must therefore equal m , so every generator preserves F . Thus Γ has a finite invariant subset of $\mathbb{P}^1$. This is excluded by non-elementarity (equivalently, by the strong irreducibility already proved for the arithmetic walk). Hence ν is non-atomic. $\square$

Lemma 2.62 (Rank-one martingale boundary). *Let ν be any μ -stationary probability on $\mathbb{P}^1$. There is a random projective point Z^- , with law ν , such that the following holds almost surely. If*

$$H_n = G_n^{-1} = g_1^{-1} \cdots g_n^{-1}$$

and $H_{n_k}/\|H_{n_k}\|$ converges to a rank-one matrix A , then $\text{im } A = Z^-$. Equivalently, all normalized subsequential limits of the inverse products have the same image line.

Proof. This is the two-dimensional specialization of the martingale boundary argument in Guivarc'h–Raugi, Lemmas 2.13–2.14 [7]; we give all steps needed here. Since μ is symmetric, the increments g_i^{-1} are again independent with law μ . The inverse product is a right product, and for every continuous f on $\mathbb{P}^1$,

$$M_n(f) := \int f d((H_n)_* \nu)$$

is a bounded martingale: conditional expectation of the next right increment replaces ν by $\int q_* \nu d\mu(q) = \nu$. Applying Theorem 2.54 to a countable dense family in $C(\mathbb{P}^1)$ shows that $(H_n)_* \nu$ converges weakly almost surely to some random probability ν_∞ .

By Lemma 2.60, $\|H_n\| \rightarrow \infty$. From any subsequence choose a further one with $H_{n_k}/\|H_{n_k}\| \rightarrow A$. Since $\det(H_{n_k}/\|H_{n_k}\|) = \|H_{n_k}\|^{-2} \rightarrow 0$ and $\|A\| = 1$, A has rank one. By Lemma 2.61, $\nu(\mathbb{P}(\ker A)) = 0$. Hence, for ν -almost every x , $H_{n_k}x \rightarrow \mathbb{P}(\text{im } A)$ projectively. Dominated convergence yields

$$(H_{n_k})_* \nu \longrightarrow \delta_{\mathbb{P}(\text{im } A)}.$$

But the left side is a subsequence of the already convergent martingale, so ν_∞ is this Dirac mass. Thus every normalized subsequential limit has the same image, which we call Z^- . Finally $\mathbb{E}M_n(f) = \nu(f)$ for every n ; bounded convergence gives $\mathbb{E}f(Z^-) = \nu(f)$. Therefore Z^- has law ν . $\square$

Lemma 2.63 (Every fixed vector escapes). *For every fixed nonzero $v \in \mathbb{R}^2$,*

$$\|G_n v\| \longrightarrow \infty \quad \text{almost surely.}$$

Proof. Suppose on some event there are $n_k \rightarrow \infty$ for which $\|G_{n_k} v\|$ stays bounded. By Lemma 2.60, after passing to a subsequence

$$\frac{G_{n_k}}{\|G_{n_k}\|} \longrightarrow A$$

with A rank one, and the boundedness forces $Av = 0$. Let $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. For $g \in SL_2(\mathbb{R})$,

$$g^{-1} = J^{-1} g^T J, \quad \|g^{-1}\| = \|g\|.$$

Consequently the corresponding normalized inverse products converge to $J^{-1} A^T J$, whose image line is exactly $\mathbb{P}(\ker A) = \mathbb{R}v$. Lemma 2.62 therefore gives $Z^- = [v]$. The law of Z^- is the non-atomic measure ν , so this event has probability zero. $\square$

Lemma 2.64 (Recurrence sign lemma). *Let (X, λ, T) be an invertible probability-preserving system and let $f \in L^1(\lambda)$. If*

$$S_n f(x) := \sum_{j=0}^{n-1} f(T^j x) \longrightarrow +\infty \quad \text{for almost every } x,$$

then $\int f d\lambda > 0$.

Proof. This is the recurrence mechanism behind Guivarc'h–Raugi Lemma 3.6, with the sign reversed [7]; we spell out the conservative-skew-product step because it is the only ergodic-theoretic point in the argument.

Let g be the almost-everywhere limit of $S_n f/n$ given by Theorem 2.56. Since $S_n f(x) \rightarrow +\infty$, the set $\{g < 0\}$ has measure zero. Thus $g \geq 0$ almost everywhere, while the integral identity in Theorem 2.56 gives $\int f d\lambda = \int g d\lambda \geq 0$. Suppose for contradiction that $\int f d\lambda = 0$. Then $g = 0$ almost everywhere.

Consider the skew product

$$\tilde{T}(x, r) = (Tx, r + f(x))$$

on $X \times \mathbb{R}$, preserving $\lambda \times dr$, and fix the finite measure strip

$$Y_\varepsilon = X \times [-\varepsilon, \varepsilon].$$

We claim that almost every point of Y_ε returns to Y_ε infinitely many times. Otherwise the set of points of Y_ε having a last return would have positive measure. Writing that set as the countable union according to the last-return time, one obtains a positive-measure set $D \subset Y_\varepsilon$ such that

$$\tilde{T}^n D \cap Y_\varepsilon = \emptyset \quad (n \geq 1).$$

Because $D \subset Y_\varepsilon$ and $\tilde{T}$ is invertible, the sets $D, \tilde{T}D, \tilde{T}^2 D, \dots$ are pairwise disjoint.

Now use the zero Birkhoff limit. Given $\delta > 0$, Lemma 2.52 lets us replace D by a positive-measure subset B and choose N so that

$$|S_k f(x)| \leq \delta k \quad (k \geq N, (x, r) \in B).$$

After one further positive-measure restriction, the finitely many initial sums $S_k f$, $0 \leq k < N$, are uniformly bounded in absolute value by some C_N . Therefore, for $n > N$,

$$\bigcup_{0 \leq k < n} \tilde{T}^k B \subset X \times [-\varepsilon - C_N - \delta n, \varepsilon + C_N + \delta n].$$

The left side has measure exactly $n(\lambda \times dr)(B)$ by pairwise disjointness, whereas the right side has measure at most $2\varepsilon + 2C_N + 2\delta n$. Choosing $2\delta < (\lambda \times dr)(B)$ and then letting $n \rightarrow \infty$ is impossible. This proves the recurrence claim.

On the other hand, for almost every (x, r) the hypothesis $S_n f(x) \rightarrow +\infty$ makes the second coordinate $r + S_n f(x)$ eventually larger than ε , so such a point can return to Y_ε only finitely many times. This contradiction rules out $\int f d\lambda = 0$, and hence $\int f d\lambda > 0$. $\square$

Proposition 2.65 (Uniform positive finite-block growth in the arithmetic SL_2 walk). *Let S, μ be as in Lemma 2.59. There are $N_0 \geq 1$ and $c_* > 0$ such that, uniformly in every projective point $x = [v]$ with $\|v\| = 1$,*

$$\int \log \|gv\| d\mu^{*N_0}(g) \geq c_*. \quad (2.7)$$

Proof. Put

$$\phi(x) = \int \sigma(g, x) d\mu(g).$$

Let ν be any stationary probability. On the two-sided Bernoulli space $\Omega = S^{\mathbb{Z}}$, with coordinates $(\omega_j)_{j \in \mathbb{Z}}$, consider

$$T(\omega, x) = (\sigma\omega, \omega_0 x), \quad (\sigma\omega)_j = \omega_{j+1}.$$

The map is invertible because every increment lies in SL_2 , and stationarity of ν shows directly that T preserves $\mu^{\mathbb{Z}} \times \nu$. With $f(\omega, x) = \sigma(\omega_0, x)$, the cocycle identity gives

$$S_n f(\omega, x) = \log \|\omega_{n-1} \cdots \omega_0 v\|.$$

Lemma 2.63 holds for every fixed x ; Fubini therefore gives $S_n f \rightarrow +\infty$ for $\mu^{\mathbb{Z}} \times \nu$ -almost every (ω, x) . The recurrence sign lemma yields

$$\chi(\nu) := \int \phi d\nu = \iint \sigma(g, x) d\mu(g) d\nu(x) > 0.$$

The set $\mathcal{S}$ of stationary probabilities is compact and nonempty, and $\nu \mapsto \chi(\nu)$ is continuous. Thus

$$\kappa := \min_{\nu \in \mathcal{S}} \chi(\nu) > 0.$$

For $n \geq 1$ set

$$F_n(x) = \mathbb{E}_x \log \|G_n v\| = \sum_{j=0}^{n-1} P_0^j \phi(x), \quad m_n = \inf_x F_n(x).$$

Choose x_n with $F_n(x_n) = m_n$ and form the occupation measure

$$\xi_n = \frac{1}{n} \sum_{j=0}^{n-1} (P_0^*)^j \delta_{x_n}.$$

Since

$$P_0^* \xi_n - \xi_n = \frac{(P_0^*)^n \delta_{x_n} - \delta_{x_n}}{n},$$

every weak subsequential limit of ξ_n is stationary. Moreover

$$\frac{m_n}{n} = \int \phi d\xi_n.$$

Compactness therefore gives $\liminf_{n \rightarrow \infty} m_n/n \geq \kappa$. Choose N_0 large enough that $m_{N_0} \geq \kappa N_0/2$, and put $c_* = m_{N_0} > 0$. $\square$

Theorem 2.66 (Mean projective contraction in rank one). *For the fixed finite symmetric non-elementary arithmetic law μ on $SL_2(\mathbb{Z})$, MC holds.*

Proof. For unit representatives v, w of distinct projective points and every $g \in SL_2(\mathbb{R})$,

$$\frac{d_{\mathbb{P}}(g[v], g[w])}{d_{\mathbb{P}}([v], [w])} = \frac{1}{\|gv\| \|gw\|}, \quad (2.8)$$

because $\|gv \wedge gw\| = |\det g| \|v \wedge w\|$ and projective representatives are renormalized to unit length. Integrating the logarithm of (2.8) against μ^{*N_0} and using Proposition 2.65 twice gives

$$\sup_{x \neq y} \int \log \frac{d_{\mathbb{P}}(gx, gy)}{d_{\mathbb{P}}(x, y)} d\mu^{*N_0}(g) \leq -2c_* < 0.$$

This is exactly MC. $\square$

Definition 2.67 (Finite-block Hölder contraction). We say that μ satisfies *FC* if there are $0 < \alpha \leq 1$, an integer $N \geq 1$, and $0 < \rho < 1$ such that

$$\sup_{x \neq y} \int \left(\frac{d_{\mathbb{P}}(gx, gy)}{d_{\mathbb{P}}(x, y)} \right)^{\alpha} d\mu^{*N}(g) \leq \rho. \quad (2.9)$$

Proposition 2.68 (Mean contraction implies FC and stationary uniqueness). *For a finitely supported law, MC implies FC. Moreover the projective Markov operator*

$$P_0 f(x) = \int f(gx) d\mu(g)$$

has a unique stationary probability ν .

Proof. Put

$$A(g; x, y) = \log \frac{d_{\mathbb{P}}(gx, gy)}{d_{\mathbb{P}}(x, y)}.$$

For every invertible g and distinct x, y one has $|A(g; x, y)| \leq 2 \log \gamma(g)$, where $\gamma(g) = \max(\|g\|, \|g^{-1}\|)$. Indeed the projective distance is the sine of the angle and the action of g is bi-Lipschitz with projective Lipschitz constants at most $\gamma(g)^2$ in either direction. Since μ^{*N_0} has finite support, there is therefore a constant $M < \infty$ such that $|A(g; x, y)| \leq M$ uniformly over its support and over $x \neq y$.

For $|a| \leq M$ and $0 < \alpha \leq 1$, Taylor's formula with remainder gives

$$e^{\alpha a} \leq 1 + \alpha a + \frac{\alpha^2 a^2}{2} e^{\alpha M} \leq 1 + \alpha a + C_M \alpha^2.$$

Integrating and using MC yields, uniformly in $x \neq y$,

$$\int e^{\alpha A(g; x, y)} d\mu^{*N_0}(g) \leq 1 - \alpha c_0 + C_M \alpha^2.$$

Choose $\alpha > 0$ so small that $C_M \alpha \leq c_0/2$. Then FC holds with $N = N_0$ and $\rho = 1 - \alpha c_0/2 < 1$. This is the finite-support version of the Taylor argument in Guivarc'h–Le Page Corollary 3.18.

It remains to record that stationary uniqueness is a consequence rather than an additional input. Since $0 < \alpha \leq 1$, the cost $d_\alpha(x, y) = d_{\mathbb{P}}(x, y)^\alpha$ is a metric on the compact projective space. Let W_α be the associated Kantorovich–Wasserstein metric on probability measures. If ξ, ζ are probability measures, choose any coupling π of them and, conditional on (x, y) , apply the same N -step matrix product to both coordinates. FC gives

$$W_\alpha((P_0^*)^N \xi, (P_0^*)^N \zeta) \leq \rho \int d_\alpha(x, y) d\pi(x, y).$$

Taking the infimum over couplings gives

$$W_\alpha((P_0^*)^N \xi, (P_0^*)^N \zeta) \leq \rho W_\alpha(\xi, \zeta). \quad (2.10)$$

Existence of a stationary probability follows from Lemma 2.53. If ν_1, ν_2 are stationary, then (2.10) gives $W_\alpha(\nu_1, \nu_2) \leq \rho W_\alpha(\nu_1, \nu_2)$, hence $\nu_1 = \nu_2$. $\square$

Lemma 2.69 (Direct Hölder spectral gap for the untwisted walk). *Assume MC and let α, N, ρ, ν be supplied by Proposition 2.68. On $C^\alpha(\mathbb{P}(V))$, with*

$$[f]_\alpha = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d_{\mathbb{P}}(x, y)^\alpha}, \quad \|f\|_\alpha = \|f\|_\infty + [f]_\alpha,$$

one has

$$[P_0^N f]_\alpha \leq \rho [f]_\alpha.$$

More importantly, this finite-block contraction already implies, without a quasi-compactness theorem, that there are $C > 0$ and $0 < \vartheta < 1$ such that

$$\|P_0^n f - \nu(f)\|_\alpha \leq C \vartheta^n \|f\|_\alpha. \quad (2.11)$$

Hence 1 is a simple isolated eigenvalue and the rest of the spectrum of P_0 on C^α is contained in a disk of radius < 1 .

Proof. Couple two copies of the projective walk by using the same N -step matrix product. Then

$$|P_0^N f(x) - P_0^N f(y)| \leq [f]_\alpha \int d_{\mathbb{P}}(gx, gy)^\alpha d\mu^{*N}(g),$$

so (2.9) gives $[P_0^N f]_\alpha \leq \rho [f]_\alpha$. Iterating,

$$[P_0^{kN} f]_\alpha \leq \rho^k [f]_\alpha. \quad (2.12)$$

Set $h_k = P_0^{kN} f - \nu(f)$. Stationarity gives $\nu(h_k) = 0$. For every x we may therefore average the difference $h_k(x) - h_k(y)$ against ν :

$$|h_k(x)| = \left| \int (h_k(x) - h_k(y)) d\nu(y) \right| \leq \text{diam}(\mathbb{P}(V))^\alpha [h_k]_\alpha.$$

Together with (2.12) this yields

$$\|P_0^{kN} f - \nu(f)\|_\infty + [P_0^{kN} f - \nu(f)]_\alpha \leq C_0 \rho^k [f]_\alpha.$$

For a general $n = kN + j$, $0 \leq j < N$, the finitely many Markov operators P_0^j are bounded on C^α because the support of μ is finite and each projective map is Lipschitz. Absorbing $\max_{0 \leq j < N} \|P_0^j\|_{C^\alpha \rightarrow C^\alpha}$ into the constant gives (2.11) with, for example, $\vartheta = \rho^{1/N}$ after enlarging C .

If $P_0 f = f$, then (2.11) gives $f = \nu(f)$, so the fixed space is exactly the constants. On the closed codimension-one subspace $\{f : \nu(f) = 0\}$ the powers of P_0 decay exponentially in operator norm; therefore the spectral radius there is at most $\vartheta < 1$. This proves the last assertion and, in particular, isolates the eigenvalue 1 without any separate Doeblin–Fortet or quasi-compactness input. $\square$

Lemma 2.70 (Analytic twisted transfer operators). *Assume MC and finite support. For complex z define*

$$P_z f(x) = \int e^{z\sigma(g,x)} f(gx) d\mu(g).$$

There is $r > 0$ such that $z \mapsto P_z$ is holomorphic as an operator on C^α for $|z| < r$. In this disk there are holomorphic functions $\kappa(z)$, $h_z \in C^\alpha$ and bounded linear functionals ν_z , normalized by $\nu_z(h_z) = 1$, and operators R_z for which

$$P_z^n = \kappa(z)^n (\nu_z \otimes h_z) + R_z^n, \quad \|R_z^n\|_{C^\alpha \rightarrow C^\alpha} \leq C(\vartheta')^n \quad (2.13)$$

with $\vartheta' < 1$, uniformly for $|z| < r$ after shrinking r . Moreover

$$\kappa(0) = 1, \quad \frac{\kappa'(0)}{\kappa(0)} = \iint \sigma(g, x) d\mu(g) d\nu(x) =: \lambda_1.$$

Proof. Because the support is finite and consists of invertible matrices, $x \mapsto \sigma(g, x)$ is uniformly Lipschitz over $g \in \text{supp } \mu$. Thus the power series for $e^{z\sigma(g,x)}$ converges in the C^α operator norm uniformly for $|z|$ in a fixed small disk; hence P_z is holomorphic.

By Lemma 2.69, the circle $|\zeta - 1| = \epsilon$ can be chosen to enclose no point of the spectrum of P_0 except the simple eigenvalue 1. For $|z|$ small the resolvent $(\zeta - P_z)^{-1}$ is given on that circle by the norm-convergent Neumann series

$$(\zeta - P_z)^{-1} = (\zeta - P_0)^{-1} \sum_{j \geq 0} ((P_z - P_0)(\zeta - P_0)^{-1})^j.$$

Define the contour operator

$$\Pi_z = \frac{1}{2\pi i} \int_{|\zeta-1|=\epsilon} (\zeta - P_z)^{-1} d\zeta.$$

The preceding norm-convergent resolvent expansion shows directly that $z \mapsto \Pi_z$ is holomorphic and that $\|\Pi_z - \Pi_0\| \rightarrow 0$. The resolvent identity, inserted into the two contour integrals, gives $\Pi_z^2 = \Pi_z$ and $P_z \Pi_z = \Pi_z P_z$; thus Π_z is the spectral projection associated with the component near 1.

We spell out why its rank cannot change. Shrink r until $\|\Pi_z - \Pi_0\| < 1$. If $u \in \text{ran } \Pi_z$ and $\Pi_0 u = 0$, then

$$\|u\| = \|(\Pi_z - \Pi_0)u\| < \|u\|,$$

so $u = 0$. Hence Π_0 is injective on $\text{ran } \Pi_z$, which has dimension at most one. Interchanging z and 0 gives the reverse inequality, so $\text{ran } \Pi_z$ is one-dimensional. Choose a fixed bounded functional ℓ with $\ell(1) = 1$. After shrinking r once more,

$$h_z = \frac{\Pi_z 1}{\ell(\Pi_z 1)}$$

is holomorphic. Since P_z commutes with Π_z , there is a scalar $\kappa(z)$ with $P_z h_z = \kappa(z) h_z$; applying ℓ shows that κ is holomorphic. The rank-one projector can then be written uniquely as $\Pi_z = \nu_z \otimes h_z$ with $\nu_z(h_z) = 1$, and $z \mapsto \nu_z$ is holomorphic as a bounded-functional-valued map.

It remains to obtain the uniform power bound on the complementary part without appealing to a separate perturbation theorem. Put

$$R_z = P_z(I - \Pi_z).$$

Since $P_z \Pi_z = \kappa(z) \Pi_z$ and Π_z commutes with P_z , we have $P_z = \kappa(z) \Pi_z + R_z$ and $\Pi_z R_z = R_z \Pi_z = 0$; consequently $P_z^n = \kappa(z)^n \Pi_z + R_z^n$. At $z = 0$, equation (2.11) gives $\|R_0^n\| \leq C \vartheta^n$. Choose one block length N_0 so large that $\|R_0^{N_0}\| < 1/4$. Since P_z and Π_z depend continuously on z , so does $R_z^{N_0}$; after shrinking r ,

$$\sup_{|z| < r} \|R_z^{N_0}\| \leq \frac{1}{2}.$$

Also $M_0 := \sup_{|z| < r, 0 \leq j < N_0} \|R_z^j\| < \infty$. Writing $n = kN_0 + j$ therefore gives

$$\|R_z^n\| \leq M_0 2^{-k} \leq C(2^{-1/N_0})^n.$$

Thus (2.13) holds with $\vartheta' = 2^{-1/N_0} < 1$. Notice that this last estimate uses only the already-proved untwisted block contraction and norm-continuity of the twisted operators.

Differentiate $P_z h_z = \kappa(z) h_z$ at $z = 0$, apply the stationary functional $\nu = \nu_0$, and use $h_0 = 1$. Since

$$P'_0 1(x) = \int \sigma(g, x) d\mu(g),$$

we obtain the displayed formula for $\kappa'(0)$. This is the Furstenberg formula for the top exponent; stationarity and Theorem 2.56 identify the same integral with the almost-sure vector-growth exponent. $\square$

Theorem 2.71 (Uniform exponential vector-growth deviations). *Assume MC. For every sufficiently small $\varepsilon > 0$ there are $c_\varepsilon, C_\varepsilon > 0$ such that, for every $n \geq 1$ and every unit vector v ,*

$$\mathbb{P} \left(\left| \frac{1}{n} \log \|G_n v\| - \lambda_1 \right| > \varepsilon \right) \leq C_\varepsilon e^{-c_\varepsilon n}. \quad (2.14)$$

The estimate is uniform in v .

Proof. Let $x = [v]$. The cocycle identity gives

$$\mathbb{E} \exp(z \log \|G_n v\|) = P_z^n 1(x).$$

For real z with $|z| < r$, (2.13) and compactness imply

$$\sup_x P_z^n 1(x) \leq C_z |\kappa(z)|^n + C(\vartheta')^n \leq C'_z e^{n\Lambda(z)},$$

where $\Lambda(z) = \log \kappa(z)$ is real analytic near 0, $\Lambda(0) = 0$, and $\Lambda'(0) = \lambda_1$. Hence for $s > 0$ small,

$$\mathbb{P}(\log \|G_n v\| \geq n(\lambda_1 + \varepsilon)) \leq C'_s \exp n\{\Lambda(s) - s(\lambda_1 + \varepsilon)\}.$$

Since $\Lambda(s) = s\lambda_1 + O(s^2)$, choose s so small that the exponent is at most $-s\varepsilon/2$. For the lower tail use $z = -s$ and Markov's inequality on $e^{-s \log \|G_n v\|}$; analyticity on a two-sided disk gives

$$\Lambda(-s) + s(\lambda_1 - \varepsilon) \leq -s\varepsilon/2.$$

The constants are independent of x , proving (2.14). $\square$

Corollary 2.72 (Top singular-value deviations). *Under the hypotheses of Theorem 2.71,*

$$\mathbb{P} \left(\left| \frac{1}{n} \log \sigma_1(G_n) - \lambda_1 \right| > \varepsilon \right) \leq C_\varepsilon e^{-c_\varepsilon n}.$$

Proof. The lower-tail event for $\sigma_1(G_n) = \|G_n\|$ implies the same lower-tail event for $\|G_n v_0\|$ for any fixed unit v_0 , so (2.14) applies. For the upper tail choose once and for all a finite $1/2$ -net $\mathcal{N}$ of the unit sphere. The elementary net estimate

$$\|A\| \leq 2 \max_{v \in \mathcal{N}} \|Av\|$$

shows that an upper deviation of $\|G_n\|$ forces an upper deviation for one of finitely many $v \in \mathcal{N}$, after replacing ε by $\varepsilon/2$ for large n . A union bound completes the proof; the finitely many small n are absorbed into C_ε . $\square$

Lemma 2.73 (SL_2 Cartan parameter and the adjoint singular values). *Let $g \in SL_2(\mathbb{R})$ and write its Cartan decomposition*

$$g = k_1 \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} k_2, \quad k_1, k_2 \in SO(2), \quad t \geq 0.$$

Choose on $\mathfrak{sl}_2(\mathbb{R})$ a K -invariant Euclidean inner product. Then the singular values of $\text{Ad}(g)$ are

$$e^{2t}, \quad 1, \quad e^{-2t}.$$

In particular $t = \log \sigma_1(g)$.

Proof. The K -invariance makes $\text{Ad}(k_1)$ and $\text{Ad}(k_2)$ orthogonal. For $a_t = \text{diag}(e^t, e^{-t})$, the Cartan involution makes $d \text{Ad}(\log a_t)$ self-adjoint, and on the root-space decomposition $\mathfrak{sl}_2 = \mathbb{R}E \oplus \mathbb{R}H \oplus \mathbb{R}F$ the eigenvalues of $\text{Ad}(a_t)$ are $e^{2t}, 1, e^{-2t}$. Orthogonal multiplication on the left and right does not change singular values. $\square$

Lemma 2.74 (Singular values of a left–right operator). *Let A, B act on a Euclidean space W and give $\text{End}(W)$ the Hilbert–Schmidt norm. For*

$$\mathcal{L}_{A,B}(X) = AXB,$$

the singular values of $\mathcal{L}_{A,B}$ are the pairwise products $\sigma_i(A)\sigma_j(B)$. Consequently, if $A = \text{Ad}(a)$ and $B = \text{Ad}(b)$ with Cartan parameters $u, v \geq 0$, then

$$\log \sigma_1(\mathcal{L}_{A,B}) = 2(u + v), \quad \log \sigma_2(\mathcal{L}_{A,B}) = 2 \max(u, v).$$

Proof. Under vectorization, $X \mapsto AXB$ is represented by $B^T \otimes A$. If $A = U_A \Sigma_A V_A^T$ and $B = U_B \Sigma_B V_B^T$ are singular-value decompositions, then

$$B^T \otimes A = (V_B \otimes U_A)(\Sigma_B \otimes \Sigma_A)(U_B^T \otimes V_A^T),$$

whose outer factors are orthogonal and whose middle factor has diagonal entries $\sigma_j(B)\sigma_i(A)$. Insert the three adjoint singular values from Lemma 2.73; the largest exponent is $2u + 2v$ and the next largest is $\max(2u, 2v)$. $\square$

Lemma 2.75 (The standard arithmetic walk satisfies RM-LD). *Let μ be the uniform law on the fixed finite symmetric non-elementary set $S \subset SL_2(\mathbb{Z})$. Then the standard two-dimensional walk satisfies RM-LD with exponents*

$$\lambda_1 = \chi > 0, \quad \lambda_2 = -\chi.$$

The transpose walk satisfies the same uniform vector-growth estimate.

Proof. Theorem 2.66 supplies MC. Hence Theorem 2.71 and Corollary 2.72 give uniform vector and top-singular-value deviations around the common exponent χ . Proposition 2.65, or equivalently the Furstenberg formula in Lemma 2.70, gives $\chi > 0$. Since $\det G_n = 1$,

$$\sigma_2(G_n) = \sigma_1(G_n)^{-1},$$

so the second singular-value estimate is automatic with center $-\chi$.

For the transpose vector estimate put $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. The identity

$$g^T = Jg^{-1}J^{-1} \quad (g \in SL_2(\mathbb{R}))$$

shows that the one-step transpose law is an orthogonal conjugate of the inverse law. Because μ is symmetric, it is therefore an orthogonal conjugate of μ itself. Reversal of the order of independent increments does not change the distribution, so the same uniform vector deviation holds for the transpose products. $\square$

Lemma 2.76 (The nilpotent conic realizes the top adjoint singular directions). *Let*

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad q(u) = u(Ju)^T \in \mathfrak{sl}_2(\mathbb{R}).$$

Equip $W = \mathfrak{sl}_2(\mathbb{R})$ with the Hilbert–Schmidt norm. Then, for unit u ,

$$\|q(u)\|_{HS} = 1, \quad q(gu) = \text{Ad}(g)q(u) \quad (g \in SL_2(\mathbb{R})).$$

If $g = k_1 \text{diag}(e^t, e^{-t})k_2$ is a Cartan decomposition, a top right singular line of $\text{Ad}(g)$ is $\mathbb{R}q(k_2^{-1}e_1)$ and a top left singular line is $\mathbb{R}q(k_1e_1)$.

Proof. Since $(Ju)^T u = 0$, the matrix $q(u)$ has trace zero, and its Hilbert–Schmidt norm is $\|u\| \|Ju\| = 1$. The symplectic identity $g^T Jg = J$ gives $Jg = g^{-T}J$, hence

$$q(gu) = gu(Jgu)^T = g u(Ju)^T g^{-1} = \text{Ad}(g)q(u).$$

For $e_1 = (1, 0)^T$, $q(e_1) = E_{12}$, and

$$\text{Ad}(\text{diag}(e^t, e^{-t}))E_{12} = e^{2t}E_{12}.$$

Conjugation by $SO(2)$ is orthogonal for the Hilbert–Schmidt norm. The stated right and left singular lines follow by transporting E_{12} through the two orthogonal Cartan factors. $\square$

Lemma 2.77 (Uniform quadratic sublevel estimate on the conic). *There are absolute constants $C, r_0 > 0$ such that for every nonzero linear functional $\ell \in W^*$, every $0 < r < r_0$, and the conic $\mathcal{C} = q(\mathbb{P}^1)$,*

$$\{[u] \in \mathbb{P}^1 : |\ell(q(u))| < r\|\ell\|\}$$

is contained in the union of at most four projective intervals, each of radius at most $C\sqrt{r}$.

Proof. In the coordinates $u = (x, y)$,

$$q(u) = \begin{pmatrix} -xy & x^2 \\ -y^2 & xy \end{pmatrix}.$$

Thus $\ell(q(u))$ is a homogeneous quadratic form $ax^2 + bxy + cy^2$. The linear map $W^* \rightarrow \text{Sym}^2((\mathbb{R}^2)^*)$ obtained in this way is an isomorphism, so after normalizing $\|\ell\| = 1$ the coefficient vector (a, b, c) stays in a compact annulus.

On the unit circle write the quadratic as

$$p(\theta) = A + B \cos 2\theta + C \sin 2\theta.$$

The normalized coefficient triple (A, B, C) again lies in a fixed compact annulus. The circle can be divided by the zeros of p' into at most four intervals on which p is monotone. On such an interval, if $|p'| \geq \sqrt{r}$ throughout the part where $|p| < r$, that part has length at most $2\sqrt{r}$. Otherwise at some point both $|p| < r$ and $|p'| < \sqrt{r}$. Since $p'' = -4(B \cos 2\theta + C \sin 2\theta)$ and the normalized coefficient triple cannot have p, p', p'' all arbitrarily small at one point, compactness gives $|p''| \geq c > 0$ in a fixed neighborhood whenever r is sufficiently small. Taylor's theorem then confines $|p| < r$ to an interval of length $O(\sqrt{r})$. The case $B = C = 0$ is harmless: then $|A|$ is uniformly bounded below and the sublevel set is empty for small r . Summing over at most four monotonicity intervals proves the claim. Passing from the circle to $\mathbb{P}^1$ only identifies antipodal points and changes the absolute constant. $\square$

2.17.1.3 Projective avoidance from singular-value large deviations

We now separate the genuinely probabilistic random-matrix input from its projective consequences. This distinction matters for the logical dependency and follows the original BFLM Section 4 architecture [2]. The projective estimates used later are not an additional theorem: once exponential large deviations for vector growth and the first two singular values are known, they follow by elementary singular-value geometry and the same estimates for the transpose walk.

Remark 2.78 (Why the projective estimates are internal). For the left–right adjoint walk used below, this definition is an interface, not an external theorem. Proposition 2.85 proves all of its clauses internally from the rank-one arithmetic walk; the present subsection then derives the BFLM projective consequences from those clauses.

Lemma 2.79 (Singular geometry). *Let $g \in \text{GL}_m(\mathbb{R})$ have singular values $\sigma_1(g) \geq \sigma_2(g) \geq \dots$. Choose unit right and left top singular vectors v_g, u_g , so $gv_g = \sigma_1(g)u_g$, and put*

$$H_g = v_g^\perp, \quad \theta_g = \mathbb{R}u_g, \quad \rho(g) = \frac{\sigma_2(g)}{\sigma_1(g)}.$$

For every unit vector x with $a = |\langle x, v_g \rangle| > 0$,

$$\|gx\| \geq a\sigma_1(g), \quad d_{\mathbb{P}}(gx, \theta_g) \leq \frac{\rho(g)}{a},$$

where $d_{\mathbb{P}}$ is the sine of the projective angle. Moreover, for unit x, y ,

$$|\langle gx, y \rangle| = |\langle x, g^T y \rangle|.$$

Proof. Write $x = av_g + w$ with $w \perp v_g$. Orthogonality of the left singular vectors gives

$$gx = a\sigma_1 u_g + gw, \quad \|gw\| \leq \sigma_2 \|w\| \leq \sigma_2.$$

The first inequality follows by orthogonality of the two summands. Dividing the norm of the component orthogonal to u_g by the total norm gives the projective estimate. The last identity is the defining property of transpose. $\square$

Theorem 2.80 (Exponential attraction and hyperplane avoidance). *Assume RM-LD. Put $\gamma = \lambda_1 - \lambda_2 > 0$. There exist $c_0, c_2 > 0$ such that, for every prescribed $c_1 > 0$ after possibly decreasing c_0 , all sufficiently large n satisfy*

(a) *uniformly in $\bar{x} \in \mathbb{P}^{m-1}$,*

$$\mathbb{P}(d_{\mathbb{P}}(G_n \bar{x}, \theta_{G_n}) > e^{-c_2 n}) \leq e^{-c_0 n};$$

(b) *uniformly in $\bar{x}, \bar{y} \in \mathbb{P}^{m-1}$,*

$$\mathbb{P}(d_{\mathbb{P}}(G_n \bar{x}, \bar{y}^\perp) < e^{-c_1 n}) \leq e^{-c_0 n}.$$

Proof. We follow the singular-value proof directly. Choose $0 < \varepsilon < \gamma/12$. RM-LD gives, outside an event of probability $O(e^{-cn})$,

$$e^{(\lambda_1 - \varepsilon)n} \leq \|G_n x\| \quad \text{for the prescribed unit } x,$$

$$\sigma_1(G_n) \leq e^{(\lambda_1 + \varepsilon)n}, \quad \sigma_2(G_n) \leq e^{(\lambda_2 + \varepsilon)n}.$$

Thus $\rho(G_n) = \sigma_2/\sigma_1 \leq e^{-(\gamma - 2\varepsilon)n}$. By Lemma 2.79, either

$$d_{\mathbb{P}}(G_n \bar{x}, \theta_{G_n}) \leq e^{-\gamma n/2}$$

or

$$d_{\mathbb{P}}(\bar{x}, H_{G_n}) \leq e^{-\gamma n/3}.$$

Indeed, otherwise the bound $\rho(G_n)/d_{\mathbb{P}}(\bar{x}, H_{G_n})$ from Lemma 2.79 is at most $e^{-\gamma n/2}$. In the second alternative, the upper singular estimate in the same lemma gives

$$\|G_n x\| \leq \sigma_1(G_n) e^{-\gamma n/3} + \sigma_2(G_n) \leq 2e^{(\lambda_1 - \gamma/4)n}$$

for n large and our choice of ε . Uniform vector-growth LD says that this event has probability $O(e^{-cn})$. Hence (a) holds, for example with $c_2 = \gamma/2$ after a harmless decrease.

For (b), use a unit representative y of $\bar{y}$. On the event in (a),

$$d_{\mathbb{P}}(G_n \bar{x}, \bar{y}^\perp) \geq d_{\mathbb{P}}(\theta_{G_n}, \bar{y}^\perp) - e^{-c_2 n}.$$

Write a singular-value decomposition $G_n = U\Sigma V$. Since $\theta_{G_n} = \mathbb{R}Ue_1$,

$$d_{\mathbb{P}}(\theta_{G_n}, \bar{y}^\perp) = |\langle Ue_1, y \rangle|.$$

Now

$$\|G_n^T y\| = \|\Sigma U^T y\| \leq \sigma_1(G_n) |\langle Ue_1, y \rangle| + \sigma_2(G_n),$$

so

$$d_{\mathbb{P}}(\theta_{G_n}, \bar{y}^\perp) \geq \frac{\|G_n^T y\|}{\sigma_1(G_n)} - \frac{\sigma_2(G_n)}{\sigma_1(G_n)}.$$

Apply RM-LD to the transpose walk and to σ_1, σ_2 . Outside an $O(e^{-cn})$ event,

$$\frac{\|G_n^T y\|}{\sigma_1(G_n)} \geq e^{-2\varepsilon n}, \quad \frac{\sigma_2(G_n)}{\sigma_1(G_n)} \leq e^{-(\gamma-2\varepsilon)n}.$$

Given the prescribed $c_1 > 0$, choose first $\varepsilon < c_1/4$ and, if necessary, replace c_1 by the smaller value $\min(c_1, c_2/2, \gamma/4)$; proving avoidance at this smaller exponential scale implies the asserted estimate after adjusting constants. For n large the last three displays give

$$d_{\mathbb{P}}(G_n \bar{x}, \bar{y}^\perp) > e^{-c_1 n}.$$

All estimates are uniform in x, y , exactly because RM-LD is uniform in the vector. Summing the exponentially small exceptional probabilities proves (b). $\square$

Corollary 2.81 (Scale-wise projective Frostman bounds). *Assume RM-LD. There exist $\tau, c, C, r_0 > 0$ such that, for every large n and every $e^{-cn} \leq r \leq r_0$,*

$$\sup_{\bar{x}, V} \mathbb{P}\{d_{\mathbb{P}}(G_n \bar{x}, V) < r\} \leq Cr^\tau,$$

and

$$\sup_V \mathbb{P}\{d_{\mathbb{P}}(\theta_{G_n}, V) < r\} \leq Cr^\tau,$$

where V ranges over projective hyperplanes.

Proof. Choose $k = \lfloor a \log(1/r) \rfloor$ with $a > 0$ small enough that $e^{-c_2 k} \leq r/4$ and $k \leq n/2$. Write $G_n = BA$ with B the last k steps. Conditional on A , Theorem 2.80(b) for the independent block B gives probability $O(e^{-c_0 k}) = O(r^\tau)$ that $B(A\bar{x})$ lies within r of a prescribed hyperplane, after changing constants. This proves the first bound.

For the attracting direction, Theorem 2.80(a) puts $G_n \bar{x}$ within $e^{-c_1 n}$ of θ_{G_n} for any fixed $\bar{x}$, outside $e^{-c_0 n}$. If θ_{G_n} lies within r of V , then, except on that exceptional event, $G_n \bar{x}$ lies within $2r$ of V . Apply the first bound. The lower restriction $r \geq e^{-cn}$ absorbs the exponential error. $\square$

Corollary 2.82 (General-position tuples). *Assume RM-LD. For some $\tau, c, C > 0$, if $G_n^{(1)}, \dots, G_n^{(m)}$ are independent n -step products, then for every $e^{-cn} \leq r \leq r_0$,*

$$\mathbb{P}(|\det(u_1, \dots, u_m)| < r) \leq Cr^\tau,$$

where u_j is any unit representative of $\theta_{G_n^{(j)}}$. The same conclusion holds when any prescribed subcollection of the attracting lines is replaced by images $G_n^{(j)} \bar{x}_j$ of fixed projective points.

Proof. Choose the vectors successively. Conditional on the first $j-1$, their span is contained in a projective hyperplane V_j . The event that the j th vector lies within distance ρ of V_j has probability at most $C\rho^\tau$ by Corollary 2.81. If every such distance is at least ρ , Gram-Schmidt gives determinant at least ρ^{m-1} . Take $\rho = r^{1/(m-1)}$ and sum the exceptional probabilities. The image-vector variant uses the first Frostman estimate instead of the attracting-line estimate. $\square$

Corollary 2.83 (The projective part of the BFLM random-matrix package is derived). *Under RM-LD, the projective hyperplane/direction nonconcentration and quantitative general-position estimates denoted (RM2) in the BFLM bootstrap hold, after decreasing the exponents.*

Proof. Combine Corollaries 2.81 and 2.82. $\square$

2.17.1.4 Projective anti-concentration and the left–right amplification

The rank-one transfer theorem has already been proved above. We now combine it with the projective consequences just established to control singular directions, and only then pass to the nine-dimensional left–right adjoint amplification.

Lemma 2.84 (Conic anti-concentration for standard singular directions). *There exist $c, \tau, C > 0$ with the following property. Let A_k be a k -step standard μ -product. Let v_{A_k} be a top right singular unit vector of $\text{Ad}(A_k)$ and u_{A_k} a top left singular unit vector. For every unit functional $\ell \in W^*$ and every $e^{-ck} \leq r \leq r_0$,*

$$\mathbb{P}(|\ell(v_{A_k})| < r) + \mathbb{P}(|\ell(u_{A_k})| < r) \leq Cr^\tau.$$

Proof. Lemma 2.75 gives RM–LD for the standard walk and for its transpose. Corollary 2.81, in dimension two, therefore says that both the top left singular direction of A_k and the top right singular direction (the top left direction for A_k^T) give probability at most $C_0 s^{\tau_0}$ to any projective interval of radius s , provided $s \geq e^{-c_0 k}$.

By Lemma 2.76, the corresponding adjoint singular directions are the conic lifts $q([u])$. Lemma 2.77 covers $|\ell(q(u))| < r$ by $O(1)$ intervals of radius $C\sqrt{r}$. Choosing $c > 0$ small enough that $C\sqrt{r} \geq e^{-c_0 k}$ throughout the stated range gives

$$\mathbb{P}(|\ell(q(u))| < r) \ll r^{\tau_0/2}.$$

Rename $\tau_0/2$ as τ . □

Proposition 2.85 (The amplified adjoint walk satisfies RM–LD unconditionally). *Let $S \subset SL_2(\mathbb{Z})$ be the fixed finite symmetric non-elementary set used in Proposition 2.176, let ν be its uniform law, and let $\rho = \text{Ad}$ on $W = \mathfrak{sl}_2(\mathbb{R})$. On $\text{End}(W) \cong \mathbb{R}^9$ let η_0 choose, with equal total probabilities, a left multiplication $L_{\rho(s)}$ or a right multiplication $R_{\rho(s)}$, with s distributed according to ν . Then η_0 satisfies RM–LD, and*

$$\lambda_1^{(9)} = 2\chi, \quad \lambda_2^{(9)} = \chi > 0, \quad (2.15)$$

where χ is the top exponent of the standard SL_2 walk. The required uniform vector estimate also holds for the transpose amplified walk.

Proof. Let K_n be the number of left multiplications among the first n amplified increments. Then $K_n \sim \text{Bin}(n, 1/2)$ and, for every $\delta > 0$,

$$\mathbb{P}(|K_n - n/2| > \delta n) \leq 2e^{-2\delta^2 n}. \quad (2.16)$$

Conditional on $K_n = k$ and on the left/right pattern, the amplified product is

$$\mathcal{T}_n(X) = \text{Ad}(A_k)X \text{Ad}(B_{n-k}),$$

where A_k and B_{n-k} are independent standard products. Put $t(A) = \log \sigma_1(A)$. Lemma 2.75 and Corollary 2.72 give, uniformly for k, j comparable with n ,

$$\mathbb{P}(|t(A_k) - \chi k| > \varepsilon k) \ll e^{-ck}, \quad \mathbb{P}(|t(B_j) - \chi j| > \varepsilon j) \ll e^{-cj}.$$

Lemmas 2.73 and 2.74 therefore imply, outside an exponentially small event,

$$\frac{1}{n} \log \sigma_1(\mathcal{T}_n) = 2\chi + O(\varepsilon),$$

and

$$\frac{1}{n} \log \sigma_2(\mathcal{T}_n) = \chi + O(\varepsilon + |K_n/n - 1/2|).$$

Together with (2.16) these are the required first and second singular-value deviations with centers 2χ and χ .

It remains to prove the uniform vector estimate; this is the point where an unnecessary amplified-projective MC hypothesis was used in the previous draft. Normalize $\|X\|_{HS} = 1$. Write v_A for a top right singular unit vector of $\text{Ad}(A_k)$ and u_B for a top left singular unit vector of $\text{Ad}(B_{n-k})$. The singular-value decomposition gives

$$\|\mathcal{T}_n X\|_{HS} \geq e^{2t(A_k) + 2t(B_{n-k})} |\langle v_A, X u_B \rangle|. \quad (2.17)$$

Because $\|X\|_{HS} = 1$, there is a unit $z \in W$ with $\|X^T z\| \geq 1/\sqrt{3}$. Hence, by Lemma 2.84, for $\delta > 0$ sufficiently small and $k \geq n/3$,

$$\mathbb{P}(\|X u_B\| < e^{-\delta n}) \ll e^{-c_1 n}.$$

Indeed the event implies $|\langle X^T z, u_B \rangle| < e^{-\delta n}$, and after normalizing the functional this is a conic sublevel event.

Condition on u_B outside that exceptional event and put $w = X u_B / \|X u_B\|$. Independence of the left and right standard products and a second application of Lemma 2.84 give

$$\mathbb{P}(|\langle v_A, X u_B \rangle| < e^{-3\delta n} \mid u_B) \leq \mathbb{P}(|\langle v_A, w \rangle| < e^{-2\delta n}) \ll e^{-c_2 n}.$$

Thus, uniformly in unit X ,

$$\mathbb{P}(|\langle v_A, X u_B \rangle| < e^{-3\delta n}) \ll e^{-c_3 n} \quad (2.18)$$

on the binomially typical event $K_n \in [n/3, 2n/3]$. Combining (2.17), (2.18), and the two standard Cartan deviations gives

$$\mathbb{P}\left(\frac{1}{n} \log \|\mathcal{T}_n X\| < 2\chi - \epsilon\right) \ll e^{-c_\epsilon n}$$

for every prescribed $\epsilon > 0$, after choosing the internal δ, ϵ small enough. The upper tail follows from $\|\mathcal{T}_n X\| \leq \sigma_1(\mathcal{T}_n)$ and the already proved σ_1 deviation. This proves uniform vector RM-LD centered at 2χ .

The transpose of $X \mapsto \text{Ad}(A)X \text{Ad}(B)$ for the Hilbert–Schmidt inner product is $X \mapsto \text{Ad}(A)^T X \text{Ad}(B)^T$. The same proof applies with top left and top right singular directions interchanged. Lemma 2.75 supplies the needed standard transpose projective estimates, while the singular values are unchanged. Hence the transpose amplified walk has the same uniform vector deviation. $\square$

Corollary 2.86 (The BFLM random-matrix engine is fully internal). *For the exact left–right amplified walk used in Proposition 2.176, every clause of RM-LD in Definition 2.50 is proved internally. No random-matrix large-deviation theorem, exterior-power theorem, or separate projective-contraction theorem remains as an external input.*

Proof. The standard rank-one contraction is Theorem 2.66; the finite-support transfer-operator argument gives the standard vector and singular-value deviations; and Proposition 2.85 proves the complete nine-dimensional amplified and transpose-amplified package. These are exactly the random-matrix clauses consumed by the projective-avoidance and BFLM bootstrap modules. $\square$

2.17.1.5 A local scalar sum–product engine

The projection argument below needs a one-dimensional growth theorem before the planar line-projection base case and its subsequent dimension-induction can be used. The series already owned such a proof in the later compact-Lie-group volume. To preserve backward-only dependency, we reproduce that proof here with independent labels. The proof is reproduced locally, so no forward reference to Volume XIV is needed. The scalar-action reduction below is checked against the energy argument in the primary matrix-algebra sum–product paper [4].

The scalar theorem

Theorem 2.87 (Internal discretized sum–product theorem). *Fix $0 < \sigma < 1$ and $C_0 \geq 1$. Let $A \subset [1, 2]$ be δ -separated, with*

$$|A| \asymp \delta^{-\sigma},$$

and assume the Frostman-type non-concentration estimate

$$|A \cap I| \leq C_0 |I|^\sigma |A| \tag{SP.19}$$

for every interval I of length at least δ . Then for every

$$0 < c < c_0(\sigma) := \frac{\sigma(1-\sigma)}{2(19+6\sigma)}$$

there is $\delta_0 = \delta_0(\sigma, c, C_0) > 0$ such that, for $0 < \delta < \delta_0$,

$$\mathcal{N}_\delta(A + A) + \mathcal{N}_\delta(A \cdot A) \gg_{\sigma, c} C_0^{-O_\sigma(1)} \delta^{-c} |A|. \tag{SP.20}$$

The proof follows a robust dense-versus-gap argument. We give it because later Volume XIV needs to know exactly where scale regularity, quotient separation and the final power gain enter.

Covering-number calculus

Lemma 2.88 (Discretization comparison). *For bounded $X, Y \subset \mathbb{R}$ and fixed $m \geq 1$,*

$$|X_\delta| \asymp \mathcal{N}_\delta(X), \quad \mathcal{N}_\delta(X + Y) \asymp |X_\delta + Y_\delta|,$$

with constants depending only on the number of summands. The same holds for differences. If $X, Y \subset [1, 2]$, then

$$\mathcal{N}_\delta(XY) \asymp |(XY)_\delta|$$

with absolute constants.

Proof. A maximal δ -separated subset of X has cardinality comparable to the least number of δ -intervals covering X . Every point of X_δ lies within δ of X , while a maximal separated subset maps to distinct points of a fixed constant multiple of the δ -grid; this gives the first comparison.

For sums, if $x = x_0 + O(\delta)$ and $y = y_0 + O(\delta)$, then $x + y = x_0 + y_0 + O(\delta)$, so replacing each input by its grid discretization changes the sumset only by a bounded enlargement of the covering intervals. The reverse inclusion follows by approximating each original summand by a grid point. Products on $[1, 2]$ are uniformly Lipschitz in each variable, so the same argument applies to xy . $\square$

We use one standard additive-combinatorial tool. For completeness, we record the exact form used below.

Lemma 2.89 (Petridis minimal-growth lemma). *Let X, B be finite nonempty subsets of an abelian group and assume*

$$K_0 := \frac{|X+B|}{|X|} \leq \frac{|Z+B|}{|Z|} \quad \text{for every nonempty } Z \subset X.$$

Then for every finite set C ,

$$|X+B+C| \leq K_0|X+C|. \quad (\text{SP.21})$$

Proof. Write $C = \{c_1, \dots, c_r\}$ in an arbitrary order. Put $X_1 = X$, and for $i > 1$ define

$$X_i := \{x \in X : c_i + x \notin \{c_1, \dots, c_{i-1}\} + X\}.$$

Then the sets $c_i + X_i$ are pairwise disjoint and

$$C + X = \bigsqcup_{i=1}^r (c_i + X_i), \quad |C + X| = \sum_{i=1}^r |X_i|.$$

We prove (SP.21) by induction on r . The case $r = 1$ is the definition of K_0 . For $r > 1$ set $C' = C \setminus \{c_r\}$ and $X_r^c = X \setminus X_r$. By the definition of X_r ,

$$c_r + X_r^c \subset C' + X, \quad c_r + X_r^c + B \subset C' + X + B.$$

Hence the points contributed for the first time by the last translate satisfy

$$|C + X + B| \leq |C' + X + B| + |X + B| - |X_r^c + B|.$$

The induction hypothesis bounds the first term by $K_0 \sum_{i < r} |X_i|$. If X_r^c is empty, the corresponding lower bound is trivial; otherwise minimality gives $|X_r^c + B| \geq K_0 |X_r^c|$. Also $|X + B| = K_0 |X|$; therefore the last two terms contribute at most $K_0 |X_r|$. Summing proves (SP.21). $\square$

Corollary 2.90 (One-summand Plünnecke inequality). *Let A, B be finite nonempty subsets of an abelian group and suppose*

$$|A + B| \leq K|A|.$$

Then there is a nonempty $X \subset A$ such that, for every integer $h \geq 1$,

$$|X + hB| \leq K^h |X|.$$

Proof. Choose nonempty $X \subset A$ minimizing $|Z + B|/|Z|$ over nonempty $Z \subset A$, and write the minimum as $K_0 \leq K$. Lemma 2.89, with $C = (h-1)B$, gives inductively

$$|X + hB| \leq K_0 |X + (h-1)B| \leq K_0^h |X| \leq K^h |X|.$$

$\square$

Proposition 2.91 (Fixed-summand Plünnecke refinement). *For every fixed $k \geq 1$ there is a constant C_k with the following property. Let $X, Y_1, \dots, Y_k$ be finite subsets of an abelian group and suppose*

$$|X + Y_i| \leq K_i |X| \quad (1 \leq i \leq k),$$

where $K_i \geq 1$. Then there is a nonempty $X' \subset X$ such that

$$|X' + Y_1 + \dots + Y_k| \leq C_k \left(\prod_{i=1}^k K_i \right) |X'|. \quad (\text{SP.22})$$

Moreover one can choose a union $X^\sharp \subset X$ with $|X^\sharp| \geq |X|/2$ and

$$|X^\sharp + Y_1 + \dots + Y_k| \leq C'_k \left(\prod_{i=1}^k K_i \right) |X^\sharp|. \quad (\text{SP.23})$$

Proof. The point is that the exact constant is irrelevant for the present application, so the different-summands theorem can be reduced to the one-summand result without a tensor-power limit. Put $L_i = \lceil K_i \rceil$, $P = \prod_i L_i$, and $M_i = P/L_i$. Let H_i be a cyclic group of order M_i (the trivial group when $M_i = 1$), and work in

$$\widehat{G} = G \times H_1 \times \cdots \times H_k.$$

Embed X as $\widehat{X} = X \times \{0\}^k$. For each i , let $\widehat{Y}_i$ consist of the points whose G -coordinate lies in Y_i , whose H_i -coordinate is arbitrary, and whose other auxiliary coordinates are zero. If $\widehat{B} = \bigcup_i \widehat{Y}_i$, then

$$|\widehat{X} + \widehat{B}| \leq \sum_i M_i |X + Y_i| \leq \sum_i M_i L_i |X| = kP |\widehat{X}|.$$

Corollary 2.90, applied to $\widehat{X}, \widehat{B}$ with $h = k$, gives a nonempty $\widehat{X}' = X' \times \{0\}^k$ such that

$$|\widehat{X}' + k\widehat{B}| \leq (kP)^k |\widehat{X}'|.$$

But

$$(X' + Y_1 + \cdots + Y_k) \times H_1 \times \cdots \times H_k \subset \widehat{X}' + k\widehat{B}.$$

Since $\prod_i M_i = P^{k-1}$,

$$|X' + Y_1 + \cdots + Y_k| \leq k^k P |X'| \leq 2^k k^k \left(\prod_i K_i \right) |X'|.$$

Thus (SP.22) holds, for instance with $C_k = 2^k k^k$.

For the large refinement, repeatedly apply the first assertion to the unused part $R \subset X$ while $|R| \geq |X|/2$. On such an R ,

$$|R + Y_i| \leq |X + Y_i| \leq 2K_i |R|.$$

Every extracted piece therefore satisfies (SP.22) with $2K_i$ in place of K_i . Stop once the extracted pieces cover at least half of X , and let $X^\#$ be their union. The image of a union is contained in the union of the images, so summing the piecewise estimates proves (SP.23), with another constant depending only on k . $\square$

Proposition 2.92 (Ruzsa triangle inequality). *For finite nonempty X, Y, Z in an abelian group,*

$$|X - Z| \leq \frac{|X - Y| |Y - Z|}{|Y|}.$$

Consequently, for bounded subsets of $\mathbb{R}$,

$$\mathcal{N}_\delta(X - Z) \ll \frac{\mathcal{N}_\delta(X - Y) \mathcal{N}_\delta(Y - Z)}{\mathcal{N}_\delta(Y)}.$$

Proof. For each $s \in X - Z$, fix one representation $s = x_s - z_s$. The map

$$(s, y) \mapsto (x_s - y, y - z_s)$$

from $(X - Z) \times Y$ into $(X - Y) \times (Y - Z)$ is injective, since the two output coordinates add to s and then determine y . This proves the finite inequality. The covering version follows from Lemma 2.88 after replacing each set by its δ -grid discretization. $\square$

Lemma 2.93 (Ruzsa covering lemma at scale δ). *Let $X, Y \subset \mathbb{R}$ be bounded and nonempty. If*

$$\mathcal{N}_\delta(X + Y) \leq K\mathcal{N}_\delta(X), \quad K \geq 1,$$

then there is a finite set $T \subset Y$ with $|T| \ll K$ such that

$$Y \subset (X - X) + T + [-C\delta, C\delta] \quad (\text{SP.24})$$

for an absolute constant C .

Proof. Choose grid discretizations X_δ, Y_δ as in Lemma 2.88. After changing K by an absolute factor,

$$|X_\delta + Y_\delta| \leq CK|X_\delta|.$$

We first use the finite Ruzsa covering argument. Choose a maximal subset $T_\delta \subset Y_\delta$ for which the translates $X_\delta + t$, $t \in T_\delta$, are pairwise disjoint. Their union is contained in $X_\delta + Y_\delta$, so

$$|T_\delta||X_\delta| \leq |X_\delta + Y_\delta| \leq CK|X_\delta|,$$

and hence $|T_\delta| \leq CK$. Maximality implies that for every $y \in Y_\delta$ there is $t \in T_\delta$ with $(X_\delta + y) \cap (X_\delta + t) \neq \emptyset$; therefore $y \in X_\delta - X_\delta + t$.

For each $t \in T_\delta$ choose a point of Y at distance $O(\delta)$ from t , and call the resulting set $T \subset Y$. Likewise every grid point of X_δ is within $O(\delta)$ of X , and every point of Y is within $O(\delta)$ of Y_δ . Thickening the preceding finite inclusion by these three rounding errors gives (SP.24). $\square$

Corollary 2.94 (Covering Plünnecke). *If*

$$\mathcal{N}_\delta(X + Y_i) \leq K_i \mathcal{N}_\delta(X),$$

then there is a δ -refinement $X' \subset X$, meaning $\mathcal{N}_\delta(X') \geq \frac{1}{2}\mathcal{N}_\delta(X)$, for which

$$\mathcal{N}_\delta(X' + Y_1 + \cdots + Y_k) \ll_k \left(\prod_i K_i \right) \mathcal{N}_\delta(X').$$

In particular,

$$\mathcal{N}_\delta(Y_1 + \cdots + Y_k) \ll_k \left(\prod_i K_i \right) \mathcal{N}_\delta(X).$$

Proof. Discretize all sets on the δ -grid, apply the large-refinement assertion (SP.23) of Proposition 2.91, and thicken the selected grid subset back to a subset of X . Lemma 2.88 absorbs the bounded rounding errors. For the final assertion, choose any point $x \in X'$; then $Y_1 + \cdots + Y_k \subset -x + (X' + Y_1 + \cdots + Y_k)$, and use $\mathcal{N}_\delta(X') \leq \mathcal{N}_\delta(X)$. $\square$

A scale-regular subset The next lemma prevents a small additive-doubling statement that holds only at the finest scale from becoming useless after multiplication by a small coefficient.

Lemma 2.95 (Uniform-tree preprocessing). *Fix $\tau_0 > 0$. Let $X \subset [1, 2]$ be δ -separated and write $|X| = \delta^{-\tau}$ with $\tau_0 \leq \tau \leq 1$. Suppose*

$$\mathcal{N}_\delta(X + X) \leq K|X|.$$

For every $\varepsilon > 0$ there is $\tilde{X} \subset X$ satisfying

$$|\tilde{X}| \geq \delta^\varepsilon |X|$$

and, simultaneously for all $\delta < t < 1$,

$$\mathcal{N}_t(\tilde{X} + \tilde{X}) \ll_{\tau_0, \varepsilon} \delta^{-10\varepsilon} K \mathcal{N}_t(\tilde{X}). \quad (\text{SP.25})$$

Proof. Choose an integer block size $j = j(\tau_0, \varepsilon)$ so large that $2j \leq 2^{\varepsilon j}$ and that the harmless constants below are dominated by $2^{\varepsilon j}$. Replacing δ by a comparable smaller scale changes covering numbers by a constant depending only on j , so assume $\delta = 2^{-mj}$.

For $0 \leq \ell \leq m$, let $\mathcal{D}_\ell$ be the partition of $[1, 2]$ into intervals of length $2^{-\ell j}$. Let X_ℓ be the occupied intervals at level ℓ . Starting with $Z_m = X_m$, construct $Z_\ell \subset X_\ell$ upward. Once $Z_{\ell+1}$ is fixed,

$$|Z_{\ell+1}| = \sum_{I \in X_\ell} \#\{J \in Z_{\ell+1} : J \subset I\}.$$

Every positive summand lies between 1 and 2^j . Split the parents according to the dyadic range $2^{q-1} \leq \#\{J \subset I\} < 2^q$, $1 \leq q \leq j$, and retain the range carrying the largest part of the displayed sum. Thus the retained children above each $I \in Z_\ell$ are comparable within a factor two, and the total descendant mass loses at most a factor 2^j at each level.

Let $\tilde{X}_\ell$ be the set of points of X whose ancestors belong to every Z_r for $r > \ell$. Then

$$|\tilde{X}_\ell| \geq (2j)^{-(m-\ell)}|X|, \quad |\tilde{X}_0| \geq (2j)^{-m}|X| \geq \delta^\varepsilon|X|.$$

Set $\tilde{X} = \tilde{X}_0$. The near-uniform branching also implies the following quantitative occupancy statement: every $I \in Z_\ell$ contains at least

$$\delta^{2\varepsilon} \frac{|X|}{|Z_\ell|} \tag{SP.26}$$

points of the original set X , once δ is sufficiently small. Indeed the average occupancy is $|\tilde{X}_\ell|/|Z_\ell|$, the dyadic selection makes all nonzero occupancies comparable by at most $2^m \leq \delta^{-\varepsilon}$, and the global retained-mass loss is at most $\delta^{-\varepsilon}$.

Fix ℓ and put $t = 2^{-\ell j}$. A member of the interval sumset $Z_\ell + Z_\ell$ has length $2t$. Choose for each such interval one representing pair $I, J \in Z_\ell$ and fix one point $y \in X \cap J$. As x runs over $X \cap I$, the sums $x + y$ are δ -separated, so by (SP.26) that coarse sum interval contains at least $\delta^{2\varepsilon}|X|/|Z_\ell|$ separated points of $X + X$. Equal-length sum intervals have bounded overlap after passing to one of a bounded number of disjoint subfamilies. Consequently

$$\mathcal{N}_\delta(X + X) \gg |Z_\ell + Z_\ell| \delta^{2\varepsilon} \frac{|X|}{|Z_\ell|}.$$

Since the left side is at most $K|X|$,

$$|Z_\ell + Z_\ell| \ll \delta^{-2\varepsilon} K |Z_\ell| \leq \delta^{-10\varepsilon} K |Z_\ell|.$$

At scale t , the occupied intervals Z_ℓ give bounded-overlap covers of $\tilde{X}$, while $Z_\ell + Z_\ell$ gives one for $\tilde{X} + \tilde{X}$. Hence (SP.25) holds at all tree scales. If t is intermediate between consecutive tree scales, the ratio of the two is at most 2^j , so monotonicity of covering numbers and a change in the implicit constant give (SP.25) for every $\delta < t < 1$. $\square$

Consequences of simultaneous small additive and multiplicative growth We now begin the proof of Theorem 2.87. Apply Lemma 2.95 to A and call the resulting set A' . Its non-concentration estimate is weakened only by the factor $\delta^{-\varepsilon}$:

$$|A' \cap I| \leq C_0 \delta^{-\varepsilon} |I|^\sigma |A'|. \tag{SP.27}$$

Define

$$K := \frac{\mathcal{N}_\delta(A' + A') + \mathcal{N}_\delta(A'A')}{|A'|}. \tag{SP.28}$$

Fix the target exponent $c < c_0(\sigma)$ from Theorem 2.87, and choose an intermediate exponent

$$c < c_1 < c_0(\sigma).$$

If $K \geq \delta^{-c_1}$, then

$$\mathcal{N}_\delta(A + A) + \mathcal{N}_\delta(AA) \geq K|A'| \gg \delta^{-c_1+\varepsilon}|A|,$$

which is at least a constant multiple of $\delta^{-c}|A|$ once $\varepsilon < c_1 - c$ and δ is small. We may therefore assume throughout the structural argument that

$$K < \delta^{-c_1}. \tag{SP.29}$$

Since $c_0(\sigma) < \sigma/4$, we may and do choose $\varepsilon < \sigma/16$. This observation will ensure that every later large subset to which Lemma 2.95 is reapplied still has cardinality $\delta^{-\tau}$ with $\tau \geq \sigma/2$. We now prove the stronger lower bound forced by the dense/gap dichotomy.

Lemma 2.96 (Popular multiplicative overlap). *There exist $b \in A'$, a set $\bar{A} \subset A'$, and a number $K^{-1} \leq \rho \leq 1$ such that*

$$|\bar{A}| \gg |\log \delta|^{-1} \rho |A'|, \tag{SP.30}$$

$$|(aA')_\delta \cap (bA')_\delta| \asymp |A'| K^{-1} \rho^{-1} \quad (a \in \bar{A}), \tag{SP.31}$$

and for each $a \in \bar{A}$,

$$\mathcal{N}_\delta(aA' \pm bA') \ll K^3 \rho |A'|. \tag{SP.32}$$

Proof. The family of grid sets $(aA')_\delta$ has total incidence $\asymp |A'|^2$. Since their union lies in a bounded thickening of $A'A'$ and

$$\mathcal{N}_\delta(A'A') \leq K|A'|,$$

Cauchy–Schwarz gives

$$\sum_{a,a' \in A'} |(aA')_\delta \cap (a'A')_\delta| \gg \frac{|A'|^4}{K|A'|} = |A'|^3 K^{-1}.$$

Average in the second variable and choose b for which the inner sum is $\gg |A'|^2 K^{-1}$. Dyadic pigeonholing the intersection sizes yields ρ and $\bar{A}$ with (SP.30)–(SP.31); there are only $O(|\log \delta|)$ dyadic ranges.

For fixed $a \in \bar{A}$, put $X = (aA')_\delta \cap (bA')_\delta$. Then

$$|X| \gg K^{-1} \rho^{-1} |A'|.$$

Both $X + aA'$ and $X + bA'$ are contained, up to bounded δ -thickening, in a translate/dilate of $A' + A'$ or $A'A'$ and hence have covering number $O(K|A'|) = O(K^2 \rho |X|)$. Corollary 2.94, followed by Proposition 2.92 for the difference case, gives

$$\mathcal{N}_\delta(aA' \pm bA') \ll K^4 \rho^2 |X| \ll K^3 \rho |A'|.$$

□

By (SP.30), $\rho \geq K^{-1}$ and (SP.29),

$$|\bar{A}| \gg |\log \delta|^{-1} \delta^{c_1} |A'| \gg \delta^{-\sigma/2}$$

for sufficiently small δ (after decreasing ε if necessary). Thus Lemma 2.95 applies to $\bar{A}$ with the fixed lower exponent $\tau_0 = \sigma/2$. Call the resulting subset A_1 . Then

$$|A_1| \gg \delta^{O(\varepsilon)} |\log \delta|^{-1} \rho |A'|, \quad (\text{SP.33})$$

and for every $\delta < t < 1$,

$$\mathcal{N}_t(A_1 + A_1) \ll K |\log \delta| \rho^{-1} \delta^{-O(\varepsilon)} \mathcal{N}_t(A_1). \quad (\text{SP.34})$$

Lemma 2.97 (Two-coefficient upper bound). *Let $d_1, d_2 \in A_1 - A_1$ and put $d = \max(|d_1|, |d_2|)$. Then*

$$\mathcal{N}_\delta(d_1 A' + d_2 A') \ll C_0 \delta^{-O(\varepsilon)} K^{12} \rho^4 d^\sigma |A'|. \quad (\text{SP.35})$$

Proof. Write $d_1 = a_1 - b_1$ and $d_2 = a_2 - b_2$ with $a_i, b_i \in A_1 \subset \bar{A}$. Apply the covering Plünnecke refinement to the four sets $a_1 A', -b_1 A', a_2 A', -b_2 A'$ and the base set $b A'$, using (SP.32). We obtain a refinement $A'' \subset A'$ with $|A''| \gg |A'|$ in covering-number scale and

$$\mathcal{N}_\delta(b A'' + d_1 A' + d_2 A') \ll K^{12} \rho^4 |A'|. \quad (\text{SP.36})$$

(The displayed exponent 12 is deliberately nonoptimized; any fixed polynomial power would suffice later.)

Put $S = d_1 A' + d_2 A'$. Since $A' \subset [1, 2]$, the set S lies in an interval of length at most Cd for an absolute C . Choose a maximal $10Cd$ -separated subset $P \subset b A''$. Maximality gives $|P| \gg \mathcal{N}_d(b A'')$, while the intervals containing the translates $p + S$ ($p \in P$) are pairwise disjoint. Hence a δ -cover of $b A'' + S$ must cover each $p + S$ separately, and

$$\mathcal{N}_\delta(b A'' + S) \geq \sum_{p \in P} \mathcal{N}_\delta(p + S) = |P| \mathcal{N}_\delta(S) \gg \mathcal{N}_d(b A'') \mathcal{N}_\delta(S).$$

Rearranging gives

$$\mathcal{N}_\delta(d_1 A' + d_2 A') \ll \mathcal{N}_d(b A'')^{-1} \mathcal{N}_\delta(b A'' + d_1 A' + d_2 A').$$

By (SP.27), every interval of length $O(d)$ contains at most $C_0 \delta^{-\varepsilon} d^\sigma |A'|$ points of A' . Since A'' is a fixed-proportion covering refinement, this forces

$$\mathcal{N}_d(b A'') \gg C_0^{-1} \delta^\varepsilon d^{-\sigma}.$$

Substitute this and (SP.36) to obtain (SP.35). □

We also need the same estimate with a bounded number of copies of one coefficient.

Lemma 2.98 (Bounded repeated coefficient, no refinement). *Fix $k \geq 2$. Let $d_1, d_2 \in A_1 - A_1$, assume $|d_2| \geq \delta$, and put $d = \max(|d_1|, |d_2|)$. Then*

$$\mathcal{N}_\delta(d_1 A_1 + \underbrace{d_2 A_1 + \cdots + d_2 A_1}_{k \text{ copies}}) \ll C_0 \delta^{-O_k(\varepsilon)} |\log \delta|^{O_k(1)} K^{13+k} \rho^{2-k} d^\sigma |A'|. \quad (\text{SP.37})$$

Proof. Choose, exactly as in the proof of Lemma 2.97, a δ -refinement $A'' \subset A'$ such that

$$\mathcal{N}_\delta(bA'' + d_1A' + d_2A') \ll K^{12}\rho^4|A'|.$$

After restricting d_1A' and d_2A' to A_1 and using (SP.33),

$$\mathcal{N}_\delta(bA'' + d_1A_1 + d_2A_1) \ll \delta^{-O(\varepsilon)}|\log \delta|K^{12}\rho^3|A_1|. \quad (\text{SP.38})$$

Set $X = d_2A_1$. Applying (SP.34) at scale $\delta/|d_2|$ and transporting by dilation gives

$$\mathcal{N}_\delta(X + X) \ll L\mathcal{N}_\delta(X), \quad L := K\delta^{-O(\varepsilon)}\rho^{-1}|\log \delta|.$$

The “in particular” part of Corollary 2.94, applied to $k+1$ identical copies of X , yields

$$\mathcal{N}_\delta(\underbrace{X + \cdots + X}_{k+1 \text{ copies}}) \ll_k L^{k+1}\mathcal{N}_\delta(X). \quad (\text{SP.39})$$

Now apply the same corollary with base X and with

$$Y_1 = \underbrace{X + \cdots + X}_{k \text{ copies}}, \quad Y_2 = bA'' + d_1A_1.$$

By (SP.39), the expansion ratio for Y_1 is at most $O_k(L^{k+1})$. If

$$R := \frac{\mathcal{N}_\delta(X + Y_2)}{\mathcal{N}_\delta(X)},$$

then (SP.38) bounds $R\mathcal{N}_\delta(X)$. Hence the global conclusion of the covering Plünnecke corollary gives

$$\begin{aligned} & \mathcal{N}_\delta(bA'' + d_1A_1 + \underbrace{d_2A_1 + \cdots + d_2A_1}_{k \text{ copies}}) \\ & \ll_k L^{k+1}R\mathcal{N}_\delta(X) \\ & \ll \delta^{-O_k(\varepsilon)}|\log \delta|^{O_k(1)}K^{13+k}\rho^{2-k}|A_1|. \end{aligned}$$

Finally the set without bA'' lies in an interval of length $O_k(d)$. The separated-translate argument of Lemma 2.97, with the same A'' , gives

$$\mathcal{N}_d(bA'') \gg C_0^{-1}\delta^\varepsilon d^{-\sigma}.$$

Dividing by this lower bound and using $|A_1| \leq |A'|$ proves (SP.37). $\square$

The quotient-set dichotomy Fix $0 < \gamma < 1/2$ and set

$$B := \left\{ \frac{a_1 - a_2}{a_3 - a_4} : a_i \in A_1, |a_3 - a_4| > \delta^\gamma \right\}.$$

Then $B \subset [-\delta^{-\gamma}, \delta^{-\gamma}]$ and $0, 1 \in B$. Let

$$s \asymp \delta^{1-2\gamma}$$

be a dyadic scale.

Lemma 2.99 (Dense-or-gap dichotomy). *At least one of the following holds.*

(A) There is $b \in B \cap [0, 1]$ for which either $b/2$ or $(b+1)/2$ has distance at least s from B .

(B) $\mathcal{N}_s(B \cap [0, 1]) \gg s^{-1}$.

Proof. Assume (A) fails. The $2s$ -neighborhood of B is then invariant, up to the same $2s$ error, under the two contractions

$$x \mapsto x/2, \quad x \mapsto (x+1)/2.$$

Starting from $0, 1 \in B$, induction shows that every dyadic rational $p/2^n$ with $0 \leq p \leq 2^n$ belongs to the $2s$ -neighborhood of B for every n for which $2^{-n} \gtrsim s$. Taking $2^{-n} \asymp s$ gives $\gg s^{-1}$ s -separated points within $O(s)$ of $B \cap [0, 1]$, proving (B). $\square$

Dense case Assume alternative (B). By averaging the s -cells meeting $B \cap [0, 1]$, choose

$$b = \frac{b_1 - b_2}{b_3 - b_4} \in B \cap [0, 1]$$

with $|b_3 - b_4| > \delta^\gamma$ such that the number of quadruples $(a_1, a_2, a_3, a_4) \in A_1^4$ satisfying

$$\left| \frac{a_1 - a_2}{a_3 - a_4} - b \right| \ll \delta^{1-2\gamma}, \quad |a_3 - a_4| > \delta^\gamma, \quad (\text{SP.40})$$

is $\ll |A_1|^4 \delta^{1-2\gamma}$.

Set $D_1 = b_1 - b_2$ and $D_2 = b_3 - b_4$. Lemma 2.97 gives

$$\mathcal{N}_\delta(D_1 A_1 + D_2 A_1) \ll C_0 \delta^{-O(\varepsilon)} K^{12} \rho^4 |D_2|^\sigma |A'|. \quad (\text{SP.41})$$

Let Q be the approximate additive-energy set of quadruples satisfying

$$D_2 a_1 + D_1 a_4 = D_2 a_2 + D_1 a_3 + O(\delta). \quad (\text{SP.42})$$

Cauchy–Schwarz in the representation function gives

$$\mathcal{N}_\delta(D_1 A_1 + D_2 A_1) \gg \frac{|A_1|^4}{|Q|}. \quad (\text{SP.43})$$

Split Q according as $|a_3 - a_4|$ is at least or at most δ^γ .

In the first range, dividing (SP.42) by $D_2(a_3 - a_4)$ shows that the quotient in (SP.40) is within $O(\delta^{1-2\gamma})$ of b . Hence

$$|Q_{\text{large}}| \ll |A_1|^4 \delta^{1-2\gamma}.$$

If this contributes at least half of Q , then (SP.43), (SP.33), and (SP.41) imply

$$K \gg C_0^{-O(1)} |\log \delta|^{-O(1)} \delta^{(2\gamma+\sigma-1)/12+O(\varepsilon)}. \quad (\text{SP.44})$$

In the second range, choose a_1, a_4 first. The Frostman estimate gives at most $C_0 \delta^{\gamma\sigma} |A|$ choices for a_3 . Once a_3 is fixed, (SP.42) confines a_2 to an interval of length $O(\delta/|D_2|)$, so there are at most $C_0 |D_2|^{-\sigma}$ choices after using $|A| \asymp \delta^{-\sigma}$. Thus

$$|Q_{\text{small}}| \ll C_0^2 |A_1|^2 |A'| \delta^{\gamma\sigma+O(\varepsilon)} |D_2|^{-\sigma}.$$

If this contributes at least half of Q , then (SP.43) and (SP.41) give

$$K \gg C_0^{-O(1)} |\log \delta|^{-O(1)} \delta^{-\gamma\sigma/12+O(\varepsilon)}. \quad (\text{SP.45})$$

Therefore the dense case yields

$$K \gg C_0^{-O(1)} |\log \delta|^{-O(1)} \delta^{O(\varepsilon)} \min \left(\delta^{(2\gamma+\sigma-1)/12}, \delta^{-\gamma\sigma/12} \right). \quad (\text{SP.46})$$

Gap case Assume alternative (A). Thus, for some

$$b = \frac{b_1 - b_2}{b_3 - b_4} \in B, \quad |b_3 - b_4| \geq \delta^\gamma,$$

either $b/2$ or $(b+1)/2$ is s -separated from B . Write the separated number as e_1/e_2 . In the first case one can take

$$e_1 = d_1, \quad e_2 = 2d_2,$$

and in the second

$$e_1 = d_1 + d_2, \quad e_2 = 2d_2,$$

for $d_1, d_2 \in A_1 - A_1$.

Let Q consist of quadruples satisfying

$$e_2 a_1 + e_1 a_4 = e_2 a_2 + e_1 a_3 + O(\delta).$$

If $|a_3 - a_4| \geq \delta^\gamma$, division shows

$$\left| \frac{a_1 - a_2}{a_3 - a_4} - \frac{e_1}{e_2} \right| \ll \delta^{1-2\gamma} \asymp s,$$

contradicting the gap property because the quotient belongs to B . Hence every quadruple in Q has $|a_3 - a_4| \leq \delta^\gamma$.

Choose a_1, a_4 first. The Frostman bound gives at most $C_0 \delta^{\gamma\sigma} |A'| \delta^{-O(\varepsilon)}$ choices for a_3 , and then at most $C_0 |e_2|^{-\sigma} \delta^{-O(\varepsilon)}$ choices for a_2 . Therefore

$$|Q| \ll C_0^{O(1)} |A_1|^2 |A'| \delta^{\gamma\sigma - O(\varepsilon)} |e_2|^{-\sigma}.$$

The energy inequality therefore gives

$$\mathcal{N}_\delta(e_1 A_1 + e_2 A_1) \gg C_0^{-O(1)} \rho^2 |\log \delta|^{-O(1)} \delta^{-\gamma\sigma + O(\varepsilon)} |e_2|^\sigma |A'|. \quad (\text{SP.47})$$

On the other hand, in Case (A.1) the set $e_1 A_1 + e_2 A_1$ is contained in $d_1 A_1 + d_2 A_1 + d_2 A_1$, and in Case (A.2) it is contained in $d_1 A_1 + d_2 A_1 + d_2 A_1 + d_2 A_1$. Since $|d_2| \geq \delta^\gamma$, Lemma 2.98, with $k = 2$ or 3, gives in the worse second case

$$\mathcal{N}_\delta(e_1 A_1 + e_2 A_1) \ll C_0 \rho^{-1} |\log \delta|^{O(1)} \delta^{-O(\varepsilon)} K^{16} |e_2|^\sigma |A'|. \quad (\text{SP.48})$$

Comparing (SP.47) and (SP.48) gives

$$K^{16} \gg C_0^{-O(1)} |\log \delta|^{-O(1)} \rho^3 \delta^{-\gamma\sigma + O(\varepsilon)}.$$

Since $\rho \geq K^{-1}$, we obtain the fully internal, deliberately nonoptimized bound

$$K \gg C_0^{-O(1)} |\log \delta|^{-O(1)} \delta^{-\gamma\sigma/19 + O(\varepsilon)}. \quad (\text{SP.49})$$

Optimization and completion of the proof By Lemma 2.99, either (SP.46) or (SP.49) holds. Choose

$$\gamma = \frac{19(1-\sigma)}{2(19+6\sigma)}.$$

Then

$$\frac{1-\sigma-2\gamma}{12} = \frac{\gamma\sigma}{19} = \frac{\sigma(1-\sigma)}{2(19+6\sigma)} = c_0(\sigma),$$

while the other dense exponent $\gamma\sigma/12$ is larger. Thus

$$K \gg C_0^{-O(1)} |\log \delta|^{-O(1)} \delta^{-c_0(\sigma)+O(\varepsilon)}.$$

Because A' retained $\delta^{O(\varepsilon)}$ of A and

$$A' + A' \subset A + A, \quad A'A' \subset AA,$$

we conclude

$$\mathcal{N}_\delta(A + A) + \mathcal{N}_\delta(AA) \gg C_0^{-O(1)} |\log \delta|^{-O(1)} \delta^{-c_0(\sigma)+O(\varepsilon)} |A|.$$

Given $c < c_0(\sigma)$, choose ε so small that the $O(\varepsilon)$ loss is less than $(c_0(\sigma) - c)/2$; for sufficiently small δ , the logarithm is absorbed by the remaining power. This is exactly (SP.20), completing the proof of Theorem 2.87. $\square$

2.17.1.6 Scalar-action growth from coefficient segments

The planar projection argument requires a statement in which a family of scalars acts on a possibly different one-dimensional set. The scalar sum-product theorem above does not by itself imply this statement. The argument below therefore supplies the missing mechanism internally: a regular Frostman window, homogeneous ring saturation, fixed-exponent stabilization, finite multiscale stitching, and a short-interval energy contradiction. The proof below keeps each quantitative transition explicit rather than replacing the chain by the phrase “iterate sum-product”.

A fully internal energy reduction The next lemma closes one part of the scalar-action mechanism without any iteration of coefficient sum-product. It says that once the coefficient set has been amplified to a whole unit interval, additive small doubling and small scalar-action growth are incompatible with the two quantitative hypotheses on X . The exponent balance is displayed explicitly because it is the part of the argument that is lost if one merely writes “by an energy argument”.

Lemma 2.100 (Unit-interval scalar action forces growth). *Fix $\kappa > 0$ and $0 < \sigma < 1$, and put*

$$\beta = \frac{1-\sigma}{1+\kappa}, \quad q = \beta\kappa = \frac{\kappa(1-\sigma)}{1+\kappa}.$$

There is $\eta_0 = \eta_0(\kappa, \sigma) > 0$ with the following property. If $0 < \eta < \eta_0$, $r > 0$ is sufficiently small, and the r -separated set $X \subset \mathbb{R}$ satisfies

$$N_\rho(X) \geq r^\eta \rho^{-\kappa} \quad (r \leq \rho \leq 1), \quad (\text{SP.50})$$

$$N_r(X) \leq r^{-\sigma-\eta}, \quad (\text{SP.51})$$

then

$$N_r(X + X) + \sup_{|a| \leq 1} N_r(X + aX) \geq r^{-\eta} N_r(X). \quad (\text{SP.52})$$

One may take any $\eta_0 < q/6$.

Proof. Write $N = N_r(X)$. Since X is r -separated, replacing X by a maximal r -separated representative of the occupied r -intervals changes all cardinalities below by absolute factors only. We therefore use N for its cardinality as well.

Assume, toward a contradiction, that both terms on the left of (SP.52) are $< r^{-\eta}N$. We first derive the local concentration estimate

$$\#(X \cap I) \ll r^{-2\eta} \rho^\kappa N \quad (\text{SP.53})$$

for every interval I of length 2ρ , $r \leq \rho \leq 1$. Choose a maximal 8ρ -separated set $P \subset X$. The lower bound (SP.50) gives $\#P \gg r^\eta \rho^{-\kappa}$. If $Z = X \cap I$, then the sets $p + Z$, $p \in P$, lie in $X + X$ and are pairwise separated by a distance comparable to ρ . Hence their r -covers have bounded overlap and

$$N_r(X + X) \gg \#P \#Z.$$

The assumed additive small-doubling bound gives (SP.53).

For $a \in [-1, 1]$, let $E(a)$ be the number of quadruples $(x, y, x', y') \in X^4$ such that

$$|x + ay - x' - ay'| \leq 4r.$$

Cauchy–Schwarz over the occupied $O(r)$ -intervals of $X + aX$ yields

$$E(a) \gg \frac{N^4}{N_r(X + aX)} \gg r^\eta N^3 \quad (\text{SP.54})$$

for every $|a| \leq 1$.

Set $\rho = r^\beta$. We bound the integral of $E(a)$ over $[-1, 1]$ by splitting according to $|y - y'| < \rho$. If $|y - y'| \geq \rho$, then for each fixed quadruple the admissible a 's form an interval of length $O(r/\rho)$, so this part contributes

$$\ll \frac{r}{\rho} N^4.$$

For the near pairs, (SP.53) shows that their number is

$$\#\{(y, y') \in X^2 : |y - y'| < \rho\} \ll r^{-2\eta} \rho^\kappa N^2.$$

For fixed x, y, y' and fixed a , the r -separation of X permits only $O(1)$ choices of x' satisfying the displayed energy inequality. Integrating first in a and then summing therefore gives the near-pair contribution

$$\ll r^{-2\eta} \rho^\kappa N^3.$$

Consequently

$$\int_{-1}^1 E(a) da \ll \left(\frac{r}{\rho} N + r^{-2\eta} \rho^\kappa \right) N^3. \quad (\text{SP.55})$$

By (SP.51) and the choice of β ,

$$\frac{r}{\rho} N \leq r^{1-\beta-\sigma-\eta} = r^{q-\eta}, \quad r^{-2\eta} \rho^\kappa = r^{q-2\eta}.$$

Thus the right side of (SP.55) is $O(r^{q-2\eta} N^3)$. On the other hand, integrating (SP.54) gives $\int_{-1}^1 E(a) da \gg r^\eta N^3$. If $3\eta < q$, these bounds are incompatible for sufficiently small r . This proves the lemma; the stated choice $\eta_0 < q/6$ leaves room for all absolute covering constants. $\square$

Lemma 2.101 (Fixed-scale additive calculus for good coefficients). *Let $X \subset \mathbb{R}$ be bounded and $r > 0$. Assume*

$$N_r(X + X) \leq LN_r(X), \quad L \geq 2,$$

and define

$$\mathcal{G}_L(X; r) = \{a \in \mathbb{R} : N_r(X + aX) \leq LN_r(X)\}.$$

If $a, b \in \mathcal{G}_L(X; r)$, then

$$N_r(X + (a + b)X) \ll L^3 N_r(X),$$

$$N_r(X + (a - b)X) \ll L^4 N_r(X).$$

Moreover, if $X \subset [-M, M]$ and $|a - b| \leq r/M$, then

$$N_r(X + bX) \ll N_r(X + aX).$$

All implied constants are absolute.

Proof. For the sum, apply Corollary 2.94 with base set X and the three summands X, aX, bX ; the hypotheses for these summands have constants L, L, L . For the difference first apply the covering Ruzsa triangle inequality with the three sets $X, -X, bX$ to obtain

$$N_r(X - bX) \ll \frac{N_r(X + X)N_r(X + bX)}{N_r(X)} \ll L^2 N_r(X).$$

A second application of the covering Plünnecke corollary to aX and $-bX$ gives the asserted L^4 bound. Finally, if $|a - b| \leq r/M$, then $(a - b)X \subset [-r, r]$, so $X + bX$ is contained in an $O(r)$ -neighborhood of $X + aX$; enlarging covering intervals by an absolute factor proves the last estimate. $\square$

Lemma 2.102 (Quantitative good-coefficient ring calculus). *Let $0 < r < 1$, let $X \subset [-M, M]$ be bounded, and put $N = N_r(X)$. Assume*

$$N_r(X + X) \leq LN, \quad L \geq 2. \tag{SP.56}$$

For $a \in \mathbb{R}$ write

$$a \in \mathcal{G}(L) \iff N_r(X + aX) \leq LN.$$

There is an absolute constant $C \geq 1$ such that the following hold.

(i) *If $a, b \in \mathcal{G}(L)$, then*

$$a + b, a - b \in \mathcal{G}(CL^4).$$

(ii) *If $a, b \in \mathcal{G}(L)$ and $|a| \leq B$ with $B \geq 1$, then*

$$ab \in \mathcal{G}(CL^8(1 + B)).$$

(iii) *If $a \neq 0$ and $a \in \mathcal{G}(L)$, then*

$$a^{-1} \in \mathcal{G}(CL \max(1, |a|^{-1})).$$

(iv) *If $a \in \mathcal{G}(L)$ and $|a - b|M \leq r$, then $b \in \mathcal{G}(CL)$.*

Consequently, for every fixed word depth s there are constants $A_s \geq 1$ and $C_s \geq 1$ with the following property. If 1 and every member of a set $C \subset [-B, B]$, $B \geq 1$, lie in $\mathcal{G}(L)$, then every coefficient obtained from at most s leaves in $C \cup (-C) \cup \{0, 1, -1\}$ by addition, subtraction and multiplication lies in

$$\mathcal{G}(C_s L^{A_s} (1+B)^{A_s}).$$

Proof. Part (i) is Lemma 2.101, after enlarging the absolute constant.

For (ii), apply Lemma 2.93 to the pair X, bX . There is a set T with $|T| \ll L$ such that

$$bX \subset (X - X) + T + [-Cr, Cr]. \quad (\text{SP.57})$$

Multiplying by a and adding X gives

$$X + abX \subset X + aX - aX + aT + [-C(1+|a|)r, C(1+|a|)r].$$

The covering Ruzsa triangle inequality, with middle set $-X$, gives

$$N_r(X - aX) \ll \frac{N_r(X + X)N_r(X + aX)}{N_r(X)} \ll L^2 N.$$

Hence covering Plünnecke, with base set X and summands aX and $-aX$, bounds the first three summands by $CL^3 N$. There are $O(L)$ translates indexed by aT , and an interval of length $O((1+B)r)$ is covered by $O(1+B)$ intervals of radius r . Thus

$$N_r(X + abX) \leq CL^8(1+B)N.$$

No change of working scale is hidden in this estimate; the scale enlargement created by multiplication has been paid for explicitly by the factor $1+B$.

For (iii), scaling by a gives the exact identity

$$X + a^{-1}X = a^{-1}(aX + X).$$

Hence

$$N_r(X + a^{-1}X) = N_{|a|r}(X + aX).$$

If $|a| \geq 1$, the right side is at most $N_r(X + aX)$. If $|a| < 1$, every r -interval is covered by $O(|a|^{-1})$ intervals of radius $|a|r$, so

$$N_{|a|r}(X + aX) \ll |a|^{-1} N_r(X + aX).$$

This proves (iii). Part (iv) follows because $(b-a)X \subset [-r, r]$, so $X + bX$ is contained in an $O(r)$ -neighbourhood of $X + aX$.

For the bounded-word conclusion, first note that $0 \in \mathcal{G}(L)$. Applying part (i) to 0 and to any $a \in C \cup \{1\}$ shows that $-a \in \mathcal{G}(CL^4)$. After replacing the initial loss parameter by CL^4 , all signed leaves in $C \cup (-C) \cup \{0, 1, -1\}$ are therefore good; this fixed replacement is absorbed into the final constants A_s, C_s . Now induct on the number of leaves. Every subexpression with at most s leaves has magnitude at most $(2B)^s$. For one leaf this is $|u| \leq B \leq 2B$. If the two child expressions use $i, j \geq 1$ leaves with $i + j \leq s$, then

$$|u + v| \leq (2B)^i + (2B)^j \leq 2(2B)^{s-1} \leq (2B)^s, \quad |uv| \leq (2B)^{i+j} \leq (2B)^s,$$

because $B \geq 1$. Thus each invocation of part (ii) costs at most a fixed power of $1+B$, depending only on s ; parts (i) and (iv) cost only fixed powers of the current doubling parameter. Since a word with at most s leaves contains at most $s-1$ binary operations, induction absorbs all losses into $C_s L^{A_s} (1+B)^{A_s}$. $\square$

Lemma 2.103 (Goodness descends to coarser scales). *Let $0 < r \leq \rho \leq 1$, let $X \subset \mathbb{R}$ be bounded, and put $N_t = N_t(X)$. Suppose*

$$N_r(X + X) \leq LN_r, \quad N_r(X + aX) \leq LN_r$$

with $L \geq 2$. Then

$$N_\rho(X + aX) \ll L^4 N_\rho(X). \quad (\text{SP.58})$$

The implied constant is absolute and is independent of a, r, ρ .

Proof. Let

$$m_\rho := \max_{t \in \mathbb{R}} N_r(X \cap [t - 2\rho, t + 2\rho]).$$

A maximal 4ρ -separated subset of X shows

$$N_r(X) \ll m_\rho N_\rho(X). \quad (\text{SP.59})$$

Choose t for which the maximum defining m_ρ is attained up to a factor two, and let $Z \subset X \cap [t - 2\rho, t + 2\rho]$ be a maximal r -separated subset. Choose a maximal 8ρ -separated subset $P \subset X + aX$. The sets $Z + p$, $p \in P$, have bounded overlap at scale r because the points of P are 8ρ -separated whereas $Z - Z \subset [-4\rho, 4\rho]$. Since $Z + P \subset X + X + aX$, it follows that

$$N_r(X + X + aX) \gg m_\rho N_\rho(X + aX). \quad (\text{SP.60})$$

Covering Plünnecke, with base set X and summands X, aX , gives

$$N_r(X + X + aX) \ll L^3 N_r(X).$$

Combine this with (SP.59)–(SP.60) and cancel m_ρ to obtain (SP.58). $\square$

Lemma 2.104 (Normalization of an approximate small segment at the induced scale). *Fix $0 < r < 1$ and let $X \subset [-M, M]$. Assume*

$$N_r(X + X) \leq LN_r(X).$$

Let $0 < \lambda \leq 1$, let $0 < \varepsilon \leq \lambda/8$, and suppose that for every $t \in [-\lambda, \lambda]$ there exists a coefficient w_t such that

$$|t - w_t|M \leq \varepsilon, \quad N_r(X + w_t X) \leq L_1 N_r(X). \quad (\text{SP.61})$$

Assume also that the chosen w_λ satisfies $|w_\lambda| \geq \lambda/2$. Put

$$\rho := \max\{r, 4\varepsilon/\lambda\}. \quad (\text{SP.62})$$

Then every $a \in [-1, 1]$ satisfies

$$N_\rho(X + aX) \ll L^{O(1)} L_1^{O(1)} \lambda^{-O(1)} N_\rho(X), \quad (\text{SP.63})$$

with absolute exponents.

Proof. Fix $a \in [-1, 1]$ and choose $w_{a\lambda}$ from (SP.61). The inverse clause of Lemma 2.102 makes w_λ^{-1} good at scale r with loss $O(L_1 \lambda^{-1})$. The multiplicative clause then makes

$$c_a := w_{a\lambda} w_\lambda^{-1}$$

good at scale r with a loss polynomial in L_1 and λ^{-1} . Furthermore,

$$\begin{aligned} |c_a - a|M &= \frac{|w_{a\lambda} - aw_\lambda|M}{|w_\lambda|} \\ &\leq \frac{|w_{a\lambda} - a\lambda|M + |a||\lambda - w_\lambda|M}{|w_\lambda|} \leq \frac{4\varepsilon}{\lambda} \leq \rho. \end{aligned}$$

Lemma 2.103 transfers the good-coefficient bound for c_a from r to ρ . Applying the same lemma with coefficient 1 transfers the additive-doubling bound to scale ρ . Finally the perturbation clause of Lemma 2.102, now at scale ρ , replaces c_a by a . This proves (SP.63). $\square$

Lemma 2.105 (Multiplicative scale propagation by an approximate contraction). *There is an absolute constant $C \geq 1$ with the following property. Let $0 < r_0 \leq \delta < 1$, let $B, L \geq 2$, and let $X \subset [-B, B]$ be bounded. Suppose $b \in \mathbb{R}$ satisfies*

$$\frac{\delta}{2} \leq |b| \leq 2\delta \tag{SP.64}$$

and, for every $t \in [r_0, 1]$,

$$N_t(X + bX) \leq LN_t(X). \tag{SP.65}$$

Then for every $\rho \in [r_0/\delta, 1]$,

$$N_{C\delta\rho}(X) \geq C^{-1}(BL)^{-1}N_{CB\delta}(X)N_\rho(X). \tag{SP.66}$$

Consequently, using the one-dimensional scale-change inequality,

$$N_{C\delta\rho}(X) \geq C^{-1}(B^2L)^{-1}N_\delta(X)N_\rho(X). \tag{SP.67}$$

All constants are independent of r_0, δ, B, L .

Proof. Choose a maximal $8B\delta$ -separated subset $P \subset X$ and a maximal 4ρ -separated subset $Q \subset X$. Maximality says that the $8B\delta$ -intervals centred at P cover X , and the 4ρ -intervals centred at Q cover X . After fixing the absolute constant $C \geq 8$, monotonicity of covering numbers gives $N_{CB\delta}(X) \leq |P|$. In one dimension every interval of radius 4ρ is covered by at most an absolute number of intervals of radius ρ , so $N_\rho(X) \ll N_{4\rho}(X) \leq |Q|$. Hence

$$|P| \gg N_{CB\delta}(X), \quad |Q| \gg N_\rho(X),$$

with absolute implied constants. For each $p \in P$, the set $p + bQ$ is $2\delta\rho$ -separated by (SP.64). Moreover $bQ \subset [-2B\delta, 2B\delta]$; hence the families $p + bQ$, $p \in P$, have bounded overlap at scale $C\delta\rho$ because the centers in P are $8B\delta$ -separated. Thus

$$N_{C\delta\rho}(X + bX) \gg N_{CB\delta}(X)N_\rho(X). \tag{SP.68}$$

Since $C\delta\rho \geq r_0$ after increasing the absolute C , (SP.65) and an absolute scale adjustment give

$$N_{C\delta\rho}(X + bX) \ll LN_{C\delta\rho}(X).$$

Combining this with (SP.68) proves (SP.66). Finally, in one dimension an interval of radius $CB\delta$ is covered by $O(B)$ intervals of radius δ , so $N_{CB\delta}(X) \gg B^{-1}N_\delta(X)$. This gives (SP.67). $\square$

Corollary 2.106 (Finite multiplicative descent). *Under the hypotheses of Lemma 2.105, suppose in addition that $k \geq 2$ is an integer with $k \leq k_0$ and that $r_0 \leq \delta^k$. Then*

$$N_{C_k \delta^k}(X) \geq (C_k B^2 L)^{-k} N_\delta(X)^k, \quad (\text{SP.69})$$

where C_k depends only on k . Hence, if $B, L \leq r_0^{-\eta}$, $\delta^k \asymp_{k_0} r_0$, and

$$N_\delta(X) \geq \delta^{-\sigma-\eta_1},$$

then

$$N_{r_0}(X) \geq r_0^{-\sigma-\eta_1+O_{k_0}(\eta)}. \quad (\text{SP.70})$$

The constant hidden in $O_{k_0}(\eta)$ is fixed before η is chosen.

Proof. Apply (SP.67) successively with $\rho = \delta, \delta^2, \dots, \delta^{k-1}$. At each step the harmless absolute scale factor can be absorbed into C_k because $k \leq k_0$. This yields (SP.69). Substitute the assumed lower bound for $N_\delta(X)$ and use $\delta^k \asymp_{k_0} r_0$; the polynomial losses from B, L contribute only $O_{k_0}(\eta)$ to the final exponent. A final bounded scale change gives (SP.70). $\square$

Lemma 2.107 (Short coefficient intervals already force scalar-action growth). *Fix $\kappa > 0$ and $0 < \sigma < 1$. Let*

$$0 < \alpha < \frac{1-\sigma}{4}, \quad \beta := \frac{1-\sigma-2\alpha}{2} > 0.$$

There is $\eta_0 = \eta_0(\kappa, \sigma, \alpha) > 0$ with the following property. Suppose $0 < \eta < \eta_0$, $r > 0$ is sufficiently small, $\lambda \geq r^\alpha$, and the r -separated set $X \subset \mathbb{R}$ satisfies

$$N_t(X) \geq r^\eta t^{-\kappa} \quad (r \leq t \leq 1), \quad (\text{SP.71})$$

$$N_r(X) \leq r^{-\sigma-\eta}. \quad (\text{SP.72})$$

Then

$$N_r(X + X) + \sup_{|a| \leq \lambda} N_r(X + aX) \geq r^{-\eta} N_r(X). \quad (\text{SP.73})$$

One may take any η_0 satisfying

$$4\eta_0 < \min\{\beta\kappa, 1 - \sigma - \alpha - \beta\}.$$

Proof. Put $N = N_r(X)$ and replace X , at an absolute covering-number cost, by a maximal r -separated representative of its occupied r -intervals. Assume for contradiction that the left side of (SP.73) is $< r^{-\eta} N$.

Exactly as in the proof of Lemma 2.100, the additive small-doubling hypothesis and (SP.71) imply the local concentration bound

$$\#(X \cap I) \ll r^{-2\eta} t^\kappa N \quad (\text{SP.74})$$

for every interval I of length $2t$, $r \leq t \leq 1$. Indeed a maximal $8t$ -separated subset of X has $\gg r^\eta t^{-\kappa}$ elements, and its translates of $X \cap I$ have bounded overlap inside $X + X$.

For $a \in [-\lambda, \lambda]$, let $E(a)$ count quadruples $(x, y, x', y') \in X^4$ with

$$|x + ay - x' - ay'| \leq 4r.$$

Cauchy–Schwarz gives, uniformly in a ,

$$E(a) \gg r^\eta N^3. \quad (\text{SP.75})$$

We integrate over the coefficient interval and set $t = r^\beta$. For $|y - y'| \geq t$, a fixed quadruple permits a only in an interval of length at most Cr/t (after intersecting with $[-\lambda, \lambda]$); hence the far-pair contribution is

$$\ll \frac{r}{t} N^4.$$

For $|y - y'| < t$, (SP.74) gives at most $O(r^{-2\eta} t^\kappa N^2)$ ordered pairs (y, y') . For fixed x, y, y' and fixed a , r -separation permits only $O(1)$ choices of x' . Thus the near-pair contribution, after integration over an interval of length 2λ , is

$$\ll \lambda r^{-2\eta} t^\kappa N^3.$$

Consequently

$$\int_{-\lambda}^{\lambda} E(a) da \ll \left(\frac{r}{t} N + \lambda r^{-2\eta} t^\kappa \right) N^3. \quad (\text{SP.76})$$

By (SP.72), $\lambda \geq r^\alpha$, and $t = r^\beta$,

$$\frac{r}{t} N \leq r^{1-\beta-\sigma-\eta}, \quad \lambda r^{-2\eta} t^\kappa \leq \lambda r^{\beta\kappa-2\eta}.$$

On the other hand, integrating (SP.75) gives

$$\int_{-\lambda}^{\lambda} E(a) da \gg \lambda r^\eta N^3.$$

We compare the two upper-bound terms directly with this quantity. Fix constants $c_E, C_E > 0$, depending only on the displayed parameters, so that the lower bound in (SP.75) is at least $c_E r^\eta N^3$ and the upper bound in (SP.76) is at most C_E times its displayed right-hand side. Put

$$m_1 := 1 - \beta - \sigma - \alpha - 2\eta, \quad m_2 := \beta\kappa - 3\eta.$$

The choice $4\eta_0 < \min\{\beta\kappa, 1 - \sigma - \alpha - \beta\}$ and $\eta < \eta_0$ gives

$$m_1 > 2\eta > 0, \quad m_2 > \eta > 0.$$

Using $\lambda \geq r^\alpha$ and (SP.72), the two ratios to the lower-bound scale $\lambda r^\eta N^3$ therefore satisfy

$$\frac{(r/t)N}{\lambda r^\eta} \leq r^{m_1}, \quad \frac{\lambda r^{-2\eta} t^\kappa}{\lambda r^\eta} = r^{m_2}.$$

Decrease the admissible threshold for r so that

$$C_E(r^{m_1} + r^{m_2}) < c_E.$$

Then (SP.76) is strictly smaller than the integral lower bound obtained from (SP.75), a contradiction. This proves the lemma with an entirely explicit positive exponent margin. $\square$

Lemma 2.108 (Approximate coefficient intervals suffice at the original scale). *Fix κ, σ, α as in Lemma 2.107, and fix $D, B \geq 1$. There is $u_0 = u_0(B, \kappa, \sigma, \alpha) > 0$ with the following property. For every $0 < u < u_0$ there is $r_0 = r_0(D, B, \kappa, \sigma, \alpha, u) > 0$ such that there do not exist $0 < r < r_0$ and an r -separated set $X \subset [-r^{-u}, r^{-u}]$ satisfying*

$$N_t(X) \geq r^u t^{-\kappa} \quad (r \leq t \leq 1), \quad N_r(X) \leq r^{-\sigma-u},$$

$$N_r(X + X) \leq r^{-u} N_r(X),$$

and, for every $t \in [-r^\alpha, r^\alpha]$, a coefficient w_t satisfying

$$|t - w_t| \leq Dr, \quad N_r(X + w_t X) \leq r^{-Bu} N_r(X). \quad (\text{SP.77})$$

Proof. Since $X \subset [-r^{-u}, r^{-u}]$, the coefficient perturbation in (SP.77) moves $w_t X$ by at most Dr^{1-u} . An interval of radius Dr^{1-u} is covered by $O_D(r^{-u})$ intervals of radius r . Hence

$$N_r(X + tX) \ll_D r^{-(B+1)u} N_r(X) \quad (|t| \leq r^\alpha).$$

Choose u_0 so small that

$$(B+3)u_0 < \eta_0(\kappa, \sigma, \alpha).$$

For the now fixed $0 < u < u_0$, decrease $r_0(D, B, \kappa, \sigma, \alpha, u)$ so that every small-scale threshold used below holds and the implied constant above is at most r^{-u} whenever $0 < r < r_0$. Put $\eta = (B+3)u$. Because $0 < r < 1$ and $\eta \geq u$, the Frostman and size hypotheses imply

$$N_t(X) \geq r^\eta t^{-\kappa}, \quad N_r(X) \leq r^{-\sigma-\eta}.$$

Moreover

$$N_r(X + X) \leq r^{-u} N_r(X) \leq r^{-\eta} N_r(X)$$

and, after the preceding absorption,

$$N_r(X + tX) \leq r^{-(B+2)u} N_r(X) < r^{-\eta} N_r(X) \quad (|t| \leq r^\alpha).$$

Lemma 2.107 gives the contradiction. $\square$

A self-contained one-dimensional ring-saturation module The coefficient-generation input for the scalar-action theorem follows from a one-dimensional saturation argument. To preserve the backward-only dependency order, we reproduce that argument here rather than referring forward to Volume XIV.

Definition 2.109 (Bounded-complexity signed ring words). For a set $S \subset \mathbb{R}$, define $\langle S \rangle_s$ recursively by

$$\langle S \rangle_1 := S \cup (-S) \cup \{0\},$$

and, for $s \geq 2$,

$$\langle S \rangle_s := \bigcup_{1 \leq j < s} \langle S \rangle_j \cup \bigcup_{\substack{i, j \geq 1 \\ i+j \leq s}} (\langle S \rangle_i + \langle S \rangle_j \cup \langle S \rangle_i - \langle S \rangle_j \cup \langle S \rangle_i \langle S \rangle_j).$$

Thus $\langle S \rangle_s$ consists of values of expressions of complexity at most s , where signed elements of S and the distinguished constant 0 have complexity one and the complexity of a binary combination is the sum of the two child complexities. Equivalently, apart from the harmless primitive 0, one may count signed leaves from S . In particular, if $x \in \langle S \rangle_i$ and $y \in \langle S \rangle_j$, then $x \pm y, xy \in \langle S \rangle_{i+j}$.

For a finite set $F \subset \mathbb{R}$ put

$$F^{(d)} = \{x_1 \cdots x_d : x_i \in F\}, \quad sF^{(d)} = \underbrace{F^{(d)} + \cdots + F^{(d)}}_{s \text{ summands}}.$$

Every lemma in the following module is proved locally.

Definition 2.110 (Discrete Frostman set). Fix $0 < \kappa < 1$ and $C_0 \geq 1$. A finite h -separated set $F \subset [-2, 2]$ is called (h, κ, C_0) -Frostman if

$$|F \cap I| \leq C_0 |I|^\kappa |F| \tag{SP.78}$$

for every interval I of length $|I| \geq h$, and

$$|F| \geq C_0^{-1} h^{-\kappa}. \tag{SP.79}$$

The first ingredient is a self-contained projection lemma. Euclidean Plancherel is part of the baseline real-analysis background declared in Volume I; the Riesz-energy and projection deductions themselves are proved here. We use the Fourier transform normalization

$$\widehat{\nu}(\xi) = \int e^{-2\pi i x \cdot \xi} d\nu(x).$$

Lemma 2.111 (Riesz-energy identity at exponent one). *Let $n \geq 2$ and let ν be a compactly supported probability measure on $\mathbb{R}^n$ with finite energy*

$$I_1(\nu) = \iint \frac{d\nu(x)d\nu(y)}{|x-y|} < \infty.$$

Then, with a constant $c_n > 0$ depending only on n ,

$$I_1(\nu) = c_n \int_{\mathbb{R}^n} |\widehat{\nu}(\xi)|^2 |\xi|^{-(n-1)} d\xi. \quad (\text{SP.80})$$

The identity is understood in $[0, \infty]$; the finite-energy hypothesis makes both sides finite.

Proof. For $r > 0$ one has the elementary Gaussian representation

$$r^{-1} = 2 \int_0^\infty e^{-\pi t^2 r^2} dt.$$

Indeed the change of variables $u = tr$ reduces the right hand side to $2 \int_0^\infty e^{-\pi u^2} du = 1$. Tonelli therefore gives directly, for the finite measure $\nu \times \nu$,

$$I_1(\nu) = 2 \int_0^\infty \iint e^{-\pi t^2 |x-y|^2} d\nu(x)d\nu(y) dt.$$

We next verify the Fourier identity used for the Gaussian kernel instead of appealing to a separate positive-definite-kernel theorem. In one dimension put

$$G_t(\xi) = \int_{\mathbb{R}} e^{-\pi t^2 x^2} e^{-2\pi i x \xi} dx.$$

Differentiation under the integral and one integration by parts give

$$G'_t(\xi) = -2\pi i \int x e^{-\pi t^2 x^2} e^{-2\pi i x \xi} dx = -\frac{2\pi \xi}{t^2} G_t(\xi).$$

Since $G_t(0) = t^{-1}$ by the Gaussian integral, the differential equation yields $G_t(\xi) = t^{-1} e^{-\pi \xi^2 / t^2}$. Tensorizing over the n coordinates gives

$$e^{-\pi t^2 |z|^2} = t^{-n} \int_{\mathbb{R}^n} e^{-\pi |\xi|^2 / t^2} e^{2\pi i z \cdot \xi} d\xi.$$

The right-hand side is absolutely integrable. Hence Fubini against the finite measure $\nu \times \nu$, followed by the definition of $\widehat{\nu}$, gives

$$\begin{aligned} \iint e^{-\pi t^2 |x-y|^2} d\nu(x)d\nu(y) &= t^{-n} \int_{\mathbb{R}^n} e^{-\pi |\xi|^2 / t^2} \left(\int e^{2\pi i x \cdot \xi} d\nu(x) \right) \left(\int e^{-2\pi i y \cdot \xi} d\nu(y) \right) d\xi \\ &= t^{-n} \int_{\mathbb{R}^n} e^{-\pi |\xi|^2 / t^2} |\widehat{\nu}(\xi)|^2 d\xi. \end{aligned}$$

Tonelli applies once more because the integrand is nonnegative. For fixed $\xi \neq 0$, the substitution $s = |\xi|/t$ evaluates the t -integral as $c_n |\xi|^{-(n-1)}$. This proves (SP.80) directly, with no smoothing or monotonicity assertion required. $\square$

Lemma 2.112 (Marstrand projection lower bound, proved in the required form). *Let $n \geq 2$, let ν be as in Lemma 2.111, and put $K = \text{supp } \nu$. Let $\Omega \subset S^{n-1}$ be any spherical cap of positive surface measure. Then there is $\theta \in \Omega$ such that*

$$|\pi_\theta(K)| \geq c_{n,\Omega} I_1(\nu)^{-1}, \quad \pi_\theta(x) = \theta \cdot x. \quad (\text{SP.81})$$

Here $|\cdot|$ denotes one-dimensional Lebesgue measure.

Proof. First suppose that ν has a smooth density. The projected measure $(\pi_\theta)_*\nu$ has a density f_θ , and one-dimensional Plancherel gives

$$\|f_\theta\|_2^2 = \int_{\mathbb{R}} |\widehat{\nu}(t\theta)|^2 dt.$$

Integrating over the whole sphere and using polar coordinates yields

$$\int_{S^{n-1}} \|f_\theta\|_2^2 d\theta \leq C_n \int_{\mathbb{R}^n} |\widehat{\nu}(\xi)|^2 |\xi|^{-(n-1)} d\xi \leq C_n I_1(\nu)$$

by Lemma 2.111. Hence the same upper bound holds with the sphere replaced by Ω . Since $\int f_\theta = 1$, Cauchy–Schwarz on the support of f_θ gives

$$1 \leq |\pi_\theta(K)|^{1/2} \|f_\theta\|_2, \quad |\pi_\theta(K)| \geq \|f_\theta\|_2^{-2}.$$

Therefore

$$\int_{\Omega} |\pi_\theta(K)| d\theta \geq \frac{|\Omega|^2}{\int_{\Omega} \|f_\theta\|_2^2 d\theta} \geq c_{n,\Omega} I_1(\nu)^{-1},$$

where we used Cauchy–Schwarz in the first step. Some $\theta \in \Omega$ satisfies (SP.81).

For a general ν , choose once and for all a spherical cap $\Omega' \Subset \Omega$ of positive surface measure. Fix a nonnegative C_c^∞ probability density ϕ supported in the unit ball and put $\phi_\eta(x) = \eta^{-n} \phi(x/\eta)$ and $\nu_\eta = \nu * \phi_\eta$. Then $\text{supp } \nu_\eta \subset K^{(\eta)}$ and $\widehat{\nu_\eta} = \widehat{\nu} \widehat{\phi_\eta}$. Since $|\widehat{\phi_\eta}| \leq 1$, Lemma 2.111 gives $I_1(\nu_\eta) \leq I_1(\nu)$. Apply the smooth argument on the fixed cap Ω' and choose $\theta_\eta \in \Omega'$ with

$$|\pi_{\theta_\eta}(K^{(\eta)})| \geq c_{n,\Omega'} I_1(\nu)^{-1}.$$

By compactness of $\overline{\Omega'}$, pass to a sequence $\eta_j \downarrow 0$ for which $\theta_{\eta_j} \rightarrow \theta \in \Omega$. Because K is compact, fix R_K with $K^{(1)} \subset B(0, R_K)$. Orthogonal projection depends Lipschitz-continuously on its direction on the compact sphere: for some dimensional constant C_n ,

$$\|\pi_\vartheta - \pi_{\vartheta'}\|_{\text{op}} \leq C_n |\vartheta - \vartheta'|.$$

Thus every $x \in K^{(\eta_j)}$ has a $y \in K$ with $|x - y| \leq \eta_j$, and consequently

$$\text{dist}(\pi_{\theta_{\eta_j}} x, \pi_\theta(K)) \leq \eta_j + C_n R_K |\theta_{\eta_j} - \theta| =: \varepsilon_j, \quad \varepsilon_j \rightarrow 0.$$

Hence $\pi_{\theta_{\eta_j}}(K^{(\eta_j)}) \subset (\pi_\theta(K))^{(\varepsilon_j)}$. Since $\pi_\theta(K)$ is compact, put $\bar{\varepsilon}_j := \sup_{k \geq j} \varepsilon_k \downarrow 0$. Continuity from above for the finite-measure neighborhoods $(\pi_\theta(K))^{(\bar{\varepsilon}_j)}$ gives

$$|\pi_\theta(K)| = \lim_{j \rightarrow \infty} |(\pi_\theta(K))^{(\bar{\varepsilon}_j)}| \geq c_{n,\Omega'} I_1(\nu)^{-1}.$$

This is (SP.81), after renaming the positive constant $c_{n,\Omega'}$ as $c_{n,\Omega}$. □

Lemma 2.113 (Frostman product measure has finite one-energy). *Let μ be a probability measure on $[-C, C]$ satisfying*

$$\mu(I) \leq C_1 |I|^\kappa \quad (\text{SP.82})$$

for every interval I . If $n\kappa > 1$, then

$$I_1(\mu^{\otimes n}) \leq C'(n, \kappa, C_1, C).$$

Proof. In the sup norm on $\mathbb{R}^n$, a ball of radius r is contained in a product of n intervals of length $2r$, so

$$\mu^{\otimes n}(B(x, r)) \leq C_2 r^{n\kappa}.$$

Euclidean and sup norms are comparable in fixed dimension. For each x , decompose the y -integral into dyadic shells $2^{-j-1} < |x - y| \leq 2^{-j}$ (and one bounded outer shell). The contribution of the j th small shell is at most $C 2^j 2^{-jn\kappa} = C 2^{-j(n\kappa-1)}$, which is summable because $n\kappa > 1$. Integrating over x proves the claim. $\square$

We next prove the elimination device which turns a positive-measure linear projection of a Cartesian power into a positive-measure homogeneous ring expression.

Lemma 2.114 (Blichfeldt principle). *Let $V \subset \mathbb{R}^n$ be measurable and let $L : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be invertible. There is a translation x such that*

$$\#(V \cap (L\mathbf{Z}^n + x)) \geq |\det L|^{-1} |V|.$$

Proof. Write $x = Ly$. For $y \in [0, 1]^n$,

$$\#(L^{-1}V \cap (\mathbf{Z}^n + y)) = \sum_{k \in \mathbf{Z}^n} 1_{L^{-1}V-k}(y).$$

Integrating over $[0, 1]^n$ and interchanging sum and integral gives $|L^{-1}V| = |\det L|^{-1} |V|$. Some y is at least the average. $\square$

Lemma 2.115 (Coefficient elimination). *Let $A \subset \mathbb{R}$ be compact with diameter $D > 0$. Let $v = (v_1, \dots, v_n) \in [1/2, 1]^n$, $n \geq 2$, and put*

$$\lambda = |v_1 A + \dots + v_n A| > 0.$$

There is an integer $N = N(n, D, \lambda)$ such that, after permuting the coordinates if necessary and putting

$$A_1 = NA^{(2)} - NA^{(2)},$$

one has

$$|v_1 A_1 + \dots + v_{n-1} A_1| \geq D\lambda. \quad (\text{SP.83})$$

Moreover, if $n \leq n_0$, $D_- \leq D \leq D_+$ and $\lambda_- \leq \lambda \leq \lambda_+$ with $D_-, \lambda_- > 0$, then N is bounded by a constant depending only on $(n_0, D_-, D_+, \lambda_-, \lambda_+)$.

Proof. Let $\pi(x) = v \cdot x$. Choose an orthogonal map U such that $U(e_n) = v/|v|$; then $U(e_n^\perp) = \ker \pi$. For $R \geq 1$ put

$$V_R = U(B_{n-1}(0, R) \times [-1, 1]).$$

We have

$$|V_R| = 2\omega_{n-1} R^{n-1}, \quad \pi(V_R) = [-|v|, |v|],$$

so the latter interval has length at most $2\sqrt{n}$. Choose $R \geq 1$ so that

$$\frac{2\omega_{n-1}R^{n-1}}{(2D)^n} \lambda > nD + 2\sqrt{n} + \lambda. \quad (\text{SP.84})$$

Lemma 2.114, applied to the lattice $(2D)\mathbf{Z}^n$, gives a translation z_0 for which there are distinct points

$$p_j = 2D\tau_j + z_0 \in V_R, \quad \tau_j \in \mathbf{Z}^n, \quad 1 \leq j \leq M,$$

with

$$M\lambda > nD + 2\sqrt{n}. \quad (\text{SP.85})$$

Indeed the number of points is at least the left factor in (SP.84) before multiplication by λ .

Every set $\pi(A^n + p_j)$ is a translate of $\pi(A^n)$ and hence has measure λ . Since

$$\text{diam } \pi(A^n) \leq D \sum_{r=1}^n v_r \leq nD$$

and $p_j \in V_R$, all these projected sets lie in one interval of length at most $nD + 2\sqrt{n}$. By (SP.85), two of them overlap. Thus for some $i \neq j$ there are $x, y \in A^n$ such that

$$\pi(x) = \pi(y + 2D\ell), \quad \ell := \tau_j - \tau_i \in \mathbf{Z}^n \setminus \{0\}.$$

Permute coordinates so that $\ell_n \neq 0$ and put $z = y + 2D\ell$. Then

$$v_n(z_n - x_n) = \sum_{r=1}^{n-1} v_r(x_r - z_r). \quad (\text{SP.86})$$

Because $|y_n - x_n| \leq D$ and $|\ell_n| \geq 1$,

$$|z_n - x_n| = |y_n - x_n + 2D\ell_n| \geq D. \quad (\text{SP.87})$$

For arbitrary $a_1, \dots, a_n \in A$, relation (SP.86) gives

$$(z_n - x_n) \sum_{j=1}^n v_j a_j = \sum_{r=1}^{n-1} v_r ((z_n - x_n) a_r + (x_r - z_r) a_n).$$

Consequently

$$\begin{aligned} & (z_n - x_n)(v_1 A + \dots + v_n A) \\ & \subset \sum_{r=1}^{n-1} v_r (z_n A - x_n A + x_r A - z_r A). \end{aligned} \quad (\text{SP.88})$$

Let $a_+, a_- \in A$ satisfy $D = a_+ - a_-$. Since $z_r = y_r + 2D\ell_r$, each coefficient term $2D\ell_r A$ is a signed sum of at most $2|\ell_r|$ copies of

$$DA = a_+ A - a_- A \subset A^{(2)} - A^{(2)}.$$

Hence every bracket on the right of (SP.88) is contained in $NA^{(2)} - NA^{(2)}$ provided

$$N \geq 4 + 4 \max_{1 \leq r \leq n} |\ell_r|. \quad (\text{SP.89})$$

Moreover

$$2D|\ell| = |p_j - p_i| \leq \text{diam}(V_R) \leq 2\sqrt{R^2 + 1},$$

so $\max_r |\ell_r| \leq \sqrt{R^2 + 1}/D$. We may therefore take

$$N = \left\lceil 4 + \frac{4\sqrt{R^2 + 1}}{D} \right\rceil.$$

Taking Lebesgue measure in (SP.88) and using (SP.87) yields (SP.83).

Finally, under the bounds in the last sentence of the lemma, the right side of (SP.84) is satisfied by one R depending only on $(n_0, D_-, D_+, \lambda_-)$; the displayed formula for N is then bounded in terms of the same data. This proves the uniformity assertion. $\square$

Proposition 2.116 (Positive-measure homogeneous ring expansion). *Fix $0 < \kappa < 1$ and $C_1, C \geq 1$. There are integers $d, s \geq 1$ and a constant $c > 0$, depending only on (κ, C_1, C) , with the following property. If μ is a probability measure supported on $[-C, C]$ satisfying (SP.82), and $K = \text{supp } \mu$, then*

$$|sK^{(d)} - sK^{(d)}| \geq c. \quad (\text{SP.90})$$

Proof. Let n be the least integer for which $n\kappa > 1$. By Lemma 2.113, $I_1(\mu^{\otimes n}) \leq C_2$. Choose a small spherical cap around the vector $(1, \dots, 1)/\sqrt{n}$. Lemma 2.112 gives a direction in that cap for which the projection of K^n has measure at least $c_2 > 0$. Multiplying all projection coefficients by one common positive scalar, we may write the resulting set as

$$v_1 K + \dots + v_n K, \quad v_j \in [1/2, 1],$$

with measure at least $c_3 > 0$.

The Frostman condition also gives a uniform lower bound for $\text{diam } K$: choose $r = (4C_1)^{-1/\kappa}$; no interval of length $2r$ can carry all the mass. Thus $\text{diam } K \geq r$, while $K \subset [-C, C]$ gives $\text{diam } K \leq 2C$.

We now record the finite elimination recursion quantitatively. Put

$$K_0 := K, \quad m_j := n - j, \quad \underline{\lambda}_0 := c_3, \quad M_0 := C.$$

At stage j , suppose that

$$K_j \subset [-M_j, M_j], \quad |v_1 K_j + \dots + v_{m_j} K_j| =: \lambda_j \geq \underline{\lambda}_j > 0. \quad (\text{SP.91})$$

Then, with $D_j := \text{diam } K_j$,

$$\frac{\underline{\lambda}_j}{m_j} \leq \frac{\lambda_j}{m_j} \leq D_j \leq 2M_j, \quad \underline{\lambda}_j \leq \lambda_j \leq 2m_j M_j. \quad (\text{SP.92})$$

Apply the uniform clause of Lemma 2.115 with

$$D_- := \frac{\underline{\lambda}_j}{m_j}, \quad D_+ := 2M_j, \quad \lambda_- := \underline{\lambda}_j, \quad \lambda_+ := 2m_j M_j.$$

It gives an integer N_j bounded by a number N_j^* determined recursively by the already fixed data (κ, C_1, C) and j . Define

$$K_{j+1} := N_j K_j^{(2)} - N_j K_j^{(2)}, \quad \underline{\lambda}_{j+1} := \frac{\underline{\lambda}_j^2}{m_j}, \quad M_{j+1} := 2N_j^* M_j^2.$$

The elimination conclusion and (SP.92) give

$$|v_1 K_{j+1} + \cdots + v_{m_j-1} K_{j+1}| \geq D_j \lambda_j \geq \frac{\lambda_j^2}{m_j} = \lambda_{j+1},$$

while $K_j^{(2)} \subset [-M_j^2, M_j^2]$ gives $K_{j+1} \subset [-M_{j+1}, M_{j+1}]$. Hence the induction hypothesis (SP.91) propagates from j to $j+1$. Because there are only $n-1$ stages, all numbers N_j^*, M_j are finite and all λ_j are strictly positive, with bounds depending only on (κ, C_1, C) .

It remains to track homogeneity. If $K_j = s_j K^{(d_j)} - s_j K^{(d_j)}$, then every product of two elements of K_j expands, by distributivity, as a signed sum of at most $4s_j^2$ monomials of degree $2d_j$ in K . Therefore K_{j+1} has the form

$$K_{j+1} = s_{j+1} K^{(2d_j)} - s_{j+1} K^{(2d_j)}$$

for an integer s_{j+1} bounded in terms of N_j^* and s_j . Starting with $d_0 = 1$ shows that after $n-1$ steps the final degree is $d = 2^{n-1}$ and the final number of summands is bounded in terms of the fixed parameters. At that stage only v_1 remains, and

$$|v_1 K_{n-1}| \geq \lambda_{n-1}.$$

Since $v_1 \leq 1$, this implies $|K_{n-1}| \geq \lambda_{n-1} > 0$. Taking $c := \lambda_{n-1}$ proves (SP.90). $\square$

We now turn positive Lebesgue measure into a quantitative approximate interval. The finite Fourier argument is included because the size of the Bohr set is the source of the power separation between segment length and approximation error.

Lemma 2.117 (Bogolyubov lemma in a cyclic group). *Let $Q \geq 2$ and let $S \subset \mathbf{Z}/Q\mathbf{Z}$ have density $\alpha = |S|/Q > 0$. There is a set of frequencies $\Gamma \subset \mathbf{Z}/Q\mathbf{Z}$ with $|\Gamma| \leq 16\alpha^{-2}$ such that the Bohr set*

$$B(\Gamma, 1/6) = \left\{ x : \left\| \frac{rx}{Q} \right\|_{\mathbb{R}/\mathbf{Z}} \leq \frac{1}{6} \text{ for all } r \in \Gamma \right\}$$

is contained in $2S - 2S$.

Proof. Write $f = 1_S$ and use normalized Fourier transform

$$\widehat{f}(r) = \frac{1}{Q} \sum_x f(x) e^{-2\pi i r x / Q}.$$

Then Parseval gives $\sum_r |\widehat{f}(r)|^2 = \alpha$ and $\widehat{f}(0) = \alpha$. Put

$$\Gamma = \{r : |\widehat{f}(r)| \geq \alpha^{3/2}/4\}.$$

Parseval gives $|\Gamma| \leq 16\alpha^{-2}$. The normalized fourfold convolution $g = f * f * \widetilde{f} * \widetilde{f}$, where $\widetilde{f}(x) = f(-x)$, satisfies

$$g(x) = \sum_r |\widehat{f}(r)|^4 e^{2\pi i r x / Q}.$$

For $r \notin \Gamma$,

$$\sum_{r \notin \Gamma} |\widehat{f}(r)|^4 \leq \frac{\alpha^3}{16} \sum_r |\widehat{f}(r)|^2 = \frac{\alpha^4}{16}.$$

On the other hand the total fourth moment is at least the contribution of $r = 0$, namely α^4 . If $x \in B(\Gamma, 1/6)$, every phase on Γ has real part at least $1/2$. Therefore

$$\Re g(x) \geq \frac{1}{2} \left(\alpha^4 - \frac{\alpha^4}{16} \right) - \frac{\alpha^4}{16} > 0.$$

Thus $g(x) > 0$, which is equivalent to $x \in 2S - 2S$. $\square$

Lemma 2.118 (A cyclic Bohr set contains a long one-dimensional progression). *Let $\Gamma \subset \mathbf{Z}/Q\mathbf{Z}$ have cardinality $D \geq 1$. For all sufficiently large Q there are integers q, L such that*

$$1 \leq q \leq Q^{1/3}, \quad L \geq c_D Q^{1/(3D)}, \quad Lq \leq Q/10,$$

and

$$\{jq : |j| \leq L\} \pmod{Q} \subset B(\Gamma, 1/6).$$

If $D = 0$, take $q = 1$ and $L = \lfloor Q/10 \rfloor$.

Proof. Assume $D \geq 1$ and enumerate $\Gamma = \{r_1, \dots, r_D\}$. Let $T = \lfloor Q^{1/(3D)} \rfloor$. Consider the $T^D + 1$ points

$$k \left(\frac{r_1}{Q}, \dots, \frac{r_D}{Q} \right) \pmod{1}, \quad 0 \leq k \leq T^D,$$

in the D -torus, partitioned into T^D boxes of side $1/T$. Two lie in the same box. Their difference supplies $1 \leq q \leq T^D \leq Q^{1/3}$ with

$$\left\| \frac{qr_j}{Q} \right\|_{\mathbb{R}/\mathbf{Z}} \leq T^{-1} \quad (1 \leq j \leq D).$$

Take $L = \lfloor T/7 \rfloor$. Then $|j| \leq L$ implies the Bohr inequalities with radius $1/6$. Finally $Lq \leq CQ^{1/3+1/(3D)} \leq CQ^{2/3} < Q/10$ for Q large. $\square$

Theorem 2.119 (Internal homogeneous ring saturation). *Fix $0 < \kappa < 1$ and $C_0 \geq 1$. There exist integers $d, s \geq 1$ and a number $\beta > 0$, depending only on (κ, C_0) , such that the following holds for all sufficiently small h . If $F \subset [-2, 2]$ is (h, κ, C_0) -Frostman, then there is a number ℓ with*

$$h^{1-\beta} \leq \ell \leq C \tag{SP.93}$$

such that

$$[-\ell, \ell] \subset sF^{(d)} - sF^{(d)} + [-\ell h^\beta, \ell h^\beta]. \tag{SP.94}$$

Every monomial in (SP.94) has the same degree d .

Proof. Around each $a \in F$ place an interval of radius $h/10$, and let μ be the probability measure which gives each such interval mass $1/|F|$ with constant density on that interval. Because F is h -separated, overlap is uniformly bounded. For an interval I of length $r \geq h$, (SP.78) gives $\mu(I) \leq Cr^\kappa$. For $r < h$, at most a bounded number of the small component intervals meet I , and

$$\mu(I) \leq C \frac{r/h}{|F|} \leq CC_0 r h^{\kappa-1} = CC_0 r^\kappa (r/h)^{1-\kappa} \leq CC_0 r^\kappa.$$

Thus μ is a genuine κ -Frostman measure with a constant depending only on (κ, C_0) , supported on a fixed bounded interval. Proposition 2.116 therefore supplies fixed d, s_0 and $c > 0$ such that, writing $K = \text{supp } \mu$,

$$E = s_0 K^{(d)} - s_0 K^{(d)}$$

has Lebesgue measure at least c and lies in a fixed interval $[-M, M]$.

Put

$$Q := \left\lceil \frac{64(M+1)}{h} \right\rceil$$

and use the half-open h -grid cells $I_m = [mh, (m+1)h)$. Let

$$S_0 := \{m \in \mathbb{Z} : I_m \cap E \neq \emptyset\}, \quad S := S_0 \pmod{Q} \subset \mathbb{Z}/Q\mathbb{Z}.$$

Since $E \subset [-M, M]$, every $m \in S_0$ satisfies $|m| \leq M/h + 1$, so reduction modulo Q is injective not only on S_0 but on the whole integer interval containing $2S_0 - 2S_0$. The occupied cells cover E , hence

$$|S| = |S_0| \geq |E|/h \geq c/h.$$

For $h < 1$ we also have $Q \leq 65(M+1)/h$, and therefore

$$|S|/Q \geq \alpha := \frac{c}{65(M+1)} > 0.$$

Lemma 2.117 gives $B(\Gamma, 1/6) \subset 2S - 2S$ with $D = |\Gamma| \leq D_0 = 16\alpha^{-2}$. Lemma 2.118 then gives integers q, L such that

$$jq \in 2S - 2S \pmod{Q} \quad (|j| \leq L), \quad L \geq cQ^{1/(3D_0)}, \quad Lq \leq Q/10.$$

For such j choose $s_1, s_2, s_3, s_4 \in S_0$ with $jq \equiv s_1 + s_2 - s_3 - s_4 \pmod{Q}$. The right-hand side has absolute value at most $4(M/h + 1) < Q/8$, while $|jq| \leq Q/10$. Their difference is a multiple of Q of absolute value less than Q , hence it is zero. Thus the congruence is an honest integer identity.

For each s_i choose $e_i \in E \cap I_{s_i}$. Then

$$|(e_1 + e_2 - e_3 - e_4) - jqh| \leq 4h.$$

Consequently every jqh , $|j| \leq L$, lies within $4h$ of $2E - 2E$. Hence every point of the interval

$$[-\ell, \ell], \quad \ell = Lqh,$$

lies within $O(qh)$ of $2E - 2E$. Since $\ell/(qh) = L$, the relative error is at most $CL^{-1} \leq Ch^{1/(3D_0)}$. Also

$$\ell = Lqh \geq Lh \geq ch^{1-1/(3D_0)}.$$

Choose once and for all

$$0 < \beta < \frac{1}{3D_0},$$

for instance $\beta = 1/(4D_0)$. For small h , constants are absorbed and the preceding two displays imply (SP.93) and a relative error at most $\frac{1}{2}h^\beta$.

Finally replace each element of K occurring in an expression by a nearest point of F . All variables remain in a fixed bounded interval, and the number and degree of monomials are fixed, so the resulting change is $O(h)$. Since $h/\ell \leq Ch^{1/(3D_0)}$, this replacement is also absorbed by ℓh^β . Moreover $2E - 2E$ expands as

$$4s_0K^{(d)} - 4s_0K^{(d)}.$$

Thus (SP.94) holds with $s = 4s_0$, and homogeneity is preserved. $\square$

Remark 2.120. The scalar sum–product chapter proves an explicit one-step discretized sum–product gain. Theorem 2.119 is not a replacement for that theorem; it closes a different logical gap. A one-step gain does not, by itself, yield a bounded-complexity segment. The argument above owns the additional saturation step needed by the adjoint transfer.

Lemma 2.121 (One-dimensional regular Frostman window from a coarse covering lower bound). *Fix $0 < \theta < 1$ and $0 < \zeta < \min\{\theta, 1 - \theta\}/100$. There are constants $c_{\text{rw}} = c_{\text{rw}}(\theta, \zeta) > 0$ and $C_{\text{rw}} = C_{\text{rw}}(\theta, \zeta) \geq 1$ with the following property. Let $0 < M < N$, let $A \subset [-N, N]$, and suppose*

$$N_M(A) \geq (N/M)^\theta. \quad (\text{SP.95})$$

Then there are $N_0 \in [M, 2N]$, an interval J of radius N_0 , and an M -separated set $F \subset A \cap J$ such that, with $\gamma = \theta - 6\zeta$,

$$\log(N_0/M) \geq c_{\text{rw}} \log(N/M), \quad (\text{SP.96})$$

$$|F| \geq C_{\text{rw}}^{-1} (N_0/M)^\gamma, \quad (\text{SP.97})$$

$$\frac{|F \cap I|}{|F|} \leq C_{\text{rw}} (|I|/N_0)^\gamma \quad (M \leq |I| \leq 2N_0) \quad (\text{SP.98})$$

for every interval I .

Proof. Write $R = N/M > 1$ and let $F_* \subset A$ be a maximal M -separated set. Maximality means that the intervals of radius M centered at the points of F_* cover A . Thus $N_M(A) \leq |F_*|$ (or the same inequality with one absolute factor if the covering convention uses diameter rather than radius). Consequently (SP.95) gives

$$|F_*| \geq c_1 R^\theta \quad (\text{SP.99})$$

with an absolute $c_1 > 0$. Put $\gamma = \theta - 6\zeta$, so $0 < \gamma < \theta < 1$.

First suppose R is larger than a constant $R_0 = R_0(\theta, \zeta)$ to be chosen below. Among all closed intervals I with $M \leq |I| \leq 2N$ and $F_* \cap I \neq \emptyset$, choose J_0 maximizing

$$\mathcal{D}(I) := \frac{|F_* \cap I|}{(|I|/M)^\gamma}.$$

A maximizer exists: after restricting the center to $[-2N, 2N]$, the numerator changes only when an endpoint crosses one of the finitely many points of F_* , so it is enough to inspect finitely many endpoint configurations. Set

$$N_0 := |J_0|, \quad F := F_* \cap J_0,$$

and let J be the interval concentric with J_0 and of radius N_0 . Then $F \subset A \cap J$ and $M \leq N_0 \leq 2N$.

The root interval $[-N, N]$ is an admissible competitor. Hence, by (SP.99),

$$\frac{|F|}{(N_0/M)^\gamma} = \mathcal{D}(J_0) \geq \frac{|F_*|}{(2N/M)^\gamma} \geq c_2 R^{\theta-\gamma}. \quad (\text{SP.100})$$

In particular (SP.97) follows after enlarging C_{rw} .

Because F is M -separated inside an interval of length N_0 , $|F| \leq 1 + N_0/M \leq 2N_0/M$. Combining this with (SP.100) yields

$$(N_0/M)^{1-\gamma} \geq (c_2/2) R^{\theta-\gamma}.$$

Choose R_0 so large that the fixed factor $c_2/2$ costs at most half of the exponent after taking logarithms. Then

$$\log(N_0/M) \geq \frac{\theta - \gamma}{2(1 - \gamma)} \log R,$$

which proves (SP.96) with $c_{\text{rw}} = (\theta - \gamma)/(2(1 - \gamma))$ in the large- R case.

It remains to prove the Frostman bound. If an interval I has $M \leq |I| \leq 2N$, maximality of J_0 gives

$$|F \cap I| \leq |F_* \cap I| \leq \mathcal{D}(J_0)(|I|/M)^\gamma = |F|(|I|/N_0)^\gamma.$$

If $2N < |I| \leq 2N_0$, then necessarily $|I|/N_0 > 1$ because $N_0 \leq 2N$, and the same estimate is trivial. This proves (SP.98).

Finally suppose $1 < R < R_0$. Take $N_0 = N$, choose any point of A for F , and let $J = [-N, N]$ enlarged, if necessary, to an interval of radius N . Then (SP.96) holds for every $c_{\text{rw}} \leq 1$, while (SP.97) and (SP.98) hold after increasing C_{rw} by the fixed factor R_0^γ . Combining the two ranges proves the lemma. $\square$

Proposition 2.122 (Variable-exponent coefficient segment from multiscale nonconcentration). *Fix $\kappa > 0$. There exist constants $b_0 = b_0(\kappa) > 0$, $C_0 = C_0(\kappa) \geq 1$ and $s_0 = s_0(\kappa) \in \mathbb{N}$, and $\varepsilon_0 = \varepsilon_0(\kappa) > 0$, such that the following holds for all sufficiently small δ . If*

$$\begin{aligned} A &\subset [-\delta^{-\varepsilon_0}, \delta^{-\varepsilon_0}], \\ N_t(A) &\geq \delta^{\varepsilon_0} t^{-\kappa} \quad (\delta \leq t \leq 1), \end{aligned}$$

then for some exponent $a_0 \in [b_0, C_0]$,

$$[-\delta^{a_0}, \delta^{a_0}] \subset \langle A \cup \{1\} \rangle_{s_0} + [-\delta^{a_0+b_0}, \delta^{a_0+b_0}]. \quad (\text{SP.101})$$

Proof. Let $\theta = \min\{\kappa/4, 1/4\}$ and $\zeta = \theta/100$, and put $\gamma = \theta - 6\zeta \in (0, 1)$. Replace A by a maximal δ -separated subset; all covering lower bounds for scales at least δ change only by absolute factors, which are absorbed by decreasing ε_0 and δ .

Set $M = \delta$ and $N = 2\delta^{-\varepsilon_0}$. The fine-scale lower bound gives $N_M(A) \gg \delta^{-\kappa+\varepsilon_0}$. If ε_0 is small enough in terms of κ , then $N_M(A) \geq (N/M)^\theta$. Apply Lemma 2.121; write N_0 and F for its output, and choose $f_0 \in F$. Put

$$h = \frac{M}{N_0}, \quad G = \frac{F - f_0}{N_0} \subset [-2, 2].$$

Then G is h -separated, and (SP.97)–(SP.98) show, after an absolute change of constant, that G is an (h, γ, C_*) -Frostman set in the sense of Definition 2.110, with C_* depending only on κ . Moreover (SP.96) implies

$$h \leq C\delta^{c_{\text{rw}}} \quad (\text{SP.102})$$

for a fixed $c_{\text{rw}} > 0$.

Apply Theorem 2.119. It gives fixed $d, s, \beta > 0$ and ℓ with $h^{1-\beta} \leq \ell \leq C$ such that

$$[-\ell, \ell] \subset sG^{(d)} - sG^{(d)} + [-\ell h^\beta, \ell h^\beta].$$

Multiplying by N_0^d yields

$$[-\Lambda, \Lambda] \subset s(F - f_0)^{(d)} - s(F - f_0)^{(d)} + [-\Lambda h^\beta, \Lambda h^\beta], \quad \Lambda = N_0^d \ell. \quad (\text{SP.103})$$

Every product of d differences $f - f_0$ expands into at most 2^d signed degree- d monomials in elements of A . Hence, after increasing a fixed depth $s_1 = s_1(\kappa)$, the middle set in (SP.103) is contained in $\langle A \cup \{1\} \rangle_{s_1}$.

The segment length is never too small:

$$\Lambda \geq N_0^d h^{1-\beta} = \delta^{1-\beta} N_0^{d-1+\beta} \geq \delta^d,$$

because $N_0 \geq \delta$. Also $\Lambda \ll \delta^{-d\varepsilon_0}$. From (SP.102), after decreasing δ , $h^\beta \leq \delta^{\beta c_{\text{rw}}/2}$. Choose first

$$b_* := \frac{\beta c_{\text{rw}}}{16}$$

and then decrease ε_0 so that $d\varepsilon_0 < b_*$. If $\Lambda \leq 1$, (SP.103) has relative error at most δ^{8b_*} . If $\Lambda > 1$, restrict the left side to $[-1, 1]$; the absolute error is then at most $\delta^{-d\varepsilon_0} \delta^{8b_*} \leq \delta^{6b_*}$ for small δ . Thus, with $L = \min\{\Lambda, 1\}$,

$$\delta^d \leq L \leq 1, \quad [-L, L] \subset \langle A \cup \{1\} \rangle_{s_1} + [-L\delta^{4b_*}, L\delta^{4b_*}]. \quad (\text{SP.104})$$

Write $L = \delta^a$, $a \in [0, d]$. If $a \geq b_*$, keep this segment; if $a < b_*$, restrict it to radius δ^{b_*} . In both cases, with $a_0 = \max\{a, b_*\}$,

$$[-\delta^{a_0}, \delta^{a_0}] \subset \langle A \cup \{1\} \rangle_{s_1} + [-\delta^{a_0+b_*}, \delta^{a_0+b_*}].$$

This proves the proposition with $b_0 = b_*$, $C_0 = d$, and $s_0 = s_1$. $\square$

Lemma 2.123 (Exponent stabilization by bounded multiplicative stitching). *Fix numbers $0 < b < C$ and an integer s_0 . Suppose that, for every sufficiently small δ and every admissible coefficient set A , there is $a_0 \in [b, C]$ for which*

$$[-\delta^{a_0}, \delta^{a_0}] \subset \langle A \cup \{1\} \rangle_{s_0} + [-\delta^{a_0+b}, \delta^{a_0+b}]. \quad (\text{SP.105})$$

Then there are fixed $a_* > 0$, $b_* > 0$ and $s_* \in \mathbb{N}$, depending only on (b, C, s_0) , such that

$$[-\delta^{a_*}, \delta^{a_*}] \subset \langle A \cup \{1\} \rangle_{s_*} + [-\delta^{a_*+b_*}, \delta^{a_*+b_*}]. \quad (\text{SP.106})$$

Proof. Fix an a_0 from (SP.105). For an exponent $\omega > 0$, write $\omega \in \Omega_s$ if there is $\xi \in \langle A \cup \{1\} \rangle_s$ such that

$$\frac{1}{2}\delta^\omega \leq |\xi| \leq 2\delta^\omega, \quad [-1, 1]\xi \subset \langle A \cup \{1\} \rangle_s + [-C_s\delta^{\omega+b/2}, C_s\delta^{\omega+b/2}], \quad (\text{SP.107})$$

where C_s is allowed to depend only on s, b, C, s_0 . Apply (SP.105) to the point $\frac{3}{4}\delta^{a_0}$. It gives a ring word $\xi \in \langle A \cup \{1\} \rangle_{s_0}$ with $|\xi - \frac{3}{4}\delta^{a_0}| \leq \delta^{a_0+b}$. For sufficiently small δ ,

$$\frac{1}{2}\delta^{a_0} \leq |\xi| \leq \delta^{a_0}.$$

Thus $[-1, 1]\xi \subset [-\delta^{a_0}, \delta^{a_0}]$, and another use of (SP.105), now pointwise on that whole interval, shows

$$[-1, 1]\xi \subset \langle A \cup \{1\} \rangle_{s_0} + [-\delta^{a_0+b}, \delta^{a_0+b}].$$

Since $\delta^{a_0+b} \leq \delta^{a_0+b/2}$, this proves $a_0 \in \Omega_{s_1}$ (indeed one may take $s_1 = s_0$ after enlarging the fixed error constant).

We claim that if $\omega \in \Omega_s$, then every

$$\omega + \gamma, \quad a_0 \leq \gamma \leq a_0 + b/4, \quad (\text{SP.108})$$

belongs to $\Omega_{s'}$ for a fixed $s' = s'(s, s_0)$. Indeed, (SP.105) supplies $q \in \langle A \cup \{1\} \rangle_{s_0}$ with $|q - \delta^\gamma| \leq \delta^{a_0+b}$. Since $\gamma \leq a_0 + b/4$, this gives $|q| \asymp \delta^\gamma$. Multiply the second inclusion in (SP.107) by the actual ring word q . Because $|q| \ll \delta^\gamma$, the approximation error is $O(\delta^{\omega+\gamma+b/2})$, and $q\xi$ is a bounded-depth ring word. This proves the claim.

Induction now gives bounded depths s_k such that

$$[(k+1)a_0, (k+1)a_0 + kb/4] \subset \Omega_{s_k} \quad (k \geq 0). \quad (\text{SP.109})$$

Let

$$K_0 = \left\lceil \frac{4C}{b} \right\rceil, \quad A_* = (K_0 + 1)C.$$

For a given $a_0 \in [b, C]$, put $k = \lfloor A_*/a_0 \rfloor - 1$. Then $(k+1)a_0 \leq A_*$ and

$$0 \leq A_* - (k+1)a_0 < a_0 \leq C \leq K_0 b/4.$$

Moreover $k \geq K_0$, because $A_*/a_0 \geq K_0 + 1$, and hence $K_0 b/4 \leq kb/4$. Therefore

$$(k+1)a_0 \leq A_* \leq (k+1)a_0 + kb/4,$$

so $A_* \in \Omega_{s_k}$ by (SP.109). Finally $k \leq A_*/b$, so all possible s_k are bounded by one number $S = S(b, C, s_0)$.

Choose the corresponding ξ . For small δ , $|\xi| \geq \frac{1}{2}\delta^{A_*}$, so the smaller interval of radius $\delta^{A_*+b/8}$ lies in $[-1, 1]\xi$. The error in (SP.107) is $O(\delta^{A_*+b/2})$, which, after absorbing the fixed constant, is at most $\delta^{A_*+b/8+b/4}$ for small δ . Thus (SP.106) holds with $a_* = A_* + b/8$, $b_* = b/4$, and $s_* = S$. $\square$

Proposition 2.124 (One-step scalar coefficient-segment engine). *Fix $\kappa > 0$. There exist constants*

$$a = a(\kappa) > 1, \quad b = b(\kappa) > 0, \quad s = s(\kappa) \in \mathbb{N},$$

and $\varepsilon_* = \varepsilon_*(\kappa) > 0$ such that, for all sufficiently small $\delta > 0$, every $A \subset \mathbb{R}$ satisfying

$$\begin{aligned} A &\subset [-\delta^{-\varepsilon_*}, \delta^{-\varepsilon_*}], \\ N_t(A) &\geq \delta^{\varepsilon_*} t^{-\kappa} \quad (\delta \leq t \leq 1) \end{aligned}$$

obeys

$$[-\delta^a, \delta^a] \subset \langle A \cup \{1\} \rangle_s + [-\delta^{a+b}, \delta^{a+b}]. \quad (\text{SP.110})$$

Here the notation is that of Definition 2.109.

Proof. Apply Proposition 2.122. It supplies fixed $b_0, C_0, s_0, \varepsilon_0$ and, for each admissible A , a variable $a_0 \in [b_0, C_0]$ satisfying (SP.101). Lemma 2.123 converts this variable exponent into fixed exponents $a, b > 0$ and a fixed ring depth s . In the present application the upper variable-exponent bound is $C_0 = d \geq 1$. Inspecting the construction in Lemma 2.123, the stabilized exponent is $a = A_* + b_0/8$ with $A_* = (K_0 + 1)C_0 > C_0 \geq 1$; hence $a > 1$. Taking $\varepsilon_* \leq \varepsilon_0$ completes the proof. $\square$

Remark 2.125 (Relation with matrix-algebra sum-product). The structure of the preceding proof parallels the small-segment and fixed-exponent stabilization mechanism in He's matrix-algebra sum-product argument [4]: a regular one-dimensional block first generates a segment with a power separation between its length and error, and bounded multiplicative stitching removes dependence of the length exponent on the particular block. The one-dimensional argument above is self-contained and will be used directly in the scalar-action theorem.

Lemma 2.126 (Finite multiscale stitching of one-step coefficient segments). *Assume Proposition 2.124. Fix $0 < \alpha < 1$. Then there are $c_0 = c_0(\kappa, \alpha) > 0$, an integer $S = S(\kappa, \alpha)$, and $r_0 > 0$ such that, whenever $0 < c < c_0$, $0 < r < r_0$, and*

$$\begin{aligned} C &\subset [-r^{-c}, r^{-c}], \\ N_t(C) &\geq r^c t^{-\kappa} \quad (r \leq t \leq 1), \end{aligned}$$

we have

$$[-r^\alpha, r^\alpha] \subset \langle C \cup \{1\} \rangle_S + [-Cr, Cr] \quad (\text{SP.111})$$

with an absolute C .

Proof. Let a, b, s, ε_* be furnished by Proposition 2.124 and put

$$q := \frac{a}{a+b} \in (0, 1).$$

Choose J so that $q^{J+1} \leq \alpha$. Define working scales

$$\delta_j := r^{q^j/(a+b)}, \quad 0 \leq j \leq J.$$

Because $a > 1$, we have $a+b > 1$, so $0 < q^j/(a+b) < 1$ and consequently $r \leq \delta_j \leq 1$ for every j . Then

$$\delta_0^{a+b} = r, \quad \delta_j^{a+b} = \delta_{j-1}^a \quad (1 \leq j \leq J), \quad \delta_J^a = r^{q^{J+1}} \geq r^\alpha. \quad (\text{SP.112})$$

All δ_j are fixed powers of r . If c is chosen smaller than $\varepsilon_* \min_{0 \leq j \leq J} q^j/(a+b)$, then for every j

$$C \subset [-\delta_j^{-\varepsilon_*}, \delta_j^{-\varepsilon_*}], \quad N_t(C) \geq \delta_j^{\varepsilon_*} t^{-\kappa} \quad (\delta_j \leq t \leq 1).$$

Thus Proposition 2.124 applies to the same coefficient set C at each scale δ_j .

For $j = 0$ it gives

$$[-\delta_0^a, \delta_0^a] \subset \langle C \cup \{1\} \rangle_s + [-r, r].$$

Suppose inductively that $[-\delta_{j-1}^a, \delta_{j-1}^a]$ is contained in bounded ring words plus an $O(r)$ error. The step at δ_j writes every element of $[-\delta_j^a, \delta_j^a]$ as a depth- s ring word plus an error of size at most $\delta_j^{a+b} = \delta_{j-1}^a$. Substitute the inductive representation for that error. Addition of two bounded-depth ring words is again a bounded-depth ring word, with depth increased by a constant depending only on the chosen convention for $\langle \cdot \rangle_s$. Since $J = J(\kappa, \alpha)$ is fixed, after $J+1$ substitutions the total depth is some $S = S(\kappa, \alpha)$ and the terminal error is $O(r)$. The last relation in (SP.112) then gives (SP.111). $\square$

Theorem 2.127 (Internal scalar-action expansion). *Fix $\kappa > 0$ and $0 < \sigma < 1$. There exist constants $c = c(\kappa, \sigma) > 0$ and $r_0 = r_0(\kappa, \sigma) > 0$ such that, whenever $0 < r < r_0$ and $C, X \subset \mathbb{R}$ are r -separated sets satisfying*

$$\begin{aligned} C \cup X &\subset [-r^{-c}, r^{-c}], \\ N_\rho(C) &\geq r^c \rho^{-\kappa} & (r \leq \rho \leq 1), \\ N_\rho(X) &\geq r^c \rho^{-\kappa} & (r \leq \rho \leq 1), \\ N_r(X) &\leq r^{-\sigma-c}, \end{aligned}$$

then

$$N_r(X + X) + \max_{a \in C} N_r(X + aX) \geq r^{-c} N_r(X). \quad (\text{SP.113})$$

Proof. Fix $\kappa > 0$ and $0 < \sigma < 1$, and choose

$$0 < \alpha < \frac{1-\sigma}{4}.$$

Write $c_{\text{st}}, S, r_{\text{st}} > 0$ for the constants furnished by Lemma 2.126; let $D \geq 1$ be the fixed constant in its terminal error interval $[-Dr, Dr]$. For this depth S , Lemma 2.102 furnishes constants A_S, C_S . If

the theorem's conclusion fails with exponent $c > 0$, put $N = N_r(X)$ and $L = r^{-c}$. Then 1 and every $a \in C$ are good coefficients at scale r with loss L . Since $C \subset [-r^{-c}, r^{-c}]$ and $r < 1$, $1 + r^{-c} \leq 2r^{-c}$, so the bounded-word conclusion of Lemma 2.102 gives

$$N_r(X + wX) \leq C_S 2^{A_S} r^{-2A_S c} N$$

for every depth- S ring word w . Fix

$$B_S := 2A_S + 1.$$

After the final scale threshold is decreased so that $C_S 2^{A_S} \leq r^{-c}$, every such word satisfies

$$N_r(X + wX) \leq r^{-B_S c} N. \quad (\text{SP.114})$$

The same contradiction hypothesis gives $N_r(X + X) \leq r^{-c} N$.

Let $u_0 = u_0(B_S, \kappa, \sigma, \alpha)$ be given by Lemma 2.108, and choose once and for all

$$0 < c < \min\{c_{\text{st}}, u_0\}.$$

Now choose the theorem's $r_0 = r_0(\kappa, \sigma) > 0$ smaller than r_{st} , the threshold $r_0(D, B_S, \kappa, \sigma, \alpha, c)$ from that approximate-interval lemma, the fixed threshold needed to absorb $C_S 2^{A_S}$ above, and the threshold ensuring $r^{-c} \geq 2$. For $0 < r < r_0$, Lemma 2.126 supplies, for every $t \in [-r^\alpha, r^\alpha]$, a depth- S ring word w_t with $|t - w_t| \leq Dr$. Equation (SP.114) supplies its action bound. The ambient, Frostman, size, additive, and action hypotheses of Lemma 2.108 therefore all hold with $u = c$ and loss exponent B_S , giving the required contradiction. $\square$

Corollary 2.128 (Scalar action from the internal segment engine). *Theorem 2.127 is the unconditional output of Proposition 2.124, Lemma 2.126, and Lemma 2.108.*

Proof. Proposition 2.124 supplies the fixed one-step coefficient segment. Lemma 2.126 iterates it through a bounded list of power scales and produces an $O(r)$ -approximation to $[-r^\alpha, r^\alpha]$. Under failure of scalar-action growth, Lemma 2.102 makes every bounded-depth coefficient used by that approximation good with a fixed power loss. These are precisely the hypotheses of Lemma 2.108, which yields the contradiction proved in Theorem 2.127. $\square$

Proposition 2.129 (Direct scalar-action consequence of the segment engine). *Proposition 2.124, followed by the finite stitching Lemma 2.126 and the short-interval energy Lemmas 2.107–2.108, implies the full scalar-action theorem at the original scale r . In particular, this proof uses the coefficient set directly and does not require passage to a regularized subset or a coarser terminal scale.*

Proof. The implication is Corollary 2.128; all scale changes appearing in the segment construction have already been absorbed by the finite stitching and the original-scale perturbation lemma. $\square$

Remark 2.130 (Coefficient-generation architecture). The local ring-saturation module first produces a variable power-scale segment from a regular Frostman window. Lemma 2.123 then converts the variable exponent into fixed exponents by bounded multiplicative stitching. Proposition 2.124, the good-coefficient calculus, and the short-interval energy argument therefore yield scalar-action growth at the original scale. The scale-propagation lemmas above remain useful auxiliary estimates for comparing this direct route with multiscale variants of the argument.

Remark 2.131 (Scalar-action proof architecture). The proof of Theorem 2.127 has the chain

$$\begin{aligned} \text{regular Frostman window} &\implies \text{homogeneous ring saturation} \implies \text{fixed one-step segment}, \\ &\implies \text{finite multiscale stitching} \implies \text{short-interval energy contradiction}. \end{aligned}$$

Each arrow is established in the preceding lemmas. The one-dimensional action theorem in [4] provides historical context for this architecture. The projection argument below uses Theorem 2.127 as its scalar input.

2.17.1.7 Discretized line projection through dimension nine

For a line $\theta \in \mathbb{P}^{d-1}$ let π_θ denote orthogonal projection onto θ , identified isometrically with $\mathbb{R}$. For an r -separated set $E \subset B(0, 1) \subset \mathbb{R}^d$, parameters $\Delta, u > 0$, and an exponent α , define the exceptional-direction set

$$\mathcal{E}_d(E; \alpha, \Delta, u, r) := \left\{ \theta \in \mathbb{P}^{d-1} : \begin{array}{l} \text{there exists } E' \subset E \text{ with } |E'| \geq r^u |E|, \\ N_r(\pi_\theta E') < r^{-\alpha/d-\Delta} \end{array} \right\}. \quad (2.115)$$

This definition makes the robust quantifier “for every large subset” explicit: a direction is bad precisely when at least one large subset violates the desired projection bound.

Proposition 2.132 (Internal planar line-projection theorem). *For every $\alpha_0, \kappa > 0$ there exist*

$$\Delta_2, u_2, \tau_2, r_2 > 0$$

such that the following holds whenever $0 < r < r_2$. Let $E \subset B(0, 1) \subset \mathbb{R}^2$ be r -separated and let $\alpha \in (\alpha_0, 2 - \alpha_0)$ satisfy

$$r^{-\alpha+u_2} \leq |E| \leq r^{-\alpha-u_2}. \quad (2.116)$$

Assume, for every $x \in \mathbb{R}^2$ and $r < \rho < r^{\tau_2}$,

$$|E \cap B(x, \rho)| \leq r^{-u_2} \rho^\kappa |E|. \quad (2.117)$$

Let ϑ be a probability measure on $\mathbb{P}^1$ satisfying

$$\vartheta(B_{\mathbb{P}}(\theta, \rho)) \leq r^{-u_2} \rho^\kappa \quad (\theta \in \mathbb{P}^1, r < \rho < r^{\tau_2}). \quad (2.118)$$

Then

$$\vartheta(\mathcal{E}_2(E; \alpha, \Delta_2, u_2, r)) \leq r^{u_2}. \quad (2.119)$$

Definition 2.133 (Discretized additive energy). For bounded sets $A, B \subset \mathbb{R}$ and $0 < r < 1$, put

$$\omega_r(+, A \times B) := N_r(\{(a, b, a', b') \in A \times B \times A \times B : |a + b - a' - b'| \leq r\}),$$

where the covering number on $\mathbb{R}^4$ is taken for the Euclidean metric. This is the tolerant additive energy at scale r ; it counts collisions only up to the resolution at which the projection theorem is formulated.

Lemma 2.134 (Discretized Balog–Szemerédi–Gowers reduction). *There is an absolute constant $C \geq 1$ with the following property. Let $A, B \subset [-R, R]$ be bounded nonempty sets, let $0 < r < 1$, and let $K \geq 1$. If*

$$\omega_r(+, A \times B) \geq K^{-1} N_r(A)^{3/2} N_r(B)^{3/2}, \quad (2.120)$$

then there are $A' \subset A$ and $B' \subset B$ such that

$$N_r(A') \geq C^{-1}K^{-C}N_r(A), \quad N_r(B') \geq C^{-1}K^{-C}N_r(B), \quad (2.121)$$

$$N_r(A' + B') \leq CK^CN_r(A)^{1/2}N_r(B)^{1/2}. \quad (2.122)$$

The constant is independent of R ; translation and dilation therefore give the same statement on every affine line.

Proof. We reduce to the finite BSG theorem proved in Volume II and keep the quantization error explicit. Partition $\mathbb{R}$ into the half-open cells $I_j = [jr, (j+1)r)$ and put

$$A_0 = \{j : I_j \cap A \neq \emptyset\}, \quad B_0 = \{j : I_j \cap B \neq \emptyset\}.$$

The one-dimensional grid has bounded overlap, so

$$C_1^{-1}|A_0| \leq N_r(A) \leq C_1|A_0|, \quad C_1^{-1}|B_0| \leq N_r(B) \leq C_1|B_0| \quad (2.123)$$

for an absolute C_1 .

If $a \in I_i$, $b \in I_j$, $a' \in I_{i'}$, and $b' \in I_{j'}$ satisfy $|a + b - a' - b'| \leq r$, then

$$|i + j - i' - j'| \leq 5.$$

Conversely, replacing the scale r by a fixed multiple of r changes every covering number by only an absolute factor. Hence (2.120) implies

$$\#\{(i, j, i', j') \in A_0 \times B_0 \times A_0 \times B_0 : |i + j - i' - j'| \leq 5\} \geq cK^{-1}|A_0|^{3/2}|B_0|^{3/2} \quad (2.124)$$

with an absolute $c > 0$.

Set

$$\tilde{B} = \bigcup_{|h| \leq 5} (B_0 + h).$$

For every quadruple counted in (2.124), if $h = i + j - i' - j'$, then

$$i + j = i' + (j' + h), \quad j \in \tilde{B}, \quad j' + h \in \tilde{B}.$$

Thus the exact additive energy of $(A_0, \tilde{B})$ is at least one eleventh of the left-hand side of (2.124): the map $(i, j, i', j') \mapsto (i, j, i', j' + h)$ has multiplicity at most eleven because $|h| \leq 5$. Since $|\tilde{B}| \leq 11|B_0|$, the finite BSG theorem of Volume II gives $A_1 \subset A_0$ and $\tilde{B}_1 \subset \tilde{B}$ with

$$|A_1| \geq K^{-O(1)}|A_0|, \quad |\tilde{B}_1| \geq K^{-O(1)}|B_0|, \quad |A_1 + \tilde{B}_1| \leq K^{O(1)}|A_0|^{1/2}|B_0|^{1/2}.$$

Among the eleven translates $B_0 + h$, choose h_0 for which $|\tilde{B}_1 \cap (B_0 + h_0)|$ is maximal. Translating this intersection back by h_0 gives $B_1 \subset B_0$ with

$$|B_1| \geq 11^{-1}|\tilde{B}_1| \quad \text{and} \quad |A_1 + B_1| = |A_1 + (B_1 + h_0)| \leq |A_1 + \tilde{B}_1|.$$

Finally choose one representative point of A in each cell indexed by A_1 and one representative point of B in each cell indexed by B_1 . Call the resulting sets A' and B' . The bounded-overlap comparison (2.123) gives (2.121); moreover every sum in $A' + B'$ lies within $2r$ of a grid sum indexed by $A_1 + B_1$, which gives (2.122) after enlarging the absolute constant. $\square$

Lemma 2.135 (Fubini–Jensen fixed-parameter popularity). *Let (Θ, ν) be a probability space, let (Z, λ) be a finite measure space with $0 < \lambda(Z) < \infty$, and let $\{Z_\theta\}_{\theta \in \Theta}$ be a measurable family of measurable subsets of Z . If*

$$\lambda(Z_\theta) \geq K^{-1} \lambda(Z) \quad (\theta \in \Theta)$$

for some $K \geq 1$, then there exist $\theta_* \in \Theta$ and a measurable set $\Theta_* \subset \Theta$ such that

$$\nu(\Theta_*) \geq \frac{1}{2K^2} \quad (2.125)$$

and

$$\lambda(Z_{\theta_*} \cap Z_\theta) \geq \frac{\lambda(Z)}{2K^2} \quad (\theta \in \Theta_*). \quad (2.126)$$

Proof. Fubini followed by Jensen gives

$$\begin{aligned} \int_{\Theta^2} \lambda(Z_{\theta_1} \cap Z_{\theta_2}) d\nu(\theta_1) d\nu(\theta_2) &= \int_Z \left(\int_{\Theta} \mathbf{1}_{Z_\theta}(z) d\nu(\theta) \right)^2 d\lambda(z) \\ &\geq \lambda(Z)^{-1} \left(\int_Z \int_{\Theta} \mathbf{1}_{Z_\theta}(z) d\nu(\theta) d\lambda(z) \right)^2 \\ &= \lambda(Z)^{-1} \left(\int_{\Theta} \lambda(Z_\theta) d\nu(\theta) \right)^2 \\ &\geq K^{-2} \lambda(Z). \end{aligned}$$

Since every intersection has measure at most $\lambda(Z)$, the set of pairs (θ_1, θ_2) for which the intersection is at least $\lambda(Z)/(2K^2)$ has $\nu^{\otimes 2}$ -measure at least $1/(2K^2)$. Fubini once more supplies one θ_* whose section has measure at least $1/(2K^2)$; that section is Θ_* . $\square$

Corollary 2.136 (Simultaneous popularity for the two BSG factors). *Let $X, Y \subset \mathbb{R}$ be bounded, and for every $\theta \in \Theta$ let $X_\theta \subset X$ and $Y_\theta \subset Y$. Fix $r > 0$ and suppose*

$$\lambda(X_\theta^{(r)}) \lambda(Y_\theta^{(r)}) \geq K^{-1} \lambda(X^{(r)}) \lambda(Y^{(r)}) \quad (\theta \in \Theta),$$

where $S^{(r)}$ is the closed r -neighbourhood of S and λ is Lebesgue measure. Then there are θ_* and Θ_* with $\nu(\Theta_*) \geq (2K^2)^{-1}$ such that for every $\theta \in \Theta_*$,

$$\lambda(X_{\theta_*}^{(r)} \cap X_\theta^{(r)}) \lambda(Y_{\theta_*}^{(r)} \cap Y_\theta^{(r)}) \geq \frac{\lambda(X^{(r)}) \lambda(Y^{(r)})}{2K^2}.$$

Proof. Apply Lemma 2.135 to $Z = X^{(r)} \times Y^{(r)}$ and $Z_\theta = X_\theta^{(r)} \times Y_\theta^{(r)}$. The product formula for Lebesgue measure gives the displayed conclusion. $\square$

Regular bad directions and the fixed-pair reduction The next lemmas isolate the parts of the planar argument which do not use the scalar-action theorem. We record them separately so that the geometric and additive-combinatorial inputs remain visible. They are stated with independent loss parameters so that no exponent is chosen retroactively.

Definition 2.137 (Bad and regular-bad planar directions). Let $0 < r < 1$, $0 < \alpha < 2$, $\varepsilon > 0$, and let $E \subset B(0, 1) \subset \mathbb{R}^2$ be bounded. Write $N = N_r(E)$ and set

$$\mathcal{B}(E; \alpha, \varepsilon, r) := \left\{ V \in \mathbb{P}^1 : \begin{array}{l} \text{there is } F \subset E \text{ with } N_r(F) \geq r^\varepsilon N, \\ N_r(\pi_V F) \leq r^{-\alpha/2-\varepsilon} \end{array} \right\}.$$

A direction V belongs to $\mathcal{B}_{\text{reg}}(E; \alpha, \kappa, \varepsilon, r)$ if there is $F \subset E$ such that

$$N_r(F) \geq r^{2\varepsilon} N, \quad (2.127)$$

$$N_r(\pi_V F) \leq r^{-\alpha/2-\varepsilon}, \quad (2.128)$$

$$N_r(\pi_V F \cap [t - \rho, t + \rho]) \leq \rho^{\kappa/8} r^{-\alpha/2-10\varepsilon} \quad (t \in \mathbb{R}, r \leq \rho \leq 1). \quad (2.129)$$

The scalar coordinate on V used in the last display is immaterial up to an absolute constant.

Lemma 2.138 (Regular-bad extraction in the plane). *Fix $0 < \alpha < 2$ and $\kappa > 0$. There is $\varepsilon_0 = \varepsilon_0(\kappa) > 0$ such that the following holds for $0 < \varepsilon < \varepsilon_0$ and all sufficiently small r . Assume*

$$r^{-\alpha+\varepsilon} \leq N_r(E) \leq r^{-\alpha-\varepsilon}, \quad (2.130)$$

$$N_r(E \cap B(x, \rho)) \leq r^{-\varepsilon} \rho^\kappa N_r(E) \quad (x \in \mathbb{R}^2, r \leq \rho \leq 1), \quad (2.131)$$

$$\vartheta(B_{\mathbb{P}}(V, \rho)) \leq r^{-\varepsilon} \rho^\kappa \quad (V \in \mathbb{P}^1, r \leq \rho \leq 1). \quad (2.132)$$

Then

$$\vartheta(\mathcal{B}(E; \alpha, \varepsilon, r) \setminus \mathcal{B}_{\text{reg}}(E; \alpha, \kappa, \varepsilon, r)) \leq r^{2\varepsilon}. \quad (2.133)$$

One may take any $\varepsilon_0 < \kappa/200$.

Proof. Put $N = N_r(E)$ and suppose that $V \in \mathcal{B}(E; \alpha, \varepsilon, r) \setminus \mathcal{B}_{\text{reg}}$. Choose a witness $F \subset E$ for badness. Let

$$T_V := \left\{ t \in V : N_r(F \cap \pi_V^{-1}([t - r, t + r])) \geq r^{\alpha/2+3\varepsilon} N \right\}$$

and put $F_1 = F \cap \pi_V^{-1}(T_V^{(2r)})$. A maximal r -separated subset of $\pi_V F$ and bounded overlap of the corresponding fibers give

$$N_r(F \setminus F_1) \ll N_r(\pi_V F) r^{\alpha/2+3\varepsilon} N \ll r^{2\varepsilon} N.$$

For small r , $r^\varepsilon - Cr^{2\varepsilon} \geq r^{2\varepsilon}$; hence $N_r(F_1) \geq r^{2\varepsilon} N$ after changing an absolute covering constant. Since V is not regular-bad, (2.129) fails for F_1 . Thus there are $t_V \in V$ and $\rho_V \in [r, 1]$ such that

$$N_r(\pi_V F_1 \cap [t_V - \rho_V, t_V + \rho_V]) > \rho_V^{\kappa/8} r^{-\alpha/2-10\varepsilon}.$$

Choose a maximal $6r$ -separated family of projected points in this interval. Every selected projected point is within $O(r)$ of T_V , so its $O(r)$ -fiber contains at least $r^{\alpha/2+3\varepsilon} N$ r -cells of F . The fibers of a $6r$ -separated family have bounded overlap. Therefore, with an absolute constant C ,

$$N_r(E \cap \pi_V^{-1}([t_V - C\rho_V, t_V + C\rho_V])) \geq C^{-1} \rho_V^{\kappa/8} r^{-8\varepsilon} N. \quad (2.134)$$

The left side is at most N , so $\rho_V \leq Cr^{64\varepsilon/\kappa}$. Absorbing C by decreasing r , all possible ρ_V lie in $[r, r^{32\varepsilon/\kappa}]$.

Assume, contrary to (2.133), that the set D of such irregular bad directions has measure at least $r^{2\varepsilon}$. Partition $[r, r^{32\varepsilon/\kappa}]$ into the finitely many intervals $[s^2, s]$ with $s = r^{2^{-j}}$ which meet that range. The number of intervals depends only on κ/ε . Hence, for all sufficiently small r , there are a scale $\rho \in [r, 1]$ and a subset $D_0 \subset D$ with $\vartheta(D_0) \geq r^{3\varepsilon}$ such that $\rho_V \in [\rho^2, \rho]$ for every $V \in D_0$. Enlarging each strip in (2.134) to width $C\rho$ and using $\rho_V^{\kappa/8} \geq \rho^{\kappa/4}$ gives

$$N_r(E \cap S_V) \geq C^{-1} \rho^{\kappa/4} r^{-8\varepsilon} N, \quad V \in D_0, \quad (2.135)$$

where S_V is a strip of width $C\rho$ perpendicular to V .

Because E is r -separated, $\lambda_2(E^{(r)}) \asymp r^2 N$, and the same bounded-overlap comparison applied inside $S_V^{(Cr)}$ turns (2.135) into

$$\lambda_2(E^{(r)} \cap S_V^{(Cr)}) \geq C^{-1} \rho^{\kappa/4} r^{-8\varepsilon} \lambda_2(E^{(r)}).$$

Normalize $\vartheta|_{D_0}$ to a probability measure and apply Lemma 2.135 to these subsets of $E^{(r)}$. There is a positive-measure set of pairs $(V_1, V_2) \in D_0^2$ for which

$$\lambda_2(E^{(r)} \cap S_{V_1}^{(Cr)} \cap S_{V_2}^{(Cr)}) \geq C^{-1} \rho^{\kappa/2} r^{-16\varepsilon} \lambda_2(E^{(r)}). \quad (2.136)$$

More precisely, the conditional probability of this event is at least $C^{-1} \rho^{\kappa/2} r^{-16\varepsilon}$. On the other hand, for fixed $V_1 \in D_0$, the conditional probability that $d_{\mathbb{P}}(V_1, V_2) < \rho^{1/2}$ is at most

$$\vartheta(D_0)^{-1} r^{-\varepsilon} \rho^{\kappa/2} \leq r^{-4\varepsilon} \rho^{\kappa/2}.$$

For sufficiently small r this is strictly smaller than the preceding lower bound. We may therefore choose a pair satisfying (2.136) and $d_{\mathbb{P}}(V_1, V_2) \geq \rho^{1/2}$.

The intersection of two width- $C\rho$ planar strips whose projective angle is at least $\rho^{1/2}$ is contained in a ball of radius $C'\rho^{1/2}$. Using (2.131) and again the bounded-overlap comparison between r -covering number and the measure of $E^{(r)}$, we obtain

$$\lambda_2(E^{(r)} \cap S_{V_1}^{(Cr)} \cap S_{V_2}^{(Cr)}) \leq C'' r^{-\varepsilon} \rho^{\kappa/2} \lambda_2(E^{(r)}).$$

This contradicts (2.136) because $r^{-16\varepsilon} \gg r^{-\varepsilon}$. The lemma follows. $\square$

Lemma 2.139 (Transverse coordinates convert a bad projection into additive energy). *There is an absolute exponent $A \geq 1$ with the following property. Let $K \geq 2$, $M \geq 1$, and $0 < r < 1$. Let $V_1, V_2 \in \mathbb{P}^1$ satisfy $d_{\mathbb{P}}(V_1, V_2) \geq K^{-1}$, and let*

$$T : \mathbb{R}^2 \longrightarrow V_1 \times V_2, \quad T(z) = (\pi_{V_1} z, \pi_{V_2} z).$$

Then T and T^{-1} have Lipschitz norm at most K^A . Suppose $A_0 \subset B(0, K)$ is r -separated and

$$K^{-1} M^2 \leq N_r(A_0) \leq K M^2, \quad (2.137)$$

and put

$$X = \pi_{V_1}(A_0), \quad Y = \pi_{V_2}(A_0).$$

Assume $N_r(X), N_r(Y) \leq KM$. Let V be a direction at projective distance at least K^{-1} from the two singular directions of the restrictions $\pi_{V|V_1}$ and $\pi_{V|V_2}$, and let $F \subset A_0$ satisfy

$$N_r(F) \geq K^{-1} M^2, \quad N_r(\pi_V F) \leq KM. \quad (2.138)$$

After identifying V_1, V_2, V with $\mathbb{R}$ by orthonormal coordinates, there is a scalar ϕ_V with $K^{-A} \leq |\phi_V| \leq K^A$ for which the tolerant additive energy

$$\mathcal{E}_{K^A r}(X, \phi_V Y) := \# \left\{ (x_1, y_1, x_2, y_2) \in X_r \times Y_r \times X_r \times Y_r : |x_1 + \phi_V y_1 - x_2 - \phi_V y_2| \leq K^A r \right\}$$

satisfies

$$\mathcal{E}_{K^A r}(X, \phi_V Y) \geq K^{-A} M^3. \quad (2.139)$$

Here X_r, Y_r are arbitrary maximal r -separated subsets of X, Y ; changing those nets changes only the power of K .

Proof. The determinant of T is the sine of the angle between V_1 and V_2 , so the hypothesis $d_{\mathbb{P}}(V_1, V_2) \geq K^{-1}$ gives the asserted $K^{O(1)}$ condition number. In the coordinates supplied by T , the scalar projection to V has the form

$$\pi_V(z) = a_V x + b_V y, \quad K^{-A} \leq |a_V|, |b_V| \leq K^A, \quad (2.140)$$

with $x = \pi_{V_1} z$, $y = \pi_{V_2} z$; the distance from the two singular directions gives the lower bounds. Divide by a_V and set $\phi_V = b_V/a_V$.

Choose a maximal $K^A r$ -separated subset P of $\pi_V F$. The projection bound in (2.138), followed by a bounded scale change, gives $|P| \leq K^A M$. Assign every r -cell of F to a nearest point of P . If n_p denotes the number assigned to p , then

$$\sum_{p \in P} n_p \geq K^{-A} M^2.$$

Cauchy–Schwarz therefore yields

$$\sum_{p \in P} n_p^2 \geq |P|^{-1} \left(\sum_{p \in P} n_p \right)^2 \geq K^{-A} M^3. \quad (2.141)$$

Every ordered pair of r -cells of F counted by the same n_p gives, after replacing their two transverse coordinates by the nearest points of X_r and Y_r , a quadruple satisfying the tolerance in the definition of $\mathcal{E}_{K^A r}$. The map from an r -cell of F to the corresponding pair in $X_r \times Y_r$ has multiplicity at most K^A , because T is K^A -bi-Lipschitz. Dividing (2.141) by this bounded multiplicity proves (2.139) after enlarging A once. $\square$

Lemma 2.140 (Fixed-pair covering–Ruzsa chain). *There is an absolute exponent $A \geq 1$ with the following property. Let $K \geq 2$, $M \geq 1$, and $0 < r < 1$. Let $X, Y \subset \mathbb{R}$ and, for every θ in an index set D , let*

$$X_\theta \subset X, \quad Y_\theta \subset Y, \quad \phi_\theta \in \mathbb{R} \setminus \{0\}.$$

Assume

$$K^{-1} M \leq N_r(X_\theta), N_r(Y_\theta) \leq K M, \quad (2.142)$$

$$K^{-1} \leq |\phi_\theta| \leq K, \quad (2.143)$$

$$N_r(X_\theta + \phi_\theta Y_\theta) \leq K M. \quad (2.144)$$

Suppose there are $\theta_* \in D$ and $D' \subset D$ such that for every $\theta \in D'$,

$$N_r(X_{\theta_*}^{(r)} \cap X_\theta^{(r)}) \geq K^{-1} M, \quad (2.145)$$

$$N_r(Y_{\theta_*}^{(r)} \cap Y_\theta^{(r)}) \geq K^{-1} M. \quad (2.146)$$

Put $X_* = X_{\theta_*}$ and $\phi_* = \phi_{\theta_*}$. Then

$$N_r(X_* + X_*) \leq K^A M, \quad (2.147)$$

$$N_r(X_* - \phi_\theta \phi_*^{-1} X_*) \leq K^A M \quad (\theta \in D'). \quad (2.148)$$

Proof. All scale changes below are by factors bounded by a power of K and are paid for by enlarging the final absolute exponent A . By the covering Ruzsa triangle inequality (Proposition 2.92), (2.144) and (2.142) give

$$N_r(X_\theta - X_\theta) + N_r(\phi_\theta Y_\theta - \phi_\theta Y_\theta) \leq K^{O(1)} M. \quad (2.149)$$

Let $I_\theta^X = X_*^{(r)} \cap X_\theta^{(r)}$. Applying the Ruzsa triangle inequality first to X_* and I_θ^X , and then to I_θ^X and X_θ , using (2.145) and (2.149), yields

$$N_r(X_* - X_\theta) \leq K^{O(1)}M. \quad (2.150)$$

The same argument with $I_\theta^Y = Y_*^{(r)} \cap Y_\theta^{(r)}$ gives

$$N_r(\phi_\theta Y_* - \phi_\theta Y_\theta) \leq K^{O(1)}M. \quad (2.151)$$

At the distinguished parameter, (2.144) says $N_r(X_* + \phi_* Y_*) \leq KM$. Multiplication by $\phi_\theta \phi_*^{-1}$ and (2.143) therefore give

$$N_r(\phi_\theta \phi_*^{-1} X_* + \phi_\theta Y_*) \leq K^{O(1)}M. \quad (2.152)$$

Now apply the covering Ruzsa triangle inequality successively along the chain

$$X_*, \quad X_\theta, \quad -\phi_\theta Y_\theta, \quad -\phi_\theta Y_*, \quad \phi_\theta \phi_*^{-1} X_*.$$

The four adjacent difference sets are controlled respectively by (2.150), (2.144), (2.151), and (2.152); every intermediate set has covering number at least $K^{-O(1)}M$ by (2.142)–(2.143). This proves (2.148).

Finally, applying the Ruzsa triangle inequality directly to $X_*, -\phi_* Y_*, X_*$ gives $N_r(X_* - X_*) \leq K^{O(1)}M$. The covering Plünnecke corollary 2.94, with base set X_* and the repeated summand $-X_*$, then gives

$$N_r(X_* + X_*) = N_r((-X_*) + (-X_*)) \leq K^{O(1)}M.$$

Absorbing all fixed powers into one absolute exponent A completes the proof. $\square$

Lemma 2.141 (Slope coefficients preserve projective nonconcentration). *There is an absolute $C \geq 1$ with the following property. Let $K \geq 2$ and let $V_1, V_2 \in \mathbb{P}^1$ satisfy $d_{\mathbb{P}}(V_1, V_2) \geq K^{-1}$. For a line V transverse to $V_1^\perp$ and $V_2^\perp$, define*

$$\phi_V = (\pi_{V|V_1})^{-1} \circ \pi_{V|V_2} : V_2 \rightarrow V_1.$$

Let D be a set of lines whose projective distance from each of the two singular directions of these restrictions is at least K^{-1} , fix $V_* \in D$, and identify $\text{End}(V_1)$ with $\mathbb{R}$ using orthonormal coordinates. Then D is the union of at most eight arcs on each of which

$$V \mapsto c(V) := -\phi_V \phi_{V_*}^{-1} \quad (2.153)$$

is K^C -bi-Lipschitz, and $K^{-C} \leq |c(V)| \leq K^C$.

Suppose in addition that D_0 is contained in one such arc, $\vartheta(D_0) \geq K^{-1}$, and

$$\vartheta(B_{\mathbb{P}}(V, \rho)) \leq K\rho^\kappa \quad (V \in \mathbb{P}^1, r \leq \rho \leq 1).$$

If C_0 is a maximal r -separated subset of $c(D_0)$, then for all $r \leq \rho \leq 1$,

$$N_\rho(C_0) \geq K^{-C(1+\kappa)}\rho^{-\kappa}. \quad (2.154)$$

Proof. Choose a linear isomorphism sending V_1 and V_2 to the coordinate axes. Its condition number is $O(K)$ because the sine of the angle between the two lines is at least K^{-1} . The induced projective map and its inverse are therefore $K^{O(1)}$ -Lipschitz. In coordinate axes a line in the allowed region is the graph of a slope s with $K^{-O(1)} \leq |s| \leq K^{O(1)}$, and ϕ_V is multiplication by s . On each

component obtained after cutting at the two singular directions and at the sign changes, the map $s \mapsto -s/s_*$ and its inverse have derivatives between two fixed powers of K . Eight arcs suffice. This proves (2.153) and the size bounds after increasing the absolute exponent C .

Cover $c(D_0)$ by $N = N_\rho(c(D_0))$ intervals of radius ρ . On a bi-Lipschitz arc, the inverse image of each such interval is contained in a projective ball of radius $K^C \rho$. If $K^C \rho \leq 1$, the direction nonconcentration estimate gives

$$K^{-1} \leq \vartheta(D_0) \leq NK^{1+C\kappa} \rho^\kappa,$$

so $N \geq K^{-2-C\kappa} \rho^{-\kappa}$. If $K^C \rho > 1$, the right side of (2.154) is less than one after increasing C , while $N \geq 1$. A maximal r -separated subset changes covering numbers only by an absolute factor. This proves (2.154). $\square$

Lemma 2.142 (A regular projection gives the scalar Frostman lower bound). *Let $0 < r < 1$, $K \geq 2$, $M \geq 1$, and $0 < \gamma \leq 1$. Suppose $X_0 \subset X \subset \mathbb{R}$ satisfy*

$$N_r(X_0) \geq K^{-1}M, \quad (2.155)$$

$$N_r(X \cap [t - \rho, t + \rho]) \leq K\rho^\gamma M \quad (t \in \mathbb{R}, r \leq \rho \leq 1). \quad (2.156)$$

Then, for every $r \leq \rho \leq 1$,

$$N_\rho(X_0) \geq C^{-1}K^{-2}\rho^{-\gamma}, \quad (2.157)$$

where C is absolute. In particular a subset retaining a polynomial fraction of a regular bad projection retains the lower covering hypothesis required by the scalar-action theorem, with exponent γ .

Proof. Cover X_0 by $N_\rho(X_0)$ intervals of radius ρ . After enlarging each interval by an absolute factor, (2.156) bounds the number of r -cells of X_0 in each interval by $CK\rho^\gamma M$. Consequently

$$K^{-1}M \leq N_r(X_0) \leq CK\rho^\gamma MN_\rho(X_0),$$

which is (2.157). $\square$

Lemma 2.143 (Harmless completion of truncated scale hypotheses). *Let $0 < r < 1$, $0 < \tau \leq 1$, $\kappa > 0$ and $u, c > 0$.*

(i) *If $\tau\kappa \leq u$ and a bounded nonempty set A satisfies*

$$N_r(A \cap B(x, \rho)) \leq r^{-u} \rho^\kappa N_r(A) \quad (r \leq \rho < r^\tau),$$

then the same inequality holds for every $r \leq \rho \leq 1$.

(ii) *If $\tau\kappa \leq u$ and a probability measure ϑ on a metric space satisfies*

$$\vartheta(B(y, \rho)) \leq r^{-u} \rho^\kappa \quad (r \leq \rho < r^\tau),$$

then the same inequality holds for every $r \leq \rho \leq 1$. The identical conclusion holds when balls are replaced by neighborhoods of any fixed family of closed sets, for example projective hyperplanes.

(iii) *If C is nonempty, $c \geq \tau\kappa$, and*

$$N_\rho(C) \geq r^c \rho^{-\kappa} \quad (r \leq \rho < r^\tau),$$

then the same lower bound holds for every $r \leq \rho \leq 1$.

Thus every truncated-scale hypothesis appearing in the planar selector can be completed to the full interval of scales by choosing τ after the loss exponent, with $\tau\kappa \leq \min\{u, c\}$.

Proof. If $\rho \geq r^\tau$ and $\tau\kappa \leq u$, then

$$r^{-u}\rho^\kappa \geq r^{-u+\tau\kappa} \geq 1.$$

The right sides in (i) and (ii) therefore dominate respectively $N_r(A)$ and the total mass 1, so the asserted upper bounds are trivial at the coarser scales. For (iii), if $\rho \geq r^\tau$ and $c \geq \tau\kappa$, then

$$r^c\rho^{-\kappa} \leq r^{c-\tau\kappa} \leq 1,$$

whereas $N_\rho(C) \geq 1$ because C is nonempty. This proves all three claims. $\square$

Quantified slice geometry

Lemma 2.144 (Well-spaced line frames and product covering). *Fix an integer $2 \leq j \leq 9$. Let $V_1, \dots, V_j$ be lines in $\mathbb{R}^d$, $d \geq j$, and assume that for some $0 < s \leq 1$,*

$$d_{\mathbb{P}}(V_i, \mathbb{P}(V_1 + \dots + V_{i-1})) \geq s \quad (2 \leq i \leq j). \quad (2.158)$$

Put $U = V_1 + \dots + V_j$ and

$$T_U : U \longrightarrow V_1 \times \dots \times V_j, \quad T_U(x) = (\pi_{V_1}x, \dots, \pi_{V_j}x).$$

Then T_U is invertible and

$$\|T_U\| + \|T_U^{-1}\| \leq C_d s^{-(j-1)}. \quad (2.159)$$

Consequently, for every bounded $A \subset \mathbb{R}^d$ and $0 < r < 1$,

$$N_r(\pi_U A) \leq C_d s^{-C_d} \prod_{i=1}^j N_r(\pi_{V_i} A). \quad (2.160)$$

Moreover, if intervals $I_i \subset V_i$ have radius at most ρ , then

$$\bigcap_{i=1}^j \pi_{V_i}^{-1}(I_i) \cap U \subset B_U(x_0, C_d s^{-(j-1)}\rho) \quad (2.161)$$

for some $x_0 \in U$ whenever the intersection is nonempty.

Proof. Choose unit vectors e_i spanning V_i and let $G = (\langle e_i, e_k \rangle)$ be their Gram matrix. Applying Gram–Schmidt in the displayed order, (2.158) says that the length of the orthogonal component of e_i outside $\text{span}(e_1, \dots, e_{i-1})$ is at least s . Hence the absolute determinant of the change-of-basis matrix from the orthonormal Gram–Schmidt frame to $(e_1, \dots, e_j)$ is at least s^{j-1} . All its entries are bounded by one. Cramer’s rule therefore bounds the inverse matrix by $C_d s^{-(j-1)}$. In orthonormal coordinates on U , the matrix of T_U has rows $e_1^\top, \dots, e_j^\top$, so its determinant is the same change-of-basis determinant. Cramer’s rule therefore gives (2.159) directly.

Choose maximal r -separated nets $P_i \subset \pi_{V_i} A$. The Cartesian product $P_1 \times \dots \times P_j$ covers $T_U(\pi_U A)$ at radius $O_d(r)$. Pulling this cover back by T_U^{-1} enlarges the radius by at most $C_d s^{-(j-1)}$. A ball of that radius in the j -dimensional space U is covered by at most $C_d s^{-C_d}$ balls of radius r . This proves (2.160). Finally the image under T_U of the intersection in (2.161) lies in a product of j intervals of radius ρ ; applying (2.159) to two points of the intersection proves the last assertion. $\square$

Definition 2.145 (Regular bad directions in dimension d). Fix $2 \leq d \leq 9$, $0 < \alpha < d$, $\kappa > 0$, $0 < r < 1$ and $\varepsilon > 0$. For an r -separated $A \subset B(0, 1) \subset \mathbb{R}^d$ write $N = N_r(A)$. A line $V \in \mathbb{P}^{d-1}$ is ε -bad for A if there is $F \subset A$ such that

$$N_r(F) \geq r^\varepsilon N, \quad N_r(\pi_V F) \leq r^{-\alpha/d-\varepsilon}. \quad (2.162)$$

It is regular bad if there is such a witness with $N_r(F) \geq r^{2\varepsilon} N$ and, with

$$\kappa_d := \frac{\kappa}{4d^2},$$

$$N_r(\pi_V F \cap [t - \rho, t + \rho]) \leq \rho^{\kappa_d} r^{-\alpha/d-10d\varepsilon} \quad (t \in V, r \leq \rho \leq 1). \quad (2.163)$$

The corresponding sets are denoted $\mathcal{B}_d(A; \alpha, \varepsilon, r)$ and $\mathcal{B}_{d,\text{reg}}(A; \alpha, \kappa, \varepsilon, r)$.

Lemma 2.146 (Regular-bad extraction in dimensions $2 \leq d \leq 9$). Fix $2 \leq d \leq 9$, $0 < \alpha < d$ and $\kappa > 0$. There is $\varepsilon_d = \varepsilon_d(d, \kappa) > 0$ such that, whenever $0 < \varepsilon < \varepsilon_d$ and r is sufficiently small, the following holds. Suppose the r -separated set $A \subset B(0, 1)$ and the probability measure ϑ on $\mathbb{P}^{d-1}$ satisfy

$$r^{-\alpha+\varepsilon} \leq N_r(A) \leq r^{-\alpha-\varepsilon}, \quad (2.164)$$

$$N_r(A \cap B(x, \rho)) \leq r^{-\varepsilon} \rho^\kappa N_r(A) \quad (r \leq \rho \leq 1), \quad (2.165)$$

$$\vartheta\{V : d_{\mathbb{P}}(V, H) < \rho\} \leq r^{-\varepsilon} \rho^\kappa \quad (r \leq \rho \leq 1) \quad (2.166)$$

for every $x \in \mathbb{R}^d$ and every projective hyperplane $H \subset \mathbb{P}^{d-1}$. Then

$$\vartheta(\mathcal{B}_d(A; \alpha, \varepsilon, r) \setminus \mathcal{B}_{d,\text{reg}}(A; \alpha, \kappa, \varepsilon, r)) \leq r^{2\varepsilon}. \quad (2.167)$$

One may take, for example, $\varepsilon_d < \kappa/(10^4 d^4)$.

Proof. Put $N = N_r(A)$ and let $V \in \mathcal{B}_d \setminus \mathcal{B}_{d,\text{reg}}$. Choose a witness F as in (2.162). Call an r -cell of $\pi_V F$ heavy when its Cr -fiber contains at least $r^{\alpha/d+3\varepsilon} N$ occupied r -cells of F , and let F_1 be the part of F over the heavy cells. Since $N_r(\pi_V F) \leq r^{-\alpha/d-\varepsilon}$, the light part has covering number $O(r^{2\varepsilon} N)$. Thus, after an absolute scale adjustment,

$$N_r(F_1) \geq r^{2\varepsilon} N. \quad (2.168)$$

Because V is not regular bad, there are $t_V \in V$ and $\rho_V \in [r, 1]$ for which (2.163) fails for F_1 . Choose a maximal $6r$ -separated family in the offending projected interval. Each selected cell is heavy and the corresponding fibers have bounded overlap. Hence, with $D_d := 10d - 3$,

$$N_r(A \cap \pi_V^{-1}([t_V - C\rho_V, t_V + C\rho_V])) \geq C_d^{-1} \rho_V^{\kappa_d} r^{-D_d\varepsilon} N. \quad (2.169)$$

The left side is at most N , so, after decreasing r ,

$$r \leq \rho_V \leq r^{D_d\varepsilon/(2\kappa_d)}. \quad (2.170)$$

Assume contrary to (2.167) that the irregular set has measure at least $r^{2\varepsilon}$. Pigeonhole the range (2.170) into the bounded logarithmic family of intervals $[s^2, s]$. We obtain a scale $\rho \in [r, 1]$ and a set D of irregular directions with

$$\vartheta(D) \geq r^{3\varepsilon} \quad (2.171)$$

and such that, after enlarging the slabs in (2.169), every $V \in D$ has a slab S_V of width $C_d \rho$ satisfying

$$\lambda_d(A^{(r)} \cap S_V) \geq C_d^{-1} \rho^{2\kappa_d} r^{-D_d \varepsilon} \lambda_d(A^{(r)}). \quad (2.172)$$

Here we used $\rho_V \geq \rho^2$ and $\lambda_d(A^{(r)}) \asymp_d r^d N$: the lower bound follows because the $r/3$ -balls centred at the r -separated points of A are disjoint, while the upper bound follows by covering $A^{(r)}$ by the r -balls centred at those same N points.

Normalize $\vartheta|_D$ to a probability measure. Applying Fubini and Jensen $d-1$ times to the measurable sets $A^{(r)} \cap S_V$ shows that the set of d -tuples $(V_1, \dots, V_d) \in D^d$ for which

$$\lambda_d\left(A^{(r)} \cap \bigcap_{i=1}^d S_{V_i}\right) \geq c_d p^d \lambda_d(A^{(r)}), \quad p = C_d^{-1} \rho^{2\kappa_d} r^{-D_d \varepsilon}, \quad (2.173)$$

has conditional probability at least $c_d p^d$.

Set $s = \rho^{1/d}$. Conditional on $V_1, \dots, V_{j-1}$, the set of V_j with distance $< s$ from their span is contained in the s -neighborhood of a projective hyperplane. By (2.166) and (2.171), its conditional probability is at most

$$r^{-4\varepsilon} s^\kappa = r^{-4\varepsilon} \rho^{\kappa/d}. \quad (2.174)$$

Since $2d\kappa_d = \kappa/(2d) < \kappa/d$, the lower probability $c_d p^d$ in (2.173) dominates the union of the $d-1$ bounds (2.174) when r is sufficiently small (the additional powers of $r^{-\varepsilon}$ only strengthen this comparison). We may therefore choose a tuple satisfying both (2.173) and (2.158).

Apply Lemma 2.144, with $j = d$, to the chosen slabs. Their intersection is contained in a ball of radius $C_d s^{-(d-1)} \rho = C_d \rho^{1/d}$. Hence (2.165) gives the upper bound

$$\lambda_d\left(A^{(r)} \cap \bigcap_{i=1}^d S_{V_i}\right) \leq C_d r^{-\varepsilon} \rho^{\kappa/d} \lambda_d(A^{(r)}).$$

The lower bound (2.173) is $\gg_d \rho^{2d\kappa_d} r^{-dD_d \varepsilon} \lambda_d(A^{(r)})$. Because $2d\kappa_d < \kappa/d$, the two inequalities are incompatible for small r . This proves (2.167). $\square$

Lemma 2.147 (Simultaneous regular codimension-one slice). *Fix $3 \leq d \leq 9$, $0 < \alpha < d$ and $\kappa > 0$. Let $0 < \varepsilon$ be sufficiently small in terms of d, κ , and suppose that $A \subset B(0, 1)$ is r -separated and satisfies (2.164)–(2.165), while ϑ satisfies (2.166). Assume the following persistent-badness hypothesis:*

$$\vartheta(\mathcal{B}_d(F; \alpha, \varepsilon, r)) > r^{2\varepsilon} \quad (2.175)$$

for every $F \subset A$ with $N_r(F) \geq r^{4d\varepsilon} N_r(A)$. Then there are a hyperplane $W \subset \mathbb{R}^d$, a translate $z + W$, an r -separated source set $A^\sharp \subset A$ contained in an $O(r)$ -slab about $z + W$, and a bounded set

$$B = z + \pi_W(A^\sharp - z) \subset A^{(Cr)} \cap (z + W)$$

such that

$$r^{-(d-1)\alpha/d + C_d \varepsilon} \leq N_r(B) \leq r^{-(d-1)\alpha/d - C_d \varepsilon} \quad (2.176)$$

and, simultaneously for every $x \in W$ and every $r \leq \rho \leq 1$,

$$N_r(B \cap B_W(x, \rho)) \leq r^{-C_d \varepsilon} \rho^{\kappa/(4d^2)} N_r(B). \quad (2.177)$$

Moreover the source-tracking map

$$p_W : A^\sharp \longrightarrow B, \quad p_W(a) = z + \pi_W(a - z),$$

has the following uniform property: for every $F \subset A^\sharp$,

$$C_d^{-1}N_r(F) \leq N_r(p_W F) \leq C_d N_r(F), \quad (2.178)$$

and $|\pi_\theta a - \pi_\theta p_W(a)| \leq C_d r$ for every line θ and every $a \in A^\sharp$. In addition one can choose lines $V_1, \dots, V_d$ with $W = V_d^\perp$ and

$$d_{\mathbb{P}}(V_j, \mathbb{P}(V_1 + \dots + V_{j-1})) \geq r^{3\varepsilon/\kappa} \quad (2 \leq j \leq d), \quad (2.179)$$

so that V_1 is regular bad for a witness containing the final nested set.

Proof. Put $N = N_r(A)$ and $v = 3\varepsilon/\kappa$. By Lemma 2.146 and (2.175), there is a regular-bad direction V_1 and a witness $A_1 \subset A$ with

$$N_r(A_1) \geq r^{2\varepsilon} N, \quad N_r(\pi_{V_1} A_1) \leq r^{-\alpha/d-\varepsilon}, \quad (2.180)$$

and satisfying the local estimate (2.163).

Inductively suppose A_{j-1} and $V_1, \dots, V_{j-1}$ have been chosen, with $2 \leq j \leq d$. At this stage $N_r(A_{j-1}) \geq r^{j\varepsilon} N$ after harmlessly weakening the exponent, so the persistent-badness hypothesis applies. The r^v -neighborhood of $\mathbb{P}(V_1 + \dots + V_{j-1})$ is contained in the r^v -neighborhood of a projective hyperplane. By (2.166), its measure is at most

$$r^{-\varepsilon} r^{v\kappa} = r^{2\varepsilon},$$

strictly smaller than the bad-direction mass in (2.175). Choose a bad V_j outside this neighborhood and let $A_j \subset A_{j-1}$ be a witness. Thus

$$N_r(A_j) \geq r^\varepsilon N_r(A_{j-1}), \quad N_r(\pi_{V_j} A_j) \leq r^{-\alpha/d-\varepsilon}, \quad (2.181)$$

and (2.179) holds. After d steps,

$$N_r(A_d) \geq r^{(d+1)\varepsilon} N \geq r^{-\alpha+C_d\varepsilon}. \quad (2.182)$$

Let $W = V_d^\perp$. Cover $\pi_{V_d} A_d$ by $O(r^{-\alpha/d-\varepsilon})$ intervals of radius r . One corresponding fiber slab therefore contains a set A'_d with

$$N_r(A'_d) \geq r^{-(d-1)\alpha/d+C_d\varepsilon}. \quad (2.183)$$

Let $z + W$ be the central affine hyperplane of that slab, set $A^\sharp := A'_d$, and define

$$B := z + \pi_W(A^\sharp - z) \subset z + W.$$

Every point of B is within Cr of the corresponding point of $A^\sharp$, so $B \subset A^{(Cr)} \cap (z + W)$. The original slab has thickness $O(r)$ in the V_d direction, so an r -ball in $z + W$ lifts to only $O_d(1)$ ambient r -balls inside that slab. The same argument applies to every subset $F \subset A^\sharp$, proving (2.178); the pointwise projection error is at most the slab thickness. In particular $N_r(B) \asymp_d N_r(A^\sharp)$, which gives the lower bound in (2.176). For the upper bound put $U = V_1 + \dots + V_{d-1}$. By Lemma 2.144, the restriction $\pi_U|_W : W \rightarrow U$ is $r^{-C_d\varepsilon}$ -bi-Lipschitz because of (2.179). Since B is within Cr of A'_d and $A_d \subset A_j$ for $1 \leq j \leq d-1$, the product-covering estimate gives

$$N_r(B) \leq r^{-C_d\varepsilon} \prod_{j=1}^{d-1} N_r(\pi_{V_j} A_j) \leq r^{-(d-1)\alpha/d-C_d\varepsilon},$$

after enlarging C_d . This completes (2.176). From now on we identify the affine hyperplane $z + W$ with W by translation.

It remains to prove the local estimate simultaneously at every scale. For $x \in W$ and $r \leq \rho \leq 1$,

$$\begin{aligned} N_r(B \cap B_W(x, \rho)) &\leq r^{-C_d \varepsilon} N_r(\pi_U(B \cap B_W(x, \rho))) \\ &\leq r^{-C_d \varepsilon} N_r(\pi_{V_1} A_1 \cap I_{C\rho}) \prod_{j=2}^{d-1} N_r(\pi_{V_j} A_j), \end{aligned}$$

where $I_{C\rho} \subset V_1$ is an interval of radius $C_d \rho$; we used $A_d \subset A_j$ for every j and the fact that B lies within Cr of A'_d , which changes the displayed covering numbers only by absolute scale factors. The regularity of V_1 and (2.181) give

$$N_r(B \cap B_W(x, \rho)) \leq r^{-C_d \varepsilon} \rho^{\kappa/(4d^2)} r^{-(d-1)\alpha/d}.$$

Divide by the lower bound (2.176) and enlarge C_d once. This is (2.177), uniformly in x and ρ . $\square$

Lemma 2.148 (Direction-law descent to a codimension-one slice). *Let $d \geq 3$, let $W \subset \mathbb{R}^d$ be a hyperplane, and let $0 < u < \kappa/4$. Suppose a probability measure ϑ on $\mathbb{P}^{d-1}$ satisfies*

$$\vartheta\{\theta : d_{\mathbb{P}}(\theta, H) < \rho\} \leq r^{-u} \rho^{\kappa} \quad (2.184)$$

for every projective hyperplane H and every $r \leq \rho \leq 1$. Fix $v \geq 2u/\kappa$ and put

$$G_v = \{\theta : d_{\mathbb{P}}(\theta, W^{\perp}) \geq r^v\}.$$

Then $\vartheta(G_v) \geq 1 - r^u$. If ν is the normalized push-forward of $\vartheta|_{G_v}$ under

$$q_W : \theta \mapsto \mathbb{R}\pi_W(e_{\theta}) \in \mathbb{P}(W),$$

where e_{θ} is any unit vector spanning θ , then for every projective hyperplane $L \subset \mathbb{P}(W)$ and $r \leq \rho \leq 1$,

$$\nu\{\theta' : d_{\mathbb{P}(W)}(\theta', L) < \rho\} \leq 2r^{-u} \rho^{\kappa} \quad (2.185)$$

for all sufficiently small r .

Proof. The complement of G_v is contained in the r^v -neighborhood of the projective hyperplane containing the point $W^{\perp}$; hence

$$\vartheta(G_v^c) \leq r^{-u} r^{v\kappa} \leq r^u.$$

Let $L = \mathbb{P}(U)$ for a codimension-one subspace $U \subset W$. For a line θ not equal to $W^{\perp}$, the elementary wedge identity

$$d_{\mathbb{P}}(\theta, \mathbb{P}(U \oplus W^{\perp})) = d_{\mathbb{P}}(\theta, W^{\perp}) d_{\mathbb{P}(W)}(q_W(\theta), L)$$

in particular implies that $d_{\mathbb{P}(W)}(q_W(\theta), L) < \rho$ forces $d_{\mathbb{P}}(\theta, \mathbb{P}(U \oplus W^{\perp})) < \rho$. Thus (2.184) bounds the unnormalized preimage by $r^{-u} \rho^{\kappa}$. Since $\vartheta(G_v) \geq 1 - r^u \geq 1/2$, normalization costs at most a factor two, proving (2.185). $\square$

Lemma 2.149 (Ambient projection versus slice projection). *Let $W \subset \mathbb{R}^d$ be a hyperplane, let $B \subset W$ be bounded, and let $\theta \in \mathbb{P}^{d-1}$ satisfy $s := d_{\mathbb{P}}(\theta, W^{\perp}) > 0$. Put $\theta' = q_W(\theta) \in \mathbb{P}(W)$. Then*

$$N_r(\pi_{\theta} B) = N_{r/s}(\pi_{\theta'}^W B) \quad (2.186)$$

up to an absolute change of the covering radius, and consequently

$$N_r(\pi_{\theta'}^W B) \ll s^{-1} N_r(\pi_{\theta} B). \quad (2.187)$$

In particular, if $s \geq r^v$, an ambient bad-projection bound transfers to the slice with a loss of at most Cr^{-v} in covering number. The same conclusion holds for every subset $B' \subset B$.

Proof. Choose a unit vector e spanning θ and write $\pi_W e = se'$ with e' a unit vector spanning θ' . For every $w \in W$,

$$\langle w, e \rangle = \langle w, \pi_W e \rangle = s \langle w, e' \rangle.$$

Thus the scalar projection of B to θ is exactly the dilation by s of its scalar projection to θ' inside W . This proves (2.186). Covering an interval of radius r/s by $O(s^{-1})$ intervals of radius r gives (2.187). $\square$

Lemma 2.150 (Finite exhaustion for the robust large-subset quantifier). *Fix $0 < r < 1$, $q > 0$, and $\varepsilon > 0$. For a bounded set F define*

$$\mathcal{E}(F; \eta) = \{\theta : \exists F' \subset F, N_r(F') \geq r^\eta N_r(F), \quad N_r(\pi_\theta F') \leq r^{-q-\eta}\}.$$

Assume that every $R \subset E$ with $N_r(R) > r^{\varepsilon/2} N_r(E)$ contains a set $G \subset R$ such that

$$\vartheta(\mathcal{E}(G; \varepsilon)) \leq r^\varepsilon. \quad (2.188)$$

Then, for all sufficiently small r ,

$$\vartheta(\mathcal{E}(E; \varepsilon/3)) \leq Cr^{\varepsilon/2}, \quad (2.189)$$

where C is absolute.

Proof. Construct successively sets $G_1, G_2, \dots$ satisfying (2.188), each inside the part of E lying outside the $2r$ -neighborhoods of the preceding G_i , until the remainder has r -covering number at most $r^{\varepsilon/2} N_r(E)$. The sets $G_i^{(r)}$ are pairwise disjoint. Put $G_0 = \bigcup_i G_i$ and

$$a_i = \frac{N_r(G_i)}{N_r(G_0)}.$$

Bounded overlap gives $N_r(G_0) \asymp \sum_i N_r(G_i)$, so after changing an absolute constant we may regard the a_i as weights summing to one.

Let $\theta \in \mathcal{E}(E; \varepsilon/3)$ and choose a witness $F \subset E$. Define

$$I = \{i : N_r(F^{(2r)} \cap G_i) \geq r^\varepsilon N_r(G_i)\}.$$

The remainder contributes at most $r^{\varepsilon/2} N_r(E)$ r -cells, and the indices outside I contribute at most $r^\varepsilon \sum_i N_r(G_i)$. Since $N_r(F) \geq r^{\varepsilon/3} N_r(E)$, for sufficiently small r it follows that

$$\sum_{i \in I} a_i \geq cr^{\varepsilon/2} \quad (2.190)$$

with an absolute $c > 0$. For every $i \in I$ the set $F^{(2r)} \cap G_i$ retains at least an r^ε fraction of G_i , and its projection is contained in an $O(r)$ -neighborhood of $\pi_\theta F$. Hence, after an absolute covering-scale adjustment, $\theta \in \mathcal{E}(G_i; \varepsilon)$.

Apply Markov's inequality to the random variable $\sum_i a_i \mathbf{1}_{\mathcal{E}(G_i; \varepsilon)}$. By (2.188) its expectation is at most r^ε , while (2.190) shows that every $\theta \in \mathcal{E}(E; \varepsilon/3)$ makes it at least $cr^{\varepsilon/2}$. Therefore

$$\vartheta(\mathcal{E}(E; \varepsilon/3)) \leq c^{-1} r^{-\varepsilon/2} r^\varepsilon = c^{-1} r^{\varepsilon/2},$$

which is (2.189). $\square$

Lemma 2.151 (Polynomial return from a tolerant scale). *Let $m \geq 1$, $0 < r < 1$, $Q \geq 1$, and let $S \subset \mathbb{R}^m$ be bounded. Then*

$$N_{Qr}(S) \leq N_r(S) \leq C_m(1+Q)^m N_{Qr}(S). \quad (2.191)$$

Consequently, if throughout a fixed finite argument the working radius is changed from r to $Q^A r$ with $Q \geq 2$ and fixed A , every covering-number estimate can be returned to radius r at a cost $Q^{O_{m,A}(1)}$ and no change of a power-law exponent is hidden.

Proof. The first inequality is monotonicity. Cover S by $N_{Qr}(S)$ balls of radius Qr . In $\mathbb{R}^m$ each such ball is covered by at most $C_m(1+Q)^m$ balls of radius r , which proves the second inequality. The last assertion follows by applying the same estimate a fixed number of times. $\square$

Definition 2.152 (Separated robust exceptional set). For $q, \zeta, \Delta > 0$ and a bounded set F put

$$\mathfrak{E}(F; q, \zeta, \Delta) := \left\{ \theta : \begin{array}{l} \text{there is } F' \subset F \text{ with } N_r(F') \geq r^\zeta N_r(F), \\ N_r(\pi_\theta F') \leq r^{-q-\Delta} \end{array} \right\}.$$

Unlike the exceptional set in (2.115), the relative-mass loss ζ and the projection gain Δ are independent. This is only bookkeeping; for an r -separated set the original exceptional set is recovered, up to absolute r -cell constants, by taking $q = \alpha/d$, $\zeta = u$, and the same projection gain Δ . Those fixed constants are harmless whenever a positive exponent gap is retained.

Lemma 2.153 (Two-parameter finite exhaustion and its persistent contrapositive). *Fix $q > 0$, $0 < \Delta_- < \Delta_+$, and $0 < \eta < 1/10$. Let E be bounded and let ϑ be a probability measure on the relevant projective space. Suppose that every $R \subset E$ satisfying*

$$N_r(R) > r^{\eta/2} N_r(E) \quad (2.192)$$

contains a set $G \subset R$ for which

$$\vartheta(\mathfrak{E}(G; q, \eta, \Delta_-)) \leq r^\eta. \quad (2.193)$$

Then, for all sufficiently small r ,

$$\vartheta(\mathfrak{E}(E; q, \eta/3, \Delta_+)) \leq Cr^{\eta/2}, \quad (2.194)$$

with an absolute C .

Equivalently, if the left side of (2.194) is larger than $Cr^{\eta/2}$, then there is an $R \subset E$ satisfying (2.192) such that

$$\vartheta(\mathfrak{E}(G; q, \eta, \Delta_-)) > r^\eta \quad \text{for every nonempty } G \subset R. \quad (2.195)$$

Proof. The proof is the weighted exhaustion argument of Lemma 2.150, with the projection threshold kept fixed. We include it because separating η from Δ is exactly what is needed in the final induction.

Choose successively $G_1, G_2, \dots$ satisfying (2.193), each inside the part of E outside the $2r$ -neighbourhoods of the preceding pieces, and stop when the remainder has r -covering number at most $r^{\eta/2} N_r(E)$. Put $G_0 = \bigcup_i G_i$ and

$$a_i = \frac{N_r(G_i)}{\sum_j N_r(G_j)}.$$

The r -neighbourhoods of the pieces have bounded overlap, so the denominator is comparable with $N_r(G_0)$ and the weights may be normalized to sum to one at an absolute cost.

Let $\theta \in \mathfrak{E}(E; q, \eta/3, \Delta_+)$ and choose a witness F . Set

$$I = \{i : N_r(F^{(2r)} \cap G_i) \geq r^\eta N_r(G_i)\}.$$

The terminal remainder contributes at most $r^{\eta/2} N_r(E)$ cells and the indices outside I contribute at most $r^\eta \sum_i N_r(G_i)$. Bounded overlap gives $\sum_i N_r(G_i) \leq C N_r(E)$ for an absolute C . Since $N_r(F) \geq r^{\eta/3} N_r(E)$, the two error-to-witness ratios are at most $r^{\eta/6}$ and $C r^{2\eta/3}$, respectively. After decreasing the scale threshold they sum to at most $1/2$; hence

$$\sum_{i \in I} a_i \geq c r^{\eta/2} \quad (2.196)$$

for small r . For $i \in I$, the set $F^{(2r)} \cap G_i$ retains an r^η fraction of G_i and its projection is contained in an $O(r)$ neighbourhood of $\pi_\theta F$. Lemma 2.151, with an absolute Q , costs only a fixed multiplicative constant in covering number. Because $\Delta_+ - \Delta_- > 0$, for sufficiently small r this constant is at most $r^{-(\Delta_+ - \Delta_-)}$. Hence the outer projection bound with gain Δ_+ implies the inner bound with gain Δ_- , and therefore $\theta \in \mathfrak{E}(G_i; q, \eta, \Delta_-)$ for all $i \in I$.

The expectation of $\sum_i a_i \mathbf{1}_{\mathfrak{E}(G_i; q, \eta, \Delta_-)}$ is at most r^η by (2.193), whereas (2.196) makes this quantity at least $c r^{\eta/2}$ on the outer exceptional set. Markov gives (2.194). The final statement is the logical contrapositive. $\square$

Lemma 2.154 (Compact uniformization of scalar action by a finite exponent grid). *Fix $0 < \alpha_0 < 1$ and $\kappa_* > 0$, and assume Theorem 2.127. There are numbers $c_{\text{unif}}, r_{\text{unif}} > 0$, depending only on (α_0, κ_*) , with the following property.*

Let

$$\theta \in [\alpha_0/2, 1 - \alpha_0/2].$$

Suppose $0 < r < r_{\text{unif}}$ and $C, X \subset \mathbb{R}$ are r -separated, contained in $[-r^{-c_{\text{unif}}}, r^{-c_{\text{unif}}}]$, and satisfy

$$N_\rho(C) \geq r^{c_{\text{unif}}} \rho^{-\kappa_*}, \quad N_\rho(X) \geq r^{c_{\text{unif}}} \rho^{-\kappa_*} \quad (r \leq \rho \leq 1), \quad (2.197)$$

$$N_r(X) \leq r^{-\theta - c_{\text{unif}}}. \quad (2.198)$$

Then

$$N_r(X + X) + \max_{a \in C} N_r(X + aX) \geq r^{-c_{\text{unif}}} N_r(X). \quad (2.199)$$

Proof. Put $h = \alpha_0/100$. Choose a finite set $\Sigma \subset (0, 1)$ with mesh at most $h/2$ such that for every θ in the displayed interval there is $\sigma \in \Sigma$ with

$$\theta + h \leq \sigma \leq \theta + 2h. \quad (2.200)$$

This is possible because $\theta + 2h \leq 1 - \alpha_0/2 + \alpha_0/50 < 1$. For each $\sigma \in \Sigma$, let c_σ, r_σ be constants from Theorem 2.127 with exponent κ_* . Take

$$c_{\text{unif}} < \min \left\{ h/10, \frac{1}{10} \min_{\sigma \in \Sigma} c_\sigma \right\}, \quad r_{\text{unif}} < \min_{\sigma \in \Sigma} r_\sigma.$$

For the σ selected by (2.200), the lower covering and boundedness assumptions with c_{unif} are stronger than those required with c_σ after decreasing r_{unif} . Moreover

$$N_r(X) \leq r^{-\theta - c_{\text{unif}}} \leq r^{-\sigma - c_\sigma},$$

because $\theta + c_{\text{unif}} < \sigma + c_\sigma$. The scalar theorem therefore gives a gain r^{-c_σ} , which is stronger than (2.199). $\square$

Proposition 2.155 (Complete planar parameter assembly from the scalar theorem). *Theorem 2.127 supplies the scalar input to Proposition 2.132. More precisely, its constants may be chosen with*

$$\Delta_2 = 6u_2, \quad 0 < u_2 \ll_{\alpha_0, \kappa} c_{\text{unif}}, \quad \tau_2 \kappa \leq u_2, \quad (2.201)$$

where c_{unif} is supplied by Lemma 2.154 with $\kappa_* = \kappa/8$.

Proof. Fix α_0, κ and put $\eta = 3u_2$. We first freeze the finite exponent ledger used below. Let $A_{\text{pre}} \geq 1$ dominate every fixed power of K produced, up to and including the fixed-pair Ruzsa stage, by Lemmas 2.139, 2.151, 2.134, Corollary 2.136, and Lemma 2.140. Let C_{sl} be the absolute exponent in Lemma 2.141 and put

$$A_{\text{sl}} := C_{\text{sl}}(A_{\text{pre}} + 2)(1 + \kappa).$$

Finally choose $A_{\text{pl}} \geq 1$ larger than $4A_{\text{pre}} + A_{\text{sl}} + C_{\text{sl}}(A_{\text{pre}} + 2)$ and than every remaining fixed exponent in the regular-to-scalar Frostman step. These are finite constants depending only on the fixed theorem parameters; no later redefinition of K is permitted. Choose

$$L > 100 \max\{A_{\text{pl}}, 1/\kappa, 1\}.$$

We then choose $u_2 > 0$ so small that

$$L^2 \eta < \min\{c_{\text{unif}}/100, \alpha_0/100, \kappa/10^4\}. \quad (2.202)$$

Finally choose τ_2 as in (2.201) and decrease r_2 whenever a proved child lemma requires it. By Lemma 2.143, the Frostman and direction hypotheses of Proposition 2.132 may from now on be used for every scale $r \leq \rho \leq 1$.

Suppose the conclusion fails. Because E is r -separated, cardinality and r -covering number are comparable by absolute constants. Thus, after decreasing r_2 and using the positive exponent gap, failure of the conclusion implies that $\mathfrak{E}(E; \alpha/2, \eta/3, 2\eta)$ has measure larger than $r^{\eta/3}$. For small r this is larger than $Cr^{\eta/2}$. The persistent contrapositive in Lemma 2.153, with $\Delta_- = \eta$ and $\Delta_+ = 2\eta$, gives a set $R \subset E$ with

$$N_r(R) > r^{\eta/2} N_r(E) \quad (2.203)$$

and such that every nonempty $G \subset R$ has

$$\vartheta(\mathfrak{E}(G; \alpha/2, \eta, \eta)) > r^\eta. \quad (2.204)$$

Indeed, if this were false, the hypothesis of the exhaustion lemma would hold. The size and local hypotheses for R with loss parameter η follow from those for E : using $u_2 = \eta/3$ and (2.203),

$$r^{-\alpha+\eta} \leq N_r(R) \leq r^{-\alpha-\eta}, \quad N_r(R \cap B(x, \rho)) \leq r^{-\eta} \rho^\kappa N_r(R)$$

after weakening inequalities, and the direction law also satisfies the $r^{-\eta} \rho^\kappa$ bound.

Apply Lemma 2.138 to R . By (2.204), after discarding at most $r^{2\eta}$ direction mass there is a regular bad direction V_1 with a witness $A_1 \subset R$ satisfying

$$N_r(A_1) \geq r^{2\eta} N_r(R), \quad N_r(\pi_{V_1} A_1) \leq r^{-\alpha/2-\eta}, \quad (2.205)$$

and

$$N_r(\pi_{V_1} A_1 \cap [t - \rho, t + \rho]) \leq \rho^{\kappa/8} r^{-\alpha/2-10\eta} \quad (r \leq \rho \leq 1). \quad (2.206)$$

Apply (2.204) to $G = A_1$. The ball of projective radius $r^{4\eta/\kappa}$ about V_1 has measure at most $r^{-\eta/3}r^{4\eta} < \frac{1}{2}r^\eta$ for small r . Hence there is a bad direction V_2 outside that ball and a witness $A_0 \subset A_1$ with

$$N_r(A_0) \geq r^\eta N_r(A_1), \quad N_r(\pi_{V_2}A_0) \leq r^{-\alpha/2-\eta}. \quad (2.207)$$

Since $A_0 \subset A_1$, the same upper bound holds for $\pi_{V_1}A_0$ and the local estimate (2.206) remains available.

Put $M = r^{-\alpha/2}$ and $K = r^{-L\eta}$. By enlarging L once, the size bounds on E, R, A_1, A_0 , the two projection bounds, and the transversality of V_1, V_2 imply

$$K^{-1}M^2 \leq N_r(A_0) \leq KM^2, \quad N_r(\pi_{V_i}A_0) \leq KM \quad (i = 1, 2), \quad d_{\mathbb{P}}(V_1, V_2) \geq K^{-1}. \quad (2.208)$$

The exponent ledger was chosen so that the coordinate map $z \mapsto (\pi_{V_1}z, \pi_{V_2}z)$ is $K^{A_{\text{pre}}}$ -bi-Lipschitz at scale r . Therefore (2.208) also gives

$$N_r(\pi_{V_i}A_0) \geq K^{-A_{\text{pre}}}M \quad (i = 1, 2),$$

after decreasing r_2 to absorb the fixed absolute covering constants.

Apply (2.204) once more, now to $G = A_0$, and let D be the resulting bad-direction set. Remove the two K^{-1} -neighbourhoods of the singular directions associated with V_1, V_2 . The direction estimate gives the explicit bound

$$2r^{-\eta/3}K^{-\kappa} = 2r^{(L\kappa-1/3)\eta}.$$

Our choice $L > 100/\kappa$ makes this at most $\frac{1}{2}r^\eta$ after decreasing r_2 . Thus the remaining set, still denoted D , has measure at least $\frac{1}{2}r^\eta$. For each $V \in D$ choose a witness $F_V \subset A_0$. By the definition of A_{pre} , the hypotheses of Lemma 2.139, followed by the polynomial scale-return lemma, yield an r -scale additive-energy lower bound with every loss between $K^{-A_{\text{pre}}}$ and $K^{A_{\text{pre}}}$. Applying Lemma 2.134 therefore gives, for every $V \in D$, sets

$$X_V \subset X := \pi_{V_1}A_0, \quad Y_V \subset Y := \pi_{V_2}A_0$$

with

$$N_r(X_V), N_r(Y_V) \geq K^{-A_{\text{pre}}}M, \quad N_r(X_V + \phi_V Y_V) \leq K^{A_{\text{pre}}}M. \quad (2.209)$$

Thickening by r turns the first two estimates into a product-density lower bound. Corollary 2.136, applied to the normalized restriction of ϑ to D , supplies one $V_* \in D$ and a set $D' \subset D$ of unnormalized measure at least $K^{-A_{\text{pre}}}r^\eta$ such that both X_{V_*} and Y_{V_*} have r -neighbourhood intersections of size at least $K^{-A_{\text{pre}}}M$ with the distinguished factors. Lemma 2.140 therefore gives, with $X_* = X_{V_*}$,

$$N_r(X_* + X_*) \leq K^{A_{\text{pre}}}M, \quad (2.210)$$

$$N_r(X_* + c(V)X_*) \leq K^{A_{\text{pre}}}M \quad (V \in D'), \quad (2.211)$$

where $c(V) = -\phi_V \phi_{V_*}^{-1}$.

Partition D' into the at most eight arcs of Lemma 2.141, and let D_0 be one arc carrying at least one eighth of its measure. Define the auxiliary bookkeeping parameter

$$K_{\text{sl}} := K^{A_{\text{pre}}+2}.$$

Because $r^\eta = K^{-1/L}$, for sufficiently small r we have

$$\vartheta(D_0) \geq \frac{1}{8}K^{-A_{\text{pre}}}r^\eta = \frac{1}{8}K^{-A_{\text{pre}}-1/L} \geq K_{\text{sl}}^{-1}.$$

Moreover the original direction estimate is $\vartheta(B(V, \rho)) \leq r^{-\eta/3} \rho^\kappa = K^{1/(3L)} \rho^\kappa \leq K_{\text{sl}} \rho^\kappa$, and all required transversality distances are at least $K^{-1} \geq K_{\text{sl}}^{-1}$. Thus Lemma 2.141, applied with K_{sl} , shows that a maximal r -separated set $C_0 \subset c(D_0)$ satisfies

$$N_\rho(C_0) \geq K^{-A_{\text{sl}}} \rho^{-\kappa} \quad (r \leq \rho \leq 1), \quad (2.212)$$

and

$$C_0 \subset [-K^{C_{\text{sl}}(A_{\text{pre}}+2)}, K^{C_{\text{sl}}(A_{\text{pre}}+2)}].$$

By (2.206), the local covering bound for $\pi_{V_1} A_1$ is at most $K^{A_{\text{pre}}} \rho^{\kappa/8} M$, since $r^{-10\eta} = K^{10/L} \leq K^{A_{\text{pre}}}$. The set $X_* \subset \pi_{V_1} A_1$ retains at least $K^{-A_{\text{pre}}} M$ r -cells. Lemma 2.142, applied with parameter $K^{A_{\text{pre}}}$, therefore gives after absorbing its absolute constant into one further power of K ,

$$N_\rho(X_*) \geq K^{-3A_{\text{pre}}} \rho^{-\kappa/8} \quad (r \leq \rho \leq 1). \quad (2.213)$$

Also

$$K^{-A_{\text{pre}}} r^{-\alpha/2} \leq N_r(X_*) \leq K^{A_{\text{pre}}} r^{-\alpha/2}. \quad (2.214)$$

Set $\kappa_* = \kappa/8$. Since $\alpha/2 \in [\alpha_0/2, 1 - \alpha_0/2]$, Lemma 2.154 applies after (2.202). Indeed every exponent appearing in the coefficient bound, in (2.212)–(2.213), and in (2.214) is at most A_{pl} , so its entire r -loss is at most $A_{\text{pl}} L \eta < c_{\text{unif}}/100$. It follows that

$$N_r(X_* + X_*) + \max_{a \in C_0} N_r(X_* + aX_*) \geq r^{-c_{\text{unif}}} N_r(X_*).$$

This contradicts (2.210)–(2.211): by (2.214), each of those upper bounds is at most $K^{2A_{\text{pre}}} N_r(X_*) = r^{-2A_{\text{pre}} L \eta} N_r(X_*)$, while $2A_{\text{pre}} L \eta < c_{\text{unif}}/2$ follows from $A_{\text{pre}} \leq A_{\text{pl}} < L/100$ and (2.202). The planar theorem follows. $\square$

Proof of Proposition 2.132. The scalar-action theorem required in the parameter assembly is Theorem 2.127. Thus all hypotheses of Proposition 2.155 are satisfied. Its proof gives the constants in (2.201) and establishes (2.119), which is exactly the conclusion of Proposition 2.132. $\square$

Remark 2.156 (Planar projection proof architecture). The proof separates the regular-bad, transverse-energy, BSG, popularity, fixed-pair Ruzsa, slope, scalar-Frostman, and coarse-scale estimates. Lemmas 2.151–2.154 supply the tolerant-scale return, robust nested bad-set selection, and uniform scalar-exponent bookkeeping. Proposition 2.155 then performs the planar parameter assembly using Theorem 2.127. These components together prove Proposition 2.132. The argument follows the projection strategy of [5], with all inputs used here stated in the present volume.

Theorem 2.157 (Internal robust line-projection theorem in dimensions $2 \leq d \leq 9$). *For every integer $d \in \{2, \dots, 9\}$ and every $\alpha_0, \kappa > 0$ there exist constants*

$$\Delta_d, u_d, \tau_d, r_d > 0$$

such that the following holds whenever $0 < r < r_d$. Let $E \subset B(0, 1) \subset \mathbb{R}^d$ be r -separated and let $\alpha \in (\alpha_0, d - \alpha_0)$ satisfy

$$r^{-\alpha+u_d} \leq |E| \leq r^{-\alpha-u_d}. \quad (2.215)$$

Assume

$$|E \cap B(x, \rho)| \leq r^{-u_d} \rho^\kappa |E| \quad (x \in \mathbb{R}^d, r < \rho < r^{\tau_d}). \quad (2.216)$$

Let ϑ be a probability measure on $\mathbb{P}^{d-1}$ such that for every projective hyperplane $H \subset \mathbb{P}^{d-1}$,

$$\vartheta\{\theta : d_{\mathbb{P}}(\theta, H) < \rho\} \leq r^{-u_d} \rho^\kappa \quad (r < \rho < r^{\tau_d}). \quad (2.217)$$

Then

$$\vartheta(\mathcal{E}_d(E; \alpha, \Delta_d, u_d, r)) \leq r^{u_d}. \quad (2.218)$$

Proposition 2.158 (Finite codimension-one induction from the planar base). *Proposition 2.132 implies Theorem 2.157 for every $d \in \{2, \dots, 9\}$. Since the planar proposition is now internally proved, the entire discretized line-projection package through dimension nine follows unconditionally from the preceding scalar and projection lemmas.*

Proof. We induct on d . The case $d = 2$ is the assumed planar proposition. Suppose the theorem is known in dimension $d - 1$, where $3 \leq d \leq 9$, for every positive lower-dimension parameter and every positive nonconcentration exponent.

Fix $\alpha_0, \kappa > 0$ in dimension d and put

$$\alpha'_0 = \frac{d-1}{d}\alpha_0, \quad \kappa' = \frac{\kappa}{4d^2}.$$

Let $\Delta'_{d-1}, u'_{d-1}, \tau'_{d-1}, r'_{d-1}$ be the constants supplied by the induction hypothesis with parameters (α'_0, κ') . Let $C_d \geq 1$ be fixed larger than every constant appearing in Lemmas 2.146, 2.147, 2.148, and 2.149. Choose $\eta_d > 0$ so small that

$$C_d \eta_d < \frac{1}{100} \min\{u'_{d-1}, \Delta'_{d-1}, \tau'_{d-1} \kappa', \kappa\}, \quad \eta_d < \frac{\kappa \Delta'_{d-1}}{100 C_d}, \quad (2.219)$$

and set

$$u_d = \eta_d/3, \quad \Delta_d = 2\eta_d, \quad \tau_d \kappa \leq u_d. \quad (2.220)$$

Decrease r_d below all thresholds in the cited lemmas and below r'_{d-1} . Lemma 2.143 again extends the ambient hypotheses to all scales $r \leq \rho \leq 1$.

Assume the d -dimensional conclusion fails. Apply Lemma 2.153 with $q = \alpha/d$, $\eta = \eta_d$, $\Delta_- = \eta_d$, and $\Delta_+ = 2\eta_d$. We obtain $R \subset E$ with $N_r(R) > r^{\eta_d/2} N_r(E)$ for which every nonempty $G \subset R$ satisfies

$$\vartheta(\mathfrak{E}(G; \alpha/d, \eta_d, \eta_d)) > r^{\eta_d}. \quad (2.221)$$

The original size and Frostman estimates imply the corresponding hypotheses for R with loss parameter η_d quantitatively. Indeed, $u_d = \eta_d/3$ and $N_r(R) > r^{\eta_d/2} N_r(E)$ give

$$r^{-\alpha+\eta_d} \leq N_r(R) \leq r^{-\alpha-\eta_d}$$

after weakening the two sides, and

$$N_r(R \cap B(x, \rho)) \leq r^{-\eta_d/3} \rho^\kappa N_r(E) \leq r^{-5\eta_d/6} \rho^\kappa N_r(R) \leq r^{-\eta_d} \rho^\kappa N_r(R).$$

The ambient direction law with prefactor r^{-u_d} is likewise stronger than the version with $r^{-\eta_d}$. Thus the hypotheses (2.164)–(2.166) hold for R with loss parameter η_d . Since (2.221) holds for every nonempty subset and $r^{\eta_d} > r^{2\eta_d}$, it also implies the persistent-badness hypothesis of Lemma 2.147. That lemma produces a hyperplane W , a source set $A^\sharp \subset R$ in an $O(r)$ slab about a translate of W , and its projected slice $B \subset W$ such that, with

$$\beta = \frac{d-1}{d}\alpha,$$

$$r^{-\beta+C_d\eta_d} \leq N_r(B) \leq r^{-\beta-C_d\eta_d}, \quad (2.222)$$

$$N_r(B \cap B_W(x, \rho)) \leq r^{-C_d\eta_d} \rho^{\kappa'} N_r(B) \quad (r \leq \rho \leq 1). \quad (2.223)$$

Because $\alpha \in (\alpha_0, d - \alpha_0)$, we have

$$\beta \in (\alpha'_0, (d - 1) - \alpha'_0).$$

The first inequality in (2.219) makes (2.222)–(2.223) stronger than the size and Frostman hypotheses of the $(d - 1)$ -dimensional theorem.

Let $u_* = C_d \eta_d$ and choose

$$v = \frac{3u_*}{\kappa}.$$

Lemma 2.148 applies after weakening the ambient prefactor to r^{-u_*} ; it removes at most r^{u_*} mass near $W^\perp$ and gives a probability measure ν on $\mathbb{P}(W)$ satisfying

$$\nu\{\theta' : d(\theta', L) < \rho\} \leq 2r^{-u_*} \rho^\kappa \leq r^{-u'_{d-1}} \rho^{\kappa'}$$

for every projective hyperplane $L \subset \mathbb{P}(W)$, after decreasing r_d . Here we used $\kappa' \leq \kappa$ and $u_* < u'_{d-1}$.

We claim that ν has too many bad directions for B . Apply (2.221) to the actual source set $A^\sharp \subset R$. For each ambient bad direction θ outside the removed r^v -neighbourhood of $W^\perp$, choose a witness $F_\theta \subset A^\sharp$. Define explicitly $B_\theta := p_W(F_\theta) \subset B$, where p_W is the source-tracking map from Lemma 2.147. By the source-tracking estimate (2.178), the map from $A^\sharp$ to B changes r -covering numbers only by a fixed factor; hence

$$N_r(B_\theta) \geq r^{2\eta_d} N_r(B) \geq r^{u'_{d-1}} N_r(B), \quad (2.224)$$

where the last inequality follows from $2\eta_d < u'_{d-1}$. Put $\theta' = q_W(\theta)$. Lemma 2.149 and $d_{\mathbb{P}}(\theta, W^\perp) \geq r^v$ give, after also using the fixed source-tracking covering constant in (2.178),

$$N_r(\pi_{\theta'}^W B_\theta) \leq C_{\text{tr}} r^{-v} r^{-\alpha/d - \eta_d}, \quad (2.225)$$

where $C_{\text{tr}} = C_{\text{tr}}(d)$ is fixed before r is chosen. The parameter ledger already gives the required exponent gap: since $C_d \geq 1$, the first inequality in (2.219) gives $2\eta_d < 2\Delta'_{d-1}/100$, while the second and $v = 3C_d \eta_d / \kappa$ give $2v < 6\Delta'_{d-1}/100$. Hence

$$2(\eta_d + v) < \frac{8}{100} \Delta'_{d-1} < \frac{1}{2} \Delta'_{d-1}.$$

After decreasing r_d we have $C_{\text{tr}} \leq r^{-\eta_d}$, and hence

$$N_r(\pi_{\theta'}^W B_\theta) \leq r^{-\alpha/d - v - 2\eta_d}.$$

Because $v + 2\eta_d < 2(v + \eta_d) < \Delta'_{d-1}/2$ and $\beta/(d - 1) = \alpha/d$, this is at most $r^{-\beta/(d-1) - \Delta'_{d-1}}$ for sufficiently small r . Thus every such θ' is exceptional for B in the sense of the $(d - 1)$ -dimensional theorem.

The ambient persistent bad set has measure $> r^{\eta_d}$. The discarded r^v -neighbourhood of $W^\perp$ has measure at most

$$r^{-u_*} r^{v\kappa} = r^{2u_*} = r^{2C_d \eta_d}.$$

Since $C_d \geq 1$, decreasing r_d makes this at most $\frac{1}{2} r^{\eta_d}$. After normalization, the push-forward of the surviving ambient bad directions therefore has ν -measure at least $\frac{1}{2} r^{\eta_d}$. Since $\eta_d < u'_{d-1}$, this is strictly larger than $r^{u'_{d-1}}$ for small r , contradicting the $(d - 1)$ -dimensional theorem. The induction closes. Because only the finitely many dimensions $3, \dots, 9$ are used, the recursive choices (2.219)–(2.220) produce positive constants after finitely many steps. $\square$

Proof of Theorem 2.157. For $d = 2$ this is Proposition 2.132, whose proof has just been completed. For each $d = 3, \dots, 9$, apply Proposition 2.158 inductively. The parameter choice in (2.219)–(2.220) is finite and positive at every stage, so it produces the constants asserted in the theorem for all eight dimensions. $\square$

Remark 2.159 (Finite-dimensional induction summary). Proposition 2.158 contains the finite parameter ledger for $d \mapsto d - 1$, including simultaneously chosen slices, direction-law descent, ambient/slice projection transfer, and all exponent margins. Proposition 2.155 supplies the planar base from the scalar-action theorem. Therefore Theorem 2.157 follows for every $d = 2, \dots, 9$.

Proposition 2.160 (Reduction to the Bourgain projection package). *Theorem 2.157 holds for $d = 9$. Consequently the Bourgain line-projection package of Definition 2.49 holds in dimension $m = 9$.*

Proof. Apply Theorem 2.157 with $d = 9$. Choose

$$\Delta_{\text{pr}} = \Delta_9, \quad u_{\text{pr}} = u_9/4, \quad \tau_{\text{pr}} = \tau_9/2, \quad r_{\text{pr}} = r_9.$$

For $0 < r < r_{\text{pr}}$, the hypotheses (2.1)–(2.3) imply (2.215)–(2.217) after shrinking r_{pr} if necessary. Since $u_{\text{pr}} < u_9$ and $0 < r < 1$, every subset of relative size at least $r^{u_{\text{pr}}}$ is also of relative size at least r^{u_9} , and an exceptional set of measure at most r^{u_9} has measure at most $r^{u_{\text{pr}}}$. The conclusion is exactly (2.4)–(2.5). $\square$

2.17.1.8 The BFLM bootstrap as an internal RM–LD consumer

The projective estimates and the discretized projection theorem are proved inside Volume XII. This subsection is deliberately formulated under the RM–LD interface of Definition 2.50, because that keeps the BFLM argument modular. In the actual application RM–LD is no longer a primitive: Proposition 2.85 proves it internally. The Bourgain line-projection package used in the dimension-increment step was formulated immediately before the analytic modules in Definition 2.49. Throughout this subsection $m \geq 2$ is fixed; in the application $m = 9$.

Corollary 2.83 derives the needed projective Frostman and general-position estimates from RM–LD. Corollary 2.160 proves the preceding projection package in the actual application dimension $m = 9$. Nothing in the bootstrap below assumes either of them as an external input.

Lemma 2.161 (Backward propagation of one Fourier coefficient). *Let $\mu_j = \eta^{(j)} * \mu_0$ on $\mathbb{T}^m$ and $A_t(j) = \{a \in \mathbb{Z}^m : |\hat{\mu}_j(a)| > t\}$. If $|\hat{\mu}_n(a_0)| \geq t$, then for every $k \leq n$ the set*

$$G_k = \{g : |\hat{\mu}_{n-k}(g^T a_0)| > t/2\}$$

has $\eta^{(k)}$ -measure at least $t/2$.

Proof. Fourier transform intertwines the toral action with transpose:

$$\hat{\mu}_n(a_0) = \sum_g \eta^{(k)}(g) \hat{\mu}_{n-k}(g^T a_0).$$

The summands have modulus at most one. If $\eta^{(k)}(G_k) < t/2$, then the absolute value of the right side is $< t/2 + (1 - t/2)(t/2) < t$, a contradiction. $\square$

Lemma 2.162 (Initial positive coarse dimension). *Assume RM-LD as in Definition 2.50. There are $c_*, C_* > 0$ such that if $|\hat{\mu}_n(a_0)| \geq t$, and $k \geq C_*(1 + |\log t|)$ with $k \leq n$, then for*

$$M = \|a_0\|, \quad N = e^{(\lambda_1 + \epsilon)k} \|a_0\|,$$

one has

$$N_M(A_{t/2}(n - k) \cap B(0, N)) \geq (N/M)^{c_*}.$$

Here N_M denotes the covering number by Euclidean M -balls.

Proof. By Lemma 2.161, a set of g of probability at least $t/2$ sends a_0 into $A_{t/2}(n - k)$. Restrict to the RM-LD good set; for $k \gg |\log t|$ its complement has probability $< t/4$, and every surviving vector has norm at most N . Suppose the image were coverable by L balls of radius M . Projectivizing, each such ball is contained in a projective ball of radius $O(Me^{-(\lambda_1 - \epsilon)k}/M) = O(e^{-(\lambda_1 - \epsilon)k})$. By Corollary 2.83, the probability of any one projective ball is at most $e^{-c_1 k}$. Thus $t/4 \leq Le^{-c_1 k}$, whence $L \geq (t/4)e^{c_1 k}$. Taking $k \geq C_*(1 + |\log t|)$ absorbs the factor t and yields the stated power of N/M . $\square$

Lemma 2.163 (Regular window inside a coarse-dimensional large spectrum). *Fix $0 < \alpha < m$ and $0 < \zeta < \min(\alpha, m - \alpha)/100$. There are $c_{\text{rw}} = c_{\text{rw}}(m, \alpha, \zeta) > 0$ and $C_{\text{rw}} = C_{\text{rw}}(m, \alpha, \zeta) \geq 1$ with the following property. Let $1 \leq M < N$, let $A \subset B(0, N)$, and suppose*

$$N_M(A) \geq (N/M)^\alpha. \quad (2.226)$$

Then there are a scale $N_0 \in [M, C_m N]$, a ball $B(x, N_0)$, and an M -separated set $F \subset A \cap B(x, N_0)$ such that

$$\log(N_0/M) \geq c_{\text{rw}} \log(N/M), \quad (2.227)$$

$$|F| \geq C_{\text{rw}}^{-1} (N_0/M)^{\alpha - 6\zeta}, \quad (2.228)$$

$$\frac{|F \cap B(y, \rho)|}{|F|} \leq C_{\text{rw}} (\rho/N_0)^{\alpha - 6\zeta} \quad (M \leq \rho \leq N_0). \quad (2.229)$$

If, in addition, $0 < t \leq 1/2$ and $A \subset A_t(j)$, then there is an M -separated set

$$E \subset (F - F) \cap A_{t^2/4}(j) \cap B(0, 2N_0)$$

with

$$|E| \geq \frac{t^2}{16C_{\text{rw}}} (N_0/M)^{\alpha - 6\zeta}, \quad (2.230)$$

$$\frac{|E \cap B(y, \rho)|}{|E|} \leq 16C_{\text{rw}} t^{-2} (\rho/N_0)^{\alpha - 6\zeta} \quad (M \leq \rho \leq N_0). \quad (2.231)$$

Thus the Fourier conditioning costs the explicit factor t^{-2} ; it is not an absolute loss.

Proof. Choose a maximal M -separated subset $F_* \subset A$. A Euclidean M -ball is covered by at most C_m balls of radius $M/2$, so

$$|F_*| \geq C_m^{-1} N_M(A) \geq C_m^{-1} (N/M)^\alpha. \quad (2.232)$$

Put $r = M/N$. Fix an integer block length T so large that

$$\frac{\log_2(4(mT + 1)C_m)}{T} < \zeta. \quad (2.233)$$

After a generic translation of the dyadic grid, organize F_* by cubes of side $2^{-qT}N$, $0 \leq q \leq k$, where $k = \lceil T^{-1} \log_2(4\sqrt{m}N/M) \rceil$. We regularize this finite tree from the terminal level upward. At level $k-1$, group the surviving parents by the dyadic interval $[2^a, 2^{a+1})$ containing their number of terminal children. Keep the class carrying the largest number of terminal leaves and, inside each retained parent, trim to the smallest child number in that class. This loses at most $2(mT+1)$ in total terminal cardinality and makes the number R_k of terminal children independent of the surviving parent. Having treated levels $q+1, \dots, k$, every surviving level- q child carries the same number of terminal descendants. We may therefore repeat exactly the same dyadic grouping at level q : maximize *terminal descendant mass*, then trim the number of retained children. Induction produces integers $R_1, \dots, R_k$ for which every surviving level- $(q-1)$ cube has exactly R_q surviving children and

$$|F_{\text{reg}}| = \prod_{q=1}^k R_q \geq |F_*| [2(mT+1)]^{-k} \geq 2^{(\alpha-2\zeta)kT}. \quad (2.234)$$

The last inequality follows from (2.232), $2^{kT} \leq N/M < 2^{(k+1)T}$ and (2.233) once N/M is above a constant depending only on the displayed parameters; the bounded remaining range is absorbed by enlarging C_{rw} . Replace F_* by this exact regular subset F_{reg} .

Write $a_q = \log_2 R_q$ and set $\gamma = \alpha - 6\zeta$. From (2.234),

$$\sum_{q=1}^k a_q \geq (\alpha - 2\zeta)kT.$$

For $0 \leq q \leq k$ put

$$S_q = \sum_{i=1}^q (a_i - \gamma T)$$

and choose q_0 with $S_{q_0} = \min_{0 \leq q \leq k} S_q$. For every $q_0 < q \leq k$,

$$\sum_{i=q_0+1}^q a_i \geq \gamma(q - q_0)T. \quad (2.235)$$

Moreover $S_k \geq 4\zeta kT$, while each increment after q_0 is at most $(m - \gamma)T$. Therefore

$$k - q_0 \geq \frac{4\zeta}{m - \gamma} k. \quad (2.236)$$

Choose one surviving cube Q_0 at level q_0 and let F be the terminal points below it. Since the tree is exact regular, every level- q descendant of Q_0 carries the same fraction $\prod_{i=q_0+1}^q R_i^{-1}$ of the uniform mass on F . By (2.235), that fraction is at most $2^{-\gamma(q-q_0)T}$. A Euclidean ball of radius ρ meets at most $C_{m,T}$ dyadic cubes of the first level whose side is below ρ . Taking $N_0 = \sqrt{m} 2^{-q_0T} N$ and absorbing the fixed dimensional factor in the constant gives (2.228)–(2.229); (2.236) gives (2.227).

Assume now $F \subset A_t(j)$. Partition the arguments of $\hat{\mu}_j(a)$, $a \in F$, into four intervals of length $\pi/2$ and retain a class F_1 of size at least $|F|/4$. If ρ_1 is the uniform probability on F_1 , then

$$\left| \sum_{a \in F_1} \rho_1(a) \hat{\mu}_j(a) \right| \geq t/\sqrt{2}. \quad (2.237)$$

Also (2.229) holds for ρ_1 with constant $4C_{\text{rw}}$. By Cauchy–Schwarz,

$$\sum_{a, a' \in F_1} \rho_1(a) \rho_1(a') \hat{\mu}_j(a - a') = \int_{\mathbb{T}^m} \left| \sum_{a \in F_1} \rho_1(a) e(a \cdot x) \right|^2 d\mu_j(x) \quad (2.238)$$

$$\geq \left| \sum_{a \in F_1} \rho_1(a) \hat{\mu}_j(a) \right|^2 \geq t^2/2. \quad (2.239)$$

Let

$$\mathcal{G} = \{(a, a') \in F_1^2 : |\hat{\mu}_j(a - a')| \geq t^2/4\}.$$

Taking real parts in (2.239) and using $|\hat{\mu}_j| \leq 1$ gives

$$(\rho_1 \otimes \rho_1)(\mathcal{G}) \geq t^2/4. \quad (2.240)$$

Hence some $a' \in F_1$ has at least $(t^2/4)|F_1|$ partners a in $\mathcal{G}$. Set

$$E = \{a - a' : a \in F_1, (a, a') \in \mathcal{G}\}.$$

Translation preserves M -separation. The definition of $\mathcal{G}$ gives $E \subset A_{t^2/4}(j)$, and $|E| \geq (t^2/4)|F_1| \geq (t^2/16)|F|$, proving (2.230). Finally a subset of relative size at least $t^2/4$ changes the normalized Frostman constant by at most $4t^{-2}$; combined with the first quadrant loss 4 this gives (2.231). $\square$

Lemma 2.164 (Frostman-preserving thinning). *Fix $0 < \kappa < \beta < m$ and $u > 0$. There is $r_0 > 0$ such that the following holds for $0 < r < r_0$. Let $E \subset B(0, 2)$ be r -separated and assume the uniform probability ρ_E satisfies*

$$\rho_E(B(x, s)) \leq Ks^\gamma \quad (r \leq s \leq 2), \quad (2.241)$$

where $\gamma > \beta$ and $1 \leq K \leq r^{-u/4}$. If $|E| \geq 2r^{-\beta}$, then there is $E_0 \subset E$ with

$$r^{-\beta} \leq |E_0| \leq 2r^{-\beta} \quad (2.242)$$

and

$$\frac{|E_0 \cap B(x, s)|}{|E_0|} \leq r^{-u} s^\kappa \quad (r \leq s \leq 1). \quad (2.243)$$

Proof. Let $L = \lceil r^{-\beta} \rceil$ and choose an L -element subset E_0 of E uniformly without replacement. For a fixed finite set $D \subset E$, the random variable $|E_0 \cap D|$ is hypergeometric with mean $L|D|/|E|$. The elementary exponential-moment proof of the Chernoff bound for sampling without replacement gives

$$\mathbb{P}\{|E_0 \cap D| > 2L|D|/|E| + A\} \leq e^{-A/3} \quad (A \geq 1). \quad (2.244)$$

Use dyadic radii $s_q = 2^q r$ up to 2, and for each q use an r -net of $B(0, 3)$ as centers. There are at most $C_m r^{-m}(1 + \log(1/r))$ resulting balls. Taking $A = 12m \log(1/r)$ in (2.244) and applying the union bound, there is a choice of E_0 for which, simultaneously for all these balls,

$$\frac{|E_0 \cap B(x, 2s_q)|}{L} \leq 2K(3s_q)^\gamma + 12mL^{-1} \log(1/r). \quad (2.245)$$

Since $\gamma > \kappa$, $K \leq r^{-u/4}$, and $L^{-1} \log(1/r) \leq r^{\beta/2}$ for small r , the right side of (2.245) is at most $r^{-u} s_q^\kappa$ after decreasing r_0 . Every ball of radius $s \in [r, 1]$ is contained in one of the tested balls with radius at most $4s$, and the fixed factor is absorbed by the same r^{-u} slack. This proves (2.243); (2.242) holds because $|E_0| = L$. $\square$

Lemma 2.165 (One-direction rectangular increment). *Assume RM-LD and the projection package of Definition 2.49. Fix $0 < \alpha_{\text{lo}} < \alpha_{\text{hi}} < m$. There are $\delta_{\text{rect}}, c_{\text{rect}}, C_{\text{rect}} > 0$ such that the following holds. Suppose $0 < t \leq 1/2$, $1 \leq M < N$, and*

$$N_M(A_t(j) \cap B(0, N)) \geq (N/M)^\alpha, \quad \alpha_{\text{lo}} \leq \alpha \leq \alpha_{\text{hi}}, \quad (2.246)$$

with

$$j \geq C_{\text{rect}} \log(N/M), \quad \log(N/M) \geq C_{\text{rect}}(1 + |\log t|). \quad (2.247)$$

Then there are $k \leq C_{\text{rect}} \log(N/M)$, scales $M_1 < N_1$, a line $\xi \in \mathbb{P}^{m-1}$, and $t_1 \geq c_{\text{rect}} t^2$ such that, with

$$\mathcal{R} = B(0, N_1) \cap \{v : \text{dist}(v, \mathbb{R}\xi) \leq M_1\},$$

one has

$$\log(N_1/M_1) \geq c_{\text{rect}} \log(N/M), \quad (2.248)$$

$$N_{M_1}(A_{t_1}(j - k) \cap \mathcal{R}) \geq (N_1/M_1)^{(\alpha + 2\delta_{\text{rect}})/m}. \quad (2.249)$$

Proof. Choose $\zeta > 0$ so small that

$$20\zeta < \min\{\alpha_{\text{lo}}, m - \alpha_{\text{hi}}, m\Delta_{\text{pr}}/4\}, \quad (2.250)$$

where Δ_{pr} is the gain in the projection package, used with a fixed hyperplane exponent smaller than the one supplied by Corollary 2.83. Lemma 2.163 applied to $A_t(j) \cap B(0, N)$ gives N_0 and an M -separated $E \subset A_{t^2/4}(j) \cap B(0, 2N_0)$. Put

$$\tilde{N}_0 = 2N_0, \quad r = M/\tilde{N}_0, \quad \beta = \alpha - 12\zeta.$$

By (2.230)–(2.231) and (2.247), after increasing C_{rect} we have

$$|E| \geq 2r^{-\beta}, \quad \frac{|E \cap B(y, \rho)|}{|E|} \leq r^{-u_{\text{pr}}/4} (\rho/\tilde{N}_0)^{\alpha-6\zeta} \quad (2.251)$$

for $M \leq \rho \leq \tilde{N}_0$. Rescale by $\tilde{N}_0^{-1}$ and apply Lemma 2.164 with a fixed $\kappa < \beta$ and with $u = u_{\text{pr}}/2$. It gives $E_* \subset E$ satisfying (2.1)–(2.2), after decreasing u_{pr} once in the harmless $O(u_{\text{pr}})$ size window. We henceforth return to the unscaled set E_* and write $\tilde{N}_0$ as the ambient radius.

Fix $0 < \omega < (\lambda_1 - \lambda_2)/20$ so small that

$$\frac{2\omega(\alpha_{\text{hi}} + 1)}{\lambda_1 - \lambda_2 - 2\omega} < \frac{\Delta_{\text{pr}}}{8}. \quad (2.252)$$

Let k be the least integer with

$$e^{(\lambda_1 - \lambda_2 - 2\omega)k} > \tilde{N}_0/M = r^{-1}. \quad (2.253)$$

Set

$$N_1 = 2e^{(\lambda_1 + \omega)k} \tilde{N}_0, \quad M_1 = 2e^{(\lambda_1 - \omega)k} M. \quad (2.254)$$

Minimality in (2.253) gives $k \asymp \log(1/r)$ and hence (2.248). Moreover (2.253) implies

$$e^{(\lambda_2 + \omega)k} \tilde{N}_0 \leq M_1, \quad (2.255)$$

up to the harmless fixed factor already inserted in (2.254).

Let G_{len} be the k -step matrices for which $|k^{-1} \log \sigma_i(g) - \lambda_i| \leq \omega$, $i = 1, 2$. RM-LD gives

$$\eta^{(k)}(G_{\text{len}}^c) \leq C_\omega e^{-c_\omega k}. \quad (2.256)$$

Let ϑ_k be the law of the attracting line of g^T . Corollary 2.83, together with (2.251), puts ϑ_k inside the projection package once C_{rect} is enlarged. Consequently there are a direction set Θ with

$$\vartheta_k(\Theta^c) \leq r^{u_{\text{pr}}} \quad (2.257)$$

and a subset $E_0 \subset E_*$ such that every $E' \subset E_0$ with $|E'| \geq r^{u_{\text{pr}}} |E_0|$ satisfies

$$N_M(\pi_\xi E') \geq r^{-\beta/m - \Delta_{\text{pr}}} \quad (\xi \in \Theta). \quad (2.258)$$

All Fourier coefficients on E_0 have modulus at least $t^2/4$. Retain one Fourier quadrant, obtaining $B \subset E_0$ with $|B| \geq |E_0|/4$ and

$$\left| \frac{1}{|B|} \sum_{b \in B} \hat{\mu}_j(b) \right| \geq \frac{t^2}{4\sqrt{2}}. \quad (2.259)$$

Writing $j_0 = j - k$ and using $\hat{\mu}_j(b) = \int \hat{\mu}_{j_0}(g^T b) d\eta^{(k)}(g)$, Jensen's inequality gives

$$\int \frac{1}{|B|} \sum_{b \in B} |\hat{\mu}_{j_0}(g^T b)|^2 d\eta^{(k)}(g) \geq t^4/32. \quad (2.260)$$

Therefore

$$G_{\text{stat}} = \left\{ g : \frac{1}{|B|} \sum_{b \in B} |\hat{\mu}_{j_0}(g^T b)|^2 \geq t^4/64 \right\} \quad (2.261)$$

has $\eta^{(k)}$ -measure at least $t^4/64$. For $g \in G_{\text{stat}}$,

$$B_g = \{b \in B : |\hat{\mu}_{j_0}(g^T b)| \geq t^2/(8\sqrt{2})\} \quad (2.262)$$

has $|B_g| \geq (t^4/128)|B|$. The separation hypothesis (2.247) makes

$$C_\omega e^{-c_\omega k} + r^{u_{\text{pr}}} < t^4/128, \quad (t^4/128)|B| \geq r^{u_{\text{pr}}} |E_0|.$$

Hence choose $g \in G_{\text{len}} \cap G_{\text{stat}}$ whose attracting line ξ belongs to Θ and for which B_g is admissible in (2.258).

By (2.255), $g^T B_g$ lies in the rectangle $\mathcal{R}$ of (2.249). The top singular direction gives

$$N_{M_1}(g^T B_g) \geq N_M(\pi_\xi B_g) \geq r^{-\beta/m - \Delta_{\text{pr}}}. \quad (2.263)$$

Also $g^T B_g \subset A_{t_1}(j - k)$ with $t_1 = t^2/(8\sqrt{2})$. Finally

$$\log(N_1/M_1) = \log(1/r) + 2\omega k + O(1).$$

Using (2.252), $\beta \geq \alpha - 12\zeta$, and (2.250), choose $0 < \delta_{\text{rect}} < m\Delta_{\text{pr}}/16$ so that, for sufficiently small r ,

$$(\beta/m + \Delta_{\text{pr}}) \log(1/r) \geq \frac{\alpha + 2\delta_{\text{rect}}}{m} \log(N_1/M_1).$$

Together with (2.263), this proves (2.249). $\square$

Lemma 2.166 (From one rectangle back to a ball). *Assume RM-LD. There are $c_{\text{rc}}, C_{\text{rc}}, \kappa_{\text{rc}} > 0$ such that the following holds. Let $0 < t_1 \leq 1/2$, $1 \leq M_1 < N_1$, let $\xi \in \mathbb{P}^{m-1}$, and put*

$$\mathcal{R} = B(0, N_1) \cap \{v : \text{dist}(v, \mathbb{R}\xi) \leq M_1\}.$$

If

$$Q = N_{M_1}(A_{t_1}(j_1) \cap \mathcal{R}), \quad j_1, \log(N_1/M_1) \geq C_{\text{rc}}(1 + |\log t_1|), \quad (2.264)$$

then there are $k_2 \leq C_{\text{rc}}(1 + |\log t_1|)$ and scales $M_2 < N_2$ satisfying

$$|\log(N_2/N_1)| + |\log(M_2/M_1)| \leq C_{\text{rc}}(1 + |\log t_1|) \quad (2.265)$$

such that, with $t_2 = (t_1/8)^{2m}$,

$$N_{M_2}(A_{t_2}(j_1 - k_2) \cap B(0, N_2)) \geq c_{\text{rc}} t_1^{\kappa_{\text{rc}}} Q^m. \quad (2.266)$$

Proof. Let $E_* \subset A_{t_1}(j_1) \cap \mathcal{R}$ be a maximal M_1 -separated set; then $|E_*| \geq c_m Q$. Fix $\omega = (\lambda_1 - \lambda_2)/20$. Choose a constant C_1 (enlarged below) and put

$$k_2 = \lceil C_1(1 + |\log t_1|) \rceil. \quad (2.267)$$

Identify $\xi^\perp$ with the transverse cross-section of $\mathcal{R}$ and partition its M_1 -ball into cubes of side

$$h = t_1^{C_1} e^{-4\omega k_2} M_1. \quad (2.268)$$

One transverse cube contains a subset $E_0 \subset E_*$ of cardinality

$$|E_0| \geq c_m t_1^{C_1(m-1)} e^{-4(m-1)\omega k_2} |E_*| \geq t_1^{C_2} Q \quad (2.269)$$

for a constant $C_2 = C_2(m, C_1, \omega)$. Any two points of E_0 have transverse difference at most $C_m h$; since E_* is M_1 -separated and $h < M_1/(10C_m)$, every nonzero difference $v = b - b'$ satisfies

$$|\langle v, \xi \rangle| \geq M_1/2, \quad \|v - \langle v, \xi \rangle \xi\| \leq C_m h. \quad (2.270)$$

Retain one Fourier quadrant from E_0 , still denoted E , so that

$$|E| \geq t_1^{C_3} Q, \quad \left| \frac{1}{|E|} \sum_{b \in E} \hat{\mu}_{j_1}(b) \right| \geq t_1/\sqrt{2}. \quad (2.271)$$

By RM-LD and Corollary 2.82, enlarge C_1 if necessary so that the following three events for an independent m -tuple $\mathbf{g} = (g_1, \dots, g_m)$ have total exceptional probability at most $(t_1/16)^{2m}$:

(G1) $e^{(\lambda_i - \omega)k_2} \leq \sigma_i(g_s) \leq e^{(\lambda_i + \omega)k_2}$ for $i = 1, 2$ and every s ;

(G2) the input line ξ is at projective distance at least $t_1^{C_1}$ from the repelling hyperplane $H(g_s^T)$ for every s ;

(G3) for unit representatives u_s of those attracting lines, $|\det(u_1, \dots, u_m)| \geq t_1^{C_1}$.

The first assertion is RM-LD, the second is projective hyperplane avoidance, and the third is Corollary 2.82; choosing C_1 larger makes all three exceptional probabilities smaller than the stated power of t_1 .

Put $j_2 = j_1 - k_2$ and

$$\Sigma_{\mathbf{g}}(b_1, \dots, b_m) = \sum_{s=1}^m g_s^T b_s.$$

Define

$$\Phi(x) = \int_G \frac{1}{|E|} \sum_{b \in E} e(g^T b \cdot x) d\eta^{(k_2)}(g).$$

Since $\mu_{j_1} = \eta^{(k_2)} * \mu_{j_2}$,

$$\int_{\mathbb{T}^m} \Phi(x) d\mu_{j_2}(x) = \frac{1}{|E|} \sum_{b \in E} \hat{\mu}_{j_1}(b).$$

Hölder and (2.271) therefore give

$$\int_{\mathbb{T}^m} |\Phi(x)|^{2m} d\mu_{j_2}(x) \geq (t_1/\sqrt{2})^{2m}. \quad (2.272)$$

Expanding the left side and using Fourier inversion produces an average of

$$\hat{\mu}_{j_2} \left(\sum_{s=1}^m g_s^T b_s - \sum_{s=m+1}^{2m} g_s^T b_s \right)$$

over $2m$ independent k_2 -step matrices and $2m$ independent points of E . Its absolute average is at least the right side of (2.272). Fixing one choice of the second m matrices and points gives a vector $b_* \in \mathbb{Z}^m$ for which

$$\int_{G^m} \frac{1}{|E|^m} \sum_{\mathbf{b} \in E^m} |\hat{\mu}_{j_2}(\Sigma_{\mathbf{g}}(\mathbf{b}) - b_*)| d\eta^{(k_2)m}(\mathbf{g}) \geq a_1, \quad a_1 = (t_1/\sqrt{2})^{2m}. \quad (2.273)$$

Put

$$\mathcal{G}_{\text{stat}} = \left\{ \mathbf{g} : \frac{1}{|E|^m} \sum_{\mathbf{b} \in E^m} |\hat{\mu}_{j_2}(\Sigma_{\mathbf{g}}(\mathbf{b}) - b_*)| \geq a_1/2 \right\}.$$

Since the inner average is at most 1, $\eta^{(k_2)m}(\mathcal{G}_{\text{stat}}) \geq a_1/2$. For every $\mathbf{g} \in \mathcal{G}_{\text{stat}}$, at least $(a_1/4)|E|^m$ tuples $\mathbf{b}$ satisfy

$$|\hat{\mu}_{j_2}(\Sigma_{\mathbf{g}}(\mathbf{b}) - b_*)| \geq a_1/4. \quad (2.274)$$

Set $t_2 = (t_1/8)^{2m}$. Since $a_1 = (t_1/\sqrt{2})^{2m}$, we have $t_2 \leq a_1/4$. Thus every tuple counted in (2.274) lies in the Fourier-threshold set $A_{t_2}(j_2)$. Moreover the averaging argument preceding (2.274) gives

$$\frac{|\mathcal{G}_{\text{stat}}|}{|E|^m} \geq \frac{a_1}{4} \geq t_2.$$

By increasing C_1 , the probability that any of (G1)–(G3) fails is less than $a_1/4$. Hence the positive-mass set $\mathcal{G}_{\text{stat}}$ contains some $\mathbf{g}$ satisfying (G1)–(G3).

We now make the separation estimate explicit. For $\mathbf{b}, \mathbf{b}' \in E^m$ put $v_s = b_s - b'_s = a_s \xi + w_s$. By (2.270), either $v_s = 0$ or $|a_s| \geq M_1/2$, and always $\|w_s\| \leq C_m h$. Write u_s for the attracting line of g_s^T . Condition (G2) and the singular-value decomposition give

$$c_s := \|\pi_{u_s}(g_s^T \xi)\| \geq t_1^{C_1} e^{(\lambda_1 - \omega)k_2}, \quad (2.275)$$

$$\|g_s^T \xi - c_s u_s\| \leq e^{(\lambda_2 + \omega)k_2}, \quad \|g_s^T w_s\| \leq e^{(\lambda_1 + \omega)k_2} C_m h. \quad (2.276)$$

By (G3), the $m \times m$ matrix whose columns are the unit vectors u_s has determinant at least $t_1^{C_1}$ and operator norm at most $\sqrt{m}$. The product formula for singular values therefore gives its least singular value at least $c_m t_1^{C_1}$. Thus

$$\left\| \sum_{s=1}^m a_s c_s u_s \right\| \geq c_m t_1^{2C_1} e^{(\lambda_1 - \omega)k_2} \max_s |a_s|. \quad (2.277)$$

The error made by replacing $a_s g_s^T \xi$ by $a_s c_s u_s$ is at most $C_m e^{(\lambda_2 + \omega)k_2} \max_s |a_s|$, and the total transverse error is at most $C_m e^{(\lambda_1 + \omega)k_2} h$. Enlarge C_1 in (2.267) so that

$$C_m e^{(\lambda_2 + \omega)k_2} \leq \frac{c_m}{4} t_1^{2C_1} e^{(\lambda_1 - \omega)k_2}, \quad (2.278)$$

this is possible because $\lambda_1 - \lambda_2 - 2\omega > 0$ and k_2 is a sufficiently large multiple of $1 + |\log t_1|$. By (2.268), the transverse error is bounded by the same quarter of the main term whenever $\max_s |a_s| \geq M_1/2$. Therefore, if $\mathbf{b} \neq \mathbf{b}'$, then

$$\left\| \sum_{s=1}^m g_s^T (b_s - b'_s) \right\| \geq c'_m t_1^{2C_1} e^{(\lambda_1 - \omega)k_2} M_1. \quad (2.279)$$

Thus distinct m -tuples produce points separated by

$$M_2 = c'_m t_1^{2C_1} e^{(\lambda_1 - \omega)k_2} M_1. \quad (2.280)$$

All outputs have norm at most

$$N_2 = C_m e^{(\lambda_1 + \omega)k_2} N_1. \quad (2.281)$$

Equations (2.267), (2.280), and (2.281) give (2.265). Finally (2.274) and the separation just proved imply

$$N_{M_2}(A_{t_2}(j_2) \cap B(0, N_2)) \geq t_2 |E|^m \geq c_{\text{rc}} t_1^{\kappa_{\text{rc}}} Q^m,$$

which is (2.266). $\square$

Proposition 2.167 (One dimension-increment step). *Assume RM-LD. Fix $0 < \alpha_{\text{lo}} < \alpha_{\text{hi}} < m$. There are $\Delta, c, C > 0$ such that, whenever $0 < t \leq 1/2$, $1 \leq M < N$,*

$$N_M(A_t(j) \cap B(0, N)) \geq (N/M)^\alpha, \quad \alpha_{\text{lo}} \leq \alpha \leq \alpha_{\text{hi}}, \quad (2.282)$$

and

$$j \geq C \log(N/M), \quad \log(N/M) \geq C(1 + |\log t|), \quad (2.283)$$

there are $j' = j - O(\log(N/M))$ and $M' < N'$ such that

$$\log(N'/M') \geq c \log(N/M), \quad (2.284)$$

$$N_{M'}(A_{t^C}(j') \cap B(0, N')) \geq (N'/M')^{\alpha + \Delta}. \quad (2.285)$$

The constants implicit in $O(\log(N/M))$ depend only on RM-LD and on $\alpha_{\text{lo}}, \alpha_{\text{hi}}, m$.

Proof. Apply Lemma 2.165. Write $R_1 = N_1/M_1$. It gives a rectangle with

$$Q := N_{M_1}(A_{t_1}(j - k) \cap \mathcal{R}) \geq R_1^{(\alpha + 2\delta_{\text{rect}})/m}, \quad \log R_1 \geq c_{\text{rect}} \log(N/M), \quad (2.286)$$

and $t_1 \geq c_{\text{rect}} t^2$. Increase the constant in (2.283) so that the hypotheses of Lemma 2.166 hold. That lemma gives $j' = j - k - k_2$ and scales $M' < N'$ with

$$N_{M'}(A_{t_2}(j') \cap B(0, N')) \geq c_{\text{rc}} t_1^{\kappa_{\text{rc}}} R_1^{\alpha+2\delta_{\text{rect}}}, \quad t_2 = (t_1/8)^{2m}. \quad (2.287)$$

By (2.265),

$$|\log(N'/M') - \log R_1| \leq C_{\text{rc}}(1 + |\log t_1|). \quad (2.288)$$

Choose the scale-separation constant C so large that the right side of (2.288) is at most $\frac{\delta_{\text{rect}}}{4(m+1)} \log R_1$ and simultaneously

$$c_{\text{rc}} t_1^{\kappa_{\text{rc}}} \geq R_1^{-\delta_{\text{rect}}/4}. \quad (2.289)$$

Then (2.287)–(2.289) imply, for $\Delta = \delta_{\text{rect}}/4$,

$$N_{M'}(A_{t_2}(j') \cap B(0, N')) \geq (N'/M')^{\alpha+\Delta}.$$

Since $t_2 \geq t^C$ after increasing C , this is (2.285). The time loss is $k + k_2 = O(\log(N/M))$, and (2.284) follows from (2.286) and (2.288). $\square$

Corollary 2.168 (Large spectrum reaches almost full dimension). *Under RM-LD, using Corollaries 2.83 and 2.160, for every $\tau > 0$ there is C_τ such that one large Fourier coefficient $|\hat{\mu}_n(a_0)| \geq t$ and $n \geq C_\tau \log(2\|a_0\|/t)$ yield, at some earlier time j and scales $M < N$, a threshold $t_1 \geq t^{C_\tau}$ with*

$$N_M(A_{t_1}(j) \cap B(0, N)) \geq t^{C_\tau} (N/M)^{m-\tau}.$$

Proof. Start with Lemma 2.162. Apply Proposition 2.167 a bounded number of times. The number of iterations depends only on m and the fixed dimension gain Δ , so thresholds remain fixed powers of t and the total backward time remains $O(\log(2\|a_0\|/t))$. $\square$

Lemma 2.169 (Almost-full large spectrum gives spatial granulation). *For every $\tau > 0$ there are $c, C > 0$ with the following property. Suppose $E \subset A_t(j) \cap B(0, N)$ is M -separated, supports an $(m - \tau)$ -Frostman probability at scale M , and $|E| \geq t^C (N/M)^{m-\tau}$. Then for some $\rho \asymp N^{-1}$ there is a $\rho^{1-c\tau}$ -separated finite set $X \subset \mathbb{T}^m$ such that*

$$\mu_j\left(\bigcup_{x \in X} B(x, C\rho)\right) \geq t^C.$$

Proof. Choose phases so that, after retaining one of four quadrants, $|E|^{-1} \sum_{a \in E} \hat{\mu}_j(a) \geq t/\sqrt{2}$. Let $F(x) = |E|^{-1} \sum_{a \in E} e(a \cdot x)$ and smooth it at scale N^{-1} . The preceding inequality gives $\int |F| d\mu_j \gg t$. Before smoothing, character orthogonality gives the exact identity

$$\int_{\mathbb{T}^m} |F(x)|^2 dx = \frac{1}{|E|^2} \sum_{a, a' \in E} \int_{\mathbb{T}^m} e((a - a') \cdot x) dx = \frac{1}{|E|}.$$

Convolution with a probability kernel at scale N^{-1} is an L^2 contraction, so the same upper bound holds after smoothing. Using $|E| \geq t^C (N/M)^{m-\tau}$, we obtain

$$\int_{\mathbb{T}^m} |F(x)|^2 dx \leq t^{-C} (M/N)^{m-\tau}.$$

Hence, for every level $\lambda > 0$,

$$|\{x : |F(x)| > \lambda\}| \leq \lambda^{-2} t^{-C} (M/N)^{m-\tau},$$

which is simply the inequality $\lambda^2 1_{\{|F|>\lambda\}} \leq |F|^2$ integrated over the torus. Choosing the same fixed power of t used in the preceding μ_j -mass estimate produces a level set of Lebesgue measure $O(\rho^{m-c\tau})$ and μ_j -mass at least a fixed power of t . Cover that level set by $O(\rho^{-m+c\tau})$ ρ -balls and extract a $\rho^{1-c\tau}$ -separated subfamily. Enlarging balls by a fixed factor covers the level set and proves the claim. $\square$

Lemma 2.170 (Concentration bootstrap). *Assume RM-LD as in Definition 2.50. Fix $\epsilon > 0$. There are $c, C > 0$ such that if X is r -separated and*

$$\mu_j\left(\bigcup_{x \in X} B(x, \rho)\right) = \beta, \quad \rho < e^{-m\lambda_1 k} r,$$

then for

$$r' = e^{-(\lambda_1 + \epsilon)k} r, \quad \rho' = e^{-(\lambda_1 - \epsilon)k} \rho,$$

there is an r' -separated X' with $|X'| \leq |X|$ and

$$\mu_{j-k}\left(\bigcup_{x \in X'} B(x, \rho')\right) \geq \beta^m - e^{-ck}.$$

Proof. Write the mass at time j as the $\eta^{(k)}$ -average of masses of inverse images at time $j - k$. Jensen gives a lower bound β^m for the average m -fold intersection of such inverse images. Restrict to m independently chosen well-shaped matrices. By RM-LD their singular scales lie in the stated exponential window; by Corollary 2.83, outside an e^{-ck} exceptional set, the corresponding attracting/repelling flags are quantitatively transverse. For a transverse m -tuple, the intersection of the inverse images of $B(x_i, \rho)$ has diameter $O(\rho')$, while two intersections associated to distinct r -separated centers are r' -separated. Choose one center in each nonempty intersection. Fubini and the exceptional-set estimate give the claimed mass inequality. $\square$

Lemma 2.171 (High concentration forces a rational center). *Assume RM-LD as in Definition 2.50. There are $\delta, c, C > 0$ such that, if $j \geq C \log(1/\rho)$ and*

$$\mu_j(B(z, \rho)) > \rho^\delta,$$

then z lies within $\rho^{9/10}$ of a rational point of denominator at most $\rho^{-1/10}$.

Proof. Choose $k \asymp \log(1/\rho)$ and, using Corollary 2.83, select $m + 1$ words $g_0, \dots, g_m$ of length k whose expanding directions are in quantitative general position and for which the ball $B(z, \rho)$ has inverse-image mass at least $\rho^{O(\delta)}$ at time $j - k$. Two of the corresponding inverse images must intersect once δ is sufficiently small. Thus for suitable points differing by $O(\rho)$ we obtain

$$(g_i - g_0)z \in \mathbb{Z}^m + O(e^{O(k)}\rho), \quad i = 1, \dots, m.$$

General position makes the integer matrix with block rows $g_i - g_0$ invertible on an $m \times m$ minor with determinant $1 \leq |D| \leq e^{O(k)}$. Cramer's rule gives

$$\text{dist}(z, D^{-1}\mathbb{Z}^m/\mathbb{Z}^m) \leq e^{O(k)}\rho.$$

Take k with a sufficiently small fixed proportionality constant relative to $\log(1/\rho)$; then $|D| \leq \rho^{-1/10}$ and the error is at most $\rho^{9/10}$. The exceptional-set probabilities are e^{-ck} and are smaller than the assumed ρ^δ after the constants are fixed. $\square$

Lemma 2.172 (Tightening around rational centers). *Assume RM-LD as in Definition 2.50. Let R_Q be the rational points of denominator at most Q and $W_{Q,r} = N_r(R_Q)$. For every $\epsilon > 0$ there are $c, k_0 > 0$ such that, whenever $k \geq k_0$ and $e^{m\lambda_1 k} r < Q^{-2}$,*

$$\mu_{j-k}(W_{Q, e^{-(\lambda_1 - \epsilon)k} r}) \geq \mu_j(W_{Q,r}) - e^{-ck}.$$

Proof. Because every matrix is integral, R_Q is invariant. Choose $L > e^{ck}$ well-shaped words $g_1, \dots, g_L$ in quantitative general position using RM-LD and Corollary 2.83. For every m -tuple among them, transversality implies that the intersection of the inverse r -neighborhoods of a fixed rational center has diameter at most $e^{-(\lambda_1 - \epsilon)k} r$. The condition $e^{m\lambda_1 k} r < Q^{-2}$ prevents intersections coming from distinct rational centers, since two distinct points of R_Q are separated by at least Q^{-2} . Hence, after removing the tightened neighborhood, the sets $g_i^{-1}W_{Q,r}$ have no m -fold intersection. Integrating the sum of their indicators and using that at least one g_i has inverse-image mass within e^{-ck} of the average gives the claim. $\square$

Theorem 2.173 (Internal Fourier-granulation bootstrap). *For the amplified $m = 9$ arithmetic walk constructed above, the torus Fourier-rigidity theorem stated immediately before Definition 2.49 holds. Equivalently, for every $0 < \lambda < \lambda_1$ there is $C > 0$ such that, whenever a nonzero Fourier coefficient of $\eta^{(n)} * \delta_x$ at $a_0 \in \mathbb{Z}^9$ has modulus larger than t and*

$$n > C \log \frac{2\|a_0\|}{t},$$

there is a rational point $y \in \mathbb{T}^9$ of denominator at most $(2\|a_0\|/t)^C$ with

$$d_{\mathbb{T}^9}(x, y) \leq e^{-\lambda n}.$$

Proof. Proposition 2.85 supplies RM-LD for the amplified walk, Corollary 2.83 supplies its projective Frostman and general-position consequences, and Corollary 2.160 supplies the required line-projection package. We now use only these proved interfaces.

Let $|\hat{\mu}_n(a_0)| > t$ and take the available backward time to be a large constant multiple of $\log(2\|a_0\|/t)$. Lemma 2.162, followed by Corollary 2.168, produces an almost-full-dimensional regular large spectrum at polynomially degraded threshold. Lemma 2.169 converts it into a positive, polynomial-in- t portion of μ_j concentrated in a union of well-separated balls.

Apply Lemma 2.170 a bounded number of times. The exponent of concentration falls from $m - O(1)$ to the fixed small δ in Lemma 2.171; because the number of iterations is fixed, the surviving mass is still $t^{O(1)}$ and all times used are $O(\log(2\|a_0\|/t))$. Lemma 2.171 therefore places that mass inside $W_{Q,r}$ for $Q \leq (2\|a_0\|/t)^C$ and a radius r which is a fixed negative exponential of the time already spent. Finally iterate Lemma 2.172; rational centers are invariant, so each additional block of random-walk time shrinks the radius by $e^{-(\lambda_1 - \epsilon)k}$ while losing only e^{-ck} mass. Choosing the blocks geometrically and taking ϵ so that the accumulated exponent is $> \lambda$ yields

$$\mu_0(W_{Q, e^{-\lambda n}}) > t^C, \quad Q \leq (2\|a_0\|/t)^C.$$

For $\mu_0 = \delta_x$ this says exactly that x lies within $e^{-\lambda n}$ of a rational point of the stated denominator, which is the torus Fourier-rigidity principle. $\square$

Remark 2.174 (BFLM analytic dependency summary). The Fourier-granulation argument, RM-LD/projective components, and the line-projection package are all internally derived. Theorem 2.66 proves the rank-one mean contraction from the earlier free-pair/tree owner and the Guivarc'h–Raugi

boundary/recurrence mechanism; Proposition 2.85 proves the amplified vector and singular-value deviations without an amplified MC hypothesis; and Corollary 2.160 supplies the required $m = 9$ projection input. Hence Theorem 2.173 follows from the projection and random-matrix inputs above, and the downstream congruence applications require no additional analytic theorem at this stage.

Remark 2.175 (Analytic dependency summary). The scalar-action theorem 2.127 feeds Proposition 2.160, which gives the Bourgain line-projection package in the required dimension $m = 9$. Together with the random-matrix estimates, this is the input to the Fourier–granulation bootstrap Theorem 2.173. Proposition 2.176 then converts that analytic package into the conductor-sensitive Fourier decay used by the correlation, flattening, and spectral arguments below.

Proposition 2.176 (Conductor Fourier decay from the internal projection package). *Let $S \subseteq \mathrm{SL}_2(\mathbb{Z})$ be the fixed symmetric non-elementary arithmetic generating set and let ν be its symmetric walk measure. Then there are $c_0, C_0, C_1 > 0$, depending only on S , such that for every odd r , every $v_0 \in \mathfrak{sl}_2(\mathbb{Z}/r\mathbb{Z})$ primitive at each $p \mid r$, and every additive character $\chi_b(v) = e(\langle v, b \rangle / r)$ of exact conductor $d > 1$,*

$$\left| \int \chi_b(\mathrm{Ad}(g)v_0) d\nu^{(\ell)}(g) \right| \leq C_0 d^{-c_0} \quad (2.34)$$

whenever $\ell \geq C_1 \log r$. Thus (2.34) supplies (2.31).

Proof. Reduce the phase to its exact conductor. If $a_0 = (r/d)$ is the common scalar divisor removed from the frequency, then the phase factors through modulus d ; after this reduction both the image of v_0 and the reduced frequency b_0 are nonzero modulo every prime dividing d . Choose the integral basis (E, H, F) of $\mathfrak{sl}_2$ and write $\rho = \mathrm{Ad} : \Gamma \rightarrow \mathrm{SL}_3(\mathbb{Z})$.

We first verify the linear hypotheses needed below. Every finite-index subgroup of the non-elementary group $\Gamma = \langle S \rangle$ is Zariski dense in SL_2 : its Zariski closure has finite index in the connected group SL_2 . Hence every invariant subspace for such a subgroup in the adjoint representation is invariant under SL_2 and therefore is an ideal in the simple Lie algebra $\mathfrak{sl}_2$. Thus $\rho(\Gamma)$ is strongly irreducible. It is proximal because Γ contains a hyperbolic element; if its eigenvalues are α, α^{-1} with $|\alpha| > 1$, the adjoint eigenvalues are $\alpha^2, 1, \alpha^{-2}$.

We shall use the standard left–right amplification of this action. Its associative algebra is all of $M_3(\mathbb{R})$. Here is a direct rank-one check, included to avoid hiding Burnside’s theorem. Since Γ is Zariski dense, the algebra generated by $\rho(\Gamma)$ contains the coefficients of the polynomial one-parameter matrices $\mathrm{Ad}(\exp(tE))$ and $\mathrm{Ad}(\exp(tF))$; hence it contains $A = \mathrm{ad} E$ and $B = \mathrm{ad} F$. In the basis (E, H, F) ,

$$A = -2e_{12} + e_{23}, \quad B = -e_{21} + 2e_{32}.$$

Now $A^2 = -2e_{13}$ and $B^2 = -2e_{31}$, while A^2B is a nonzero multiple of e_{12} and B^2A a nonzero multiple of e_{32} . Subtracting these from A, B gives e_{23}, e_{21} , and products then give all diagonal and remaining matrix units. Thus the generated algebra is $M_3(\mathbb{R})$.

Put $\Lambda = M_3(\mathbb{Z})$ and let Γ_0 act on $M_3(\mathbb{R})$ by the left and right maps $L_{\rho(s)}, R_{\rho(s)}$ for $s \in S$. Any invariant subspace under a finite-index subgroup is a two-sided ideal for the algebra generated by a finite-index subgroup of $\rho(\Gamma)$, hence is 0 or $M_3(\mathbb{R})$. So the Γ_0 -action is strongly irreducible. If $g \in \Gamma$ is the hyperbolic element above, then $L_{\rho(g)}R_{\rho(g)}$ has a unique eigenvalue of largest modulus, namely $|\alpha|^4$; hence the amplified action is proximal.

Define the integral functional

$$\phi(T) = \langle Tv_0, b_0 \rangle \quad (T \in \Lambda).$$

For every $p \mid d$ it is primitive modulo p : choose coordinates i, j for which $(v_0)_i$ and $(b_0)_j$ are nonzero modulo p and take the matrix unit e_{ji} . Therefore $x = \phi/d \in \Lambda^*/\Lambda^*$ has exact denominator d .

Let S_0 consist of the left and right multiplication maps attached to S , with equal weights. Starting from the identity matrix, the push-forward of the ℓ -step S_0 walk under $\gamma \mapsto \gamma(I)$ is exactly the ℓ -step S walk under $g \mapsto \rho(g)$. Indeed, if the assertion holds at time j , one further left or right multiplication has distribution $\nu * \nu^{(j)}$ or $\nu^{(j)} * \nu$, and both equal $\nu^{(j+1)}$. Consequently the Fourier coefficient in (2.34) equals

$$\int e\left(\left\langle \gamma^* \frac{\phi}{d}, I \right\rangle\right) d\eta_{S_0}^{(\ell)}(\gamma). \quad (2.35)$$

The frequency $I \in \Lambda$ is fixed, independently of d .

The arithmetic hypotheses above place the standard walk in the rank-one scope of Theorem 2.66; Proposition 2.85 and Corollary 2.86 therefore supply the full RM-LD package for S_0 without an additional hypothesis. Theorem 2.173, together with the internally proved projective and line-projection packages, proves the torus Fourier-rigidity principle for this exact walk.

Suppose (2.35) were larger than d^{-c_0} . Apply that torus Fourier-rigidity theorem with $t = d^{-c_0}$ and the fixed frequency I . It produces a rational point a/q_1 with

$$q_1 \leq Cd^{C_{Tc_0}}, \quad \left\| \frac{\phi}{d} - \frac{a}{q_1} \right\| \leq e^{-\lambda \ell}. \quad (2.36)$$

Choose $c_0 > 0$ so small that $C_{Tc_0} < 1/4$. For large d this gives $q_1 \leq Cd^{1/4} < d$. Since ϕ/d has exact denominator d , it cannot equal a rational point of denominator $q_1 < d$. In fixed lattice coordinates, after putting both points over the common denominator dq_1 , some coordinate has a nonzero integral numerator; hence the two distinct points are separated by at least $c/(dq_1)$. Moreover

$$dq_1 \leq Cd^{5/4} \leq Cr^{5/4}.$$

It is therefore enough to choose C_1 once and for all so that $\lambda C_1 > 5/4$ (with a further fixed enlargement absorbing the constant C/c). Then

$$e^{-\lambda \ell} \leq r^{-\lambda C_1} < c/(dq_1),$$

contradicting (2.36). The finitely many bounded values of d are absorbed into C_0 . Hence (2.34) holds. $\square$

Remark 2.177 (Orbital-spreading interface). Proposition 2.47 reduces the rank-one commutator propagation to a quantitative orbital-spreading estimate. Concretely, let $\mu = \nu_q^{(\ell)}$ be the logarithmic-time arithmetic walk and suppose $\mu(xH) \geq e^{-\varepsilon W}$. For $r = \prod_{p \in D} p^{t_p}$ and the primitive coefficient $v_0 \in \mathfrak{s}(r)$ of the isolated seed, the required estimate is that for one fixed C

$$\left| \text{supp} \left(\sum_{j=1}^C \text{Ad}(h_j) v_0 \right) \right| \geq e^{-\delta W} r^3, \quad h_j \in H, \quad (2.30)$$

with $\delta \rightarrow 0$ as $\varepsilon \rightarrow 0$, uniformly in the number of prime factors.

The arbitrary-modulus architecture obtains this estimate from a conductor-sensitive Fourier bound for the adjoint random walk and then applies Parseval and Cauchy–Schwarz. Lemmas 2.44–2.45 and Proposition 2.47 carry out the subsequent bracket counting. Proposition 2.176 provides the needed conductor Fourier bound from the BFLM bootstrap and the left–right matrix-algebra random-matrix action. Thus the logical chain is

$$\begin{aligned} \text{projection/random-matrix control} &\implies \text{conductor Fourier decay} \\ &\implies \text{orbital spreading} \implies \text{deep congruence propagation.} \end{aligned}$$

The exact group-theoretic part of the compatibility argument *can* be proved immediately.

Theorem 2.178 (Exact prime-power Goursat compatibility). *Let*

$$G = \prod_{p \in I} G_{p,n_p}, \quad G_{p,n} = \mathrm{SL}_2(\mathbb{Z}/p^n\mathbb{Z}),$$

with pairwise distinct allowed primes. An exact subdirect graph obstruction between two disjoint prime blocks either remains nontrivial after reduction modulo the relevant primes, or its nontrivial part is supported inside a single residue characteristic. In particular there is no exact common quotient linking two different residue characteristics.

Proof. Let $H \leq G_J \times G_K$ be the exact subdirect subgroup giving the graph obstruction. Goursat's construction produces normal subgroups $N_J \triangleleft G_J$ and $N_K \triangleleft G_K$ and an isomorphism

$$G_J/N_J \cong G_K/N_K.$$

If this quotient has a nontrivial composition factor surviving modulo the residue primes, then the obstruction descends to the squarefree residue product and is of the type already classified in Volume X.

Suppose instead that every simple composition factor of the common quotient comes from principal congruence kernels. A principal congruence kernel at residue characteristic p is a finite p -group through its standard filtration; hence all of its composition factors are cyclic of order p . A nontrivial quotient common to a p -block and an r -block with $p \neq r$ would therefore have a simple composition factor of order both p and r , which is impossible. Thus any residual vertical quotient is confined to one residue characteristic. This proves the exact compatibility statement. $\square$

Remark 2.179 (How the quantitative prime-power lift is assembled). The entropy decomposition and seed-isolation steps, the exact prime-power Goursat compatibility above, and the rank-one bracket calculation supply the group-theoretic and congruence-layer part of the lift. Proposition 2.176 now proves the conductor-sensitive Fourier decay (2.31) for the adjoint arithmetic walk. Proposition 2.48 converts it to the required orbital spreading with no divisor-count loss, and Proposition 2.47 completes the deep principal-congruence propagation in $O(c_1^{-1})$ stages. Together these results prove the quantitative prime-power correlation lift without any additional analytic input.

2.18 Captured mass makes every proper projection almost full

Let ν_q be the fixed symmetric lazy arithmetic walk. The next lemma is the point at which the well-founded induction is used.

Lemma 2.180 (Inductive mixing forces a dense proper projection). *Assume the lazy walk on every proper divisor $d \mid q$ has spectral gap at least $\sigma > 0$. Let $I \subsetneq \mathcal{P}$ be nonempty and put*

$$q_I = \prod_{p \in I} p^{n_p}, \quad M_I = |G(q_I)| = e^{W(I)}.$$

Suppose

$$\nu_q^{(\ell)}(xH) \geq e^{-\varepsilon_e W} \tag{5.1}$$

with

$$\ell \geq C_{\mathrm{mix}} W, \quad \sigma C_{\mathrm{mix}} > 1.$$

Then, for all sufficiently large W and sufficiently small ε_c ,

$$\log |H_I| \geq W(I) - 2\varepsilon_c W. \quad (5.2)$$

The same lower bound holds for $(H^C)_I$.

Proof. Projection cannot decrease the captured mass, so

$$\nu_{q_I}^{(\ell)}(x_I H_I) \geq e^{-\varepsilon_c W}.$$

Let u_I be uniform probability on $G(q_I)$. Spectral gap gives, in counting-measure L^2 ,

$$\|\nu_{q_I}^{(\ell)} - u_I\|_2 \leq e^{-\sigma \ell}.$$

Hence Cauchy–Schwarz yields

$$e^{-\varepsilon_c W} \leq \frac{|H_I|}{M_I} + |H_I|^{1/2} e^{-\sigma \ell} \leq \frac{|H_I|}{M_I} + e^{W/2 - \sigma C_{\text{mix}} W}.$$

Choose C_{mix} once so that $\sigma C_{\text{mix}} > 1$; then the second term is e^{-cW} for some fixed $c > 0$. After choosing $\varepsilon_c < c/2$ and W large, it is at most $\frac{1}{2}e^{-\varepsilon_c W}$. Therefore

$$|H_I| \geq \frac{1}{2}e^{-\varepsilon_c W} M_I.$$

Taking logarithms and absorbing $\log 2$ into $\varepsilon_c W$ proves (5.2). Since $H \subseteq H^C$, the same estimate holds for $(H^C)_I$. $\square$

Theorem 2.181 (Captured arbitrary-modulus filling theorem). *Fix $\varepsilon_0 > 0$. There are constants $\delta_0, c_K > 0$ and an integer $C = C(\varepsilon_0)$, independent of q and $\omega(q)$, with the following property. Let*

$$q = \prod_{p \in \mathcal{P}} p^{n_p}, \quad G(q) = \text{SL}_2(\mathbb{Z}/q\mathbb{Z}), \quad W = \log |G(q)|,$$

and let $H \subseteq G(q)$ be symmetric, contain the identity, and be a K -approximate subgroup. Assume

$$\log K \leq c_K W, \quad \nu_q^{(\ell)}(xH) \geq e^{-\delta_0 W}, \quad \ell \geq C_{\text{start}} W, \quad (2.37)$$

and assume that the arithmetic walk has the already-owned uniform spectral gap on every proper divisor of q (with the prime-power case supplied by Volume XI). Then

$$|H^C| \geq e^{-\varepsilon_0 W} |G(q)|. \quad (2.38)$$

Equivalently, a captured approximate subgroup with sufficiently small tripling cannot remain genuinely intermediate in size.

Proof. We assemble the preceding lemmas in the same small-exponent/large-exponent order as the original arbitrary-modulus argument. The point of spelling out the assembly is that the bounded and deep pieces live in complementary quotients, so their cardinalities multiply; no correlation estimate is being silently assumed.

There are two already-owned boundary cases. If q is a prime power, the Volume-XI spectral theorem and the same Cauchy–Schwarz estimate as in Lemma 2.180 imply directly that a set carrying the mass in (2.37) has size at least $e^{-2\delta_0 W} |G(q)|$. If q is squarefree, the corresponding uniform spectral theorem is owned by Volume X and gives the same conclusion. Thus we may

assume that q has at least two residue characteristics and at least one exponent greater than one; every residue quotient used below is then either an earlier squarefree owner or a proper divisor covered by the induction hypothesis.

Choose a small parameter $\eta > 0$, eventually much smaller than ε_0 , and apply Lemma 2.8. Thus

$$\mathcal{P} = B \sqcup T \sqcup D, \quad W(T) \leq \eta W,$$

with bounded exponents on B and large exponents on D . All constants below are fixed after η is fixed.

The bounded-depth part. First suppose $W(B) \geq \eta W$. The collision estimate Lemma 2.33, followed by proper-divisor mixing exactly as in Lemma 2.41, gives a squarefree residue image of a bounded product of $H^{-1}H$ having density $e^{-o_\eta(W(B))}$ in the B -residue product. (When D is nonempty one may use the deeper slice from that lemma; when D is negligible take $d = 1$ in the same collision calculation.) Lemma 2.40 then deletes a set $E_B \subseteq B$ with $W(E_B) \leq \eta W$ and makes a bounded power surjective onto the squarefree residue product of $B \setminus E_B$. Proposition 2.38 lifts this surjectivity through all bounded congruence layers, after enlarging E_B by another set of weight at most ηW . Hence there is a bounded product set $Y_B \subseteq H^{C_B}$ whose projection onto

$$G_{B_0}, \quad B_0 := B \setminus E_B,$$

is the whole group. Choose a transversal $Y'_B \subseteq Y_B$ with exactly one representative above each element of G_{B_0} ; then

$$|Y'_B| = |G_{B_0}|. \quad (2.39)$$

If $W(B) < \eta W$, put all of B into the exceptional block and take $Y'_B = \{1\}$; the loss is already within the final budget.

The deep part. Suppose $W(D) \geq \eta W$. Proposition 2.36 and the unconditional correction Proposition 2.42 give, after discarding $O(\eta W)$ conductor weight, an element x_0 which is trivial on the already retained bounded block and has exact depths

$$c_1 n_p \leq v_p(x_0 - 1) < c_2 n_p$$

on the retained deep block. Proposition 2.176 and Proposition 2.48 supply the orbital-spreading hypothesis (2.24). The strengthened Proposition 2.47 then gives a bounded product $Y_D \subseteq H^{C_D}$, lying in the kernel of the retained bounded quotient, whose projection to the retained deep block D_0 satisfies

$$|\pi_{D_0}(Y_D)| \geq e^{-\eta W} |G_{D_0}|. \quad (2.40)$$

Choose one representative for every element of this projected set and call the resulting subset Y'_D . Because the seed-isolation step made the deep construction trivial on B_0 , every element of Y'_D lies in the kernel of the projection to G_{B_0} . If $W(D) < \eta W$, discard D and take $Y'_D = \{1\}$.

Multiplying the complementary pieces. The map

$$Y'_D \times Y'_B \longrightarrow G(q), \quad (y_D, y_B) \mapsto y_D y_B,$$

is injective. Indeed, equality of two products first gives equality of the B_0 -coordinates; our choice of the transversal Y'_B then identifies the bounded representatives, after which the deep representatives coincide. This is the exact cardinality multiplication used in the source proof, rather than an independence assumption. Consequently a bounded power of H has size at least

$$e^{-\eta W} |G_{B_0}| |G_{D_0}|. \quad (2.41)$$

The omitted transition block, bounded-block exceptions, deep-seed exceptions, and first/terminal deep-layer losses have total logarithmic size at most $C_*\eta W$ after the parameter hierarchy in Proposition 2.47 is chosen. Thus

$$|H^C| \geq e^{-C_*\eta W} |G(q)|.$$

Choose $\eta < \varepsilon_0/(2C_*)$, then choose the capture exponent δ_0 and the approximate-group constant c_K small enough for all preceding collision, mixing, seed, and bounded-product estimates. The finitely many bounded conductors are absorbed by increasing C . This proves (2.38). $\square$

Corollary 2.182 (Quantitative prime-power correlation lift). *Under the hypotheses of Theorem 2.181, for every $\eta_0 > 0$ there is a bounded power $A = H^C$ such that*

$$\text{Corr}_{\mathcal{P}}(A) \leq C_0 \log K + \eta_0 W, \quad (2.42)$$

with C_0 and C independent of q and $\omega(q)$. In particular this is the correlation interface required in (1.1).

Proof. Apply Theorem 2.181 with target loss $\varepsilon_0 = \eta_0/2$. Since every coordinate projection of A is contained in its ambient factor,

$$\text{Corr}_{\mathcal{P}}(A) = \sum_{p \in \mathcal{P}} \log |A_p| - \log |A| \leq W - \log |A| \leq \frac{\eta_0}{2} W.$$

This is stronger than (2.42); adding the nonnegative allowance $C_0 \log K$ gives the stated form. The important point is that the estimate is paid once after the complementary bounded/deep pieces have been multiplied, not once per prime factor. $\square$

Remark 2.183 (Source architecture). Theorem 2.181 is the rank-one version of the small-exponent/large-exponent assembly in Bourgain–Varjú’s proof of their arbitrary-modulus product-growth proposition. The source proof first fills the bounded-exponent quotient, constructs a controlled-depth element on the large-exponent quotient, fills the latter by adjoint-orbit spreading and commutators, and then multiplies the two complementary contributions. The preceding sections separate those operations into explicit student-level lemmas and propositions; Corollary 2.182 merely translates the resulting almost-full bounded power into the correlation language used by the recombination chapter.

2.19 Proper-divisor induction and the global obstruction

2.20 Purpose

Let

$$q = \prod_{p \in \mathcal{P}} p^{n_p}, \quad G(q) = \text{SL}_2(\mathbf{Z}/q\mathbf{Z}) \cong \prod_{p \in \mathcal{P}} G_p, \quad G_p = \text{SL}_2(\mathbf{Z}/p^{n_p}\mathbf{Z}),$$

and write

$$w_p = \log |G_p|, \quad W = \sum_{p \in \mathcal{P}} w_p = \log |G(q)|.$$

A naive prime-by-prime argument would apply the prime-power product theorem separately to every projection of one globally captured inverse set. That is not legitimate: a lower mass bound of size $e^{-\varepsilon W}$ need not be strong enough to make the projection to a small factor G_p satisfy the one-prime captured-set hypotheses.

The present argument removes that step. The replacement is a simultaneous induction on the modulus. At modulus q we assume the spectral-gap theorem for every proper divisor $d \mid q$; the base case, when q is a prime power, is Volume XI. A captured approximate subgroup is then tested only on one or two *proper CRT subproducts*. No prime-by-prime captured-mass comparison occurs.

The structural argument below uses the quantitative correlation bound

$$\text{Corr}_{\mathcal{P}}(A) \leq C_0 \log K + \eta W. \quad (1.1)$$

The squarefree correlation theorem is proved in Volume X, the exact no-common-quotient calculation is proved above, and the preceding sections establish the entire vertical prime-power lift, including the conductor Fourier estimate, orbital spreading, and bounded-stage deep Lie propagation. Theorem 2.182 assembles these ingredients into (1.1). Consequently the proper-divisor induction that follows uses only the stated Volume-X/Volume-XI inputs and the induction hypothesis on proper conductors.

2.21 Well-founded divisor induction

Definition 2.184 (Divisor complexity). For $q = \prod p^{n_p}$ put

$$\Omega(q) := \sum_{p \mid q} n_p.$$

Every proper divisor $d \mid q$ has $\Omega(d) < \Omega(q)$ after replacing d by its exact conductor. Thus induction on Ω simultaneously handles deletion of a prime factor and reduction of a prime-power depth.

Remark 2.185. The proof below only projects to divisors supported on a proper subset of the prime factors, but the Ω -formulation is the one needed later in the exact-conductor spectral endgame, where a proper conductor may retain the same prime support while lowering one exponent.

2.22 A two-block partition of conductor weight

Lemma 2.186 (Dominant factor or two macroscopic blocks). *Fix $0 < \theta < 1/4$. If $|\mathcal{P}| \geq 2$, then at least one of the following alternatives is available:*

(i) *there is $p_0 \in \mathcal{P}$ with*

$$w_{p_0} > (1 - \theta)W;$$

(ii) *there is a partition $\mathcal{P} = I \sqcup J$ into nonempty proper blocks satisfying*

$$W(I) \geq \frac{\theta}{2}W, \quad W(J) \geq \frac{\theta}{2}W.$$

Proof. Assume no factor has weight $> (1 - \theta)W$. Add factors in any order until the accumulated weight first reaches $\theta W/2$, and call the selected set I . By minimality and the bound on every individual factor,

$$W(I) \leq \frac{\theta}{2}W + (1 - \theta)W = \left(1 - \frac{\theta}{2}\right)W.$$

Hence the complementary nonempty block J has weight at least $\theta W/2$. The same is true of I by construction. $\square$

2.23 Block correlation is dominated by prime correlation

For a partition $\mathcal{P} = I \sqcup J$ and $A \subseteq G(q)$ define

$$\text{Corr}_{I,J}(A) := \log |A_I| + \log |A_J| - \log |A|.$$

Lemma 2.187 (Coarsening decreases correlation defect). *For every nonempty $A \subseteq G(q)$,*

$$0 \leq \text{Corr}_{I,J}(A) \leq \text{Corr}_{\mathcal{P}}(A),$$

where

$$\text{Corr}_{\mathcal{P}}(A) = \sum_{p \in \mathcal{P}} \log |A_p| - \log |A|.$$

Proof. The first inequality follows from the injective map $A \rightarrow A_I \times A_J$. For the second, use

$$|A_I| \leq \prod_{p \in I} |A_p|, \quad |A_J| \leq \prod_{p \in J} |A_p|.$$

Taking logarithms proves the claim. $\square$

2.24 The structural two-block obstruction

Theorem 2.188 (Two-block obstruction for a medium slow approximate subgroup). *Fix $0 < \gamma < 1/4$. There are constants*

$$0 < \delta \ll \theta \ll \gamma, \quad \varepsilon_K > 0,$$

with the following property. Let $H \subseteq G(q)$ be symmetric, $1 \in H$, and suppose

$$e^{\gamma W} \leq |H| \leq e^{(1-\gamma)W}. \quad (4.1)$$

Assume H is a K -approximate subgroup with

$$\log K \leq \varepsilon_K W,$$

and let $A = H^C$ for one fixed exponent C , so

$$|A| \leq K^{c_C} |H|. \quad (4.2)$$

If $|\mathcal{P}| \geq 2$, then at least one of the following holds.

(A) proper-projection deficiency: *there is a nonempty proper block $I \subsetneq \mathcal{P}$ such that*

$$\log |A_I| \leq W(I) - \delta W; \quad (4.3)$$

(B) macroscopic cross-block correlation: *there is a partition $\mathcal{P} = I \sqcup J$ into nonempty proper blocks such that*

$$\text{Corr}_{I,J}(A) \geq \delta W. \quad (4.4)$$

The constants are independent of q and of the number of prime factors.

Proof. Choose, for example, $\theta = \gamma/10$ and $\delta = \gamma/20$, and then choose ε_K so small that $c_C \varepsilon_K < \gamma/20$.

Apply Lemma 2.186. The two alternatives of that lemma are not asserted to be mutually exclusive; the argument needs only one available branch. First suppose a dominant factor p_0 satisfies $w_{p_0} > (1 - \theta)W$. If (4.3) fails for $I = \{p_0\}$, then

$$\log |A| \geq \log |A_{p_0}| > w_{p_0} - \delta W > (1 - \theta - \delta)W.$$

On the other hand, (4.1)–(4.2) give

$$\log |A| \leq (1 - \gamma)W + c_C \varepsilon_K W.$$

Our parameter choices make the first lower bound strictly larger than the second upper bound, a contradiction. Hence (A) holds.

Now suppose Lemma 2.186(ii) gives $\mathcal{P} = I \sqcup J$. Assume (A) fails for both blocks and (B) also fails. Then

$$\begin{aligned} \log |A| &= \log |A_I| + \log |A_J| - \text{Corr}_{I,J}(A) \\ &> (W(I) - \delta W) + (W(J) - \delta W) - \delta W \\ &= (1 - 3\delta)W. \end{aligned}$$

Again (4.2) gives $\log |A| \leq (1 - \gamma + c_C \varepsilon_K)W$, and the chosen constants satisfy $3\delta + c_C \varepsilon_K < \gamma$. This contradiction proves the theorem. $\square$

Corollary 2.189 (Correlation control leaves only a proper-projection obstruction). *Assume, in addition, that the correlation theorem of Chapter 2 supplies*

$$\text{Corr}_{\mathcal{P}}(A) < \delta W. \tag{4.5}$$

Then every medium slowly growing approximate subgroup has a proper block I satisfying (4.3).

Proof. By Lemma 2.187, (4.5) rules out alternative (B) in Theorem 2.188. $\square$

2.25 The global captured obstruction is impossible

Theorem 2.190 (No medium captured slow approximate subgroup at a composite modulus). *Fix $\gamma > 0$. Assume:*

- (i) *the prime-power spectral theorem of Volume XI;*
- (ii) *the uniform spectral theorem for every proper divisor of q as the induction hypothesis;*
- (iii) *the conductor-sensitive correlation interface the correlation theorem for the bounded power $A = H^C$.*

Then, after choosing the inverse-theorem parameter sufficiently small and the logarithmic starting time sufficiently large, there is no symmetric K -approximate subgroup $H \subseteq G(q)$ satisfying simultaneously

$$e^{\gamma W} \leq |H| \leq e^{(1-\gamma)W}, \quad \log K \leq \varepsilon_K W, \quad \nu_q^{(\ell)}(xH) \geq e^{-\varepsilon_c W}.$$

The assertion is uniform in q .

Proof. If q is a prime power, this is the Volume-XI base case. Assume $|\mathcal{P}| \geq 2$.

Apply Corollary 2.189. We obtain a nonempty proper block I such that

$$\log |(H^C)_I| \leq W(I) - \delta W. \quad (6.1)$$

On the other hand, Lemma 2.180 gives

$$\log |(H^C)_I| \geq W(I) - 2\varepsilon_c W. \quad (6.2)$$

Choosing $2\varepsilon_c < \delta$ makes (6.1) and (6.2) incompatible. $\square$

Remark 2.191 (Why no local-mass comparison is needed). The proof never asks the projection of H to a small prime-power factor to carry mass $p^{-O(\varepsilon n_p)}$. It projects only to one proper CRT block supplied by Theorem 2.188; the induction hypothesis makes the walk mix on that entire block, and the captured mass then forces that block projection to be almost full in one step.

Remark 2.192 (Why there is no divisor loss). The induction hypothesis is used once at the modulus selected by the structural obstruction. The spectral constant is the same uniform σ at every proper divisor; it is not multiplied down an induction chain. Likewise, the correlation defect is charged once through the correlation theorem above. Thus neither $\omega(q)$ nor the number of divisors enters any exponent.

2.26 Worked example: why a 50/50 split is not a terminal dichotomy

Suppose $G = G_I \times G_J$ with $W(I) = W(J) = W/2$, and a medium approximate set has one small projection and one dense projection. Neither block occupies “almost all” of the conductor weight. A pure small/dense dichotomy therefore says nothing. The correct argument instead measures

$$\text{Corr}_{I,J}(A) = \log |A_I| + \log |A_J| - \log |A|.$$

If this defect is small and both proper projections are large, A itself is almost full, contradicting the medium-size bound. If the defect is large, the Volume-X correlation theorem detects the cross-block obstruction. Thus the mesoscopic 50/50 case is handled directly rather than rounded into a false pure alternative.

2.27 Worked example: a dominant prime-power factor

If one factor has weight $> (1 - \theta)W$, the complementary CRT block is small but nonempty. Failure of a macroscopic deficiency on the dominant projection already forces A to have size $> \exp((1 - \theta - \delta)W)$, contradicting the medium approximate-group upper bound once $\theta, \delta \ll \gamma$. No iteration over the remaining primes is required.

Chapter 3

Global Nonconcentration and L^2 -Flattening

3.1 How the inverse theorem enters

A failure of L^2 flattening produces one medium approximate subgroup carrying nonnegligible walk mass. Chapter 2 proves that such an object cannot exist at a composite modulus once all proper divisors are already mixed. This turns the well-founded divisor induction into a one-step flattening theorem with constants independent of $\omega(q)$.

3.2 Normalization and earlier inputs

Let $G = G(q)$, $M = |G|$, and let ν be the fixed symmetric arithmetic step measure, lazified if necessary. All L^2 norms use counting measure:

$$\|f\|_2^2 = \sum_{g \in G} |f(g)|^2,$$

so uniform probability has norm $M^{-1/2}$.

We use three already-owned inputs.

- (i) The noncommutative failed-flattening inverse theorem from the earlier flattening volume.
- (ii) The fixed-generator logarithmic-girth/free-pair estimate, which gives an initial power saving in L^2 at time $C_0 \log M$.
- (iii) The proper-divisor induction theorem of Chapter 2, including its two-block structural obstruction. The same chapter supplies both the corrected conductor-weight bookkeeping and the cross-prime correlation interface used by that theorem.

3.3 The exact inverse-theorem interface

Earlier input 3.1 (Earlier owner: weighted failed-flattening inverse theorem). There is an absolute $C \geq 1$ such that if μ is a symmetric probability measure on a finite group and $K \geq 2$ satisfies

$$\|\mu * \mu\|_2 \geq K^{-1} \|\mu\|_2,$$

then there is a symmetric K^C -approximate subgroup H and x in the group such that

$$K^{-C} \|\mu\|_2^{-2} \leq |H| \leq K^C \|\mu\|_2^{-2}$$

and

$$\mu(xH) \geq K^{-C}.$$

Remark 3.2. This is not reproved here. The important point is its exact use: captured mass and the L^2 scale give the two cardinality bounds. No lower-size estimate is smuggled in from a separate entropy statement.

Corollary 3.3 (Medium inverse set in the nonflat range). *Fix $0 < \gamma < 1/4$. There is $\varepsilon_0(\gamma) > 0$ such that if*

$$M^{-1/2+\gamma} \leq \|\mu\|_2 \leq M^{-\gamma}$$

and flattening fails at scale $K = M^{\varepsilon_0}$, then Theorem 3.1 yields

$$M^\gamma \leq |H| \leq M^{1-\gamma}$$

and

$$\mu(xH) \geq M^{-C\varepsilon_0}.$$

Proof. The size is $M^{O(\varepsilon_0)} \|\mu\|_2^{-2}$. The displayed L^2 range therefore puts its logarithm between $(2\gamma - O(\varepsilon_0)) \log M$ and $(1 - 2\gamma + O(\varepsilon_0)) \log M$. Choose ε_0 so small that the $O(\varepsilon_0)$ loss is below γ . $\square$

3.4 One-step arbitrary-modulus flattening

Theorem 3.4 (One-step L^2 flattening). *Fix $\gamma > 0$. In the well-founded induction on $\Omega(q) = \sum_{p^n \parallel q} n$, assume the uniform spectral theorem for every proper divisor of q . Then there is $\varepsilon = \varepsilon(\gamma) > 0$, independent of q and $\omega(q)$, such that whenever*

$$M^{-1/2+\gamma} \leq \|\nu^{(\ell)}\|_2 \leq M^{-\gamma}, \quad \ell \geq C_{\text{start}} \log M,$$

one has

$$\|\nu^{(2\ell)}\|_2 \leq M^{-\varepsilon} \|\nu^{(\ell)}\|_2.$$

Proof. Set $\mu = \nu^{(\ell)}$. Suppose the conclusion fails. Apply Theorem 3.1 with $K = M^\varepsilon$ and then Corollary 3.3. We obtain a medium K^C -approximate subgroup H with a coset carrying mass at least $M^{-C\varepsilon}$.

Choose ε so small that the inverse-theorem approximate-group constant satisfies the small- K hypothesis of the proper-divisor induction theorem of Chapter 2 and so that the captured exponent $C\varepsilon$ is below the projection-deficiency constant there. The current time is at least $C_{\text{start}} \log M$, where C_{start} is chosen larger than the uniform mixing threshold required by the proper-divisor theorem on every proper divisor. Thus the approximate subgroup H satisfies every hypothesis of the proper-divisor theorem's “no medium captured slow approximate subgroup” theorem. This is a contradiction. Hence flattening holds. $\square$

3.5 Initial power saving

Earlier input 3.5 (Earlier arithmetic owner: logarithmic-time initial saving). There are $C_0, \gamma_0 > 0$, depending only on the fixed arithmetic generating set, such that for every allowed q ,

$$\|\nu^{(C_0 \log M)}\|_2 \leq M^{-\gamma_0}.$$

Since convolution by a probability measure is an L^2 contraction, the same bound holds at every later time. We therefore put

$$C_{\text{start}} := \max\{C_0, C_{\text{mix}}\},$$

where C_{mix} is the proper-divisor mixing constant fixed in the proper-divisor theorem.

Remark 3.6. The proof belongs to the earlier arithmetic/girth part of the series: a bounded-word free pair and integer height control give a logarithmic injectivity window, and the earlier free-group return-probability decay gives the power saving. It is uniform in the factorization of q .

3.6 Bounded-stage global flattening

Theorem 3.7 (Flattening to the conductor threshold). *For every sufficiently small $\gamma > 0$ there are C, J independent of q such that for some*

$$L \leq C \log M$$

one has

$$\|\nu^{(L)}\|_2 \leq M^{-1/2+\gamma}.$$

The number of flattening stages J is bounded solely in terms of the fixed exponents.

Proof. Start at $\ell_0 = C_{\text{start}} \log M$. Proposition 3.5 and L^2 contraction give the same initial power saving at this later time. If the norm is already at most $M^{-1/2+\gamma}$, stop. Otherwise Theorem 3.4 multiplies the norm by at most $M^{-\varepsilon}$ when time is doubled. After at most

$$J \leq \left\lceil \frac{1/2 - \gamma - \gamma_0}{\varepsilon} \right\rceil + 1$$

stages the exponent reaches $1/2 - \gamma$. Since J is constant, $L = 2^J \ell_0 = O(\log M)$. □

3.7 Worked example: exponent iteration

Suppose at the starting logarithmic time one has $\|\nu^{(\ell_0)}\|_2 \leq M^{-0.12}$, and one-step flattening improves the exponent by $\varepsilon = 0.04$ whenever the norm is above $M^{-1/2+0.05} = M^{-0.45}$. After at most nine doublings the exponent reaches at least 0.48, so the target is crossed. The number of stages is a constant; hence the final time remains $O(\log M)$ rather than acquiring a factor depending on the divisor count.

Chapter 4

Exact-Conductor Quasirandomness and the Spectral Endgame

4.1 Closing the strong induction

Assume the spectral theorem for every proper divisor of q . Chapters 2 and 3 give the full-conductor flattening estimate at q . An eigenvalue whose representation has proper exact conductor is handled by the induction hypothesis; only exact conductor q reaches the new trace–multiplicity argument. This is the key reason no sum over divisors appears.

4.2 Local dimension input and tensorization

Let $G(q) = \mathrm{SL}_2(\mathbf{Z}/q\mathbf{Z})$ with allowed prime factors. By Chapter 1, an irreducible representation has exact conductor

$$d = \prod_p p^{e_p}$$

and factors as an external tensor product of exact local-conductor representations.

Earlier input 4.1 (Earlier owners: local conductor-sensitive dimension). There are $\beta_0 > 0$ and $C_0 \geq 1$ such that, for every sufficiently large allowed p and every nontrivial irreducible representation ρ_p of exact conductor p^e ,

$$\dim \rho_p \geq C_0^{-1} |G(p^e)|^{\beta_0}.$$

For $e = 1$ this is the prime-field quasirandom bound; for $e \geq 2$ it is the Volume-XI conductor-sensitive bound.

Proposition 4.2 (No-divisor-loss global multiplicity bound). *After enlarging the finite exceptional-prime set, there is $\beta > 0$ such that every nontrivial irreducible representation of exact conductor d satisfies*

$$\dim \rho \geq |G(d)|^\beta.$$

The constant is independent of $\omega(d)$.

Proof. By tensor factorization,

$$\dim \rho = \prod_{p|d} \dim \rho_p \geq C_0^{-\omega(d)} \prod_{p|d} |G(p^{e_p})|^{\beta_0}.$$

Choose the allowed-prime threshold p_0 so large that $C_0 \leq |G(p^e)|^{\beta_0/2}$ for every $p \geq p_0$ and $e \geq 1$. Then each local factor is at least $|G(p^e)|^{\beta_0/2}$. Multiplying gives the claim with $\beta = \beta_0/2$. Thus the apparent factor $C_0^{-\omega(d)}$ is absorbed locally, not estimated by a divisor-count bound. $\square$

4.3 Trace and exact conductor

Let T_q be convolution by the fixed symmetric lazy step measure on $\ell^2(G(q))$.

Lemma 4.3 (Regular-representation multiplicity). *If an eigenvalue λ of T_q occurs in an irreducible representation ρ , then in the regular representation it occurs with multiplicity at least $\dim \rho$.*

Proof. The regular representation decomposes as

$$\bigoplus_{\rho \in \widehat{G}} (\dim \rho) \rho.$$

The convolution operator acts by the same matrix $\widehat{\nu}(\rho)$ on each of the $\dim \rho$ copies. Hence every eigenvalue of that matrix repeats at least $\dim \rho$ times. $\square$

Lemma 4.4 (Return probability / trace identity). *For every $L \geq 1$,*

$$|G(q)| \nu_q^{(2L)}(e) = \text{Tr}(T_q^{2L}), \quad \nu_q^{(2L)}(e) = \|\nu_q^{(L)}\|_2^2.$$

Consequently, if λ occurs in exact conductor d , then

$$(\dim \rho) |\lambda|^{2L} \leq |G(d)| \|\nu_d^{(L)}\|_2^2.$$

Proof. The kernel of T^{2L} has constant diagonal entry $\nu^{(2L)}(e)$, giving the trace identity. Symmetry gives $\nu^{(2L)}(e) = \sum_g \nu^{(L)}(g)^2$. If ρ has exact conductor d , work on the quotient $G(d)$ and apply Lemma 4.3. $\square$

4.4 Well-founded induction on exact conductor, not a sum over divisors

Theorem 4.5 (Uniform spectral gap for arbitrary moduli). *There is $\sigma > 0$ such that the lazy Cayley convolution operator on every allowed $G(q)$ satisfies*

$$\|T_q\|_{\ell_0^2 \rightarrow \ell_0^2} \leq 1 - \sigma.$$

Proof. We first choose the constant that the induction will preserve. Let $\sigma_{\text{pp}} > 0$ be the uniform prime-power gap supplied by Volume XI. Let $\beta > 0$ be the exponent in Proposition 4.2, and let C be the uniform logarithmic-time constant in Theorem 3.7. Fix once and for all

$$0 < \gamma < \beta/4, \quad \sigma_{\text{fc}} := 1 - \exp\left(-\frac{\beta}{4C}\right) > 0,$$

and set

$$\sigma := \min(\sigma_{\text{pp}}, \sigma_{\text{fc}}).$$

We prove the theorem with this same σ at every stage. This removes any possibility of losing spectral gap while descending through a long divisor chain.

Induction parameter. Use well-founded induction on divisibility, measured by

$$\Omega(q) = \sum_{p^n \parallel q} n.$$

The one-prime family $q = p^n$ is the Volume-XI base case and already has gap at least $\sigma_{\text{pp}} \geq \sigma$. At a modulus with at least two residue characteristics, assume that every proper divisor $d \mid q$ satisfies the spectral theorem with the same preselected σ .

Case 1: the eigenspace has proper exact conductor. Let λ be a nonconstant eigenvalue on $G(q)$ and let $d \mid q$ be the exact conductor of its irreducible representation. If $d < q$, then $\Omega(d) < \Omega(q)$. Chapter 1 identifies the component with the pullback of the corresponding component on $G(d)$, so the induction hypothesis gives immediately

$$|\lambda| \leq 1 - \sigma.$$

No trace estimate at modulus q is charged to this component.

Case 2: the eigenspace has full exact conductor. Suppose now that $d = q$. Every proper-divisor spectral hypothesis needed by Theorem 2.190 and Chapter 3 is available from Case 1's induction hypothesis. Put $M = |G(q)|$. Theorem 3.7 produces a time

$$L \leq C \log M, \quad \|\nu_q^{(L)}\|_2 \leq M^{-1/2+\gamma}.$$

Proposition 4.2 gives $\dim \rho \geq M^\beta$, while Lemma 4.4 gives the exact-conductor trace inequality. Combining them yields

$$M^\beta |\lambda|^{2L} \leq M \|\nu_q^{(L)}\|_2^2 \leq M^{2\gamma},$$

so, because $2\gamma < \beta/2$,

$$|\lambda|^{2L} \leq M^{-\beta/2}.$$

Taking logarithms and using $L \leq C \log M$ gives

$$|\lambda| \leq \exp\left(-\frac{\beta \log M}{4L}\right) \leq \exp\left(-\frac{\beta}{4C}\right) = 1 - \sigma_{\text{fc}} \leq 1 - \sigma.$$

Both cases preserve the same constant σ chosen before the induction. Thus the induction closes for every allowed modulus, with no dependence on the length of the divisor chain. $\square$

Remark 4.6 (No divisor sum). The proof never sums traces over all $d \mid q$. Every eigenspace is assigned its unique exact conductor and is treated there. Proper conductors are handled by the well-founded proper-divisor induction; the full conductor is handled once by global multiplicity plus flattening. In particular, a proper conductor obtained merely by lowering one p -adic exponent is covered even when it has the same prime support as q . This is the promised “exact-conductor trace estimate with no summation loss over divisors.”

4.5 Super-approximation and consequences

Earlier input 4.7 (Earlier arithmetic owner: prime-field strong approximation). For the fixed non-elementary arithmetic subgroup $\Gamma < \text{SL}_2(\mathbf{Z})$, reduction modulo every sufficiently large allowed prime is $\text{SL}_2(\mathbf{F}_p)$. Volume XI then lifts this surjectivity to every p^n , and Chapter 1 combines the local images by CRT.

Corollary 4.8 (Rank-one super-approximation). *Let $\Gamma < \mathrm{SL}_2(\mathbf{Z})$ be the fixed non-elementary arithmetic subgroup satisfying the strong-approximation hypotheses used throughout the series. Outside the finite exceptional-prime set, the Cayley graphs of its congruence images modulo arbitrary q form a uniform expander family.*

Proof. Earlier input 4.7, together with the Volume-XI lifting theorem and CRT, identifies the congruence image with the required $G(q)$ outside the finite exceptional set. Theorem 4.5 supplies a uniform spectral gap for the fixed generating set. $\square$

Remark 4.9 (Historical source boundary: what is uniform in the arbitrary-modulus theorem). Bourgain–Varjú formulate the arbitrary-modulus expansion theorem for a *fixed* finite symmetric set $S \subset \mathrm{SL}_d(\mathbf{Z})$ whose generated subgroup is Zariski dense, while the modulus q varies [1]. Corollary 4.8 has the same uniformity variable in our rank-one setting: S is fixed once and for all. The rank $\Omega(q) = \sum n_p$ used in the proof of Theorem 4.5 is only an internal device for well-founded induction; it is not presented as the quantitative modulus weight of the historical proof. All quantitative size charges in this volume use logarithmically weighted conductor data, as recorded in the front-matter proof architecture.

Corollary 4.10 (Property (τ)). *The fixed arithmetic group has property (τ) with respect to the full allowed congruence family.*

Proof. This is the definition/equivalent spectral formulation of property (τ) applied to Corollary 4.8. $\square$

Corollary 4.11 (Uniform logarithmic mixing and diameter). *For the lazy walk, total-variation mixing time is $O(\log |G(q)|)$ uniformly in q . The Cayley-graph diameter is also $O(\log |G(q)|)$.*

Proof. A uniform spectral gap contracts ℓ_0^2 exponentially. Starting from a point mass, after $C \log |G|$ steps the L^2 error is $O(|G|^{-1/2})$, hence total variation is bounded by Cauchy–Schwarz. If a ball of radius L missed a vertex, the walk at time L could not be close to uniform; taking the mixing constant large gives the diameter bound. $\square$

4.6 Worked example: exact conductor $p^2 r$

Let $q = p^3 r^2$ and suppose an irreducible constituent factors as $\rho_p \boxtimes \rho_r$ with exact local conductors p^2 and r . Its exact conductor is $d = p^2 r$. Tensor factorization gives

$$\dim \rho = \dim \rho_p \dim \rho_r \geq |G(p^2)|^\beta |G(r)|^\beta = |G(d)|^\beta.$$

If $d < q$, the eigenvalue is controlled by the already-established spectral theorem at the proper divisor d ; it is not included in a divisor sum at modulus q . Only constituents of exact conductor q are charged to the new full-conductor trace estimate.

Historical references

- [1] J. Bourgain and P. P. Varjú, *Expansion in $SL_d(\mathbf{Z}/q\mathbf{Z})$, q arbitrary*, Invent. Math. 188 (2012), 151–173.
- [2] J. Bourgain, A. Furman, E. Lindenstrauss, and S. Mozes, *Stationary measures and equidistribution for orbits of non-abelian semigroups on the torus*, J. Amer. Math. Soc. 24 (2011), 231–280.
- [3] J. Bourgain, *The discretized sum-product and projection theorems*, J. Anal. Math. 112 (2010), 193–236.
- [4] W. He, *Discretized sum-product estimates in matrix algebras*, J. Anal. Math. 139 (2019), 637–676.
- [5] W. He, *Orthogonal projections of discretized sets*, J. Fractal Geom. 7 (2020), 271–317.
- [6] Y. Guivarc’h and É. Le Page, *Spectral gap properties for linear random walks and Pareto’s asymptotics for affine stochastic recursions*, Ann. Inst. H. Poincaré Probab. Statist. 52 (2016), 503–574; arXiv:1204.6004.
- [7] Y. Guivarc’h and A. Raugi, *Frontière de Furstenberg, propriétés de contraction et théorèmes de convergence*, Publications des séminaires de mathématiques et informatique de Rennes, Séminaire de probabilités 1 (1980), exposé no. 2, 1–36.
- [8] P. P. Varjú, *Expansion in $SL_d(\mathcal{O}_K/I)$, I square-free*, J. Eur. Math. Soc. 14 (2012), 273–305.

The Bourgain–Gamburd Method
Volume XIII: Expansion, Orbit Counting, and the Affine Sieve

18 August 2026

Scope and historical note

This volume combines congruence expansion with a dimension-one combinatorial sieve. The expansion side supplies a level of distribution for congruence conditions along a thin orbit. The sieve side is developed internally using the dimension-one Rosser–Iwaniec parameter $\beta_R = 2$, exact finite-depth switching recurrences, positive delay majorants, and a fixed-ratio fundamental lemma. The polynomial

$$F(g) = 1 + Q_0 g_{11}$$

is kept concrete throughout so that admissibility, local densities, and the sieve dimension can be calculated explicitly. Historical references identify the original affine-sieve and Rosser sources; the active chapters contain the proof route used in this volume.

Proof scope of the sieve chapter. The historical affine-sieve argument supplies the expansion-to-level-of-distribution interface, while this volume proves the dimension-one combinatorial sieve mechanism in its own right. Accordingly, the Rosser–Iwaniec recursion below is not used as a one-line black box: its support rules, switching recurrence, and fundamental-lemma estimate are proved in Chapter 3.

How to read this volume

There are two independent engines. Expansion supplies a level of distribution for congruence conditions along the thin orbit; the combinatorial sieve converts that distribution into almost-prime values. Chapter 3 is where the second engine really lives. Do not read the Rosser weights as formal inclusion–exclusion notation: each deleted branch corresponds to a first failure of the support level, and the entire lower-bound argument rests on proving that the total mass of these deep ordered branches is tiny. Chapter 4 then plugs the concrete local densities into that abstract sieve machine.

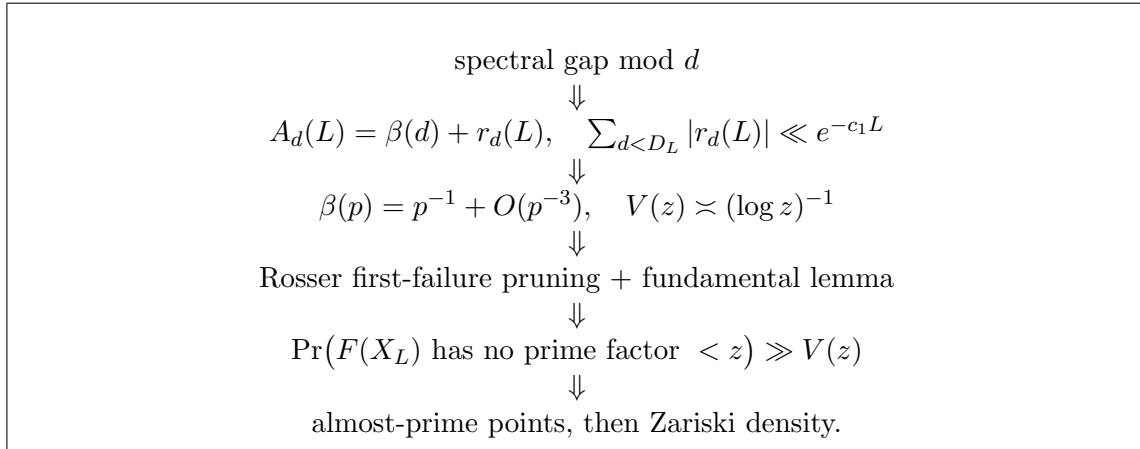


Figure 1: The affine-sieve proof as two engines joined at the level of distribution: expansion supplies the remainder bound; the combinatorial sieve consumes it.

Series position. This volume uses the congruence expansion supplied through Volume XII. Its main outputs are a self-contained dimension-one Rosser sieve and the affine-sieve almost-prime application.

Quantifier and asymptotic conventions. Unless a statement explicitly says otherwise, every constant occurring in a quantitative theorem is chosen in the order in which it is quantified in that theorem. Thus a phrase such as “for all sufficiently small δ ” means: there is a displayed or locally quantified threshold $\delta_0 > 0$, depending only on the parameters listed before that phrase, such that the assertion holds for every $0 < \delta < \delta_0$. The phrases “sufficiently large N ” and “sufficiently large p ” have the analogous meaning.

If $X \ll_\theta Y$ or $X = O_\theta(Y)$, the implicit constant depends only on θ and on the fixed ambient data of the current volume; it is independent of the variables quantified after θ . The words *absolute*, *uniform*, and *bounded* never introduce an unlisted dependence. Whenever an inequality between exponents or scales is essential to the argument, that inequality is written explicitly in the proof; these conventions are not a substitute for such a calculation.

Contents

Scope and historical note	i
How to read this volume	ii
1 Congruence Orbits and Expansion	1
1.1 Purpose	1
1.2 Equidistribution on congruence quotients	1
1.3 The local divisor density	2
1.4 Worked residue-count example	2
2 Sieve Dimension and Local Factors	3
2.1 An internal Mertens package	3
2.2 Level of distribution from expansion	5
2.3 Worked diagnostic: why the $O(p^{-3})$ term is harmless	5
3 Rosser Weights and the Combinatorial Lower Sieve	6
3.1 Purpose and exact scope	6
3.2 Weighted sieve data	6
3.3 Rosser lower and upper weights	7
3.4 Buchstab switching before Rosser pruning	8
3.5 The switching identity	9
3.6 The Rosser–Iwaniec terminal-tail theorem	10
3.6.1 Local Mertens control	10
3.6.2 Finite-depth Rosser tails and the exact depth recursion	10
3.6.3 The positive delay envelope	12
3.6.4 A source-faithful discrete comparison	16
3.7 The fixed-ratio fundamental lemma	19
3.8 Lower and upper sieve theorems	19
3.9 A diagnostic example	20
3.10 Worked first-failure example	20
4 The Affine Sieve and Almost-Prime Orbits	21
4.1 The abstract random-walk affine sieve	21
4.2 The concrete rank-one polynomial	23
4.3 An elementary algebraic point-count lemma	23
4.4 A prime of controlled exponential size	24
4.5 Exponential escape from each proper algebraic subvariety	25
4.6 Worked saturation calculation	26

Chapter 1

Congruence Orbits and Expansion

Proof architecture. The sieve sees the random walk only through finite congruence quotients. Uniform expansion supplies a quantitative replacement for uniform counting; exact local counting of the condition $F(g) \equiv 0$ supplies the local density; Chinese remaindering makes the squarefree local factors multiplicative.

1.1 Purpose

This module proves the arithmetic and probabilistic inputs needed by the internal Rosser sieve. It supplies the local-density and level-of-distribution inputs for the concrete rank-one affine-sieve model used later.

Let $\Gamma < \mathrm{SL}_2(\mathbb{Z})$ be a fixed finitely generated non-elementary arithmetic subgroup with a symmetric lazy finitely supported probability measure μ . Let X_L be the length- L random walk. Let Q_0 be divisible by the finite exceptional set of primes excluded by strong approximation and by all small primes discarded in the uniform expansion statements. We sieve the integral polynomial

$$F(g) = 1 + Q_0 g_{11}.$$

Notice that $F(g) \neq 0$ for $g \in \mathrm{SL}_2(\mathbb{Z})$ because $Q_0 \geq 2$.

Earlier input (Volume XII: arbitrary-modulus spectral gap). *There is $\sigma > 0$ such that, for every allowed modulus q coprime to Q_0 , the averaging operator of μ on the nonconstant subspace of $\ell^2(G_q)$, $G_q = \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$, has norm at most $1 - \sigma$.*

Earlier input (Earlier arithmetic reduction / strong approximation). *For every modulus q coprime to Q_0 , the reduction of Γ is the whole G_q .*

1.2 Equidistribution on congruence quotients

Theorem 1.2.1 (Uniform random-walk equidistribution). *Let $\nu_{L,q}$ be the distribution of X_L modulo an allowed q , and let u_q be uniform measure on G_q . Then*

$$\|\nu_{L,q} - u_q\|_2 \leq (1 - \sigma)^L \|\delta_e - u_q\|_2 \leq e^{-cL}$$

for $c = -\log(1 - \sigma) > 0$. Consequently, for every $B \subseteq G_q$,

$$\left| \Pr(X_L \bmod q \in B) - \frac{|B|}{|G_q|} \right| \leq |B|^{1/2} e^{-cL}.$$

Proof. The first inequality is the spectral-gap estimate applied L times on the orthogonal complement of constants. The second follows from $\|\delta_e - u_q\|_2 \leq 1$. Pairing $\nu_{L,q} - u_q$ with 1_B and applying Cauchy–Schwarz proves the subset estimate. $\square$

1.3 The local divisor density

For d squarefree and $(d, Q_0) = 1$, put

$$B_d = \{g \in G_d : F(g) \equiv 0 \pmod{d}\}, \quad \beta(d) = |B_d|/|G_d|.$$

Proposition 1.3.1 (Prime local density). *For an allowed prime p ,*

$$\beta(p) = \frac{p}{p^2 - 1} = \frac{1}{p} + O(p^{-3}).$$

Proof. Modulo p , Q_0 is invertible and $F(g) = 0$ is equivalent to $g_{11} = a_p := -Q_0^{-1} \neq 0$. Write $g = \begin{pmatrix} a_p & b \\ c & d \end{pmatrix}$. For each of the p^2 pairs (b, c) there is a unique

$$d = \frac{1 + bc}{a_p}$$

with determinant one. Hence $|B_p| = p^2$. Since $|\mathrm{SL}_2(\mathbb{F}_p)| = p(p^2 - 1)$, the formula follows. $\square$

Proposition 1.3.2 (Multiplicativity of local densities). *If d is squarefree and $(d, Q_0) = 1$, then*

$$\beta(d) = \prod_{p|d} \beta(p).$$

Proof. The Chinese remainder theorem identifies G_d with $\prod_{p|d} G_p$, and the congruence $F(g) \equiv 0 \pmod{d}$ with the independent conjunction of the congruences modulo the prime factors. Both numerator and denominator therefore factor. $\square$

1.4 Worked residue-count example

For $p = 5$ and a nonzero prescribed upper-left entry a , the pairs $(b, c) \in \mathbb{F}_5^2$ give exactly 25 matrices, because $d = (1 + bc)/a$ is forced. Since $|\mathrm{SL}_2(\mathbb{F}_5)| = 120$, the local density is $25/120 = 5/24$, exactly $p/(p^2 - 1)$.

Chapter 2

Sieve Dimension and Local Factors

Why this chapter is internal. The lower sieve needs quantitative control of the Euler product, not merely the slogan that it has dimension one. We therefore derive the needed prime harmonic estimate from the von Mangoldt identity, Chebyshev bounds, and Stirling, and only then identify the sieve dimension.

2.1 An internal Mertens package

We include the elementary argument because the sieve-dimension estimate must be proved explicitly before the sieve is used.

Lemma 2.1.1 (Chebyshev bound for the von Mangoldt sum). *With $\psi(x) = \sum_{n \leq x} \Lambda(n)$, one has $\psi(x) = O(x)$.*

Proof. For every prime power p^k with $m < p^k \leq 2m$, Legendre's formula gives

$$v_p \binom{2m}{m} = \sum_{j \geq 1} \left(\left\lfloor \frac{2m}{p^j} \right\rfloor - 2 \left\lfloor \frac{m}{p^j} \right\rfloor \right).$$

Each summand is 0 or 1, and the term $j = k$ is 1 when $m < p^k \leq 2m$. Hence every logarithmic contribution occurring in $\psi(2m) - \psi(m)$ is present in the prime factorization of the central binomial coefficient, so

$$\psi(2m) - \psi(m) \leq \log \binom{2m}{m}.$$

Since $\binom{2m}{m} \leq \sum_{j=0}^{2m} \binom{2m}{j} = 2^{2m}$, the right side is at most $2m \log 2$. Apply this with $m = \lfloor x/2 \rfloor, \lfloor x/4 \rfloor, \dots$ and sum the resulting dyadic increments. The geometric series is $O(x)$, proving $\psi(x) = O(x)$. $\square$

Lemma 2.1.2 (First Mertens estimate). *As $x \rightarrow \infty$,*

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + O(1).$$

Proof. The identity $\log n = \sum_{d|n} \Lambda(d)$ gives

$$\log N! = \sum_{d \leq N} \Lambda(d) \left\lfloor \frac{N}{d} \right\rfloor = N \sum_{d \leq N} \frac{\Lambda(d)}{d} + O(\psi(N)).$$

By Lemma 2.1.1 and Stirling's formula $\log N! = N \log N - N + O(\log N)$,

$$\sum_{d \leq N} \frac{\Lambda(d)}{d} = \log N + O(1).$$

The contribution of prime powers p^k , $k \geq 2$, is absolutely bounded:

$$\sum_p \sum_{k \geq 2} \frac{\log p}{p^k} < \infty.$$

Removing it proves the claim. □

Lemma 2.1.3 (Prime harmonic sum). *There is a constant M such that*

$$\sum_{p \leq x} \frac{1}{p} = \log \log x + M + O(1/\log x).$$

Proof. Apply partial summation to Lemma 2.1.2 with the weight $1/\log t$:

$$\sum_{p \leq x} \frac{1}{p} = \frac{A(x)}{\log x} + \int_2^x \frac{A(t)}{t(\log t)^2} dt, \quad A(t) = \sum_{p \leq t} \frac{\log p}{p}.$$

Insert $A(t) = \log t + O(1)$. The main integral is $\log \log x + O(1)$; the error integral converges with tail $O(1/\log x)$. □

Worked level-of-distribution budget

From $|r_d(L)| \ll d^{3/2} e^{-cL}$ one obtains, for $D_L = e^{\eta L}$,

$$\sum_{d < D_L} |r_d(L)| \ll e^{-cL} \sum_{d < D_L} d^{3/2} \ll \exp\left((-c + \tfrac{5}{2}\eta)L\right).$$

Thus every fixed $\eta < 2c/5$ gives exponential decay. This is the exact place where the expansion constant becomes a level-of-distribution exponent; there is no appeal to a qualitative “sufficiently small” parameter.

Theorem 2.1.4 (Euler product and sieve dimension one). *Let*

$$V(z) = \prod_{\substack{p < z \\ p \nmid Q_0}} (1 - \beta(p)).$$

Then for some $C_F > 0$,

$$V(z) = \frac{C_F}{\log z} (1 + O(1/\log z)).$$

In particular, uniformly for $z \geq w \geq z_0$,

$$\sum_{w \leq p < z} \beta(p) = \log \frac{\log z}{\log w} + O(1/\log w), \quad \frac{V(w)}{V(z)} \ll \frac{\log z}{\log w}.$$

Proof. Proposition 1.3.1 gives $\beta(p) = p^{-1} + O(p^{-3})$ and $\sum_p \beta(p)^2 < \infty$. Hence Lemma 2.1.3 gives

$$\sum_{p < z} \beta(p) = \log \log z + C_1 + O(1/\log z).$$

Using $\log(1 - u) = -u + O(u^2)$,

$$\log V(z) = -\log \log z + C_2 + O(1/\log z).$$

Exponentiation proves the first assertion; taking ratios and differences gives the remaining estimates. Deleting the finitely many primes dividing Q_0 only changes C_F . $\square$

2.2 Level of distribution from expansion

Let

$$A_d(L) = \Pr(d \mid F(X_L)).$$

Theorem 2.2.1 (Uniform progression count). *For every squarefree d coprime to Q_0 ,*

$$A_d(L) = \beta(d) + r_d(L), \quad |r_d(L)| \leq C d^{3/2} e^{-cL}.$$

Proof. Apply Theorem 1.2.1 to B_d . Since $|B_d| \leq |G_d|$ and $|G_d| \leq d^3$, the subset discrepancy is at most $d^{3/2} e^{-cL}$. The uniform density is $|B_d|/|G_d| = \beta(d)$ by definition and Proposition 1.3.2. $\square$

Theorem 2.2.2 (Exponential level of distribution). *There exists $\eta_0 > 0$ such that, for every fixed $0 < \eta < \eta_0$ and $D_L = e^{\eta L}$,*

$$\sum_{\substack{d < D_L \\ d \text{ squarefree} \\ (d, Q_0) = 1}} |r_d(L)| \ll e^{-c_1 L}$$

for some $c_1 = c_1(\eta) > 0$.

Proof. The progression estimate has an exponential gain in the walk length and a polynomial loss in the modulus. Summing the latter explicitly gives

$$\sum_{d < D} |r_d(L)| \ll e^{-cL} \sum_{d < D} d^{3/2} \ll D^{5/2} e^{-cL}.$$

Now put $D = D_L = e^{\eta L}$. The right-hand side becomes

$$\exp\left(-\left(c - \frac{5}{2}\eta\right)L\right).$$

Choose once and for all $\eta_0 < 2c/5$. For every fixed $0 < \eta < \eta_0$, the number

$$c_1 := c - \frac{5}{2}\eta$$

is positive, and the preceding display is $O(e^{-c_1 L})$. This is the claimed exponential level of distribution. The calculation also records the precise budget: the exponent $3/2$ in the single-modulus discrepancy costs $5\eta/2$ after summation over $d < D_L$. $\square$

2.3 Worked diagnostic: why the $O(p^{-3})$ term is harmless

Because $\sum_p p^{-3} < \infty$, replacing $1/p$ by $p/(p^2 - 1) = 1/p + O(p^{-3})$ changes $\sum_{p < z} \beta(p)$ and $\log V(z)$ only by convergent constants. It changes the leading constant in $V(z)$ but not the exponent 1 of $\log z$.

Chapter 3

Rosser Weights and the Combinatorial Lower Sieve

The bottleneck. Inclusion–exclusion itself is easy; the difficulty is to truncate it at level D without losing the main term. Rosser’s first-failure tree does precisely this. The critical point is the terminal-branch factorial estimate: the Rosser–Iwaniec switching tails are governed by a positive delay system whose solutions have superfactorial decay. The delicate point is the discrete-to-continuous comparison. We organize the induction by Rosser depth, as in Iwaniec, and use the stronger two-sided local Mertens estimate already available in this affine-sieve setting to split the proof into a precise large-endpoint comparison and a direct moderate-endpoint factorial estimate.

3.1 Purpose and exact scope

First reading versus second reading. For the first pass, read the Bonferroni inequality, the first-failure recursion, the terminal-branch factorial estimate, and the fixed-ratio fundamental lemma. These four steps already explain why truncating inclusion–exclusion leaves a positive main term. The delay equations and Stieltjes comparison should then be read a second time as the quantitative device proving that the first-failure tail is uniformly small in the Rosser depth.

We prove a deliberately non-sharp, fixed-ratio form of the dimension-one fundamental lemma directly from the Rosser–Buchstab switching tree. The critical estimate is the terminal-branch bound: after the exact finite-depth recurrences are written down, positive delay majorants and a discrete-to-continuous comparison show that the surviving first-failure mass decays superfactorially. For one sufficiently large absolute sieve ratio s_0 , the lower Rosser main term is therefore a fixed positive fraction of the Euler product. This is stronger than the affine-sieve application later needs, and no external fundamental-lemma theorem is imported.

3.2 Weighted sieve data

Let $\mathcal{P}$ be a set of primes and put

$$P(z) = \prod_{\substack{p < z \\ p \in \mathcal{P}}} p.$$

A finite weighted sequence is a finite set $\mathcal{A}$ with nonnegative weights $a(x)$. Write

$$X = \sum_{x \in \mathcal{A}} a(x), \quad A_d = \sum_{\substack{x \in \mathcal{A} \\ d|x}} a(x).$$

We assume that for squarefree $d \mid P(z)$,

$$A_d = Xg(d) + r_d, \tag{3.1}$$

where g is multiplicative, $0 \leq g(p) < 1$, and

$$V(y) = \prod_{\substack{p < y \\ p \in \mathcal{P}}} (1 - g(p)).$$

The sifted mass is

$$S(\mathcal{A}, z) = \sum_{\substack{x \in \mathcal{A} \\ (x, P(z))=1}} a(x).$$

Definition 3.2.1 (Strong dimension-one hypothesis). We say that g has strong dimension one if there are constants C_0, z_0 such that for $z \geq w \geq z_0$,

$$\left| \sum_{w \leq p < z} g(p) - \log \frac{\log z}{\log w} \right| \leq \frac{C_0}{\log w}. \tag{3.2}$$

$$\frac{V(w)}{V(z)} \leq C_0 \frac{\log z}{\log w}. \tag{3.3}$$

Primes below z_0 are absorbed into constants.

3.3 Rosser lower and upper weights

Fix $D \geq z \geq 2$ and set $s = \log D / \log z$. We use the *dimension-one Rosser–Iwaniec parameter*

$$\beta_R = 2.$$

Thus a prefix is cut when its product times the square of its smallest prime reaches D . This is the normalization used in Iwaniec’s dimension-one linear sieve; it is important here, because replacing 2 by 1 changes the boundary problem for the delay recursion. If $d = p_1 \cdots p_r$ with $z > p_1 > \cdots > p_r$, define $\rho_d^\pm$ by

$$\rho_{p_1 \cdots p_r}^\pm = 0 \iff p_1 \cdots p_r p_r^2 \geq D \quad \text{and} \quad r \equiv \nu_\pm \pmod{2},$$

where $\nu_+ = 1$ and $\nu_- = 0$; otherwise $\rho_d^\pm = 1$. Put

$$\sigma_d^\pm = \prod_{j=1}^r \rho_{p_1 \cdots p_j}^\pm, \quad \lambda_d^\pm = \mu(d) \sigma_d^\pm,$$

with $\lambda_1^\pm = 1$.

Lemma 3.3.1 (First-failure Bonferroni inequality). *For every integer m ,*

$$\sum_{d|(m, P(z))} \lambda_d^- \leq \mathbf{1}_{(m, P(z))=1} \leq \sum_{d|(m, P(z))} \lambda_d^+.$$

Proof. Order the prime divisors of $(m, P(z))$ as $q_1 > \dots > q_t$. Ordinary inclusion–exclusion is exact after all t primes. The Rosser rule follows the alternating inclusion–exclusion chain until the first prefix on the parity appropriate to $+$ or $-$ for which $q_1 \dots q_j q_j^2 \geq D$, and then deletes that branch and all descendants. At an odd truncation the alternating sum is an upper bound for zero; at an even truncation it is a lower bound. Applying this independently at the first deleted prefix proves the two pointwise inequalities. Equivalently, pair every deleted descendant with the unique first deleted prefix; the remaining tail has the sign $+$ in the upper tree and $-$ in the lower tree. $\square$

Lemma 3.3.2 (Support below the level). *If $\lambda_d^\pm \neq 0$, then $d < D$.*

Proof. Write $d = p_1 \dots p_r$. If r has the active parity, the last surviving condition says $dp_r^2 < D$, hence $d < D$. If r has the opposite parity, the prefix d/p_r has active parity and survives, so $(d/p_r)p_{r-1}^2 < D$. Since $p_r < p_{r-1}$,

$$d = (d/p_r)p_r < (d/p_r)p_{r-1} < D/p_{r-1} < D.$$

Thus again $d < D$. $\square$

Summing Lemma 3.3.1 against $a(x)$ gives

$$\sum_{d|P(z)} \lambda_d^- A_d \leq S(\mathcal{A}, z) \leq \sum_{d|P(z)} \lambda_d^+ A_d. \quad (3.4)$$

By Lemma 3.3.2, only $d < D$ occurs.

3.4 Buchstab switching before Rosser pruning

Lemma 3.4.1 (Buchstab identity). *For $2 \leq w < z$,*

$$S(\mathcal{A}, z) = S(\mathcal{A}, w) - \sum_{\substack{w \leq p < z \\ p \in \mathcal{P}}} S(\mathcal{A}_p, p),$$

where $\mathcal{A}_p$ is the weighted subsequence of elements divisible by p , with the factor p removed only notationally.

Proof. Every element counted by $S(\mathcal{A}, w)$ either has no prime factor from $[w, z]$, in which case it is counted by $S(\mathcal{A}, z)$, or has a unique least prime factor p in $[w, z]$. In the latter case it is counted exactly once by $S(\mathcal{A}_p, p)$. This partitions the w -sifted elements. $\square$

Iterating Lemma 3.4.1 produces a tree indexed by strictly decreasing primes. Rosser’s rule below is a parity-sensitive pruning of this Buchstab tree at the first prefix whose size reaches the support level D .

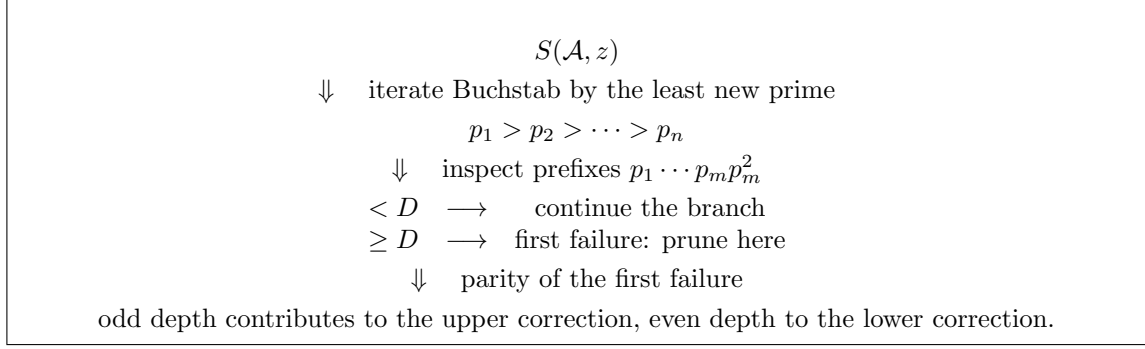


Figure 3.1: Rosser pruning of the Buchstab tree. The terminal branch is owned by its unique first prefix crossing the support level D ; parity determines whether it corrects the upper or lower sieve.

3.5 The switching identity

For $n \geq 1$ define the terminal branch sum

$$V_n(D, z) = \sum_{\substack{z > p_1 > \cdots > p_n \\ p_1 \cdots p_m p_m^2 < D \ (m < n, m \equiv n \pmod{2}) \\ p_1 \cdots p_n p_n^2 \geq D}} g(p_1) \cdots g(p_n) V(p_n). \quad (3.5)$$

Proposition 3.5.1 (Exact Rosser switching identity). *Let*

$$V^\pm(D, z) = \sum_{d|P(z)} \lambda_d^\pm g(d).$$

Then

$$V^+(D, z) = V(z) + \sum_{\substack{n \geq 1 \\ n \text{ odd}}} V_n(D, z), \quad V^-(D, z) = V(z) - \sum_{\substack{n \geq 1 \\ n \text{ even}}} V_n(D, z).$$

Proof. Expand $V(z) = \prod_{p < z} (1 - g(p))$ by inclusion–exclusion and compare it term by term with the Rosser-pruned expansion. Every term present in one expansion and absent from the other has a unique first prefix at which the Rosser rule switches off. If that first prefix has length n , all earlier prefixes of the same parity satisfy the strict inequality in (3.5), while the n th satisfies the terminal inequality. Summing all descendants of that prefix over primes below p_n contributes exactly $V(p_n)$. The sign is $+$ for an odd first failure and $-$ for an even first failure. This partitions the symmetric difference of the two expansions and gives the two identities. $\square$

Lemma 3.5.2 (Terminal depth). *If $n + 2 \leq s$, then $V_n(D, z) = 0$.*

Proof. A nonempty summand would satisfy

$$D \leq p_1 \cdots p_n p_n^2 < z^{n+2}.$$

Since $D = z^s$, this forces $n + 2 > s$. $\square$

3.6 The Rosser–Iwaniec terminal-tail theorem

The rest of this chapter follows the architecture of Iwaniec’s original *Rosser’s sieve*. We deliberately keep only the dimension-one estimates needed later. The original paper treats a general dimension parameter and optimizes the continuous sieve functions; here the stronger two-sided hypothesis in Definition 3.2.1 allows us to suppress those optimizations while retaining the essential discrete comparison.

3.6.1 Local Mertens control

Lemma 3.6.1 (Local Mertens control and the exact Rosser kernel). *Assume Definition 3.2.1. After enlarging z_0 there is a constant $C_2 = C_2(C_0)$ with the following properties. For every $p \geq z_0$,*

$$g(p) \leq \frac{C_2}{\log p}. \quad (\text{K1})$$

If $z \geq w \geq z_0$ and $R = \log z / \log w$, then

$$\frac{V(w)}{V(z)} = R \exp \left\{ O_{C_0} \left(\frac{1}{\log w} \right) \right\}. \quad (\text{K2})$$

Finally the weighted prime kernel telescopes exactly:

$$\sum_{w \leq p < z} g(p)V(p) = V(w) - V(z). \quad (\text{K3})$$

Proof. Apply (3.2) with $w = p$ and $z = p + 1$. The main logarithmic term is $O((p \log p)^{-1})$, so positivity gives (K1), after changing the constant. Enlarge z_0 so that $g(p) \leq 1/2$ for $p \geq z_0$. Then

$$\log \frac{V(w)}{V(z)} = \sum_{w \leq p < z} -\log(1 - g(p)) = \sum_{w \leq p < z} g(p) + O \left(\sum_{w \leq p < z} g(p)^2 \right).$$

On a logarithmic shell $e^k \leq p < e^{k+1}$, (K1) and (3.2) give respectively $g(p) \ll k^{-1}$ and $\sum g(p) \ll k^{-1}$, hence the square-sum on that shell is $O(k^{-2})$. Summing from $k \asymp \log w$ proves $\sum_{p \geq w} g(p)^2 \ll 1/\log w$, and (K2) follows from (3.2). For (K3), insert the prime factors of the Euler product one at a time: at a prime p the product drops by exactly $g(p)V(p)$. Telescoping gives the identity. $\square$

3.6.2 Finite-depth Rosser tails and the exact depth recursion

For $R \geq 0$ define the normalized odd and even terminal tails

$$\mathcal{T}_{R,z}^+(s) = \frac{1}{V(z)} \sum_{r=0}^R V_{2r+1}(z^s, z), \quad \mathcal{T}_{R,z}^-(s) = \frac{1}{V(z)} \sum_{r=1}^R V_{2r}(z^s, z),$$

where an empty sum is zero. This is Iwaniec’s finite-depth organization: R counts *pairs* of switching steps rather than raw depth. It is the right organization because deleting the first prime from an even terminal branch produces an odd branch of depth index $R - 1$, whereas deleting it from an odd branch produces an even branch of the same index.

Lemma 3.6.2 (Exact finite-depth Rosser recurrences). *Put*

$$s_p = \frac{\log(z^s/p)}{\log p}.$$

Then the following identities hold.

1. *If $1 < s \leq 3$, then*

$$\mathcal{T}_{0,z}^+(s) = \frac{V(z^{s/3})}{V(z)} - 1,$$

and $\mathcal{T}_{0,z}^+(s) = 0$ for $s > 3$.

2. *For $s \geq 2$ and $R \geq 1$,*

$$\mathcal{T}_{R,z}^-(s) = \sum_{z^{s/(2R+2)} \leq p < z} g(p) \frac{V(p)}{V(z)} \mathcal{T}_{R-1,p}^+(s_p). \quad (\text{R-})$$

3. *For $s \geq 3$ and $R \geq 1$,*

$$\mathcal{T}_{R,z}^+(s) = \sum_{z^{s/(2R+3)} \leq p < z} g(p) \frac{V(p)}{V(z)} \mathcal{T}_{R,p}^-(s_p). \quad (\text{R+})$$

4. *For $1 < s \leq 3$ and $R \geq 1$, put*

$$z_s = (z^s)^{1/3} = z^{s/3}.$$

Then

$$\mathcal{T}_{R,z}^+(s) = \mathcal{T}_{0,z}^+(s) + \frac{V(z_s)}{V(z)} \mathcal{T}_{R,z_s}^+(3). \quad (\text{R0})$$

Proof. The first identity is exactly the depth-one terminal condition $p^3 \geq z^s$: by the Euler-product telescope (K3),

$$\frac{1}{V(z)} \sum_{z^{s/3} \leq p < z} g(p)V(p) = \frac{V(z^{s/3})}{V(z)} - 1.$$

For (R-) fix the first prime $p = p_1$ in an even terminal chain of length $2r$. Removing p changes the level from z^s to z^s/p , the endpoint from z to p , and leaves an odd first-failure chain of index $r - 1$. The support condition $2r \leq 2R$ forces $p \geq (z^s)^{1/(2R+2)}$. Conversely every child chain counted on the right reconstructs a unique parent chain, so the identity is exact. The proof of (R+) is identical: an odd chain of length $2r + 1$ becomes an even chain of index r , and the support condition gives $p \geq (z^s)^{1/(2R+3)}$. Finally, let $1 < s \leq 3$ and put $z_s = D^{1/3} = z^{s/3}$. The depth-one branches contribute the first term in (R0). A longer odd branch must first survive the odd prefix of length one, namely $p_1^3 < D$, so necessarily $p_1 < z_s$. After this restriction the raw longer-branch sum is exactly the odd tail with the same level D and endpoint z_s ; relative to that endpoint the parameter is $\log D / \log z_s = 3$. Our tails are normalized by the Euler product at their endpoint, so changing the normalization from $V(z_s)$ back to $V(z)$ contributes the factor $V(z_s)/V(z)$. The depth-one tail at parameter 3 and endpoint z_s is zero (the endpoint inequality is strict), so no term is double counted. This proves (R0). $\square$

3.6.3 The positive delay envelope

The discrete recurrences above have a continuous model. We need only a positive envelope with a coarse superfactorial tail, not the optimized linear sieve asymptotic.

Lemma 3.6.3 (Positive dimension-one delay functions). *There are positive functions $q^+, q^- : (0, \infty) \rightarrow (0, \infty)$ with*

$$q^+(s) = \frac{1}{2} \quad (0 < s \leq 3), \quad q^-(s) = 1 \quad (0 < s \leq 2), \quad (\text{Q0})$$

and, on the remaining intervals,

$$(q^+)'(s) = -\frac{s}{(s-1)^2} q^-(s-1) \quad (s > 3), \quad (q^-)'(s) = -\frac{s}{(s-1)^2} q^+(s-1) \quad (s > 2). \quad (\text{Q1})$$

They satisfy

$$q^\pm(s) = \int_s^\infty \frac{t}{(t-1)^2} q^\mp(t-1) dt \quad (\text{Q2})$$

in the ranges of (Q1), and there are constants $c, C > 0$ such that

$$q^+(s) + q^-(s) \leq C e^{-cs \log(3s)} \quad (s \geq 3). \quad (\text{Q3})$$

Moreover

$$q^\pm(s-1) \leq C s^2 q^\pm(s) \quad (s \geq 4). \quad (\text{Q4})$$

Proof. The method of steps gives unique continuous solutions of (Q0)–(Q1). The point is to prove that they stay positive and decay. Set

$$a(s) = \frac{q^+(s) + q^-(s)}{s^2}, \quad b(s) = \frac{q^+(s) - q^-(s)}{s^2}.$$

For $s > 3$, (Q1) becomes

$$sa'(s) = -2a(s) - a(s-1), \quad sb'(s) = -2b(s) + b(s-1). \quad (\text{Q5})$$

The polynomial

$$G(s) = s^2 - 2s + \frac{1}{2}$$

satisfies

$$(sG(s))' = 2G(s) + G(s+1), \quad (\text{Q6})$$

and is positive and increasing for $s \geq 2$. Multiplying (Q5) by the conjugate solutions G and 1, respectively, and integrating one unit gives

$$sa(s)G(s) = \int_{s-1}^s a(x)G(x+1) dx, \quad sb(s) = -\int_{s-1}^s b(x) dx \quad (\text{Q7})$$

for $s \geq 3$. To see that no integration constant is hidden here, one can compute the initial piece explicitly. On $2 \leq s \leq 3$,

$$q^-(s) = \frac{1}{2} - \frac{1}{2} \log(s-1) + \frac{1}{2(s-1)},$$

and direct substitution at $s = 3$ gives both identities in (Q7).

The initial functions are positive on $[2, 3]$, hence $|b| < a$ there. Choose $\eta < 1$ with $|b| \leq \eta a$ on that compact interval. If equality first occurred at some $u > 3$, then the second identity in (Q7) and the first one, using $G(x+1) \geq G(u)$ on $[u-1, u]$, would give

$$|ub(u)| \leq \eta \int_{u-1}^u a(x) dx \leq \eta ua(u),$$

contradicting $|b(u)| = \eta a(u)$. Thus $|b| \leq \eta a$ for all $s \geq 2$, and $q^\pm = \frac{1}{2}s^2(a \pm b)$ remain positive and comparable to s^2a .

We next record two quantitative consequences of (Q7). This is the rank-one specialization of Iwaniec's one-step comparison, and we spell out the half-step twice because it will later control the Stieltjes error. From (Q5) and positivity, $a'(s) < 0$, so a is decreasing; moreover G is increasing on $[2, \infty)$. Restricting the first identity in (Q7) to the first half of its integration interval gives

$$sa(s)G(s) \geq \int_{s-1}^{s-1/2} a(x)G(x+1) dx \geq \frac{1}{2}a(s-1/2)G(s).$$

Hence $a(s-1/2) \leq 2sa(s)$. Applying the same estimate with $s-1/2$ in place of s yields

$$a(s-1) \leq 2(s-1/2)a(s-1/2) \leq 4s^2a(s). \quad (\text{Q8})$$

The finitely many values before this argument enters the range $G > 0$ are absorbed by enlarging the constant.

For the opposite inequality, choose $0 < \varepsilon < 1/4$, put

$$r(s) = (es)^{\varepsilon s}, \quad \Theta(s) = s^2a(s)r(s), \quad \theta(s) = \frac{r'(s)}{r(s)} = \varepsilon(\log s + 2).$$

We claim, after decreasing ε if necessary, that Θ is decreasing for $s \geq 3$. Otherwise let u be the first sufficiently large point with $\Theta'(u) = 0$. Since

$$\frac{\Theta'}{\Theta}(s) = \theta(s) - \frac{a(s-1)}{sa(s)}, \quad (\text{Q9})$$

we have

$$\theta(u) \frac{ua(u)}{a(u-1)} = 1. \quad (\text{Q10})$$

Before u the function Θ is decreasing. Therefore (Q7) yields

$$ua(u)G(u) \leq a(u-1)(u-1)^2r(u-1)G(u+1) \int_{u-1}^u \frac{dx}{x^2r(x)}.$$

Because θ is increasing,

$$r(u-1) \int_{u-1}^u \frac{dx}{r(x)} \leq \frac{1}{\theta(u-1)} \left(1 - \frac{r(u-1)}{r(u)} \right).$$

Combining this with (Q10) gives

$$1 \leq \frac{\theta(u)}{\theta(u-1)} \frac{G(u+1)}{G(u)} \left(1 - \frac{r(u-1)}{r(u)} \right). \quad (\text{Q11})$$

But

$$\frac{\theta(u)}{\theta(u-1)} = 1 + O\left(\frac{1}{u \log u}\right), \quad \frac{G(u+1)}{G(u)} = 1 + O\left(\frac{1}{u}\right),$$

whereas

$$\frac{r(u-1)}{r(u)} = \exp\left(-\int_{u-1}^u \theta(t) dt\right) \gg_{\varepsilon} u^{-\varepsilon}.$$

Since $\varepsilon < 1$, the last saving dominates the two $O(1/u)$ losses, so the right side of (Q11) is < 1 for all sufficiently large u , a contradiction. On the remaining compact interval the claim follows by choosing ε still smaller, because for $\varepsilon = 0$ the second term in (Q9) is strictly negative.

Thus (Q9) gives

$$a(s-1) \geq \varepsilon s(\log s + 2)a(s) \tag{Q12}$$

for $s \geq 3$. Iterating (Q12) over the integers gives

$$a(s) \leq C \prod_{s_0 < m \leq s} \frac{1}{\varepsilon m \log(3m)} \leq C e^{-cs \log(3s)}.$$

Together with $|b| \leq \eta a$, this proves (Q3), while (Q8) gives (Q4). Finally (Q3) implies $q^{\pm}(s) \rightarrow 0$; integrating (Q1) from s to infinity then gives (Q2). $\square$

Lemma 3.6.4 (A global coarse delay supersolution). *There are positive functions H^+, H^- and constants $\delta, c, C > 0$ such that*

$$\begin{aligned} \int_{s-1}^{\infty} \frac{H^+(u)}{u} du &\leq (1-\delta)H^-(s) & (s \geq 2), \\ \int_{s-1}^{\infty} \frac{H^-(u)}{u} du &\leq (1-\delta)H^+(s) & (s \geq 3), \end{aligned} \tag{H1}$$

and

$$H^+(s) + H^-(s) \leq C e^{-cs \log(3s)} \quad (s \geq 4). \tag{H2}$$

For $s \geq 4$ the functions $H^{\pm}(s)/s$ are decreasing, the two signs are comparable, and

$$H^{\pm}(s-1) \leq C s^2 H^{\pm}(s), \quad H^{\pm}(s-1) \leq C s^2 H^{\mp}(s). \tag{H3}$$

Proof. From (Q2),

$$q^{\pm}(s) - \int_{s-1}^{\infty} \frac{q^{\mp}(u)}{u} du = \int_{s-1}^{\infty} \frac{q^{\mp}(u)}{u^2} du > 0. \tag{H4}$$

Let $Q_0(s) = e^{-c_0 s \log(3s)}$ with $0 < c_0 < \min(c/4, 1/4)$. For a constant $A > 0$ to be chosen below, define explicitly

$$H_A^+(s) = A q^+(s) + Q_0(s), \quad H_A^-(s) = A q^-(s) + Q_0(s). \tag{H0}$$

We shall choose one A and then write $H^{\pm} = H_A^{\pm}$.

We record the elementary estimate that will be used below. If $\phi(u) = c_0 u \log(3u)$, then $\phi'(u) = c_0(\log(3u) + 1)$ is increasing, so for large s

$$\begin{aligned} \int_{s-1}^{\infty} \frac{Q_0(u)}{u} du &\leq \frac{1}{s-1} \int_{s-1}^{\infty} e^{-\phi(u)} du \\ &\leq \frac{e^{-\phi(s-1)}}{(s-1)\phi'(s-1)}. \end{aligned}$$

Moreover

$$\phi(s) - \phi(s-1) = c_0 \log(3s) + O(c_0),$$

and therefore the last expression is

$$\ll \frac{s^{c_0-1}}{\log(3s)} Q_0(s) = o(Q_0(s)).$$

After increasing the fixed starting point, it is at most $Q_0(s)/4$.

The point requiring care is that the relative slack in (H4) need not be uniform as $s \rightarrow \infty$. This is why the slower auxiliary envelope Q_0 is present. Choose c_0 smaller than the exponent in (Q3), say also $c_0 < 10^{-2}$. Then $q^\pm(s) = o(Q_0(s))$. Later, in the moderate-endpoint estimate, we shall simply choose its (arbitrarily weakenable) exponent larger than this fixed c_0 .

We now choose A without a circular tail/compactness dependence. First fix S_0 so large that, for both signs and every $s \geq S_0$,

$$\int_{s-1}^{\infty} \frac{Q_0(u)}{u} du \leq \frac{1}{4} Q_0(s). \quad (\text{H5})$$

On their respective fixed compact domains, define

$$\Delta^-(s) := q^-(s) - \int_{s-1}^{\infty} \frac{q^+(u)}{u} du \quad (2 \leq s \leq S_0),$$

$$\Delta^+(s) := q^+(s) - \int_{s-1}^{\infty} \frac{q^-(u)}{u} du \quad (3 \leq s \leq S_0).$$

These functions are strictly positive by (H4). The corresponding Q_0 -defects,

$$Q_0(s) - \int_{s-1}^{\infty} \frac{Q_0(u)}{u} du,$$

are bounded below on the same compact intervals. Hence one may choose a single finite A so large that both inequalities in (H1) are strict throughout their compact starting ranges for the functions in (H0).

Freeze this A . Since $q^\pm = o(Q_0)$, choose $S_1 \geq S_0$ so that $Aq^\pm(s) \leq Q_0(s)$ for $s \geq S_1$. On $s \geq S_1$, (H4) and (H5) give, for example,

$$H^\pm(s) - \int_{s-1}^{\infty} \frac{H^\mp(u)}{u} du \geq Q_0(s) - \int_{s-1}^{\infty} \frac{Q_0(u)}{u} du \geq \frac{3}{4} Q_0(s) \geq \frac{3}{8} H^\pm(s).$$

On the remaining compact interval $[S_0, S_1]$, both the q -slack (H4) and the Q_0 -slack (H5) are nonnegative and the first is strictly positive; the ratio of the total slack to $H^\pm$ therefore has a positive minimum. Taking δ smaller than the three positive margins obtained on $[2, S_0]$, $[S_0, S_1]$, and $[S_1, \infty)$ proves (H1) globally. This also makes clear that A is frozen before δ and before the later transfer constant C_* .

The strict separation $|b| \leq \eta a$ makes q^+ and q^- comparable on the tail; compactness handles the remaining interval. The upper decay (H2) follows because both $Aq^\pm$ and Q_0 have superfactorial upper bounds (with the slower exponent chosen for the sum). The same-sign and cross-sign one-step estimates in (H3) follow from Lemma 3.6.3, (Q8), comparability of the two signs, and the elementary corresponding estimates for Q_0 . We retain separately the lower bound $H^\pm \geq Q_0$, which will be used in the tail comparison below. Finally, (Q1) shows directly that each $q^\pm$ is nonincreasing once it leaves its constant initial interval, and Q_0 is decreasing. Hence $H^\pm(s)/s$ is decreasing on the tail; the residual compact range is absorbed into the stated constants. $\square$

3.6.4 A source-faithful discrete comparison

The next proof uses the exact depth recursion above, as in Iwaniec. The simplification is that our strong dimension-one hypothesis gives the two-sided local Mertens approximation (K2), so a large-endpoint Stieltjes comparison and a moderate-endpoint factorial estimate suffice; no optimized linear-sieve error term is needed.

Lemma 3.6.5 (Moderate-endpoint factorial bound). *Fix $B > 0$. There are $c_B > 0$ and s_B such that, if $D = z^s$, $s \geq s_B$, and $\log z \leq s^B$, then*

$$\frac{1}{V(z)} \sum_{n \geq 1} V_n(D, z) \leq e^{-c_B s \log s}. \quad (\text{M1})$$

Proof. Put $G(z) = \sum_{p < z} g(p)$. Definition 3.2.1 gives $G(z) \ll 1 + \log \log(3z) \ll_B \log(3s)$, while (K2), from the fixed point z_0 , gives $V(z)^{-1} \ll 1 + \log z \ll s^B$. By Lemma 3.5.2, every terminal chain has length $n > s - 2$. Dropping all other constraints and using the elementary symmetric-sum inequality,

$$\frac{1}{V(z)} \sum_{n \geq 1} V_n(D, z) \leq V(z)^{-1} \sum_{n > s-2} \frac{G(z)^n}{n!}.$$

Stirling's inequality and the ratio test bound the last tail by $\exp\{-s \log s + O_B(s \log \log(3s))\}$, and the polynomial factor $V(z)^{-1}$ is absorbed. This proves (M1), with for example any fixed $c_B < 1$ once s_B is large enough. $\square$

The comparison has three endpoint ranges. Before the induction, keep the following ledger in view. For a parent parameter s and $Y = s \log z$, the child parameter is $u_p = Y / \log p - 1$. The proof treats

$$\boxed{1 < u_p < 3} \quad (\text{compact endpoint}), \quad \boxed{s - 1 \leq u_p \leq U = Y^{1/4}} \quad (\text{Stieltjes range}),$$

and

$$\boxed{u_p > U} \quad (\text{superfactorial tail}).$$

The first range is absorbed by the exact endpoint recurrence (R0), the second is where (K2)–(K3) turn the prime sum into the delay integral, and the third is summable only because the positive envelopes have the decay (H2). This separation is what prevents the induction constant from accumulating with the Rosser depth R .

Lemma 3.6.6 (Uniform discrete-to-delay transfer). *There is a constant C_* such that, uniformly in R, z ,*

$$\begin{aligned} \mathcal{T}_{R,z}^-(s) &\leq C_* \frac{H^-(s)}{s} & (s \geq 2), \\ \mathcal{T}_{R,z}^+(s) &\leq C_* \frac{H^+(s)}{s} & (s \geq 3). \end{aligned} \quad (\text{D1})$$

For $1 < s < 3$ the upper tail is bounded by a constant depending only on the strong dimension-one data. In particular, for some $c, C > 0$,

$$\mathcal{T}_{R,z}^+(s) + \mathcal{T}_{R,z}^-(s) \leq C e^{-cs \log(3s)} \quad (\text{D2})$$

uniformly in R whenever $s \geq 4$.

Proof. We induct on R , proving the minus estimate first and then the plus estimate at the same depth. There are two endpoint regimes.

If $\log z \leq s^8$ and s is large, Lemma 3.6.5 is stronger than (D1), because the functions $H^\pm$ contain the slower envelope $e^{-c_0 s \log(3s)}$. If s is bounded as well, then z ranges in a fixed bounded interval; only finitely many primes occur, every finite-depth tail stabilizes after their number of levels, and (D1) follows after enlarging C_* . We may therefore assume

$$L := \log z > s^8, \quad Y := \log(z^s) = sL. \quad (\text{D3})$$

Consider first (R-), and write

$$u_p = \frac{Y}{\log p} - 1, \quad w(t) = e^{Y/(t+1)}.$$

The induction hypothesis gives $\mathcal{T}_{R-1,p}^+(u_p) \leq C_* H^+(u_p)/u_p$ whenever $u_p \geq 3$. We also need a careful endpoint form on the compact strip $1 < u_p < 3$. Apply (R0) with child endpoint p and put $p_u = p^{u_p/3}$. By (K2), uniformly for large p ,

$$\frac{V(p_u)}{V(p)} = \frac{3}{u_p} + O\left(\frac{1}{\log p}\right), \quad \mathcal{T}_{0,p}^+(u_p) = \frac{3}{u_p} - 1 + O\left(\frac{1}{\log p}\right).$$

Using the already proved depth- $(R-1)$ estimate at parameter 3 therefore gives

$$\mathcal{T}_{R-1,p}^+(u_p) \leq \left(\frac{3}{u_p} - 1\right) + \frac{C_* H^+(3)}{u_p} + O\left(\frac{1 + C_*}{\log p}\right). \quad (\text{D0})$$

On $(1, 3)$ the function q^+ is constant while Q_0 is strictly decreasing, so $H^+(u) - H^+(3)$ is positive and comparable to $3 - u$ on every compact subinterval ending at 3. Consequently, after enlarging C_* once, the first two terms on the right of (D0) are at most $C_* H^+(u_p)/u_p$. The finitely many small endpoints are absorbed into the same constant.

The remaining $O(C_*/\log p)$ term is kept separate. The strip $u_p < 3$ implies $s < 4$ and $\log p > Y/4$. Hence its total contribution to the parent sum is

$$\ll \frac{C_*}{Y} \sum_{p > e^{Y/4}} g(p) \frac{V(p)}{V(z)} \ll \frac{C_*}{Y},$$

by (K3) and (K2), because the parent parameter is then confined to the fixed compact interval $2 \leq s < 4$. For large Y this is absorbed by the positive compact lower bound for $C_* H^-(s)/s$; bounded Y leaves only finitely many prime endpoints and is absorbed into C_* . Thus, apart from an error already absorbed at the parent level, it is enough to estimate

$$C_* \sum_{p < z} g(p) \frac{V(p)}{V(z)} \frac{H^+(u_p)}{u_p}. \quad (\text{D4})$$

The lower cutoff in (R-) was dropped, which only enlarges this positive sum.

Let $a = s - 1$ and $U = Y^{1/4}$. The atomic measure in (D4) has distribution function

$$F(t) = \sum_{w(t) \leq p < z} g(p) \frac{V(p)}{V(z)} = \frac{V(w(t))}{V(z)} - 1 \quad (\text{D5})$$

as long as $w(t) \geq z_0$, by (K3). Formula (K2) gives, uniformly for $a \leq t \leq U$,

$$F(t) = \frac{t-a}{s} + E(t), \quad |E(t)| \leq C \frac{(t+1)^2}{sY}. \quad (\text{D6})$$

Indeed $\log w(t) = Y/(t+1)$ and $V(w(t))/V(z) = ((t+1)/s) \exp\{O((t+1)/Y)\}$.

Put $f(t) = H^+(t)/t$. For large t this is decreasing; on the fixed compact transition interval its total variation is bounded once and for all. Write $F(t) = (t-a)/s + E(t)$. Stieltjes integration by parts gives

$$\int_a^U f dE = f(U)E(U) - f(a)E(a) - \int_a^U E(t) df(t).$$

Here $E(a) = 0$. If a is beyond the compact transition range, monotonicity of f makes its total variation at most $f(a)$; if a lies in the transition range, compactness gives the same estimate after multiplication by one fixed constant because $f(a)$ is bounded below there. Combining this with (D6) yields

$$\int_a^U f dF \leq \frac{1}{s} \int_a^U \frac{H^+(t)}{t} dt + C \frac{U^2}{sY} \frac{H^+(a)}{a}. \quad (\text{D7})$$

By (H1) the first term is at most $(1-\delta)H^-(s)/s$. The ratio bound (H3) and $U = Y^{1/4}$ show that the second term is

$$\ll \frac{sU^2}{Y} \frac{H^-(s)}{s} = \frac{s}{Y^{1/2}} \frac{H^-(s)}{s} \leq \frac{\delta}{4} \frac{H^-(s)}{s} \quad (\text{D8})$$

after enlarging the fixed bounded regime, because (D3) implies $Y > s^9$.

For $t > U$ while $w(t) \geq z_0$, the coarser bound (3.3) gives $F(t) \ll (t+1)/s$. On a dyadic block $[T, 2T]$ the mass of the corresponding atoms is at most $F(2T) \ll T/s$, whereas monotonicity gives $f(t) \leq f(T) = H^+(T)/T$. Hence that block contributes $\ll H^+(T)/s$. By (H2),

$$\sum_{j \geq 0} H^+(2^j U) \leq C \sum_{j \geq 0} \exp\{-c 2^j U \log(3 \cdot 2^j U)\} \ll \exp\{-c' U \log(3U)\}.$$

Consequently

$$\frac{1}{s} \sum_{j \geq 0} H^+(2^j U) = o\left(\frac{H^-(s)}{s}\right), \quad (\text{D9})$$

because $H^-(s) \geq Q_0(s)$ and $U = Y^{1/4} \gg s^{9/4}$ make the displayed superfactorial bound $o(Q_0(s))$. Once $w(t) < z_0$, only a fixed finite set of small primes remains. Their normalized coefficients are $O(V(z)^{-1}) = O(1+L)$ by (K2) from the fixed point z_0 . For each such prime one has $t \asymp Y$, and therefore even this polynomial-in- L coefficient times $H^+(t)/t$ is exponentially smaller than $Q_0(s)/s$ by (H2). Combining (D7)–(D9) proves the minus estimate with a strict margin. More precisely, after the bounded regimes have been absorbed once, the three ranges contribute at most

$$C_* \left(1 - \frac{\delta}{2}\right) \frac{H^-(s)}{s}$$

for a fixed $\delta > 0$. This is the quantitative reason the same C_* may be reused at depth R : the induction consumes margin, not an additional multiplicative constant.

For the plus estimate with $s \geq 3$, use (R+) and the minus estimate just proved at the same depth. The identical Stieltjes argument, with H^- and H^+ interchanged, has the same fixed margin and gives (D1). This closes the induction on R without replacing C_* by a depth-dependent sequence. For $1 < s < 3$, (R0), (K2), and the estimate at $s = 3$ give a uniform bound; the finitely many small endpoints are again absorbed into the constant. Finally (D2) follows from (D1) and (H2). $\square$

Theorem 3.6.7 (Terminal-branch factorial estimate). *There are constants $c, C > 0$ such that, for $D = z^s$ and $s \geq 4$,*

$$\sum_{n \geq 1} V_n(D, z) \leq CV(z)e^{-cs \log(3s)}. \quad (\text{T})$$

Proof. For fixed D, z only finitely many strictly decreasing prime chains occur. Hence, for R larger than the number of primes below z , the two finite tails in Lemma 3.6.6 contain every odd and even terminal depth. Letting R reach this stabilizing value and applying (D2) gives (T). $\square$

Remark 3.6.8 (Relation with Iwaniec's original proof). The exact $\beta = 2$ first-failure rule and recurrences Lemma 3.6.2 are the dimension-one specialization of Iwaniec's Rosser recursion. His paper proves a sharper result by comparing the discrete tails with optimized continuous sieve functions and by inserting a logarithmic slack factor in the delicate intermediate endpoint range. Our application needs only one sufficiently large fixed sieve ratio. The stronger two-sided local Mertens estimate (K2), already proved for the present affine sieve, lets us replace that optimization by the large-endpoint Stieltjes comparison (D5)–(D9) plus the moderate-endpoint factorial estimate. The positive delay functions $q^\pm$ and their conjugate-polynomial argument are retained because they are the mechanism that makes the depth induction uniform.

3.7 The fixed-ratio fundamental lemma

Theorem 3.7.1 (Internal fixed-ratio fundamental lemma). *Assume Definition 3.2.1. There exists an absolute $s_0 = s_0(C_0) \geq 4$ such that if $D = z^s$ with $s \geq s_0$, then*

$$\frac{3}{4}V(z) \leq V^-(D, z) \leq V(z) \leq V^+(D, z) \leq \frac{5}{4}V(z).$$

Proof. Choose s_0 so that the right side of Theorem 3.6.7 is at most $V(z)/4$. Proposition 3.5.1 then gives the result, since all V_n are nonnegative. $\square$

3.8 Lower and upper sieve theorems

Theorem 3.8.1 (Lower-bound sieve at level D). *Assume Definition 3.2.1, $D = z^s$, $s \geq s_0$, and the sieve data (3.1). Then*

$$S(\mathcal{A}, z) \geq \frac{3}{4}XV(z) - \sum_{\substack{d < D \\ d|P(z)}} |r_d|.$$

Proof. Insert $A_d = Xg(d) + r_d$ into the lower inequality in (3.4). Since $|\lambda_d^-| \leq 1$ and the support is $d < D$,

$$S(\mathcal{A}, z) \geq XV^-(D, z) - \sum_{d < D} |r_d|.$$

Apply Theorem 3.7.1. $\square$

Theorem 3.8.2 (Upper-bound companion). *Under the same hypotheses,*

$$S(\mathcal{A}, z) \leq \frac{5}{4}XV(z) + \sum_{\substack{d < D \\ d|P(z)}} |r_d|.$$

Proof. Use the upper inequality in (3.4), $|\lambda_d^+| \leq 1$, and Theorem 3.7.1. $\square$

3.9 A diagnostic example

Suppose $g(p) = p/(p^2 - 1)$ for all sufficiently large primes. Then $g(p) = p^{-1} + O(p^{-3})$. Once the elementary Mertens estimate

$$\sum_{p \leq x} \frac{1}{p} = \log \log x + C + O(1/\log x)$$

is available, Definition 3.2.1 follows, as does $V(z) \asymp 1/\log z$. Thus Theorem 3.8.1 gives a positive sifted mass whenever the total remainder below D is $o(X/\log z)$. Chapters 1–2 have already proved the corresponding remainder and Euler-product estimates for the affine random-walk sequence used in this volume.

Bibliographic note

The switching architecture is historically Rosser’s and the quantitative terminal-tree organization is associated with Iwaniec’s refinement of Rosser’s sieve. The proof above is written internally and intentionally sacrifices the sharp linear-sieve constants; citations serve only attribution.

3.10 Worked first-failure example

Suppose $D = z^{10}$ and use the dimension-one Rosser parameter $\beta = 2$. A terminal branch of length $n \leq 8$ is impossible, because $p_1 \cdots p_n p_n^2 < z^{n+2} \leq z^{10} = D$. Thus every correction to the Euler product begins only at depth at least nine. The continuous-delay majorant and its discrete transfer quantify how little total mass can survive at these deep first-failure levels.

Chapter 4

The Affine Sieve and Almost-Prime Orbits

Final mechanism. Set $D = e^{\eta L}$ and $z = D^{1/s_0}$. Expansion makes the total remainder below D exponentially small; the Rosser sieve leaves mass $\asymp 1/\log z \asymp 1/L$; exponential word-height bounds convert the absence of prime factors below z into a fixed bound on $\Omega(F)$. A separate congruence point-count argument shows that this almost-prime set cannot lie in any proper algebraic subvariety.

4.1 The abstract random-walk affine sieve

We first isolate the exact theorem used in the concrete application.

Theorem 4.1.1 (Abstract fixed-ratio affine sieve). *Let ν_L be probability measures on a countable set $\mathcal{O}$, and let $F : \mathcal{O} \rightarrow \mathbb{Z} \setminus \{0\}$. Suppose there are a finite set of bad primes $\mathcal{B}$, constants $\eta, h, c_0 > 0$, a multiplicative function β on squarefree integers coprime to $\mathcal{B}$, and a fixed number $s_0 \geq 4$ with the following properties.*

(i) *For squarefree d coprime to $\mathcal{B}$,*

$$\nu_L(d \mid F) = \beta(d) + r_d(L), \quad \sum_{d < e^{\eta L}} |r_d(L)| \leq e^{-c_0 L}.$$

(ii) *The Euler product $V(z) = \prod_{p < z, p \notin \mathcal{B}} (1 - \beta(p))$ satisfies the dimension-one hypotheses of Chapter 3, and the internal Rosser fundamental lemma there is valid at ratio s_0 .*

(iii) *Every prime in $\mathcal{B}$ is excluded from every value $F(x)$ in the support of the measures.*

(iv) *$|F(x)| \leq Ce^{hL}$ on $\text{supp } \nu_L$.*

Then there are constants $c > 0$ and an integer

$$R \leq \left\lceil \frac{hs_0}{\eta} \right\rceil + 1$$

such that, for all large L ,

$$\nu_L(\Omega(|F|) \leq R) \geq \frac{c}{L}.$$

Proof. There are three scales to keep separate: the walk length L , the distribution level D , and the sieving threshold z . We choose them in that order:

$$D = e^{\eta L}, \quad z = D^{1/s_0} = e^{\eta L/s_0}.$$

Thus the Rosser ratio is exactly $\log D / \log z = s_0$, the fixed ratio at which hypothesis (ii) supplies the internal lower-sieve theorem.

Step 1: surviving sifted mass. Apply Theorem 3.8.1 to the weighted sequence defined by ν_L . Its total mass is one, so hypothesis (i) gives

$$\nu_L((F, P(z)) = 1) \geq c_1 V(z) - e^{-c_0 L}.$$

The dimension-one hypothesis gives

$$V(z) \asymp \frac{1}{\log z} = \frac{s_0}{\eta L}.$$

Because an exponential is eventually smaller than every reciprocal polynomial, there is L_0 such that $e^{-c_0 L} \leq \frac{1}{2} c_1 V(z)$ for $L \geq L_0$. Hence

$$\nu_L((F, P(z)) = 1) \gg \frac{1}{L}.$$

This is the positive mass that will be converted into almost-prime values.

Step 2: convert roughness into a bound on prime factors. Fix an x counted by the preceding event. Any prime divisor of $F(x)$ below z would either belong to the finite bad set or divide $P(z)$. Hypothesis (iii) excludes the first alternative, while $(F(x), P(z)) = 1$ excludes the second. Thus every prime divisor of $F(x)$ is at least z . Counting prime factors with multiplicity gives

$$z^{\Omega(|F(x)|)} \leq |F(x)|.$$

Hypothesis (iv) supplies the opposite estimate, so

$$\Omega(|F(x)|) \log z \leq \log C + hL.$$

Using $\log z = \eta L / s_0$ yields the explicit ledger

$$\Omega(|F(x)|) \leq \frac{hs_0}{\eta} + \frac{s_0 \log C}{\eta L}.$$

For all sufficiently large L , the second term is < 1 . Therefore every point in the sifted event satisfies

$$\Omega(|F(x)|) \leq \left\lceil \frac{hs_0}{\eta} \right\rceil + 1,$$

and Step 1 gives the asserted probability lower bound. $\square$

Remark 4.1.2 (Historical ownership of the two affine-sieve engines). The Bourgain–Gamburd–Sarnak affine-sieve framework isolates precisely the same two quantitative inputs used here: a level of distribution for congruence classes and a sieve-dimension estimate, and then invokes an elementary combinatorial sieve [1]. In their general notation the sieve dimension is the number t of relevant irreducible factors. Our concrete polynomial $F(g) = 1 + Q_0 g_{11}$ has one such factor, so the present application is dimension one. There is nevertheless a deliberate ownership difference: Chapter 3 proves the fixed-ratio dimension-one Rosser mechanism internally, following the Rosser–Iwaniec architecture [2], rather than importing the elementary sieve theorem as a black box. Accordingly, the historical paper owns the affine expansion-to-sieve interface; the active proof of the dimension-one combinatorial engine is in this volume.

4.2 The concrete rank-one polynomial

Return to the group and polynomial

$$F(g) = 1 + Q_0 g_{11}$$

of Chapters 1–2.

Lemma 4.2.1 (Exponential height bound). *There are $C, h > 0$ such that $|F(g)| \leq Ce^{hL}$ for every group element reachable by a word of length L in the fixed support of μ .*

Proof. Let M bound the operator norm of every generator and its inverse. A word of length L has operator norm at most M^L , so every entry has absolute value at most M^L . Thus $|F(g)| \leq 1 + Q_0 M^L \leq Ce^{L \log M}$. $\square$

Proposition 4.2.2 (Explicit admissibility and bad-prime removal). *The polynomial F has no congruence obstruction on the orbit. More precisely:*

- (a) *if $p \mid Q_0$, then $F(g) \equiv 1 \pmod{p}$ for every g ;*
- (b) *if $p \nmid Q_0$ is allowed, then the reduction of the orbit is all of $\mathrm{SL}_2(\mathbb{F}_p)$ and the zero locus of F has density $p/(p^2 - 1) < 1$.*

Proof. Part (a) is immediate from the definition. For (b), strong approximation is an earlier result, and the local density was counted explicitly in Proposition 1.3.1. In particular there are residues on the orbit for which F is nonzero modulo p . $\square$

Corollary 4.2.3 (Concrete saturation bound). *There is a fixed integer $R = R(\Gamma, \mu, Q_0)$ such that*

$$\Pr(\Omega(|1 + Q_0(X_L)_{11}|) \leq R) \gg \frac{1}{L}.$$

One may take

$$R = \left\lceil \frac{hs_0}{\eta} \right\rceil + 1$$

for any η admitted by the level-of-distribution theorem of Chapters 1–2.

Proof. Theorems 2.2.2 and 2.1.4 supply hypotheses (i)–(ii) of Theorem 4.1.1, Proposition 4.2.2 supplies (iii), and Lemma 4.2.1 supplies (iv). $\square$

4.3 An elementary algebraic point-count lemma

Lemma 4.3.1 (Schwartz–Zippel over a finite field). *If $P \in \mathbb{F}_p[x_1, \dots, x_m]$ is nonzero of total degree at most D , then*

$$|\{x \in \mathbb{F}_p^m : P(x) = 0\}| \leq Dp^{m-1}.$$

Proof. Induct on m . The case $m = 1$ is the elementary fact that a nonzero polynomial of degree at most D has at most D roots. For $m > 1$ write

$$P(x', x_m) = \sum_{j=0}^r Q_j(x') x_m^j, \quad Q_r \neq 0,$$

where $x' = (x_1, \dots, x_{m-1})$ and $r \leq D$. The total degree bound gives $\deg Q_r \leq D - r$. By induction the exceptional set

$$E = \{x' \in \mathbb{F}_p^{m-1} : Q_r(x') = 0\}$$

has size at most $(D - r)p^{m-2}$. For $x' \notin E$, the one-variable polynomial $P(x', X)$ is nonzero of degree r , hence has at most r zeros. For $x' \in E$ we use the trivial bound of p possible values of x_m . Therefore

$$|Z(P)| \leq r(p^{m-1} - |E|) + p|E| \leq rp^{m-1} + (D - r)p^{m-1} = Dp^{m-1}.$$

□

Lemma 4.3.2 (Proper subvarieties occupy $O(1/p)$ of $\mathrm{SL}_2(\mathbb{F}_p)$). *Let $\mathcal{V} \subsetneq \mathrm{SL}_2$ be a fixed proper algebraic subvariety over $\mathbb{Q}$. There are $C_{\mathcal{V}}$ and a finite set of primes such that for all remaining p ,*

$$|\mathcal{V}(\mathbb{F}_p) \cap \mathrm{SL}_2(\mathbb{F}_p)| \leq C_{\mathcal{V}}p^2.$$

Proof. Choose a polynomial $P(a, b, c, d)$ whose class in $\mathbb{Q}[a, b, c, d]/(ad - bc - 1)$ is nonzero and which vanishes on $\mathcal{V}$. On the chart $a \neq 0$, substitute $d = (1 + bc)/a$ and multiply by a fixed power of a . This produces a nonzero polynomial $\tilde{P}(a, b, c)$ over $\mathbb{Q}$; it remains nonzero modulo all but finitely many primes. Lemma 4.3.1 gives $O(p^2)$ zeros on the chart. The complementary chart $a = 0$ itself contains only $p(p - 1) = O(p^2)$ matrices, since $bc = -1$ and d is free. Add the two bounds. □

4.4 A prime of controlled exponential size

Lemma 4.4.1 (Elementary prime-location lemma). *There is an absolute $A > 1$ such that, for every sufficiently large x , there is a prime p with*

$$x < p < x^A.$$

Proof. We first prove the coarse lower bound $\pi(y) \gg y/\log y$. Since

$$\sum_{k=0}^{2n} \binom{2n}{k} = 4^n,$$

the largest term satisfies

$$\binom{2n}{n} \geq \frac{4^n}{2n+1}.$$

On the other hand, for every prime $p \leq 2n$ the exponent of p in $\binom{2n}{n}$ is at most $\log(2n)/\log p$; hence

$$\log \binom{2n}{n} \leq \sum_{p \leq 2n} \frac{\log(2n)}{\log p} \log p = \pi(2n) \log(2n).$$

Thus $\pi(y) \geq cy/\log y$ for all sufficiently large y . Choose any fixed integer $A > 2$. If no prime lay in $(x, x^A]$, then $\pi(x^A) = \pi(x) \leq x$, whereas the lower bound just proved gives

$$\pi(x^A) \geq \frac{cx^A}{A \log x} > x$$

for sufficiently large x , a contradiction. □

4.5 Exponential escape from each proper algebraic subvariety

Theorem 4.5.1 (Algebraic-subvariety escape). *For every fixed proper $\mathcal{V} \subsetneq \mathrm{SL}_2$ there is $c_{\mathcal{V}} > 0$ such that*

$$\Pr(X_L \in \mathcal{V}(\mathbb{Z})) \ll_{\mathcal{V}} e^{-c_{\mathcal{V}}L}.$$

Proof. Let $c > 0$ be the spectral mixing exponent from Theorem 1.2.1. Choose $0 < \kappa < c/(2A)$, where A is from Lemma 4.4.1. For each large L , choose an allowed prime

$$e^{\kappa L} < p_L < e^{A\kappa L};$$

the finitely many excluded primes do not matter. If $X_L \in \mathcal{V}(\mathbb{Z})$, its reduction lies in $\mathcal{V}(\mathbb{F}_{p_L})$. By congruence mixing and Lemma 4.3.2,

$$\Pr(X_L \in \mathcal{V}) \leq \frac{C_{\mathcal{V}}}{p_L} + |\mathcal{V}(\mathbb{F}_{p_L})|^{1/2} e^{-cL} \ll e^{-\kappa L} + p_L e^{-cL}.$$

Because $p_L \leq e^{A\kappa L}$ and $A\kappa < c/2$, both terms decay exponentially. $\square$

Theorem 4.5.2 (Zariski density of almost-prime points). *The set*

$$\{g \in \Gamma : \Omega(|1 + Q_0 g_{11}|) \leq R\}$$

is Zariski dense in SL_2 .

Proof. Fix a proper algebraic subvariety $\mathcal{V}$. Corollary 4.2.3 says that the probability of the R -almost-prime event is $\gg 1/L$, whereas Theorem 4.5.1 says the probability of lying in $\mathcal{V}$ is exponentially small. Hence for all sufficiently large L there is an endpoint with R -almost-prime F -value outside $\mathcal{V}$. Since this is true for every proper $\mathcal{V}$, the full almost-prime set is not contained in any proper algebraic subvariety. $\square$

Theorem 4.5.3 (Concrete rank-one affine sieve theorem). *For the fixed non-elementary Γ , lazy symmetric walk μ , and $F(g) = 1 + Q_0 g_{11}$, there is a fixed integer R such that the R -almost-prime points of the orbit are Zariski dense. Quantitatively, at random-walk time L the probability of an R -almost-prime value is $\gg 1/L$.*

Proof. The quantitative assertion is Corollary 4.2.3; Zariski density is Theorem 4.5.2. $\square$

Proposition 4.5.4 (Stability under finitely many bad primes). *Let $\mathcal{E}$ be any additional finite set of primes. Put*

$$Q_{\mathcal{E}} = Q_0 \prod_{p \in \mathcal{E}} p, \quad F_{\mathcal{E}}(g) = 1 + Q_{\mathcal{E}} g_{11}.$$

Then the conclusions of Corollary 4.2.3 and Theorem 4.5.2 hold for $F_{\mathcal{E}}$, with constants allowed to depend on $\mathcal{E}$.

Proof. For every prime p in the enlarged bad set one has $F_{\mathcal{E}}(g) \equiv 1 \pmod{p}$ identically, so those primes never divide a sieved value. At every remaining allowed prime, multiplication of Q_0 by the fixed unit $Q_{\mathcal{E}}/Q_0$ merely changes the nonzero residue $-Q_0^{-1}$ defining the linear condition on g_{11} ; the count in Proposition 1.3.1 is therefore unchanged. Consequently the one-prime local density, the dimension-one Euler product, and the level of distribution have the same form after deleting finitely many primes. The height estimate of Lemma 4.2.1 changes only its fixed constant. Applying the abstract sieve theorem and the algebraic-subvariety escape argument again proves the stated conclusions. $\square$

4.6 Worked saturation calculation

If $|F(g)| \leq e^{hL}$, $D = e^{\eta L}$, and $z = D^{1/s_0}$, then every sifted value satisfies $z^{\Omega(F(g))} \leq e^{hL}$. Hence $\Omega(F(g)) \leq hs_0/\eta$. All constants are fixed before L tends to infinity, so the almost-prime order is genuinely uniform.

Historical references

- [1] J. Bourgain, A. Gamburd, and P. Sarnak, *Affine linear sieve, expanders, and sum-product*, Invent. Math. 179 (2010), 559–644.
- [2] H. Iwaniec, *Rosser's sieve*, Acta Arith. 36 (1980), 171–202.

The Bourgain–Gamburd Method
Volume XIV: Compact Product Growth and Spectral Gap in $SU(2)$

18 August 2026

Guide to the volume

The compact $SU(2)$ argument is organized as a single chain. Rank-one geometry and Diophantine nonconcentration feed a metric inverse theorem; a scalar discretized sum–product theorem and metric Larsen–Pink produce a Lie segment; transverse BCH filling yields compact product growth; this product theorem drives L^2 flattening; and the representation theory of $SU(2)$ converts flattening into a spectral gap. The final chapter removes the auxiliary free-pair model and returns to arbitrary finite algebraic dense generating sets.

Historical scope distinction. Two compact results should not be conflated. The 2008 $SU(2)$ theorem gives spectral gap for free generators satisfying the noncommutative Diophantine condition (in particular algebraic free generators), whereas the later $SU(d)$ theorem treats arbitrary finite algebraic generating sets whose subgroup is dense. This volume proves the free-pair mechanism first and then proves the reduction back to an arbitrary finite algebraic dense generating set, rather than importing that reduction from the earlier free case.

Only the publication chapters listed in the table of contents belong to the reading order of this volume. **Series position.** This volume uses the inverse/flattening philosophy of Volumes II and VI, now in compact metric form. Its main outputs are compact product growth and the algebraic-generator spectral-gap theorem in $SU(2)$.

Quantifier and asymptotic conventions. Unless a statement explicitly says otherwise, every constant occurring in a quantitative theorem is chosen in the order in which it is quantified in that theorem. Thus a phrase such as “for all sufficiently small δ ” means: there is a displayed or locally quantified threshold $\delta_0 > 0$, depending only on the parameters listed before that phrase, such that the assertion holds for every $0 < \delta < \delta_0$. The phrases “sufficiently large N ” and “sufficiently large p ” have the analogous meaning.

If $X \ll_\theta Y$ or $X = O_\theta(Y)$, the implicit constant depends only on θ and on the fixed ambient data of the current volume; it is independent of the variables quantified after θ . The words *absolute*, *uniform*, and *bounded* never introduce an unlisted dependence. Whenever an inequality between exponents or scales is essential to the argument, that inequality is written explicitly in the proof; these conventions are not a substitute for such a calculation.

Contents

1	Rank-One Compact Geometry	1
1.1	The round model of $SU(2)$	1
1.2	Maximal tori, normalizers, and centralizers	2
1.3	BCH with a quantitative remainder	3
1.4	Quantitative local multiplication	4
2	Algebraic and Diophantine Nonconcentration	7
2.1	Algebraic height gives exponential separation	7
2.2	The free-pair random walk	8
2.3	Ball non-concentration at all logarithmic scales	9
2.4	Torus tubes force commuting word differences	9
2.5	Cyclic centralizers and tube escape	10
3	Discretized Metric Inverse Theory	11
3.1	Metric discretization in a compact three-dimensional group	11
3.2	Metric Ruzsa calculus	11
3.3	Discretized multiplicative energy	12
3.4	Robust weighted BSG at finite resolution	13
3.5	Failed flattening gives a metric approximate group	14
3.6	The lower size comes from anti-ball information, not BSG	15
4	Internal Discretized Real Sum–Product	17
4.1	Why a scalar theorem is needed	17
4.2	The scalar theorem	17
4.3	Covering-number calculus	18
4.4	A scale-regular subset	20
4.5	Small additive and multiplicative growth	21
4.6	The quotient-set dichotomy	24
4.7	Dense case	25
4.8	Gap case	26
4.9	Optimization and completion of the proof	26
4.10	Interface for the $SU(2)$ rich-torus stage	27
5	Metric Larsen–Pink and Rich Tori	29
5.1	Rank-one geometry	29
5.2	Local transverse-intersection counting	30
5.3	A specialized metric Larsen–Pink estimate	31
5.4	From Larsen–Pink to a rich maximal torus	32
5.5	From a rich torus to a regular one-dimensional slice	33

6	From a Rich Torus to a Lie Segment	35
6.1	Purpose and interfaces	35
6.2	An internal homogeneous ring-saturation theorem	36
6.3	From a torus tube to a regular parameter set	42
6.4	Root-plane scalarization and homogeneous adjoint words	45
6.5	Signed adjoint sums are actual group words	46
6.6	Scale ladder in $SU(2)$	48
6.7	Uniformity across the variable-exponent scale range	50
7	Transverse BCH Filling	53
7.1	Inputs and target	53
7.2	Symmetrizing and conjugating the segment	53
7.3	A quantitative inverse-function lemma	54
7.4	The transverse product map	55
7.5	Transfer from exact filling to a bounded power of A	55
8	The Compact Local Product Theorem	57
8.1	Covering preliminaries	57
8.2	The intermediate-dimensional local product theorem	58
8.3	Obstruction form	60
8.4	Transition to flattening	60
9	Obstruction Classification and Random-Walk Interfaces	61
9.1	Standing internal inputs	61
9.2	A captured set cannot fail scale regularity silently	62
9.3	The torus branch is an actual torus-coset concentration	62
9.4	Canonical obstruction alternative	63
10	Compact L^2 Flattening and Multiscale Iteration	65
10.1	Interfaces and normalization	65
10.2	A weak product theorem derived from the local-product chapter	66
10.3	Converting captured mass into the weak product hypotheses	68
10.4	One-step compact L^2 flattening	69
10.5	Propagation of the non-concentration inputs	70
10.5.1	Comparison of neighboring smoothing scales	71
10.6	Bounded-stage multiscale flattening	71
10.7	Proof architecture	72
11	Internal Density of Matrix Coefficients	73
11.1	The concrete model of $SU(2)$	73
11.2	An internal multivariable Bernstein theorem	74
11.3	Radial extension from the sphere	75
12	$SU(2)$ Representation Theory and the Spectral Endpoint	77
12.1	The irreducible representations of $SU(2)$	77
12.2	The character formula	79
12.3	Schur orthogonality and density of coefficients	79
12.4	Peter–Weyl and Plancherel in the exact form used later	80
12.5	The smoothing kernel in a high-frequency block	81
12.6	The Hilbert–Schmidt spectral endgame	81

12.7 Scope of the model spectral theorem	82
13 Arbitrary Algebraic Dense Generators	83
13.1 The target and the strategy	83
13.2 Zariski density and change of field	83
13.3 Some algebraic place is unbounded	84
13.4 Producing a proximal word	85
13.5 Move the fixed points and apply ping-pong	86
13.6 A free subgroup of $SU(2)$ is automatically dense here	86
13.7 The model walk satisfies the earlier compact inputs	88
13.8 Transfer the gap back to the original generators	88
13.9 Mechanism of the reduction	90

Chapter 1

Rank-One Compact Geometry

This chapter fixes the geometric normalization used throughout the compact argument. We identify $SU(2)$ with the round three-sphere, compute small-ball volume, classify maximal tori and their normalizers, record the quantitative BCH estimate, and prove the transverse local multiplication lemma that later converts three one-dimensional directions into genuine three-dimensional volume.

1.1 The round model of $SU(2)$

Every element of $SU(2)$ has a unique form

$$U(z, w) = \begin{pmatrix} z & w \\ -\bar{w} & \bar{z} \end{pmatrix}, \quad |z|^2 + |w|^2 = 1. \quad (1.1)$$

We identify $(z, w) \in \mathbb{C}^2 \simeq \mathbb{R}^4$ with a point of the unit three-sphere S^3 .

Theorem 1.1 (Unit-quaternion model and Haar normalization). *The map*

$$S^3 \longrightarrow SU(2), \quad (z, w) \longmapsto U(z, w),$$

is a diffeomorphism. Under the standard identification of S^3 with the unit quaternions, it is a group isomorphism. The round metric on S^3 is bi-invariant for this group law, and normalized Haar probability on $SU(2)$ is

$$dm_G = \frac{1}{2\pi^2} d\text{vol}_{S^3},$$

where $2\pi^2$ is the round volume of the unit three-sphere.

Proof. Let $U = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SU(2)$. The first column has norm one. Orthogonality of the columns and determinant one force the second column to be $(b, d) = (-\bar{c}, \bar{a})$; after writing $z = a$ and $w = b$ this is exactly (1.1). Conversely a matrix in (1.1) has orthonormal columns and determinant $|z|^2 + |w|^2 = 1$. Thus the displayed map is a smooth bijection with smooth inverse given by the first row, hence a diffeomorphism.

Write a quaternion as $q = z + w\mathbf{j}$, with the rule $\mathbf{j}z = \bar{z}\mathbf{j}$. Direct multiplication gives

$$(z + w\mathbf{j})(z' + w'\mathbf{j}) = (zz' - w\bar{w}') + (zw' + w\bar{z}')\mathbf{j},$$

and multiplication of the corresponding matrices gives the same two complex coordinates. Hence the map is a group isomorphism from the unit quaternions.

Left and right multiplication by a unit quaternion preserve the Euclidean norm on $\mathbb{H} \simeq \mathbb{R}^4$, hence restrict to round isometries of S^3 . Therefore round volume is both left- and right-invariant. By uniqueness of Haar measure on a compact group it is a scalar multiple of Haar measure. Since the round volume of S^3 is $2\pi^2$, division by this number gives Haar probability. $\square$

We henceforth use the geodesic distance d of this round bi-invariant metric.

Proposition 1.2 (Exact and asymptotic volume of small balls). *For $0 \leq r \leq \pi$,*

$$m_G(B(e, r)) = \frac{r - \sin r \cos r}{\pi}.$$

Consequently there exist absolute $c, C, r_0 > 0$ such that

$$cr^3 \leq m_G(B(x, r)) \leq Cr^3 \quad (x \in \mathrm{SU}(2), 0 < r \leq r_0).$$

More precisely,

$$m_G(B(e, r)) = \frac{2}{3\pi}r^3 + O(r^5).$$

Proof. Geodesic polar coordinates about a point of the unit S^3 have volume element $4\pi \sin^2 t dt$: the geodesic sphere of radius t is a round two-sphere of radius $\sin t$ and area $4\pi \sin^2 t$. Therefore

$$\mathrm{vol}(B(e, r)) = 4\pi \int_0^r \sin^2 t dt = 2\pi(r - \sin r \cos r).$$

Divide by $2\pi^2$. Bi-invariance makes ball volume independent of the center. Finally $\sin r \cos r = r - \frac{2}{3}r^3 + O(r^5)$, which gives the expansion and hence the uniform two-sided comparison for small r . $\square$

Example 1.3 (Why dimension three matters). At resolution $\delta \ll r$, a ball of radius r therefore requires $\asymp (r/\delta)^3$ balls of radius δ . This cubic exponent is the ambient benchmark against which the one-dimensional torus and two-dimensional conjugacy-class estimates in the product theorem are compared.

1.2 Maximal tori, normalizers, and centralizers

Put

$$T = \left\{ \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} : \theta \in \mathbb{R} \right\}, \quad w = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Theorem 1.4 (Rank-one tori and their normalizers). *Every maximal torus of $\mathrm{SU}(2)$ is conjugate to T . Moreover*

$$N_{\mathrm{SU}(2)}(T) = T \sqcup wT, \quad N_{\mathrm{SU}(2)}(T)/T \cong \mathbb{Z}/2\mathbb{Z}.$$

Every nontrivial connected proper subgroup of $\mathrm{SU}(2)$ has identity component a maximal torus.

Proof. Every $g \in \mathrm{SU}(2)$ is unitary, hence unitarily diagonalizable. If g is noncentral its two eigenvalues are distinct and have the form $e^{i\theta}, e^{-i\theta}$. A connected abelian subgroup consists of commuting normal matrices and is therefore simultaneously unitarily diagonalizable, so it is contained in a conjugate of T . Thus maximal connected abelian subgroups are precisely the conjugates of T .

If $n \in N_{\mathrm{SU}(2)}(T)$, then conjugation by n permutes the two common eigenspaces $\mathbb{C}e_1, \mathbb{C}e_2$ of T . If it preserves them individually, n is diagonal and lies in T . If it swaps them, $w^{-1}n$ preserves them individually, so $n \in wT$. Conversely $wT w^{-1} = T$ and T normalizes itself, proving the normalizer formula.

For the last assertion, let $H < \mathrm{SU}(2)$ be a connected proper positive-dimensional subgroup. Its Lie algebra is a proper nonzero Lie subalgebra of the three-dimensional Lie algebra $\mathfrak{su}(2)$. Under the standard identification $\mathfrak{su}(2) \simeq \mathbb{R}^3$ the bracket is a nonzero scalar multiple of the cross product. A two-dimensional plane is not closed under cross product, so there is no 2-dimensional Lie subalgebra. Hence $\dim H = 1$, and a connected compact one-dimensional group is a circle; by the first paragraph it is a maximal torus. $\square$

Proposition 1.5 (Centralizers of noncentral elements). *If $g \in \mathrm{SU}(2)$ is noncentral, then $C_{\mathrm{SU}(2)}(g)$ is the unique maximal torus containing g .*

Proof. The two eigenvalues of g are distinct. Any matrix commuting with g preserves each of its two eigenspaces. In an orthonormal eigenbasis such a commuting element of $\mathrm{SU}(2)$ is diagonal of the form $\mathrm{diag}(e^{it}, e^{-it})$. Thus the centralizer is a conjugate of T . Any torus containing g lies in the centralizer, so uniqueness follows. $\square$

Lemma 1.6 (Quantitative normalizer test). *Fix a unit vector $X \in \mathfrak{su}(2)$ and let $T_X = \exp(\mathbb{R}X)$. There are constants $c, C > 0$ such that for every $g \in \mathrm{SU}(2)$,*

$$c \, \mathrm{dist}(g, N(T_X)) \leq \min\{\|\mathrm{Ad}(g)X - X\|, \|\mathrm{Ad}(g)X + X\|\} \leq C \, \mathrm{dist}(g, N(T_X)).$$

Here the norm on $\mathfrak{su}(2)$ is the $\mathrm{Ad}(\mathrm{SU}(2))$ -invariant Euclidean norm.

Proof. Under $\mathrm{Ad}(\mathrm{SU}(2)) = \mathrm{SO}(3)$, the normalizer $N(T_X)$ is exactly the inverse image of the stabilizer of the unoriented axis $\{\pm X\}$. The function

$$F(g) = \min\{\|\mathrm{Ad}(g)X - X\|, \|\mathrm{Ad}(g)X + X\|\}$$

vanishes precisely on this compact embedded submanifold. In a tubular neighborhood, the differential of the orbit map in directions normal to the normalizer is an isomorphism onto the tangent space of the two-sphere at $\pm X$, so the inverse function theorem gives two-sided linear comparison with distance to $N(T_X)$. Outside a fixed tubular neighborhood, both F and the distance to $N(T_X)$ are bounded below and above by positive constants; compactness enlarges the constants to a global comparison. $\square$

1.3 BCH with a quantitative remainder

We use the operator norm inherited from $M_2(\mathbb{C})$; all norms on the finite-dimensional space $\mathfrak{su}(2)$ are equivalent.

Proposition 1.7 (Baker–Campbell–Hausdorff to second order). *There exist $r_0, C > 0$ such that whenever $X, Y \in \mathfrak{su}(2)$ and $\|X\| + \|Y\| \leq r_0$,*

$$\log(\exp X \exp Y) = X + Y + \frac{1}{2}[X, Y] + R(X, Y),$$

with

$$\|R(X, Y)\| \leq C(\|X\| + \|Y\|)^3.$$

The logarithm is the principal matrix logarithm in a fixed neighborhood of the identity.

Proof. For $\|X\| + \|Y\|$ small, the matrix exponential series gives

$$e^X e^Y = I + (X + Y) + \left(\frac{1}{2}X^2 + XY + \frac{1}{2}Y^2 \right) + E_3,$$

where $\|E_3\| \leq C_1(\|X\| + \|Y\|)^3$ by absolute convergence of the matrix power series. Put $Z = e^X e^Y - I$. For $\|Z\| < 1/2$,

$$\log(I + Z) = Z - \frac{1}{2}Z^2 + E'_3, \quad \|E'_3\| \leq C_2\|Z\|^3.$$

Since $Z = (X + Y) + O((\|X\| + \|Y\|)^2)$, substituting the quadratic part yields

$$\begin{aligned} \log(e^X e^Y) &= X + Y + \frac{1}{2}X^2 + XY + \frac{1}{2}Y^2 - \frac{1}{2}(X + Y)^2 + O((\|X\| + \|Y\|)^3) \\ &= X + Y + \frac{1}{2}(XY - YX) + O((\|X\| + \|Y\|)^3). \end{aligned}$$

This is the asserted formula. $\square$

Corollary 1.8 (Commutator expansion). *For $\|X\| + \|Y\| \leq r_0$,*

$$\log(e^X e^Y e^{-X} e^{-Y}) = [X, Y] + O((\|X\| + \|Y\|)^3).$$

Proof. Apply Proposition 1.7 three times and collect terms of degree at most two. All linear terms cancel and the quadratic term is $[X, Y]$; the accumulated remainder remains cubic. $\square$

1.4 Quantitative local multiplication

For $X_1, X_2, X_3 \in \mathfrak{su}(2)$ define

$$\Phi(t_1, t_2, t_3) = e^{t_1 X_1} e^{t_2 X_2} e^{t_3 X_3}.$$

Choose an oriented orthonormal basis of $\mathfrak{su}(2)$ and write $\det(X_1, X_2, X_3)$ for the determinant of their coordinate columns.

Lemma 1.9 (A quantitative inverse-function lemma). *Let $F : B_{\mathbb{R}^3}(0, r) \rightarrow \mathbb{R}^3$ be C^1 , $F(0) = 0$, and let $L = DF(0)$. Suppose L is invertible, $\|L^{-1}\| \leq M$, and*

$$\sup_{\|x\| \leq r} \|DF(x) - L\| \leq \frac{1}{2M}.$$

Then

$$B\left(0, \frac{r}{2M}\right) \subset F(B(0, r)).$$

Proof. Write $F(x) = Lx + R(x)$. The derivative bound gives $\|R(x) - R(y)\| \leq (2M)^{-1}\|x - y\|$. For fixed z with $\|z\| \leq r/(2M)$ define

$$\Psi_z(x) = L^{-1}(z - R(x)).$$

For $\|x\| \leq r$,

$$\|\Psi_z(x)\| \leq M\|z\| + M\|R(x) - R(0)\| \leq r/2 + r/2 = r,$$

and $\|\Psi_z(x) - \Psi_z(y)\| \leq \frac{1}{2}\|x - y\|$. Thus Ψ_z is a contraction of the closed ball into itself. Its fixed point satisfies $F(x) = z$. $\square$

Theorem 1.10 (Transverse local product map). *Assume $\|X_i\| \leq 1$ and*

$$|\det(X_1, X_2, X_3)| \geq \kappa > 0.$$

Then there are absolute $c_0, c_1 > 0$ such that for every $0 < r \leq c_0 \kappa$,

$$B_{\mathrm{SU}(2)}(e, c_1 \kappa r) \subset \{e^{t_1 X_1} e^{t_2 X_2} e^{t_3 X_3} : |t_i| \leq r\}.$$

In particular the differential of Φ at the origin is nonsingular.

Proof. In exponential coordinates put $F(t) = \log \Phi(t)$. By Proposition 1.7,

$$F(t) = t_1 X_1 + t_2 X_2 + t_3 X_3 + O(\|t\|^2),$$

with a uniform C^1 remainder on a fixed small cube. Thus $DF(0) = L$, where $Le_i = X_i$. Since $\|X_i\| \leq 1$, the two largest singular values of L are bounded by an absolute constant, while their product times the smallest singular value equals $|\det L| \geq \kappa$. Hence $\|L^{-1}\| \leq C/\kappa$. The C^1 Taylor estimate gives $\|DF(t) - L\| \leq C'\|t\|$. If $r \leq c_0 \kappa$ with c_0 small enough, this is at most $1/(2\|L^{-1}\|)$ on the cube, after replacing the cube by a comparable Euclidean ball. Lemma 1.9 therefore gives $B_{\mathrm{su}(2)}(0, c\kappa r) \subset F([-r, r]^3)$. On a fixed exponential neighborhood, $\exp$ and $\log$ are bi-Lipschitz, so this contains a group ball $B_{\mathrm{SU}(2)}(e, c_1 \kappa r)$. $\square$

Theorem 1.11 (Quantitative local volume growth away from degeneracy). *Under the hypotheses of Theorem 1.10, for $0 < r \leq c_0\kappa$ one has*

$$m_G(\{e^{t_1 X_1} e^{t_2 X_2} e^{t_3 X_3} : |t_i| \leq r\}) \geq c\kappa^3 r^3.$$

Equivalently, for $0 < \delta \leq c\kappa r$ the image requires at least

$$c \left(\frac{\kappa r}{\delta} \right)^3$$

balls of radius δ .

More generally, if X is a unit torus direction and group elements g, h are chosen so that

$$\text{dist}(g, N(T_X)) \geq \rho \quad \text{and} \quad \text{dist}(\text{Ad}(h)X, \text{span}(X, \text{Ad}(g)X)) \geq \rho,$$

then, after changing absolute constants,

$$|\det(X, \text{Ad}(g)X, \text{Ad}(h)X)| \geq c\rho^2,$$

so the preceding conclusions hold with $\kappa = c\rho^2$.

Proof. Theorem 1.10 places the ball $B(e, c_1\kappa r)$ inside the image. Proposition 1.2 gives Haar measure at least $c(\kappa r)^3$, and the same ball-volume comparison gives the covering-number lower bound.

For the final assertion, Lemma 1.6 implies that the distance of $\text{Ad}(g)X$ from the line $\mathbb{R}X$ is at least $c\rho$. Gram–Schmidt writes the absolute determinant as the product of the length of X , the distance of $\text{Ad}(g)X$ from $\mathbb{R}X$, and the distance of $\text{Ad}(h)X$ from the plane spanned by the first two vectors. These three factors are at least $1, c\rho, \rho$. $\square$

Remark 1.12 (Geometric interface). The theorem isolates exactly the geometric input used downstream. The rich-torus chapter produces a one-dimensional torus direction; the Lie-segment chapter turns it into a short segment; the transverse-filling chapter produces conjugates with a quantitative determinant; and the local-product chapter compares the resulting three-dimensional covering lower bound with small-tripling upper bounds. No sum-product, Larsen–Pink, or flattening theorem is used in this chapter.

Chapter 2

Algebraic and Diophantine Nonconcentration

The spectral-gap argument needs quantitative separation from two geometric obstructions: very small balls and thin tubes around maximal tori. For an algebraic free pair we obtain both estimates from height growth, freeness, and the cyclic structure of centralizers. The final chapter will explain how an arbitrary algebraic dense generating set supplies such a bounded-word free pair, so no freeness hypothesis is imposed on the final theorem.

2.1 Algebraic height gives exponential separation

Let $S \subset \mathrm{SU}(2)$ be a fixed finite symmetric set whose matrix entries are algebraic. Choose a number field K containing all entries of all elements of S . Enlarge K if necessary so that every field embedding needed below is visible inside one fixed normal closure. Let $d = [K : \mathbb{Q}]$.

Lemma 2.1 (Denominator clearing and conjugate growth). *There are constants $D \in \mathcal{N}$ and $M \geq 2$, depending only on S , such that:*

- (i) *every entry of Ds , $s \in S$, is an algebraic integer;*
- (ii) *for every embedding $\sigma : K \hookrightarrow \mathbb{C}$ and every $s \in S$, every entry of $\sigma(Ds)$ has absolute value at most M ;*
- (iii) *if $w = s_1 \cdots s_n$ is a word of length n , then every entry of $D^n w$ is an algebraic integer and every conjugate of every such entry has absolute value at most $(2M)^n$.*

Proof. There are finitely many algebraic entries, so one common positive integer D clears all of their fractional ideals. There are finitely many embeddings and finitely many matrices, so a common conjugate bound M exists. The entries of $(Ds_1) \cdots (Ds_n) = D^n w$ are sums of products of algebraic integers, hence algebraic integers. At each matrix multiplication an entry is a sum of two products; induction gives the crude uniform bound $(2M)^n$ for every conjugate entry. $\square$

Theorem 2.2 (Liouville-type separation for algebraic matrix expressions). *There is $C_S > 0$ such that the following holds. Let A_n be any nonzero matrix whose entries are K -linear expressions obtained as differences of two S -words of lengths at most n . Then*

$$\|A_n\| \geq e^{-C_S n}.$$

In particular, if $u, v \in \langle S \rangle$ have S -word length at most n and $u \neq v$, then

$$d(u, v) \geq ce^{-C_S n}, \tag{2.1}$$

for the fixed bi-invariant Riemannian metric, after changing C_S and $c > 0$.

Proof. Choose a nonzero entry α of A_n . After multiplying by D^n (and, if the two word lengths differ, by the harmless integral power needed to reach the common exponent), we obtain a nonzero algebraic integer $\beta = D^n \alpha$. Its field norm is a nonzero rational integer, hence

$$1 \leq |N_{K/\mathbb{Q}}(\beta)| = \prod_{\sigma: K \hookrightarrow \mathbb{C}} |\sigma(\beta)|.$$

Lemma 2.1 bounds every factor except the distinguished embedding by C_0^n for a fixed C_0 . Therefore

$$|\beta| \geq C_0^{-(d-1)n}, \quad |\alpha| \geq D^{-n} C_0^{-(d-1)n} = e^{-C_S n}.$$

This gives the matrix-norm assertion.

Apply it to $u - v$. On the compact matrix group $\mathrm{SU}(2)$, the Riemannian metric and the ambient Euclidean matrix norm are locally bi-Lipschitz, while pairs outside a fixed local neighborhood have a fixed positive separation. Enlarging C_S therefore gives (2.1) uniformly. $\square$

Corollary 2.3 (Commutator separation). *There is C'_S such that if u, v have S -word length at most n , then either $[u, v] = 1$ or*

$$d([u, v], 1) \geq e^{-C'_S n}.$$

Proof. The commutator has word length at most $4n$. Apply Theorem 2.2 to the distinct elements $[u, v]$ and 1. $\square$

2.2 The free-pair random walk

Now suppose $S_0 = \{a, a^{-1}, b, b^{-1}\} \subset \mathrm{SU}(2)$ freely generates a copy of the free group F_2 . Let

$$\nu = \frac{1}{4}(\delta_a + \delta_{a^{-1}} + \delta_b + \delta_{b^{-1}}), \quad q_n = \nu^{*n}.$$

We write $|g|$ for reduced word length in F_2 .

Lemma 2.4 (Elementary spectral-radius bound on the four-regular tree). *For every $g \in F_2$ and every $n \geq 0$,*

$$q_n(g) \leq \vartheta^n, \quad \vartheta = \frac{\sqrt{3}}{2} < 1. \tag{2.2}$$

Proof. Let A be the adjacency operator of the Cayley tree and put $\phi(x) = 3^{-|x|/2}$. If $x \neq e$, one neighbor is closer to the root and three are farther, so

$$(A\phi)(x) = \sqrt{3}\phi(x) + \frac{3}{\sqrt{3}}\phi(x) = 2\sqrt{3}\phi(x).$$

At the root, $(A\phi)(e) = 4/\sqrt{3} < 2\sqrt{3}$. The weighted Schur test for the symmetric nonnegative kernel of A therefore gives $\|A\|_{\ell^2 \rightarrow \ell^2} \leq 2\sqrt{3}$. The Markov operator is $P = A/4$, hence $\|P\| \leq \sqrt{3}/2$. Finally

$$q_n(g) = \langle P^n \delta_e, \delta_g \rangle \leq \|P\|^n \leq \vartheta^n.$$

$\square$

Lemma 2.5 (Concentration cannot increase after further convolution). *If μ, λ are probability measures on a group and E is Borel, then*

$$(\mu * \lambda)(E) \leq \sup_y \mu(Ey^{-1}).$$

For a bi-invariant metric, right translates of balls are balls of the same radius. A right translate of a left maximal-torus tube $xT^{(r)}$ is a left tube of the same radius around a conjugate maximal torus.

Proof. Fubini gives $(\mu * \lambda)(E) = \int \mu(Ey^{-1}) d\lambda(y)$. The geometric assertions follow from bi-invariance and $xTy^{-1} = xy^{-1}(yTy^{-1})$. $\square$

2.3 Ball non-concentration at all logarithmic scales

Theorem 2.6 (Logarithmic-scale ball non-concentration). *There are constants $c_0, \kappa_b, C > 0$, depending only on the fixed algebraic free pair, with the following property. For all sufficiently small δ , every $n \geq c_0 \log(1/\delta)$, every $x \in \text{SU}(2)$, and every $\delta \leq r \leq r_0$,*

$$q_n(B(x, r)) \leq Cr^{\kappa_b}. \quad (2.3)$$

The same estimate holds for the smoothed measure $q_n * P_\delta$, with a changed constant.

Proof. Let C_S be a separation constant large enough for Theorem 2.2 and Corollary 2.3. Choose a small constant $c_* > 0$ with $4C_S c_* < 1/4$. For a sufficiently small r , put

$$m = \lfloor c_* \log(1/r) \rfloor.$$

Then $4r < ce^{-C_S m}$ for all sufficiently small r . Thus by (2.1), a ball of radius $2r$ contains at most one group element representable by a word of length at most m . Lemma 2.4 gives

$$q_m(B(x, 2r)) \leq \vartheta^m \leq Cr^{\kappa_b}, \quad \kappa_b = \frac{1}{2}c_* \log(1/\vartheta) > 0,$$

after decreasing the exponent to absorb the floor. For r in the remaining fixed compact interval between this small threshold and r_0 , enlarge C .

Choose $c_0 > c_*$. If $n \geq c_0 \log(1/\delta)$ and $r \geq \delta$, then $n \geq m$; write $q_n = q_m * q_{n-m}$ and apply Lemma 2.5. This proves (2.3). Finally

$$(q_n * P_\delta)(B(x, r)) \leq q_n(B(x, r + \delta)) \leq q_n(B(x, 2r)),$$

which proves the smoothed version. $\square$

2.4 Torus tubes force commuting word differences

Lemma 2.7 (Tube differences are close to one torus). *There is an absolute C such that if $g, h \in xT^{(r)}$, where T is a maximal torus and r is sufficiently small, then*

$$g^{-1}h \in T^{(Cr)}.$$

If $u, v \in T^{(Cr)}$, then

$$d([u, v], 1) \leq Cr.$$

Proof. Choose $t_g, t_h \in T$ with $d(g, xt_g), d(h, xt_h) \leq r$. Bi-invariance and the Lipschitz continuity of multiplication on the compact group give $d(g^{-1}h, t_g^{-1}t_h) \leq Cr$. For the second assertion choose $t_u, t_v \in T$ within Cr of u, v . Since $[t_u, t_v] = 1$ and the commutator map is smooth on the compact group, it is globally Lipschitz on a fixed small product neighborhood, giving the result. $\square$

Proposition 2.8 (Torus-tube occupancy forces exact commutation). *There is $c_* > 0$ such that, for all sufficiently small r , if $m \leq c_* \log(1/r)$ and*

$$\mathcal{W} \subset \text{supp}(q_m) \cap xT^{(r)},$$

then all elements of

$$\{g_0^{-1}g : g \in \mathcal{W}\}$$

commute pairwise, for any fixed $g_0 \in \mathcal{W}$.

Proof. Each difference has word length at most $2m$. By Lemma 2.7, any two differences have commutator within Cr of the identity. Such a commutator has word length at most $8m$. Choose c_* smaller, if necessary, so that $Cr < e^{-8C_S m}$ for $m \leq c_* \log(1/r)$. Corollary 2.3 then says that the commutator cannot be nontrivial. Hence it is the identity. $\square$

2.5 Cyclic centralizers and tube escape

Lemma 2.9 (Centralizers in F_2 are cyclic). *If $1 \neq u \in F_2$, then $C_{F_2}(u)$ is infinite cyclic. Moreover every left coset of a cyclic subgroup contains at most $4n + 1$ elements of word length at most n .*

Proof. Let the Cayley tree of F_2 carry its usual free action. A nontrivial element u acts hyperbolically: after conjugating $u = cvc^{-1}$ with v cyclically reduced, its invariant axis is the bi-infinite geodesic obtained by concatenating copies of v . Any element commuting with u preserves this axis and its orientation (an orientation reversal would conjugate u to u^{-1} rather than commute with it). The restriction homomorphism from the centralizer to the group of translations of the simplicial axis is injective because the free action has trivial vertex stabilizers. A subgroup of the integer translation group is cyclic. Thus the centralizer is cyclic, generated by the primitive translation along the axis.

For the counting statement write a generator as $z = cvc^{-1}$ with v cyclically reduced. For $k \neq 0$,

$$|z^k| = 2|c| + |k||v|,$$

so $|z^k| \leq 2n$ allows at most $4n + 1$ integers k ; translating by a fixed group element can only replace the radius n by at most a harmless factor two when the coset meets the radius- n ball. The stated coarse bound follows, and is more than sufficient below. $\square$

Theorem 2.10 (Quantitative escape from maximal-torus tubes). *There are $c_0, \kappa_t, C > 0$ such that for all sufficiently small δ , all $n \geq c_0 \log(1/\delta)$, all maximal tori T , all $x \in \mathrm{SU}(2)$, and all $\delta \leq r \leq r_0$,*

$$q_n(xT^{(r)}) \leq Cr^{\kappa_t}. \quad (2.4)$$

The same estimate holds for $q_n * P_\delta$ after changing C .

Proof. For small r , choose $m = \lfloor c_* \log(1/r) \rfloor$ with c_* satisfying Proposition 2.8. Let $\mathcal{W} = \mathrm{supp}(q_m) \cap xT^{(2r)}$. If it is empty there is nothing to prove. Otherwise fix $g_0 \in \mathcal{W}$. Proposition 2.8 says that the differences $g_0^{-1}g$, $g \in \mathcal{W}$, commute pairwise. If one is nontrivial, all the others lie in its cyclic centralizer by Lemma 2.9; if none is nontrivial then $\mathcal{W}$ is a singleton. Hence $\mathcal{W}$ lies in one left coset of a cyclic subgroup and contains at most Cm elements of the word ball of radius m .

By Lemma 2.4,

$$q_m(xT^{(2r)}) \leq Cm\vartheta^m \leq Cr^{\kappa_t}$$

for some $\kappa_t > 0$, after decreasing the exponent to absorb the logarithmic factor. As in Theorem 2.6, fixed larger values of r are absorbed into the constant.

For $n \geq c_0 \log(1/\delta)$ with $c_0 > c_*$ and $r \geq \delta$, we have $n \geq m$. Write $q_n = q_m * q_{n-m}$ and apply Lemma 2.5; right translation changes a left T -tube into a left tube about a conjugate maximal torus, over which the estimate is uniform. Finally smoothing by P_δ enlarges the tube radius from r to at most $2r$. $\square$

Remark 2.11 (From the free-pair model to arbitrary generators). The estimates in this chapter are stated for an algebraic free pair because freeness makes the word-counting and centralizer arguments transparent. The final chapter constructs such a pair from bounded words in any finite algebraic dense symmetric generating set. The flattening and representation-theoretic chapters are then applied to its symmetric four-point walk, and the final Dirichlet-form comparison transfers the spectral gap back to the original walk.

Chapter 3

Discretized Metric Inverse Theory

We now translate failure of L^2 -flattening into finite-scale approximate-group structure. The argument develops metric Ruzsa calculus, discretized multiplicative energy, and a weighted Balog–Szemerédi–Gowers theorem at resolution δ . A separate anti-ball estimate supplies the lower effective-size bound; keeping these two roles distinct is essential in the later product/obstruction argument.

3.1 Metric discretization in a compact three-dimensional group

Let $G = \mathrm{SU}(2)$ with the round bi-invariant metric and normalized Haar probability from the compact-geometry chapter. Write

$$B_r = B(1, r), \quad P_r = \frac{1_{B_r}}{m_G(B_r)}.$$

For a set A , $A^{(r)}$ denotes its closed r -neighborhood and $\mathcal{N}(A, r)$ the least number of radius- r balls needed to cover it. The compact-geometry chapter gives

$$cr^3 \leq m_G(B_r) \leq Cr^3 \tag{3.1}$$

for r below a fixed r_0 .

Lemma 3.1 (Maximal nets and volume). *If $X \subset G$ is maximal r -separated, then the balls $B(x, r)$ cover G , the balls $B(x, r/2)$ are disjoint, and $|X| \asymp r^{-3}$. More generally, if $X_A \subset A$ is a maximal r -separated subset, then*

$$|X_A| \asymp \mathcal{N}(A, r)$$

with absolute comparison constants.

Proof. Maximality gives the covering statement and separation gives disjoint half-balls. The volume comparison (3.1), applied to the disjoint half-balls and to a covering by full balls, gives all asserted comparisons. $\square$

3.2 Metric Ruzsa calculus

Theorem 3.2 (Metric Ruzsa triangle inequality). *There is an absolute C such that for nonempty compact $A, B, C_0 \subset G$ and all sufficiently small r ,*

$$\mathcal{N}(AC_0^{-1}, Cr) \mathcal{N}(B, r) \leq C \mathcal{N}(AB^{-1}, r) \mathcal{N}(BC_0^{-1}, r). \tag{3.2}$$

Consequently all fixed-length covering-number Ruzsa inequalities obtained from the exact finite-set triangle inequality remain valid with a fixed enlargement of the resolution.

Proof. Choose a maximal $10Cr$ -separated set $X \subset AC_0^{-1}$ and a maximal r -separated set $Y \subset B$. For each $x \in X$ choose one representation $x = a_x c_x^{-1}$. Map

$$(x, y) \mapsto (a_x y^{-1}, y c_x^{-1}).$$

Round the first coordinate to a fixed radius- r net of AB^{-1} and the second to a radius- r net of BC_0^{-1} . If two pairs have the same two rounded values, their exact first factors are $O(r)$ -close and their exact second factors are $O(r)$ -close. Multiplying the two coordinates and using bi-invariance gives $d(x, x') = O(r)$. Taking the separation constant in the definition of X larger than this fixed error forces $x = x'$. With x fixed, closeness of the first factors forces y, y' to lie in one fixed $O(r)$ -ball. Since Y is r -separated, the number of such y' is bounded by an absolute packing constant. Thus the rounding map has bounded multiplicity and

$$|X| |Y| \leq CN(AB^{-1}, r) \mathcal{N}(BC_0^{-1}, r).$$

Lemma 3.1 converts $|X|$ and $|Y|$ to the two covering numbers on the left. $\square$

Proposition 3.3 (Metric Ruzsa covering lemma). *Let $A, B \subset G$ be compact and suppose*

$$\mathcal{N}(AB, r) \leq K \mathcal{N}(B, r).$$

Then there is $X \subset A$ with $|X| \leq CK$ such that

$$A \subset (XBB^{-1})^{(Cr)}. \quad (3.3)$$

Proof. Choose a maximal subset $X \subset A$ for which the sets $x(B'_r)$, $x \in X$, have disjoint small thickenings, where B'_r is the union of $r/10$ -balls about a maximal r -separated subset of B . Each such thickening has Haar measure $\gg \mathcal{N}(B, r)r^3$, while their union lies in $(AB)^{(Cr)}$, whose Haar measure is $\ll \mathcal{N}(AB, r)r^3$. Hence $|X| \leq CK$. Maximality says that for every $a \in A$ there are $x \in X$ and $b_1, b_2 \in B$ for which ab_1 and xb_2 are $O(r)$ -close. Bi-invariance then gives $a \in (xb_2b_1^{-1})^{(Cr)}$. $\square$

3.3 Discretized multiplicative energy

For a probability measure μ put

$$\text{Col}_r(\mu) = \iint 1_{\{d(x, y) \leq r\}} d\mu(x) d\mu(y),$$

and

$$\text{E}_r(\mu) = \int 1_{\{d(x_1 x_2, x_3 x_4) \leq r\}} d\mu(x_1) \cdots d\mu(x_4).$$

Proposition 3.4 (Energy- L^2 comparison). *There are absolute $c, C > 0$ such that for every probability measure μ and $0 < r < r_0$,*

$$cr^{-3} \text{Col}_{cr}(\mu) \leq \|\mu * P_r\|_2^2 \leq Cr^{-3} \text{Col}_{Cr}(\mu), \quad (3.4)$$

$$cr^{-3} \text{E}_{cr}(\mu) \leq \|\mu * \mu * P_r\|_2^2 \leq Cr^{-3} \text{E}_{Cr}(\mu). \quad (3.5)$$

Proof. For $a, b \in G$,

$$\langle L_a P_r, L_b P_r \rangle = \frac{m_G(B(a, r) \cap B(b, r))}{m_G(B_r)^2}.$$

This vanishes when $d(a, b) > 2r$, is at most Cr^{-3} always, and is at least cr^{-3} when $d(a, b) \leq cr$, by (3.1) and the inclusion of a fixed sub-ball in the intersection. Expanding the squared L^2 norm of $\mu * P_r = \int L_a P_r d\mu(a)$ gives (3.4). Apply the same argument to $\mu * \mu$ to obtain (3.5). $\square$

3.4 Robust weighted BSG at finite resolution

We record precisely how the earlier Volume-II theorem is transported. A finite weighted configuration consists of points $x_i \in G$ with weights $p_i \geq 0$, $\sum p_i = 1$. Put $s^2 = \sum p_i^2$.

Lemma 3.5 (Stability of bounded group words). *For $r \geq 0$ write $x \sim_r y$ when $d(x, y) \leq r$. Let $w = w(z_1, \dots, z_n)$ be a group word, and let $m_i(w)$ be the total number of occurrences of $z_i^{\pm 1}$ in w . If $x_i \sim_{r_i} y_i$ for $1 \leq i \leq n$, then*

$$d(w(x_1, \dots, x_n), w(y_1, \dots, y_n)) \leq \sum_{i=1}^n m_i(w) r_i. \quad (3.6)$$

Consequently, if a formal identity between two words of total length at most J is obtained from q relations, each valid up to distance r , then the resulting identity is valid up to distance at most $2Jqr$.

Proof. Bi-invariance gives, for all $x, x', y, y' \in G$,

$$d(xy, x'y') \leq d(x, x') + d(y, y'), \quad d(x^{-1}, y^{-1}) = d(x, y).$$

Induction on the word length proves (3.6). For the final assertion, replace the q approximate relations one at a time in the two formal words. At each replacement at most $2J$ occurrences are affected, so the triangle inequality gives the stated $2Jqr$ bound. $\square$

Theorem 3.6 (Robust weighted BSG). *There is an absolute C_0 such that the following holds. Let $\{(x_i, p_i)\}_{i=1}^N$ be a finite weighted configuration and suppose*

$$\sum_{d(x_i x_j, x_k x_l) \leq r} p_i p_j p_k p_l \geq K^{-2} s^2, \quad K \geq 2. \quad (3.7)$$

Put $L = K \log(2NK)$. Then there are a finite symmetric set A , $1 \in A$, a point $a \in G$, and a finite set X with

$$|A| \leq L^{C_0} s^{-2}, \quad (3.8)$$

$$\sum_{x_i \in (aA)^{(C_0 r)} p_i \geq L^{-C_0}, \quad (3.9)$$

$$A^2 \subset (XA)^{(C_0 r)}, \quad |X| \leq L^{C_0}. \quad (3.10)$$

The analogous right-coset capture also holds.

Proof. Let $\mathcal{E}_r$ be the set of quadruples (i, j, k, l) occurring in (3.7). The weighted proof in Volume II uses the relation $x_i x_j = x_k x_l$ only through a finite list of bounded-length group-word identities: the popularity/pruning stage is purely a weighted graph argument; the common-neighbor and path-counting stage produces quotient and product identities of bounded word length; and the final Ruzsa/covering stage uses only products, inverses, and cancellation. Let J_{BSG} be the maximum word length and let q_{BSG} be the maximum number of energy relations used in any one of those identities. These are absolute integers because the Volume-II proof is fixed once and for all.

Run the same weighted graph selection with $\mathcal{E}_r$ in place of the exact energy graph. No counting estimate changes: the hypothesis (3.7) is exactly the lower bound used in the popularity and dyadic-pigeonhole steps. Whenever the discrete proof concludes an exact word identity from at most q_{BSG} energy relations, Lemma 3.5 gives instead

$$d(w_1, w_2) \leq C_{\text{BSG}} r, \quad C_{\text{BSG}} := 2J_{\text{BSG}} q_{\text{BSG}}. \quad (3.11)$$

Thus the two large weighted subsets produced by the graph argument have all the same quotient/product covering estimates as in Volume II, with resolution $C_{\text{BSG}}r$ rather than exact equality.

The subsequent Ruzsa steps are now genuinely metric statements: Theorem 3.2 and Proposition 3.3 convert those bounded-resolution quotient/product estimates into a symmetric set $A \ni 1$ and a set X satisfying

$$A^2 \subset (XA)^{(Cr)}, \quad |X| \leq L^C.$$

The weight truncation and dyadic selection in Volume II give simultaneously

$$|A| \leq L^C s^{-2}, \quad \sum_{x_i \in (aA)^{(Cr)}} p_i \geq L^{-C},$$

and the right-handed version follows by applying the same argument to inverses. All constants occurring here depend only on the fixed integers $J_{\text{BSG}}, q_{\text{BSG}}$ and on the absolute constants in the metric Ruzsa lemmas. Enlarging one absolute exponent to C_0 yields exactly (3.8)–(3.10). $\square$

Remark 3.7 (Why this is not a new black box). The combinatorial BSG theorem is not being cited externally here: its complete proof is already owned by Volume II. The only new assertion is stability under replacing exact product equality by $O(r)$ -equality, and the preceding proof identifies the mechanism that makes every step stable. The metric Ruzsa tools needed for that stability were proved in Sections 1–2.

3.5 Failed flattening gives a metric approximate group

Lemma 3.8 (A finite r -cell discretization). *For each small r there is a Borel partition $G = \bigsqcup_{i=1}^N C_i$ and representatives $x_i \in C_i$ such that*

$$C_i \subset B(x_i, r), \quad N \asymp r^{-3},$$

and each ball of radius Cr meets only $O_C(1)$ cells. If $p_i = \mu(C_i)$, then

$$\sum_i p_i^2 \asymp_C \text{Col}_{Cr}(\mu) \tag{3.12}$$

with fixed changes of the radius, and approximate product energy of the representatives is comparable in the same sense to $\text{E}_{Cr}(\mu)$.

Proof. Take a maximal r -separated net and its Voronoi partition. A cell lies in the radius- r ball about its center. Bounded local doubling gives the uniform cell-incidence bound. Pairs in one cell are $2r$ -close, while every Cr -close pair lies in one of only $O_C(1)$ neighboring cell pairs; Cauchy–Schwarz over this bounded-degree adjacency graph gives (3.12). The product statement follows identically because replacing each factor by its cell representative moves a product by $O(r)$ in a bi-invariant metric. $\square$

Theorem 3.9 (Metric failed-flattening inverse theorem). *There is an absolute C_1 such that the following holds. Let q be a probability measure, put $f = q * P_\delta$, and suppose*

$$\|q * q * P_\delta\|_2 > \delta^\zeta \|f\|_2, \quad 0 < \zeta < 1. \tag{3.13}$$

For sufficiently small δ there exist a symmetric measurable set $H = H^{-1}$ containing 1, a finite set X , and $a \in G$ such that

$$H^2 \subset HX \cap XH, \quad |X| \leq \delta^{-C_1 \zeta}, \tag{3.14}$$

$$(q * P_\delta)(aH) \geq \delta^{C_1 \zeta}, \tag{3.15}$$

$$m_G(H) \leq \delta^{-C_1 \zeta} \|f\|_2^{-2}. \tag{3.16}$$

Moreover H is a union of $O(\delta)$ -balls and, for every fixed m ,

$$\mathcal{N}(H^m, C_m \delta) \leq \delta^{-C_m \zeta} \mathcal{N}(H, \delta). \quad (3.17)$$

The theorem intentionally asserts only an upper effective-support bound.

Proof. Proposition 3.4 and Lemma 3.8, with fixed changes from δ to $C\delta$, turn (3.13) into

$$\sum_{d(x_i x_j, x_k x_l) \leq C\delta} p_i p_j p_k p_l \geq c \delta^{2\zeta} \sum_i p_i^2.$$

Apply Theorem 3.6 with $K = C\delta^{-\zeta}$. Since $N \asymp \delta^{-3}$,

$$L = K \log(2NK) \leq \delta^{-2\zeta}$$

for all sufficiently small δ after enlarging the final absolute exponent. We obtain a finite metric approximate group A , a captured coset, and

$$|A| \leq \delta^{-C\zeta} s^{-2}, \quad s^2 = \sum_i p_i^2. \quad (3.18)$$

By Proposition 3.4, Lemma 3.8, and (3.4),

$$s^2 \asymp \delta^3 \|f\|_2^2$$

again at fixed comparable resolutions.

Let $V = B(1, R\delta)$ with R a sufficiently large fixed constant and put $H = AV = VA$; the equality holds because a ball around the identity is conjugation-invariant for a bi-invariant metric. The set is symmetric and contains 1. From $A^2 \subset (X_0 A)^{(C\delta)}$, cover the fixed-radius error ball multiplied by V^2 by $O(1)$ left translates of V . This enlarges X_0 by only an absolute factor and yields the exact inclusion $H^2 \subset XH$. Apply the same argument on the right for $H^2 \subset HX$.

The captured discrete cell mass in (3.9) implies, after enlarging R , that $q((aH)^{(C\delta)}) \geq \delta^{C\zeta}$. Convolution by P_δ and a further fixed enlargement of H turn this into (3.15). Since H is covered by $C|A|$ radius- $C\delta$ balls, (3.1), (3.18), and $s^2 \asymp \delta^3 \|f\|_2^2$ give

$$m_G(H) \leq C|A|\delta^3 \leq \delta^{-C_1\zeta} \|f\|_2^{-2},$$

which is (3.16). Finally the exact approximate-group inclusion (3.14) gives $H^m \subset X_m H$ with $|X_m| \leq \delta^{-C_m \zeta}$ for every fixed m , and therefore (3.17). $\square$

3.6 The lower size comes from anti-ball information, not BSG

Proposition 3.10 (Independent lower effective-size estimate). *Assume H , a satisfy (3.15) and that for some $\kappa > 0$*

$$(q * P_\delta)(B(x, r)) \leq C r^\kappa \quad (\delta \leq r \leq r_0). \quad (3.19)$$

Then for every such r ,

$$\mathcal{N}(H, r) \geq c \delta^{C_1 \zeta} r^{-\kappa}. \quad (3.20)$$

In particular

$$\mathcal{N}(H, \delta) \geq c \delta^{-\kappa + C_1 \zeta}.$$

Thus if $C_1 \zeta < \kappa/2$, the inverse set has a genuine positive-power lower effective size, obtained independently of Theorem 3.9.

Proof. Cover H by $M = \mathcal{N}(H, r)$ radius- r balls. Then aH is covered by M radius- r balls with translated centers, up to a fixed harmless enlargement. Equations (3.15) and (3.19) give

$$\delta^{C_1\zeta} \leq (q * P_\delta)(aH) \leq CMr^\kappa.$$

Rearrange. □

Remark 3.11 (Separation of upper and lower size information). The separation between Theorem 3.9 and Proposition 3.10 is deliberate. The weighted inverse theorem of Volume II already warns that a lower collision-scale bound is not unconditional: failed flattening may be detected on a small heavy dyadic level. The compact argument therefore combines captured mass with the independent Diophantine anti-ball estimate from the preceding chapter. This keeps the upper effective-support estimate logically separate from the lower-size information and avoids a circular use of the inverse theorem.

Chapter 4

Internal Discretized Real Sum–Product

The compact product theorem eventually reduces one-dimensional concentration inside a maximal torus to additive and multiplicative growth of a real set. This chapter proves the required discretized sum–product theorem internally, with the covering-number and Frostman bookkeeping in the precise form needed after passage to a local torus coordinate.

4.1 Why a scalar theorem is needed

The previous Volume-XIV local-product draft compressed the one-dimensional growth input into a few sentences inside a “three-dimensional small-doubling lemma.” That compression was not acceptable: the compact product theorem needs a genuine power gain at a real one-dimensional scale, and every loss has to remain visible when the argument is transferred to a torus coordinate.

We prove a quantitative discretized sum–product theorem for subsets of $[1, 2]$. The formulation is deliberately stronger than the eventual $SU(2)$ application needs, because it gives an explicit gain exponent and isolates the only hypotheses that the rich-torus stage must verify.

For $X \subset \mathbb{R}$ and $t > 0$, let $\mathcal{E}_t(X)$ be the least number of intervals of length t needed to cover X . Let

$$X_t := \mathcal{N}_t(X) \cap t\mathbb{Z},$$

where $\mathcal{N}_t(X)$ is the closed t -neighborhood. Up to absolute constants,

$$|X_t| \asymp \mathcal{E}_t(X).$$

All implicit constants below may depend on the fixed dimension parameter σ , but never on δ .

4.2 The scalar theorem

Theorem 4.1 (Internal discretized sum–product theorem). *Fix $0 < \sigma < 1$ and $C_0 \geq 1$. Let $A \subset [1, 2]$ be δ -separated, with*

$$|A| \asymp \delta^{-\sigma},$$

and assume the Frostman-type non-concentration estimate

$$|A \cap I| \leq C_0 |I|^\sigma |A| \tag{4.1}$$

for every interval I of length at least δ . Then for every

$$0 < c < c_0(\sigma) := \frac{\sigma(1 - \sigma)}{2(19 + 6\sigma)}$$

there is $\delta_0 = \delta_0(\sigma, c, C_0) > 0$ such that, for $0 < \delta < \delta_0$,

$$\mathcal{E}_\delta(A + A) + \mathcal{E}_\delta(A \cdot A) \gg_{\sigma, c} C_0^{-O_\sigma(1)} \delta^{-c} |A|. \quad (4.2)$$

The proof follows a robust dense-versus-gap argument. We give it because later Volume XIV needs to know exactly where scale regularity, quotient separation and the final power gain enter.

4.3 Covering-number calculus

Lemma 4.2 (Discretization comparison). *For bounded $X, Y \subset \mathbb{R}$ and fixed $m \geq 1$,*

$$|X_\delta| \asymp \mathcal{E}_\delta(X), \quad \mathcal{E}_\delta(X + Y) \asymp |X_\delta + Y_\delta|,$$

with constants depending only on the number of summands. The same holds for differences. If $X, Y \subset [1, 2]$, then

$$\mathcal{E}_\delta(XY) \asymp |(XY)_\delta|$$

with absolute constants.

Proof. A maximal δ -separated subset of X has cardinality comparable to the least number of δ -intervals covering X . Every point of X_δ lies within δ of X , while a maximal separated subset maps to distinct points of a fixed constant multiple of the δ -grid; this gives the first comparison.

For sums, if $x = x_0 + O(\delta)$ and $y = y_0 + O(\delta)$, then $x + y = x_0 + y_0 + O(\delta)$, so replacing each input by its grid discretization changes the sumset only by a bounded enlargement of the covering intervals. The reverse inclusion follows by approximating each original summand by a grid point. Products on $[1, 2]$ are uniformly Lipschitz in each variable, so the same argument applies to xy . $\square$

We use one standard additive-combinatorial tool. For completeness, we record the exact form used below.

Lemma 4.3 (Petridis minimal-growth lemma). *Let X, B be finite nonempty subsets of an abelian group and assume*

$$K_0 := \frac{|X + B|}{|X|} \leq \frac{|Z + B|}{|Z|} \quad \text{for every nonempty } Z \subset X.$$

Then for every finite set C ,

$$|X + B + C| \leq K_0 |X + C|. \quad (4.3)$$

Proof. Write $C = \{c_1, \dots, c_r\}$ in an arbitrary order. Put $X_1 = X$, and for $i > 1$ define

$$X_i := \{x \in X : c_i + x \notin \{c_1, \dots, c_{i-1}\} + X\}.$$

Then the sets $c_i + X_i$ are pairwise disjoint and

$$C + X = \bigsqcup_{i=1}^r (c_i + X_i), \quad |C + X| = \sum_{i=1}^r |X_i|.$$

We prove (4.3) by induction on r . The case $r = 1$ is the definition of K_0 . For $r > 1$ set $C' = C \setminus \{c_r\}$ and $X_r^c = X \setminus X_r$. By the definition of X_r ,

$$c_r + X_r^c \subset C' + X, \quad c_r + X_r^c + B \subset C' + X + B.$$

Hence the points contributed for the first time by the last translate satisfy

$$|C + X + B| \leq |C' + X + B| + |X + B| - |X_r^c + B|.$$

The induction hypothesis bounds the first term by $K_0 \sum_{i < r} |X_i|$. If X_r^c is empty, the corresponding lower bound is trivial; otherwise minimality gives $|X_r^c + B| \geq K_0 |X_r^c|$. Also $|X + B| = K_0 |X|$; therefore the last two terms contribute at most $K_0 |X_r|$. Summing proves (4.3). $\square$

Corollary 4.4 (One-summand Plünnecke inequality). *Let A, B be finite nonempty subsets of an abelian group and suppose*

$$|A + B| \leq K|A|.$$

Then there is a nonempty $X \subset A$ such that, for every integer $h \geq 1$,

$$|X + hB| \leq K^h|X|.$$

Proof. Choose nonempty $X \subset A$ minimizing $|Z + B|/|Z|$ over nonempty $Z \subset A$, and write the minimum as $K_0 \leq K$. Lemma 4.3, with $C = (h - 1)B$, gives inductively

$$|X + hB| \leq K_0|X + (h - 1)B| \leq K_0^h|X| \leq K^h|X|.$$

□

Proposition 4.5 (Fixed-summand Plünnecke refinement). *For every fixed $k \geq 1$ there is a constant C_k with the following property. Let $X, Y_1, \dots, Y_k$ be finite subsets of an abelian group and suppose*

$$|X + Y_i| \leq K_i|X| \quad (1 \leq i \leq k),$$

where $K_i \geq 1$. Then there is a nonempty $X' \subset X$ such that

$$|X' + Y_1 + \dots + Y_k| \leq C_k \left(\prod_{i=1}^k K_i \right) |X'|. \quad (4.4)$$

Moreover one can choose a union $X^\sharp \subset X$ with $|X^\sharp| \geq |X|/2$ and

$$|X^\sharp + Y_1 + \dots + Y_k| \leq C'_k \left(\prod_{i=1}^k K_i \right) |X^\sharp|. \quad (4.5)$$

Proof. The point is that the exact constant is irrelevant for the present application, so the different-summands theorem can be reduced to the one-summand result without a tensor-power limit. Put $L_i = \lceil K_i \rceil$, $P = \prod_i L_i$, and $M_i = P/L_i$. Let H_i be a cyclic group of order M_i (the trivial group when $M_i = 1$), and work in

$$\widehat{G} = G \times H_1 \times \dots \times H_k.$$

Embed X as $\widehat{X} = X \times \{0\}^k$. For each i , let $\widehat{Y}_i$ consist of the points whose G -coordinate lies in Y_i , whose H_i -coordinate is arbitrary, and whose other auxiliary coordinates are zero. If $\widehat{B} = \bigcup_i \widehat{Y}_i$, then

$$|\widehat{X} + \widehat{B}| \leq \sum_i M_i |X + Y_i| \leq \sum_i M_i L_i |X| = kP|\widehat{X}|.$$

Corollary 4.4, applied to $\widehat{X}, \widehat{B}$ with $h = k$, gives a nonempty $\widehat{X}' = X' \times \{0\}^k$ such that

$$|\widehat{X}' + k\widehat{B}| \leq (kP)^k |X'|.$$

But

$$(X' + Y_1 + \dots + Y_k) \times H_1 \times \dots \times H_k \subset \widehat{X}' + k\widehat{B}.$$

Since $\prod_i M_i = P^{k-1}$,

$$|X' + Y_1 + \dots + Y_k| \leq k^k P |X'| \leq 2^k k^k \left(\prod_i K_i \right) |X'|.$$

Thus (4.4) holds, for instance with $C_k = 2^k k^k$.

For the large refinement, repeatedly apply the first assertion to the unused part $R \subset X$ while $|R| \geq |X|/2$. On such an R ,

$$|R + Y_i| \leq |X + Y_i| \leq 2K_i|R|.$$

Every extracted piece therefore satisfies (4.4) with $2K_i$ in place of K_i . Stop once the extracted pieces cover at least half of X , and let $X^\sharp$ be their union. The image of a union is contained in the union of the images, so summing the piecewise estimates proves (4.5), with another constant depending only on k . $\square$

Proposition 4.6 (Ruzsa triangle inequality). *For finite nonempty X, Y, Z in an abelian group,*

$$|X - Z| \leq \frac{|X - Y||Y - Z|}{|Y|}.$$

Consequently, for bounded subsets of $\mathbb{R}$,

$$\mathcal{E}_\delta(X - Z) \ll \frac{\mathcal{E}_\delta(X - Y)\mathcal{E}_\delta(Y - Z)}{\mathcal{E}_\delta(Y)}.$$

Proof. For each $s \in X - Z$, fix one representation $s = x_s - z_s$. The map

$$(s, y) \mapsto (x_s - y, y - z_s)$$

from $(X - Z) \times Y$ into $(X - Y) \times (Y - Z)$ is injective, since the two output coordinates add to s and then determine y . This proves the finite inequality. The covering version follows from Lemma 4.2 after replacing each set by its δ -grid discretization. $\square$

Corollary 4.7 (Covering Plünnecke). *If*

$$\mathcal{E}_\delta(X + Y_i) \leq K_i \mathcal{E}_\delta(X),$$

then there is a δ -refinement $X' \subset X$, meaning $\mathcal{E}_\delta(X') \geq \frac{1}{2}\mathcal{E}_\delta(X)$, for which

$$\mathcal{E}_\delta(X' + Y_1 + \cdots + Y_k) \ll_k \left(\prod_i K_i \right) \mathcal{E}_\delta(X').$$

In particular,

$$\mathcal{E}_\delta(Y_1 + \cdots + Y_k) \ll_k \left(\prod_i K_i \right) \mathcal{E}_\delta(X).$$

Proof. Discretize all sets on the δ -grid, apply the large-refinement assertion (4.5) of Proposition 4.5, and thicken the selected grid subset back to a subset of X . Lemma 4.2 absorbs the bounded rounding errors. For the final assertion, choose any point $x \in X'$; then $Y_1 + \cdots + Y_k \subset -x + (X' + Y_1 + \cdots + Y_k)$, and use $\mathcal{E}_\delta(X') \leq \mathcal{E}_\delta(X)$. $\square$

4.4 A scale-regular subset

The next lemma prevents a small additive-doubling statement that holds only at the finest scale from becoming useless after multiplication by a small coefficient.

Lemma 4.8 (Uniform-tree preprocessing). *Fix $\tau_0 > 0$. Let $X \subset [1, 2]$ be δ -separated and write $|X| = \delta^{-\tau}$ with $\tau_0 \leq \tau \leq 1$. Suppose*

$$\mathcal{E}_\delta(X + X) \leq K|X|.$$

For every $\varepsilon > 0$ there is $\tilde{X} \subset X$ satisfying

$$|\tilde{X}| \geq \delta^\varepsilon |X|$$

and, simultaneously for all $\delta < t < 1$,

$$\mathcal{E}_t(\tilde{X} + \tilde{X}) \ll_{\tau_0, \varepsilon} \delta^{-10\varepsilon} K \mathcal{E}_t(\tilde{X}). \quad (4.6)$$

Proof. Choose an integer block size $j = j(\tau_0, \varepsilon)$ so large that $2j \leq 2^{\varepsilon j}$ and that the harmless constants below are dominated by $2^{\varepsilon j}$. Replacing δ by a comparable smaller scale changes covering numbers by a constant depending only on j , so assume $\delta = 2^{-mj}$.

For $0 \leq \ell \leq m$, let $\mathcal{D}_\ell$ be the partition of $[1, 2]$ into intervals of length $2^{-\ell j}$. Let X_ℓ be the occupied intervals at level ℓ . Starting with $Z_m = X_m$, construct $Z_\ell \subset X_\ell$ upward. Once $Z_{\ell+1}$ is fixed,

$$|Z_{\ell+1}| = \sum_{I \in X_\ell} \#\{J \in Z_{\ell+1} : J \subset I\}.$$

Every positive summand lies between 1 and 2^j . Split the parents according to the dyadic range $2^{q-1} \leq \#\{J \subset I\} < 2^q$, $1 \leq q \leq j$, and retain the range carrying the largest part of the displayed sum. Thus the retained children above each $I \in Z_\ell$ are comparable within a factor two, and the total descendant mass loses at most a factor $2j$ at each level.

Let $\tilde{X}_\ell$ be the set of points of X whose ancestors belong to every Z_r for $r > \ell$. Then

$$|\tilde{X}_\ell| \geq (2j)^{-(m-\ell)}|X|, \quad |\tilde{X}_0| \geq (2j)^{-m}|X| \geq \delta^\varepsilon|X|.$$

Set $\tilde{X} = \tilde{X}_0$. The near-uniform branching also implies the following quantitative occupancy statement: every $I \in Z_\ell$ contains at least

$$\delta^{2\varepsilon} \frac{|X|}{|Z_\ell|} \tag{4.7}$$

points of the original set X , once δ is sufficiently small. Indeed the average occupancy is $|\tilde{X}_\ell|/|Z_\ell|$, the dyadic selection makes all nonzero occupancies comparable by at most $2^m \leq \delta^{-\varepsilon}$, and the global retained-mass loss is at most $\delta^{-\varepsilon}$.

Fix ℓ and put $t = 2^{-\ell j}$. A member of the interval sumset $Z_\ell + Z_\ell$ has length $2t$. Choose for each such interval one representing pair $I, J \in Z_\ell$ and fix one point $y \in X \cap J$. As x runs over $X \cap I$, the sums $x + y$ are δ -separated, so by (4.7) that coarse sum interval contains at least $\delta^{2\varepsilon}|X|/|Z_\ell|$ separated points of $X + X$. Equal-length sum intervals have bounded overlap after passing to one of a bounded number of disjoint subfamilies. Consequently

$$\mathcal{E}_\delta(X + X) \gg |Z_\ell + Z_\ell| \delta^{2\varepsilon} \frac{|X|}{|Z_\ell|}.$$

Since the left side is at most $K|X|$,

$$|Z_\ell + Z_\ell| \ll \delta^{-2\varepsilon} K|Z_\ell| \leq \delta^{-10\varepsilon} K|Z_\ell|.$$

At scale t , the occupied intervals Z_ℓ give bounded-overlap covers of $\tilde{X}$, while $Z_\ell + Z_\ell$ gives one for $\tilde{X} + \tilde{X}$. Hence (4.6) holds at all tree scales. If t is intermediate between consecutive tree scales, the ratio of the two is at most 2^j , so monotonicity of covering numbers and a change in the implicit constant give (4.6) for every $\delta < t < 1$. $\square$

4.5 Consequences of simultaneous small additive and multiplicative growth

We now begin the proof of Theorem 4.1. Apply Lemma 4.8 to A and call the resulting set A' . Its non-concentration estimate is weakened only by the factor $\delta^{-\varepsilon}$:

$$|A' \cap I| \leq C_0 \delta^{-\varepsilon} |I|^\sigma |A'|. \tag{4.8}$$

Define

$$K := \frac{\mathcal{E}_\delta(A' + A') + \mathcal{E}_\delta(A'A')}{|A'|}. \quad (4.9)$$

Fix the target exponent $c < c_0(\sigma)$ from Theorem 4.1, and choose an intermediate exponent

$$c < c_1 < c_0(\sigma).$$

If $K \geq \delta^{-c_1}$, then

$$\mathcal{E}_\delta(A + A) + \mathcal{E}_\delta(AA) \geq K|A'| \gg \delta^{-c_1+\varepsilon}|A|,$$

which is at least a constant multiple of $\delta^{-c}|A|$ once $\varepsilon < c_1 - c$ and δ is small. We may therefore assume throughout the structural argument that

$$K < \delta^{-c_1}. \quad (4.10)$$

Since $c_0(\sigma) < \sigma/4$, we may and do choose $\varepsilon < \sigma/16$. This observation will ensure that every later large subset to which Lemma 4.8 is reapplied still has cardinality $\delta^{-\tau}$ with $\tau \geq \sigma/2$. We now prove the stronger lower bound forced by the dense/gap dichotomy.

Lemma 4.9 (Popular multiplicative overlap). *There exist $b \in A'$, a set $\bar{A} \subset A'$, and a number $K^{-1} \leq \rho \leq 1$ such that*

$$|\bar{A}| \gg |\log \delta|^{-1} \rho |A'|, \quad (4.11)$$

$$|(aA')_\delta \cap (bA')_\delta| \asymp |A'| K^{-1} \rho^{-1} \quad (a \in \bar{A}), \quad (4.12)$$

and for each $a \in \bar{A}$,

$$\mathcal{E}_\delta(aA' \pm bA') \ll K^3 \rho |A'|. \quad (4.13)$$

Proof. The family of grid sets $(aA')_\delta$ has total incidence $\asymp |A'|^2$. Since their union lies in a bounded thickening of $A'A'$ and

$$\mathcal{E}_\delta(A'A') \leq K|A'|,$$

Cauchy–Schwarz gives

$$\sum_{a,a' \in A'} |(aA')_\delta \cap (a'A')_\delta| \gg \frac{|A'|^4}{K|A'|} = |A'|^3 K^{-1}.$$

Average in the second variable and choose b for which the inner sum is $\gg |A'|^2 K^{-1}$. Dyadic pigeonholing the intersection sizes yields ρ and $\bar{A}$ with (4.11)–(4.12); there are only $O(|\log \delta|)$ dyadic ranges.

For fixed $a \in \bar{A}$, put $X = (aA')_\delta \cap (bA')_\delta$. Then

$$|X| \gg K^{-1} \rho^{-1} |A'|.$$

Both $X + aA'$ and $X + bA'$ are contained, up to bounded δ -thickening, in a translate/dilate of $A' + A'$ or $A'A'$ and hence have covering number $O(K|A'|) = O(K^2 \rho |X|)$. Corollary 4.7, followed by Proposition 4.6 for the difference case, gives

$$\mathcal{E}_\delta(aA' \pm bA') \ll K^4 \rho^2 |X| \ll K^3 \rho |A'|.$$

□

By (4.11), $\rho \geq K^{-1}$ and (4.10),

$$|\bar{A}| \gg |\log \delta|^{-1} \delta^{c_1} |A'| \gg \delta^{-\sigma/2}$$

for sufficiently small δ (after decreasing ε if necessary). Thus Lemma 4.8 applies to $\bar{A}$ with the fixed lower exponent $\tau_0 = \sigma/2$. Call the resulting subset A_1 . Then

$$|A_1| \gg \delta^{O(\varepsilon)} |\log \delta|^{-1} \rho |A'|, \quad (4.14)$$

and for every $\delta < t < 1$,

$$\mathcal{E}_t(A_1 + A_1) \ll K |\log \delta| \rho^{-1} \delta^{-O(\varepsilon)} \mathcal{E}_t(A_1). \quad (4.15)$$

Lemma 4.10 (Two-coefficient upper bound). *Let $d_1, d_2 \in A_1 - A_1$ and put $d = \max(|d_1|, |d_2|)$. Then*

$$\mathcal{E}_\delta(d_1 A' + d_2 A') \ll C_0 \delta^{-O(\varepsilon)} K^{12} \rho^4 d^\sigma |A'|. \quad (4.16)$$

Proof. Write $d_1 = a_1 - b_1$ and $d_2 = a_2 - b_2$ with $a_i, b_i \in A_1 \subset \bar{A}$. Apply the covering Plünnecke refinement to the four sets $a_1 A', -b_1 A', a_2 A', -b_2 A'$ and the base set $b A'$, using (4.13). We obtain a refinement $A'' \subset A'$ with $|A''| \gg |A'|$ in covering-number scale and

$$\mathcal{E}_\delta(b A'' + d_1 A' + d_2 A') \ll K^{12} \rho^4 |A'|. \quad (4.17)$$

(The displayed exponent 12 is deliberately nonoptimized; any fixed polynomial power would suffice later.)

Put $S = d_1 A' + d_2 A'$. Since $A' \subset [1, 2]$, the set S lies in an interval of length at most Cd for an absolute C . Choose a maximal $10Cd$ -separated subset $P \subset b A''$. Maximality gives $|P| \gg \mathcal{E}_d(b A'')$, while the intervals containing the translates $p + S$ ($p \in P$) are pairwise disjoint. Hence a δ -cover of $b A'' + S$ must cover each $p + S$ separately, and

$$\mathcal{E}_\delta(b A'' + S) \geq \sum_{p \in P} \mathcal{E}_\delta(p + S) = |P| \mathcal{E}_\delta(S) \gg \mathcal{E}_d(b A'') \mathcal{E}_\delta(S).$$

Rearranging gives

$$\mathcal{E}_\delta(d_1 A' + d_2 A') \ll \mathcal{E}_d(b A'')^{-1} \mathcal{E}_\delta(b A'' + d_1 A' + d_2 A').$$

By (4.8), every interval of length $O(d)$ contains at most $C_0 \delta^{-\varepsilon} d^\sigma |A'|$ points of A' . Since A'' is a fixed-proportion covering refinement, this forces

$$\mathcal{E}_d(b A'') \gg C_0^{-1} \delta^\varepsilon d^{-\sigma}.$$

Substitute this and (4.17) to obtain (4.16). □

We also need the same estimate with a bounded number of copies of one coefficient.

Lemma 4.11 (Bounded repeated coefficient, no refinement). *Fix $k \geq 2$. Let $d_1, d_2 \in A_1 - A_1$, assume $|d_2| \geq \delta$, and put $d = \max(|d_1|, |d_2|)$. Then*

$$\mathcal{E}_\delta(d_1 A_1 + \underbrace{d_2 A_1 + \cdots + d_2 A_1}_{k \text{ copies}}) \ll C_0 \delta^{-O_k(\varepsilon)} |\log \delta|^{O_k(1)} K^{13+k} \rho^{2-k} d^\sigma |A'|. \quad (4.18)$$

Proof. Choose, exactly as in the proof of Lemma 4.10, a δ -refinement $A'' \subset A'$ such that

$$\mathcal{E}_\delta(b A'' + d_1 A' + d_2 A') \ll K^{12} \rho^4 |A'|.$$

After restricting $d_1 A'$ and $d_2 A'$ to A_1 and using (4.14),

$$\mathcal{E}_\delta(b A'' + d_1 A_1 + d_2 A_1) \ll \delta^{-O(\varepsilon)} |\log \delta| K^{12} \rho^3 |A_1|. \quad (4.19)$$

Set $X = d_2 A_1$. Applying (4.15) at scale $\delta/|d_2|$ and transporting by dilation gives

$$\mathcal{E}_\delta(X + X) \ll L \mathcal{E}_\delta(X), \quad L := K \delta^{-O(\varepsilon)} \rho^{-1} |\log \delta|.$$

The “in particular” part of Corollary 4.7, applied to $k + 1$ identical copies of X , yields

$$\mathcal{E}_\delta(\underbrace{X + \cdots + X}_{k+1 \text{ copies}}) \ll_k L^{k+1} \mathcal{E}_\delta(X). \quad (4.20)$$

Now apply the same corollary with base X and with

$$Y_1 = \underbrace{X + \cdots + X}_{k \text{ copies}}, \quad Y_2 = bA'' + d_1A_1.$$

By (4.20), the expansion ratio for Y_1 is at most $O_k(L^{k+1})$. If

$$R := \frac{\mathcal{E}_\delta(X + Y_2)}{\mathcal{E}_\delta(X)},$$

then (4.19) bounds $R\mathcal{E}_\delta(X)$. Hence the global conclusion of the covering Plünnecke corollary gives

$$\begin{aligned} & \mathcal{E}_\delta(bA'' + d_1A_1 + \underbrace{d_2A_1 + \cdots + d_2A_1}_{k \text{ copies}}) \\ & \ll_k L^{k+1} R\mathcal{E}_\delta(X) \\ & \ll \delta^{-O_k(\varepsilon)} |\log \delta|^{O_k(1)} K^{13+k} \rho^{2-k} |A_1|. \end{aligned}$$

Finally the set without bA'' lies in an interval of length $O_k(d)$. The separated-translate argument of Lemma 4.10, with the same A'' , gives

$$\mathcal{E}_d(bA'') \gg C_0^{-1} \delta^\varepsilon d^{-\sigma}.$$

Dividing by this lower bound and using $|A_1| \leq |A'|$ proves (4.18). $\square$

4.6 The quotient-set dichotomy

Fix $0 < \gamma < 1/2$ and set

$$B := \left\{ \frac{a_1 - a_2}{a_3 - a_4} : a_i \in A_1, |a_3 - a_4| > \delta^\gamma \right\}.$$

Then $B \subset [-\delta^{-\gamma}, \delta^{-\gamma}]$ and $0, 1 \in B$. Let

$$s \asymp \delta^{1-2\gamma}$$

be a dyadic scale.

Lemma 4.12 (Dense-or-gap dichotomy). *At least one of the following holds.*

(A) *There is $b \in B \cap [0, 1]$ for which either $b/2$ or $(b+1)/2$ has distance at least s from B .*

(B) $\mathcal{E}_s(B \cap [0, 1]) \gg s^{-1}$.

Proof. Assume (A) fails. The $2s$ -neighborhood of B is then invariant, up to the same $2s$ error, under the two contractions

$$x \mapsto x/2, \quad x \mapsto (x+1)/2.$$

Starting from $0, 1 \in B$, induction shows that every dyadic rational $p/2^n$ with $0 \leq p \leq 2^n$ belongs to the $2s$ -neighborhood of B for every n for which $2^{-n} \gtrsim s$. Taking $2^{-n} \asymp s$ gives $\gg s^{-1}$ s -separated points within $O(s)$ of $B \cap [0, 1]$, proving (B). $\square$

4.7 Dense case

Assume alternative (B). By averaging the s -cells meeting $B \cap [0, 1]$, choose

$$b = \frac{b_1 - b_2}{b_3 - b_4} \in B \cap [0, 1]$$

with $|b_3 - b_4| > \delta^\gamma$ such that the number of quadruples $(a_1, a_2, a_3, a_4) \in A_1^4$ satisfying

$$\left| \frac{a_1 - a_2}{a_3 - a_4} - b \right| \ll \delta^{1-2\gamma}, \quad |a_3 - a_4| > \delta^\gamma, \quad (4.21)$$

is $\ll |A_1|^4 \delta^{1-2\gamma}$.

Set $D_1 = b_1 - b_2$ and $D_2 = b_3 - b_4$. Lemma 4.10 gives

$$\mathcal{E}_\delta(D_1 A_1 + D_2 A_1) \ll C_0 \delta^{-O(\varepsilon)} K^{12} \rho^4 |D_2|^\sigma |A'|. \quad (4.22)$$

Let Q be the approximate additive-energy set of quadruples satisfying

$$D_2 a_1 + D_1 a_4 = D_2 a_2 + D_1 a_3 + O(\delta). \quad (4.23)$$

Cauchy–Schwarz in the representation function gives

$$\mathcal{E}_\delta(D_1 A_1 + D_2 A_1) \gg \frac{|A_1|^4}{|Q|}. \quad (4.24)$$

Split Q according as $|a_3 - a_4|$ is at least or at most δ^γ .

In the first range, dividing (4.23) by $D_2(a_3 - a_4)$ shows that the quotient in (4.21) is within $O(\delta^{1-2\gamma})$ of b . Hence

$$|Q_{\text{large}}| \ll |A_1|^4 \delta^{1-2\gamma}.$$

If this contributes at least half of Q , then (4.24), (4.14), and (4.22) imply

$$K \gg C_0^{-O(1)} |\log \delta|^{-O(1)} \delta^{(2\gamma+\sigma-1)/12+O(\varepsilon)}. \quad (4.25)$$

In the second range, choose a_1, a_4 first. The Frostman estimate gives at most $C_0 \delta^{\gamma\sigma} |A|$ choices for a_3 . Once a_3 is fixed, (4.23) confines a_2 to an interval of length $O(\delta/|D_2|)$, so there are at most $C_0 |D_2|^{-\sigma}$ choices after using $|A| \asymp \delta^{-\sigma}$. Thus

$$|Q_{\text{small}}| \ll C_0^2 |A_1|^2 |A'| \delta^{\gamma\sigma+O(\varepsilon)} |D_2|^{-\sigma}.$$

If this contributes at least half of Q , then (4.24) and (4.22) give

$$K \gg C_0^{-O(1)} |\log \delta|^{-O(1)} \delta^{-\gamma\sigma/12+O(\varepsilon)}. \quad (4.26)$$

Therefore the dense case yields

$$K \gg C_0^{-O(1)} |\log \delta|^{-O(1)} \delta^{O(\varepsilon)} \min \left(\delta^{(2\gamma+\sigma-1)/12}, \delta^{-\gamma\sigma/12} \right). \quad (4.27)$$

4.8 Gap case

Assume alternative (A). Thus, for some

$$b = \frac{b_1 - b_2}{b_3 - b_4} \in B, \quad |b_3 - b_4| \geq \delta^\gamma,$$

either $b/2$ or $(b+1)/2$ is s -separated from B . Write the separated number as e_1/e_2 . In the first case one can take

$$e_1 = d_1, \quad e_2 = 2d_2,$$

and in the second

$$e_1 = d_1 + d_2, \quad e_2 = 2d_2,$$

for $d_1, d_2 \in A_1 - A_1$.

Let Q consist of quadruples satisfying

$$e_2 a_1 + e_1 a_4 = e_2 a_2 + e_1 a_3 + O(\delta).$$

If $|a_3 - a_4| \geq \delta^\gamma$, division shows

$$\left| \frac{a_1 - a_2}{a_3 - a_4} - \frac{e_1}{e_2} \right| \ll \delta^{1-2\gamma} \asymp s,$$

contradicting the gap property because the quotient belongs to B . Hence every quadruple in Q has $|a_3 - a_4| \leq \delta^\gamma$.

Choose a_1, a_4 first. The Frostman bound gives at most $C_0 \delta^{\gamma\sigma} |A'| \delta^{-O(\varepsilon)}$ choices for a_3 , and then at most $C_0 |e_2|^{-\sigma} \delta^{-O(\varepsilon)}$ choices for a_2 . Therefore

$$|Q| \ll C_0^{O(1)} |A_1|^2 |A'| \delta^{\gamma\sigma - O(\varepsilon)} |e_2|^{-\sigma}.$$

The energy inequality therefore gives

$$\mathcal{E}_\delta(e_1 A_1 + e_2 A_1) \gg C_0^{-O(1)} \rho^2 |\log \delta|^{-O(1)} \delta^{-\gamma\sigma + O(\varepsilon)} |e_2|^\sigma |A'|. \quad (4.28)$$

On the other hand, in Case (A.1) the set $e_1 A_1 + e_2 A_1$ is contained in $d_1 A_1 + d_2 A_1 + d_2 A_1$, and in Case (A.2) it is contained in $d_1 A_1 + d_2 A_1 + d_2 A_1 + d_2 A_1$. Since $|d_2| \geq \delta^\gamma$, Lemma 4.11, with $k = 2$ or 3 , gives in the worse second case

$$\mathcal{E}_\delta(e_1 A_1 + e_2 A_1) \ll C_0 \rho^{-1} |\log \delta|^{O(1)} \delta^{-O(\varepsilon)} K^{16} |e_2|^\sigma |A'|. \quad (4.29)$$

Comparing (4.28) and (4.29) gives

$$K^{16} \gg C_0^{-O(1)} |\log \delta|^{-O(1)} \rho^3 \delta^{-\gamma\sigma + O(\varepsilon)}.$$

Since $\rho \geq K^{-1}$, we obtain the fully internal, deliberately nonoptimized bound

$$K \gg C_0^{-O(1)} |\log \delta|^{-O(1)} \delta^{-\gamma\sigma/19 + O(\varepsilon)}. \quad (4.30)$$

4.9 Optimization and completion of the proof

By Lemma 4.12, either (4.27) or (4.30) holds. Choose

$$\gamma = \frac{19(1 - \sigma)}{2(19 + 6\sigma)}.$$

Then

$$\frac{1 - \sigma - 2\gamma}{12} = \frac{\gamma\sigma}{19} = \frac{\sigma(1 - \sigma)}{2(19 + 6\sigma)} = c_0(\sigma),$$

while the other dense exponent $\gamma\sigma/12$ is larger. Thus

$$K \gg C_0^{-O(1)} |\log \delta|^{-O(1)} \delta^{-c_0(\sigma) + O(\varepsilon)}.$$

Because A' retained $\delta^{O(\varepsilon)}$ of A and

$$A' + A' \subset A + A, \quad A'A' \subset AA,$$

we conclude

$$\mathcal{E}_\delta(A + A) + \mathcal{E}_\delta(AA) \gg C_0^{-O(1)} |\log \delta|^{-O(1)} \delta^{-c_0(\sigma) + O(\varepsilon)} |A|.$$

Given $c < c_0(\sigma)$, choose ε so small that the $O(\varepsilon)$ loss is less than $(c_0(\sigma) - c)/2$; for sufficiently small δ , the logarithm is absorbed by the remaining power. This is exactly (4.2), completing the proof of Theorem 4.1. $\square$

4.10 Interface for the $SU(2)$ rich-torus stage

The theorem above is the scalar engine. The next stage must not apply it blindly: it must manufacture its Frostman hypothesis from the metric Larsen–Pink estimates.

Proposition 4.13 (Bi-Lipschitz transfer to a rank-one torus coordinate). *Let $T < SU(2)$ be a maximal torus, and restrict to a sufficiently small fixed arc T_0 on which*

$$\theta \mapsto \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$$

is a coordinate chart. For every compact subarc $J \Subset T_0$, the angle map and its inverse are uniformly bi-Lipschitz. Consequently, if $B \subset J$ and the metric covering numbers of B satisfy

$$N(B \cap B_T(x, r), \delta) \leq Cr^\sigma N(B, \delta) \quad (\delta \leq r \leq 1),$$

then, for every maximal δ -separated subset $A \subset \theta(B)$, there is an absolute constant $C_1 \geq 1$ (depending only on the chosen compact chart) such that

$$|A \cap I| \leq C_1 C |I|^\sigma |A| \quad (|I| \geq C_1 \delta).$$

After replacing δ by $C_1 \delta$, this is exactly the non-concentration hypothesis of Theorem 4.1.

Proof. On a fixed compact subarc avoiding the coordinate cut, the derivative of the chart is bounded above and below. The mean-value theorem therefore gives two-sided Lipschitz control. Balls in T correspond to intervals in the angle coordinate with comparable radii, and separated sets/covering numbers are preserved up to constants. The claimed transfer is immediate. $\square$

Remark 4.14 (How the scalar theorem enters the compact argument). The scalar theorem gives a power gain once a rich torus slice is quantitatively non-concentrated. The remaining geometric steps are carried out in the next four chapters: metric Larsen–Pink produces a rich torus; regularization and sum–product create a Lie segment; conjugation supplies three transverse directions; and quantitative BCH turns those directions into three-dimensional covering growth.

Chapter 5

Metric Larsen–Pink and Rich Tori

Assume a discretized set in $SU(2)$ has small bounded-product growth while avoiding the elementary small-ball obstruction. Rank-one Larsen–Pink counting then forces substantial mass near one maximal torus. This chapter proves the specialized metric intersection estimate and extracts the rich torus that drives the one-dimensional amplification in the next chapter.

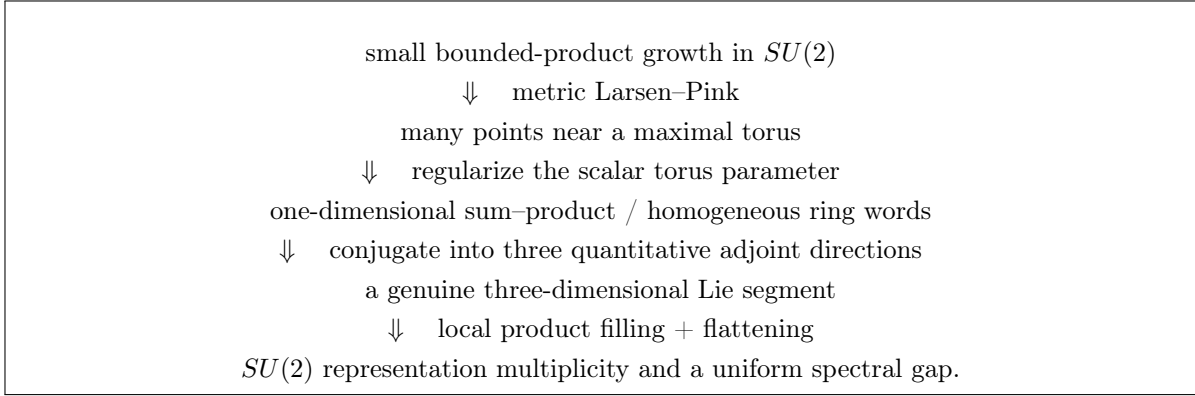


Figure 5.1: The compact-group endpoint of the Bourgain–Gamburd architecture. A rich one-dimensional torus slice is amplified into three Lie directions and then into spectral expansion.

5.1 Rank-one geometry

Fix once and for all a sufficiently small symmetric exponential neighborhood $U \subset SU(2)$. All bounded products occurring below are assumed to remain in a slightly larger fixed exponential neighborhood. We use a smooth left-invariant Riemannian metric. On bounded subsets, left and right translations, inversion, conjugation, $\exp$, and $\log$ have uniformly bounded derivatives.

We identify $\mathfrak{su}(2)$ with $\mathbb{R}^3$ so that, after rescaling the inner product,

$$[X, Y] = X \times Y, \quad \text{Ad}(SU(2)) = SO(3).$$

A one-dimensional subspace $\mathbb{R}n \subset \mathfrak{su}(2)$ exponentiates to a maximal torus T_n . Inside U , every proper positive-dimensional connected subgroup is of this form.

Lemma 5.1 (Distance to a torus and rotation of its axis). *There are constants $c, C > 0$ such that for $g \in U$ and a unit vector $n \in \mathfrak{su}(2)$,*

$$c \, \text{dist}(g, T_n) \leq \| \text{Ad}(g)n - n \| \leq C \, \text{dist}(g, T_n).$$

After shrinking U , the same conclusion holds with the angle between the two lines $\mathbb{R}n$ and $\mathbb{R}\text{Ad}(g)n$ in place of the norm difference.

Proof. The orbit map

$$SU(2)/T_n \longrightarrow S^2, \quad gT_n \longmapsto \text{Ad}(g)n$$

is a smooth diffeomorphism. Its differential at the identity coset is the map $n^\perp \rightarrow n^\perp$, $X \mapsto X \times n$, which is an isometry up to the fixed normalization. Compactness of the unit sphere of axes and the inverse function theorem give uniform two-sided estimates in a fixed neighborhood. $\square$

Lemma 5.2 (Three quantitative adjoint directions). *Let $0 < \rho < \rho_0$ and let $A \subset U$ have the property that for every maximal torus $T < SU(2)$ there exists $a \in A$ with $\text{dist}(a, T) \geq \rho$. For every unit $n \in \mathfrak{su}(2)$ there exist*

$$g_1, g_2, g_3 \in (A \cup \{1\})^2$$

such that

$$|\det(\text{Ad}(g_1)n, \text{Ad}(g_2)n, \text{Ad}(g_3)n)| \geq c\rho^3. \quad (5.1)$$

Consequently, for every unit vector $v \in \mathfrak{su}(2)$, one of these g_i satisfies

$$|\langle v, \text{Ad}(g_i)n \rangle| \geq c\rho^3. \quad (5.2)$$

Proof. Take $g_1 = 1$. By the hypothesis applied to T_n , choose $a \in A$ with $\text{dist}(a, T_n) \geq \rho$. Lemma 5.1 gives

$$\text{dist}(\text{Ad}(a)n, \mathbb{R}n) \geq c\rho.$$

Put $g_2 = a$ and let $P = \text{span}(n, \text{Ad}(a)n)$. Its unit normal w is well defined and the ordered pair spanning P has area at least $c\rho$.

Apply the hypothesis to T_w and choose $b \in A$ with $\text{dist}(b, T_w) \geq \rho$. Then $\text{Ad}(b)P$ makes angle at least $c\rho$ with P . If both $\text{Ad}(b)n$ and $\text{Ad}(ba)n$ had distance less than $c'\rho^2$ from P , then, because n and $\text{Ad}(a)n$ have condition number $O(\rho^{-1})$, the whole plane $\text{Ad}(b)P$ would be $O(c'\rho)$ -close to P , contradicting the preceding angle bound when c' is small. Thus one of b or ba may be chosen as g_3 so that

$$\text{dist}(\text{Ad}(g_3)n, P) \geq c\rho^2.$$

Multiplying this by the $c\rho$ area of the first two vectors proves (5.1). Since the three columns have norm one, the least singular value of their matrix is at least a constant multiple of its determinant. Equation (5.2) follows by applying the transpose matrix to v . $\square$

5.2 Local transverse-intersection counting

We isolate the only differential-geometric counting fact needed in the Larsen–Pink argument. A (r, L) *hypersurface chunk* means a piece which, in a ball of radius r and in exponential coordinates, is the graph of a C^2 function with C^2 norm at most L . Curves are defined similarly. Two tangent spaces are κ -transverse when the sine of the relevant angle is at least κ .

Lemma 5.3 (Transverse hypersurfaces meet in a controlled curve). *There is an absolute C with the following property. Let M, N be (r, L) hypersurface chunks in a three-dimensional exponential chart, with their normals κ -separated at the centers. If*

$$r \leq c\kappa/L,$$

then for every left translate zN arising from points in the two r -balls, $M \cap (zN)^{(C\delta)}$ is contained in the $C\delta\kappa^{-C}L^C$ -neighborhood of a C^2 curve chunk whose complexity is at most $\kappa^{-C}L^C$.

Proof. Pass to exponential coordinates and left-trivialize tangent spaces. Translation and the group law have uniformly bounded C^2 norms in our fixed chart. The two surface equations therefore have gradients whose wedge product is at least $c\kappa$ at the center and, by the restriction on r , at least $c\kappa/2$ throughout the relevant ball. The quantitative implicit function theorem applied to the pair of equations gives a one-dimensional C^2 common zero set, with first and second derivatives bounded by a fixed power of L/κ . If both equations are only satisfied to accuracy $O(\delta)$, the same theorem, or equivalently the normal-coordinate form of the proof, places the point within $O(\delta\kappa^{-C}L^C)$ of that curve. $\square$

Lemma 5.4 (A transverse curve and hypersurface have bounded fibers). *Let P be an (r, L) curve chunk and N an (r, L) hypersurface chunk. Suppose the tangent to P has scalar product at least κ with a unit normal to N at the relevant center, and $r \leq c\kappa/L$. Then*

$$\mathcal{N}(P \cap (zN)^{(C\delta)}, \delta) \leq \kappa^{-C}L^C$$

for every translate zN arising inside the chart.

Proof. Use arclength t on P and let F be a defining function for zN with unit normal at the center. The transversality assumption gives $|\frac{d}{dt}F| \geq c\kappa$ throughout the small ball. Thus each connected portion on which $|F| \leq C\delta$ has length $O(\delta/\kappa)$. The controlled-complexity graph charts split P into at most $\kappa^{-C}L^C$ such portions. A δ -separated set in them has the asserted cardinality. $\square$

Lemma 5.5 (Pair counting through product fibers). *Let X, Y be δ -separated subsets of two r -balls in U . Suppose the corresponding hypersurface chunks are κ -transverse and have complexity L . Then*

$$|X||Y| \leq C\mathcal{N}(XY^{-1}, \delta) \kappa^{-C}L^C \max_P \mathcal{N}(X \cap P^{(\delta\kappa^{-C}L^C)}, \delta),$$

where P ranges over the curve chunks supplied by Lemma 5.3.

Proof. Cover XY^{-1} by δ -balls. If (x, y) maps into the ball centered at z , then x lies within $O(\delta)$ of zy . Hence x lies in the intersection of the first hypersurface with an $O(\delta)$ -neighborhood of z times the second. Lemma 5.3 places all such x near one controlled curve. Once x is fixed, the equation $xy^{-1} = z + O(\delta)$ determines y up to a bounded number of δ -balls because inversion and translation are bi-Lipschitz. Summing over output balls proves the estimate. $\square$

5.3 A specialized metric Larsen–Pink estimate

For a set S , write $S^{(t)}$ for its t -neighborhood. The following theorem is the rank-one substitute for the general submanifold Larsen–Pink machinery.

Theorem 5.6 (Metric Larsen–Pink for a regular $SU(2)$ conjugacy class). *There are constants C, m_0 depending only on the fixed neighborhood U . Let $0 < \delta \ll \rho \ll 1$, let $A \subset U$ be symmetric, contain 1, and be δ -separated. Put $n = |A|$. Assume:*

(i) *for every maximal torus T there is $a_T \in A$ with $\text{dist}(a_T, T) \geq \rho$;*

(ii) *for every $1 \leq m \leq m_0$,*

$$\mathcal{N}(A^m, \delta) \leq \rho^{-C}n; \tag{5.3}$$

(iii) *$a \in U$ satisfies $\text{dist}(a, \{1, -1\}) \geq \rho$.*

Let $\mathcal{C}_a$ be the conjugacy class of a . Then, for every fixed $m \leq m_0/4$,

$$\mathcal{N}(A^m \cap \mathcal{C}_a^{(\delta)}, \delta) \leq \rho^{-C}n^{2/3}. \tag{5.4}$$

Proof. Let X be a maximal δ -separated subset of the set on the left, and put $q = |X|$. Project each point of X to a nearest point of $\mathcal{C}_a$. Because the projection moves points by $O(\delta)$, a bounded-coloring argument extracts a subset of size $\gg q$ whose projected points are still $c\delta$ -separated. Replacing X by this subset and changing δ by an absolute factor, we may therefore suppose $X \subset \mathcal{C}_a$ exactly. The regularity assumption on a implies that $\mathcal{C}_a$ is a smooth hypersurface. In the vector model $SU(2) \simeq S^3$, it is a level set of trace; equivalently its tangent plane is the orthogonal complement, in left trivialization, of the axis of the centralizer torus. Since a is ρ -away from the center, its local C^2 complexity is at most ρ^{-C} .

Set $r = \rho^{C_1}$, with C_1 large compared with all complexity exponents below. Cover the fixed chart by $O(r^{-3}) = \rho^{-O(1)}$ balls of radius r and keep one containing a subset $X_0 \subset X$ with

$$|X_0| \geq \rho^C q. \quad (5.5)$$

Let n_0 be the unit normal of the resulting conjugacy-class chunk. Lemma 5.2 provides a word $h \in (A \cup \{1\})^2$ for which the conjugate chunk $h\mathcal{C}_a h^{-1}$ is ρ^C -transverse to the original chunk after left trivialization. Put $Y = hX_0 h^{-1}$. Then $|Y| = |X_0|$ and, by (5.3),

$$\mathcal{N}(X_0 Y^{-1}, \delta) \leq \rho^{-C} n.$$

Taking C_1 sufficiently large makes transversality stable throughout the two r -balls. Lemma 5.5 therefore gives

$$|X_0|^2 \leq \rho^{-C} n \ell, \quad (5.6)$$

where ℓ is the largest δ -covering number of X_0 in the $\delta\rho^{-C}$ -neighborhood of one of the resulting controlled curve chunks. Absorbing the thickness into the exponent of ρ and localizing once more, we may choose a curve chunk P and a δ -separated set $W \subset X_0$ with

$$|W| \geq \rho^C \ell. \quad (5.7)$$

Let v be the unit tangent to P at its center. Apply Lemma 5.2 to the normal axis n_0 . Equation (5.2) supplies $k \in (A \cup \{1\})^2$ such that the conjugate hypersurface $k\mathcal{C}_a k^{-1}$ has normal making scalar product at least ρ^C with v . With $Z = kX_0 k^{-1}$, Lemma 5.4 and the same product-fiber counting argument give

$$|W| |X_0| \leq \rho^{-C} n. \quad (5.8)$$

Indeed WZ^{-1} lies in a fixed bounded product power of A , so (5.3) controls the number of output δ -balls, while every output fiber is ρ^{-C} by Lemma 5.4.

Combining (5.6), (5.7), and (5.8) gives

$$|X_0|^3 \leq \rho^{-C} n^2.$$

Together with (5.5), this yields $q \leq \rho^{-C} n^{2/3}$, which is (5.4). $\square$

5.4 From Larsen–Pink to a rich maximal torus

Lemma 5.7 (Almost centralizing a regular element means lying near its torus). *Let $a \in U$ satisfy $d(a, 1) \geq \rho$ and let $T = C_{SU(2)}(a)$. If u lies in a fixed sufficiently small neighborhood and*

$$d(uau^{-1}, a) \leq t,$$

then

$$d(u, T) \leq C\rho^{-1}t.$$

Proof. Write $a = \exp(\theta n)$ with n unit. Since $a \in U$ and $d(a, 1) \geq \rho$, $|\theta| \asymp d(a, 1) \geq c\rho$. Conjugation gives $uau^{-1} = \exp(\theta \operatorname{Ad}(u)n)$. The exponential map is bi-Lipschitz in the fixed chart, so the hypothesis implies

$$|\theta| \|\operatorname{Ad}(u)n - n\| \leq Ct.$$

Lemma 5.1 converts this to the stated distance from $T_n = T$. $\square$

Corollary 5.8 (Rich-torus extraction). *Under the hypotheses of Theorem 5.6, there is a maximal torus T such that*

$$\mathcal{N}(A^{-1}A \cap T^{(\delta\rho^{-C})}, \delta) \geq \rho^C n^{1/3}. \quad (5.9)$$

Proof. Choose any maximal torus T_0 . By hypothesis there is $a \in A$ with $d(a, T_0) \geq \rho$, and hence $d(a, 1) \geq \rho$ because $1 \in T_0$. After shrinking U , this a is regular and its centralizer $T = C_{SU(2)}(a)$ is a maximal torus.

Consider the conjugation map on A ,

$$\Phi(x) = xax^{-1}.$$

Its image lies in $A^3 \cap \mathcal{C}_a$. Theorem 5.6, with $m = 3$, shows that it meets at most $\rho^{-C} n^{2/3}$ balls of radius $O(\delta)$. Consequently one such ball contains $\Phi(x)$ for at least $\rho^C n^{1/3}$ distinct $x \in A$. Fix one of them, x_0 . For every other x in this fiber put $u = x_0^{-1}x \in A^{-1}A$. Then

$$d(uau^{-1}, a) \leq C\delta.$$

Lemma 5.7 puts every such u in $T^{(C\delta\rho^{-1})} \subset T^{(\delta\rho^{-C})}$. The map $x \mapsto x_0^{-1}x$ is an isometry for the left-invariant metric, so the selected u 's remain δ -separated. This proves (5.9). $\square$

5.5 From a rich torus to a regular one-dimensional slice

The rich-torus conclusion is exactly the geometric input needed by the scalar sum–product theorem. A local torus coordinate turns a δ -separated subset of the tube in (5.9) into a real set. To invoke the scalar theorem one must still prove the corresponding all-scale Frostman upper bound for a suitable torus slice. That upper bound is obtained by applying the same Larsen–Pink counting argument to smaller torus balls (equivalently, to one-dimensional torus chunks) and pigeonholing a regular scale. This scale-selection step, followed by the sum–product amplification and conversion to a Lie segment, is performed in the next chapter.

Chapter 6

From a Rich Torus to a Lie Segment

A rich torus is not yet enough for three-dimensional product growth. We regularize the torus slice, apply the scalar sum-product theorem, realize the resulting scalar combinations by homogeneous adjoint words, and use BCH together with a fine commutator scale ladder to manufacture a genuine power-scale segment in the Lie algebra. The scale is selected only after the regularized exponent is known, which avoids any fixed-exponent saturation assumption.

6.1 Purpose and interfaces

the rich-torus chapter produces, under small bounded-product growth and quantitative avoidance of maximal tori, a maximal torus T for which a bounded product set has many points in a thin tube around T . The purpose of this chapter is to turn that one-dimensional abundance into a genuine segment of group elements. Four points require proof rather than prose:

- (E1) a thin torus tube must be converted into an actual one-dimensional parameter set with an all-scale Frostman estimate;
- (E2) the parameter set must generate, by *bounded homogeneous ring expressions*, a segment with a power improvement between its length and its approximation error;
- (E3) those scalar ring expressions must be represented by signed sums of actual adjoint operators $\text{Ad}(w)$ with w a bounded word in the original set;
- (E4) signed sums of adjoint operators must be converted, by BCH, into bounded group words; finally, a fine mesh of small-element scales must be built so that the small element can be selected after the variable local segment exponent is known, while keeping the final error at $O(\delta)$.

The first two points are one-dimensional. The adjoint/BCH transfer uses elementary local geometry of $SU(2)$, and the final scale choice is supplied by a bounded commutator mesh rather than by an imported fixed-exponent saturation theorem.

Throughout, $G = SU(2)$ is equipped with a fixed bi-invariant Riemannian metric and U is a fixed sufficiently small exponential chart about 1. Constants implicit in $O(\cdot)$ may depend on U and on fixed dimension parameters, but not on the small scale.

For a finite set $F \subset \mathbb{R}$ let

$$F^{(d)} = \{x_1 \cdots x_d : x_i \in F\}, \quad sF^{(d)} = \underbrace{F^{(d)} + \cdots + F^{(d)}}_{s \text{ summands}}.$$

The adjective *homogeneous* refers to the fact that every monomial has the same degree d . This homogeneity is essential in the adjoint transfer.

Example 6.1 (Three visible Lie directions from one torus). Let

$$H_0 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad t_\theta = \exp(\theta H_0) = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}.$$

Thus a short torus arc satisfies $t_\theta = I + \theta H_0 + O(\theta^2)$. Consider the two $SU(2)$ matrices

$$g = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \quad h = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}.$$

Direct multiplication gives

$$\text{Ad}(g)H_0 = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}, \quad \text{Ad}(h)H_0 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Together with H_0 these are three linearly independent elements of $\mathfrak{su}(2)$. Consequently

$$gt_\theta g^{-1} = I + \theta \text{Ad}(g)H_0 + O(\theta^2), \quad ht_\theta h^{-1} = I + \theta \text{Ad}(h)H_0 + O(\theta^2).$$

This model calculation is the geometry behind the quantitative argument: abundance on one torus becomes useful only after bounded conjugations place copies of its infinitesimal direction in a well-conditioned basis of the three-dimensional Lie algebra. The actual proof below obtains those conjugating words from the original set and controls all errors uniformly.

6.2 An internal homogeneous ring-saturation theorem

We begin with a finite-scale Frostman hypothesis.

Definition 6.2 (Discrete Frostman set). Fix $0 < \kappa < 1$ and $C_0 \geq 1$. A finite h -separated set $F \subset [-2, 2]$ is called (h, κ, C_0) -Frostman if

$$|F \cap I| \leq C_0 |I|^\kappa |F| \tag{6.1}$$

for every interval I of length $|I| \geq h$, and

$$|F| \geq C_0^{-1} h^{-\kappa}. \tag{6.2}$$

The first ingredient is a self-contained projection lemma. Euclidean Plancherel is part of the baseline real-analysis background declared in Volume I; the Riesz-energy and projection deductions themselves are proved here. We use the Fourier transform normalization

$$\widehat{\nu}(\xi) = \int e^{-2\pi i x \cdot \xi} d\nu(x).$$

Lemma 6.3 (Riesz-energy identity at exponent one). Let $n \geq 2$ and let ν be a compactly supported probability measure on $\mathbb{R}^n$ with finite energy

$$I_1(\nu) = \iint \frac{d\nu(x)d\nu(y)}{|x-y|} < \infty.$$

Then, with a constant $c_n > 0$ depending only on n ,

$$I_1(\nu) = c_n \int_{\mathbb{R}^n} |\widehat{\nu}(\xi)|^2 |\xi|^{-(n-1)} d\xi. \tag{6.3}$$

The identity is understood in $[0, \infty]$; the finite-energy hypothesis makes both sides finite.

Proof. For $r > 0$ one has the elementary Gaussian representation

$$r^{-1} = 2 \int_0^\infty e^{-\pi t^2 r^2} dt.$$

Indeed the change of variables $u = tr$ reduces the right hand side to $2 \int_0^\infty e^{-\pi u^2} du = 1$. Tonelli therefore gives directly, for the finite measure $\nu \times \nu$,

$$I_1(\nu) = 2 \int_0^\infty \iint e^{-\pi t^2 |x-y|^2} d\nu(x) d\nu(y) dt.$$

We next verify the Fourier identity used for the Gaussian kernel instead of appealing to a separate positive-definite-kernel theorem. In one dimension put

$$G_t(\xi) = \int_{\mathbb{R}} e^{-\pi t^2 x^2} e^{-2\pi i x \xi} dx.$$

Differentiation under the integral and one integration by parts give

$$G'_t(\xi) = -2\pi i \int_{\mathbb{R}} x e^{-\pi t^2 x^2} e^{-2\pi i x \xi} dx = -\frac{2\pi \xi}{t^2} G_t(\xi).$$

Since $G_t(0) = t^{-1}$ by the Gaussian integral, the differential equation yields $G_t(\xi) = t^{-1} e^{-\pi \xi^2 / t^2}$. Tensorizing over the n coordinates gives

$$e^{-\pi t^2 |z|^2} = t^{-n} \int_{\mathbb{R}^n} e^{-\pi |\xi|^2 / t^2} e^{2\pi i z \cdot \xi} d\xi.$$

The right-hand side is absolutely integrable. Hence Fubini against the finite measure $\nu \times \nu$, followed by the definition of $\widehat{\nu}$, gives

$$\begin{aligned} \iint e^{-\pi t^2 |x-y|^2} d\nu(x) d\nu(y) &= t^{-n} \int_{\mathbb{R}^n} e^{-\pi |\xi|^2 / t^2} \left(\int e^{2\pi i x \cdot \xi} d\nu(x) \right) \left(\int e^{-2\pi i y \cdot \xi} d\nu(y) \right) d\xi \\ &= t^{-n} \int_{\mathbb{R}^n} e^{-\pi |\xi|^2 / t^2} |\widehat{\nu}(\xi)|^2 d\xi. \end{aligned}$$

Tonelli applies once more because the integrand is nonnegative. For fixed $\xi \neq 0$, the substitution $s = |\xi|/t$ evaluates the t -integral as $c_n |\xi|^{-(n-1)}$. This proves (6.3) directly, with no smoothing or monotonicity assertion required. $\square$

Lemma 6.4 (Marstrand projection lower bound, proved in the required form). *Let $n \geq 2$, let ν be as in Lemma 6.3, and put $K = \text{supp } \nu$. Let $\Omega \subset S^{n-1}$ be any spherical cap of positive surface measure. Then there is $\theta \in \Omega$ such that*

$$|\pi_\theta(K)| \geq c_{n,\Omega} I_1(\nu)^{-1}, \quad \pi_\theta(x) = \theta \cdot x. \quad (6.4)$$

Here $|\cdot|$ denotes one-dimensional Lebesgue measure.

Proof. First suppose that ν has a smooth density. The projected measure $(\pi_\theta)_* \nu$ has a density f_θ , and one-dimensional Plancherel gives

$$\|f_\theta\|_2^2 = \int_{\mathbb{R}} |\widehat{\nu}(t\theta)|^2 dt.$$

Integrating over the whole sphere and using polar coordinates yields

$$\int_{S^{n-1}} \|f_\theta\|_2^2 d\theta \leq C_n \int_{\mathbb{R}^n} |\widehat{\nu}(\xi)|^2 |\xi|^{-(n-1)} d\xi \leq C_n I_1(\nu)$$

by Lemma 6.3. Hence the same upper bound holds with the sphere replaced by Ω . Since $\int f_\theta = 1$, Cauchy–Schwarz on the support of f_θ gives

$$1 \leq |\pi_\theta(K)|^{1/2} \|f_\theta\|_2, \quad |\pi_\theta(K)| \geq \|f_\theta\|_2^{-2}.$$

Therefore

$$\int_\Omega |\pi_\theta(K)| d\theta \geq \frac{|\Omega|^2}{\int_\Omega \|f_\theta\|_2^2 d\theta} \geq c_{n,\Omega} I_1(\nu)^{-1},$$

where we used Cauchy–Schwarz in the first step. Some $\theta \in \Omega$ satisfies (6.4).

For a general ν , choose once and for all a spherical cap $\Omega' \Subset \Omega$ of positive surface measure. Fix a nonnegative C_c^∞ probability density ϕ supported in the unit ball and put $\phi_\eta(x) = \eta^{-n} \phi(x/\eta)$ and $\nu_\eta = \nu * \phi_\eta$. Then $\text{supp } \nu_\eta \subset K^{(\eta)}$ and $\widehat{\nu}_\eta = \widehat{\nu} \widehat{\phi}_\eta$. Since $|\widehat{\phi}_\eta| \leq 1$, Lemma 6.3 gives $I_1(\nu_\eta) \leq I_1(\nu)$. Apply the smooth argument on the fixed cap Ω' and choose $\theta_\eta \in \Omega'$ with

$$|\pi_{\theta_\eta}(K^{(\eta)})| \geq c_{n,\Omega'} I_1(\nu)^{-1}.$$

By compactness of $\overline{\Omega'}$, pass to a sequence $\eta_j \downarrow 0$ for which $\theta_{\eta_j} \rightarrow \theta \in \Omega$. Because K is compact, the projected sets $\pi_{\theta_{\eta_j}}(K^{(\eta_j)})$ lie in an $o(1)$ -neighborhood of $\pi_\theta(K)$. Outer regularity of Lebesgue measure gives (6.4), after renaming the positive constant $c_{n,\Omega'}$ as $c_{n,\Omega}$. $\square$

Lemma 6.5 (Frostman product measure has finite one-energy). *Let μ be a probability measure on $[-C, C]$ satisfying*

$$\mu(I) \leq C_1 |I|^\kappa \tag{6.5}$$

for every interval I . If $n\kappa > 1$, then

$$I_1(\mu^{\otimes n}) \leq C'(n, \kappa, C_1, C).$$

Proof. In the sup norm on $\mathbb{R}^n$, a ball of radius r is contained in a product of n intervals of length $2r$, so

$$\mu^{\otimes n}(B(x, r)) \leq C_2 r^{n\kappa}.$$

Euclidean and sup norms are comparable in fixed dimension. For each x , decompose the y -integral into dyadic shells $2^{-j-1} < |x - y| \leq 2^{-j}$ (and one bounded outer shell). The contribution of the j th small shell is at most $C 2^j 2^{-jn\kappa} = C 2^{-j(n\kappa-1)}$, which is summable because $n\kappa > 1$. Integrating over x proves the claim. $\square$

We next prove the elimination device which turns a positive-measure linear projection of a Cartesian power into a positive-measure homogeneous ring expression.

Lemma 6.6 (Blichfeldt principle). *Let $V \subset \mathbb{R}^n$ be measurable and let $L : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be invertible. There is a translation x such that*

$$\#(V \cap (L\mathbb{Z}^n + x)) \geq |\det L|^{-1} |V|.$$

Proof. Write $x = Ly$. For $y \in [0, 1]^n$,

$$\#(L^{-1}V \cap (\mathbb{Z}^n + y)) = \sum_{k \in \mathbb{Z}^n} 1_{L^{-1}V-k}(y).$$

Integrating over $[0, 1]^n$ and interchanging sum and integral gives $|L^{-1}V| = |\det L|^{-1} |V|$. Some y is at least the average. $\square$

Lemma 6.7 (Coefficient elimination). *Let $A \subset \mathbb{R}$ be compact with diameter $D > 0$. Let $v = (v_1, \dots, v_n) \in [1/2, 1]^n$, $n \geq 2$, and put*

$$\lambda = |v_1 A + \dots + v_n A| > 0.$$

There is an integer $N = N(n, D, \lambda)$ such that, after permuting the coordinates if necessary and putting

$$A_1 = N A^{(2)} - N A^{(2)},$$

one has

$$|v_1 A_1 + \dots + v_{n-1} A_1| \geq D\lambda. \quad (6.6)$$

Moreover N is bounded uniformly when (n, D, λ) ranges in a compact subset of $\{2, 3, \dots\} \times (0, \infty)^2$.

Proof. Let $\pi(x) = v \cdot x$. Choose an orthogonal map U sending $e_n^\perp$ onto $\ker \pi$. For $R \geq 1$ define the slab

$$V_R = U(B_{n-1}(0, R) \times [-1, 1]).$$

Its volume is $2\omega_{n-1}R^{n-1}$, while $\pi(V_R)$ lies in a fixed interval of length $O_n(1)$ because π vanishes on the first $n-1$ directions of the displayed product. Apply Lemma 6.6 to the lattice $(2D)\mathbb{Z}^n$. By choosing R large enough in terms of n, D, λ , we obtain points

$$p_j = 2D\tau_j + z \in V_R, \quad \tau_j \in \mathbb{Z}^n,$$

for $1 \leq j \leq M$, where

$$M\lambda > nD + O_n(1). \quad (6.7)$$

The translates $A^n + p_j$ lie in $A^n + V_R$. Each projected set $\pi(A^n + p_j)$ has measure λ , whereas all of them lie in one interval of length at most $nD + O_n(1)$. By (6.7), two projections overlap. Thus for some $i \neq j$ there are $x, y \in A^n$ with

$$\pi(x) = \pi(y + 2D\ell), \quad \ell = \tau_j - \tau_i \neq 0.$$

Permute coordinates so that $\ell_n \neq 0$, and set $z = y + 2D\ell$. Then

$$v_n(z_n - x_n) = \sum_{r=1}^{n-1} v_r(x_r - z_r). \quad (6.8)$$

Because $|y_n - x_n| \leq D$ and $\ell_n \neq 0$,

$$|z_n - x_n| \geq D. \quad (6.9)$$

Multiplying the set $v_1 A + \dots + v_n A$ by $z_n - x_n$ and using (6.8) gives

$$\begin{aligned} & (z_n - x_n)(v_1 A + \dots + v_n A) \\ & \subset \sum_{r=1}^{n-1} v_r (z_n A - x_n A + x_r A - z_r A). \end{aligned}$$

Now $x_r, x_n, y_r, y_n \in A$, and $z_r = y_r + 2D\ell_r$. Since compactness of A implies $D = \max A - \min A \in A - A$, every coefficient $2D\ell_r$ is a sum or difference of $2|\ell_r|$ elements of $A - A$. Consequently every set $z_n A - x_n A + x_r A - z_r A$ is contained in $NA^{(2)} - NA^{(2)}$ once

$$N \geq 4 + 4 \max_r |\ell_r|.$$

The vector $2D\ell = p_j - p_i$ lies in $V_R - V_R$, hence $|\ell_r| \leq C_n \sqrt{R^2 + 1}/D$. Thus one may choose N depending only on n, D, λ . Taking Lebesgue measure and using (6.9) proves (6.6). The construction of R and the resulting bound for N are continuous in positive lower and upper bounds for D and λ , proving the uniformity statement. $\square$

Proposition 6.8 (Positive-measure homogeneous ring expansion). *Fix $0 < \kappa < 1$ and $C_1, C \geq 1$. There are integers $d, s \geq 1$ and a constant $c > 0$, depending only on (κ, C_1, C) , with the following property. If μ is a probability measure supported on $[-C, C]$ satisfying (6.5), and $K = \text{supp } \mu$, then*

$$|sK^{(d)} - sK^{(d)}| \geq c. \quad (6.10)$$

Proof. Let n be the least integer for which $n\kappa > 1$. By Lemma 6.5, $I_1(\mu^{\otimes n}) \leq C_2$. Choose a small spherical cap around the vector $(1, \dots, 1)/\sqrt{n}$. Lemma 6.4 gives a direction in that cap for which the projection of K^n has measure at least $c_2 > 0$. Multiplying all projection coefficients by one common positive scalar, we may write the resulting set as

$$v_1 K + \dots + v_n K, \quad v_j \in [1/2, 1],$$

with measure at least $c_3 > 0$.

The Frostman condition also gives a uniform lower bound for $\text{diam } K$: choose $r = (4C_1)^{-1/\kappa}$; no interval of length $2r$ can carry all the mass. Thus $\text{diam } K \geq r$. The diameter is of course at most $2C$. Apply Lemma 6.7 to remove one coefficient. The resulting set K_1 is a bounded homogeneous signed sum of degree-two monomials in K and satisfies a positive projection-measure lower bound depending only on the fixed parameters. Its diameter is bounded above because K is bounded, and bounded below because a subset of a bounded linear sum with positive Lebesgue measure must itself have positive diameter. We may therefore apply Lemma 6.7 again. After $n - 1$ eliminations only one coefficient remains and we have a set of positive measure.

It remains only to check homogeneity. If $K_j = s_j K^{(d_j)} - s_j K^{(d_j)}$, then every product of two elements of K_j expands, by distributivity, as a signed sum of $O(s_j^2)$ monomials of degree $2d_j$ in K . Hence $NK_j^{(2)} - NK_j^{(2)}$ has the same form $s_{j+1} K^{(2d_j)} - s_{j+1} K^{(2d_j)}$ with bounded s_{j+1} . Starting with $d_0 = 1$ shows that after $n - 1$ steps the final degree is $d = 2^{n-1}$ and the number of summands is bounded in terms of the fixed parameters. Absorbing the final nonzero scalar coefficient into the lower bound for Lebesgue measure proves (6.10). $\square$

We now turn positive Lebesgue measure into a quantitative approximate interval. The finite Fourier argument is included because the size of the Bohr set is the source of the power separation between segment length and approximation error.

Lemma 6.9 (Bogolyubov lemma in a cyclic group). *Let $Q \geq 2$ and let $S \subset \mathbb{Z}/Q\mathbb{Z}$ have density $\alpha = |S|/Q > 0$. There is a set of frequencies $\Gamma \subset \mathbb{Z}/Q\mathbb{Z}$ with $|\Gamma| \leq 16\alpha^{-2}$ such that the Bohr set*

$$B(\Gamma, 1/6) = \left\{ x : \left\| \frac{rx}{Q} \right\|_{\mathbb{R}/\mathbb{Z}} \leq \frac{1}{6} \text{ for all } r \in \Gamma \right\}$$

is contained in $2S - 2S$.

Proof. Write $f = 1_S$ and use normalized Fourier transform

$$\widehat{f}(r) = \frac{1}{Q} \sum_x f(x) e^{-2\pi i r x / Q}.$$

Then Parseval gives $\sum_r |\widehat{f}(r)|^2 = \alpha$ and $\widehat{f}(0) = \alpha$. Put

$$\Gamma = \{r : |\widehat{f}(r)| \geq \alpha^{3/2}/4\}.$$

Parseval gives $|\Gamma| \leq 16\alpha^{-2}$. The normalized fourfold convolution $g = f * f * \widetilde{f} * \widetilde{f}$, where $\widetilde{f}(x) = f(-x)$, satisfies

$$g(x) = \sum_r |\widehat{f}(r)|^4 e^{2\pi i r x / Q}.$$

For $r \notin \Gamma$,

$$\sum_{r \notin \Gamma} |\widehat{f}(r)|^4 \leq \frac{\alpha^3}{16} \sum_r |\widehat{f}(r)|^2 = \frac{\alpha^4}{16}.$$

On the other hand the total fourth moment is at least the contribution of $r = 0$, namely α^4 . If $x \in B(\Gamma, 1/6)$, every phase on Γ has real part at least $1/2$. Therefore

$$\Re g(x) \geq \frac{1}{2} \left(\alpha^4 - \frac{\alpha^4}{16} \right) - \frac{\alpha^4}{16} > 0.$$

Thus $g(x) > 0$, which is equivalent to $x \in 2S - 2S$. $\square$

Lemma 6.10 (A cyclic Bohr set contains a long one-dimensional progression). *Let $\Gamma \subset \mathbb{Z}/Q\mathbb{Z}$ have cardinality $D \geq 1$. For all sufficiently large Q there are integers q, L such that*

$$1 \leq q \leq Q^{1/3}, \quad L \geq c_D Q^{1/(3D)}, \quad Lq \leq Q/10,$$

and

$$\{jq : |j| \leq L\} \pmod{Q} \subset B(\Gamma, 1/6).$$

If $D = 0$, take $q = 1$ and $L = \lfloor Q/10 \rfloor$.

Proof. Assume $D \geq 1$ and enumerate $\Gamma = \{r_1, \dots, r_D\}$. Let $T = \lfloor Q^{1/(3D)} \rfloor$. Consider the $T^D + 1$ points

$$k \left(\frac{r_1}{Q}, \dots, \frac{r_D}{Q} \right) \pmod{1}, \quad 0 \leq k \leq T^D,$$

in the D -torus, partitioned into T^D boxes of side $1/T$. Two lie in the same box. Their difference supplies $1 \leq q \leq T^D \leq Q^{1/3}$ with

$$\left\| \frac{qr_j}{Q} \right\|_{\mathbb{R}/\mathbb{Z}} \leq T^{-1} \quad (1 \leq j \leq D).$$

Take $L = \lfloor T/7 \rfloor$. Then $|j| \leq L$ implies the Bohr inequalities with radius $1/6$. Finally $Lq \leq CQ^{1/3+1/(3D)} \leq CQ^{2/3} < Q/10$ for Q large. $\square$

Theorem 6.11 (Internal homogeneous ring saturation). *Fix $0 < \kappa < 1$ and $C_0 \geq 1$. There exist integers $d, s \geq 1$ and a number $\beta > 0$, depending only on (κ, C_0) , such that the following holds for all sufficiently small h . If $F \subset [-2, 2]$ is (h, κ, C_0) -Frostman, then there is a number ℓ with*

$$h^{1-\beta} \leq \ell \leq C \tag{6.11}$$

such that

$$[-\ell, \ell] \subset sF^{(d)} - sF^{(d)} + [-\ell h^\beta, \ell h^\beta]. \tag{6.12}$$

Every monomial in (6.12) has the same degree d .

Proof. Around each $a \in F$ place an interval of radius $h/10$, and let μ be the probability measure which gives each such interval mass $1/|F|$ with constant density on that interval. Because F is h -separated, overlap is uniformly bounded. For an interval I of length $r \geq h$, (6.1) gives $\mu(I) \leq Cr^\kappa$. For $r < h$, at most a bounded number of the small component intervals meet I , and

$$\mu(I) \leq C \frac{r/h}{|F|} \leq CC_0 r h^{\kappa-1} = CC_0 r^\kappa (r/h)^{1-\kappa} \leq CC_0 r^\kappa.$$

Thus μ is a genuine κ -Frostman measure with a constant depending only on (κ, C_0) , supported on a fixed bounded interval. Proposition 6.8 therefore supplies fixed d, s_0 and $c > 0$ such that, writing $K = \text{supp } \mu$,

$$E = s_0 K^{(d)} - s_0 K^{(d)}$$

has Lebesgue measure at least c and lies in a fixed interval $[-M, M]$.

Choose an integer $Q \asymp h^{-1}$ so large that reduction modulo Q is injective on the integer interval corresponding to $[-8M/h, 8M/h]$. Let $S \subset \mathbb{Z}/Q\mathbb{Z}$ be the set of h -grid cells which meet E . Since $|E| \geq c$, after the harmless change from h to a comparable grid spacing we have $|S|/Q \geq \alpha > 0$, where α depends only on the fixed parameters. Lemma 6.9 gives $B(\Gamma, 1/6) \subset 2S - 2S$ with $D = |\Gamma| \leq D_0 = 16\alpha^{-2}$. Lemma 6.10 gives q, L such that $jq \in 2S - 2S$ for $|j| \leq L$, $L \geq cQ^{1/(3D_0)}$, and $Lq \leq Q/10$. Because there is no modular wrap in this range, these are honest integer identities.

Each occupied cell has a representative point of E within $O(h)$ of its grid point. Consequently every jqh , $|j| \leq L$, lies within $O(h)$ of $2E - 2E$. Hence every point of the interval

$$[-\ell, \ell], \quad \ell = Lqh,$$

lies within $O(qh)$ of $2E - 2E$. Since $\ell/(qh) = L$, the relative error is at most $CL^{-1} \leq Ch^{1/(3D_0)}$. Also

$$\ell = Lqh \geq Lh \geq ch^{1-1/(3D_0)}.$$

Choose once and for all

$$0 < \beta < \frac{1}{3D_0},$$

for instance $\beta = 1/(4D_0)$. For small h , constants are absorbed and the preceding two displays imply (6.11) and a relative error at most $\frac{1}{2}h^\beta$.

Finally replace each element of K occurring in an expression by a nearest point of F . All variables remain in a fixed bounded interval, and the number and degree of monomials are fixed, so the resulting change is $O(h)$. Since $h/\ell \leq Ch^{1/(3D_0)}$, this replacement is also absorbed by ℓh^β . Moreover $2E - 2E$ expands as

$$4s_0 K^{(d)} - 4s_0 K^{(d)}.$$

Thus (6.12) holds with $s = 4s_0$, and homogeneity is preserved. $\square$

Remark 6.12. The scalar sum-product chapter proves an explicit one-step discretized sum-product gain. Theorem 6.11 is not a replacement for that theorem; it closes a different logical gap. A one-step gain does not, by itself, yield a bounded-complexity segment. The argument above owns the additional saturation step needed by the adjoint transfer.

6.3 From a torus tube to a regular parameter set

Let T be a maximal torus. On a sufficiently short arc we use arclength coordinate $\theta \mapsto t(\theta)$, normalized so that on the root plane $V = \mathfrak{t}^\perp \simeq \mathbb{C}$,

$$\text{Ad}(t(\theta))|_V = e^{i\omega\theta} \tag{6.13}$$

for one fixed nonzero real ω (the precise normalization is irrelevant).

Lemma 6.13 (Pruning a torus tube). *Let $0 < \delta \leq \eta \ll 1$ and let B be a δ -separated subset of $T^{(\eta)}$ lying in the fixed chart. There is a subset $B' \subset B$ and a nearest-point projection $p : B' \rightarrow T$ such that*

- (i) $p(B')$ is $c\eta$ -separated;

$$(ii) |B'| \geq c(\delta/\eta)^3 |B|;$$

$$(iii) d(b, p(b)) \leq \eta \text{ for } b \in B'.$$

Proof. Choose a maximal 20η -separated subset $B' \subset B$. Every ball of radius 20η in the three-dimensional group contains at most $C(\eta/\delta)^3$ points of the δ -separated set B , so maximality gives (ii). If $b_1, b_2 \in B'$ and their nearest torus projections were closer than 10η , then $d(b_1, b_2) < 12\eta$ after changing harmless metric constants, contradicting the chosen separation. Thus the projections are $c\eta$ -separated. The tube condition gives (iii). $\square$

Lemma 6.14 (Minimal-scale Frostman regularization). *Fix $0 < \kappa_0 < 1$. Let S be an η -separated subset of an interval I_0 of length R_0 , and suppose*

$$|S| \geq c_0(R_0/\eta)^{\kappa_0}. \quad (6.14)$$

There are an interval $I \subset I_0$ of length $R \geq \eta$, a subset $F_0 = S \cap I$, and $\kappa = \kappa_0/2$ such that, after the affine rescaling $I \rightarrow [-1, 1]$ and writing $h = \eta/R$,

$$(i) h \leq C(\eta/R_0)^{c(\kappa_0)};$$

$$(ii) \text{ the rescaled set } F \text{ is } Ch\text{-separated and } |F| \geq ch^{-\kappa};$$

$$(iii) \text{ for every interval } J \text{ of length at least } Ch, |F \cap J| \leq C|J|^\kappa |F|;$$

$$(iv) \text{ diam } F \geq c.$$

All constants depend only on (κ_0, c_0) .

Proof. For simplicity use dyadic subintervals of I_0 ; changing constants handles arbitrary intervals. Choose $R \geq \eta$ minimal for which some dyadic interval I of length R satisfies

$$|S \cap I| \geq c_1(R/R_0)^\kappa |S|, \quad \kappa = \kappa_0/2, \quad (6.15)$$

where c_1 is a fixed small dyadic constant. Such an interval exists at scale R_0 . For any dyadic $J \subset I$ of length r with $\eta \leq r < R$, minimality gives

$$|S \cap J| < C(r/R_0)^\kappa |S|.$$

Divide by (6.15) to obtain

$$|F_0 \cap J| \leq C(r/R)^\kappa |F_0|.$$

Choose the dyadic scale r_d with $r \leq r_d < 2r$. Any interval of length r meets at most two adjacent dyadic cells of length r_d , so it is contained in their union. Applying the preceding estimate to those two cells gives (iii), with only an absolute factor, after rescaling.

For (i), packing gives $|F_0| \leq CR/\eta$. Combining this with (6.15) and (6.14) gives

$$\frac{R}{\eta} \geq c(R/R_0)^\kappa (R_0/\eta)^{\kappa_0}.$$

Writing $x = R/R_0$ and $H = \eta/R_0$ yields $x^{1-\kappa} \geq cH^{1-\kappa_0}$. Hence

$$h = \frac{\eta}{R} = \frac{H}{x} \leq CH^{\frac{\kappa_0 - \kappa}{1 - \kappa}} = CH^{\frac{\kappa_0}{2 - \kappa_0}},$$

which is (i). The same inequality, read in the other direction, gives $|F_0| \geq c(R/\eta)^\kappa$, proving (ii). Finally, if the rescaled F lay in an interval of sufficiently small fixed length c , then (iii) applied to that interval would give $|F| \leq Cc^\kappa |F| < |F|$, a contradiction. This proves (iv). $\square$

Lemma 6.15 (Short-arc preparation with Taylor room). *Fix $0 < \kappa_* < 1$. Let S be an η -separated subset of a fixed torus chart of length 1 with*

$$|S| \geq \eta^{-\kappa_*}.$$

Let $\beta_* > 0$ be any fixed target exponent. There exist numbers $\kappa \in (0, 1)$ and $c_* > 0$, depending only on (κ_*, β_*) , with the following property. For all sufficiently small η , one can find an interval I of length R and a subset $F_0 = S \cap I$ such that, after rescaling I affinely to $[-1, 1]$ and putting $h = \eta/R$,

(i) *F is (Ch, κ, C) -Frostman;*

(ii) *$h \leq \eta^{c_*}$;*

(iii) *after decreasing β_* if necessary by a constant factor depending only on κ_* ,*

$$R + h \leq h^{\beta_*/2}. \quad (6.16)$$

The same conclusion holds if the initial chart has a fixed bounded length rather than 1.

Proof. Choose

$$a = \frac{\kappa_*}{4}$$

and partition the chart into $O(\eta^{-a})$ intervals of length $R_0 \asymp \eta^a$. One interval contains a subset S_0 with

$$|S_0| \geq c\eta^{-\kappa_*+a}.$$

Writing $H = \eta/R_0 \asymp \eta^{1-a}$, this becomes

$$|S_0| \geq cH^{-\kappa_0}, \quad \kappa_0 = \frac{\kappa_* - a}{1 - a} > 0.$$

Apply Lemma 6.14 inside that interval. It gives a subinterval of length $R \leq R_0$, relative mesh $h = \eta/R$, and a Frostman set with exponent $\kappa = \kappa_0/2$. It also gives $h \leq CH^{c(\kappa_0)}$, proving (ii) after changing the exponent c_* .

Since $R \leq R_0 = \eta^a$ and $R \leq R_0$ also implies $h = \eta/R \geq \eta/R_0 \asymp \eta^{1-a}$, we have

$$R \leq C\eta^a \leq h^{\beta_*/2}$$

provided $\beta_* \leq 2a/(1-a)$; we are always free to decrease a positive exponent in a power-saving conclusion. Also $h \leq h^{\beta_*/2}$ for $0 < h < 1$ once $\beta_* < 2$. Absorbing the fixed factor 2 by one further harmless decrease of β_* proves (6.16). $\square$

Lemma 6.16 (Rich-torus count gives a fixed one-dimensional exponent). *Fix $0 < \sigma_- < 3$. Suppose at a base scale Δ one has*

$$n = \mathcal{N}(A, \Delta) \geq \Delta^{-\sigma_- + \epsilon}, \quad \rho = \Delta^\nu,$$

and the rich-torus chapter supplies a maximal torus T and a Δ -separated set $B \subset A^{O(1)} \cap T^{(\eta)}$ satisfying

$$|B| \geq \rho^{C_1} n^{1/3}, \quad c\Delta\rho^{-C_2} \leq \eta \leq C\Delta\rho^{-C_2}.$$

(The tube width is therefore taken at the natural Larsen–Pink output scale, rather than being allowed to shrink arbitrarily.) If $\epsilon, \nu > 0$ are sufficiently small in terms of (σ_-, C_1, C_2) , then after Lemma 6.13 the projected torus set contains an η -separated subset S with

$$|S| \geq \eta^{-\kappa_*}, \quad \kappa_* = \frac{\sigma_-}{4}. \quad (6.17)$$

Proof. Lemma 6.13, with $\delta = \Delta$, loses at most a factor $C(\eta/\Delta)^3 \leq \rho^{-3C_2}$. Hence

$$|S| \geq c\rho^{C_1+3C_2}n^{1/3} \geq c\Delta^{-\sigma_-/3+\epsilon/3+(C_1+3C_2)\nu}.$$

On the other hand the two-sided tube-width comparison gives

$$\eta^{-\sigma_-/4} \leq C\Delta^{-\frac{\sigma_-}{4}(1-C_2\nu)}.$$

The gap between $\sigma_-/3$ and $\sigma_-/4$ is $\sigma_-/12$, so choosing ϵ and ν small enough makes the exponent in the first display strictly more negative than the exponent in the second. For small Δ this proves (6.17). $\square$

6.4 Root-plane scalarization and homogeneous adjoint words

Lemma 6.17 (Rank-one scalarization of torus differences). *Let I be an arc of T of length $R \ll 1$, let $t_* = t(\theta_*) \in I$, and write $t_s = t(\theta_* + Rs)$ for $|s| \leq 2$. Let $b_s, b_* \in G$ satisfy $d(b_s, t_s), d(b_*, t_*) \leq \eta$, and put $h = \eta/R$. Define*

$$\Delta_s = \text{Ad}(b_s) - \text{Ad}(b_*).$$

Then on the root plane V there is a fixed complex scalar c_* with $|c_*| \asymp 1$ such that

$$\Delta_s|_V = Rc_*s \text{Id}_{\mathbb{C}} + O(R(R+h)) \quad (6.18)$$

uniformly for $|s| \leq 2$. On the toral axis, $\Delta_s = O(\eta)$.

Proof. For the exact torus points, (6.13) gives

$$\text{Ad}(t_s) - \text{Ad}(t_*) = e^{i\omega\theta_*}(e^{i\omega Rs} - 1) = R(i\omega e^{i\omega\theta_*})s + O(R^2)$$

on V . The adjoint map is Lipschitz on the fixed compact chart, so replacing t_s, t_* by b_s, b_* changes the operator by $O(\eta) = O(Rh)$. This proves (6.18). Exact torus elements act trivially on the toral axis, so only the $O(\eta)$ perturbation remains there. $\square$

Lemma 6.18 (Homogeneous operator monomials). *Fix $d \geq 1$. Let P_V denote orthogonal projection onto the root plane V and define $\Xi_0 = c_*^d P_V$ on all of $\mathfrak{su}(2)$. Under the hypotheses of Lemma 6.17, for $s_1, \dots, s_d \in [-2, 2]$,*

$$\Delta_{s_1} \cdots \Delta_{s_d} = R^d(s_1 \cdots s_d)\Xi_0 + O_d(R^d(R+h)) \quad (6.19)$$

in full operator norm on $\mathfrak{su}(2)$. Moreover the left side expands as a signed sum of at most 2^d operators $\text{Ad}(w)$, where every w is a word of length d in the letters b_{s_j}, b_* .

Proof. On V , write $\Delta_{s_j} = Rc_*s_jI + E_j$ with $\|E_j\| \leq CR(R+h)$. Expanding the product gives the asserted main term and an $O_d(R^d(R+h))$ error there. On the toral axis, the first factor has size $O(\eta) = O(Rh)$, while every remaining factor has full operator norm $O(R+\eta) = O(R)$. Hence the product maps the toral axis to a vector of norm $O_d(R^dh)$, which is again contained in the stated error. The same estimates control the off-diagonal blocks created by the $O(\eta)$ perturbations, so (6.19) holds in full operator norm.

For the word expansion, distribute

$$(\text{Ad}(b_{s_1}) - \text{Ad}(b_*)) \cdots (\text{Ad}(b_{s_d}) - \text{Ad}(b_*)).$$

Each of its 2^d terms is signed and, by multiplicativity of the adjoint representation, is $\text{Ad}(w)$ for the corresponding product word w . $\square$

Proposition 6.19 (Operator segment from a regularized rich torus). *Fix $0 < \kappa < 1$. Let $F \subset [-2, 2]$ be an (h, κ, C_0) -Frostman parameter set. Assume that there are elements $b_s \in G$ for $s \in F$ and $b_* \in G$ satisfying the hypotheses of Lemma 6.17 with the same parameters R, h and with $t_s = t(\theta_* + Rs)$. Assume in addition that*

$$R + h \leq h^{\beta/2}, \quad (6.20)$$

where β is the exponent in Theorem 6.11 (we may decrease β once and for all). Then there exist bounded integers d, s, m such that, for some L satisfying

$$L \geq cR^d h^{1-\beta}, \quad (6.21)$$

for every $0 \leq t \leq L$ there is an operator

$$Q_t = \sum_{j=1}^m \epsilon_j \operatorname{Ad}(w_j), \quad \epsilon_j \in \{\pm 1\},$$

with all w_j bounded words in the b_s and there is a fixed rank-two operator Ξ on all of $\mathfrak{su}(2)$, with kernel equal to the toral axis and $\|\Xi\| = 1$, such that

$$\|Q_t - t\Xi\| \leq Lh^{\beta/3} \quad (0 \leq t \leq L). \quad (6.22)$$

Proof. Apply Theorem 6.11. It gives d, s and $\ell \geq h^{1-\beta}$ such that every $z \in [-\ell, \ell]$ is within ℓh^β of a signed sum of s monomials of degree d in parameters from F . Replace each scalar monomial by the corresponding operator monomial in Lemma 6.18 and then expand each operator monomial into signed adjoint words. All main terms have the same factor $R^d \Xi_0$, precisely because the scalar theorem is homogeneous. Thus, after dividing the direction by the fixed number $\|\Xi_0\| \asymp 1$ and absorbing that constant into the endpoint, they approximate a segment in the full-space rank-two direction $\Xi = \Xi_0 / \|\Xi_0\|$, whose kernel is the toral axis, of length $L \asymp R^d \ell$.

The scalar approximation contributes $O(Lh^\beta)$. The finitely many Taylor/off-torus errors contribute $O(R^d \ell(R + h))$, which is $O(Lh^{\beta/2})$ by (6.20). Decreasing the displayed exponent to $\beta/3$ absorbs fixed constants. The number $m \leq 2^d \cdot O(s)$ and all word lengths depend only on the fixed Frostman parameters. $\square$

6.5 Signed adjoint sums are actual group words

Lemma 6.20 (BCH realization of a signed adjoint sum). *Fix m . There are $u_0, C_m > 0$ such that the following holds. Let $X \in \mathfrak{su}(2)$ be unit, $0 < u < u_0$, and put $g = \exp(uX)$. For $w_1, \dots, w_m \in U^{O(1)}$ and signs $\epsilon_j \in \{\pm 1\}$,*

$$W = \prod_{j=1}^m w_j g^{\epsilon_j} w_j^{-1}$$

satisfies

$$\log W = u \sum_{j=1}^m \epsilon_j \operatorname{Ad}(w_j) X + O_m(u^2). \quad (6.23)$$

In particular, if $Q = \sum \epsilon_j \operatorname{Ad}(w_j)$ and Z is a Lie-algebra vector with $\|QX - Z\| \leq e$, then

$$d(W, \exp(uZ)) \leq C_m(ue + u^2). \quad (6.24)$$

Proof. Conjugation commutes with the exponential map, so

$$w_j g^{\epsilon_j} w_j^{-1} = \exp(\epsilon_j u \operatorname{Ad}(w_j) X).$$

The BCH formula on the fixed chart gives

$$\log(e^{Z_1} \cdots e^{Z_m}) = Z_1 + \cdots + Z_m + O_m \left(\left(\max_j \|Z_j\| \right)^2 \right)$$

when the Z_j are small. Taking $Z_j = \epsilon_j u \operatorname{Ad}(w_j) X$ proves (6.23). The exponential map is Lipschitz on the chart, giving (6.24). $\square$

Lemma 6.21 (Normalizer and torus agree locally). *After shrinking the fixed identity chart U once and for all, for every maximal torus $T < SU(2)$ and every $g \in U$ one has*

$$d(g, N(T)) = d(g, T). \quad (6.25)$$

Proof. All maximal tori are conjugate, and the metric is bi-invariant. For one fixed torus T_0 , the nonidentity Weyl component $N(T_0) \setminus T_0$ is compact and does not contain the identity, hence has a positive distance d_0 from 1. The same d_0 works for every conjugate torus. Choose $U \subset B(1, d_0/4)$. Since $1 \in T$, every $g \in U$ has $d(g, T) < d_0/4$, whereas every point of $N(T) \setminus T$ is at distance greater than $d_0/2$ from g . Thus the nearest component of the normalizer is T itself. $\square$

Lemma 6.22 (Moving a vector off a toral kernel). *Let T have unit axis n and let $N(T)$ be its normalizer. Suppose A contains 1 and contains an element a with $d(a, N(T)) \geq \rho$. Then for every unit $X \in \mathfrak{su}(2)$ there is $x \in \{1, a\}$ such that*

$$\operatorname{dist}(\operatorname{Ad}(x)X, \mathbb{R}n) \geq c\rho. \quad (6.26)$$

Proof. If $\operatorname{dist}(X, \mathbb{R}n) \geq c\rho$, take $x = 1$. Otherwise X is within $c\rho$ of one of $\pm n$. The stabilizer in $SU(2)$ of the unoriented axis $\mathbb{R}n$ under the adjoint action is exactly $N(T)$. On the fixed compact group, the smooth map $a \mapsto \operatorname{dist}(\operatorname{Ad}(a)n, \mathbb{R}n)$ is therefore quantitatively comparable, in a sufficiently small neighborhood of $N(T)$, to $d(a, N(T))$; away from that neighborhood it has a fixed positive lower bound. Thus $\operatorname{dist}(\operatorname{Ad}(a)n, \mathbb{R}n) \geq c'\rho$. The $O(\rho)$ perturbation from n to X is absorbed by taking the first constant c sufficiently small. $\square$

Proposition 6.23 (Rich torus plus a small element gives a Lie segment). *Fix the Frostman/richness exponents appearing above. Then there are constants $b_0 > 0$, $q_0 > 0$ and a bounded word exponent C with the following property. Let $\Delta > 0$ be small and suppose a bounded product set of a symmetric $A \subset U$, $1 \in A$, contains:*

- (i) *a torus-tube set for which Proposition 6.19 gives an operator segment of length $L = \Delta^\alpha$ and relative error at most Δ^{b_0} ;*
- (ii) *an element $g = \exp(uX) \in A^C$ with $u = \Delta^\gamma$;*
- (iii) *an element of a bounded power of A at distance at least Δ^{q_0} from the normalizer $N(T)$ of the rich torus;*
- (iv) *$q_0 < b_0/4$ and $\gamma \geq \alpha + b_0$.*

Then there is a unit vector ξ such that

$$\{\exp(t\xi) : 0 \leq t \leq \Delta^{\alpha+\gamma+q_0}\} \subset (A^C)^{(C\Delta^{\alpha+\gamma+b_0})}. \quad (6.27)$$

In particular the approximation error is smaller than the guaranteed segment length by the fixed relative factor $\Delta^{b_0-q_0}$.

Proof. Let Ξ be the full-space rank-two operator in Proposition 6.19. Its kernel is the toral axis. By Lemma 6.22, after conjugating g by one bounded word we may replace X by a unit vector X' satisfying

$$\lambda := \|\Xi X'\| \geq c\Delta^{q_0}.$$

Put $\xi = \Xi X'/\lambda$. For $0 \leq t \leq L$, Proposition 6.19 gives a signed adjoint sum Q_t with

$$\|Q_t X' - t\lambda\xi\| \leq CL\Delta^{b_0}.$$

Lemma 6.20 therefore constructs a bounded group word within

$$C(uL\Delta^{b_0} + u^2)$$

of $\exp(ut\lambda\xi)$. As t runs over $[0, L]$, the target parameter covers $[0, u\lambda L]$. In particular,

$$[0, u\lambda L] \supset [0, c\Delta^{\alpha+\gamma+q_0}].$$

After shrinking the endpoint by the fixed factor c , this gives the segment claimed in (6.27).

For the error, the first term is $\Delta^{\alpha+\gamma+b_0}$. The hypothesis $\gamma \geq \alpha + b_0$ gives $u^2 = \Delta^{2\gamma} \leq \Delta^{\alpha+\gamma+b_0}$. Finally, since $q_0 < b_0$, the error exponent exceeds the guaranteed endpoint exponent $\alpha + \gamma + q_0$ by $b_0 - q_0 > 0$. $\square$

6.6 Scale ladder in $SU(2)$

The preceding proposition is local: it asks for an element at a sufficiently small prescribed scale. We now prove that the scale-invariant hypotheses of the compact product argument manufacture such an element by repeated commutators.

Lemma 6.24 (Arbitrarily fine initial scale increment). *Fix $0 < \sigma < 3$ and a target increment $0 < r < 1/4$. For sufficiently small δ , suppose*

$$\mathcal{N}(A, t) \geq \delta^\varepsilon t^{-\sigma} \quad (\delta \leq t \leq 1) \quad (6.28)$$

with $\varepsilon > 0$ sufficiently small in terms of (σ, r) . Then there are $x, y \in A$ with

$$\delta^r \leq d(x, y) \leq 2\delta^{\sigma r/4}. \quad (6.29)$$

If $a_0 = xy^{-1}$ and

$$\kappa = -\frac{\log d(a_0, 1)}{\log \delta^{-1}},$$

then, after taking δ sufficiently small,

$$\frac{\sigma r}{5} \leq \kappa \leq 2r. \quad (6.30)$$

Proof. At the fine scale δ^r , the lower bound gives

$$\mathcal{N}(A, \delta^r) \geq \delta^{-\sigma r + \varepsilon}.$$

At the coarser scale $\delta^{\sigma r/4}$, three-dimensional packing in the fixed chart gives

$$\mathcal{N}(A, \delta^{\sigma r/4}) \leq C\delta^{-3\sigma r/4}.$$

For $\varepsilon < \sigma r/8$, the fine covering number is larger for small δ . Hence one coarse ball contains two points which remain separated at the fine scale, proving (6.29). Bi-invariance gives the same bounds for $a_0 = xy^{-1}$. Taking logarithms, the fixed factor 2 contributes $o(1)$ to the exponent as $\delta \downarrow 0$; shrinking δ proves (6.30). $\square$

Lemma 6.25 (Rank-one commutator scale multiplication). *There are $r_0, c, C > 0$ such that the following holds. Let $a_0 = \exp(Y)$ and $a = \exp(X)$ with $\|X\|, \|Y\| < r_0$. Let $T_0 = C_G(a_0)$ and assume a unit vector in the direction of $\text{Ad}(x)X$ is at distance at least ρ from $\mathfrak{t}_0$. Then*

$$c\rho \|Y\| \|X\| \leq d([a_0, xax^{-1}], 1) \leq C\|Y\| \|X\| \quad (6.31)$$

provided $\|X\| + \|Y\| \leq c\rho$.

Proof. Put $X' = \text{Ad}(x)X$. BCH gives

$$\log[a_0, e^{X'}] = (\text{Ad}(a_0) - 1)X' + O(\|Y\| \|X\| (\|Y\| + \|X\|)).$$

In $SU(2)$, $\ker(\text{Ad}(a_0) - 1) = \mathfrak{t}_0$. On the orthogonal root plane, $\text{Ad}(a_0)$ is a rotation through angle comparable to $\|Y\|$, so its difference from the identity has singular value comparable to $\|Y\|$. The transversality assumption therefore gives

$$\|(\text{Ad}(a_0) - 1)X'\| \geq c\rho \|Y\| \|X\|.$$

The error term is at most half of this main term under the displayed smallness assumption. The upper bound follows from the same BCH formula. The exponential chart is bi-Lipschitz, so Lie-algebra norm and group distance are comparable. $\square$

Proposition 6.26 (Fine commutator mesh of small-element exponents). *Fix $0 < \sigma < 3$ and $0 < r < 1/4$. Assume (6.28) and suppose A is δ^ε -away from every maximal torus (equivalently, by Lemma 6.21, from every maximal-torus normalizer inside U). If ε is sufficiently small in terms of (σ, r) , then there are a number κ satisfying (6.30), a bounded integer $J = J(\sigma, r)$, and elements $a_1, \dots, a_J \in A^C$ such that, writing*

$$\lambda_j = -\frac{\log d(a_j, 1)}{\log \delta^{-1}}, \quad (6.32)$$

one has

$$|\lambda_j - j\kappa| \leq Cj\varepsilon + o_\delta(1) \quad (1 \leq j \leq J). \quad (6.33)$$

Consequently, after choosing ε and then δ sufficiently small, every interval $J_0 \subset [4r, 1 - 4r]$ of length at least $6r$ contains one of the exponents λ_j . The word exponent C is bounded in terms of (σ, r) only.

Proof. Take $a_1 = a_0$ from Lemma 6.24, so $\lambda_1 = \kappa$. Keep the centralizer torus $T_0 = C_G(a_0)$ fixed throughout the construction. Inductively suppose $a_j = \exp(X_j)$ has been constructed. If the direction of X_j is already at distance at least $c\delta^\varepsilon$ from $\mathfrak{t}_0$, take $x_j = 1$; otherwise choose $x_j \in A$ with $d(x_j, N(T_0)) \geq \delta^\varepsilon$ and apply Lemma 6.22. In either case

$$\text{dist}(\text{Ad}(x_j)X_j, \mathfrak{t}_0) \geq c\delta^\varepsilon \|X_j\|.$$

Set

$$a_{j+1} = [a_0, x_j a_j x_j^{-1}].$$

Because $\kappa \geq \sigma r/5$ while ε is chosen much smaller than σr , the smallness condition in Lemma 6.25 holds uniformly for every bounded j . That lemma gives

$$c\delta^\varepsilon d(a_0, 1) d(a_j, 1) \leq d(a_{j+1}, 1) \leq C d(a_0, 1) d(a_j, 1).$$

Taking logarithms yields

$$\lambda_{j+1} = \kappa + \lambda_j + O(\varepsilon) + o_\delta(1),$$

and induction proves (6.33). We stop before $j\kappa$ reaches $1 + 2r$; since $\kappa \geq \sigma r/5$, this requires only $J = O_\sigma(r^{-1})$ steps, so all word lengths remain bounded.

For the mesh statement, consecutive ideal exponents $j\kappa$ have gap at most $2r$ by (6.30). The accumulated perturbation in (6.33) can be made smaller than $r/4$ throughout the bounded ladder. The sequence starts below $2r + o(1)$ and continues beyond $1 - 2r$ before stopping. Hence any interval inside $[4r, 1 - 4r]$ of length at least $6r$ contains some λ_j . $\square$

6.7 Uniformity across the variable-exponent scale range

The operator endpoint selected by the regularization is allowed to vary with the inherited base scale. The following uniform bounds are enough because Proposition 6.26 provides a fine mesh of small-element scales *after* that endpoint is known.

Lemma 6.27 (Uniform local exponent bounds). *Fix the rich-torus/Frostman parameters and choose $\nu > 0$ sufficiently small. There are constants $A_2 < \infty$ and $b_0 > 0$ such that at every inherited base scale Δ the operator segment produced by Proposition 6.19 may be written $L = \Delta^\alpha$ with*

$$0 \leq \alpha \leq A_2, \quad \|Q_t - t\Xi\| \leq L\Delta^{b_0} \quad (0 \leq t \leq L). \quad (6.34)$$

These constants depend only on the fixed dimension/richness parameters, not on Δ .

Proof. Take the natural tube width $\eta \asymp \Delta\rho^{-C_2} = \Delta^e$, $e = 1 - C_2\nu > 0$. The short-arc regularization gives $\eta \leq R \leq \eta^a$ and $h \leq \eta^{c_*}$ for fixed positive a, c_* . Since Theorem 6.11 gives $L \gg R^d h^{1-\beta}$, we have

$$L \gg \eta^{1-\beta} R^{d-1+\beta} \geq \eta^d \gg \Delta^{2d}$$

for ν small and Δ small; thus one may take, for example, $A_2 = 2d + 1$ after absorbing fixed constants. Proposition 6.19 gives relative error $O(h^{\beta/3})$. Since $h \leq \eta^{c_*} \ll \Delta^{ec_*}$, any fixed $b_0 < ec_*\beta/3$ works after one final decrease to absorb constants. $\square$

Lemma 6.28 (Inheritance of bounded-power growth to an intermediate scale). *Let $\delta \leq \Delta \ll 1$. Suppose, for a finite list of integers m , one has*

$$\mathcal{N}(A^{m+1}, \delta) \leq \delta^{-K_m \varepsilon} \mathcal{N}(A, \delta).$$

Then

$$\mathcal{N}(A^m, \Delta) \leq C\delta^{-K_m \varepsilon} \mathcal{N}(A, \Delta) \quad (6.35)$$

for the same finite list.

Proof. Choose a maximal 2Δ -separated subset $Y \subset A^m$, so $|Y| \asymp \mathcal{N}(A^m, \Delta)$. Cover A by $O(\mathcal{N}(A, \Delta))$ balls of radius Δ and choose one, say $B(x, \Delta)$, for which

$$\mathcal{N}(A \cap B(x, \Delta), \delta) \geq c \frac{\mathcal{N}(A, \delta)}{\mathcal{N}(A, \Delta)}.$$

By bi-invariance, the sets $y(A \cap B(x, \Delta))$, $y \in Y$, are contained in A^{m+1} and have uniformly bounded overlap after δ -discretization: their centers yx remain 2Δ -separated while each set lies in a Δ -ball about its center. Consequently

$$\mathcal{N}(A^{m+1}, \delta) \geq c\mathcal{N}(A^m, \Delta) \frac{\mathcal{N}(A, \delta)}{\mathcal{N}(A, \Delta)}.$$

Insert the assumed upper bound and cancel $\mathcal{N}(A, \delta)$ to obtain (6.35). $\square$

Theorem 6.29 (Quantitative $SU(2)$ Lie segment). *Fix $0 < \sigma < 3$. There are constants $\tau, \varepsilon_1, \nu > 0$ and an integer C with the following property for all sufficiently small δ . Let $A \subset U$ be symmetric, $1 \in A$, and assume:*

(i) *for every $t \in [\delta, 1]$,*

$$\mathcal{N}(A, t) \geq \delta^{\varepsilon_1} t^{-\sigma};$$

(ii) *A is δ^{ε_1} -away from every maximal torus;*

(iii) for the finite list of powers used by the rich-torus chapter and the present module,

$$\mathcal{N}(A^m, \delta) \leq \delta^{-C_m \varepsilon_1} \mathcal{N}(A, \delta);$$

(iv) at the inherited scale below, the remaining geometric hypotheses of the rich-torus chapter hold with $\rho = \Delta^\nu$.

Then there is a unit vector $\xi \in \mathfrak{su}(2)$ such that

$$\{\exp(t\xi) : 0 \leq t \leq \delta^{1-\tau}\} \subset (A^C)^{(C\delta)}. \quad (6.36)$$

Proof. Let A_2, b_0 be the constants in Lemma 6.27. Set

$$q_0 = \frac{b_0}{8}, \quad b_* = b_0 - q_0 > 0.$$

Choose a fixed base-scale exponent $\theta > 0$ so small that

$$\theta(A_2 + b_0) \leq \frac{1}{4}, \quad (6.37)$$

and put

$$\Delta = \delta^\theta, \quad \tau = \frac{\theta b_*}{4}.$$

Finally choose the mesh parameter $r > 0$ so small that

$$40r \leq \theta b_*. \quad (6.38)$$

All these choices depend only on the fixed parameters. First choose $\nu > 0$ sufficiently small for Lemmas 6.16 and 6.27. Then choose $\varepsilon_1 > 0$ much smaller than

$$\theta\nu, \quad \sigma r, \quad \theta q_0,$$

and than every remaining fixed exponent threshold in Lemmas 6.16 and 6.15. In particular $\varepsilon_1 \leq \theta\nu$, so the original torus-avoidance scale δ^{ε_1} is at least the inherited Larsen–Pink scale $\Delta^\nu = \delta^{\theta\nu}$.

We first construct the local operator segment at scale Δ . Hypothesis (i) gives

$$\mathcal{N}(A, \Delta) \geq \delta^{\varepsilon_1} \Delta^{-\sigma} = \Delta^{-\sigma + \varepsilon_1/\theta}.$$

Lemma 6.28 transfers the bounded-power estimates in (iii) from δ to Δ , with only the displayed fixed power loss. Thus the rich-torus chapter applies at base scale Δ ; Lemma 6.16, Lemma 6.15, and Proposition 6.19 then give an operator segment

$$L = \Delta^\alpha, \quad 0 \leq \alpha \leq A_2,$$

with relative error at most Δ^{b_0} . Because $\varepsilon_1 \leq \theta q_0$, hypothesis (ii), together with Lemma 6.21, is stronger than the local normalizer-avoidance threshold $\Delta^{q_0} = \delta^{\theta q_0}$.

At this point α is known. This is where we use the fine commutator mesh rather than requiring α to have been fixed in advance. Define the interval

$$I_\alpha = [1 - \theta(\alpha + b_0) + 2r, 1 - \tau - \theta(\alpha + q_0) - 2r]. \quad (6.39)$$

Its length is

$$|I_\alpha| = \theta(b_0 - q_0) - \tau - 4r = \frac{3}{4}\theta b_* - 4r > 6r$$

by (6.38). The lower endpoint is at least $3/4 + 2r$ by (6.37), while the upper endpoint is at most $1 - \tau - 2r < 1 - 4r$ by (6.38). Hence $I_\alpha \subset [4r, 1 - 4r]$ after harmlessly decreasing r once more.

Apply Proposition 6.26 at the original scale δ with this value of r . It provides a bounded word $a \in A^{C_0}$ for which

$$\lambda = -\frac{\log d(a, 1)}{\log \delta^{-1}} \in I_\alpha.$$

Since $\lambda < 1$, a lies in the exponential chart for small δ ; write $a = \exp(uX)$ with $\|X\| = 1$ and $u = d(a, 1) = \delta^\lambda$. Relative to the inherited scale $\Delta = \delta^\theta$ this is

$$u = \Delta^\gamma, \quad \gamma = \frac{\lambda}{\theta}.$$

The lower endpoint of I_α and (6.37) imply

$$\gamma \geq \frac{1}{\theta} - (\alpha + b_0) \geq \alpha + b_0,$$

so every hypothesis of Proposition 6.23 is satisfied.

That proposition gives a group segment whose guaranteed endpoint is at least

$$\Delta^{\alpha+\gamma+q_0} = \delta^{\theta(\alpha+q_0)+\lambda} \geq \delta^{1-\tau-2r}.$$

Because $1 - \tau - 2r < 1 - \tau$, this segment contains the shorter target interval $[0, \delta^{1-\tau}]$. Its approximation radius is at most

$$C\Delta^{\alpha+\gamma+b_0} = C\delta^{\theta(\alpha+b_0)+\lambda} \leq C\delta^{1+2r} \leq \delta$$

for sufficiently small δ , where the penultimate inequality uses the lower endpoint of I_α . Enlarging the harmless final constant gives (6.36). $\square$

Remark 6.30 (Why the variable exponent is no longer a blocker). The inherited scale $\Delta = \delta^\theta$ is chosen *before* the operator exponent α is known. Once regularization reveals α , the commutator mesh supplies a small element with exponent λ in the interval (6.39). The two endpoints of that interval encode exactly the two inequalities needed downstream: the lower endpoint makes the BCH/operator error $O(\delta^{1+2r})$, while the upper endpoint keeps the produced segment at least $\delta^{1-\tau-2r}$ long. Thus no fixed-exponent saturation theorem is needed.

Chapter 7

Transverse BCH Filling

The preceding chapter produces one quantitative Lie segment. By conjugating it in three suitably transverse adjoint directions and applying a quantitative inverse-function estimate, we turn three one-dimensional segments into a three-dimensional ball. Finally we transfer the exact filling back to a bounded product of the original discretized set.

7.1 Inputs and target

Fix a sufficiently small exponential chart $U \subset SU(2)$ and a bi-invariant Riemannian metric. Identify $\mathfrak{su}(2)$ with $\mathbb{R}^3$ using the invariant inner product. We use the following two already-owned outputs.

Input A. Lie segment from the Lie-segment chapter. For some fixed $\tau > 0$, $C_0 < \infty$, and a unit vector $\xi \in \mathfrak{su}(2)$,

$$\{\exp(t\xi) : 0 \leq t \leq r\} \subset (A^{C_0})^{(C_0\delta)}, \quad r = \delta^{1-\tau}. \quad (7.1)$$

Input B. Three quantitative adjoint directions from the rich-torus chapter. If A is ρ -away from every maximal torus, then there are $g_1, g_2, g_3 \in (A \cup \{1\})^2$ such that, putting $X_i = \text{Ad}(g_i)\xi$,

$$|\det(X_1, X_2, X_3)| \geq c\rho^3. \quad (7.2)$$

Our target is a ball

$$B_G(1, \delta^{1-\tau_F}) \subset (A^C)^{(C\delta)} \quad (7.3)$$

for some $\tau_F > 0$. This is the precise input the local-product stage needs.

7.2 Symmetrizing and conjugating the segment

Lemma 7.1 (One-sided to symmetric segment). *Assume $A = A^{-1}$ and (7.1). Then*

$$\{\exp(t\xi) : |t| \leq r\} \subset (A^{C_0})^{(C_0\delta)}. \quad (7.4)$$

Proof. For $t \geq 0$ this is (7.1). If $t < 0$, choose $a \in A^{C_0}$ with $d(a, \exp((-t)\xi)) \leq C_0\delta$. Inversion is an isometry for a bi-invariant metric and A^{C_0} is symmetric, so $d(a^{-1}, \exp(t\xi)) \leq C_0\delta$. $\square$

Lemma 7.2 (Conjugation transfers the segment). *Let $g \in A^k$ and $X = \text{Ad}(g)\xi$. Under (7.4),*

$$\{\exp(tX) : |t| \leq r\} \subset (A^{C_0+2k})^{(C_0\delta)}. \quad (7.5)$$

Proof. The identity $g \exp(t\xi)g^{-1} = \exp(t \operatorname{Ad}(g)\xi)$ is exact. Conjugating a $C_0\delta$ -approximation by g preserves its metric error because the metric is bi-invariant, and turns a word in A^{C_0} into one in A^{C_0+2k} . $\square$

Corollary 7.3 (Three approximate transverse segments). *Under Inputs A–B there is C_1 such that, for $i = 1, 2, 3$,*

$$\{\exp(tX_i) : |t| \leq r\} \subset (A^{C_1})^{(C_1\delta)}, \quad (7.6)$$

where the unit vectors X_i satisfy (7.2).

Proof. Apply Lemma 7.2 to the three words g_i supplied by the rich-torus chapter. $\square$

7.3 A quantitative inverse-function lemma

The proof below is included to make the power of the transversality parameter visible.

Lemma 7.4 (Quantitative local surjectivity). *Let $F : B_{\mathbb{R}^3}(0, R) \rightarrow \mathbb{R}^3$ be C^1 , $F(0) = 0$, and let $M = DF(0)$. Suppose*

$$s_{\min}(M) \geq \kappa, \quad \sup_{|x| \leq R} \|DF(x) - M\| \leq \frac{\kappa}{4}. \quad (7.7)$$

Then

$$B_{\mathbb{R}^3}(0, \kappa R/8) \subset F(B_{\mathbb{R}^3}(0, R/2)). \quad (7.8)$$

Proof. Fix y with $|y| \leq \kappa R/8$. Since $\|M^{-1}\| \leq \kappa^{-1}$, define on $B(0, R/2)$

$$\Psi_y(x) = x + M^{-1}(y - F(x)).$$

For $x, z \in B(0, R/2)$, the mean-value formula and (7.7) give

$$|\Psi_y(x) - \Psi_y(z)| \leq \|M^{-1}\| \sup_{w \in B(0, R/2)} \|M - DF(w)\| |x - z| \leq \frac{1}{4} |x - z|.$$

Moreover $|\Psi_y(0)| \leq R/8$. Hence, for $|x| \leq R/2$,

$$|\Psi_y(x)| \leq |\Psi_y(x) - \Psi_y(0)| + |\Psi_y(0)| \leq R/8 + R/8 = R/4.$$

Thus Ψ_y maps the closed ball $B(0, R/2)$ into itself and is a contraction. Its fixed point satisfies $F(x) = y$. This proves (7.8). $\square$

Lemma 7.5 (Determinant controls the least singular value). *If X_1, X_2, X_3 are unit vectors and M is the matrix with these columns, then*

$$s_{\min}(M) \geq \frac{|\det M|}{3}. \quad (7.9)$$

Proof. Let $s_1 \geq s_2 \geq s_3$ be the singular values. Since the Frobenius norm of M is $\sqrt{3}$, one has $s_1, s_2 \leq \sqrt{3}$. Therefore $|\det M| = s_1 s_2 s_3 \leq 3s_3$. $\square$

7.4 The transverse product map

Define

$$\Phi(t_1, t_2, t_3) = \exp(t_1 X_1) \exp(t_2 X_2) \exp(t_3 X_3). \quad (7.10)$$

For r sufficiently small, its image lies in the exponential chart and we put

$$F = \log \circ \Phi.$$

Lemma 7.6 (Derivative and distortion of the product chart). *There are absolute constants $r_0, C > 0$ such that, whenever $r < r_0$,*

$$DF(0)(t_1, t_2, t_3) = t_1 X_1 + t_2 X_2 + t_3 X_3 \quad (7.11)$$

and

$$\sup_{|t| \leq r} \|DF(t) - DF(0)\| \leq Cr. \quad (7.12)$$

Here $|t|$ is the Euclidean norm in $\mathbb{R}^3$.

Proof. The differential of the exponential map at 0 and of the logarithm at 1 are the identity, so the product rule gives (7.11). On a fixed compact subchart, multiplication, exponential, and logarithm have uniformly bounded second derivatives. Taylor's theorem therefore gives (7.12), uniformly in the choice of unit vectors X_i . $\square$

Proposition 7.7 (Three transverse segments fill a ball). *Assume (7.2). There is an absolute $c > 0$ such that, if*

$$r \leq c\rho^3, \quad (7.13)$$

then

$$B_G(1, c\rho^3 r) \subset \{\exp(t_1 X_1) \exp(t_2 X_2) \exp(t_3 X_3) : |t_i| \leq r\}. \quad (7.14)$$

Proof. Let $M = DF(0)$. Lemma 7.5 and (7.2) give $s_{\min}(M) \geq c_1 \rho^3$. By Lemma 7.6, the derivative variation on $B(0, r)$ is at most Cr . Under (7.13), after decreasing c , this is at most $c_1 \rho^3/4$. Lemma 7.4, with $R = r$, shows that

$$B_{\text{su}(2)}(0, c_2 \rho^3 r) \subset F(B(0, r/2)).$$

On the fixed exponential chart, $\exp$ is bi-Lipschitz at this scale, so the image contains $B_G(1, c\rho^3 r)$. Since $B(0, r/2)$ is contained in the coordinate cube $[-r, r]^3$, this proves (7.14). $\square$

7.5 Transfer from exact filling to a bounded power of A

Lemma 7.8 (Products of approximate factors). *For fixed m , if $d(x_j, a_j) \leq e$ for $1 \leq j \leq m$, then in a bi-invariant metric*

$$d(x_1 \cdots x_m, a_1 \cdots a_m) \leq me. \quad (7.15)$$

Proof. Insert and remove the factors one at a time. Left and right translations are isometries, so each replacement changes the product by at most e ; the triangle inequality gives the result. $\square$

Theorem 7.9 (Quantitative transverse filling for the compact product argument). *Assume the Lie-segment chapter gives (7.1) with $r = \delta^{1-\tau}$, and assume A is ρ -away from every maximal torus with*

$$\rho \geq \delta^{\varepsilon_0}. \quad (7.16)$$

If $\varepsilon_0 < \min\{(1-\tau)/4, \tau/12\}$, then there are $C < \infty$ and $\tau_F > 0$ such that

$$B_G(1, \delta^{1-\tau_F}) \subset (A^C)^{(C\delta)}. \quad (7.17)$$

One may take any fixed

$$0 < \tau_F < \tau - 3\varepsilon_0; \quad (7.18)$$

for example $\tau_F = (\tau - 3\varepsilon_0)/2 > 3\tau/8$.

Proof. the rich-torus chapter and Corollary 7.3 give unit vectors X_i with determinant at least $c\rho^3$ and three segment approximations (7.6). The filling condition $r \leq c\rho^3$ follows for small δ because

$$r = \delta^{1-\tau}, \quad \rho^3 \geq \delta^{3\varepsilon_0},$$

and $1 - \tau > 3\varepsilon_0$ by hypothesis. Proposition 7.7 therefore gives a ball of radius at least

$$c\rho^3 r \geq c\delta^{1-\tau+3\varepsilon_0}.$$

For any fixed τ_F satisfying (7.18), the exponent $1 - \tau_F$ is strictly larger than $1 - \tau + 3\varepsilon_0$. Hence, for sufficiently small δ , the last ball contains $B_G(1, \delta^{1-\tau_F})$; this absorbs the fixed constant c without changing any equality of exponents.

Every point in the filling image is a product of three exact segment factors. Replace those factors by their $C_1\delta$ -approximants in A^{C_1} from (7.6). By Lemma 7.8, the resulting product lies within $3C_1\delta$ and belongs to A^{3C_1} . This proves (7.17) after renaming the constant. $\square$

Chapter 8

The Compact Local Product Theorem

We combine scalar growth, rich-torus extraction, Lie-segment amplification, and transverse filling. The key covering argument translates the resulting power-scale ball by a separated subset of the original set; this supplies the missing factor that produces strict covering growth throughout the intermediate-dimensional range.

8.1 Covering preliminaries

Write $\mathcal{N}(E, r)$ for the least number of radius- r metric balls needed to cover E . All sets lie in a fixed compact chart, and all scale comparisons are for sufficiently small r . We use the elementary local volume estimate

$$\mathcal{N}(B_G(1, R), \delta) \asymp (R/\delta)^3 \quad (0 < \delta \leq R \ll 1). \quad (8.1)$$

This follows from normal coordinates and bounded Jacobians.

Lemma 8.1 (Separated translates of a local ball). *Let $0 < \delta \ll R \ll 1$. Suppose*

$$B_G(1, R) \subset (A^k)^{(C_0\delta)}. \quad (8.2)$$

Then

$$\mathcal{N}(A^{k+1}, C_1\delta) \geq c\mathcal{N}(A, 8R)(R/\delta)^3, \quad (8.3)$$

where $C_1, c > 0$ depend only on the fixed chart and C_0 .

Proof. Choose a maximal $8R$ -separated set $Y \subset A$. Then $|Y| \asymp \mathcal{N}(A, 8R)$. The balls $yB_G(1, 2R)$, $y \in Y$, are pairwise disjoint. By (8.2), every point of $yB_G(1, R)$ lies within $C_0\delta$ of $yA^k \subset A^{k+1}$. Hence the disjoint union

$$E := \bigcup_{y \in Y} yB_G(1, R)$$

is contained in $(A^{k+1})^{(C_0\delta)}$. If q balls of radius $C_1\delta$ cover A^{k+1} , then the same centers, with radius $(C_0 + C_1)\delta$, cover E . Thus

$$q \geq \mathcal{N}(E, (C_0 + C_1)\delta).$$

Because the centers in Y are $8R$ -separated, the components $yB_G(1, R)$ are separated by at least $6R$. For $\delta \ll R$, a ball of radius $(C_0 + C_1)\delta$ meets at most one component, while each component requires $\gg (R/\delta)^3$ such balls by (8.1). Hence

$$q \gg |Y|(R/\delta)^3 \asymp \mathcal{N}(A, 8R)(R/\delta)^3,$$

which is (8.3). □

Lemma 8.2 (Scale regularity converts filling into a dimension gain). *Fix $0 < \sigma < 3$. Suppose, for some $\eta > 0$,*

$$\mathcal{N}(A, t) \geq \delta^\eta t^{-\sigma} \quad (\delta \leq t \leq 1), \quad (8.4)$$

and suppose (8.2) holds with

$$R = \delta^{1-\tau_F}, \quad \tau_F > 0. \quad (8.5)$$

Then

$$\mathcal{N}(A^{k+1}, C_1 \delta) \geq c \delta^{-\sigma-(3-\sigma)\tau_F+\eta}. \quad (8.6)$$

Proof. Apply Lemma 8.1 and (8.4) at scale $8R$:

$$\begin{aligned} \mathcal{N}(A^{k+1}, C_1 \delta) &\gg \delta^\eta R^{-\sigma} (R/\delta)^3 \\ &= \delta^\eta \delta^{-\sigma(1-\tau_F)} \delta^{-3\tau_F} \\ &= \delta^{-\sigma-(3-\sigma)\tau_F+\eta}. \end{aligned}$$

□

8.2 The intermediate-dimensional local product theorem

The exact finite-set owner behind the next lemma is BG-II.2.001: bounded signed-word growth from small tripling. What requires proof here is the passage from that finite statement to covering numbers. We make that passage explicit so that no metric inverse theorem is hidden inside the phrase “small tripling.”

Lemma 8.3 (Fixed metric powers from small tripling). *For every fixed integer $m \geq 3$ there are constants $C_m, c_m > 0$ with the following property. Let $A \subset U$ be compact, symmetric, and contain 1. If*

$$\mathcal{N}(A^3, \delta) \leq K \mathcal{N}(A, \delta),$$

then

$$\mathcal{N}(A^m, C_m \delta) \leq c_m K^{C_m} \mathcal{N}(A, \delta). \quad (8.7)$$

The constants depend on m and on the fixed local geometry of U , but not on A, K , or δ .

Proof. The earlier metric Ruzsa triangle theorem was proved by replacing each finite set in the exact Ruzsa injection by a maximal separated net. It has the following consequence: every fixed finite chain of Ruzsa-triangle inequalities remains valid for covering numbers, with only a fixed enlargement of the resolution and an absolute packing multiplicity at each step.

Now replay the proof of BG-II.2.001. That proof starts from $|B^3| \leq K|B|$ for a finite symmetric set $B \ni 1$ and, using only finitely many Ruzsa triangle inequalities and the inclusions $B \subset B^2 \subset B^3$, obtains for each fixed m

$$|B^m| \leq K^{O_m(1)} |B|.$$

Replace every occurrence of a finite-set Ruzsa inequality in this fixed proof by its metric version. Because the number of steps is bounded in terms of m , the successive resolution inflations multiply to a single constant C_m , the bounded packing multiplicities multiply to a single constant c_m , and the powers of K remain bounded by K^{C_m} after enlarging C_m once. The initial hypothesis is exactly the covering-number analogue of the finite tripling hypothesis, while $1 \in A$ gives $\mathcal{N}(A, C\delta) \leq \mathcal{N}(A^2, C\delta) \leq \mathcal{N}(A^3, C\delta)$ whenever those inclusions are used. This proves (8.7). □

Fixed multiplicative changes of the resolution are harmless here: in the three-dimensional compact chart, maximal-net/volume comparison gives $\mathcal{N}(E, \delta) \ll_C \mathcal{N}(E, C\delta)$, while monotonicity gives the reverse comparison in the other direction. Thus a power-saving assertion at $C\delta$ can be normalized back to δ at the cost of an absolute factor.

Theorem 8.4 (Compact local product growth in the intermediate range). *Fix $0 < \sigma < 3$. There are constants $c_* = c_*(\sigma) > 0$, $\eta_* = \eta_*(\sigma) > 0$, $C_* = C_*(\sigma) \geq 1$ and $\delta_* = \delta_*(\sigma) > 0$ such that, for every $0 < \delta < \delta_*$, the following holds. Let $A \subset U$ be symmetric, $1 \in A$, and assume*

$$\mathcal{N}(A, \delta) \leq \delta^{-\sigma-\eta_*}, \quad (8.8)$$

together with the all-scale lower bound

$$\mathcal{N}(A, t) \geq \delta^{\eta_*} t^{-\sigma} \quad (\delta \leq t \leq 1). \quad (8.9)$$

Assume also that, for every maximal torus T , some $a_T \in A$ satisfies

$$d(a_T, T) \geq \delta^{\eta_*}. \quad (8.10)$$

Then

$$\mathcal{N}(A^3, C_*\delta) \geq \delta^{-c_*} \mathcal{N}(A, \delta). \quad (8.11)$$

The displayed constants may also depend on the fixed compact chart, but not on δ or on A .

Proof. Let $\tau > 0$ and $\varepsilon_1 > 0$ be the fixed constants supplied by the Lie-segment chapter for this σ . We first fix

$$\tau_F := \tau/2.$$

Choose $\eta_* > 0$ so small that

$$\eta_* < \min \left\{ \varepsilon_1, \frac{1-\tau}{4}, \frac{\tau}{12}, \frac{(3-\sigma)\tau_F}{16} \right\}. \quad (8.12)$$

With this choice, the torus-avoidance hypothesis of the transverse-filling chapter may be used with $\rho = \delta^{\eta_*}$ and $\varepsilon_0 = \eta_*$. In particular $\tau_F = \tau/2 < \tau - 3\eta_*$, so this value of τ_F is independent of the subsequent small-tripling exponent.

Now choose $c_* > 0$ sufficiently small that the fixed-power losses furnished by (8.7) meet all the finitely many upper-growth thresholds in Modules D–E and, for the power $k+1$ occurring below,

$$C_{k+1}c_* < \frac{(3-\sigma)\tau_F}{4}. \quad (8.13)$$

Such a choice is possible because the Lie-segment chapter's power thresholds are fixed once σ is fixed.

Argue by contradiction. If (8.11) fails, fixed-scale comparison lets us normalize the failure to

$$\mathcal{N}(A^3, \delta) < C\delta^{-c_*} \mathcal{N}(A, \delta).$$

The fixed multiplicative constant in this tripling bound is harmless: apply (8.7) with $K = C\delta^{-c_*}$ and absorb the resulting fixed factors into the constants implicit in the exponents C_m . Thus, for every one of the finitely many bounded powers required in the rich-torus, Lie-segment, and transverse-filling chapters,

$$\mathcal{N}(A^m, C\delta) \leq \delta^{-C_m c_*} \mathcal{N}(A, \delta). \quad (8.14)$$

Because the final Module-E parameter order gives $\varepsilon_1 \leq \theta\nu$, our choice $\eta_* \leq \varepsilon_1$ implies

$$\delta^{\eta_*} \geq \delta^{\varepsilon_1} \geq \delta^{\theta\nu} = \Delta^\nu.$$

Thus the lower regularity and torus-avoidance assumptions (8.9)–(8.10) are at least as strong as both the original Module-E hypotheses and the inherited Module-D scale $\rho = \Delta^\nu$. Therefore the rich-torus, Lie-segment, and transverse-filling chapters apply, and the transverse-filling chapter with the already fixed $\tau_F = \tau/2$ gives $k < \infty$ such that

$$B_G(1, R) \subset (A^k)^{(C\delta)}, \quad R = \delta^{1-\tau_F}. \quad (8.15)$$

Lemma 8.2 and (8.8) now give

$$\begin{aligned} \frac{\mathcal{N}(A^{k+1}, C_1\delta)}{\mathcal{N}(A, \delta)} &\geq c\delta^{-(3-\sigma)\tau_F+2\eta_*} \\ &\geq \delta^{-3(3-\sigma)\tau_F/4} \end{aligned}$$

for sufficiently small δ , by (8.12) (with ample room for the fixed constant c).

On the other hand (8.14), applied to $m = k + 1$, bounds the same ratio by $\delta^{-C_{k+1}c_*}$. By (8.13) this is strictly smaller than the lower power above for sufficiently small δ , a contradiction. Hence (8.11) holds. $\square$

8.3 Obstruction form

Corollary 8.5 (Growth-or-concentration trichotomy). *Fix $0 < \sigma < 3$, and let $c_*, \eta_*, C_*, \delta_*$ be the constants in Theorem 8.4. For every $0 < \delta < \delta_*$ and every symmetric $A \subset U$ with $1 \in A$ and $\mathcal{N}(A, \delta) \leq \delta^{-\sigma-\eta_*}$, at least one of the following holds:*

- (i) **product growth:** $\mathcal{N}(A^3, C_*\delta) \geq \delta^{-c_*}\mathcal{N}(A, \delta)$;
- (ii) **multiscale concentration:** for some $t \in [\delta, 1]$, $\mathcal{N}(A, t) < \delta^{\eta_*}t^{-\sigma}$;
- (iii) **proper-subgroup concentration:** for some maximal torus T , $A \subset T^{(\delta^{\eta_*})}$.

Proof. If (ii) and (iii) both fail, the all-scale lower bound and torus-avoidance hypotheses of Theorem 8.4 hold, so (i) follows. $\square$

Remark 8.6 (Endpoint dimensions). The strict inequality $\sigma < 3$ is exactly where the gain $(3 - \sigma)\tau_F$ in Lemma 8.2 is positive. The near-full-dimensional endpoint is handled separately in the flattening/spectral endgame; it is not hidden inside the intermediate-dimensional theorem. Likewise, very small effective support is handled by the initial non-concentration/escape stage before this local theorem is invoked.

8.4 Transition to flattening

The covering gain above explains why the power-scale ball is useful even when it contains fewer δ -cells than A : translating the ball by an R -separated subset of A contributes the additional factor $\mathcal{N}(A, R)$ and yields the strict exponent gain $(3 - \sigma)\tau_F$. The next chapter packages the corresponding failure alternatives in the form needed by the L^2 inverse theorem.

Chapter 9

Obstruction Classification and Random-Walk Interfaces

The local product theorem is now converted into the exact trichotomy required by flattening: growth, concentration in a small ball, or concentration near a maximal torus. We formulate the alternatives in the captured-set language produced by the metric inverse theorem and show how the previously proved random-walk nonconcentration estimates eliminate the two geometric obstruction branches.

9.1 Standing internal inputs

Let $G = SU(2)$ with a fixed bi-invariant Riemannian metric. Fix a sufficiently small identity chart U . Write $\mathcal{N}(E, r)$ for the radius- r covering number.

We use the following results proved earlier in this volume.

P1. Metric failed-flattening output. For a probability q , $f = q * P_\delta$, and a flattening failure with loss δ^ζ , there are a symmetric $H \subset U$, $1 \in H$, a finite X , and $a \in G$ such that

$$H^2 \subset HX \cap XH, \quad |X| \leq \delta^{-C_0\zeta}, \quad (9.1)$$

$$(q * P_\delta)(aH) \geq \delta^{C_0\zeta}, \quad (9.2)$$

$$\mathcal{N}(H^m, C_m\delta) \leq \delta^{-C_m\zeta} \mathcal{N}(H, \delta) \quad (m \text{ fixed}). \quad (9.3)$$

The inverse theorem supplies no unconditional lower size.

P2. Ball non-concentration. There are $\kappa_b > 0, C_b < \infty$ such that

$$(q * P_\delta)(B(x, r)) \leq C_b r^{\kappa_b} \quad (\delta \leq r \leq r_0). \quad (9.4)$$

P3. Torus-coset escape. There are $\kappa_t > 0, C_t < \infty$ such that, for every maximal torus T and $x \in G$,

$$(q * P_\delta)(xT^{(r)}) \leq C_t r^{\kappa_t} \quad (\delta \leq r \leq r_0). \quad (9.5)$$

P4. Module-G intermediate-dimensional product theorem. For every compact $J \Subset (0, 3)$ there are positive uniform constants such that a symmetric $A \subset U$ with effective dimension $\sigma \in J$, the required all-scale lower covering bound, and avoidance of all maximal tori has a power gain in $\mathcal{N}(A^3, O(\delta))$. Equivalently, failure of that power gain forces either failure of the all-scale lower bound at some intermediate scale or containment in a small maximal-torus tube.

9.2 A captured set cannot fail scale regularity silently

Lemma 9.1 (Covering failure produces an actual mass concentration). *Let μ be a probability measure and suppose $\mu(aH) \geq \beta$. Then for every $r \geq \delta$ there exists $x \in G$ such that*

$$\mu(B(x, 2r)) \geq \frac{\beta}{\mathcal{N}(H, r)}. \quad (9.6)$$

Consequently, if

$$\mathcal{N}(H, r) < \delta^\eta r^{-\sigma}, \quad (9.7)$$

then some $2r$ -ball has mass

$$\mu(B(x, 2r)) > \beta \delta^{-\eta} r^\sigma. \quad (9.8)$$

Proof. Cover H by $M = \mathcal{N}(H, r)$ radius- r balls. Their left translates by a cover aH ; bi-invariance shows that each translated radius- r ball is again a radius- r ball. The total μ -mass of the cover is at least β , so one member has mass at least β/M . Enlarging the radius to $2r$ absorbs harmless choices of open/closed balls. Substituting (9.7) gives (9.8). $\square$

Corollary 9.2 (Anti-ball estimates force the all-scale lower bound). *Assume (9.2) and (9.4). Fix $0 < \sigma < \kappa_b$ and $\eta > 0$. If $\zeta > 0$ is chosen after η so that*

$$C_0 \zeta < \eta, \quad (9.9)$$

then, for all sufficiently small δ and every $\delta \leq r \leq r_0$,

$$\mathcal{N}(H, r) \geq \delta^\eta r^{-\sigma}. \quad (9.10)$$

Proof. If (9.10) failed, Lemma 9.1 with $\mu = q * P_\delta$ and $\beta = \delta^{C_0 \zeta}$ would give

$$\mu(B(x, 2r)) > \delta^{C_0 \zeta - \eta} r^\sigma.$$

The anti-ball estimate gives the opposite bound $\ll r^{\kappa_b}$. Hence failure would force

$$\delta^{C_0 \zeta - \eta} \ll r^{\kappa_b - \sigma} \leq r_0^{\kappa_b - \sigma}.$$

But (9.9) makes the exponent $C_0 \zeta - \eta$ negative, so the left-hand side tends to $+\infty$ as $\delta \rightarrow 0$, uniformly in $r \leq r_0$. This is impossible once δ is sufficiently small. Thus (9.10) holds at every scale. $\square$

Remark 9.3 (Quantifier order). For the eventual flattening argument one chooses a target effective-dimension interval first, then the product-theorem regularity exponent, and only afterwards the flattening-loss exponent ζ . Thus the preceding choice of ζ is compatible with the required quantifier order.

9.3 The torus branch is an actual torus-coset concentration

Lemma 9.4 (Captured torus containment creates torus-coset mass). *If $H \subset T^{(r)}$ for a maximal torus T , then*

$$(q * P_\delta)(aT^{(r)}) \geq (q * P_\delta)(aH). \quad (9.11)$$

In particular, under (9.2), the torus branch gives mass at least $\delta^{C_0 \zeta}$ to one left torus tube.

Proof. The inclusion $aH \subset aT^{(r)}$ is immediate. Apply the positive measure $q * P_\delta$ to both sides. $\square$

Corollary 9.5 (Torus escape excludes the torus branch). *Assume (9.2) and (9.5). If a product-theorem obstruction gives $H \subset T^{(\delta^\eta)}$, choose $\zeta > 0$ after η so that*

$$C_0 \zeta < \eta \kappa_t. \quad (9.12)$$

Then the torus branch is impossible for sufficiently small δ .

Proof. By Lemma 9.4, the captured mass is at least $\delta^{C_0 \zeta}$. By torus escape it is at most $C_t \delta^{\eta \kappa_t}$. Condition (9.12) makes the lower bound asymptotically larger than the upper bound. $\square$

9.4 Canonical obstruction alternative

Theorem 9.6 (Rank-one product/obstruction alternative in captured form). *Fix $0 < \sigma < 3$, and let $c_*, \eta_*, C_*, \delta_*$ be the constants supplied by Theorem 8.4. There exists $\delta_0 = \delta_0(\sigma) > 0$ such that the following holds for $0 < \delta < \min\{\delta_*, \delta_0\}$. Let $H \subset U$ be the symmetric captured set in (9.1)–(9.3), assume $1 \in H$, and suppose*

$$\mathcal{N}(H, \delta) \leq \delta^{-\sigma-\eta_*}. \quad (9.13)$$

Then at least one of the following alternatives holds:

(i) **product growth:**

$$\mathcal{N}(H^3, C_*\delta) \geq \delta^{-c_*} \mathcal{N}(H, \delta); \quad (9.14)$$

(ii) **small-ball concentration:** *there exist $r \in [\delta, 1]$ and $x \in G$ such that*

$$\mathcal{N}(H, r) < \delta^{\eta_*} r^{-\sigma}, \quad (q * P_\delta)(B(x, 2r)) > \delta^{C_0\zeta - \eta_*} r^\sigma; \quad (9.15)$$

(iii) **torus-tube concentration:** *there is a maximal torus T such that*

$$H \subset T^{(\delta^{\eta_*})}, \quad (q * P_\delta)(aT^{(\delta^{\eta_*})}) \geq \delta^{C_0\zeta}. \quad (9.16)$$

Proof. Apply Corollary 8.5 to $A = H$. Its size hypothesis is exactly (9.13). Alternative (i) of that corollary is (9.14).

If its multiscale alternative holds, there is $r \in [\delta, 1]$ with the first inequality in (9.15). The capture estimate (9.2) gives $\mu(aH) \geq \delta^{C_0\zeta}$ for $\mu = q * P_\delta$; therefore Lemma 9.1, with $\beta = \delta^{C_0\zeta}$ and $\eta = \eta_*$, gives the second inequality in (9.15).

If the torus alternative of Corollary 8.5 holds, then $H \subset T^{(\delta^{\eta_*})}$ for some maximal torus T . Lemma 9.4 and (9.2) give the second assertion in (9.16). These are precisely the three alternatives of the cited corollary. $\square$

Corollary 9.7 (Obstruction elimination). *In Theorem 9.6, assume in addition that $0 < \sigma < \kappa_b$ and choose $\zeta > 0$ so that*

$$C_0\zeta < \min\{\eta_*, \eta_*\kappa_t\}. \quad (9.17)$$

Then, for sufficiently small δ , alternatives (ii) and (iii) are impossible.

Proof. Alternative (ii) contradicts Corollary 9.2 with $\eta = \eta_*$, because the first inequality in (9.17) is exactly (9.9). Alternative (iii) contradicts Corollary 9.5, because the second inequality in (9.17) is exactly (9.12) with $\eta = \eta_*$. $\square$

Corollary 9.8 (Approximate-group contradiction and one-step flattening interface). *Assume the hypotheses and parameter order of Corollary 9.7. If the fixed-power estimate (9.3) is available for $m = 3$ with constant C_3 and*

$$C_3\zeta < c_*, \quad (9.18)$$

then the growth alternative (i) is also impossible for sufficiently small δ . Hence the assumed one-step flattening failure cannot occur.

Proof. By Corollary 9.7, only alternative (i) of Theorem 9.6 can remain. That alternative gives

$$\mathcal{N}(H^3, C_*\delta) \geq \delta^{-c_*} \mathcal{N}(H, \delta).$$

Fixed-scale comparison changes $C_*\delta$ to the resolution $C_3\delta$ in (9.3) at only an absolute multiplicative cost. Thus (9.3) gives the opposite estimate

$$\mathcal{N}(H^3, C_*\delta) \ll \delta^{-C_3\zeta} \mathcal{N}(H, \delta).$$

Because (9.18) gives $c_* - C_3\zeta > 0$, these two inequalities are incompatible as $\delta \rightarrow 0$. $\square$

Remark 9.9 (Upper and lower effective size). The trichotomy preserves an important logical distinction: failed flattening supplies the upper effective-support bound, whereas the lower-size information is supplied independently by anti-ball nonconcentration. Neither estimate is inferred from the other.

Chapter 10

Compact L^2 Flattening and Multiscale Iteration

We combine the metric inverse theorem with the compact product/obstruction alternative to obtain a one-step L^2 improvement. Because the product theorem is naturally parameterized by an effective dimension, we first derive a scale-invariant weak product form uniform on the compact range of dimensions that occurs in the iteration. Neighboring-scale comparison then propagates the nonconcentration hypotheses through a bounded multiscale iteration.

10.1 Interfaces and normalization

Let $G = SU(2)$, equipped with a fixed bi-invariant Riemannian metric and normalized Haar probability measure m_G . For a Borel set $E \subset G$, $\mathcal{N}(E, r)$ denotes its radius- r covering number. Write

$$P_r = \frac{1_{B(1, r)}}{m_G(B(1, r))}.$$

All scales are sufficiently small, and changing a radius by a fixed multiplicative constant changes covering numbers and the estimates below only by an absolute constant.

The argument uses four results established in the preceding chapters.

- I1. Fixed-dimension product theorem (the local-product chapter).** For every compact interval $J \subseteq (0, 3)$ there is a uniform $c_J > 0$ such that the local-product theorem may be used for every $\sigma \in J$ with a growth exponent at least c_J , provided its regularity and torus-avoidance exponents are chosen sufficiently small in terms of J . This uniformity is obtained by taking the minimum of the finitely many explicit parameter margins occurring in the scalar-to-local-product chain after restricting σ to J .
- I2. Metric failed-flattening inverse theorem.** There is an absolute C_0 with the following parameterized consequence. Let q be a probability measure, put $f = q * P_\delta$, and suppose

$$\|q * q * P_\delta\|_2 > \delta^\zeta \|f\|_2. \quad (10.1)$$

Then there are a symmetric set $H = H^{-1}$ which is a union of $O(\delta)$ -balls, a finite set X , and $a \in G$ such that

$$H^2 \subset HX \cap XH, \quad |X| \leq \delta^{-C_0\zeta}, \quad (10.2)$$

$$(q * P_\delta)(aH) \geq \delta^{C_0\zeta}, \quad (10.3)$$

$$m_G(H) \leq \delta^{-C_0\zeta} \|f\|_2^{-2}. \quad (10.4)$$

In particular every fixed product power obeys

$$\mathcal{N}(H^m, C\delta) \leq \delta^{-C_m \zeta} \mathcal{N}(H, \delta). \quad (10.5)$$

This is the exact qualitative content needed below: captured mass together with an upper effective-support bound.

I3. Ball non-concentration. For the logarithmic random-walk lengths used below there is $\kappa_b > 0$ such that

$$(q * P_\delta)(B(x, r)) \leq Cr^{\kappa_b} \quad (x \in G, \delta \leq r \leq r_0). \quad (10.6)$$

The same estimate with an additional harmless factor $\delta^{-o(1)}$ is sufficient.

I4. Torus-coset escape. There is $\kappa_t > 0$ such that for every maximal torus T , every $x \in G$, and $\delta \leq r \leq r_0$,

$$(q * P_\delta)(xT^{(r)}) \leq Cr^{\kappa_t}. \quad (10.7)$$

The use of left cosets is essential because the inverse theorem captures mass on aH rather than on H itself.

Interfaces I2–I4 were proved in the preceding chapters; here we use only the displayed consequences.

10.2 A weak product theorem derived from the local-product chapter

Failed flattening naturally gives only a fixed Frostman lower exponent, while the effective support dimension can vary. We therefore prove the version of the product theorem in which the lower exponent and the upper effective dimension are allowed to differ.

Theorem 10.1 (Weak compact product theorem). *Fix $0 < \kappa < \theta < 3$. There exists $\varepsilon = \varepsilon(\kappa, \theta) > 0$ such that for all sufficiently small δ the following holds. Let $A \subset U$ be symmetric, $1 \in A$, and assume*

$$\mathcal{N}(A, \delta) \leq \delta^{-\theta-\varepsilon}, \quad (10.8)$$

$$\mathcal{N}(A, r) \geq \delta^\varepsilon r^{-\kappa} \quad (\delta \leq r \leq r_0), \quad (10.9)$$

and for every maximal torus T there is $a_T \in A$ with

$$d(a_T, T) > \delta^\varepsilon. \quad (10.10)$$

Then

$$\mathcal{N}(A^3, C\delta) \geq \delta^{-\varepsilon} \mathcal{N}(A, \delta). \quad (10.11)$$

Proof. Put

$$\Theta = \frac{\theta + 3}{2}, \quad \theta < \Theta < 3.$$

Choose $\varepsilon_3 > 0$ so small that the local-product chapter is valid uniformly for $\sigma \in [\kappa/2, \Theta]$ with regularity, avoidance, and growth exponent ε_3 . Shrink ε_3 further so that

$$\varepsilon_3 < \frac{\kappa}{4}, \quad \varepsilon_3 < \frac{\Theta - \kappa}{4}.$$

Let K be the least nonnegative integer such that

$$\sigma_K := \Theta - (2K + 1)\varepsilon_3 \leq \kappa.$$

Then $\sigma_j \in [\kappa/2, \Theta]$ for $0 \leq j \leq K$. Set

$$a = \frac{\varepsilon_3}{4\Theta} \in (0, 1).$$

Let C_R be a fixed exponent in the earlier metric small-tripling propagation statement: if $\mathcal{N}(A^3, C\delta) \leq K_0 \mathcal{N}(A, \delta)$, then $\mathcal{N}(A^4, C'\delta) \leq K_0^{C_R} \mathcal{N}(A, \delta)$. Choose $\varepsilon > 0$ so small that

$$\varepsilon < \min \left\{ \frac{\Theta - \theta}{2}, \frac{a^K \varepsilon_3}{4}, \frac{a^K \varepsilon_3}{4C_R} \right\}. \quad (10.12)$$

Assume for contradiction that

$$\mathcal{N}(A^3, C\delta) < \delta^{-\varepsilon} \mathcal{N}(A, \delta). \quad (10.13)$$

The small-tripling propagation theorem then gives

$$\mathcal{N}(A^4, C'\delta) \leq \delta^{-C_R \varepsilon} \mathcal{N}(A, \delta). \quad (10.14)$$

We construct scales

$$\delta = \rho_0 \leq \rho_1 \leq \dots$$

with the two invariants

$$\rho_j \leq \delta^{a^j}, \quad (10.15)$$

$$\mathcal{N}(A, \rho_j) \leq \rho_j^{-\Theta + 2j\varepsilon_3}. \quad (10.16)$$

For $j = 0$, (10.15) is trivial, while (10.8) and $\theta + \varepsilon < \Theta$ imply $\mathcal{N}(A, \delta) \leq \delta^{-\Theta}$, proving (10.16).

Suppose the invariants hold at $j \leq K$ and write

$$\sigma_j = \Theta - (2j + 1)\varepsilon_3.$$

Then (10.16) is exactly the upper-size hypothesis

$$\mathcal{N}(A, \rho_j) \leq \rho_j^{-\sigma_j - \varepsilon_3}$$

required by the local-product chapter at base scale ρ_j . Moreover, (10.9), (10.15), and $\varepsilon \leq a^K \varepsilon_3 \leq a^j \varepsilon_3$ give, for every $r \geq \rho_j$,

$$\mathcal{N}(A, r) \geq \delta^\varepsilon r^{-\kappa} \geq \rho_j^{\varepsilon_3} r^{-\kappa}. \quad (10.17)$$

The same exponent comparison turns (10.10) into $\rho_j^{\varepsilon_3}$ -avoidance from every maximal torus.

If

$$\mathcal{N}(A, r) \geq \rho_j^{\varepsilon_3} r^{-\sigma_j} \quad (\rho_j \leq r \leq r_0) \quad (10.18)$$

held, the local-product chapter would give

$$\mathcal{N}(A^3, C\rho_j) \geq \rho_j^{-\varepsilon_3} \mathcal{N}(A, \rho_j). \quad (10.19)$$

We now show that this is incompatible with (10.14). Choose a $5\rho_j$ -separated subset $Y \subset A^3$ with $|Y| \gg \mathcal{N}(A^3, C\rho_j)$, and choose a radius- ρ_j ball containing a δ -separated subset $Z \subset A$ of cardinality

$$|Z| \gg \frac{\mathcal{N}(A, \delta)}{\mathcal{N}(A, \rho_j)}.$$

After translating Z inside that ball, bi-invariance of the metric shows that $YZ \subset A^4$ is $c\delta$ -separated: distinct Y -fibres are separated at scale ρ_j , and separation inside a fibre is inherited from Z . Hence

$$\mathcal{N}(A^4, C\delta) \gg \frac{\mathcal{N}(A, \delta)}{\mathcal{N}(A, \rho_j)} \mathcal{N}(A^3, C\rho_j) \gg \rho_j^{-\varepsilon_3} \mathcal{N}(A, \delta). \quad (10.20)$$

By (10.15), $\rho_j^{-\varepsilon_3} \geq \delta^{-a^j \varepsilon_3} \geq \delta^{-a^K \varepsilon_3}$, contradicting (10.14) and (10.12). Therefore (10.18) must fail whenever $j < K$.

Choose $\rho_{j+1} \geq \rho_j$ for which it fails:

$$\mathcal{N}(A, \rho_{j+1}) < \rho_j^{\varepsilon_3} \rho_{j+1}^{-\sigma_j}. \quad (10.21)$$

Comparing with the original lower bound (10.9) gives

$$\rho_{j+1}^{\sigma_j - \kappa} < \delta^{-\varepsilon} \rho_j^{\varepsilon_3} \leq \delta^{a^j \varepsilon_3 - \varepsilon}.$$

Since $0 < \sigma_j - \kappa \leq \Theta$ and $\varepsilon \leq a^j \varepsilon_3/2$, we obtain

$$\rho_{j+1} \leq \delta^{a^j \varepsilon_3 / (2\Theta)} = \delta^{2a^{j+1}} \leq \delta^{a^{j+1}},$$

which is the next instance of (10.15). Since $\rho_j \leq \rho_{j+1}$, (10.21) also yields

$$\mathcal{N}(A, \rho_{j+1}) \leq \rho_{j+1}^{\varepsilon_3 - \sigma_j} = \rho_{j+1}^{-\Theta + (2j+2)\varepsilon_3},$$

which is (10.16) at $j+1$.

At $j = K$ we have $\sigma_K \leq \kappa$. Then (10.17) automatically implies (10.18), because $r^{-\kappa} \geq r^{-\sigma_K}$ for $0 < r \leq 1$. The local-product chapter must therefore give (10.19), and the transfer argument again contradicts (10.14). This contradiction proves (10.11). $\square$

Remark 10.2 (What is new here). This is the compact rank-one counterpart of the scale-rescaling argument used to pass from a fixed-dimensional product theorem to the weaker spectral-gap version. The proof uses the local-product chapter as the only new geometric input; it does not reintroduce the superseded three-dimensional small-doubling shortcut.

10.3 Converting captured mass into the weak product hypotheses

Lemma 10.3 (Captured mass forces all-scale covering). *Assume (10.3) and (10.6). Then for every $\delta \leq r \leq r_0$,*

$$\mathcal{N}(H, r) \geq c \delta^{C_0 \zeta} r^{-\kappa_b}. \quad (10.22)$$

Proof. Cover H by $M = \mathcal{N}(H, r)$ radius- r balls. Since H is a union of $O(\delta)$ -balls, aH is contained in the union of the corresponding M balls enlarged to radius Cr . Therefore, by (10.6),

$$\delta^{C_0 \zeta} \leq (q * P_\delta)(aH) \leq CM r^{\kappa_b}.$$

Rearranging gives (10.22). $\square$

Lemma 10.4 (The effective support stays below full dimension). *Suppose $\|f\|_2 \geq \delta^{-\gamma}$ and (10.4) holds. If $0 < \zeta \leq \gamma/(4C_0)$, then*

$$\mathcal{N}(H, C\delta) \leq \delta^{-3-C_0\zeta+2\gamma} \leq \delta^{-(3-\gamma)-C_0\zeta}. \quad (10.23)$$

Proof. Because H is a union of $O(\delta)$ -balls, local volume comparison gives $\mathcal{N}(H, C\delta) \ll m_G(H) \delta^{-3}$. From (10.4) and the lower bound on $\|f\|_2$,

$$m_G(H) \leq \delta^{-C_0\zeta+2\gamma}.$$

This proves the first inequality. The second follows from $2\gamma - C_0\zeta \geq \gamma$. $\square$

Lemma 10.5 (Local difference core of the inverse set). *Fix symmetric identity neighborhoods V, U with $V^{-1}V \subset U$, where U is the identity chart required by Theorem 10.1. Choose once and for all a finite cover*

$$G = \bigcup_{i=1}^M g_i V.$$

For the inverse-theorem set H , define

$$B = (H^{-1}H) \cap U.$$

Then $B = B^{-1}$, $1 \in B$, and for every $\delta \leq r \leq r_0$,

$$\mathcal{N}(B, Cr) \geq c_M \mathcal{N}(H, r). \quad (10.24)$$

Moreover, for every fixed m ,

$$\mathcal{N}(B^m, C\delta) \leq \delta^{-C'_m \zeta} \mathcal{N}(B, C\delta). \quad (10.25)$$

Finally, if (10.3) and (10.7) hold and $\eta\kappa_t > C_0\zeta$, then for every maximal torus T there exists $b_T \in B$ with

$$d(b_T, T) > \delta^\eta. \quad (10.26)$$

Proof. For a fixed r , write $H_i = H \cap g_i V$. Since the H_i cover H ,

$$\mathcal{N}(H, r) \leq \sum_{i=1}^M \mathcal{N}(H_i, r),$$

so some H_i has covering number at least $M^{-1}\mathcal{N}(H, r)$. Choose $h_i \in H_i$. The left translate $h_i^{-1}H_i$ lies in $H^{-1}H \cap V^{-1}V \subset B$, and translation is an isometry. This proves (10.24), up to the fixed covering-radius constants.

Because $B \subset H^{-1}H = H^2$, one has $B^m \subset H^{2m}$. The fixed-power inverse-theorem bound (10.5), together with $\mathcal{N}(B, C\delta) \gg \mathcal{N}(H, \delta)$ from (10.24), gives (10.25).

For the torus assertion, suppose instead that $B \subset T^{(\delta^\eta)}$. For every nonempty H_i choose $h_i \in H_i$. Then every $h \in H_i$ satisfies $h_i^{-1}h \in B$, and therefore

$$H_i \subset h_i T^{(C\delta^\eta)}.$$

Consequently

$$aH \subset \bigcup_{i=1}^M ah_i T^{(C\delta^\eta)}.$$

By the torus-coset escape estimate,

$$(q * P_\delta)(aH) \leq CM\delta^{\eta\kappa_t},$$

contradicting (10.3) for sufficiently small δ when $\eta\kappa_t > C_0\zeta$. Hence (10.26) holds. $\square$

10.4 One-step compact L2 flattening

Theorem 10.6 (One-step discretized L^2 flattening). *Fix $\gamma > 0$. Assume the logarithmic-scale ball and torus-coset estimates (10.6)–(10.7) for the family of random-walk measures under consideration. Then there exists $\zeta = \zeta(\gamma, \kappa_b, \kappa_t) > 0$ such that, for all sufficiently small δ , whenever*

$$\|q * P_\delta\|_2 \geq \delta^{-\gamma}, \quad (10.27)$$

we have

$$\|q * q * P_\delta\|_2 \leq \delta^\zeta \|q * P_\delta\|_2. \quad (10.28)$$

Here q may be any logarithmic-length convolution power for which Interfaces I2–I4 hold.

Proof. If $\gamma \geq 3/2$, the claim is vacuous at sufficiently small scale by the trivial bound $\|q * P_\delta\|_2 \ll \delta^{-3/2}$, so assume $0 < \gamma < 3/2$. Set

$$\theta = 3 - \gamma, \quad \kappa = \min \left\{ \frac{\kappa_b}{2}, \frac{\theta}{2} \right\}.$$

Then $0 < \kappa < \theta < 3$. Let $\varepsilon_w = \varepsilon(\kappa, \theta)$ be the exponent from Theorem 10.1. Choose $\zeta > 0$ so small that

$$C_* \zeta < \min \left\{ \frac{\gamma}{4}, \frac{\varepsilon_w}{20}, \frac{\kappa_t \varepsilon_w}{20}, \frac{\kappa_b \varepsilon_w}{20} \right\}, \quad (10.29)$$

where C_* dominates C_0 and the finitely many fixed-power constants needed below.

Suppose (10.28) fails. Interface I2 supplies H, a, X satisfying (10.2)–(10.4). Form the local difference core $B = (H^{-1}H) \cap U$ of Lemma 10.5.

First, B has the required upper effective dimension. Lemma 10.4, (10.24), and $B \subset H^2$ give, after increasing the fixed constant in the radius,

$$\mathcal{N}(B, C\delta) \leq \delta^{-C_* \zeta} \mathcal{N}(H, C\delta) \leq \delta^{-3+2\gamma-C_* \zeta}.$$

By (10.29), this is at most $\delta^{-\theta-\varepsilon_w}$ for sufficiently small δ .

Second, the captured mass gives the all-scale lower bound. Lemma 10.3 and (10.24) imply, for $\delta \leq r \leq r_0$,

$$\mathcal{N}(B, Cr) \gg \delta^{C_0 \zeta} r^{-\kappa_b} \geq \delta^{\varepsilon_w} r^{-\kappa},$$

using $\kappa \leq \kappa_b$ and $C_0 \zeta < \varepsilon_w$. Third, Lemma 10.5 with $\eta = \varepsilon_w$ and (10.29) gives the required δ^{ε_w} -avoidance from every maximal torus.

Thus Theorem 10.1 applies to B (after the harmless fixed-radius renormalization) and yields

$$\mathcal{N}(B^3, C\delta) \geq \delta^{-\varepsilon_w} \mathcal{N}(B, C\delta). \quad (10.30)$$

On the other hand, the approximate-group power control in Lemma 10.5 gives

$$\mathcal{N}(B^3, C\delta) \leq \delta^{-C'_3 \zeta} \mathcal{N}(B, C\delta), \quad (10.31)$$

with $C'_3 \zeta < \varepsilon_w/2$ by construction. Equations (10.30) and (10.31) are incompatible for small δ . Therefore (10.28) holds. $\square$

Remark 10.7 (Endpoint logic). The proof shows explicitly why the two endpoint obstructions do not have to be inserted into the local-product chapter itself. The upper effective-support estimate keeps H a fixed power away from three-dimensional Haar size; captured mass plus ball non-concentration gives a fixed positive lower dimension. The weak product theorem then handles every intermediate effective dimension at once.

10.5 Propagation of the non-concentration inputs

The one-step theorem must be reusable after convolution length doubles. The following simple principle is the bookkeeping reason this causes no loss.

Lemma 10.8 (Set concentration cannot increase under further convolution). *Let μ, λ be probability measures on G . For every Borel set E ,*

$$(\mu * \lambda)(E) \leq \sup_{x \in G} \mu(Ex^{-1}). \quad (10.32)$$

Consequently, any uniform ball or left-coset tube bound for μ is inherited by $\mu * \lambda$ with the same exponent and constants up to fixed metric enlargements.

Proof. By Fubini,

$$(\mu * \lambda)(E) = \int_G \mu(Ey^{-1}) d\lambda(y) \leq \sup_y \mu(Ey^{-1}).$$

Right translation preserves ball radii for the bi-invariant metric. A translate of a left coset tube is again a coset tube of a conjugate maximal torus, so the same supremum argument applies to the torus family. $\square$

Corollary 10.9 (Power-scale propagation). *Suppose that for each $r \in [\delta, r_0]$ a prefix convolution power q_r of the random walk, of length $O(\log(1/r))$, satisfies ball and torus-coset bounds $\ll r^{\kappa_0}$. Then every longer convolution power q of length $O(\log(1/\delta))$ satisfies (10.6)–(10.7) uniformly for all $r \in [\delta, r_0]$, after decreasing κ_0 by a fixed factor if desired.*

Proof. Write $q = q_r * \lambda_r$. Because the ball kernel P_δ is conjugation-invariant, it is central in the convolution algebra, and hence $q * P_\delta = (q_r * P_\delta) * \lambda_r$. Apply Lemma 10.8 to the smoothed prefix. Smoothing enlarges an r -ball or r -tube only to radius $r + O(\delta) \asymp r$. $\square$

10.5.1 Comparison of neighboring smoothing scales

Lemma 10.10 (Smoothing-scale comparison). *There is an absolute C such that, for every finite positive measure μ and every $0 < \delta \leq \rho \leq r_0$,*

$$\|\mu * P_\rho\|_2 \leq C \|\mu * P_\delta\|_2, \quad \|\mu * P_\delta\|_2 \leq C \left(\frac{\rho}{\delta}\right)^3 \|\mu * P_\rho\|_2. \quad (10.33)$$

In particular, for $\rho = \delta^c$, $0 < c < 1$, information at one of the scales δ, δ^c propagates to the other with an explicit power loss depending only on c .

Proof. Choose a maximal δ -separated set $\{x_i\}_{i=1}^M$ in $B(1, \rho)$. By the three-dimensional local volume comparison, $M \asymp (\rho/\delta)^3$, the balls $B(x_i, 2\delta)$ cover $B(1, \rho)$, and their overlap is bounded after replacing 2δ by a fixed multiple of δ . Hence

$$P_\rho \leq \frac{C}{M} \sum_{i=1}^M L_{x_i} P_{C\delta}.$$

Convolving with μ , taking L^2 norms, and using translation invariance gives the first inequality after the harmless fixed-radius comparison between $P_{C\delta}$ and P_δ . For the second, the inclusions of balls and $m_G(B(1, r)) \asymp r^3$ give the pointwise kernel bound

$$P_\delta \leq C(\rho/\delta)^3 P_\rho.$$

Positivity is preserved by convolution with μ , so the same pointwise inequality holds after convolution; taking L^2 norms proves the claim. $\square$

10.6 Bounded-stage multiscale flattening

Theorem 10.11 (Bounded-stage flattening to the prescribed L^2 threshold). *Fix $\gamma > 0$. Let ν be the symmetric generating measure and suppose the Diophantine ball and torus-coset estimates are available at logarithmic lengths as in Corollary 10.9. There are constants $J = J(\gamma)$ and $C = C(\gamma)$ such that for every sufficiently small δ one can find*

$$\ell \leq C \log(1/\delta) \quad (10.34)$$

with

$$\|\nu^{(\ell)} * P_\delta\|_2 < \delta^{-\gamma}. \quad (10.35)$$

Moreover ℓ is obtained from an initial logarithmic length by at most J doublings.

Proof. Start with a logarithmic length $\ell_0 \asymp \log(1/\delta)$ large enough that the non-concentration inputs hold. The trivial smoothing bound is

$$\|\nu^{(\ell_0)} * P_\delta\|_2 \leq \|P_\delta\|_2 \asymp \delta^{-3/2}. \quad (10.36)$$

If this is already below $\delta^{-\gamma}$, stop. Otherwise Theorem 10.6 gives

$$\|\nu^{(2\ell_0)} * P_\delta\|_2 \leq \delta^\zeta \|\nu^{(\ell_0)} * P_\delta\|_2.$$

By Lemma 10.8 and Corollary 10.9, the hypotheses remain valid after each doubling. After j doublings, as long as the threshold has not been crossed,

$$\|\nu^{(2^j \ell_0)} * P_\delta\|_2 \ll \delta^{-3/2+j\zeta}.$$

Choose

$$J > \frac{3/2 - \gamma}{\zeta} + 1.$$

Then the threshold must be crossed for some $j \leq J$. Since J is independent of δ , $2^j \ell_0 \leq C(\gamma) \log(1/\delta)$, proving (10.34)–(10.35). $\square$

Corollary 10.12 (Uniform logarithmic-in-scale flattening). *For every fixed $\gamma > 0$ there is $C(\gamma)$ such that, for every sufficiently small δ , some $\ell \leq C(\gamma) \log(1/\delta)$ satisfies*

$$\|\nu^{(\ell)} * P_\delta\|_2 < \delta^{-\gamma}.$$

The constants are uniform over the scale once the algebraic generating set and the non-concentration exponents are fixed.

Proof. This is Theorem 10.11; the point of the separate statement is to record the exact form needed by the representation-theoretic spectral endpoint. $\square$

10.7 Proof architecture

The compact flattening argument can be summarized as

$$\begin{aligned} \text{fixed-dimension product growth} &\implies \text{uniform weak product growth} \\ &\implies \text{one-step } L^2 \text{ flattening} \\ &\implies \text{bounded-stage multiscale flattening.} \end{aligned}$$

The passage to weak product growth is what allows the effective dimension to vary across the multiscale decomposition. The resulting logarithmic-scale L^2 bound is precisely the input used by the representation-theoretic spectral endpoint.

Chapter 11

Internal Density of Matrix Coefficients

For completeness we prove, without invoking Stone–Weierstrass, the uniform density statement used in the representation-theoretic chapter. The proof identifies $SU(2)$ with S^3 , proves a multivariable Bernstein approximation theorem directly, and observes that polynomial coordinate functions restrict to finite linear combinations of irreducible matrix coefficients.

11.1 The concrete model of $SU(2)$

Write

$$G = SU(2) = \left\{ \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} : |a|^2 + |b|^2 = 1 \right\}.$$

If $a = x_1 + ix_2$ and $b = x_3 + ix_4$, the map

$$G \longrightarrow S^3 \subset \mathbb{R}^4, \quad g \longmapsto (x_1, x_2, x_3, x_4)$$

is a homeomorphism. The four real coordinate functions are complex linear combinations of $a, \bar{a}, b, \bar{b}$, hence of matrix coefficients of the standard representation and its dual.

Lemma 11.1 (Products of coefficients are coefficients). *Let $c_{u,v}^\pi(g) = \langle \pi(g)u, v \rangle$ and $c_{u',v'}^\sigma(g) = \langle \sigma(g)u', v' \rangle$ be matrix coefficients of finite-dimensional unitary representations. Then*

$$c_{u,v}^\pi(g) c_{u',v'}^\sigma(g) = c_{u \otimes u', v \otimes v'}^{\pi \otimes \sigma}(g).$$

Complex conjugates of coefficients are coefficients of dual representations.

Proof. The tensor-product Hermitian form gives

$$\langle (\pi \otimes \sigma)(g)(u \otimes u'), v \otimes v' \rangle = \langle \pi(g)u, v \rangle \langle \sigma(g)u', v' \rangle.$$

The dual-representation assertion follows from unitarity and the usual identification of a conjugate coefficient with the corresponding dual coefficient. $\square$

Lemma 11.2 (Complete reducibility for finite-dimensional unitary representations). *Every finite-dimensional unitary representation of a group is an orthogonal direct sum of irreducible subrepresentations.*

Proof. Induct on the dimension. If the representation is irreducible there is nothing to prove. Otherwise choose a nonzero proper invariant subspace W . Unitarity makes $W^\perp$ invariant: for $v \in W^\perp$, $w \in W$, and g in the group,

$$\langle gv, w \rangle = \langle v, g^{-1}w \rangle = 0.$$

Apply the induction hypothesis to W and $W^\perp$. $\square$

Corollary 11.3 (Coordinate polynomials are finite-dimensional coefficients). *Every complex polynomial in x_1, x_2, x_3, x_4 , restricted to $S^3 \cong SU(2)$, is a finite linear combination of matrix coefficients of finite-dimensional unitary representations of $SU(2)$. By Lemma 11.2, it is therefore a finite linear combination of irreducible matrix coefficients.*

Proof. Each coordinate is a linear combination of coefficients of the standard representation and its dual. Apply Lemma 11.1 to monomials and then take finite linear combinations. Lemma 11.2 splits every finite-dimensional unitary representation into irreducibles, so each resulting coefficient is a finite sum of irreducible coefficients. $\square$

11.2 An internal multivariable Bernstein theorem

We first prove the approximation result needed below rather than cite Stone–Weierstrass or a general polynomial approximation theorem.

For a continuous function $F : [0, 1]^d \rightarrow \mathbb{C}$, define

$$B_n F(x) = \sum_{k_1=0}^n \cdots \sum_{k_d=0}^n F\left(\frac{k_1}{n}, \dots, \frac{k_d}{n}\right) \prod_{j=1}^d \binom{n}{k_j} x_j^{k_j} (1-x_j)^{n-k_j}.$$

This is an ordinary polynomial in $x_1, \dots, x_d$.

Theorem 11.4 (Multivariable Bernstein approximation, proved internally). *For every $F \in C([0, 1]^d)$,*

$$\|B_n F - F\|_\infty \longrightarrow 0.$$

Proof. Fix $x = (x_1, \dots, x_d)$. Let $K_1, \dots, K_d$ be independent random variables with $K_j \sim \text{Bin}(n, x_j)$. Then the defining formula is exactly

$$B_n F(x) = \mathbb{E} F(K_1/n, \dots, K_d/n).$$

Let $M = \|F\|_\infty$. Given $\varepsilon > 0$, uniform continuity supplies $\eta > 0$ such that $|F(y) - F(x)| < \varepsilon/2$ whenever $\|y - x\|_\infty < \eta$. For each j ,

$$\mathbb{E}(K_j/n) = x_j, \quad \text{Var}(K_j/n) = \frac{x_j(1-x_j)}{n} \leq \frac{1}{4n}.$$

For each j , the elementary estimate

$$\mathbb{P}(|K_j/n - x_j| \geq \eta) \leq \frac{\text{Var}(K_j/n)}{\eta^2}$$

follows from $\eta^2 1_{\{|K_j/n - x_j| \geq \eta\}} \leq |K_j/n - x_j|^2$. The union bound therefore gives, uniformly in x ,

$$\mathbb{P}\left(\max_j |K_j/n - x_j| \geq \eta\right) \leq \frac{d}{4n\eta^2}.$$

Split the expectation into this exceptional event and its complement. Then

$$|B_n F(x) - F(x)| \leq \frac{\varepsilon}{2} + 2M \frac{d}{4n\eta^2}.$$

For n sufficiently large the second term is at most $\varepsilon/2$, uniformly in x . Hence $\|B_n F - F\|_\infty \leq \varepsilon$ for all sufficiently large n . $\square$

Corollary 11.5 (Polynomial approximation on a cube). *If $Q = [-1, 1]^d$ and $F \in C(Q)$, then for every $\varepsilon > 0$ there is a complex polynomial p in d real variables with*

$$\sup_{x \in Q} |F(x) - p(x)| < \varepsilon.$$

Proof. Transport Q affinely to $[0, 1]^d$, apply Theorem 11.4, and transport the resulting polynomial back. $\square$

11.3 Radial extension from the sphere

Lemma 11.6 (Elementary continuous extension from S^3). *Every $f \in C(S^3)$ has an explicit continuous extension F to the cube $[-1, 1]^4$.*

Proof. Choose a continuous cutoff $\psi : [0, \infty) \rightarrow [0, 1]$ with $\psi(t) = 0$ for $t \leq 1/3$ and $\psi(t) = 1$ for $t \geq 2/3$. Put

$$F(x) = \begin{cases} \psi(\|x\|)f(x/\|x\|), & x \neq 0, \\ 0, & x = 0. \end{cases}$$

For $\|x\| \leq 1/3$ this is identically zero, so continuity at the origin is immediate; away from the origin it is a composition of continuous maps. On S^3 one has $\psi(1) = 1$, so $F|_{S^3} = f$. $\square$

Theorem 11.7 (Internal density of irreducible matrix coefficients). *Finite linear combinations of irreducible matrix coefficients of $SU(2)$ are uniformly dense in $C(SU(2))$.*

Proof. Identify $SU(2)$ with S^3 . Let $f \in C(S^3)$ and $\varepsilon > 0$. Extend f to $F \in C([-1, 1]^4)$ by Lemma 11.6. Corollary 11.5 supplies a polynomial $p(x_1, x_2, x_3, x_4)$ satisfying

$$\sup_{[-1, 1]^4} |F - p| < \varepsilon.$$

Restricting to S^3 gives

$$\|f - p|_{S^3}\|_\infty < \varepsilon.$$

By Corollary 11.3, $p|_{S^3}$ is a finite linear combination of irreducible matrix coefficients. Since f and ε were arbitrary, the desired span is uniformly dense. $\square$

Remark 11.8 (Why prove Bernstein approximation directly?). Stone–Weierstrass would imply the same density statement, but the explicit Bernstein proof keeps the approximation step self-contained and quantitative enough for the exact representation-theory interface used here.

Chapter 12

$SU(2)$ Representation Theory and the Spectral Endpoint

This chapter develops the representation theory needed to convert logarithmic-scale flattening into a spectral gap. We classify the irreducible representations of $SU(2)$, prove the character and orthogonality formulas, establish the required Plancherel decomposition, and use high-frequency multiplicity in a Hilbert–Schmidt argument. The result is first stated for the model algebraic free walk; the final chapter transfers it to an arbitrary algebraic dense generating set.

12.1 The irreducible representations of $SU(2)$

Write

$$G = SU(2) = \left\{ \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} : |a|^2 + |b|^2 = 1 \right\}.$$

Let V_m be the vector space of homogeneous polynomials of degree m in two variables,

$$V_m = \text{Sym}^m(\mathbb{C}^2), \quad \dim V_m = m + 1,$$

with the natural action induced from the standard representation of $SU(2)$. We use the unitary structure induced from the tensor power of the standard Hermitian form.

Proposition 12.1 (Weights of the symmetric powers). *For*

$$t_\theta = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix},$$

the monomials $z_1^{m-j} z_2^j$, $0 \leq j \leq m$, are weight vectors of weights $m - 2j$. Thus the weights of V_m are

$$m, m - 2, \dots, -m,$$

each with multiplicity one.

Proof. The standard basis vectors have torus weights 1 and -1 . Tensoring m times and then passing to the symmetric power adds the weights. Explicitly,

$$t_\theta \cdot z_1^{m-j} z_2^j = e^{i(m-2j)\theta} z_1^{m-j} z_2^j.$$

□

Let $\mathfrak{sl}_2(\mathbb{C})$ have the standard basis

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad E = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad F = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

with

$$[H, E] = 2E, \quad [H, F] = -2F, \quad [E, F] = H.$$

Every finite-dimensional continuous representation of $SU(2)$ differentiates and complexifies to a representation of $\mathfrak{sl}_2(\mathbb{C})$.

Lemma 12.2 (Highest-weight string). *Let V be a finite-dimensional $\mathfrak{sl}_2(\mathbb{C})$ -module, and suppose*

$$Hv = nv, \quad Ev = 0,$$

with $v \neq 0$. Then for every $j \geq 1$,

$$EF^jv = j(n - j + 1)F^{j-1}v. \quad (12.1)$$

If $F^rv \neq 0$ and $F^{r+1}v = 0$, then $n = r \in \mathbb{Z}_{\geq 0}$, and $v, Fv, \dots, F^rv$ are nonzero vectors of distinct weights $n, n-2, \dots, -n$.

Proof. The commutator identity

$$[E, F^j] = \sum_{k=0}^{j-1} F^k [E, F] F^{j-1-k}$$

and $HF^qv = (n - 2q)F^qv$ give (12.1) by summing the resulting arithmetic progression. Putting $j = r + 1$ and using $F^{r+1}v = 0$ yields

$$0 = (r + 1)(n - r)F^rv,$$

so $n = r$. Distinct H -weights are linearly independent, and the displayed formula prevents an earlier vector in the string from vanishing. $\square$

Theorem 12.3 (Classification of irreducible $SU(2)$ representations). *Every finite-dimensional irreducible complex representation of $SU(2)$ is isomorphic to exactly one $V_m = \text{Sym}^m(\mathbb{C}^2)$, $m \geq 0$. In particular,*

$$\dim V_m = m + 1.$$

Proof. First V_m is irreducible. Indeed the torus acts unitarily, so any nonzero invariant subspace contains a torus weight vector: average a nonzero vector against a suitable torus character to project onto one of its nonzero weight components. Applying the raising and lowering operators E, F to any nonzero weight vector in V_m reaches the highest-weight vector and then, by Lemma 12.2, the entire weight string. Hence the invariant subspace is all of V_m .

Conversely let V be irreducible. Restriction to the compact torus is a finite-dimensional unitary representation of the circle, hence decomposes into integral weight spaces. Choose a maximal weight n and a nonzero vector v of that weight. Then $Ev = 0$, because E raises weight by 2. Since V is finite-dimensional, $F^{r+1}v = 0$ for some minimal r . Lemma 12.2 gives $n = r \geq 0$, and the span

$$W = \text{span}\{v, Fv, \dots, F^rv\}$$

is stable under H, E, F . It is therefore stable under every exponential $\exp X$; exponentials from an identity neighborhood generate the connected group $SU(2)$, and irreducibility gives $W = V$. The action on this basis has the same highest-weight string as on V_n , so $V \simeq V_n$. Distinct n have distinct highest torus weights, hence are nonisomorphic. $\square$

12.2 The character formula

Theorem 12.4 (Dimension and explicit character formula). *Every $g \in SU(2)$ is conjugate to a unique t_θ with $0 \leq \theta \leq \pi$. For the irreducible V_m ,*

$$\chi_m(t_\theta) = \sum_{j=0}^m e^{i(m-2j)\theta} = \frac{\sin((m+1)\theta)}{\sin \theta}, \quad (12.2)$$

where the quotient is interpreted continuously at $\theta = 0, \pi$. In particular $\chi_m(1) = m+1$.

Proof. A unitary matrix is unitarily diagonalizable; determinant one forces its eigenvalues to be $e^{i\theta}, e^{-i\theta}$, and choosing $0 \leq \theta \leq \pi$ makes the conjugacy parameter unique. Proposition 12.1 gives the finite geometric sum in (12.2). Multiplying by $e^{im\theta}$ and summing the ratio $e^{-2i\theta}$ gives

$$\sum_{j=0}^m e^{i(m-2j)\theta} = \frac{e^{i(m+1)\theta} - e^{-i(m+1)\theta}}{e^{i\theta} - e^{-i\theta}},$$

which is the sine quotient. The endpoint values follow by continuity, or directly from the weights. $\square$

Example 12.5 (The first nontrivial character calculation). For $m = 2$, the irreducible representation V_2 has dimension 3 and weights $2, 0, -2$. Hence

$$\chi_2(t_\theta) = e^{2i\theta} + 1 + e^{-2i\theta} = 1 + 2\cos(2\theta).$$

At $\theta = \pi/6$ this gives

$$\chi_2(t_{\pi/6}) = 1 + 2\cos(\pi/3) = 2, \quad \frac{\chi_2(t_{\pi/6})}{\dim V_2} = \frac{2}{3}.$$

The normalized character has already dropped a definite distance below 1 at a fixed angular scale. For large m , the next lemma quantifies the complementary small-angle regime $\theta \lesssim (m+1)^{-1}$ that is needed in the spectral endpoint.

Lemma 12.6 (Normalized characters near the identity). *There are absolute $c, C > 0$ such that, whenever $0 \leq \theta \leq c/(m+1)$,*

$$\frac{\chi_m(t_\theta)}{m+1} \geq 1 - C(m+1)^2\theta^2. \quad (12.3)$$

Proof. The character is real and, by the weight sum,

$$\frac{\chi_m(t_\theta)}{m+1} = \frac{1}{m+1} \sum_{j=0}^m \cos((m-2j)\theta).$$

Using $\cos x \geq 1 - x^2/2$ and $|m-2j| \leq m$ gives the claim. $\square$

12.3 Schur orthogonality and density of coefficients

For a unitary representation π and vectors u, v , write $c_{u,v}^\pi(g) = \langle \pi(g)u, v \rangle$.

Lemma 12.7 (Averaging an operator). *Let $\pi : G \rightarrow U(V)$ and $\sigma : G \rightarrow U(W)$ be irreducible. For $T : W \rightarrow V$ define*

$$\mathcal{A}(T) = \int_G \pi(g)T\sigma(g)^{-1} dm_G(g).$$

Then $\mathcal{A}(T)$ intertwines σ with π . Hence it is zero when $\pi \not\cong \sigma$ and is a scalar multiple of the identity when $\pi = \sigma$.

Proof. Haar invariance gives $\pi(h)\mathcal{A}(T) = \mathcal{A}(T)\sigma(h)$. The conclusion is Schur's lemma; its elementary proof is included here for completeness. If an intertwiner has an eigenvalue λ , its kernel after subtracting λI is a nonzero invariant subspace; irreducibility makes it all of V . For nonisomorphic irreducibles, the kernel and image of a nonzero intertwiner would force an isomorphism. $\square$

Theorem 12.8 (Schur orthogonality). *Choose orthonormal bases and write $\pi_{ij}(g)$ for the corresponding matrix coefficients. For irreducibles π, σ ,*

$$\int_G \pi_{ij}(g) \overline{\sigma_{k\ell}(g)} dm_G(g) = \begin{cases} d_\pi^{-1} \delta_{ik} \delta_{j\ell}, & \pi \simeq \sigma, \\ 0, & \pi \not\simeq \sigma. \end{cases} \quad (12.4)$$

Proof. Apply Lemma 12.7 to rank-one maps and evaluate matrix entries. In the case $\pi = \sigma$, the averaged operator is scalar; taking traces determines the scalar as d_π^{-1} times the appropriate matrix entry of the original rank-one map. This gives (12.4). $\square$

Theorem 12.9 (Internal density of matrix coefficients). *Finite linear combinations of irreducible matrix coefficients are uniformly dense in $C(SU(2))$.*

Proof. This is Theorem 11.7 of the preceding chapter, where the density is proved directly by multivariable Bernstein approximation on the concrete S^3 model. $\square$

12.4 Peter–Weyl and Plancherel in the exact form used later

For $f \in L^1(G)$ define

$$\widehat{f}(m) = \int_G f(g) \pi_m(g^{-1}) dm_G(g) \in \text{End}(V_m).$$

Theorem 12.10 (Peter–Weyl and Plancherel for $SU(2)$). *The functions*

$$\sqrt{m+1} \pi_{ij}^{(m)}, \quad m \geq 0, \quad 1 \leq i, j \leq m+1,$$

form an orthonormal basis of $L^2(SU(2))$. Consequently, for $f \in L^2(G)$,

$$\|f\|_2^2 = \sum_{m=0}^{\infty} (m+1) \|\widehat{f}(m)\|_{\text{HS}}^2, \quad (12.5)$$

and whenever convolution is defined,

$$\widehat{f * h}(m) = \widehat{h}(m) \widehat{f}(m) \quad (12.6)$$

for the stated transform convention.

Proof. Schur orthogonality gives orthonormality. Theorem 12.9 gives uniform, hence L^2 , density of their span, so they form an orthonormal basis. Expanding f in that basis and applying Parseval gives (12.5). Formula (12.6) follows by Fubini and the substitution $x = yz$. $\square$

Corollary 12.11 (High-frequency multiplicity). *In the left regular representation, the irreducible V_m occurs with multiplicity $m+1$. Equivalently, if a convolution operator has a singular value s in its V_m Fourier block, then its Hilbert–Schmidt mass on $L^2(G)$ receives a contribution at least*

$$(m+1)s^2.$$

Proof. For each fixed column j , the span of the coefficients $\pi_{ij}^{(m)}$ over i is a copy of V_m under left translation, and the $m+1$ columns are orthogonal by Theorem 12.8. The contribution statement is the m -th nonnegative term in (12.5). $\square$

12.5 The smoothing kernel in a high-frequency block

Let P_δ be the normalized indicator of the metric ball $B(1, \delta)$, as in the flattening chapter. The metric is bi-invariant, so P_δ is conjugation invariant.

Lemma 12.12 (Small balls act almost as the identity). *There is an absolute $c_* > 0$ such that, for*

$$\delta \leq \frac{c_*}{m+1},$$

one has

$$\widehat{P}_\delta(m) = \alpha_m(\delta) \text{Id}_{V_m}, \quad \alpha_m(\delta) \geq \frac{3}{4}. \quad (12.7)$$

Proof. Conjugation invariance makes $\widehat{P}_\delta(m)$ an intertwiner of the irreducible V_m , so it is scalar by Lemma 12.7. Taking traces gives

$$\alpha_m(\delta) = \frac{1}{m+1} \int_G P_\delta(g) \chi_m(g^{-1}) dm_G(g).$$

If $g \in B(1, \delta)$, its conjugacy angle satisfies $|\theta(g)| \leq C\delta$ in a fixed identity chart. Lemma 12.6 therefore gives

$$\frac{\chi_m(g^{-1})}{m+1} \geq 1 - C'(m+1)^2 \delta^2.$$

Choose c_* so that the final error is at most $1/4$ and average over the ball. $\square$

12.6 The Hilbert–Schmidt spectral endgame

Theorem 12.13 (Flattening implies a uniform gap for the model walk). *Let ν be a symmetric finitely supported probability measure whose support generates a dense subgroup of $SU(2)$. Assume that the Module- H logarithmic flattening conclusion holds for this walk: for some fixed $0 < \gamma < 1/2$ and $C_H < \infty$, every sufficiently small δ admits an integer $\ell \leq C_H \log(1/\delta)$ such that*

$$\|\nu^{(\ell)} * P_\delta\|_2 \leq \delta^{-\gamma}. \quad (12.8)$$

Then there is $\eta > 0$ such that for every nontrivial irreducible V_m ,

$$\|\widehat{\nu}(m)\|_{\text{op}} \leq 1 - \eta. \quad (12.9)$$

Hence convolution by ν has a spectral gap on $L_0^2(SU(2))$.

Proof. Because ν is symmetric, every $\widehat{\nu}(m)$ is self-adjoint. Fix a large m and put

$$d_m = m+1, \quad \delta = \frac{c_*}{d_m},$$

with c_* from Lemma 12.12. Choose ℓ by (12.8) and let λ be an eigenvalue of $\widehat{\nu}(m)$ with $|\lambda| = \|\widehat{\nu}(m)\|_{\text{op}}$. By Theorem 12.10 and Lemma 12.12,

$$\delta^{-2\gamma} \geq \|\nu^{(\ell)} * P_\delta\|_2^2 \quad (12.10)$$

$$\geq d_m \|\widehat{\nu^{(\ell)} * P_\delta(m)}\|_{\text{HS}}^2 \quad (12.11)$$

$$\geq d_m \left(\frac{3}{4}\right)^2 |\lambda|^{2\ell}. \quad (12.12)$$

Since $\delta \asymp d_m^{-1}$,

$$|\lambda|^{2\ell} \leq C d_m^{-(1-2\gamma)}. \quad (12.13)$$

Also $\ell \leq C_H \log(d_m/c_*) \leq 2C_H \log d_m$ for large m . If $|\lambda| \geq e^{-\eta_\infty}$, then

$$|\lambda|^{2\ell} \geq d_m^{-4C_H\eta_\infty}.$$

Choosing

$$\eta_\infty < \frac{1-2\gamma}{8C_H}$$

contradicts (12.13) for all sufficiently large m . Thus the high-frequency blocks have a uniform absolute gap.

It remains to handle finitely many m . For each nontrivial irreducible, density of the subgroup generated by $\text{supp } \nu$ implies $\|\widehat{\nu}(m)\|_{\text{op}} < 1$. Indeed equality in the triangle inequality for an average of unitaries would give a common invariant line on which every generator acts by the same unit scalar; using the inverses and density would produce a one-dimensional subrepresentation of $SU(2)$. The only continuous one-dimensional representation of $SU(2)$ is trivial: its differential vanishes on $[\mathfrak{su}(2), \mathfrak{su}(2)] = \mathfrak{su}(2)$, and connectedness then forces the character itself to be constant. This contradicts $m > 0$. Taking the minimum gap over this finite set and the high-frequency gap proves (12.9).

Finally Peter–Weyl decomposes $L_0^2(G)$ into these nontrivial blocks, so their uniform operator bound is exactly the spectral gap for convolution by ν . $\square$

Remark 12.14 (Why the multiplicity is decisive). The flattening estimate in the flattening chapter may be extremely weak: any fixed $\gamma < 1/2$ suffices. At frequency m , however, a near-unit eigenvalue is repeated with multiplicity $m+1$. Taking $\delta \asymp (m+1)^{-1}$ turns that multiplicity into the factor δ^{-1} , which beats the flattening loss $\delta^{-2\gamma}$ and forces a uniform spectral contraction.

Corollary 12.15 (Uniform exponential L^2 mixing for the model walk). *Under the hypotheses of Theorem 12.13, there are $C < \infty$ and $c > 0$ such that for every $f \in L^2(G)$,*

$$\left\| \nu^{(n)} * f - \int_G f dm_G \right\|_2 \leq C e^{-cn} \left\| f - \int_G f dm_G \right\|_2. \quad (12.14)$$

In particular, for every fixed $r > 0$,

$$\|\nu^{(n)} * P_r - 1\|_2 \leq C e^{-cn} \|P_r - 1\|_2.$$

Proof. Theorem 12.13 gives operator norm at most $1 - \eta$ on $L_0^2(G)$. Iterate and use $(1 - \eta)^n \leq e^{-\eta n}$. $\square$

12.7 Scope of the model spectral theorem

Theorem 12.13 gives a uniform spectral gap for the algebraic free-pair model walk once the logarithmic flattening/nonconcentration package has been established. The final chapter constructs such a bounded-word model walk from an arbitrary finite algebraic dense symmetric generating set and compares its Dirichlet form with that of the original walk.

Chapter 13

Arbitrary Algebraic Dense Generators

We remove the model-walk hypothesis. Starting from an arbitrary finite symmetric algebraic dense set, we pass to a suitable algebraic place, construct a proximal word, move its attracting and repelling points to obtain ping-pong, and thereby produce a bounded-word free dense pair. The compact nonconcentration, flattening, and spectral results apply to that model walk, and a direct Dirichlet-form comparison transfers the gap back to the original generators.

13.1 The target and the strategy

Let $S \subset SU(2)$ be a finite symmetric set, every matrix entry of every $s \in S$ algebraic, and assume

$$\overline{\langle S \rangle} = SU(2).$$

Write $\Gamma = \langle S \rangle$. We shall construct elements $a, b \in \Gamma$ with the following properties:

- (i) a, b freely generate a nonabelian free subgroup;
- (ii) $\overline{\langle a, b \rangle} = SU(2)$;
- (iii) a, b still have algebraic entries;
- (iv) a and b are words of some finite length $L = L(S)$ in S .

The symmetric four-point walk

$$\nu = \frac{1}{4}(\delta_a + \delta_{a^{-1}} + \delta_b + \delta_{b^{-1}})$$

then satisfies the previously proved algebraic free-pair nonconcentration inputs, so the flattening and representation-theory chapters give a spectral gap for ν . A direct Dirichlet-form comparison transfers that gap back to the uniform S -walk.

The only genuinely new issue is the construction of the free pair. Compact $SU(2)$ itself has no proximal elements in its standard complex representation. Algebraicity supplies a different place of the number field at which the same matrix group is unbounded. At that place, rank-one projective dynamics becomes proximal and elementary ping-pong applies.

13.2 Zariski density and change of field

Lemma 13.1 (The compact real form is complex Zariski dense). *The subset $SU(2) \subset SL_2(\mathbb{C})$ is Zariski dense. Consequently Γ is Zariski dense in SL_2 over the number field generated by its matrix entries.*

Proof. Let H be the complex Zariski closure of $SU(2)$ in $SL_2(\mathbb{C})$. Its Lie algebra $\mathfrak{h}$ is a complex Lie subalgebra of $\mathfrak{sl}_2(\mathbb{C})$ containing the real tangent space $\mathfrak{su}(2)$ at the identity. But the complex span of $\mathfrak{su}(2)$ is all of $\mathfrak{sl}_2(\mathbb{C})$. Hence $\mathfrak{h} = \mathfrak{sl}_2(\mathbb{C})$, so the identity component of H has the same dimension as $SL_2(\mathbb{C})$ and therefore equals it. Thus $H = SL_2(\mathbb{C})$.

Since Γ is analytically dense in $SU(2)$, every polynomial vanishing on Γ vanishes on $SU(2)$ and hence on $SL_2(\mathbb{C})$. This proves the claim. $\square$

Lemma 13.2 (Zariski density survives scalar extension). *Let $K \subset L$ be fields and $X \subset K^n$. If X is Zariski dense over K in an affine K -variety V , then its image is Zariski dense in V_L over L .*

Proof. Fix a degree bound d , and let W_d be the finite-dimensional K -space of regular functions on V represented by polynomials of degree at most d . Zariski density says that the evaluation map

$$W_d \longrightarrow K^X, \quad f \longmapsto (f(x))_{x \in X},$$

is injective. Because W_d is finite-dimensional, there are finitely many points $x_1, \dots, x_m \in X$ for which the finite evaluation map

$$E_d : W_d \longrightarrow K^m, \quad f \longmapsto (f(x_1), \dots, f(x_m))$$

is already injective: choose a basis of W_d and successively add evaluation rows until the row rank is $\dim_K W_d$.

After scalar extension, the same evaluation matrix has the same nonzero maximal minor, now viewed in L . Hence

$$E_d \otimes_K L : W_d \otimes_K L \longrightarrow L^m$$

is injective. The degree- $\leq d$ filtered piece of the coordinate ring also commutes with this scalar extension, so every degree- $\leq d$ regular function on V_L is represented in $W_d \otimes_K L$. Thus no nonzero such function vanishes on all of X . Since d was arbitrary, the Zariski closure of X in V_L is all of V_L . $\square$

13.3 Some algebraic place is unbounded

Let K be a number field containing all entries of S . For every place v of K , write K_v for the completion and Γ_v for the same matrix words viewed inside $SL_2(K_v)$.

Lemma 13.3 (Bounded at every place would force finiteness). *There exists a place v of K for which Γ_v is unbounded.*

Proof. Suppose instead that Γ_v is bounded for every v . At all but finitely many nonarchimedean places, every entry of every $s \in S \cup S^{-1}$ is integral; hence $\Gamma_v \subset SL_2(\mathcal{O}_v)$ there. Enlarge a finite set of places Σ so that it contains all archimedean places and every finite place where some generator is nonintegral.

For $v \in \Sigma$, boundedness of Γ_v gives a constant C_v such that every entry x of every $\gamma \in \Gamma$ satisfies $|x|_v \leq C_v$. Outside Σ we have $|x|_v \leq 1$. The finite-place inequalities say that all such x lie in one fixed fractional ideal $I \subset K$: indeed, at each of the finitely many bad prime ideals they give a fixed lower bound on the valuation, while elsewhere the valuation is nonnegative.

Under the Minkowski embedding

$$I \hookrightarrow \mathbb{R}^{r_1} \times \mathbb{C}^{r_2},$$

I is a lattice. The archimedean bounds $|x|_v \leq C_v$ confine every matrix entry to a compact box. A lattice meets a compact set in only finitely many points. Thus there are only finitely many possibilities for each of the four matrix entries of an element of Γ , so Γ itself is finite. This contradicts its density in the infinite group $SU(2)$. $\square$

By Lemma 13.2, Γ_v is still Zariski dense in SL_2 over K_v .

13.4 An unbounded Zariski-dense rank-one group contains a proximal word

We first define the dynamical term used in the strategy above. Let k be a local field and let k'/k be a finite extension. An element $g \in SL_2(k)$ is called *proximal over k'* if its action on $\mathbb{P}^1(k')$ has two distinct fixed points g^+, g^- and, after diagonalization over k' , its two eigenvalues have different absolute values. The fixed point corresponding to the eigenvalue of larger modulus is attracting and the other is repelling. Passing from k to k' enlarges only the projective space on which ping-pong is performed; the group elements themselves remain in $SL_2(k)$.

Lemma 13.4 (Large trace is proximal after at most a quadratic extension). *Let k be a characteristic-zero local field and $g \in SL_2(k)$. Suppose*

$$|\operatorname{tr} g|_k > 2$$

when k is archimedean, and $|\operatorname{tr} g|_k > 1$ when k is nonarchimedean. If k'/k is a splitting field of $X^2 - (\operatorname{tr} g)X + 1$, then $[k' : k] \leq 2$ and the two eigenvalues $\alpha, \alpha^{-1} \in k'$ have different absolute values. Hence g is proximal over k' .

Proof. The splitting field has degree at most two. Extend the absolute value of k to k' . If the reciprocal eigenvalues had equal absolute value, both would have absolute value one. Since

$$\operatorname{tr} g = \alpha + \alpha^{-1},$$

the ordinary triangle inequality would give $|\operatorname{tr} g| \leq 2$ at an archimedean place, while the ultrametric inequality would give $|\operatorname{tr} g| \leq 1$ at a nonarchimedean place. Both contradict the hypothesis. Thus, after relabelling, $|\alpha| > 1 > |\alpha^{-1}|$, and the two eigenlines give the attracting and repelling fixed points. $\square$

Lemma 13.5 (Projective compactness produces a proximal word). *Let k be a local field of characteristic zero and let $\Lambda \leq SL_2(k)$ be unbounded and Zariski dense. Then there are a finite extension k'/k of degree at most two and an element $g \in \Lambda$ that is proximal over k' .*

Proof. Choose $\gamma_n \in \Lambda$ escaping every compact subset. Projective matrix space $\mathbb{P}(M_2(k))$ is compact. Equivalently, divide γ_n by an entry of maximal absolute value and pass to a subsequence. Thus there are $c_n \in k^\times$ and a nonzero matrix M such that

$$c_n \gamma_n \longrightarrow M, \quad |c_n| \longrightarrow 0.$$

Because $\det(c_n \gamma_n) = c_n^2 \rightarrow 0$ while $M \neq 0$, the limit M has rank one.

Write $M = u \otimes \varphi$ with $u \neq 0$ and $\varphi \neq 0$. Choose $v \in k^2$ with $\varphi(v) \neq 0$. There is $h_0 \in SL_2(k)$ carrying u to v : extend u and v to bases and rescale the second basis vector to make the change-of-basis determinant one. Hence

$$\operatorname{tr}(h_0 M) = \varphi(h_0 u) = \varphi(v) \neq 0.$$

So the regular function $h \mapsto \operatorname{tr}(hM)$ is not identically zero on SL_2 . Zariski density supplies $h \in \Lambda$ with $\operatorname{tr}(hM) \neq 0$. For $g_n = h\gamma_n$ we then have

$$c_n \operatorname{tr}(g_n) \longrightarrow \operatorname{tr}(hM) \neq 0,$$

and therefore $|\operatorname{tr}(g_n)| \rightarrow \infty$. Choose n beyond the threshold of Lemma 13.4 and let k' split the characteristic polynomial of g_n . That lemma gives the required proximal element. Notice that no p -adic logarithm or square-root theorem is used. $\square$

Lemma 13.6 (North–south dynamics). *Let $g \in SL_2(k)$ be proximal with attracting and repelling fixed points g^+, g^- . Given disjoint neighborhoods U^+ of g^+ and U^- of g^- in $\mathbb{P}^1(k)$, there is N_0 such that for $N \geq N_0$,*

$$g^N(\mathbb{P}^1(k) \setminus U^-) \subset U^+, \quad g^{-N}(\mathbb{P}^1(k) \setminus U^+) \subset U^-.$$

Proof. Conjugate g so that $g^+ = [1 : 0]$, $g^- = [0 : 1]$, and $g = \text{diag}(\lambda, \lambda^{-1})$ with $|\lambda| > 1$. In the affine coordinate $z = y/x$ away from g^- , g^N acts by $z \mapsto \lambda^{-2N}z$. Uniform contraction on compact subsets away from g^- gives the first inclusion; the second follows by applying the same argument to g^{-1} . $\square$

13.5 Move the fixed points and apply ping–pong

Lemma 13.7 (A Zariski-dense group moves a pair off itself). *Let $\Lambda \leq SL_2(k)$ be Zariski dense and let $x \neq y$ in $\mathbb{P}^1(\bar{k})$. There is $r \in \Lambda$ such that*

$$\{rx, ry\} \cap \{x, y\} = \emptyset.$$

Proof. For fixed $u, v \in \{x, y\}$, the condition $ru = v$ cuts out a proper algebraic subset of SL_2 (after extending scalars to a field containing x, y). The union of the four such subsets is still proper. By scalar-extension Zariski density, Λ is not contained in it. $\square$

Lemma 13.8 (Rank-one ping–pong). *Let g be proximal and choose r as in Lemma 13.7 for $x = g^+$, $y = g^-$. For all sufficiently large N , the two elements*

$$a = g^N, \quad b = rg^Nr^{-1}$$

freely generate a nonabelian free group.

Proof. Choose pairwise disjoint neighborhoods

$$U_a^+, U_a^-, U_b^+, U_b^-$$

of g^+, g^-, rg^+, rg^- , respectively. By Lemma 13.6, after increasing N ,

$$a^{\pm 1}(\mathbb{P}^1(k) \setminus U_a^\mp) \subset U_a^\pm, \quad b^{\pm 1}(\mathbb{P}^1(k) \setminus U_b^\mp) \subset U_b^\pm.$$

Pick a point outside all four neighborhoods. Reading a nonempty reduced word from right to left, its last letter puts the point into the attracting domain belonging to that letter, and every subsequent reduced letter sends it into its own disjoint attracting domain. The result cannot equal the starting point. Thus no nonempty reduced word is the identity. $\square$

Applying Lemma 13.3 and then Lemma 13.5 to Γ_v gives a proximal word $g \in \Gamma$ over a field extension k'/K_v of degree at most two. Scalar-extension Zariski density lets Lemma 13.7 move its attracting and repelling lines inside $\mathbb{P}^1(k')$, and Lemmas 13.6–13.8 then give two elements $a, b \in \Gamma$ which are finite words in S and freely generate F_2 . The field extension is used only to see the projective dynamics; all group elements are still words in the original matrices, so a, b retain algebraic entries in the original compact embedding.

13.6 A free subgroup of $SU(2)$ is automatically dense here

The compact-return step uses two elementary structural facts, which we prove here at the point of use.

Lemma 13.9 (Normalizer of the standard torus). *Let*

$$T = \left\{ \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} : \theta \in \mathbb{R} \right\} < SU(2).$$

Then

$$N_{SU(2)}(T) = T \sqcup wT, \quad w = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

so $N(T)/T$ has order two.

Proof. The common eigenspaces of T on $\mathbb{C}^2$ are the two coordinate lines. A normalizing matrix must permute them. If it fixes both, it is diagonal and lies in T . If it swaps them, multiplying by w^{-1} fixes both and hence lies in T . Conversely w exchanges the two diagonal characters and normalizes T . $\square$

Lemma 13.10 (Proper closed subgroups of $SU(2)$ are virtually abelian). *Every proper closed subgroup $H < SU(2)$ is either finite or contains a one-dimensional torus with finite index.*

Proof. A closed subgroup of a Lie group is a Lie subgroup. If H° had dimension three, it would be open in connected $SU(2)$ and hence equal to $SU(2)$. There is no two-dimensional Lie subalgebra of $\mathfrak{su}(2)$: under the cross-product model, the bracket of two independent vectors in a two-plane is perpendicular to that plane. Thus $\dim H^\circ$ is zero or one.

If it is zero, compactness makes the discrete group H finite. If it is one, $H^\circ = T_0$ is a maximal torus. Since the identity component is characteristic, $H \subset N(T_0)$. Conjugating T_0 to the standard torus and using Lemma 13.9 shows that $N(T_0)/T_0$ has order two. Hence T_0 has finite index in H . $\square$

Lemma 13.11 (Virtually abelian groups cannot contain the ping-pong pair). *Every finitely generated virtually abelian group has polynomial word growth. Consequently it contains no subgroup freely generated by two fixed elements.*

Proof. Let H be generated by a finite set R and let $A \leq H$ be abelian of finite index. Replacing A by the intersection of its finitely many conjugates, we may assume that A is normal, still abelian and finite index. Choose coset representatives $q_1, \dots, q_m$. Schreier rewriting gives a finite generating set

$$\{q_i r q_j^{-1} : r \in R, A q_i r = A q_j\}$$

for A : following successive cosets along an R -word telescopes the word into these generators followed by its final representative. Thus A is a quotient of $\mathbb{Z}^d$ for some finite d , and the number of its elements representable by words of length $O(n)$ is $O((1+n)^d)$. Multiplying by the finitely many cosets gives the same polynomial bound for the word ball of H .

If fixed elements $a, b \in H$ freely generated F_2 , all reduced words of length at most n in $a^{\pm 1}, b^{\pm 1}$ would be distinct. There are $1 + 4 \sum_{j=0}^{n-1} 3^j$ such words, and if a, b have bounded R -length they all lie in an R -ball of radius $O(n)$. This contradicts polynomial growth. $\square$

Lemma 13.12 (The free pair is dense in the original compact embedding). *For the a, b constructed above,*

$$\overline{\langle a, b \rangle} = SU(2).$$

Proof. Let H be the closure. It contains the infinite free group $\langle a, b \rangle$, so H is not finite. If H were proper, Lemma 13.10 would make it virtually abelian, contradicting Lemma 13.11. Therefore $H = SU(2)$. $\square$

Theorem 13.13 (Algebraic dense free-pair reduction). *Let $S \subset SU(2)$ be finite, symmetric, algebraic, and dense-generating. Then there are $L < \infty$ and $a, b \in S^{2L}$ such that a, b have algebraic entries and freely generate a dense subgroup of $SU(2)$. In particular, both a and b may be chosen to have even S -word length.*

Proof. Let

$$\Gamma_0 = \langle S^2 \rangle.$$

Every even S -word belongs to Γ_0 , and after fixing $s_0 \in S$, every $s \in S$ belongs to the same right coset $\Gamma_0 s_0$ because $ss_0^{-1} \in S^2$. Hence $[\Gamma : \Gamma_0] \leq 2$. Moreover Γ_0 is normal: conjugating a product of two generators by one generator gives an even word, and these products generate Γ_0 .

The closure $K_0 = \overline{\Gamma_0}$ is therefore a closed normal subgroup of $SU(2)$: normality first holds under conjugation by the dense subgroup Γ , and continuity of conjugation extends it to all of $SU(2)$. Its Lie algebra is an ideal of the simple Lie algebra $\mathfrak{su}(2)$, so either K_0 is zero-dimensional or $K_0 = SU(2)$. The first case would make K_0 finite; then Γ , being a union of at most two cosets of Γ_0 , would also be finite, contradicting density. Therefore Γ_0 is dense.

Now apply Lemmas 13.3, 13.5, 13.7, 13.8, and 13.12 to the finite algebraic symmetric set S^2 , which densely generates Γ_0 . The resulting a, b are finite S^2 -words, hence even S -words. If their S^2 -lengths are at most L , then $a, b \in S^{2L}$. $\square$

13.7 The model walk satisfies the earlier compact inputs

Proposition 13.14 (Interface to the flattening and representation-theory chapters). *Let a, b be supplied by Theorem 13.13 and let*

$$T = \{a, a^{-1}, b, b^{-1}\}, \quad \nu = \frac{1}{4} \sum_{t \in T} \delta_t.$$

Then, with $q_n = \nu^{*n}$, Theorems 2.6 and 2.10 apply to q_n with constants depending only on the fixed pair a, b . Consequently Corollary 10.12 supplies the logarithmic-scale L^2 flattening hypothesis of Theorem 12.13, and that theorem gives an absolute spectral gap for convolution by ν on $L_0^2(SU(2))$.

Proof. The pair a, b has algebraic entries and freely generates F_2 , so it satisfies exactly the standing free-pair hypothesis introduced before Theorem 2.6. The height/separation constants in that chapter may therefore be recomputed for $T = \{a, a^{-1}, b, b^{-1}\}$; freeness supplies the regular-tree atom bound and the cyclic-centralizer count used in the torus argument. This proves the two cited nonconcentration theorems for $q_n = \nu^{*n}$.

The pair is also dense by Theorem 13.13. Thus the ball and torus estimates just obtained are precisely Interfaces I3–I4 of the flattening chapter, while the remaining interfaces there are the already proved metric inverse and local-product theorems. Corollary 10.12 therefore applies to this same walk ν . Finally Theorem 12.13 has only two walk-specific hypotheses: dense generation and that logarithmic flattening conclusion. Both have now been verified, so its spectral-gap conclusion applies to ν . $\square$

13.8 Transfer the gap back to the original generators

For a finite symmetric set $R \subset SU(2)$ define the Dirichlet form

$$\mathcal{E}_R(f) = \frac{1}{2|R|} \sum_{r \in R} \|f - rf\|_2^2,$$

where $(rf)(x) = f(r^{-1}x)$.

Lemma 13.15 (Two-sided bounded-word comparison). *Suppose every $t \in T$ is an even S -word of length at most $2L$. Put*

$$M_S f = \frac{1}{|S|} \sum_{s \in S} sf,$$

and define

$$\mathcal{E}_S^+(f) = \frac{1}{2|S|} \sum_{s \in S} \|f - sf\|_2^2, \quad \mathcal{E}_S^-(f) = \frac{1}{2|S|} \sum_{s \in S} \|f + sf\|_2^2.$$

Then

$$\mathcal{E}_T(f) \leq C_{L,S} \mathcal{E}_S^+(f), \quad \mathcal{E}_T(f) \leq C_{L,S} \mathcal{E}_S^-(f).$$

For $\|f\|_2 = 1$ these forms satisfy

$$\mathcal{E}_S^+(f) = 1 - \langle M_S f, f \rangle, \quad \mathcal{E}_S^-(f) = 1 + \langle M_S f, f \rangle.$$

Proof. For the first inequality, if $t = s_1 \cdots s_\ell$ with $\ell \leq 2L$, telescoping gives

$$f - tf = \sum_{j=1}^{\ell} s_1 \cdots s_{j-1} (f - s_j f).$$

Unitary invariance and Cauchy–Schwarz bound its squared norm by $2L$ times the sum of the corresponding $\|f - s_j f\|_2^2$ terms.

For the second inequality use that ℓ is even and write the alternating identity

$$f - tf = \sum_{j=1}^{\ell} (-1)^{j-1} s_1 \cdots s_{j-1} (f + s_j f).$$

The same Cauchy–Schwarz argument now bounds $\|f - tf\|_2^2$ by $2L$ times the sum of the relevant $\|f + s_j f\|_2^2$ terms. Sum over the four elements of T and bound the total number of occurrences of each generator by a constant depending only on L and $|S|$. The displayed identities for the quadratic forms follow by expanding the squares and using symmetry of S . $\square$

Theorem 13.16 (Full algebraic-generator spectral gap in $SU(2)$). *Let $S \subset SU(2)$ be a finite symmetric set with algebraic entries and $\overline{\langle S \rangle} = SU(2)$. Let*

$$\mu_S = \frac{1}{|S|} \sum_{s \in S} \delta_s.$$

Then there is $\eta_S > 0$ such that

$$\|\mu_S * f\|_2 \leq (1 - \eta_S) \|f\|_2 \quad (f \in L_0^2(SU(2))).$$

Thus the uniform S -averaging operator itself, not merely its lazy version, has an absolute spectral gap.

Proof. Choose the even-word free dense algebraic pair T by Theorem 13.13. By Proposition 13.14 and the representation-theory chapter there is $c_T > 0$ such that

$$\mathcal{E}_T(f) \geq c_T \|f\|_2^2 \quad (f \in L_0^2).$$

Lemma 13.15 therefore gives, for every unit vector $f \in L_0^2$,

$$\mathcal{E}_S^+(f) \geq c_S, \quad \mathcal{E}_S^-(f) \geq c_S$$

with $c_S = c_T / C_{L,S} > 0$. Since M_S is self-adjoint,

$$-1 + c_S \leq \langle M_S f, f \rangle \leq 1 - c_S$$

for every unit $f \in L_0^2$. The variational characterization of the norm of a self-adjoint operator gives

$$\|M_S|_{L_0^2}\| \leq 1 - c_S.$$

This is the claimed absolute gap. $\square$

Corollary 13.17 (Exponential L^2 mixing). *With η_S as in Theorem 13.16, for every $f \in L^2(SU(2))$ and every integer $n \geq 0$,*

$$\left\| \mu_S^{(n)} * f - \int f dm_G \right\|_2 \leq (1 - \eta_S)^n \|f - \int f dm_G\|_2.$$

In particular the same estimate has the form $e^{-c'_S n}$ with $c'_S = -\log(1 - \eta_S) > 0$.

Proof. Write $f = \int f dm_G + f_0$ with $f_0 \in L^2_0(SU(2))$. Constants are fixed by convolution with the probability measure μ_S , while Theorem 13.16 gives $\|\mu_S * f_0\|_2 \leq (1 - \eta_S)\|f_0\|_2$. Iterating this inequality n times proves the displayed bound. The exponential form follows from the stated definition of c'_S . $\square$

13.9 Mechanism of the reduction

The rank-one reduction proved in this chapter avoids invoking any form of the Tits alternative as a black box. The mechanism is

$$\begin{aligned} \text{algebraicity} &\Rightarrow \text{unbounded algebraic place} \Rightarrow \text{proximal word} \Rightarrow \text{local-field ping-pong}, \\ &\Rightarrow \text{free dense pair} \Rightarrow \text{Dirichlet-form transfer}. \end{aligned}$$

Together with the compact nonconcentration, flattening, and representation-theoretic results, this completes the spectral-gap theorem for arbitrary finite algebraic dense symmetric generators.