Hypergraph Rank and Expansion

Kevin Pratt

CMU-CS-23-135

September 2023

Computer Science Department
School of Computer Science
Carnegie Mellon University
Pittsburgh PA 15213

Thesis Committee:
Ryan O'Donnell, Chair
Pravesh Kothari
Prasad Tetali
Alex Lubotzky (Weizmann)
Chris Umans (Caltech)

*Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy.*

This research was sponsored by the U.S. Army Contracting Office under award number W911NF-20-S-0004 and the National Science Foundation under award numbers 1717606 and 190310. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, the U.S. government or any other entity.

Keywords: Algorithm design, graphs and hypergraphs, algebraic algorithms, fast matrix multiplication, expander graphs, high–dimensional expanders

Abstract

The linear–algebraic notions of matrix rank and expansion in graphs are ubiquitous throughout computer science, with applications to algebraic complexity theory, communication complexity, and derandomization. In recent years, high–dimensional generalizations of these notions to tensors and hypergraphs have led to progress on a wide variety of problems, including the improved analysis of Markov chains, algorithms and barriers for fast matrix multiplication, and the discovery of good locally testable codes and quantum LDPC codes. Yet compared to their linear–algebraic counterparts, these notions are very poorly understood. This thesis studies such notions of tensor rank and expansion in hypergraphs and their applications to algorithm design.

Our contributions are threefold. First, we give new applications of tensor rank to the area of parameterized and exact algorithms, unifying several prior algorithmic tools and obtaining faster, more space–efficient algorithms for a handful of problems. We then study algorithms for fast matrix multiplication. In this area we identify new limitations on current algorithmic approaches via connections to additive combinatorics and group theory. Finally, we give several new group–theoretic families of sparse spectral high–dimensional expanders.

Acknowledgments

This thesis would not have been possible without the great advisors and mentors that I have been fortunate to have crossed paths with. Ryan O'Donnell, I am extremely grateful to you for nurturing my growth as an independent researcher, for all of your feedback and insights, and for your encouragement and kindness. Chris Umans, thank you for the many hours spent explaining, re-explaining, and discussing fast matrix multiplication with me. Don Sheehy, thank you for introducing me to the world of computer science research while I was at the University of Connecticut. I would not have started this journey if I did not meet you. John Bowers, thank you for taking me on for an undergraduate summer research project. This was a formative experience for me that I often reflect on fondly.

I thank all of the great friends I have made in the CMU Computer Science Department. Our camaraderie brought me happiness when I needed it most. I hope we stay in touch for years to come. I am particularly grateful to have known Costin Bădescu, Mark Gillespie, Alex Wang, and Jalani Williams as roommates.

Finally, I thank my family for their never-ending support, and I apologize for my explanations of what I spent the last six years doing. Unfortunately, this thesis will not be of any help there.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis studies notions of *rank* and *expansion* in tensors and hypergraphs, and their applications to theoretical computer science. These notions have been at the heart of a number of recent advances, including the improved analysis of Markov chain mixing [ALGV19], algorithms and barriers for fast matrix multiplication [S+69, BCC+17a], and the discovery of locally testable codes with optimal parameters [DEL+22]. While the classical linear–algebraic notions of matrix rank and expansion in graphs are generally well–understood, we know very little about their higher–dimensional counterparts. For example, finding lower bounds on the rank of *explicit* tensors is a central challenge in algebraic complexity theory [Raz13]. Good upper bounds on the ranks of particular tensors are challenging to establish, and have important algorithmic applications to problems including fast matrix multiplication [S+69]. In the field of additive combinatorics, understanding the connections between different notions of tensor rank is at the heart of the inverse conjectures for the Gowers norms [CM23]. Finding hypergraphs with certain expansion properties is a major challenge that has recently led to the discovery of good locally testable codes [DEL+22] and quantum LDPC codes [PK21]. All of these challenges are absent for matrices and graphs: it is trivial to find an explicit matrix of full rank, matrix rank can be computed efficiently, all "sensible" notions of matrix rank are equivalent, and random graphs have excellent expansion properties.

In the rest of this section, we give an introduction to the main concepts in this thesis and highlight a few of their applications. We then give a summary of our contributions and of the rest of this thesis.

1.1 Notions of tensor rank

Let $\mathbb{F}$ be a field. For the purposes of this thesis, a *tensor T* of *order d*, or a *d-tensor*, is a multilinear map

$$T : \underbrace{\mathbb{F}^n \times \cdots \times \mathbb{F}^n}_{d \text{ times}} \to \mathbb{F}.$$

By choosing coordinates, we can view T concretely as the multilinear polynomial

$$T = \sum_{\alpha \in [n]^d} c_\alpha \prod_{i=1}^{d} x_{i,\alpha_i}$$

where $c_\alpha \in \mathbb{F}$ and $x_{i,j}$ are variables. One can then visualize T as the $n \times \cdots \times n$ "cube" filled with the coefficients c_α, in the same way that one identifies a bilinear form with a matrix.

Example 1.1.1. A particular family of tensors we will be interested in are the *matrix multiplication tensors*. These are the trilinear polynomials in the three sets of variables $\{x_{ij}\}_{i,j \in [n]}$, $\{y_{ij}\}_{i,j \in [n]}$, $\{z_{ij}\}_{i,j \in [n]}$ defined by

$$\langle n, n, n \rangle := \sum_{i,j,k \in [n]} x_{ij} y_{jk} z_{ki}.$$

We first recall the familiar case when $d = 2$. A tensor of order 2 is a bilinear form, which we can identify with the unique matrix $A \in \mathbb{F}^{n \times n}$ such that $T(x, y) = x^T A y$ for $x, y \in \mathbb{F}^n$. The rank of T, denoted $\mathbf{R}(T)$, is the minimum number r such that $T = \sum_{i=1}^{r} u_i \otimes v_i$ for some linear forms $u_i, v_i : \mathbb{F}^n \to \mathbb{F}$. In the special case when T is symmetric (meaning $T(x, y) = T(y, x)$ for all $x, y \in \mathbb{F}^n$) and $\mathrm{char}(\mathbb{F}) \neq 2$, we can identify T with a homogeneous polynomial of degree 2 in $\mathbb{F}[x_1, \ldots, x_n]$; the rank of T is then equal to the minimum r such that we can express this polynomial as a sum of r squares of linear forms.

This suggests the following generalization of rank for tensors of order greater than two: a tensor has rank one if it can be written as $v_1 \otimes \cdots \otimes v_d$ for some linear forms v_i, and its rank $\mathbf{R}(T)$ is the minimum number of rank one tensors whose span contains T. In the case that T is a symmetric tensor of order d, (equivalently, a homogeneous polynomial of degree d), we define its *Waring rank* $\mathbf{R}_S(T)$ as the minimum r such that we can write $T = \sum_{i=1}^{r} c_i \ell_i^{\otimes d}$ for some linear forms ℓ_i and $c_i \in \mathbb{F}$. Although these two definitions are the same for symmetric tensor of order 2 (i.e. symmetric matrices), this was recently found to be false for higher orders [Shi18]. These are probably the most common notions of tensor rank, having been studied since the 1800's [Syl52].

Example 1.1.2. Over $\mathbb{F} = \mathbb{Q}$, we have the identity

$$xyz = \frac{1}{24}\left[(x+y+z)^3 - (x+y-z)^3 - (x-y+z)^3 - (-x+y+z)^3\right]$$

so $\mathbf{R}_S(xyz) \leq 4$. In fact, one can show that $\mathbf{R}_S(xyz) = 4$ [RS11].

The importance of tensor rank to algorithm design was made clear by work of Strassen [S$^+$69]. Strassen observed that for a tensor T of order 3, its tensor rank is equal (up to a constant factor) to the minimum number of (nonscalar) multiplications needed to compute T. As a consequence, he showed that the number of arithmetic operations needed to multiply two matrices is determined by the rank of the *matrix multiplication tensors*. On the flip side, work of Raz showed that sufficiently strong lower bounds on the ranks of explicit tensors would imply lower bounds in complexity theory that seem to be very far from our current reach [Raz13]. To illustrate our lack of understanding of tensor rank: we know that almost all 3-tensors have rank $\Omega(n^2)$, yet we are currently unable to prove that any explicit tensor has rank greater than $O(n)$.

The story does not end there, however. Here are three alternative characterizations of matrix rank we could have started with:

1. The maximum number r such that BAC is the $r \times r$ identity matrix, for some matrices B, C.
2. The maximum over all subspaces U, V such that $u^T A v = 0$ for all $u \in U$ and $v \in V$, of $\mathrm{Codim}(U) + \mathrm{Codim}(V)$.
3. The quantity $-\log_p \mathbf{E}_{x,y \in \mathbb{F}^n} \chi(x^T A y)$, in the case where $\mathbb{F} = \mathbb{F}_p$ and $\chi : \mathbb{F} \to \mathbb{C}$ is a nontrivial additive character (without loss of generality, we may assume that $\chi(m) = e^{2\pi m i/p}$).

The natural generalizations of these to higher-order tensors lead to the notions of *subrank*, *partition rank*, and *analytic rank*. While these are all equivalent for matrices, they are mutually inequivalent for tensors of order 3 and higher, and have applications to distinct areas of computer science and math. Partition rank (or more precisely, its symmetric analogue of *Schmidt rank*) was first studied in algebraic number theory due to applications to counting integer points on varieties [Sch85]. It was then re-introduced under the name of *slice rank* in a solution to the *cap–set problem* of determining the size of the largest subset of $\mathbb{F}_3^n$ with no 3-term arithmetic progessions. In commutative algebra, partition rank was recently used to resolve *Stillman's conjecture* [AH20]. Subrank was introduced by Strassen's work on fast matrix multiplication [Str86], and the relation between partition rank and subrank is at the heart of recent barrier for to fast matrix multiplication. Analytic rank was introduced by Gowers and Wolf [GW11] due to connections to the Gowers inverse conjectures in additive combinatorics.

In the following subsections we give some motivating applications of tensor rank to algorithm design.

1.1.1 The exponent of matrix multiplication

Let ω be the smallest real number such that there exists an algorithm for multiplying two $n \times n$ matrices over $\mathbb{F}$ using $n^{\omega+\varepsilon}$ arithmetic operations for any $\varepsilon > 0$. One easily has the bounds $2 \leq \omega \leq 3$, and a long line of research has led to the upper bound of $\omega < 2.373$ [WXXZ23]. In this section, we summarize the history behind this bound and some of the key ideas involved. We will give a more in-depth discussion of these ideas in Chapter 3.

Strassen showed that determining the value of ω reduces to determining the tensor rank of the matrix multiplication tensors $\langle n, n, n \rangle$. Specifically, he showed that

$$\omega = \lim_{n \to \infty} \log_n \mathbf{R}(\langle n, n, n \rangle).$$

In [S+69] the bound of $\omega < 2.82$ was shown. Stassen's key observation, fundamental to all subsequent improvements, was that *Kronecker products* of matrix multiplication tensors are themselves matrix multiplication tensors. Here the Kronecker product of two tensors is the natural generalization of the Kronecker product of matrices (Definition 3.1.1). As tensor rank is easily shown to be submultiplicative under this operation, in order to obtain nontrivial bounds on ω, it suffices to give a bound on the rank $\langle n, n, n \rangle$ better than n^3 for some particular n. Stassen executed this idea by showing that $\mathbf{R}(\langle 2, 2, 2 \rangle) \leq 7 < 8$, and hence $\omega \leq \log_2 7 < 2.81$.

All subsequent improvements can be understood as bounding the rank of tensors that are progressively less-and-less directly related to matrix multiplication. In 1979, Bini et al. [BCRL79] observed that it suffices to bound the *border rank* of the matrix multiplication tensor. Informally, the border rank of a tensor is the minimum rank a tensor approaching T. Although a sequence of matrices cannot have rank less than that of their limit point, this can happen for tensors. This led to the bound of $\omega < 2.78$.

The next major improvement was due to Schöenhage [Sch81]. Shoenhage showed that it sufficed to bound the rank of a direct sum of matrix multiplication tensors, instead of a single matrix multiplication tensor. Here a direct sum of tensors is the natural generalization of a direct sum of matrices. Specifically, he established the *asymptotic sum inequality*

$$\sum_i (a_i b_i c_i)^{\omega/3} \leq \mathbf{R}(T).$$

Using this, he obtained the bound of $\omega < 2.53$. The asymptotic sum inequality has been the cornerstone for all subsequent improvements.

The next milestone was Strassen's *laser method* [Str86]. This time, one starts with a tensor T which is a *sum* of matrix multiplication tensors — not necessarily a direct sum. In other words, the summands may "interfere" with one another. Strassen showed that under certain conditions one then obtain from T a large direct sum of matrix multiplication tensors. One then applies the asymptotic sum inequality to this resulting tensor. Doing so, he obtained the bound of $\omega < 2.48$. By refining this method, Coppersmith and Winograd [CW87] showed that $\omega < 2.376$. The current best bound of $\omega < 2.372$ [WXXZ23] follows from further refinements to this method.

In 2005, a group–theoretic approach was proposed by Cohn and Umans [CU03]. It is this approach that we will focus on in this thesis. For a finite group G, we let $\mathrm{Irr}(G)$ denote the set of irreducible complex representations of G. For $X \subseteq G$, let $Q(X) = \{xx'^{-1} : x, x' \in X\}$ be the quotient set of X.

Definition 1.1.3. We say that $S, T, U \subseteq G$ satisfy the *triple product property* (or TPP for short) if for all $s \in Q(S), t \in Q(T), u \in Q(U)$,

$$stu = 1 \implies s = t = u = 1.$$

Remark 1.1.4. A helpful way to think of the triple product property is the following. Consider the complete tripartite hypergraph with parts X_1, X_2, X_3 of size n. Let $f_i : E(X_i, X_{(i+1) \bmod 3}) \to G$ for $i = 1, 2, 3$ be functions from the bipartitions of X to the group G. Then one can show that the existence of sets satisfying the TPP is equivalent to the existence of f_i such that $f_1((i,j))f_2((k,l))f_3(m,p) = 1$ if and only if $j = k, l = m, p = i$. In other words, f_i can be thought of as a "cocycle" satisfying nonequality constraints.

The bounds on ω from this approach come from the following theorem.

Theorem 1.1.5. *[CU03, Theorem 4.1] If S, T, U satisfy the TPP in G, then*

$$(|S||T||U|)^{\omega/3} \leq \sum_i d_i^\omega$$

where $d_i \in \mathbb{N}$ are the dimensions of the irreducible representations of G.

In [CKSU05], it was shown that one can prove $\omega < 2.48$ using this approach — recovering a bound given earlier by Strassen. In fact, it turns out that the current best bounds on ω can be obtained from Theorem 3.2.2 by simulating the laser method inside of the group–theoretic framework.

1.1.2 Waring rank and subgraph counting

In [Pra19] we studied the following problem.

Question 1.1.6. Let $\varepsilon > 0$. What is the minimum over all $\{c_{i_1,\ldots,i_d}\} \in [1-\varepsilon, 1+\varepsilon]$ of the Waring rank of

$$\sum_{1 \le i_1 < i_2 < \cdots < i_d \le n} c_{i_1,\ldots,i_d} x_{i_1} x_{i_2} \cdots x_{i_d}?$$

Denote this quantity by $A_\varepsilon(n,d)$. In [Pra19] we gave an algorithm that computes a $(1+\varepsilon)$ approximation of the number of simple cycles (i.e. a cycle which does not revisit any vertex) of length d in a graph on n vertices with running time $A_\varepsilon(n,d) \cdot \mathrm{poly}(n)$[1]. By establishing the inequality

$$A_\varepsilon(n,d) \le 4.075^d \varepsilon^{-2} \log n$$

we obtained a $4.075^d\,\mathrm{poly}(n, \varepsilon^{-1})$-time algorithm for this problem. More generally, we obtained an algorithm with this same runtime for approximately counting d-vertex subgraphs bounded treewidth. This improved on a $5.44^d\,\mathrm{poly}(n)$-time algorithm of Alon et al. [ADH$^+$08].

We conjecture the following, which would imply that $A_\varepsilon(n,d) \ge 2.58^d$. For a polynomial f, let $\mathrm{Derivs}_k(f)$ be the vector space spanned by all partial derivatives of f of order k.

Conjecture 1.1.7. *Suppose that* $\{c_{i_1,\ldots,i_d}\} \in (\mathbb{F} - \{0\})^{\binom{n}{d}}$. *Then*

$$\dim \mathrm{Derivs}_k\Big(\sum_{1 \le i_1 < i_2 < \cdots < i_d \le n} c_{i_1,\ldots,i_d} x_{i_1} x_{i_2} \cdots x_{i_d}\Big) \ge \min\left(\binom{2d-k}{k}, \binom{d+k}{d-k}\right)$$

If true this would be tight, as the above inequality is reversed when $c_{i_1,\ldots,i_d} = \prod_{j \le k}(i_j - i_k)^2$ (Theorem 2.3.19).

1.2 Expansion in graphs and hypergraphs

Let $G = (V,E)$ be an undirected graph. For simplicity, assume that G is regular of degree d. Associated with G is the adjacency matrix A, which acts on the space of functions from $V(G)$ to $\mathbb{R}$ via

$$Af(x) = \sum_{y \sim x} f(y).$$

This matrix is symmetric and has real entries, and hence by the spectral theorem has real eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$. It is easily checked that $\lambda_1 = d$. We say that G is a λ-expander if all other eigenvalues are contained in the interval $[-\lambda, \lambda]$. For fixed d (a regime of interest in many applications), the Alon–Boppanna theorem (c.f. [HLW06, Theorem 5.3] says that $2\sqrt{d-1} - o(1) \le \lambda$; the closer λ is to this lower bound, the "better" the expander it is, with graphs achieving the asymptotic minimum $2\sqrt{d-1}$ being called *Ramanujan*. Furthermore, there are explicit families of Ramanujan graphs [LPS88]. A result of Friedman [Fri08] shows that with high probability a random graph G comes close to matching this, with $\lambda(G) \le 2\sqrt{d-1} + o(1)$.

[1] Here we are assuming constant time arithmetic operations for simplicity.

Intuitively, expander graphs are highly "jumbled" and behave pseudo-randomly. One key fact which shows this is the expander mixing lemma (c.f. [HLW06, Lemma 2.5]), which states that for all $A, B \subseteq V(G)$,

$$|E(A,B) - \frac{d|A||B|}{n}| \leq \lambda(|A||B|)^{1/2}$$

where $E(A,B)$ is the number of edges with one vertex in A and one in B. As $\frac{d|A||B|}{n}$ is the expected number of edges between A and B in a random graph with the same density as that of G, expansion can be seen as a pseudorandomness property. Expander graphs are indispensable to computer science, with some important applications including the proof that $SL = L$ [Rei08], the construction of good LDPC codes [SS96], and Dinur's proof of the PCP theorem [Din07].

Motivated by the importance of expander graphs, it is natural to try to extend the above definition to hypergraphs. One such generalization is that of *local spectral expansion*. Let X be a downward-closed 3-uniform hypergraph (a 2-dimensional *simplicial complex*). As was the case of graphs, it is convenient to assume that X is "regular". By this we now mean that all vertices are contained in the same number of edges in X, and all edges are contained in the same number of triangles in X. We define the *link* of a vertex $v \in X$ is the induced subgraph on the neighboring vertices of v. We will write $X(0), X(1), X(2)$ for the sets of vertices, edges, and triangles in X, respectively.

Definition 1.2.1. We say that X is a λ-local spectral expander if the links of all vertices in X are λ-expander graphs.

This notion was first used in work of Garland to prove the vanishing of the cohomology groups of certain simplicial complexes [Gar73], which had applications to representation theory and arithmetic geometry. It was subsequently used by Żuk to prove Kazhdan's property (T) for random discrete groups [Żuk03]. It has recently been used to obtained improved bounds on the mixing time of Markov chains [AGV18], and it has inspired the discoveries of locally testable codes and quantum LDPC codes [PK21, DEL$^+$22]. We will elaborate on these two recent applications in the next subsection.

However, we could have started with the following alternative notions of expansion in d-regular graphs:

1. (Discrepancy) For some $\mu > 0$, for all $A, B \subseteq V(G)$,

$$|E(A,B) - \frac{d|A||B|}{n}| \leq \mu(|A||B|)^{1/2}$$

2. (Edge isoperimetry) For some $\varepsilon > 0$,

$$h(G) := \min_{A \subseteq V(G)} \frac{E(A,\overline{A})}{\min(A,\overline{A})} \geq \varepsilon$$

3. (Random walk mixing) For some ε', the transition matrix of the lazy random walk on G, given by $\frac{1}{d+1}(A+I)$, has second largest eigenvalue at most ε'.

These three notions of expansion are qualitatively equivalent. The random walk definition is simply a rephrasing of our first definition. The discrepancy definition is equivalent to the spectral definition by the expander mixing lemma and its converse [BL06]. The equivalence between edge expansion and spectral expansion the content of the discrete Cheeger inequalities [HLW06, Theorem 4.11].

These definitions generalize to the inequivalent notions of *hypergraph discrepancy* or *quasirandomness*, *coboundary expansion*, and the rapid mixing of the *up-down random walks*. Although these four notions are inequivalent, there are some important relations between them. For example, by the *trickling–down* theorem [Opp18], sufficiently strong local spectral expansion implies rapid mixing of the up-down walks.

In the next subsections we give some applications related to coboundary expansion and the mixing of up-down walks, so we elaboreate on those definition snow. Coboundary expansion is the following natural generalization of edge expansion. Let $C^k(X) := \{f : X(k) \to \mathbb{F}_2\}$ be the space of k-cochains on X. Let $\delta_k : C^k(X) \to C^{k+1}(X)$ be the coboundary operator, given by $\delta_k f(\tau) = \sum_{\sigma \subseteq \tau, \sigma \in X(k)} f(\tau)$. Denote by $B^k(X) := \operatorname{im}(\delta_{k-1})$ the space of k-coboundaries, and by $Z^k(X) := \ker(\delta_k)$ the space of k-cocycles. The k-cohomology group of X is then $H^k(X) := Z^k(X)/B^k(X)$. We let $\|f\|$ be the fractional Hamming weight of f, and we define the distance between $f, g \in C^k(X)$ to be $\|f - g\|$. If V is a subspace of $C^k(X)$, we write $d(f, V)$ as shorthand for $\min_{v \in V} d(f, v)$.

Definition 1.2.2. The k-th coboundary constant of X is

$$h_k(X) := \min_{f \in C^k \setminus B^k} \frac{\|\delta f\|}{d(f, B^k)}.$$

We say that X is a ε-coboundary expander if $h_k(X) \geq \varepsilon$ for all k.

This generalizes the definition of edge isoperimetry in a graph. Indeed, consider $h_0(X)$. If $f \in C^0(X)$, then $\|\delta f\|$ is the fractional number of edges from the support of f to its complement. Since $X(-1) = \{\emptyset\}$, the only 0-coboundaries are the two constant functions on $V(G)$, and the distance of f to such a function is simply the minimum of the fractional support size of f and the fractional size of the complement of its support. Coboundary expansion has close connections to the areas of property testing [KL14] and coding theory [PK21].

On the other hand, a natural generalization of the random-walk definition is that the rapid mixing of the *up-down* walks on X. There are two natural random walks on X, obtained by walking from a k-face to a uniformly random $k+1$ face containing it, and then back down to a uniformly random k face. When $k = 0$, this is just the non-lazy random walk on the graph with vertex set $X(0)$ and edges $X(1)$. This notion has applications to sampling algorithms [AGV18].

1.2.1 Sampling spanning trees

Given an undirected connected graph G, let T_G be the set of spanning trees of G. In this section we describe a polynomial-time algorithm given in [ALGV19] for approximately sampling a uniformly random spanning tree from G. Although efficient algorithms for this problem were known prior to this work, a generalization of these ideas led to the resolution of the Mihail-Vazirani conjecture; see [GK23, Section 10] for more discussion.

The algorithm of [ALGV19] is the following. First, find an arbitrary spanning tree T in G. Then, delete a uniformly random edge from T. This yields a forest F consisting of two trees (one may be empty). Next, re-sample a uniformly random spanning tree containing the edges of F —

in other words, choose a random edge going between the two components of F. This defines a random walk on the spanning tres in G, and it is not hard to show that its stationary distibution is the uniform distribution on T_G.

In [AGV18] it was shown that the spectral gap of this random walk is at least $\frac{1}{n-1}$. As a consequence, this gives a polynomial time algorithm for sampling a spanning tree from a distribution that is ε-close to the uniform distribution on T_G in total variation distance. Their approach is based on analyzing local spectral expansion of a "spanning tree hypergraph" associated to G in order to bound the mixing time of the up-down random walks on this hypergraph via the trickling down theorem.

1.2.2 Good locally testable codes

A linear code C is a subspace of $\mathbb{F}_2^n$. The *distance* of C is the minimum fractional Hamming weight of any nonzero vector in C, and the *rate* of C is its fractional dimension $\dim C/n$. We say that C is (q, κ)-*locally testable* if there exists an algorithm that, given $x \in \mathbb{F}_2^n$, queries x in at most q coordinates and

- if $x \in C$, always outputs "$x \in C$";
- if $x \notin C$, outputs "$x \notin C$ with probability at least $\kappa \cdot d(x, C)$.

This is a desirable property of a code, as it allows us to first perform an efficient check if $x \in C$. If we find that $x \notin C$, there is no point in attempting to decode x. Locally testable codes were introduced due to connections to probabilistically checkable proofs, and understanding them may be a stepping stone towards *linear* PCP's [GS06].

In [DEL+22], such a code with constant distance, rate and locality was found. While the construction does not explicitly use any of the aforementioned notions of HDXs, it uses notions similar to the coboundary expansion and mixing of higher order random walks.

1.3 Overview and summary of results

In the first part of this thesis, we focus on designing faster algorithms for some widely applicable algebraic problems. We begin in Chapter 2 by studying such problems that arise in the area of *parameterized* and *exact* algorithms. In these areas, one is interested in designing faster exponential-time algorithms for NP-complete problems.

We first study the following problem: given evaluation access to a polynomial f, decide if the monomial expansion of f contains a multilinear monomial (i.e. a monomial with no variable appearing to a power higher than 1) in its support. By exploiting connections between this problem and Waring rank, we give a faster algorithm for approximately counting subgraphs of bounded treewidth, improving on work of [ADH+08]. We give a very simple polynomial space algorithm for counting the number of cycles of length d in a graph in time $n^{d+O(1)}$, answering a question of Koutis and Williams [KW09]. We also show that many earlier combinatorial algorithms in this area can be understood fundamentally as Waring rank bounds. We then give an improved algorithm for some important special cases of this problem if we assume white-box access to a circuit computing f. This gives improved aglorithms for several problems, including detecting cycles of length d, detecting d-internal outbranchings (improving on [GRWZ18a]), and more. This section is based on the papers [Pra19, BP21].

We then study the complexity of matrix multiplication. Our interest is primarily in the group–theoretic approach of Cohn and Umans [CU03]. We begin by giving a brief overview of the techniques in this area in Chapter 3. In Chapter 4, we explore the viability of matrix groups within the group–theoretic framework. Our main result here is that one cannot obtain $\omega = 2$ using groups of Lie type, answering a question of [CU03]. By relating this new "quasirandomness" barrier with prior barriers, we then give a counterexample to a conjecture of Petrov [Pra22]. We also explore a framework for possibly obtaining nontrivial bounds on ω using continuous Lie groups. This is based on the paper [BCG+22]. In Chapter 5 we explore problems in additive and extremal combinatorics that we see as stepping stones towards understanding the power of the group–theoretic framework.

In the final chapter, we give new examples of spectral high–dimensional expanders. Despite their utility, examples of sparse high-dimensional expanders are very scarce. Prior to our work, we knew of just two examples. Our examples are obtained by generalizing [KO18] to groups of Lie type (more precisely, Chevalley groups). This section is based on the paper [OP22].

While the themes of tensor rank and hypergraph expansion are mostly separate in this thesis, we will see some connections. For example, the quasirandomness barrier that we identify for the group-theoretic approach is closely connected to the hypergraph mixing generalization of expander graphs. In the other direction, in Section 6.4 we will use lower bounds on the *Schmidt rank* of certain polynomials to establish local spectral expansion of certain hypergraphs. This makes use of a line of work on the equidistribution of "high rank" polynomial maps [GT07, Mil19].

Chapter 2

Applications of Tensor Rank to Parameterized and Exact Algorithms

2.1 Introduction

The *Waring rank* of a homogeneous n-variate degree-d polynomial $f \in \mathcal{S}_d^n := \mathbb{C}[x_1, \dots, x_n]_d$, denoted $\mathbf{R}_S(f)$, is the minimum r such that

$$f = \ell_1^d + \cdots + \ell_r^d, \tag{2.1}$$

for some linear forms $\ell_1, \dots, \ell_r \in \mathcal{S}_1^n$. The study of Waring rank is a classical problem in algebraic geometry and invariant theory, with pioneering work done in the second half of the 19th century by A. Clebsch, J.J. Sylvester, and T. Reye, among others [IK99, Introduction]. It has enjoyed a recent resurgence of popularity within algebraic geometry [IK99, Lan12] and has connections in computer science to the limiting exponent of matrix multiplication ω [CHI$^+$18], the Mulmuley-Sohoni Geometric Complexity Theory program [BIP19], and several other areas in algebraic complexity [Lan17, EGOW18]. This chapter adds *parameterized algorithms* to this list, showing that several methods in this area (color-coding methods [AYZ95, AG07, HWZ08], the group-algebra/determinant sum approach [Kou08, Wil09a, Bjö10a], and inclusion-exclusion methods) fundamentally result from rank upper bounds for a specific family of polynomials. In a situation similar to that of ω, better explicit upper bounds on the Waring rank of these polynomials yield faster algorithms for certain problems in a black-box manner, and lower bounds on the Waring rank of these polynomials imply barriers such algorithms face.

This connection should not come as a complete surprise, as many algorithms work by solving a question about the coefficients of some efficiently-computable "generating polynomial" determined by the input. The insight of this chapter, which has been largely unexploited, is that in general this is a question about Waring rank.

Let $e_{n,d} := \sum_{1 \le i_1 < i_2 < \cdots < i_d \le n} x_{i_1} \cdots x_{i_d}$ denote the elementary symmetric polynomial of degree d in n variables. We will study the following questions:

Question 2.1.1. What is $A(n, d)$, the minimum Waring rank among all $g \in \mathcal{S}_d^n$ with the property that $\operatorname{supp}(g) = \operatorname{supp}(e_{n,d})$?[1]

[1] Here $\operatorname{supp}(\sum_{\alpha \in \mathbb{N}^n} c_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n}) := \{\alpha \in \mathbb{N}^n : c_\alpha \neq 0\}$.

Question 2.1.2. What is $A^+(n,d)$, the the minimum Waring rank among all $g \in \mathbb{R}_{\geq 0}[x_1,\ldots,x_n]$ with the property that $\operatorname{supp}(g) = \operatorname{supp}(e_{n,d})$?
Question 2.1.3. For $0 \leq \varepsilon < 1$, what is $A^\epsilon(n,d)$, the minimum Waring rank among all $g \in \mathbb{R}[x_1,\ldots,x_n]$ with the property that $\operatorname{supp}(g) = \operatorname{supp}(e_{n,d})$ and the nonzero coefficients of g are in the range $1 \pm \varepsilon$?

We now illustrate the algorithmic relevance of these questions with a new and very simple $\binom{n}{\lfloor d/2 \rfloor}\operatorname{poly}(n)$-time and $\operatorname{poly}(n)$-space algorithm for exactly counting simple cycles (i.e., closed walks with no repeated vertices) of length d in an n-vertex graph. This is the fastest polynomial space algorithm for this problem, improving on a $2^d\binom{n}{\lfloor d/2 \rfloor}\operatorname{poly}(n)$-time algorithm of Fomin et al. [FLR$^+$12] which in turn improved on a $2^d(d/2)!\binom{n}{\lfloor d/2 \rfloor}\operatorname{poly}(n)$-time algorithm of Vassilevska Williams and Williams [VW09].

Given a directed graph G, let A_G be the symbolic matrix with entry (i,j) equal to the variable x_i if there is an edge from vertex v_i to vertex v_j, and zero otherwise. By the trace method,

$$f_G \coloneqq \operatorname{tr}(A_G^d) = \sum_{\substack{\text{closed walks} \\ (v_{i_1}, v_{i_2}, \ldots, v_{i_d}) \in G}} x_{i_1} \cdots x_{i_d} \in \mathcal{S}_d^n. \tag{2.2}$$

Now we denote by $g(\partial\mathbf{x})$ the partial differential operator $g(\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_n})$. Note that if $f = \sum_\alpha a_\alpha x_1^{\alpha_1}\cdots x_n^{\alpha_n}$ and $g = \sum_\alpha b_\alpha x_1^{\alpha_1}\cdots x_n^{\alpha_n}$ are both elements of $\mathcal{S}_d^n$,

$$\begin{aligned} g(\partial\mathbf{x})f &= \sum_\alpha b_\alpha \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n} \sum_\alpha a_\alpha x_1^{\alpha_1}\cdots x_n^{\alpha_n} \\ &= \sum_\alpha \alpha_1! \cdots \alpha_n! a_\alpha b_\alpha. \end{aligned}$$

The algorithm is based on two easy observations:
Observation 2.1.4. The number of simple cycles of length d in G equals $e_{n,d}(\partial\mathbf{x})f_G$.
Observation 2.1.5. If $g = a_1\ell_1^d + \cdots + a_r\ell_r^d$, where $\ell_i = c_{i,1}x_1 + \cdots + c_{i,n}x_n$ for $i = 1,\ldots,r$, then for all $f \in \mathcal{S}_d^n$,

$$g(\partial\mathbf{x})f = d! \sum_{i=1}^r a_i f(c_{i,1},\ldots,c_{i,n}).$$

It is immediate that we can compute the number of simple cycles in G of length d using $\mathbf{R}_S(e_{n,d}) = A^0(n,d)$ evaluations of f_G. Now, it was recently shown in [Lee16] that

$$\mathbf{R}_S(e_{n,d}) \leq \binom{n}{\leq \lfloor d/2 \rfloor} \coloneqq \sum_{i=0}^{\lfloor d/2 \rfloor} \binom{n}{i}.$$

Explicitly, for $S \subseteq [n]$ and $i \in [n]$, define the indicator function $\delta_{S,i} \coloneqq -1$ if $i \in S$, and $\delta_{S,i} \coloneqq 1$ otherwise. Then for d odd, $2^{d-1}d! \cdot e_{n,d}$ equals

$$\sum_{\substack{S \subseteq [n] \\ |S| \leq \lfloor d/2 \rfloor}} (-1)^{|S|} \binom{n - \lfloor d/2 \rfloor - |S| - 1}{\lfloor d/2 \rfloor - |S|} (\delta_{S,1}x_1 + \delta_{S,2}x_2 + \cdots + \delta_{S,n}x_n)^d.$$

(A similar formula holds for d even.) It follows that the number of length-d simple cycles in G equals

$$\frac{1}{2^{d-1}} \sum_{\substack{S\subseteq[n] \\ |S|\leq\lfloor d/2\rfloor}} (-1)^{|S|} \binom{n-\lfloor d/2\rfloor-|S|-1}{\lfloor d/2\rfloor-|S|} f_G(\delta_{S,1},\dots,\delta_{S,n}). \tag{2.3}$$

This gives a closed form for the number of length-d simple cycles in G that is easily seen to be computable in the stated time and space bounds. This algorithm is much simpler, both computationally and conceptually, than those of previous approaches.[2]

The above argument shows something very general: given $f \in \mathcal{S}_d^n$ as a black-box, we can compute $e_{n,d}(\partial\mathbf{x})f$ (that is, the sum of the coefficients of the *multilinear monomials* in f) using $\binom{n}{\leq\lfloor d/2\rfloor}$ queries. This answers a "significant" open problem asked by Koutis and Williams [KW09] in a completely black-box way.[3] Moreover, it follows from a special case of our Theorem 2.1.6 that *any* algorithm must make $\mathbf{R}_S(e_{n,d}) \geq \Omega(\binom{n}{\leq\lfloor d/2\rfloor})$ [Lee16] queries to compute $e_{n,d}(\partial\mathbf{x})f$ in the black-box setting:

Theorem 2.1.6. *Fix $g \in \mathcal{S}_d^n$ and let $f \in \mathcal{S}_d^n$ be given as a black-box. The minimum number of queries to f needed to compute $g(\partial\mathbf{x})f$ is $\mathbf{R}_S(g)$, assuming unit-cost arithmetic operations.*

In light of this lower bound, one might next ask for a $(1\pm\varepsilon)$ approximation of $e_{n,d}(\partial\mathbf{x})f$. This prompts our main algorithmic result, which is based on an answer to Question 2.1.3:

Theorem 2.1.7. *Let $f \in \mathbb{R}_{\geq 0}[x_1,\dots,x_n]_d$ be given as a black-box. There is a randomized algorithm which given any $0<\varepsilon<1$ computes a number z such that with probability 2/3,*

$$(1-\varepsilon)\cdot e_{n,d}(\partial\mathbf{x})f < z < (1+\varepsilon)\cdot e_{n,d}(\partial\mathbf{x})f.$$

This algorithm runs in time $4.075^d \cdot \varepsilon^{-2}\log(\varepsilon^{-1})\cdot \mathrm{poly}(n,s_f)$ and uses $\mathrm{poly}(n,s_f,\log(\varepsilon^{-1}))$ space. Here s_f is the maximum bit complexity of f on the domain $\{\pm1\}^n$.

The algorithm and the proof behind Theorem 2.1.7 are simple and can be found in Section 2.4. Applying this theorem to to the graph polynomial f_G, an algorithm for approximately counting simple cycles of length d is immediate.[4] More generally, we have the following:

Theorem 2.1.8. *Let G and H be graphs where $|G|=n$, $|H|=d$, and H has treewidth $\mathrm{tw}(H)$. There is a randomized algorithm which given any $0<\varepsilon<1$ computes a number z such that with probability $2/3$,*

$$(1-\varepsilon)\cdot\mathrm{Sub}(H,G) < z < (1+\varepsilon)\cdot\mathrm{Sub}(H,G).$$

This algorithm runs in time $4.075^d\cdot n^{\mathrm{tw}(H)+O(1)}\cdot\varepsilon^{-2}\log(\varepsilon^{-1})$. Here $\mathrm{Sub}(H,G)$ denotes the number of subgraphs of G isomorphic to H.

The previous fastest algorithm ran in time $5.44^d n^{\mathrm{tw}(H)+O(1)}\varepsilon^{-2}$-time algorithm of Alon et al. [ADH+08], improving on a $5.44^{d\log\log d} n^{\mathrm{tw}(H)+O(1)}\varepsilon^{-2}$-time algorithm of Alon and Gutner

[2] We note that the use of inclusion-exclusion (or "Möbius inversion" [Ned09]) in numerous exact-counting algorithms, such as Ryser's formula for computing the permanent [Rys64] and algorithms for counting Hamiltonian cycles [KGK77] and set packings [BH06], implicitly relies on a natural but suboptimal bound on $\mathbf{R}_S(e_{n,d})$; namely the one given by Equation (2.5) below. We elaborate on this in Example 2.2.3.

[3] An alternate solution to this problem was given contemporaneously in [ACDM18].

[4] In fact, Theorem 2.1.7 gives the fastest *polynomial space* algorithm for approximately counting cycles that we are aware of.

[AG07]. The first parameterized algorithm for a variant of this problem was given by Arvind and Raman [AR02] and had runtime $d^{O(d)}n^{\mathrm{tw}(H)+O(1)}$. In the special case that H has pathwidth $\mathrm{pw}(H)$, an algorithm of Brand et al. [BDH18] runs in time $4^d n^{\mathrm{pw}(H)+O(1)}\varepsilon^{-2}$. We stress that this application is only a motivating example – Theorem 2.1.7 is extremely general and can be applied to approximately count set partitions and packings [BH06], dominating sets [KW09], repetition-free longest common subsequences [BBDS12], and functional motifs in biological networks [GS13].

In the rest of this section we outline our approach. This will suggest a path to derandomize and improve the base of the exponent in Theorem 2.1.7 (and hence Theorem 2.1.8) from 4.075 to 2. Specifically, we raise the following question:

Question 2.1.9. Is $A^{\varepsilon}(n,d) \leq 2^d \cdot \mathrm{poly}(n,\varepsilon^{-1})$?

Prior to this work it was believed [KW15] that a derandomization of polynomial identity testing would be needed to obtain, for instance, a deterministic $2^d\mathrm{poly}(n)$-time algorithm just for *detecting* simple paths of length d in a graph. On the contrary, an explicit affirmative answer to the above question would give a $2^d\mathrm{poly}(n,\varepsilon^{-1})$-time deterministic algorithm for *approximately counting* simple paths.

Remark 2.1.10. A focus on approximating $g(\partial\mathbf{x})f$ in the case that f and g are real stable has recently led to several advances in algorithms and combinatorics; see e.g. [Gur06]. In particular, a result of Anari et al. [AOGSS17] shows that in this case $e_{n,d}(\partial\mathbf{x})f$ can be approximated (up to a factor of $e^{d+\varepsilon}$) deterministically in polynomial time given black-box access to f. This chapter shows that the general (i.e., *unstable*) case raises interesting questions as well.

2.1.1 Our approach and connections to previous work

To continue with the previous example, note that the graph polynomial f_G is supported on a multilinear monomial if and only if G contains a cycle of length d. This motivates the following problem of well-recognized algorithmic importance [Gur04, Kou08, Wil09a]:

Problem 2.1.11. Given black-box access to $f \in \mathcal{S}_d^n$ over $\mathbb{C}^n$, decide if f is supported on a multilinear monomial.

It is not hard to see that any algorithm for computing $g(\partial\mathbf{x})f$, where g is supported on exactly the set of degree-d multilinear monomials, can be used to solve Problem 2.1.11 with one-sided error (Proposition 2.2.11 (a)). This suggests studying upper bounds on $A(n,d)$ (Question 2.1.1) as an approach to solve Problem 2.1.11. Perhaps surprisingly though, it turns out that several known methods in parameterized algorithms can be understood as giving constructive upper bounds on $A(n,d)$, and better upper bounds to $A(n,d)$ would improve upon these methods. For example, the seminal color-coding method of Alon, Yuster, and Zwick [AYZ95] can be recovered from an upper bound on $A(n,d)$ of $O(5.44^d \log n)$, and an improvement to color-coding given by Hüffner et al. [HWZ08] follows from an upper bound on $A(n,d)$ of $O(4.32^d \log n)$ (Remark 2.3.35). The group-algebra/determinant sum approach of [Wil09a, Kou08, Bjö10a] reduces to answering a generalization of Question 2.1.1 (see Definition 2.3.24) in the case that the underlying field is not $\mathbb{C}$ but of characteristic 2. (In Theorem 2.3.28 we give the essentially optimal upper bound of $2^d - 1$ for this variant, which in turn can be used to recover [Wil09a, Kou08, Bjö10a]). Prior to this work, no connection of this precision between these methods was known.

Question 2.1.1 provides insight into lower bounds on previous methods as well. For example, the bounds on $\mathbf{R}_S(e_{n,d})$ given in [Lee16] directly yield asymptotically sharper lower bounds than those given by Alon and Gutner [AG09, Theorem 1] on the size of *perfectly balanced hash families* used by exact-counting color-coding algorithms (Theorem 2.4.8). Curiously, this improvement is ultimately a consequence of *Bézout's theorem* in algebraic geometry. Question 2.1.1 and a classical lower bound on Waring rank (Theorem 2.2.4) explain why *disjointness matrices* arose in the context of lower bounds on color-coding [AG09] and the group-algebra approach [KW09]: they are the partial derivatives matrices of the elementary symmetric polynomials.

Our main answers to Question 2.1.1 are the following. By our Theorems 2.3.4, 2.3.17 and 2.3.34, it follows that

$$2^{d-1} \leq A(n,d) \leq \min(6.75^d, O(4.075^d \log n)).$$

Perhaps surprisingly, this gives an upper bound on $A(n,d)$ *independent* of n. On the negative side, our lower bound on $A(n,d)$ rules out Question 2.1.1 as an approach to obtain algorithms faster than $2^d\text{poly}(n)$ for Problem 2.1.11; moreover, we show in Theorem 2.2.13 that there is also a lower bound of 2^{d-1} on the number of queries needed to solve Problem 2.1.11 with one-sided error.

It is easily seen by Observation 2.1.5 that constructive upper bounds on $A^+(n,d)$ yield deterministic algorithms for determining if f is supported on a multilinear monomial in the case that f has nonnegative real coefficients (as, e.g., the graph polynomial f_G has), and constructive upper bounds on $A^\varepsilon(n,d)$ yield deterministic algorithms for approximating $e_{n,d}(\partial\mathbf{x})f$. This broadly generalizes the use of color-coding in designing approximate counting and deterministic decision algorithms.

Our bounds on $A(n,d)$ also hold for $A^+(n,d)$. Remarkably, we show in Example 2.3.43 that if $A^+(33700,4) \leq 10$ then $A^+(n,d) \leq O(3.9999^d \log n)$. It follows from our Theorem 2.3.4 and Theorem 2.3.34 that

$$2^{d-1} \leq A^\varepsilon(n,d) \leq O(4.075^d \varepsilon^{-2} \log n),$$

and from our Corollary 2.3.12 that $\lim_{n\to\infty} A^\varepsilon(n,d) = \infty$ for all $d > 1$ and $\varepsilon < 1/2$ – unlike $A^+(n,d)$, $A^\varepsilon(n,d)$ depends on n. As an aside, it is immediate that

$$\underline{\mathbf{R}}(e_{n,d}) \leq \lim_{\varepsilon\to 0} A^\varepsilon(n,d) \leq \mathbf{R}_S(e_{n,d}),$$

where $\underline{\mathbf{R}}(g)$ denotes the *Waring border rank* of g, i.e., the minimum r such that there exists a sequence of polynomials of Waring rank at most r converging to g in the Euclidean topology.

2.1.2 Paper overview

For ease of exposition, we work over $\mathbb{C}$ unless specified otherwise. Most of our theorems can be extended to infinite (or sufficiently large) fields of arbitrary characteristic by replacing the polynomial ring with the ring of divided power polynomials (see [IK99, Appendix A]). Except for in Section 2.4, we assume that arithmetic operations can be performed with infinite precision and at unit cost.

In Section 2.2 we introduce concepts related to Waring rank (in particular the *Apolarity Lemma*) in order to better understand the following problems:

Problem 2.1.12. Fix $g \in \mathcal{S}^n_d$. Given black-box access to $f \in \mathcal{S}^n_d$,

a) Compute $g(\partial \mathbf{x})f$.

b) Compute a $(1 \pm \varepsilon)$ approximation of $g(\partial \mathbf{x})f$ (assuming $f, g \in \mathbb{R}_{\geq 0}[x_1, \dots, x_n]$).

c) Determine if $\operatorname{supp}(f) \cap \operatorname{supp}(g) = \emptyset$.

The fundamental connection between Waring rank and Problem 2.1.12 (a) is given by our Theorem 2.1.6. Using similar ideas, we show that at least 2^{d-1} queries are required to test if $\operatorname{supp}(f) \cap \operatorname{supp}(e_{n,d}) = \emptyset$ with one-sided error in Theorem 2.2.13. We then introduce the new concepts of support rank, ε-support rank, and nonnegative support rank, which give upper bounds on the complexity of randomized and deterministic algorithms for the above problems. A related notion of support rank for tensors has previously appeared in the context of ω and quantum communication complexity [CU13, BCZ17, WGE16], but we are unaware of previous work on support rank in the symmetric (polynomial) case. In the case when $d = 2$ these notions are related to the well-studied concepts of sign rank, zero-nonzero rank, and approximate rank of matrices [BDYW11, ALSV13].

In Section 2.3 we study $A(n, d)$ and its variants. We start in Section 2.3.1 by proving negative results, showing that $A(n, d) \geq 2^{d-1}$ (Theorem 2.3.4), and that for sufficiently large n, $A(n, 2) = 3$ (Proposition 2.3.9) and $A(n, 3) \geq 5$ (Corollary 2.3.7). Using bounds on the ε-*rank* of the identity matrix [Alo03], we show in Corollary 2.3.12 that for $1/\sqrt{n} \leq \varepsilon < 1/2$,

$$\Omega(\log n \cdot \varepsilon^{-2}/\log(\varepsilon^{-1})) \leq A^{\varepsilon}(n, 2) \leq O(\log n \cdot \varepsilon^{-2}).$$

While it may at first seem like we are splitting hairs by focusing on particular values of d, we will later show in Example 2.3.43 that, for example, proving that $A^+(n, 4) \leq 10$ for sufficiently large n would yield improved upper bounds on $A^+(n, d)$ for *all* n and d.

Curiously, our lower bound on $A(n, 3)$ is a consequence of the classical *Cayley-Salmon theorem* in algebraic geometry, and our general lower bound on $A(n, d)$ ultimately follows from Bézout's theorem via [RS11]. On this note, we show in Proposition 2.3.6 that Question 2.1.1 is equivalent to a question about the geometry of linear spaces contained in the *Fermat hypersurface* $\{x \in \mathbb{C}^n : \sum_{i=1}^n x_i^d = 0\}$.

The rest of Section 2.3 is focused on general upper bounds on $A(n, d)$ and its variants. Proposition 2.3.14 will give a simple explanation as to why *determinant sums* (as in the title of [Bjö10a]) can be computed in a parameterized way: for all $d \times n$ matrices A and B, the Waring rank of

$$\sum_{\substack{\alpha \in \{0,1\}^n \\ |\alpha| = d}} \det(A_\alpha B_\alpha) x_1^{\alpha_1} \cdots x_n^{\alpha_n} \tag{2.4}$$

is at most $\mathbf{R}_S(\det_d)$. A special case of this example is used in Theorem 2.3.17 to show that $A^+(n, d) < 6.75^d$. In order to improve this, it would suffice to find a better upper bound on the Waring rank of a single polynomial: the determinant of a symbolic $d \times d$ Hankel matrix. We show in Theorem 2.3.19 that the method of partial derivatives cannot give lower bounds on the Waring rank of this polynomial better than 2.6^d.

Next we define rank for polynomials over a field k of arbitrary characteristic – as it is, our definition of rank is not valid in positive characteristic (example: try to write xy as a sum of squares of linear forms over a field of characteristic two). Using this we define $A_{\mathsf{k}}(n, d)$, which equals

$A(n,d)$ when $\operatorname{char}(\mathsf{k}) = 0$. We note in Theorem 2.3.25 that $A_{\mathsf{k}}(n,d) \geq 2^{d-1}$. Theorem 2.3.28 shows that this lower bound is essentially optimal when $\operatorname{char}(\mathsf{k}) = 2$, as then $A_{\mathsf{k}}(n,d) \leq 2^d - 1$; specifically, this rank upper bound holds for Equation (2.4) in the case that $A = B$. This is a simple consequence of the fact that the permanent and the determinant agree in characteristic 2. We explain in this section how the group-algebra approach of [Kou08, Wil09a] and the basis of [Bjö10a] reduce to a slightly weaker fact than this upper bound. A precise connection between support rank and a certain "product-property" of abelian group algebras critical to [Kou08, Wil09a] is given by Theorem 2.3.29.

In Section 2.3.3 we present a method for translating upper bounds on $A^+(n_0, d_0)$ for some *fixed* n_0 and d_0 into upper bounds on $A^+(n,d)$ for *all* n and d (Theorem 2.3.42). This method also allows us to recursively bound $A^\varepsilon(n,d)$ for fixed d (Theorem 2.3.32). This approach can be seen as a vast generalization of color-coding methods, and is based on a *direct power sum* operation on polynomials and a combinatorial tool generalizing *splitters* that we call a *perfect splitter*. We use this to show that $A^\varepsilon(n,d) \leq O(4.075^d \varepsilon^{-2} \log n)$ in Theorem 2.3.34.

In Section 2.4 we give applications of the previous section. We start by giving the proof Theorem 2.1.7, which is then used to prove Theorem 2.1.8. We end with an improved lower bound on the size of perfectly-balanced hash families in Theorem 2.4.8.

We conclude by giving several standalone problems.

2.2 Preliminaries and methods

We use multi-index notation: for $f \in \mathcal{S}^n_d$, we write $f = \sum_{\alpha \in \mathbb{N}^n} c_\alpha x^\alpha$, where $x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. For $\alpha \in \mathbb{N}^n$, we let $|\alpha| := \sum_{i=1}^n \alpha_i$ and $\alpha! := \alpha_1! \alpha_2! \cdots \alpha_n!$. We then define $\mathbb{N}^n_d := \{\alpha \in \mathbb{N}^n : |\alpha| = d\}$, and similarly $\{0,1\}^n_d := \{\alpha \in \{0,1\}^n : |\alpha| = d\}$. Given $\beta \in \mathbb{N}^n$ we say that $\alpha \geq \beta$ if $\alpha_i \geq \beta_i$ for all $i \in [n]$. We denote by ∂_i the differential operator $\frac{\partial}{\partial x_i}$, and we let $\partial^\alpha := \partial_1^{\alpha_1} \cdots \partial_n^{\alpha_n}$. We let $\mathbf{V}(f) := \{p \in \mathbb{C}^n : f(p) = 0\}$ denote the hypersurface defined by f. For $\ell = \sum_{i=1}^n a_i x_i \in \mathcal{S}^n_1$, we let $\ell^* := (a_1, \ldots, a_n) \in \mathbb{C}^n$. For $X \subseteq \mathbb{C}^n$, the ideal of polynomials in $\mathcal{S}^n$ vanishing on X is denoted by $I(X)$. The ideal generated by $f_1, \ldots, f_k \in \mathcal{S}^n$ is denoted by $\langle f_1, \ldots, f_k \rangle$. Given an ideal $I \subseteq \mathcal{S}^n$ we let I_d denote the subspace of I of degree-d polynomials.

The set of $n \times m$ matrices with entries in a field k is denoted by $\mathsf{k}^{n \times m}$. For a matrix $A \in \mathsf{k}^{n \times m}$ and a multi-index $\alpha \in \mathbb{N}^n$, we let A_α be the $n \times |\alpha|$ matrix whose first α_1 columns are the first column of A, next α_2 columns are the second column of A, etc. We let $\det_d, \operatorname{per}_d \in \mathsf{k}[x_{ij} : i,j \in [d]]_d$ denote the degree-d determinant and permanent polynomials, respectively. Recall that the permanent is defined by

$$\operatorname{per}_d = \sum_{\sigma \in \mathfrak{S}_d} \prod_{i=1}^d x_{i,\sigma(i)},$$

where $\mathfrak{S}_d$ denotes the symmetric group on d letters.

The subsequent theorems are classical and easily verified. The first is the crux of this chapter. The second shows that Waring rank is always defined (i.e., finite).

Theorem 2.2.1. *Let $f \in \mathcal{S}^n_d$ and let $j \geq d$.*

a) [IK99, Lemma 1.15(i)] For all $\ell_1, \dots, \ell_r \in \mathcal{S}_1^n$,

$$f(\partial \mathbf{x}) \sum_{i=1}^{r} \ell_i^j = d! \sum_{i=1}^{r} f(\ell_i^*) \ell_i^{j-d}.$$

b) [CGLM08, Lemma 3.5] For all $g \in \mathcal{S}_d^n$, $f(\partial \mathbf{x}) g = g(\partial \mathbf{x}) f$.

Theorem 2.2.2. *[IK99, Corollary 1.16] $\mathbf{R}_S(f) \leq \dim \mathcal{S}_d^n = \binom{n+d-1}{d}$.*

Importantly, Theorems 2.2.1 (a) and 2.2.1 (b) imply that $g(\partial \mathbf{x}) f$ can be computed with $\mathbf{R}_S(g)$ queries in Problem 2.1.12 (a), as noted in Observation 2.1.5. We will show in the next subsection that this is optimal, even if we are allowed to query f adaptively.

Example 2.2.3. The following Waring decomposition of $e_{n,d}$ is easily seen by inclusion-exclusion:

$$d! \cdot e_{n,d} = \sum_{\substack{\alpha \in \{0,1\}^n \\ |\alpha| \leq d}} (-1)^{|\alpha|+d} \binom{n-|\alpha|}{d-|\alpha|} \left(\sum_{i=1}^{n} \alpha_i x_i \right)^d . \tag{2.5}$$

In fact, this decomposition is *synonymous* with inclusion-exclusion in many exact algorithms, as we now illustrate. For $A \in \mathbb{C}^{n \times n}$, let

$$\mathrm{Prod}_A \coloneqq (A_{1,1} x_1 + \cdots + A_{1,n} x_n) \cdots (A_{n,1} x_1 + \cdots + A_{n,n} x_n) \in \mathcal{S}_n^n .$$

It is easily seen that the coefficient of $x_1 \cdots x_n$ in Prod_A equals the permanent of A. In other words, $\mathrm{per}(A) = e_{n,n}(\partial \mathbf{x}) \mathrm{Prod}_A$. It follows directly from Theorem 2.2.1 and Equation (2.5) that

$$\mathrm{per}(A) = \sum_{\alpha \in \{0,1\}^n} (-1)^{|\alpha|+n} \mathrm{Prod}_A(\alpha),$$

which is Ryser's formula for computing the permanent [Rys64]. As another example, applying Theorem 2.2.1 and Equation (2.5) to the closed-walk generating polynomial Section 2.5, one finds that the number of Hamiltonian cycles in G equals

$$\sum_{\alpha \in \{0,1\}^n} (-1)^{|\alpha|+n} \mathrm{tr}(A_G^n)(\alpha),$$

which was first given in [KGK77] and rediscovered several times thereafter [Kar82, Bax93]. As a third example, let $S_1, \dots, S_m \subseteq [k \cdot r]$, where $|S_i| = r$ for all i. Note that that the coefficient of $x_1 \cdots x_{kr}$ in $\mathrm{Part}_{S_1,\dots,S_m} \coloneqq \left(\sum_{i=1}^m \prod_{j \in S_i} x_j \right)^k$ equals the number of ordered partitions of $[kr]$ into k of the sets S_i. Therefore the number of such partitions equals

$$\sum_{\alpha \in \{0,1\}^{kr}} (-1)^{|\alpha|+kr} \mathrm{Part}_{S_1,\dots,S_m}(\alpha),$$

which was given in [BH06, BHK09]. The fastest known algorithms for computing the permanent and counting Hamiltonian cycles and set partitions follow from the straightforward evaluation of the above formulas. A similar perspective on these algorithms appeared earlier in [Bar96].

Understanding these algorithms from the perspective of Waring decompositions is extremely insightful, and was our initial motivation. For example, it is clear from the above argument that *any* Waring decomposition of $x_1 \cdots x_n$ yields an algorithm for the above problems – there is nothing special about Equation (2.5). This immediately raises the question: what is $\mathbf{R}_S(x_1 \cdots x_n)$? This was only answered recently in [RS11], where a lower bound on the *degree* of a form's *apolar subscheme* was used to show that $\mathbf{R}_S(x_1 \cdots x_n) = 2^{n-1}$.[5] This lower bound shows that the above algorithms are, in a restricted sense, optimal. Similar observations have been made in [Gur03, Gly13].

Although the Waring decomposition of Equation (2.5) is essentially optimal in the case when $n = d$, it is far from optimal in general. Indeed, Equation (2.5) only shows that $\mathbf{R}_S(e_{n,d}) \leq \binom{n}{\leq d}$, whereas it was shown in [Lee16] that for d odd, $\mathbf{R}_S(e_{n,d}) = \binom{n}{\leq \lfloor d/2 \rfloor}$, and for d even,

$$\binom{n}{\leq d/2} - \binom{n-1}{d/2} \leq \mathbf{R}_S(e_{n,d}) \leq \binom{n}{\leq d/2}.$$

2.2.1 Apolarity and the method of partial derivatives

Fix $g \in \mathcal{S}^n_d$. For integers $u, v \geq 0$ such that $u + v = d$, let $Cat_g(u, v) : \mathcal{S}^n_u \to \mathcal{S}^n_v$ be given by

$$Cat_g(u, v)(f) \coloneqq f(\partial \mathbf{x}) g.$$

These maps, called *catalecticants*, were first introduced by J.J. Sylvester in 1852 [Syl52]. Their importance is due in large part to the following method for obtaining Waring rank lower bounds, known as the *method of partial derivatives* in complexity theory [Lan17, Section 6.2.2].

Theorem 2.2.4. *[IK99, pg. 11] For all $g \in \mathcal{S}^n_d$ and integers $u, v \geq 0$ such that $u + v = d$,*

$$\mathbf{R}_S(g) \geq \operatorname{rank}(Cat_g(u, v)).$$

Remark 2.2.5. As a matrix, $Cat_g(u, v)$ has $\binom{n+u-1}{u}$ columns, indexed by the degree-u monomials in $x_1, \ldots, x_n$, and $\binom{n+v-1}{v}$ rows, indexed by the degree-v monomials in $x_1, \ldots, x_n$. Therefore the best rank lower bound Theorem 2.2.4 can give is $\binom{n+\lceil d/2 \rceil - 1}{\lceil d/2 \rceil}$, which is obtained when $u = \lceil d/2 \rceil, v = \lfloor d/2 \rfloor$. In contrast, it is known [Lan12, Section 3.2] that the rank for *almost all* $g \in \mathcal{S}^n_d$ is at least $\lceil \binom{n+d-1}{d} / n \rceil$ (with respect to a natural distribution on forms), so the method of partial derivatives is far from optimal. Finding methods for proving better lower bounds is a significant barrier and a topic of great interest from both an algebraic-geometric and complexity-theoretic perspective; see [Lan17, Section 10.1] and [EGOW18].

Example 2.2.6. It is a classical fact from linear algebra that for $g \in \mathcal{S}^n_2$, $\mathbf{R}_S(g) = \operatorname{rank}(Cat_g(1, 1))$. Explicitly, this says that $g = \sum_{1 \leq i \leq j \leq n} A_{ij} x_i x_j$ can be written as a sum of at most r squares of linear forms if and only if the matrix $A = (A_{ij})$ has rank at most r. Hence Waring rank can be viewed as a higher dimensional generalization of symmetric matrix rank.

Let $g^{\perp}_j \coloneqq \ker Cat_g(j, d - j)$ be the set of degree-j forms annihilating g under the differentiation action. The next fact is known as the *Apolarity Lemma* in the Waring rank literature.

[5] A lower bound of $\binom{n}{\lfloor n/2 \rfloor}$ can be shown easily using the *method of partial derivatives*, presented in the next subsection.

Lemma 2.2.7. *[Tei14, Theorem 4.2] Let $\ell_1, \dots, \ell_r \in \mathcal{S}_1^n$ be pairwise linearly independent. Then for all $g \in \mathcal{S}_d^n$, $g \in \operatorname{span}\{\ell_1^d, \dots, \ell_r^d\}$ if and only if $I(\{\ell_1^*, \dots, \ell_r^*\})_d \subseteq g_d^\perp$.*

A complete answer to the complexity of Problem 2.1.12 (a) is now in hand.

Theorem 2.2.8. *Fix $g \in \mathcal{S}_d^n$ and let $f \in \mathcal{S}_d^n$ be given as a black-box. The minimum number of queries to f needed to compute $g(\partial \mathbf{x})f$ is $\mathbf{R}_S(g)$, assuming unit-cost arithmetic operations.*

Proof. The upper bound is immediate from Theorem 2.2.1 (b). To prove the lower bound we first show the following: for any pairwise linearly independent points $v_1, \dots, v_m \in \mathbb{C}^n$ where $m < \mathbf{R}_S(g)$, there exists a $p \in \mathcal{S}_d^n$ such that $p \in I(\{v_1, \dots, v_m\})$ but $g(\partial \mathbf{x})p \neq 0$. If this were not the case, there exist pairwise linearly independent points $v_1, \dots, v_m$ such that $I(\{v_1, \dots, v_m\})_d \subseteq g_d^\perp$. But this implies that g has rank at most m by the Apolarity Lemma, a contradiction.

So now given any $f \in \mathcal{S}_d^n$, suppose that our algorithm queries f at $v_1, \dots, v_m$, which can be assumed to be pairwise linearly independent. By the above argument, there exists some $p \in \mathcal{S}_d^n$ such that $(p+f)(v_i) = p(v_i) + f(v_i) = f(v_i)$ for all $i \in [m]$, and hence the algorithm cannot distinguish f from $p + f$, but at the same time $g(\partial \mathbf{x})f \neq g(\partial \mathbf{x})(p+f)$. ∎

2.2.2 Support rank, nonnegative support rank, and ε-support rank

We now introduce variants of Waring rank of algorithmic relevance.

Definition 2.2.9. The support rank and nonnegative support rank of $f \in \mathcal{S}_d^n$ are given by

$$\mathbf{R}_{\text{supp}}(f) := \min(\mathbf{R}_S(g) : g \in \mathcal{S}_d^n, \operatorname{supp}(g) = \operatorname{supp}(f)),$$
$$\mathbf{R}_{\text{supp}}^{+}(f) := \min(\mathbf{R}_S(g) : g \in \mathbb{R}_{\geq 0}[x_1, \dots, x_n]_d, \operatorname{supp}(g) = \operatorname{supp}(f)).$$

Furthermore, if $f \in \mathbb{R}_{\geq 0}[x_1, \dots, x_n]_d$, the ε-support rank of f is given by

$$\mathbf{R}_{\text{supp}}^{\varepsilon}(f) := \min(\mathbf{R}_S(g) : g \in \mathbb{R}[x_1, \dots, x_n]_d, \forall \alpha \in \mathbb{N}_d^n,$$
$$(1-\varepsilon) \cdot \partial^\alpha f \leq \partial^\alpha g \leq (1+\varepsilon) \cdot \partial^\alpha f).$$

Note that condition in the definition of $\mathbf{R}_{\text{supp}}^{\varepsilon}$ is simply that the coefficient of x^α in g is bounded by a factor of $(1 \pm \varepsilon)$ times the coefficient of x^α in f.

Roughly speaking, support rank corresponds to decision algorithms, nonnegative support rank to *deterministic* decision algorithms, and ε-support rank to deterministic *approximate counting* algorithms. This is now formalized.

Definition 2.2.10. For $g \in \mathcal{S}_d^n$ and $0 < \delta < 1$, a *g-support intersection certification algorithm* with one-sided error δ is an algorithm which, given any $f \in \mathcal{S}_d^n$ as a black-box, outputs "$\operatorname{supp}(f) \cap \operatorname{supp}(g) = \emptyset''$ on all instances f where $\operatorname{supp}(f) \cap \operatorname{supp}(g) = \emptyset$, and correctly outputs "$\operatorname{supp}(f) \cap \operatorname{supp}(g) \neq \emptyset''$ with probability at least $1 - \delta$ on all instances where $\operatorname{supp}(f) \cap \operatorname{supp}(g) \neq \emptyset$.

Proposition 2.2.11. *a) For all $g \in \mathcal{S}_d^n$ and $\delta > 0$, there is a g-support intersection certification algorithm with one-sided error δ that makes $\mathbf{R}_{\text{supp}}(g)$ queries.*

b) For a fixed $g \in \mathcal{S}_d^n$ and all $f \in \mathbb{R}_{\geq 0}[x_1, \dots, x_n]_d$ given as a black-box, there is a deterministic algorithm that decides if $\operatorname{supp}(g) \cap \operatorname{supp}(f)$ using $\mathbf{R}_{\text{supp}}^{+}(g)$ queries.

c) *For a fixed $g \in \mathbb{R}_{\geq 0}[x_1, \ldots, x_n]_d$ and all $f \in \mathbb{R}_{\geq 0}[x_1, \ldots, x_n]_d$ given as a black-box, there is a deterministic algorithm that computes a $(1 \pm \varepsilon)$-approximation to $g(\partial \mathbf{x})f$ using $\mathbf{R}^{\varepsilon}_{\text{supp}}(g)$ queries.*

Proof. a. Let $U \subseteq \mathbb{C}$, where $|U| \geq d/\delta$. Let $a_1, \ldots, a_n$ be indeterminates. Note that $g(\partial \mathbf{x})f(a_1x_1, \ldots, a_nx_n)$ is not identically zero in $\mathbb{C}[a_1, \ldots, a_n]$ if and only if $\text{supp}(f) \cap \text{supp}(g) \neq \emptyset$. Then by choosing $a_1, \ldots, a_n$ uniformly at random from U, it follows that $g(\partial \mathbf{x})\ f(a_1x_1, \ldots, a_nx_n)$ will evaluate to zero whenever $\text{supp}(f) \cap \text{supp}(g) = \emptyset$, and whenever $\text{supp}(f) \cap \text{supp}(g) \neq \emptyset$ this does not evaluate to zero with probability at least $1 - \delta$ by the Schwarz-Zippel lemma. By Theorem 2.2.1, $g(\partial \mathbf{x})f(a_1x_1, \ldots, a_nx_n)$ can be computed using $\mathbf{R}_S(g)$ queries, and the conclusion follows.

b. If both f and g have nonnegative coefficients, then $g(\partial \mathbf{x})f > 0$ if and only if $\text{supp}(f) \cap \text{supp}(g) \neq \emptyset$. The result follows from Theorem 2.2.1.

c. This is immediate from Theorem 2.2.1. ■

It follows from a variation of the proof of Theorem 2.1.6 that Proposition 2.2.11 (a) is optimal for monomials:

Proposition 2.2.12. *For all $\alpha \in \mathbb{N}^n$ and all $\delta < 1$, any x^α-support intersection certification algorithm with one-sided error δ makes at least $\mathbf{R}_{\text{supp}}(x^\alpha) = \prod_{i=1}^n (1 + \alpha_i)/\min_{i \in [n]}(1 + \alpha_i)$ queries.*

Proof. The upper bound follows from Theorem 2.2.1 (b); in fact, this shows that we can compute $\partial^\alpha f$ exactly using $\mathbf{R}_S(x^\alpha)$ queries.

For the lower bound, given any $f \in \mathcal{S}^n_d$ where $\alpha \in \text{supp}(f)$, suppose a support intersection certification algorithm queries f at pairwise linearly independent points $v_1, \ldots, v_m$, where $m < \mathbf{R}_S(x^\alpha)$. Then by the Apolarity Lemma, there exists a $p \in \mathcal{S}^n_d$ such that $p \in I(\{v_1, \ldots, v_m\})$ but $\partial^\alpha p \neq 0$ (see the proof of Theorem 2.1.6). Note that the condition that $\partial^\alpha p \neq 0$ is equivalent to saying that $\alpha \in \text{supp}(p)$. Therefore there exists some $\lambda \in \mathbb{C}$ such that $\alpha \notin \text{supp}(f + \lambda p)$. But note that $(f + \lambda p)(v_i) = f(v_i) + \lambda p(v_i) = f(v_i)$ for all $i \in [m]$, and hence the algorithm cannot distinguish between f and $f + \lambda p$. Since the algorithm has no false negatives, it must always give the incorrect answer on f. We conclude by the matching upper and lower bounds on $\mathbf{R}_S(x^\alpha)$ given in [CCG11]. ■

Theorem 2.2.13. *Any $e_{n,d}$-support intersection certification algorithm with one-sided error δ makes at least 2^{d-1} queries.*

Proof. Suppose for contradiction that such an algorithm made fewer queries. Then given f as a black-box, we run this algorithm with access to $f(x_1, \ldots, x_d, 0, \ldots, 0)$. By definition, this algorithm always answers correctly if the coefficient of $x_1 \cdots x_d$ is zero, and answers correctly with probability at least $1 - \delta$ if this coefficient is nonzero. But this gives an $x_1 \cdots x_d$-support intersection certification algorithm with one-sided error δ making fewer than 2^{d-1} queries. Since $\mathbf{R}_S(x_1 \cdots x_d) = 2^{d-1}$ [RS11], this contradicts Proposition 2.2.12. ■

2.3 Support ranks of elementary symmetric polynomials

We are now ready to study $A(n,d)$ and its variants, which we now recall.

Problem 2.3.1. Determine $A(n,d) := \mathbf{R}_{\text{supp}}(e_{n,d})$, $A^+(n,d) := \mathbf{R}^+_{\text{supp}}(e_{n,d})$ and $A^\varepsilon(n,d) := \mathbf{R}^\varepsilon_{\text{supp}}(e_{n,d})$.

Obviously $A(n,d) \le A^+(n,d) \le A^\varepsilon(n,d)$, and for all n, $A(n,1) = 1$. It follows from [RS11] that $A^\varepsilon(n,n) = 2^{n-1}$ and from [Lee16] that $A^\varepsilon(n,d) \le \binom{n}{\le \lfloor d/2 \rfloor}$; the latter turns out to be arbitrarily far from optimal, however.

We will be interested in Problem 2.3.1 as n goes to infinity. To facilitate this, we adopt the notation $A(\mathbb{N},d) := \lim_{n\to\infty} A(n,d)$, defining $A^+(\mathbb{N},d)$ and $A^\varepsilon(\mathbb{N},d)$ analogously. We will show in Proposition 2.3.3 (a) that $A(n,d)$, $A^+(n,d)$, and $A^\varepsilon(n,d)$ are nondecreasing in n, in Proposition 2.3.14 that $A^+(\mathbb{N},d)$ is finite for each d, and in Corollary 2.3.12 that $A^\varepsilon(\mathbb{N},d)$ is infinite for $\varepsilon < 1/2$ and $d > 1$.

For notational convenience, we define

$$\begin{aligned}
\mathfrak{E}(n,d) &:= \{f \in \mathcal{S}^n_d : \operatorname{supp}(f) = \operatorname{supp}(e_{n,d})\},\\
\mathfrak{E}^+(n,d) &:= \{f \in \mathfrak{E}(n,d) : \forall \alpha \in \{0,1\}^n_d,\ \partial^\alpha f \in \mathbb{R}^+\},\\
\mathfrak{E}^\varepsilon(n,d) &:= \{f \in \mathfrak{E}^+(n,d) : \forall \alpha \in \{0,1\}^n_d,\ \partial^\alpha f \in (1 \pm \varepsilon))\}.
\end{aligned}$$

Remark 2.3.2. Our upper bounds to Problem 2.3.1 will be obtained by the following general method. We start with some $f \in \mathcal{S}^m_d$ whose rank is known. We then find $L_1,\dots,L_m \in \mathcal{S}^n_1$, where $n \gg m$, so that $f(L_1,\dots,L_m) \in \mathfrak{E}(n,d)$. This will show that

$$A(n,d) \le \mathbf{R}_S(f(L_1,\dots,L_m)) \le \mathbf{R}_S(f).$$

For example, we first show that $A^+(\mathbb{N},d) < 6.75^d$ by taking f to be the determinant of a generic Hankel matrix, and $\ell_1,\dots,\ell_n$ to be given by rank-1 Hankel matrices (points on the *rational normal scroll*). We later use this method to show that $A^\varepsilon(n,d) \le O(4.075^d \varepsilon^{-2} \log n)$ by taking f to be a "direct sum" of $e_{\lfloor 1.55d \rfloor,d}$ and $L_1,\dots,L_n$ to be given by a $(1+\varepsilon)$-balanced splitter. We note in Remark 2.3.35 that color-coding can be viewed as taking f to be a direct sum of $x_1x_2\cdots x_d$ and $L_1,\dots,L_m$ to be a perfect hash family. A simple geometric property that f and $L_1,\dots,L_m$ must satisfy in this method is given by Proposition 2.3.5.

2.3.1 Lower bounding $A(n,d)$ and the $d=2$ case

We start with some simple relations between different values of $A(n,d)$ that will be used throughout this section.

Proposition 2.3.3. *For all $n \ge d$,*

a) $A(n,d) \le A(n+1,d)$,

b) $A(n,d) \le A(n+1,d+1)$.

Moreover, these statements remain valid when "A" is replaced with A^+ and A^ε.

Proof. a. Suppose $f \in \mathfrak{E}(n+1,d)$, and let f' be obtained from f by setting $x_{n+1} = 0$. Then clearly $\mathbf{R}_S(f') \le \mathbf{R}_S(f)$ and $f' \in \mathfrak{E}(n,d)$. Therefore $A(n,d) \le A(n+1,d)$.

b. If $f \in \mathfrak{E}(n+1, d+1)$, then $\partial_{n+1} f \in \mathfrak{E}(n, d)$. Hence $A(n,d) \leq \mathbf{R}_S(\partial_{n+1} f) \leq \mathbf{R}_S(f)$, where the final inequality follows from Theorem 2.2.1 (a).

It is easy to see that the same arguments hold if we replace $A(n,d)$ with $A^+(n,d)$ or $A^\varepsilon(n,d)$. ■

Theorem 2.3.4. *For all $n \geq d$,*

$$2^{d-1} \leq A(n,d) \leq A^+(n,d) \leq A^\varepsilon(n,d).$$

Proof. It was shown in [RS11] that $\mathbf{R}_S(x_1 \cdots x_d) = 2^{d-1}$, and therefore $A(d,d) = 2^{d-1}$. The theorem is then immediate from Proposition 2.3.3 (a). ■

We now give an insightful geometric characterization of $A(n,d)$.

Proposition 2.3.5. *$A(n,d) \leq r$ if and only if for some m there exists $f \in \mathcal{S}_d^m$ and points $v_1, \ldots, v_n$ in $\mathbb{C}^m$ such that $\mathbf{R}_S(f) \leq r$ and f vanishes on the span of any $d-1$ of the points $v_1, \ldots, v_n$, but not on the span of any d of them.*

Proof. Suppose that $A(n,d) \leq r$. By definition, there exists a $f \in \mathfrak{E}(n,d)$ with $\mathbf{R}_S(f) \leq r$. It follows that f vanishes on the span of the span of any $d-1$ of the standard basis vectors in $\mathbb{C}^n$, but not on the span of any d of them.

Conversely, suppose there exists such an f and points $v_1, \ldots, v_n$, and let

$$f' := f(x_1 v_1 + \cdots + x_n v_n).$$

It is immediate that $\mathbf{R}_S(f') \leq \mathbf{R}_S(f)$. Additionally, f' must be multilinear as f vanishes on the span of any $d-1$ of the points $v_1, \ldots, v_n$. But then for $\alpha \in \{0,1\}_d^n$, the coefficient of x^α in f' is given by $f'(\alpha) = f(\sum_{i=1}^n \alpha_i v_i)$. If this was zero f would vanish on the span of the d points $\{v_i : i \in \operatorname{supp}(\alpha)\}$, a contradiction. This shows that $f' \in \mathfrak{E}(n,d)$, proving the claim. ■

Proposition 2.3.6. *$A(n,d) \leq r$ if and only if there exist n points in $\mathbb{C}^r$ such that the span of any $d-1$ of them is contained in $\mathbf{V}(\sum_{i=1}^r x_i^d)$, but the span of any d of them is not.*

Proof. If $A(n,d) \leq r$, then for some $f \in \mathfrak{E}(n,d)$ and linear forms $\ell_1, \ldots, \ell_r$, $f = \sum_{i=1}^r \ell_i^d$. Let $v_j := ((\ell_1{}^*)_j, (\ell_2{}^*)_j, \ldots, (\ell_r{}^*)_j)$ for all $j \in [n]$. Since f is multilinear, $\sum_{i=1}^r x_i^d$ must vanish on the span of any $d-1$ of the points $v_1, \ldots, v_n$, and since each multilinear monomial has a nonzero coefficient, $\sum_{i=1}^r x_i^d$ does not vanish on the span of any d of $v_1, \ldots, v_n$.

Conversely, suppose that there exists such a set of points. Since $\sum_{i=1}^r x_i^d$ has rank r, by Proposition 2.3.5 we conclude that $A(n,d) \leq r$. ■

Corollary 2.3.7. $5 \leq A(8,3) \leq A(\mathbb{N},3)$.

Proof. Suppose for contradiction that $A(8,3) = 4$. By Proposition 2.3.6, this implies that there are 8 points in $\mathbb{C}^4$ such that the planes spanned by any two of them are contained in $\mathbf{V}(x_1^3 + x_2^3 + x_3^3 + x_4^3)$, but the span of any three of them is not. Note that this is only possible if no three points are coplanar, and hence the $\binom{8}{2} = 28$ planes spanned by any two points are distinct. But by the Cayley-Salmon theorem, $\mathbf{V}(x_1^3 + x_2^3 + x_3^3 + x_4^3)$ contains exactly $27 < 28$ lines in the projective space $\mathbb{CP}^3$ [Gat14, Lemma 11.1], a contradiction. ■

Remark 2.3.8. A similar proof fails to show that $6 \leq A(\mathbb{N}, 3)$, as $\mathbb{P}(\mathbf{V}(x_1^3 + \cdots + x_5^3))$ contains infinitely many lines (see [Gat14, Exercise 11.10.b]).

The $d = 2$ case of Problem 2.3.1 is solved using linear algebra.

Proposition 2.3.9. $A(\mathbb{N}, 2) = 3$.

Proof. It suffices by Example 2.2.6 to show that for $n \geq 3$, the minimum rank of a symmetric $n \times n$ matrix with zeros on the diagonal and nonzero values elsewhere is 3. There is a lower bound of 3 since the principal 3×3 minor of any such matrix is easily seen to be nonzero. An upper bound of 3 is given by the matrix $((i - j)^2)_{i,j \in [n]}$. ■

To understand $A^\varepsilon(n, 2)$ we will need the following fact:

Theorem 2.3.10. *[Alo03, Theorem 9.3] Let B be an n-by-n real matrix with $b_{i,i} = 1$ for all i and $|b_{i,j}| \leq \varepsilon$ for all $i \neq j$. Then if $1/\sqrt{n} \leq \varepsilon < 1/2$,*

$$\operatorname{rank}(B) \geq \Omega\left(\frac{\log n \cdot \varepsilon^{-2}}{\log(\varepsilon^{-1})}\right).$$

Proposition 2.3.11. *a) If $1/\sqrt{n} \leq \varepsilon < 1/2$,*

$$A^\varepsilon(n, 2) \geq \Omega\left(\frac{\log n \cdot \varepsilon^{-2}}{\log(\varepsilon^{-1})}\right).$$

b) For all $\varepsilon > 0$,

$$A^\varepsilon(n, 2) \leq O\left(\log n \cdot \varepsilon^{-2}\right).$$

Proof. It follows from Example 2.2.6 that $A^\varepsilon(n, 2)$ is the minimum rank among all real symmetric matrices A with $A_{i,i} = 0$ and $A_{i,j} \in [1 - \varepsilon, 1 + \varepsilon]$ for all $i \neq j$. Note that given any such A, the matrix $J - A$ (where J denotes the all-ones matrix) has diagonal entries equal to 1, off-diagonal entries bounded in absolute value by ε, and rank at most $\operatorname{rank}(A) + 1$. Conversely, given any symmetric matrix B with $b_{i,i} = 1$ for all i and $|b_{i,j}| \leq \varepsilon$ for all $i \neq j$, the matrix $J - B$ has zeros on the diagonal, off-diagonal entries in the range $[1 - \varepsilon, 1 + \varepsilon]$, and rank at most $\operatorname{rank}(B) + 1$. So it suffices to determine the minimum rank of such a matrix B. Given this observation, (a) is immediate from Theorem 2.3.10.

To show (b), let $m := O(\log n/\varepsilon^2)$. By the Johnson-Lindenstrauss Lemma, there exist unit vectors $v_1, \ldots, v_n \in \mathbb{R}^m$ such that $|v_i \cdot v_j| \leq \varepsilon$ for all $i \neq j$. It follows that the matrix $(v_i^T \cdot v_j)_{i,j \in [n]}$ has the desired properties and rank at most m. ■

Corollary 2.3.12. *For all $0 < \varepsilon < 1/2$ and $d \geq 2$, $A^\varepsilon(\mathbb{N}, d) = \infty$.*

Proof. Fix $0 < \varepsilon < 1/2$. By Proposition 2.3.11 (a),

$$A^\varepsilon(n, 2) \geq \Omega\left(\frac{\log n \cdot \varepsilon^{-2}}{\log(\varepsilon^{-1})}\right)$$

for all $n \geq \varepsilon^{-2}$, and so $A^\varepsilon(\mathbb{N}, 2) = \infty$. Now suppose that $A^\varepsilon(\mathbb{N}, d)$ is bounded above for some $d > 2$. Then by Proposition 2.3.3, for all n

$$A^\varepsilon(n, 2) \leq A^\varepsilon(n + d - 2, d) \leq A^\varepsilon(\mathbb{N}, d),$$

a contradiction. ■

2.3.2 Upper bounds via the determinant

The relevance of the determinant to Problem 2.3.1 is immediate from Proposition 2.3.5. The obvious but key observation is that for all n, d with $n \geq d$, a generic set of n rank-1 $d \times d$ matrices has the property that the sum of any d of them is invertible, and hence the span of any $d-1$ of them is contained in $\mathbf{V}(\det_d)$ but the span of any d of them is not. Applying Proposition 2.3.5, we conclude that $A(n,d) \leq \mathbf{R}_S(\det_d)$. We now make this more explicit.

Definition 2.3.13. Let $d \leq n$. For $A, B \in \mathbb{C}^{d\times n}$, let

$$g_{A,B} \coloneqq \sum_{\alpha \in \{0,1\}^n_d} \det_d(A_\alpha B_\alpha) x^\alpha. \tag{2.6}$$

Proposition 2.3.14. *For all $A, B \in \mathbb{C}^{d\times n}$,*

$$\mathbf{R}_S(g_{A,B}) \leq \mathbf{R}_S(\det_d) \leq (5/6)^{\lfloor d/3 \rfloor} 2^{d-1} d!.$$

Furthermore, $A^+(\mathbb{N}, d) \leq \mathbf{R}_S(\det_d) \leq (5/6)^{\lfloor d/3 \rfloor} 2^{d-1} d!$ and $A^+(\mathbb{N}, d)$ exists.

Proof. Let $X = \operatorname{diag}(x_1, \ldots, x_n)$. By the Cauchy-Binet formula it follows that $\det_d((A \cdot X) \cdot B^T) = g_{A,B}$. The first statement then follows from the fact that $\mathbf{R}_S(\det_d) \leq (5/6)^{\lfloor d/3 \rfloor} 2^{d-1} d!$ [Tei14, Example 1.14].

Note that by taking A and B to have positive minors[6], $g_{A,B} \in \mathfrak{E}^+(n,d)$. This shows that $A^+(\mathbb{N}, d) \leq \mathbf{R}_S(\det_d)$. Since Proposition 2.3.3 (a) shows that $(A^+(n,d))_n$ is nondecreasing, it follows that the limit $A^+(\mathbb{N}, d)$ exists. ■

Remark 2.3.15. The asymptotically best known lower bound on $\mathbf{R}_S(\det_d)$ is $\binom{d}{\lfloor d/2 \rfloor}^2$, which follows from the method of partial derivatives [Gur03] [Lan12, Theorem 9.3.2.1]. Therefore one cannot hope to improve the upper bound given by Proposition 2.3.14 exponentially beyond 4^d by finding a better upper bound on the Waring rank of the determinant.

Definition 2.3.16. Let $h_d \in \mathcal{S}_d^{2d-1}$ be the determinant of a symbolic Hankel matrix (that is, the determinant of the $d \times d$ matrix whose (i,j)th entry is the variable x_{i+j}).

Theorem 2.3.17.

$$A^+(\mathbb{N}, d) \leq \mathbf{R}_S(h_d) \leq \binom{3d-2}{d} < 6.75^d.$$

Proof. Let $a_1, a_2, \ldots, a_n$ be distinct elements of $\mathbb{R}$, let

$$A = (a_i^{j-1})_{i \in [n], j \in [d]} \in \mathbb{C}^{d \times n},$$

and let $X = \operatorname{diag}(x_1, \ldots, x_n)$. By the Cauchy-Binet formula,

$$\det_d((A \cdot X) \cdot A^T) = g_{A,A} = \sum_{\alpha \in \{0,1\}^n_d} \det_d(A_\alpha A_\alpha) x^\alpha = \sum_{\alpha \in \{0,1\}^n_d} \det_d(A_\alpha)^2 x^\alpha.$$

[6] For instance, by taking the columns of A and B to be given by real Vandermonde vectors.

Since A is a Vandermonde matrix, $\det_d(A_\alpha)^2 > 0$ for all $\alpha \in \{0,1\}^n_d$. Hence $g_{A,A} \in \mathfrak{E}^+(n,d)$. Now observe that $(A \cdot X) \cdot A^T$ is a Hankel matrix; explicitly, it equals

$$\sum_{i=1}^n (1, a_i^1, \ldots, a_i^{d-1})^T (1, a_i^1, \ldots, a_i^{d-1}) x_i.$$

Therefore $\det_d(AXA^T) = h_d(AXA^T)$, and so $A^+(\mathbb{N}, d) \le \mathbf{R}_S(h_d)$. Since h_d is a degree-d polynomial in $2d-1$ variables, by the dimension bound of Theorem 2.2.2 we have that $\mathbf{R}_S(h_d) \le \binom{3d-2}{d}$, and therefore $A^+(\mathbb{N}, d) \le \binom{3d-2}{d}$. The theorem follows from Stirling's approximation. ■

Remark 2.3.18. The above theorem can be slightly improved by using the state-of-the-art bound [Jel13] on the maximum Waring rank in $\mathcal{S}^n_d$ of

$$\binom{n+d-2}{d-1} - \binom{n+d-6}{d-3},$$

valid when $n, d \ge 3$, which shows that

$$A^+(n,d) \le \mathbf{R}_S(h_d) \le \binom{3d-3}{d-1} - \binom{3d-7}{d-3}.$$

It follows from Remark 2.2.5 that the lower bound on $\mathbf{R}_S(h_d)$ given by the method of partial derivatives is at most $\binom{\lceil 5d/2 \rceil - 1}{\lceil d/2 \rceil} < 3.5^d$. The next theorem shows that the actual lower bound obtained by the method of partial derivatives is exponentially worse than this.

Theorem 2.3.19. *For all integers $d, u, v > 0$ such that $u + v = d$,*

$$\operatorname{rank}(Cat_{h_d}(u,v)) \le \binom{\lceil 3d/2 \rceil}{\lfloor d/2 \rfloor} < 2.6^d.$$

Proof. First note that if $A = \text{Vandermonde}(a_1, \ldots, a_n; d) = (a_i^{j-1}) \in \mathbb{C}^{d \times n}$ with $a_1, \ldots, a_n$ distinct, $g_{A,A}$ equals h_d up to a change of variables. This implies that $\operatorname{rank}(Cat_{h_d}(u,v)) = \operatorname{rank}(Cat_{g_{A,A}}(u,v))$. So we will equivalently work with $f := g_{A,A}$. Furthermore we assume that $u \le v$; this is without loss of generality as $Cat_f(u,v) = Cat_f(v,u)^T$. We will then show that $\operatorname{rank}(Cat_f(u,v)) \le m := \binom{2v+u}{u}$. As this is maximized when $u = \lfloor d/2 \rfloor$, $v = \lceil d/2 \rceil$, the theorem follows.

The matrix $Cat_f(u,v)$ has rows indexed by monomials x^α, where $\alpha \in \mathbb{N}^{2d-1}_u$, and columns indexed by monomials x^β, where $\beta \in \mathbb{N}^{2d-1}_v$. Because f is multilinear, the entries in a row indexed by a non-multilinear monomial x^α will be zero, as x^α annihilates f under differentiation. Similarly, any column indexed by a non-multilinear monomial will have all entries equal to zero. Therefore it suffices to consider the submatrix M of $Cat_f(u,v)$ indexed by multilinear monomials. We identify the row/column corresponding to x^α with the set $\operatorname{supp}(\alpha) \subseteq [2d-1]$.

Note that M_{IJ} (the entry of M at row I and column J) equals 0 if I and J have a nonempty intersection, and equals $\prod_{i \ne j \in I \cup J} (a_i - a_j)^2$ otherwise. Hence the row indexed by I is a multiple of $\prod_{i \ne j \in I} (a_i - a_j)^2$, and similarly the column indexed by J is a multiple of $\prod_{i \ne j \in J} (a_i - a_j)^2$.

Therefore $M = D_1QD_2$ for some invertible (diagonal) matrices D_1 and D_2, and so it suffices to upper bound the rank of Q.

Next, observe that $Q_{IJ} = \prod_{i\in I, j\in J}(a_i - a_j)^2$. Write $I = \{i_1, \dots, i_u\}$, $J = \{j_1, \dots, j_v\}$. We now claim that there exist $g_1, h_1, \dots, g_m, h_m$ with $g_i \in \mathcal{S}^u$, $h_i \in \mathcal{S}^v$, such that

$$Q_{IJ} = \sum_{k=1}^{m} g_k(a_{i_1}, \dots, a_{i_u})h_k(a_{j_1}, \dots, a_{j_v}). \tag{2.7}$$

To see this, view Q_{IJ} as a polynomial in the variables $a_{i_1}, \dots, a_{i_u}$ with coefficients in $\mathbb{C}[a_{j_1}, \dots, a_{j_v}]$. This is a symmetric polynomial in u variables, where the maximum degree of any variable in any monomial is $2v$. Therefore Q_{IJ} can be written as in Equation (2.7) as a sum over symmetrizations of monomials with total degree at most u and maximum individual degree $2v$, for some coefficients h_k in $\mathbb{C}[a_{j_1}, \dots, a_{j_v}]$. The number of such symmetrizations of monomials is the number of partitions having maximum part size $2v$ and at most u parts, which is $\binom{2v+u}{u} = m$.

Having shown this, it follows that

$$Q = \sum_{k=1}^{m} (g_k(a_{i_1}, \dots, a_{i_u}))^T_{I\subseteq[2d-1], |I|=u} (h_k(a_{j_1}, \dots, a_{j_v}))_{J\subseteq[2d-1], |J|=v},$$

and so Q has rank at most m. We conclude by Stirling's approximation. ■

Remark 2.3.20. Numerical evidence suggests that equality holds in Theorem 2.3.19 when $u = \lfloor d/2 \rfloor$. This would imply that $\mathbf{R}_S(h_d) = \Omega(2.59^d)$.

2.3.3 $A(n, d)$ in positive characteristic and abelian group algebras

We briefly introduce a generalization of Waring rank to $\mathcal{S}^n_d(\mathsf{k}) := \mathsf{k}[x_1, \dots, x_n]_d$, where k is a field of arbitrary characteristic. This notion has been studied extensively as early as 1916 [Mac94], and directly corresponds to Waring rank in the case that $\mathrm{char}(\mathsf{k}) = 0$. For a thorough algebraic-geometric treatment of this subject, see [IK99]. Assume k is algebraically closed unless stated otherwise.

Definition 2.3.21. For $\ell = \sum_{i=1}^n a_i x_i \in \mathcal{S}^n_1(\mathsf{k})$, let

$$\ell^{[d]} := \sum_{\alpha\in\mathbb{N}^n_d} a_1^{\alpha_1} \cdots a_n^{\alpha_n} x^\alpha \in \mathcal{S}^n_d(\mathsf{k}).$$

Note that $\ell^{[d]}$ is just ℓ^d without any multinomial coefficients. We remark that the projectivization of the set $\{\ell^{[d]} : \ell \in \mathcal{S}^n_1(\mathsf{k})\}$ is the classical *Veronese variety* in algebraic geometry [IK99, Corollary A.10].

Definition 2.3.22. For $f \in \mathcal{S}^n_d(\mathsf{k})$, let $\mathbf{R}^\nu(f)$ be the minimum r such that there exist linear forms $\ell_1, \dots, \ell_r$ with

$$f = \sum_{i=1}^{r} \ell_i^{[d]},$$

and let

$$\mathbf{R}^\nu_{\mathrm{supp}}(f) := \min(\mathbf{R}^\nu(g) : g \in \mathcal{S}^n_d(\mathsf{k}), \mathrm{supp}(g) = \mathrm{supp}(f)).$$

The next proposition shows that the $d = j$ case of Theorem 2.2.1 (a) holds (ignoring a factorial) with the above definition of rank in the case that g is multilinear. Recall that this fact is key for algorithmic upper bounds.

Proposition 2.3.23. *Suppose that $g = \sum_{i=1}^r \ell_i^{[d]} \in \mathcal{S}_d^n$ is multilinear. Then for all $f \in \mathcal{S}_d^n$,*

$$g(\partial \mathbf{x}) f = \sum_{i=1}^r f(\ell_i^*).$$

Proof. Suppose that $g = \sum_\alpha b_\alpha x^\alpha$ and $\ell_i = (\sum_{j=1}^n c_{i,j} x_j)^{[d]}$. Note that $b_\alpha = \sum_{i=1}^r c_{i,1}^{\alpha_1} \cdots c_{i,n}^{\alpha_n}$. If $f = \sum_\alpha a_\alpha x^\alpha$, then since g is multilinear, $g(\partial \mathbf{x}) f = \sum_\alpha a_\alpha b_\alpha$. On the other hand,

$$\sum_{i=1}^r f(c_{i,1}, \ldots, c_{i,n}) = \sum_{i=1}^r \sum_\alpha a_\alpha c_{i,1}^{\alpha_1} \cdots c_{i,n}^{\alpha_n} = \sum_\alpha a_\alpha b_\alpha.$$

■

Definition 2.3.24. Let $A_{\mathsf{k}}(n,d) := \mathbf{R}^\nu_{\text{supp}}(e_{n,d})$.

It is easy to see that if $\mathsf{k} = \mathbb{C}$ and if g is multilinear, $\mathbf{R}_S(g) = \mathbf{R}^\nu(g)$. This implies that $A_{\mathbb{C}}(n,d) = A(n,d)$, and so the above definition really does generalize $A(n,d)$.

Theorem 2.3.25. *For all $n \geq d$, $A_{\mathsf{k}}(n,d) \geq 2^{d-1}$.*

Proof. It follows from an argument similar to that of Proposition 2.3.3 (a) that $A(d,d) \leq A(n,d)$ for all $n \geq d$. As it was shown in [RS11] that $\mathbf{R}^\nu(x_1 \cdots x_d) \geq 2^{d-1}$, the conclusion follows. ■

Definition 2.3.26. Given $A \in \mathsf{k}^{d \times n}$, let

$$g_A := \sum_{\alpha \in \mathbb{N}_d^n} \text{per}_d(A_\alpha) x^\alpha. \tag{2.8}$$

Lemma 2.3.27. *Let k be arbitrary and let $A \in \mathsf{k}^{d \times n}$. Then $\mathbf{R}^\nu(g_A) \leq 2^d - 1$.*

Proof. For $1 \leq i \leq d$, let $L_i := \sum_{j=1}^n A_{ij} y_j \in \mathsf{k}[y_1, \ldots, y_n]$. Now consider

$$\sum_{\alpha \in \mathbb{N}_d^n} L_1^{\alpha_1} \cdots L_n^{\alpha_n} x^\alpha \in \mathsf{k}[y_1, \ldots, y_n][x_1, \ldots, x_n].$$

Note that the coefficient of $y_1 \cdots y_d$ in this polynomial is equal to g_A. It then follows from inclusion-exclusion (or Equation (2.5)) that this coefficient equals

$$\sum_{\alpha \in \{0,1\}^d} (-1)^{|\alpha| + d} (\sum_{i=1}^n x_i \sum_{j=1}^d \alpha_j A_{i,j})^{[d]}. \tag{2.9}$$

■

Theorem 2.3.28. *If k is infinite and $\text{char}(\mathsf{k}) = 2$, $A_{\mathsf{k}}(n,d) \leq 2^d - 1$.*

Proof. Let $A \in \mathsf{k}^{d \times n}$ be a matrix with non-vanishing $d \times d$ minors. Since $\text{char}(\mathsf{k}) = 2$,

$$g_A = \sum_{\alpha \in \mathbb{N}_d^n} \det_d(A_\alpha) x^\alpha.$$

If $\alpha \notin \{0,1\}^n$ then A_α has a repeated column and so $\det(A_\alpha) = 0$. Otherwise $\det(A_\alpha) \neq 0$. Therefore g_A has the desired support. The conclusion follows from Lemma 2.3.27. ■

Theorem 2.3.28 gives the following $2^d \text{poly}(n)$-time algorithm for testing if a polynomial $f \in \mathcal{S}_d^n(\mathsf{k})$ over a large enough field of characteristic 2 is supported on any multilinear monomial. For $U \subseteq \mathsf{k}$, where $|U| \geq 2d$, choose $a = (a_1, \ldots, a_n) \in U^n$ uniformly at random, and take $A \in \mathsf{k}^{d \times n}$ to have nonvanishing $d \times d$ minors. Then compute

$$\sum_{\alpha \in \{0,1\}^d} f(a_1 \sum_{j=1}^d \alpha_j A_{1,j}, \ldots, a_n \sum_{j=1}^d \alpha_j A_{n,j}). \tag{2.10}$$

It follows from Proposition 2.3.23, Theorem 2.3.28, and the Schwarz-Zippel lemma that this quantity is nonzero with probability at least $1/2$ when f is supported on a multilinenar monomial, and zero otherwise. If $f = \sum_\alpha b_\alpha x^\alpha$, this algorithm computes

$$\sum_{\alpha \in \{0,1\}_d^n} b_\alpha a^\alpha \det(A_\alpha).$$

The “option 2” implementation of “decide-multilinear” in [Kou08] is obtained exactly if instead we choose $A \in \mathbb{Z}_2^{d \times n}$ uniformly at random and take $a_1, \ldots, a_n = 1$. Similarly, the algorithm of [Wil09a] is obtained by choosing both $A \in \mathbb{Z}_2^{d \times n}$ and $a_1, \ldots, a_n \in \mathsf{k}$ uniformly at random. Additionally, the algorithm of [Bjö10a] for detecting Hamiltonian cycles reduces to computing Equation (2.10) where $a_1, \ldots, a_n = 1$, $A \in \mathsf{k}^{d \times n}$ is chosen uniformly at random, and the generating polynomial f has the property that $\deg f \approx 3d/4$. This explains the relevance of “determinant sums” to [Bjö10a] and shows that [Wil09a, Kou08] were in fact also computing “determinant sums”. This connection was made earlier in [BDH18].

The algorithms of [Wil09a, Kou08] were presented in terms of a property of abelian group algebras. The following theorem elucidates the connection between support rank and this property.

Theorem 2.3.29. *Let G be an abelian group, and let $y_1, \ldots, y_n \in \mathsf{k}[G]$. For $\alpha \in \mathbb{N}^n$, let $f_\alpha := \prod_{i=1}^n y_i^{\alpha_i}$. Define*

$$T := \{\alpha \in \mathbb{N}_d^n : f_\alpha(\text{Id}_G) \neq 0\}.$$

Then $\mathbf{R}^\nu_{\text{supp}}(\sum_{\alpha \in T} x^\alpha) \leq |G|$.

Proof. Let ρ be the regular representation of G; this extends linearly to a representation of $\mathsf{k}[G]$. Consider the $|G| \times |G|$ matrices $\rho(y_1), \ldots, \rho(y_n)$. Since G is abelian, there exists an invertible matrix A so that $\rho(y_i) = A \Lambda_i A^{-1}$ for all $i \in [n]$ and some diagonal matrices $\Lambda_1, \ldots, \Lambda_n$.

By assumption, we have that for all $\alpha \in \mathbb{N}_d^n$, $f_\alpha(\text{Id}_G) \neq 0$ if and only if $\alpha \in T$. Note that $f_\alpha(\text{Id}_G) \neq 0$ if and only if for some $\lambda \neq 0$ and all $i \in |G|$, $\rho(f_\alpha)_{i,i} = \lambda$. Letting $D \in \mathsf{k}^{|G| \times |G|}$ be a diagonal matrix with nonzero trace, it follows that $\text{tr}(D \cdot \rho(f_\alpha)) \neq 0$ if and only if $\alpha \in T$. Note that

$$\begin{aligned} \text{tr}(D \cdot \rho(f_\alpha)) = \text{tr}(D \cdot \rho(\prod_{i=1}^n y_i^{\alpha_i})) &= \text{tr}(D \cdot \prod_{i=1}^n \rho(y_i)^{\alpha_i}), \\ &= \text{tr}(D \cdot \prod_{i=1}^n (A \Lambda_i A^{-1})^{\alpha_i}), \\ &= \text{tr}(D \cdot \prod_{i=1}^n \Lambda_i^{\alpha_i}). \end{aligned}$$

Let $M_i := D^{1/n}\Lambda_i$. By the above discussion, for all $\alpha \in \mathbb{N}_d^n$, $\mathrm{tr}(\prod_{i=1}^n M_i^{\alpha_i}) \neq 0$ if and only if $\alpha \in T$.

Define the linear forms $\ell_i = \sum_{j=1}^n (M_j)_{i,i} x_i$ for all $i \in |G|$. We now claim that $P := \sum_{i=1}^n \ell_i^{[d]}$ has the desired support. To see this, consider the coefficient of x^α in P, where $|\alpha| = d$. By definition, this is equal to

$$\sum_{i=1}^{|G|} (M_1)_{i,i}^{\alpha_1} \cdots (M_n)_{i,i}^{\alpha_n} = \mathrm{tr}(\prod_{i=1}^n M_i^{\alpha_i}),$$

and hence the claim holds. ∎

Theorem 2.3.29 allows to to recover the approach of [Kou08, Wil09a] from a support-rank perspective. Let $G = \mathbb{Z}_2^d$, and let $v_1, \ldots, v_n \in G$ be chosen independently and random. Then let $y_i := \mathrm{Id}_G + v_i \in \mathsf{k}[G]$ for all i in the statement of Theorem 2.3.29. The key fact used in [Kou08, Wil09a] was that when $\mathrm{char}(\mathsf{k}) = 2$, $f_\alpha(\mathrm{Id}_G) = 0$ whenever $\alpha \notin \{0,1\}_d^n$, and for any $\alpha \in \{0,1\}_d^n$, $f_\alpha(\mathrm{Id}_G) \neq 0$ with probability at least $1/4$. The algorithms of [Kou08, Wil09a] then follow by using the decomposition given by Theorem 2.3.29. Note that this algorithm does not use a decomposition of a multilinear polynomial supported on all multilinear monomials, but rather it samples a multilinear polynomial that is supported on a given multilinear monomial with *constant probability*.

2.3.4 A recursive approach for bounding $A(n,d)$

In this section we provide a recursive method for upper bounding $A^+(n,d)$ and $A^\varepsilon(n,d)$. We will start with a recursive bound on $A^\varepsilon(n,d)$ for varying n and fixed d, and later build upon this to give a recursive bound on $A^+(n,d)$ for all n and d.

A recursive bound on $A^\varepsilon(n,d)$ for fixed d

We will first need the following tool introduced in [AG07].

Definition 2.3.30. For $\delta > 1$, a δ-balanced (n,k,l)-splitter $\mathcal{F}$ is a family of functions from $[n]$ to $[l]$ such that for some real number c, for all $S \subseteq [n]$ where $|S| = k$, the number of functions in $\mathcal{F}$ that are injective on S is between c/δ and $c\delta$.

A δ-balanced (n,k,k)-splitter will be called a δ-balanced (n,k)-perfect hash family. If $\mathcal{F}$ only satisfies the property that for each $S \subseteq [n]$, where $|S| = k$, there exists *some* function in $\mathcal{F}$ that is injective on S, we call $\mathcal{F}$ an (n,k,l)-splitter.

The next fact essentially appears in [AG07]; we reproduce the proof for completeness. Here $(n)_k := n(n-1)\cdots(n-k+1)$ denotes the falling factorial.

Lemma 2.3.31. *For $1 < \delta \leq 2$, there exists a δ-balanced (n,k,l)-splitter of size*

$$O\left(\frac{l^k \cdot k \log n}{(l)_k(\delta-1)^2}\right).$$

Proof. Set $p := \frac{(l)_k}{l^k}$ and $M := \lceil \frac{8(k\log n+1)}{p(\delta-1)^2} \rceil$. Choose M independent random functions from $[n]$ to $[l]$. For any $S \subseteq [n]$ of size k, the expected number of functions that are injective on S is pM.

By the Chernoff bound, the probability that the number of functions that are injective on S is less than pM/δ or greater than $pM\delta$ is at most $2e^{-(\delta-1)^2pM/8}$. Then by a union bound the expected number of such sets for which the number of 1-1 functions is not as desired is at most

$$\binom{n}{k}2e^{-(\delta-1)^2pM/8} \leq \binom{n}{k}2e^{-(k\log n+1)} < 1. \qquad \blacksquare$$

Theorem 2.3.32. *Suppose $f \in \mathfrak{E}^{\varepsilon_0}(n_0, d)$ where $0 < \varepsilon_0 < 1$. Then for all $\varepsilon_0 < \varepsilon < 1$ and all $n \geq d$,*

$$A^\varepsilon(n, d) \leq O\left(\frac{\mathbf{R}_S(f) \cdot n_0^d \cdot d\log n}{(n_0)_d(\delta-1)^2}\right),$$

where $\delta := \min(\frac{1+\varepsilon}{1+\varepsilon_0}, \frac{1-\varepsilon_0}{1-\varepsilon})$.

Proof. If $n \leq n_0$ the theorem follows from Proposition 2.3.3 (a). Hence we will assume that $n > n_0$.

Let $\mathcal{F} = \{\pi_i : i \in [M]\}$ be a δ-balanced (n, d, n_0)-splitter of minimal size M. For all $(i, j) \in [M] \times [n_0]$, define the linear forms $L_{i,j} = \sum_{k\in\pi_i^{-1}(j)} x_k$. Now we claim that for some constant c,

$$f' := \frac{1}{c}\sum_{i=1}^{M} f(L_{i,1}, L_{i,2}, \ldots, L_{i,n_0}) \in A^\varepsilon(n, d).$$

First notice that since f is multilinear and $L_{i,1}, \ldots, L_{i,n_0}$ are linear forms with disjoint supports for all i, f' is also multilinear. Next, by virtue of the fact that $f \in \mathfrak{E}^{\varepsilon_0}(n_0, d)$, the coefficient of any multilinear monomial x^α in $f(L_{i,1}, L_{i,2}, \ldots, L_{i,n_0})$ is in the range $[1-\varepsilon_0, 1+\varepsilon_0]$ if and only if π_i is injective on $\mathrm{supp}(\alpha)$. Then because $\mathcal{F}$ is a δ-balanced splitter, there are between c/δ and $c\delta$ such contributions to the coefficient of x^α in the above sum, for some fixed real number c. But this implies that the coefficient of x^α in f' is between $(1-\varepsilon_0)/\delta$ and $(1+\varepsilon_0)\delta$, which by our choice of δ implies that $f \in \mathfrak{E}^\varepsilon(n, d)$. By subadditivity of rank, $\mathbf{R}_S(f') \leq M \cdot \mathbf{R}_S(f)$, and the theorem follows by the bound on M given by Lemma 2.3.31. ■

Remark 2.3.33. As Waring rank can be strictly subadditive, it is possible that the final step of the above lemma is far from optimal; see also Remark 2.3.38.

Theorem 2.3.34. *For all $0 < \varepsilon < 1$, $A^\varepsilon(n, d) \leq O(4.075^d\varepsilon^{-2}\log n)$.*

Proof. Let $c \geq 1$ be a constant to be determined later. Taking $n_0 = \lceil cd \rceil$, $f = e_{n_0,d}$, $\varepsilon_0 = \varepsilon/2$ in Theorem 2.3.32,

$$A^\varepsilon(n, d) \leq O\left(\frac{\mathbf{R}_S(e_{n_0,d}) \cdot n_0^d \cdot d\log n}{(n_0)_d(\delta-1)^2}\right)$$

where $\delta = \min(\frac{1+\varepsilon}{1+\varepsilon/2}, \frac{1-\varepsilon/2}{1-\varepsilon}) = \frac{1+\varepsilon}{1+\varepsilon/2} \geq \varepsilon/3 + 1$. Combining this with the upper bound on $\mathbf{R}_S(e_{n_0,d})$ given in [Lee16],

$$A^\varepsilon(n, d) \leq O\left(\left(\sqrt{2e} \cdot c\left(\frac{c-1}{e}\right)^{c-1}\left(\frac{e}{c-1/2}\right)^{c-1/2}\right)^d \varepsilon^{-2}d^2\log n\right).$$

Using a computer we found that this is minimized when $c \approx 1.55$, in which case we obtain an upper bound of $O(4.075^d\varepsilon^{-2}\log n)$. ■

Remark 2.3.35. If we take $f = x_1 x_2 \cdots x_d$ and use the upper bound on $\mathbf{R}_S(x_1 \cdots x_d)$ given by Equation (2.5), it follows from Theorem 2.3.32 that

$$A^{\varepsilon}(n,d) \leq (2^d - 1)\frac{d^d}{d!}\varepsilon^{-2} = O((2e)^d \varepsilon^{-2}) = O(5.44^d \cdot \varepsilon^{-2}).$$

The decomposition implicit in the above bound is as follows. Let $\mathcal{F}$ be an $(1+\varepsilon)$-balanced (n,d)-perfect hash family. For $\pi \in \mathcal{F}$ and $i \in [d]$, let $L_{\pi,i} \coloneqq \sum_{j \in \pi^{-1}(i)} x_j$. Then for some $c > 0$,

$$\frac{1}{c}\sum_{\pi \in \mathcal{F}} \sum_{\alpha \in \{0,1\}^d} (-1)^{|\alpha|+d} \left(\sum_{i=1}^{d} \alpha_i L_{\pi,i}\right)^d \in \mathfrak{E}^{\varepsilon}(n,d).$$

Applying this to the cycle-generating polynomial Section 2.5, one finds that a $(1 \pm \varepsilon)$-approximation of the number of length-d cycles in the graph G is given by

$$\frac{1}{c \cdot d!}\sum_{\pi \in \mathcal{F}} \sum_{\alpha \in \{0,1\}^d} (-1)^{|\alpha|+d} f_G(\alpha_{\pi(1)}, \ldots, \alpha_{\pi(n)}).$$

This is equivalent to the color-coding algorithm for counting cycles described in [AG09], except we use inclusion-exclusion instead of dynamic programming to count the number of colorful simple cycles for a given coloring. Similarly, by replacing $\mathcal{F}$ with an (n,d)-perfect hash family one obtains an algorithm for detecting simple cycles that parallels the one given in [AYZ95]. We note that using inclusion-exclusion rather than dynamic programming reduces the space complexity of the counting step from exponential to polynomial.

Furthermore, this bound is naturally derived by an application of color-coding. Using each function in a $(1+\varepsilon)$-balanced (n,d)-perfect hash family we color the variables $x_1, \ldots, x_n$ using d colors. To each color we associate the linear form equal to the sum of the variables of that color. Since these linear forms have disjoint support, their product is multilinear. Summing the resulting products of linear forms for each function in the family, any given multilinear monomial appears with coefficient between $c/(1+\varepsilon)$ and $c(1+\varepsilon)$. The resulting polynomial is a sum of products of $|\mathcal{F}|$ linear forms, which can be written as a sum of powers of $O(|\mathcal{F}|2^d)$ linear forms using Equation (2.5).

An improvement to color-coding was made in [HWZ08] based on the idea of using $n_0 \coloneqq \lceil 1.3d \rceil$ colors rather than d. We recover this result as follows. By applying Theorem 2.3.32 with $f = e_{n_0,d}$ and using the suboptimal bound on $\mathbf{R}_S(e_{n_0,d})$ given by Equation (2.5),

$$A^{+}(n,d) \leq O\left(\binom{1.3d}{d}\frac{(1.3d)^d}{(1.3d)_d} d \log n\right) = O(4.32^d \log n).$$

In fact, the choice of $n_0 = \lceil 1.3d \rceil$ is optimal if we are using the rank bound of Equation (2.5); this follows from the same calculation done in [GRWZ18a, Section 8]. The algorithm resulting from this bound was virtually described in [GRWZ18a, AFS09].

A recursive bound on $A^+(n,d)$ for all n and d

Definition 2.3.36. For $g \in \mathcal{S}_d^n$ and $s,t \in \mathbb{N}$, let

$$g^{\circledast(s,t)} := \sum_{i=1}^{s} \prod_{j=1}^{t} g(x_{i,j,1}, x_{i,j,2}, \ldots, x_{i,j,n}) \in \mathbb{C}[x_{i,j,k} : (i,j,k) \in [s] \times [t] \times [n]].$$

In words, $g^{\circledast(s,t)}$ is obtained from g by taking the t-fold product of g with itself using disjoint sets of variables, and then taking the s-fold sum of the resulting polynomial using disjoint sets of variables.

Lemma 2.3.37. *For all $g \in \mathcal{S}_d^n$, $\mathbf{R}_S(g^{\circledast(s,t)}) \le s((d+1)\mathbf{R}_S(g))^t$.*

Proof. By subadditivity of Waring rank, $\mathbf{R}_S(g^{\circledast(s,t)}) \le s\mathbf{R}_S(g^{\circledast(1,t)})$. Now letting $r = \mathbf{R}_S(g)$, there exist linear forms $\ell_{i,j} \in \mathbb{C}[x_{1,i,1}, \ldots, x_{1,i,n}]$ for $(i,j) \in [t] \times [r]$ so that

$$g^{\circledast(1,t)} = \prod_{i=1}^{t} \sum_{j=1}^{r} \ell_{i,j}^d = \sum_{v \in [r]^t} \prod_{i=1}^{t} \ell_{i,v_i}^d.$$

Using the fact that $\mathbf{R}_S(\prod_{i=1}^t x_i^d) \le (d+1)^t$ (which follows from e.g. Equation (2.5)[7]), it follows that $\mathbf{R}_S(g^{\circledast(s,t)}) \le s\mathbf{R}_S(g^{\circledast(1,t)}) \le s((d+1)\mathbf{R}_S(g))^t$. ■

Remark 2.3.38. The first step of the above lemma is to apply subadditivity of Waring rank to polynomials in disjoint sets of variables. *Strassen's direct sum conjecture* claims that rank is actually additive in this case; see [CCC15] for more. It was recently shown in [Shi17] that the tensor version of this conjecture is false; if the polynomial version is also false, the upper bound of Lemma 2.3.37 may not be optimal.

Definition 2.3.39. An (n,d,n_0,d_0)-perfect splitter, where $n \ge d$, $n_0 \ge d_0$, and $d_0 \mid d$, is a family of functions $\mathcal{F} = \{\pi : [n] \to [d/d_0] \times [n_0]\}$ such that for all $S \subseteq [n]$ where $|S| = d$, there exists a $\pi \in \mathcal{F}$ such that for all $i \in [d/d_0]$, $\pi(S)$ contains d_0 elements whose first coordinate is i, and any two elements in $\pi(S)$ with the same first coordinate have differing second coordinates.

In other words, we want the elements of $\pi(S)$ to be "split evenly" by their first coordinate, and those elements with the same first coordinate should have different second coordinates. As special cases, an (n,d,d,d)-perfect splitter is a (n,d)-perfect hash family, and when $n_0 \ge n$, an (n,d,n_0,d_0)-perfect splitter is a (n,d,d_0)-splitter.

Definition 2.3.40. For $n \ge d$, $n_0 \ge d_0$, and $d_0 \mid d$, let

$$\sigma(n,d,n_0,d_0) := \left\lceil \left(\frac{n_0^{d_0}}{(n_0)_{d_0}} \right)^{d/d_0} \frac{d_0!^{d/d_0} (d/d_0)^d}{d!} d \log n \right\rceil.$$

Proposition 2.3.41. *There exists an (n,d,n_0,d_0)-perfect splitter of size $\sigma(n,d,n_0,d_0)$.*

[7] The slightly better bound of $(d+1)^{t-1}$ given in [RS11] can be used here.

Proof. We will consider the probability that a random function π has the desired effect on a fixed subset $S \subseteq [n]$, where $|S| = d$. The conclusion will then follow from a union bound.

Let $\pi : [n] \to [d/d_0] \times [n_0]$ be chosen uniformly at random. The probability that each integer in $[d/d_0]$ appears equally often as the first coordinate in $\pi(S)$ equals

$$p_1 := \frac{d!}{d_0!^{d/d_0}(d/d_0)^d}.$$

Assuming this happens, the probability that all elements in $\pi(S)$ with a given first coordinate are assigned different second coordinates equals

$$p_2 := \frac{(n_0)_{d_0}}{n_0^{d_0}},$$

and so with probability p_2^{d/d_0} this happens for all d/d_0 choices of the first coordinate. Hence if we generate $c = \lceil (p_1 p_2^{d/d_0})^{-1} \rceil$ independent and uniformly random functions, some function has the desired effect on S with probability at least $1 - e^{-1}$. Therefore if we generate $\lceil cd \log n \rceil$ random functions, the expected number of subsets for which no function has the desired effect on equals

$$\binom{n}{d} e^{-\lceil d \log n \rceil} < 1. \qquad \blacksquare$$

Theorem 2.3.42. *Let $f \in \mathfrak{E}^+(n_0, d_0)$. Then for all integers n, d where $n \geq d$,*

$$A^+(n, d) \leq s((d_0 + 1)\mathbf{R}_S(f))^{\lceil d/d_0 \rceil},$$

where

$$s = \sigma(n + \lceil d/d_0 \rceil d_0 - d, \lceil d/d_0 \rceil d_0, n_0, d_0).$$

Proof. We start with the case that $d = t \cdot d_0$ for some $t \in \mathbb{N}$. Let $\mathcal{F} = \{\pi_i : i \in [s]\}$ be an (n, d, n_0, d_0)-perfect splitter of minimal size. For $(i, j, k) \in [s] \times [t] \times [n_0]$, let $L_{i,j,k} := \sum_{m \in \pi_i^{-1}(j,k)} x_m$. We now claim that $g^{\circledast(s,t)}(L_{i,j,k}) \in \mathfrak{E}^+(n, d)$. To see this, first note that for any i, the linear forms $\{L_{i,j,k} : (j, k) \in [t] \times [n_0]\}$ have disjoint support. Since f is multilinear, it follows that

$$f_i := f(L_{i,1,1}, \ldots, L_{i,1,n_0}) \cdots f(L_{i,t,1}, \ldots, L_{i,t,n_0})$$

is multilinear for all i, and therefore so is $f^{\circledast(s,t)}(L_{i,j,k})$.

Now consider the coefficient of some degree-d multilinear monomial x^α in f_i. Since f has nonnegative coefficients, this will be nonnegative. Furthermore, if π_i splits the set $\mathrm{supp}(\alpha)$ evenly by first coordinate and all elements in $\pi_i(\mathrm{supp}(\alpha))$ with the same first coordinates have different coordinates, this coefficient will be strictly positive by definition of the linear forms $L_{i,j,k}$. Since $\mathcal{F}$ is a perfect splitter, each degree-d multilinear monomial will then appear with a positive coefficient. Therefore by Proposition 2.3.41,

$$A^+(n, d) \leq \mathbf{R}_S(f^{\circledast(s,t)}) \leq s((d_0 + 1)\mathbf{R}_S(f))^{d/d_0}.$$

Now suppose that $d_0 \nmid d$. By Proposition 2.3.3 (b), we have that

$$A^+(n, d) \leq A^+(n + \lceil d/d_0 \rceil d_0 - d, \lceil d/d_0 \rceil d_0),$$

which is at most $s((d_0 + 1)\mathbf{R}_S(f))^{\lceil d/d_0 \rceil}$ by a reduction to the case when $d_0 \mid d$. $\blacksquare$

Note that by taking $d_0 = d$ in the above theorem, we find that

$$A^+(n,d) \le O\left(\frac{A^+(n_0,d)\cdot n_0^d \cdot d\log n}{(n_0)_d}\right),$$

recovering Theorem 2.3.32 in the case of nonnegative support rank.

Example 2.3.43. Theorem 2.3.42 suggests bounding $A^+(\mathbb{N},d)$ for small values of d as an approach to improve the upper bounds of this section. For example, suppose that $A^+(\mathbb{N},4) \le 10$. Then we have that for all $n_0 \ge 4$ and all n, d,

$$\begin{aligned} A(n,d) &\le \sigma(n+4\lceil d/4\rceil - d, \lceil d/4\rceil 4, n_0, 4)5^{\lceil d/4\rceil -1}10^{\lceil d/4\rceil} \\ &= O\left(\left(\frac{n_0^4}{n_{0(4)})}\right)^{d/4} \frac{4!^{d/4}(d/4)^d}{d!}\log\binom{n}{d}50^{d/4}\right) \\ &= O\left(\left(\frac{n_0^4}{n_{0(4)})}\right)^{d/4} (e\cdot 1200^{1/4}/4)^d d\log n\right) \\ &= O\left(\left(\frac{n_0^4}{n_{0(4)})}\right)^{d/4} 3.9998^d d\log n\right). \end{aligned}$$

Taking $n_0 \ge 33700$, we conclude that $A(n,d) \le O(3.9999^d \log n)$.

In contrast, the best upper bound we know on $A^+(\mathbb{N},4)$ is 79, which follows from Remark 2.3.18. When used in Theorem 2.3.42 this only shows that $A^+(n,d) \le O(6.706^d \log n)$.

2.4 Applications

We now recall and prove Theorem 2.1.7.

Theorem 2.4.1. *Let $f \in \mathbb{R}_{\ge 0}[x_1,\dots,x_n]_d$ be given as a black-box. There is a randomized algorithm which given any $0 < \varepsilon < 1$ computes a number z such that with probability 2/3,*

$$(1-\varepsilon)\cdot e_{n,d}(\partial \mathbf{x})f < z < (1+\varepsilon)\cdot e_{n,d}(\partial\mathbf{x})f.$$

This algorithm runs in time $4.075^d \cdot \varepsilon^{-2}\log(\varepsilon^{-1})\cdot \operatorname{poly}(n,s_f)$ and uses $\operatorname{poly}(n,s_f,\log(\varepsilon^{-1}))$ space. Here s_f is the maximum bit complexity of f on the domain $\{\pm 1\}^n$.

Proof. Set $n_0 := \lceil 1.55d\rceil$, $p := (n_0)_d/n_0^d$, and $M := \lceil 3\varepsilon^{-2}/p\rceil$. Let $\mathcal{F}$ be a family of M independent and uniformly random functions from $[n]$ to $[n_0]$. For $\pi \in \mathcal{F}$ and $i \in [n_0]$, define the linear form $L_{\pi,i} := \sum_{j\in\pi^{-1}(i)} x_j$. The algorithm will compute and return

$$\frac{1}{pM}\sum_{\pi\in\mathcal{F}} e_{n_0,d}(L_{\pi,1}(\partial\mathbf{x}),\dots,L_{\pi,n_0}(\partial\mathbf{x}))f.$$

By Theorem 2.2.1 (a) and the upper bound on $e_{n_0,d}$ given in [Lee16], for d odd this equals

$$\frac{1}{pM\cdot 2^{d-1}}\sum_{\pi\in\mathcal{F}}\sum_{\substack{S\subseteq[n_0]\\ |S|\le\lfloor d/2\rfloor}} (-1)^{|S|}\binom{n_0-\lfloor d/2\rfloor - |S| - 1}{\lfloor d/2\rfloor - |S|} f(\delta_{S,\pi(1)},\dots,\delta_{S,\pi(n)}),$$

and for d even equals

$$\frac{1}{pM \cdot 2^{d-1}(n_0 - d)} \sum_{\pi \in \mathcal{F}} \sum_{\substack{S \subseteq [n_0] \\ |S| \leq d/2}} (-1)^{|S|} \binom{n_0 - d/2 - |S| - 1}{d/2 - |S|}$$

$$\cdot (n_0 - 2|S|) f(\delta_{S,\pi(1)}, \ldots, \delta_{S,\pi(n)}),$$

where $\delta_{S,i} := -1$ if $i \in S$ and $\delta_{S,i} := 1$ otherwise. Hence this quantity can be computed using

$$M \sum_{i=0}^{\lfloor d/2 \rfloor} \binom{n_0}{i} \leq O\left(d \frac{\lceil 1.55d \rceil^d}{(\lceil 1.55d \rceil)_d} \binom{\lceil 1.55d \rceil}{\lfloor d/2 \rfloor} \varepsilon^{-2} \right) \leq O(4.075^d \varepsilon^{-2})$$

queries to f on $\{\pm 1\}^n$. The stated time and space bounds then follow from the straightforward evaluation of the above formulas.

We now prove that this quantity gives the desired approximation of $e_{n,d}(\partial \mathbf{x}) f$. Write $f = \sum_{\alpha \in \mathbb{N}^n} a_\alpha x^\alpha$ and fix some $\pi \in \mathcal{F}$. Let

$$e_{n_0,d}(L_{\pi,1}, \ldots, L_{\pi,n_0}) = \sum_{\alpha \in \{0,1\}^n_d} b_\alpha x^\alpha$$

and

$$Y_\pi := e_{n_0,d}(L_{\pi,1}(\partial \mathbf{x}), \ldots, L_{\pi,n_0}(\partial \mathbf{x})) f = \sum_{\alpha \in \{0,1\}^n} a_\alpha b_\alpha.$$

First observe that for any fixed $\alpha \in \{0,1\}^n_d$, $b_\alpha = 1$ with probability p and $b_\alpha = 0$ with probability $1 - p$. By linearity of expectation, it follows that $\mathbb{E}[Y_\pi] = p \cdot e_{n,d}(\partial \mathbf{x}) f$. Moreover,

$$\begin{aligned} \mathrm{Var}[Y_\pi] &= \sum_\alpha \mathrm{Var}[a_\alpha b_\alpha] + \sum_{\beta \neq \alpha} \mathrm{Cov}[a_\alpha b_\alpha, a_\beta b_\beta] \\ &= \sum_\alpha a_\alpha^2 \mathrm{Var}[b_\alpha] + \sum_{\beta \neq \alpha} a_\alpha a_\beta \mathrm{Cov}[b_\alpha, b_\beta]. \end{aligned}$$

As the probability that $b_\alpha = b_\beta = 1$ is at most p for all α, β, we have that

$$\mathrm{Cov}[b_\alpha, b_\beta] = \mathbb{E}[b_\alpha b_\beta] - \mathbb{E}[b_\alpha]\mathbb{E}[b_\beta] \leq p,$$

and hence $\mathrm{Var}[Y_\pi] \leq p(e_{n,d}(\partial \mathbf{x}) f)^2$.

Now let $Z := \frac{1}{M} \sum_{\pi \in \mathcal{F}} Y_\pi$. Then $\mathbb{E}[Z] = p \cdot e_{n,d}(\partial \mathbf{x}) f$ and

$$\mathbf{Var}[Z] = \mathbf{Var}[Y_\pi]/M \leq p \cdot (e_{n,d}(\partial \mathbf{x}) f)^2 / M.$$

By Chebychev's inequality, the probability that Z is smaller or bigger than its expectation by $\varepsilon \cdot p \cdot e_{n,d}(\partial \mathbf{x}) f$ is at most ε^{-2}/pM, which by our choice of M is at most $1/3$. Dividing by p we obtained the desired approximation. ■

Remark 2.4.2. In order to derandomize Theorem 2.1.7, it would suffice to give a near-optimal construction of a $(1+\varepsilon)$-balanced $(n, d, 1.55d)$-splitter, as first defined in [AG07]. We note that such a construction was given for ("unbalanced") $(n, k, \alpha k)$-splitters for all $\alpha \geq 1$ in [GRWZ18a]. Furthermore, note that for any *fixed* values of n and d, Theorem 2.1.7 can be made deterministic by taking $\mathcal{F}$ to be a $(1+\varepsilon)$-balanced $(n, d, 1.55d)$-splitter of optimal size.

2.4.1 Approximately counting subgraphs of bounded treewidth

We now give an application of Theorem 2.1.7. First, we recall the notion of treewidth:

Definition 2.4.3. A *tree decomposition* of a graph $G = (V, E)$ is given by a tree T with *nodes* $X_1, \ldots, X_n$, where $X_i \subseteq V$, with the following properties:

1. Each vertex in G is contained in at least one node in T.
2. If X_i and X_j both contain a vertex v, then all nodes in T on the path from X_i and X_j contain v.
3. If $(u, v) \in E$, then there is a node in T containing both u and v.

The *width* of a tree decomposition is the size of the largest node in T minus one. The *treewidth* of g, denoted $\mathrm{tw}(G)$, is the minimum width among all tree decompositions of G.

Definition 2.4.4. For graphs G, H, where $|G| = n$ and $|H| = d$, let

$$P_{H,G}(x_1, \ldots, x_n) \coloneqq \sum_{\Phi \in \mathrm{Hom}(H,G)} \prod_{v \in V(H)} x_{\Phi(v)} \in \mathcal{S}_d^n.$$

The key fact is that $P_{H,G}$ can be computed by a small arithmetic circuit in the case when H has small treewidth. For this we use the following lemma, proven in [BDH18, FLR$^+$12].

Lemma 2.4.5. *[BDH18, Lemma 16] Let G and H be graphs where $|G| = n$ and $|H| = d$. Then there is an arithmetic formula C of size $O(d \cdot n^{\mathrm{tw}(H)+1})$ computing $P_{H,G}$. Furthermore, this formula can be constructed in time $O(1.76^d) + |C| \cdot \mathrm{polylog}(|C|)$.*

Theorem 2.4.6. *Let G and H be graphs where $|G| = n$, $|H| = d$, and H has treewidth $\mathrm{tw}(H)$. There is a randomized algorithm which given any $0 < \varepsilon < 1$ computes a number z such that with probability $2/3$,*

$$(1 - \varepsilon) \cdot \mathrm{Sub}(H, G) < z < (1 + \varepsilon) \cdot \mathrm{Sub}(H, G).$$

This algorithm runs in time $4.075^d \cdot n^{\mathrm{tw}(H)+O(1)} \cdot \varepsilon^{-2} \log(\varepsilon^{-1})$. Here $\mathrm{Sub}(H, G)$ denotes the number of subgraphs of G isomorphic to H.

Proof. We first construct a formula C computing $P_{H,G}$ using Lemma 2.4.5. Note that C can be evaluated on inputs in $\{\pm 1\}^n$ in time $O(n^{\mathrm{tw}(H)+1})$, and the maximum bit-complexity of $P_{H,G}$ on $\{\pm 1\}^n$ is $\log f(1, 1, \ldots, 1) = \log(|\mathrm{Hom}(H, G)|) \leq d \log n$.

Next note that $e_{n,d}(\partial \mathbf{x}) P_{H,G}$ equals the number of injective homomorphisms from H to G. Using Theorem 2.1.7 and the formula C we first compute a $(1 \pm \varepsilon)$ approximation to this number in time $4.075^d n^{\mathrm{tw}(H)+O(1)} \varepsilon^{-2} \log \varepsilon^{-1}$. In order to obtain a $(1 \pm \varepsilon)$ approximation to $\mathrm{Sub}(H, G)$ we divide this by $|\mathrm{Aut}(H, H)|$, which can be computed exactly in $O(1.01^d)$ time by using a $\mathrm{poly}(d)$-time reduction to graph isomorphism [Mat79] and the quasi-polynomial time graph isomorphism algorithm of [Bab16].

The total time taken is

$$O(1.76^d) + |C| \cdot \mathrm{polylog}(|C|) + 4.075^d \cdot n^{\mathrm{tw}(H)+O(1)} \cdot \varepsilon^{-2} \mathrm{polylog}(\varepsilon^{-1}) + O(1.01^d),$$

$$\leq 4.075^d \cdot n^{\mathrm{tw}(H)+O(1)} \cdot \varepsilon^{-2} \mathrm{polylog}(\varepsilon^{-1}).$$

■

2.4.2 Lower bounds on perfectly balanced hash families

In this section we show how the bounds on $\mathbf{R}_S(e_{n,d})$ given in [Lee16] imply lower bounds on the size of perfectly balanced hash families.

Definition 2.4.7. [AG09, Definition 1] Let $n > l \geq k > 0$. A family of functions $\mathcal{F} = \{\pi : [n] \to [l]\}$ is said to be a perfectly-k balanced hash family if for some $c \in \mathbb{N}$ and all $S \subseteq [n]$ of size k, there are c functions in $\mathcal{F}$ that are injective on S.

Theorem 2.4.8. *Let $\mathcal{F}$ be a perfectly-k balanced hash family from $[n]$ to $[l]$. Then*

a. If k is odd,

$$|\mathcal{F}| \geq \frac{\sum_{i=0}^{\lfloor k/2 \rfloor} \binom{n}{i}}{\sum_{i=0}^{\lfloor k/2 \rfloor} \binom{l}{i}}.$$

b. If k is even,

$$|\mathcal{F}| \geq \frac{\left(\sum_{i=0}^{k/2} \binom{n}{i}\right) - \binom{n-1}{k/2}}{\sum_{i=0}^{k/2} \binom{l}{i}}.$$

Proof. Suppose that k is odd, and let $\mathcal{F}$ be a perfectly k balanced hash family from $[n]$ to $[l]$. For each $\pi \in \mathcal{F}$ define the linear forms $L_{\pi,i} := \sum_{j \in \pi^{-1}(i)} x_j$. Consider the polynomial

$$f := \sum_{\pi \in \mathcal{F}} e_{k,l}(L_{\pi,1}, \ldots, L_{\pi,l}).$$

Since $\mathcal{F}$ is a perfectly balanced hash family it follows that, up to scaling, $f = e_{n,k}$, and hence $\mathbf{R}_S(f) = \sum_{i=0}^{\lfloor k/2 \rfloor} \binom{n}{i}$. On the other hand, by subadditivity of rank, we have that $\mathbf{R}_S(f) \leq |\mathcal{F}| \mathbf{R}_S(e_{k,l}) = |\mathcal{F}| \sum_{i=0}^{\lfloor k/2 \rfloor} \binom{l}{i}$. Hence

$$|\mathcal{F}| \geq \frac{\sum_{i=0}^{\lfloor k/2 \rfloor} \binom{n}{i}}{\sum_{i=0}^{\lfloor k/2 \rfloor} \binom{l}{i}}.$$

The case for k even is shown similarly. ∎

2.4.3 A lower bound on maximum support rank

In the previous section we saw that the support rank of $e_{n,d}$ is significantly lower than its rank. Could it be that this holds for all polynomials? Note that it is true for almost all polynomials: a random element of S^n_d has full support with probability 1, and hence has support rank at most $\mathbf{R}_S((\sum x_i)^d) = 1$. There are polynomials with support rank at least n, for example $\sum x_i^d$. Are there any polynomials whose support rank is much larger than this? Can the maximum support rank it be as high as the maximum Waring rank of roughly n^d (Remark 2.3.18)?[8]

In this section we will show that this is the case. This is a consequence of the following result of Rónyai, Babai, and Ganapathy. A *zero–nonzero pattern* of a function $f : \mathbb{F}^n \to \mathbb{F}^m$ and an input $x \in \mathbb{F}^n$ is defined as the support of $f(x)$. By the number of zero–nonzero patterns of f we mean the number of distinct zero–nonzero patterns over all $x \in \mathbb{F}^n$.

[8] We thank J.M. Landsberg for bringing this question to our attention.

Theorem 2.4.9. *[RBG01, Theorem 1.1] Let $f = (f_1, \dots, f_m)$ be a sequence of polynomials in n variables over an arbitrary field $\mathbb{F}$, with $deg(f_i) = d_i$. Then the number of zero–patterns of f is at most $\binom{n+\sum_{i=1}^m d_i}{n}$.*

Corollary 2.4.10. *For sufficiently large d, there exists $g \in S_d^n$ such that $\mathbf{R}_{supp}(g) \geq \frac{1}{nd}\binom{n+d-1}{d}$.*

Proof. Consider the polynomial map f whose image is the set of all polynomials of Waring rank at most r. This gives a sequence of $m = \binom{n+d-1}{d}$ polynomials in nr variables of degree d. Hence the total number of zero patterns of f is at most $\binom{nr+md}{nr}$. On the other hand, the total number of supports in S_d^n is $2^{\binom{n+d}{d}}$. So if r is the maximum support rank, then

$$2^m \leq \binom{nr+md}{nr} \leq (1 + md/nr)^{nr} e^{nr}$$

so

$$2^{m/nr} \leq e(1 + md/nr)$$

This is not satisfied when $m/nr \geq d$ for all $d \geq 8$. Hence for large enough d must have $r \geq m/nd$ as claimed. ■

2.5 Faster white-box algorithms

In the last section, we saw that $A^+(n,d) \leq O(4.075^d \cdot \log n)$. As a consequence, we gave a $4.075^d \operatorname{poly}(n)$-time *deterministic* algorithm for detecting simple cycles of length d in an n vertex graph. In this section we show how this runtime can be improved to $\varphi^{2d} \operatorname{poly}(n)$ where $\varphi \approx 1.61$ is the golden ratio. This improvement comes from assuming access to a circuit computing the generating polynomial

$$\operatorname{tr}(A_G^d) = \sum_{\substack{\text{closed walks}\\(v_{i_1}, v_{i_2}, \dots, v_{i_d}) \in G}} x_{i_1} \cdots x_{i_d} \in \mathcal{S}_d^n.$$

Similar to before, the starting point is the following. Let $A \in \mathbb{R}^{d \times n}$ be a matrix any d columns of which are linearly independent. Let $X = A \cdot \text{diag}(x_1, \dots, x_n) \cdot A^T$. By the Cauchy-Binet Theorem,

$$\det X = \sum_{S \in \binom{[n]}{d}} \det(A_S)^2 \prod_{i \in S} x_i.$$

(Here A_S refers to the $d \times d$ submatrix of A with columns indexed by the set S.) Since any d columns in A are linearly independent, $\det(A_S)^2 > 0$ for all $S \in \binom{[n]}{d}$. Then note that the result of differentiating $\operatorname{tr}(A_G^d)$ by $\det(A_S)^2 \prod_{i \in S} x_i$ is positive if there is a simple cycle on the vertices $\{v_i : i \in S\}$, and zero otherwise. It follows that $\langle \det X, \operatorname{tr}(A_G^d) \rangle > 0$ if and only if G contains a simple cycle of length d. Here we use the notation

$$\langle f, g \rangle := f(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n})g. \tag{2.11}$$

Motivated by this example, we consider the algorithmic task of computing the inner product $\langle f, g\rangle$ in the special case where f is the determinant of a symbolic matrix (a matrix whose entries are homogeneous linear polynomials) and g is given as an arithmetic circuit. We will start by giving a simple algorithm (Theorem 2.5.9) for the special case when g is computed by a *skew* circuit, meaning one of the two operands to each multiplication gate in the circuit is a variable or a scalar. In Theorem 2.5.13 we show how the runtime of this algorithm can be significantly improved under the additional assumption that X is a Hankel matrix. We do this by making use of identities in the space of minors of a Hankel matrix originally studied in commutative algebra [Con98]. Aside from that, our Theorem 2.5.9 and Theorem 2.5.13 only make use of elementary linear algebra. Theorem 2.5.9 and Theorem 2.5.13 lead mechanically to new and improved algorithms for several well-studied problems.

As an application of Theorem 2.5.13, we give in Corollary 2.5.14 a deterministic $\varphi^{2d}\operatorname{poly}(n) < 2.62^d\operatorname{poly}(n)$-time algorithm for the aforementioned example of detecting simple cycles of length d in a graph, where $\varphi := \frac{1+\sqrt{5}}{2}$ is the golden ratio. This brushes up against the fastest-known deterministic algorithm for this problem which has runtime $2.55^d\operatorname{poly}(n)$ [Tsu19], and unexpectedly matches the runtime of a previous algorithm [FLPS16] while using a new (significantly condensed) approach. Whereas recent algorithms for this problem have relied on explicit constructions of pseudorandom objects such as perfect hash families, universal sets, and representative sets, our algorithm exploits algebraic–combinatorial identities.

Algorithmic results

In contrast to the black-box setting of the last section, we now consider the white-box setting where g is specified by an arithmetic circuit C. We first prove the following:

Theorem 2.5.1. *Let C be a skew arithmetic circuit computing $g \in \mathcal{S}_d^n$, and let $X = (\ell_{i,j})_{i,j\in[d]}$ be a symbolic matrix with entries in $\mathcal{S}_1^n$. Then we can compute $\langle \det X, g\rangle$ with $4^d|C|\operatorname{poly}(d)$ arithmetic operations.*

Theorem 2.5.9 yields faster algorithms for the *k-matroid intersection* and *matroid k-parity* problems. These are the following problems:

Problem 2.5.2 (Matroid k-Parity)**.** Suppose we are given a matrix $B \in \mathbb{Q}^{km\times kn}$ representing a matroid M with groundset $[kn]$, and a partition π of $[kn]$ into parts of size k. Decide if the union of any m parts in π are independent in M.

Problem 2.5.3 (k-Matroid Intersection)**.** Suppose we are given matrices $B_1, \ldots, B_k \in \mathbb{Q}^{m\times n}$ representing matroids $M_1, \ldots, M_k$ with the common groundset $[n]$. Decide if $M_1, \ldots, M_k$ share a common base.

As a simple application of Theorem 2.5.9, we showed in [BP21] that these can be solved in time $4^{km}\operatorname{poly}(N)$, where N denotes the size of the input. When $k = 2$ these are the classic matroid parity and intersection problems and can be solved in polynomial time, but for $k > 2$ are NP hard. The first algorithms for general k faster than naïve enumeration were given by Barvinok in [Bar95], and had runtimes $(km)^{2k+1}4^{km}\operatorname{poly}(N)$ and $(km)^{2k}4^{k^2m}\operatorname{poly}(N)$, respectively. Parameterized algorithms for these problems were also given by Marx in [Mar09] where they were used to give fixed-parameter tractable algorithms for several other problems. The fastest algorithms prior to our work were due to Fomin et al. [FLPS16] and had runtime $2^{km\omega}\operatorname{poly}(N)$.

By combining Theorem 2.5.9 with a known construction of the determinant as a skew circuit [MV97], we obtain a faster deterministic algorithm for the following problem:

Problem 2.5.4 (SING)**.** Given matrices $A_1, \dots, A_n \in \mathbb{Q}^{d\times d}$, decide if their span contains an invertible matrix. Equivalently, decide if $\det \sum_{i=1}^n x_i A_i \not\equiv 0$.

In [BP21] we showed that SING can be solved in $4^d \operatorname{poly}(N)$. The fastest previous algorithm, given by Gurvits in [Gur03], had runtime $2^n n! \operatorname{poly}(N)$ and made use of an upper bound of $2^d d!$ on $\mathbf{R}_S(\det_n)$. This problem was originally studied by Edmonds for its application to matching problems [Edm67] and is of fundamental importance to complexity theory [KI04]. As a result variants of it have attracted attention, leading to a recent breakthrough in the non-commutative setting [GGOW19].

Of particular interest will be the case of Theorem 2.5.9 when X is a Hankel matrix, meaning that $(X)_{i,j} = (X)_{i+k,j-k}$ for all $k = 0, \dots, j-i$. Our main result shows the following exponential improvement in this special case:

Theorem 2.5.5. *Let C be a skew arithmetic circuit computing $g \in \mathcal{S}^n_d$, and let $X = (\ell_{i,j})_{i,j\in[d]}$ be a symbolic Hankel matrix with entries in $\mathcal{S}^n_1$. Then we can compute $\langle \det X, g\rangle$ with $\varphi^{2d} \operatorname{poly}(d)|C|$ arithmetic operations.*

The improvement in Theorem 2.5.13 over Theorem 2.5.9 is due to the fact that the vector space of partial derivatives of the determinant has dimension about 4^d, whereas the space of partial derivatives of the generic Hankel determinant has dimension less than φ^{2d}.

Theorem 2.5.13 yields the following applications:

Corollary 2.5.6. *The following admit deterministic algorithms running in time $\varphi^{2d} \operatorname{poly}(n)$:*

1. *Deciding whether a given directed n-vertex graph has a directed spanning tree with at least d non-leaf vertices,*
2. *Deciding whether a given edge-colored, directed n-vertex graph has a directed spanning tree containing at least d colors,*
3. *Deciding whether a given planar, edge-colored, directed n-vertex graph has a perfect matching containing at least d colors.*

The previous fastest algorithms for these problems had runtimes $3.19^d \cdot \operatorname{poly}(n)$, $4^d \operatorname{poly}(n)$, and $4^d \operatorname{poly}(n)$, respectively [Bra]. This built upon work of Gutin et al. [GRWZ18b], which gave runtimes $3.41^d \operatorname{poly}(n)$, $4.32^d \operatorname{poly}(n)$, and $4.32^d \operatorname{poly}(n)$. Problem (1) is the best studied among these, with [GRWZ18b, Table 1] listing eleven articles on this problem in the last fourteen years. It is noteworthy that our improvements do not rely on any problem-specific adaptations.

Theorem 2.5.13 also yields a $\varphi^{2d} \operatorname{poly}(n)$-time algorithm for detecting simple cycles (and paths, and more generally subgraphs of bounded treewidth). While it is known that simple cycles of length d in an n-vertex directed graph can be detected in randomized time $2^d \operatorname{poly}(n)$ [Wil09b] (and $1.66^d \operatorname{poly}(n)$ for undirected graphs [Bjö10b]), it is a major open problem to achieve this runtime deterministically. The fastest deterministic algorithm was given by Tsur [Tsu19] and runs in time $2.55^d \operatorname{poly}(n)$. This improved on a long line of work which started 35 years ago with a $d! \operatorname{poly}(n)$-time algorithm of Monien [Mon85]. It is important to note that our algorithm only works for unweighted graphs (or weighted graphs with integer weights bounded by $\operatorname{poly}(n)$), while the fastest known and some earlier algorithms work for weighted graphs. Our approach differs from several previous algorithms, which have been based on paradigms such as *color coding* and *divide and color* [CFK+15, Chapter 5].

In the next section we prove Theorem 2.5.9 and Theorem 2.5.13. These algorithms work by inductively evaluating a circuit, storing at each gate the result of differentiating $\det X$ by the polynomial computed at that gate. The key to our algorithm is that the spaces of partial derivatives of $\det X$ has dimension at most 4^d, and if X is a Hankel matrix this dimension is at most φ^{2d}. So while one might naïvely represent an element in these spaces as a linear combination of $\binom{n+d}{d}$ monomials, doing so generally includes a significant amount of unnecessary information. Instead, we express elements in these spaces as linear combinations of minors (or maximal minors in the Hankel case) of X. Ultimately our algorithms will compute at each gate a vector indexed by (pairs of) increasing sequences representing a linear combinations of minors.

2.5.1 Proofs

We start by giving an algorithm for computing (2.11) in the case that g is the determinant of a symbolic matrix and f is computed by a skew arithmetic circuit C. This is a warmup for the special case when g is the determinant of a symbolic Hankel matrix. We denote by $|C|$ the total number of gates in C.

We assume all arithmetic operations can be done in $\mathrm{poly}(n)$ time for ease of exposition; this is innocuous as our applications will only involve numbers with $\mathrm{poly}(n)$-bounded bit-length.

Let $\mathbb{N}_d^k$ be the set of k-tuples with elements in $[d]$, and let $I(d,k) \subseteq \mathbb{N}_d^k$ be the set of strictly increasing sequences of length k with elements in $[d]$; when $k = 0$ we include the empty sequence. Given a $d \times d$ matrix X and tuples $\alpha, \beta \in I(d,k)$, we denote by $X[\alpha|\beta]$ the minor (determinant of a submatrix) of X with rows indexed by α and columns indexed by β. We declare the "empty minor" $X[\,|\,]$ to equal one. We use the notation $a_1, \dots, \hat{a}_i, \dots, a_k$ to denote the sequence $a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_k$ obtained by omitting a_i.

Note that since square $k \times k$ submatrices of X can be identified by pairs of elements in $I(d,k)$ (their row and column indices), the vector space spanned by all minors of X has dimension at most $\sum_{k=0}^{d} |I(d,k)|^2 = \sum_{k=0}^{d} \binom{d}{k}^2 = \binom{2d}{d}$.

For $f \in \mathcal{S}_d^n$, $\mathrm{Derivs}(f)$ denotes the vector space spanned by the partial derivatives of f of all orders (this includes f itself). For example, $\mathrm{Derivs}(x_1x_2)$ is the vector space spanned by x_1x_2, x_1, x_2, and 1. The next observation is a simple bound on this quantity for determinants of symbolic matrices, and has been essentially observed several times previously (e.g. [Sha15, Lemma 1.3]).

Proposition 2.5.7. *Let $X = (\ell_{i,j})_{i,j\in[d]}$ be a symbolic matrix with entries in $\mathcal{S}_1^n$. Then the space $\mathrm{Derivs}(\det X)$ is contained in the space of minors of X. Hence*

$$\dim \mathrm{Derivs}(\det X) \le \sum_{i=0}^{d} \binom{d}{i}^2 = \binom{2d}{d} < 4^d.$$

Proof. Let $\mathfrak{S}_d$ denote the symmetric group on d elements. By the Leibniz formula for the determinant and the product rule, for any $l \in [n]$,

$$\begin{aligned}\frac{\partial \det X}{\partial x_l} &= \sum_{\sigma\in\mathfrak{S}_d} \operatorname{sgn}(\sigma)\sum_{i=1}^{d} \frac{\partial \ell_{i,\sigma(i)}}{\partial x_l}\prod_{j\neq i}\ell_{j,\sigma(j)} \\ &= \sum_{1\leq i,j\leq d} \frac{\partial \ell_{i,j}}{\partial x_l} \sum_{\sigma\in\mathfrak{S}_d,\sigma(i)=j} \operatorname{sgn}(\sigma)\prod_{m\neq i}\ell_{m,\sigma(m)} \\ &= \sum_{1\leq i,j\leq d} (-1)^{i+j}\frac{\partial \ell_{i,j}}{\partial x_l} X[1,\dots,\hat{\imath},\dots,d|1,\dots,\hat{\jmath},\dots,d].\end{aligned}$$

Note that since the entries of X are linear forms, $\frac{\partial \ell_{i,j}}{\partial x_l}$ is a scalar. To see the last equality, consider the martix $X^{(ij)}$ obtained by setting the (i,j)th entry of X to 1, and all other entries in the ith row of X to zero. Then $\det X^{(ij)} = \sum_{\sigma\in\mathfrak{S}_d,\sigma(i)=j}\operatorname{sgn}(\sigma)\prod_{m\neq i}\ell_{m,\sigma(m)}$, but at the same time by Laplace expansion, $\det X^{(ij)} = (-1)^{i+j}X[1,\dots,\hat{\imath},\dots,d|1,\dots,\hat{\jmath},\dots,d]$.

This shows that the space of order-1 partial derivatives of $\det X$ is contained in the span of the degree-$(d-1)$ minors of X. The proposition follows by repeated application of this fact. ■

Lemma 2.5.8. *Given as input a symbolic matrix $X = (\ell_{i,j})_{i,j\in[d]}$ with entries in $\mathcal{S}_1^n$, a linear combination P of minors of X, and $l \in [n]$, we can compute a representation for $\frac{\partial P}{\partial x_l}$ as a linear combination of minors of X with $4^d \operatorname{poly}(d)$ arithmetic operations.*

Proof. Let $P = \sum_{k=0}^{d}\sum_{\alpha,\beta\in I(d,k)} c_{\alpha,\beta}X[\alpha|\beta]$ and let $a_{i,j}^{(l)}$ be the coefficient of x_l in $\ell_{i,j}$ (so the input consists of l and the vectors $(c_{\alpha,\beta}) \in \mathbb{R}^{\binom{2d}{d}}$, $(a_{i,j}^{(k)}) \in \mathbb{R}^{d^2 n}$). Then by the same considerations as in the proof of Proposition 2.5.7, $\frac{\partial P}{\partial x_l}$ equals

$$\sum_{k=1}^{d}\sum_{\alpha,\beta\in I(d,k)}\sum_{1\leq i,j\leq k} c_{\alpha,\beta}(-1)^{i+j}a_{i,j}^{(l)}X[\alpha_1,\dots,\hat{\alpha}_i,\dots,\alpha_k|\beta_1,\dots,\hat{\beta}_j,\dots,\beta_k].$$

Note that for $\alpha,\beta \in I(d,k)$, the coefficient of $X[\alpha|\beta]$ in the above equals

$$\sum_{1\leq i,j\leq k}\sum_{\substack{\alpha',\beta'\in I(d,k+1)\\ \alpha=\alpha'_1,\dots,\hat{\alpha}'_i,\dots,\alpha'_{k+1}\\ \beta=\beta'_1,\dots,\hat{\beta}'_j,\dots,\beta'_{k+1}}} (-1)^{i+j}a_{i,j}^{(l)}c_{\alpha',\beta'}.$$

The numbers of pairs of sequences α', β' considered by the inner sum is naïvely at most d^4, and hence the coefficient of each minor can be computed with $O(d^6)$ arithmetic operations. Since there are $\binom{2d}{d}$ minors, all coefficients can be computed with the stated number of operations. ■

Theorem 2.5.9. *Let C be a skew arithmetic circuit computing $g \in \mathcal{S}_d^n$, and let $X = (\ell_{i,j})_{i,j\in[d]}$ be a symbolic matrix with entries in $\mathcal{S}_1^n$. Then we can compute $\langle \det X, g\rangle$ with $4^d|C|\operatorname{poly}(d)$ arithmetic operations.*

Proof. Say that gate v in C computes the polynomial C_v. We will compute the inner product (2.11) inductively: at gate v we will compute and store C_v^∂, a representation for $C_v(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}) \det A$ as a linear combination of minors of X. C_v^∂ will be stored as a vector of length $\binom{2d}{d}$ indexed by pairs of row and column sets. At the end of the algorithm we will have computed $f(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}) \det X = \langle f, \det X \rangle$ at the output gate.

We start by computing and storing $\frac{\partial}{\partial x_l} \det X$ at input gate x_l, which by Lemma 2.5.8 can be done in $4^d \operatorname{poly}(n, d)$ time. Now suppose that gate v takes input from gates v' and v'', and that we have already computed $C_{v'}^\partial$ and $C_{v''}^\partial$. To compute C_v^∂, there are two cases to consider:

1. $C_v = x_i \cdot C_{v'}$. Then $C_v^\partial = \frac{\partial}{\partial x_i} C_{v'}(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}) \det A = \frac{\partial}{\partial x_i} C_{v'}^\partial$. Using Lemma 2.5.8 this can be done with $O(4^d d^3)$ operations.
2. $C_v = C_{v'} + C_{v''}$. Since differentiation is linear, $C_v^\partial = C_{v'}^\partial + C_{v''}^\partial$. Since $C_{v'}^\partial$ and $C_{v''}^\partial$ are vectors of length $\binom{2d}{d}$, it takes $\binom{2d}{d} \operatorname{poly}(n)$ additions to add them.

Hence at each gate we use $O(4^d d^3)$ arithmetic operations, for a total of $O(4^d d^3 |C|)$. ∎

We now show how Theorem 2.5.9 can be applied to obtain a deterministic algorithm for detecting simple cycles in graphs. This is not competitive, but it motivates a following improvement.

Proposition 2.5.10. *Let G be a graph on n vertices. We can decide in $4^d \operatorname{poly}(n)$ time if G contains a simple cycle of length d.*

Proof. Let $V \in \mathbb{Q}^{d \times n}$ be the Vandermonde matrix with $(V)_{i,j} = j^i$. Let $X = V \cdot \operatorname{diag}(x_1, \ldots, x_n) \cdot V^T$. By the Cauchy-Binet Theorem,

$$\det X = \sum_{\alpha \in I(n,d)} V[1, \ldots, d|\alpha]^2 \prod_{i \in S} x_i.$$

Since any d columns in V are linearly independent, $V[1, \ldots, d|\alpha]^2 > 0$ for all $\alpha \in I(n, d)$. Furthermore, observe that $tr(A_G^d)$ has nonnegative coefficients and contains a square-free monomial if and only if G contains a simple cycle of length d. It follows that $\langle \det A, tr(A_G^d) \rangle \neq 0$ if and only if G contains such a cycle. In addition, $tr(A_G^d)$ can be naïvely computed by a skew circuit of size $O(dn^3)$. The theorem follows by applying Theorem 2.5.9, noting that we only perform arithmetic on $\operatorname{poly}(n)$-bit integers. ∎

Note that the (i, j)th entry in the matrix X in the proof of Proposition 2.5.10 is equal to $\sum_{k=1}^n k^{i+j} x_k$, and therefore X is Hankel. We now show how this additional structure can be exploited to give a significant improvement.

Fix linear forms $\ell_1, \ldots, \ell_{2d-1} \in \mathcal{S}_1^n$, and let C_d be the symbolic matrix

$$\begin{bmatrix} \ell_1 & \ell_2 & \ell_3 & \cdots & \cdots & \cdots & \ell_{2d-2} & \ell_{2d-1} \\ \ell_2 & \ell_3 & \cdots & \cdots & \cdots & \cdots & \ell_{2d-1} & 0 \\ \ell_3 & \cdots & \cdots & \cdots & \cdots & \ell_{2d-1} & 0 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 & 0 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \cdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ \ell_{2d-2} & \ell_{2d-1} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \ell_{2d-1} & 0 & 0 & \cdots & \cdots & \cdots & 0 & 0 \end{bmatrix} \tag{2.12}$$

The minors of the form $C_d[1, 2, \ldots, k|b_1, \ldots, b_k]$, where $k \leq d$ and $b_k \leq 2d - k$, are called *maximal*. For brevity we denote such a minor by $C_d[b_1, \ldots, b_k]$. Let H_d be the submatrix of C_d with row and column subscripts $1, \ldots, d$. It is readily seen that H_d is a Hankel matrix.

Proposition 2.5.11. Derivs$(\det H_d)$ *is contained in the space of maximal minors of C_d, and the number of maximal minors of C_d is at most φ^{2d}.*

Proof. The maximal minors of C_d span the space of minors of H_d by Corollary 2.2(c) of [Con98]. Hence by Proposition 2.5.7, they span the space of partial derivatives of $\det H_d$. The second claim follows by noting that the number of maximal minors of degree k equals $|I(2d-k, k)| = \binom{2d-k}{k}$. Hence the total number of maximal minors equals $\sum_{k=0}^{d} \binom{2d-k}{k} < \varphi^{2d}$, using in the final inequality the facts that the dth Fibonacci number satisfies $F_d = \sum_{k=0}^{\lfloor \frac{d-1}{2} \rfloor} \binom{d+k-1}{k}$, and that $F_d \leq \varphi^{d-1}$. ∎

Lemma 2.5.12. *Given as input a linear combination P of maximal minors of C_d and $l \in [2d-1]$, we can compute a representation for $\frac{\partial P}{\partial x_l}$ as a linear combination of maximal minors of C_d with $\varphi^{2d}\,\mathrm{poly}(d)$ arithmetic operations.*

Proof. For brevity we will write $[\alpha]$ for the minor $C_d[\alpha]$. Let $P = \sum_{k=0}^{d} \sum_{\beta \in I(2d-k,k)} c_\beta[\beta]$, and say that the coefficient of x_l in $(C_d)_{i,j}$ is $a_{i,j}^{(l)}$. As in Lemma 2.5.8,

$$\frac{\partial P}{\partial x_l} = \sum_{k=1}^{d} \sum_{\beta \in I(2d-k,k)} c_\beta \sum_{1 \leq i,j \leq k} (-1)^{i+\beta_j} a_{i,\beta_j}^{(l)} [1, \ldots, \hat{i}, \ldots, k|\beta_1, \ldots, \hat{\beta}_j, \ldots, \beta_k].$$

Note that the only minors appearing with nonzero coefficients are of the form $[1, \ldots, \hat{i}, \ldots, k|\gamma]$ for $k \in [d], i \in [k]$ and $\gamma \in I(2d-k, k-1)$. Call the coefficient of this minor in the above $b(i, \gamma)$. Then

$$b(i, \gamma) = \sum_{1 \leq j \leq k} \sum_{\substack{\beta \in I(2d-k,k) \\ \gamma = (\beta_1, \ldots, \hat{\beta}_j, \ldots, \beta_k)}} c_\beta (-1)^{i+\beta_j} a_{i,\beta_j}^{(l)}.$$

We can enumerate over all such sequences β considered by the inner sum in time $O(d^2)$, and hence $b(i, \gamma)$ can be computed with $O(d^3)$ additions and multiplications. We can thus compute

$$\frac{\partial P}{\partial x_l} = \sum_{k=1}^{d} \sum_{i=1}^{k} \sum_{\gamma \in I(2d-k,k-1)} b(i, \gamma)[1, \ldots, \hat{i}, \ldots, k|\gamma] \tag{2.13}$$

with $d^4 \sum_{k=1}^d |I(2d-k, k-1)| \leq \varphi^{2d} \operatorname{poly}(d)$ arithmetic operations. Note that this expresses $\frac{\partial P}{\partial x_l}$ as a linear combination of minors that are not necessarily maximal. We now fix this.

We first claim that for all $i \in [k]$ and $\beta \in I(2d-k, k-1)$,

$$[1, \ldots, \hat{i}, \ldots, k|\beta] = \sum_{J \subseteq [k-1], |J|=k-i} [e(J) + (1, \ldots, k-1)|\beta]$$

where $e(J)$ is the indicator vector of the set J. This holds since when $J = \{i, \ldots, k-1\}$, $e(J) + (1, \ldots, k-1) = (1, \ldots, \hat{i}, \ldots, k)$, and for all other J, $e(J) + (1, \ldots, k-1)$ will have a repeated value and hence $[e(J) + (1, \ldots, k-1)|\beta] = 0$.

Given this claim, it follows from [Con98, Lemma 2.1(a)] that

$$[1, \ldots, \hat{i}, \ldots, k|\beta] = \sum_{J \subseteq [k-1], |J|=k-i} [\beta + e(J)],$$

and so letting Q_k be the degree-k part of Equation (2.13),

$$Q_k = \sum_{i=1}^{k+1} \sum_{\beta \in I(2d-k-1,k)} b(i, \beta) \sum_{J \subseteq [k], |J|=k+1-i} [\beta + e(J)].$$

We now show how to efficiently compute the coefficients of the maximal minors in this expression from the already computed $b(i, \gamma)$'s.

Let $0 \leq k \leq d-1$ be fixed. For $\beta \in I(2d-k-1, k)$ and integers i, j where $0 \leq i \leq j \leq k$, let $D(\beta, i, j, k) \subseteq \{0,1\}^k$ be the set of binary vectors of length k containing exactly i ones, whose last $k-j$ entries are zero, and whose summation with β is strictly increasing everywhere except possibly at positions j and $j+1$ (that is, we may have $w_j + \beta_j = w_{j+1} + \beta_{j+1}$). Define

$$A^k(i,j) \coloneqq \sum_{\beta \in I(2d-k-1,k)} b(k+1-i, \beta) \sum_{w \in D(\beta,i,j,k)} [\beta + w].$$

Note that $\sum_{i=0}^k A^k(i,k) = Q_k$, so it suffices to show how to compute $A^k(i,j)$ for all i, j. We do this with a dynamic program. When we store $A^k(i,j)$ we will store all coefficients of maximal minors arising in the above definition, even though such a minor might contain a repeated column and hence equal zero. The minors arising in this definition are specified by sequences of length k with maximum value $2d-k$ that are strictly increasing everywhere but possibly at one position. Hence the number of such sequences is at most $k\binom{2d-k}{k}$.

For the base cases, we have

$$A^k(0,j) = \sum_{\beta \in I(2d-k-1,k)} b(k+1, \beta)[\beta],$$
$$A^k(i,i) = \sum_{\beta \in I(2d-k-1,k)} b(k+1-i, \beta)[\beta + e(\{1, \ldots, i\})].$$

These vectors are initialized in time $O(k\binom{2d-k}{k})$. Now suppose we have computed quantities $A^k(i,j-1)$ and $A^k(i-1,j-1)$. Then $A^k(i,j)$ equals

$$\sum_{\beta\in I(2d-k-1,k)} b(k+1-i,\beta)\left(\sum_{\substack{w\in B(\beta,i,j,k),\\ w_j=0}}[\beta+w]+\sum_{\substack{w\in D(\beta,i,j,k),\\ w_j=1}}[\beta+w]\right)$$

$$=\sum_{\beta\in I(2d-k-1,k)} b(k+1-i,\beta)\sum_{\substack{w\in D(\beta,i,j-1,k),\\ \beta+w \text{ is strictly increasing}}}[\beta+w]$$

$$+\sum_{\beta\in I(2d-k-1,k)} b(k+1-i,\beta)\sum_{w\in D(\beta,i-1,j-1,k)}[\beta+w+e(\{j\})].$$

The first part of the sum can be computed from $A^k(i,j-1)$ by setting the coefficient of any maximal minor with a repeated column equal zero, and the second sum can be computed from $A^k(i-1,j-1)$ by setting the coefficient of $[\beta]$ to that of $[\beta-e(\{j\})]$. Hence $A^k(i,j)$ can be computed with $O(k\binom{2d-k}{k})$ arithmetic operations. It follows that we can represent $\frac{\partial P}{\partial x_l}=\sum_{i=0}^{d-1}Q_i$ in the space of maximal minors using $\varphi^{2d}\operatorname{poly}(d)$ arithmetic operations. ■

With this we have the following analog of Theorem 2.5.9. We omit the proof as it is almost exactly the same, we just work in the space of maximal minors rather than minors, using Lemma 2.5.12 to differentiate instead of Lemma 2.5.8.

Theorem 2.5.13. *Let C be a skew arithmetic circuit computing $g\in\mathcal{S}_d^n$, and let $X=(\ell_{i,j})_{i,j\in[d]}$ be a symbolic Hankel matrix with entries in $\mathcal{S}_1^n$. Then we can compute $\langle\det X,g\rangle$ with $\varphi^{2d}\operatorname{poly}(d)|C|$ arithmetic operations.*

Corollary 2.5.14. *Let G be a graph on n vertices. We can decide in $\varphi^{2d}\operatorname{poly}(n)$ time if G contains a simple cycle of length d.*

Proof. Let $V\in\mathbb{Q}^{d\times n}$ be the Vandermonde matrix with $(B)_{i,j}=j^i$, and $X=V\cdot diag(x_1,\dots,x_n)\cdot V^T$. By the argument of Proposition 2.5.10, $\langle\det X,\operatorname{tr}(A_G)^d\rangle\neq 0$ if and only if G contains a simple cycle of length d. Note that the (i,j)th entry in X equals $\sum_{k=1}^n k^{i+j}x_k$, and therefore X is Hankel. We conclude by applying Theorem 2.5.13 to compute $\langle\det X,tr(A_G)^d\rangle$, as $tr(A_G^d)$ can be computed by a skew circuit of size $\operatorname{poly}(n)$. ■

2.6 Further questions

Question 2.6.1. For all integers u,v such that $u+v=d$, what is the minimum rank of a matrix with rows indexed by subsets of $[n]$ of size u and columns indexed by subsets of $[n]$ of size v, such that entry (I,J) is nonzero if and only if $I\cap J=\emptyset$, and entry (I,J) equals entry (K,L) whenever $I\cup J=K\cup L$? It follows from the method of partial derivatives that this quantity is a lower bound on $A(n,d)$. Theorem 2.3.19 shows that this is at most 2.6^d.

Question 2.6.2. How many points are there in $\mathbb{C}^n$ such that the spaces spanned by any $d-1$ of them are contained in $\mathbf{V}(e_{n,d})$, but the spaces spanned by any d of them are not? It is easy to see that $\mathbf{V}(e_{3,2})$ contains infinitely many such points; could it be that for all d and some fixed $c \in \mathbb{N}$, $\mathbf{V}(e_{d+c,d})$ contains infinitely many such points? This would imply that $A(\mathbb{N}, d) \leq 2^d \mathrm{poly}(d)$.

Question 2.6.3. Similarly, how many matrices in $\mathbb{C}^{n \times n}$ have the property that the span of any $d-1$ of them is contained in $\mathbf{V}(\mathrm{per}_d)$, but not the span of any d of them? If there exist infinitely many points then it follows from Proposition 2.3.5 and the fact that $\mathbf{R}_S(\mathrm{per}_d) \leq 4^{d-1}$ [Lan12] that $A(\mathbb{N}, d) \leq 4^{d-1}$.

Question 2.6.4. Do all (g, ε)-support intersection certification algorithms require $\mathbf{R}_{\mathrm{supp}}(g)$ queries? Proposition 2.2.12 shows that this is the case for monomials. Similarly, are $\mathbf{R}^{\varepsilon}_{\mathrm{supp}}(g)$ queries required to compute a $(1 \pm \varepsilon)$ approximation of $f(\partial \mathbf{x})g$ in the general black-box setting? Theorem 2.1.6 shows that this is true when $\varepsilon = 0$.

Remark 2.6.5. Theorem 2.3.42 can be made algorithmic by using an explicit construction of a perfect splitter. The only such constructions we know however are far from optimal; that is, they give families of functions much larger than $\sigma(n, d, n_0, d_0)$ in general.

Chapter 3

A Brief History of Fast Matrix Multiplication

The exponent of matrix multiplication is the smallest number ω such that for each $\varepsilon > 0$, there exists an algorithm for multiplying two $n \times n$ matrices using $O(n^{\omega+\varepsilon})$ field operations[1]. It is clear that $\omega \geq 2$, and a long line of work has led to the best upper bound currently known of $\omega < 2.372$ [WXXZ23]. It is a longstanding and well-known open problem to resolve the conjecture that $\omega = 2$.

In this chapter we give an overview of this area, which we will study in more detail in the following two chapters. We begin in the next subsection by giving a brief summary of the traditional approach to obtaining upper bounds, which was pioneered by Strassen [S$^+$69, Str86], Schöenhage [Sch81], and Coppersmith and Winograd [CW87]. In the following subsection we describe the group-theoretic approach of Cohn and Umans [CU03], which is known to capture the traditional approach. We then discuss two major barriers to obtaining $\omega = 2$ within the group–theoretic framework. Despite these barriers, we seem to be very far from ruling out obtaining $\omega = 2$ with this framework.

3.1 The laser method

Recall that the matrix multiplication tensor $\langle n, m, p \rangle$ is the trilinear form

$$\langle n, m, p \rangle := \sum_{(i,j,k) \in [n] \times [m] \times [p]} x_{ij} y_{jk} z_{kl}.$$

Strassen [S$^+$69] noted that

$$\omega = \lim_{n \to \infty} \mathbf{R}(\langle n, n, n \rangle)^{1/n},$$

and since then efforts at bounding ω have focused on bounding the rank of the matrix multiplication tensor.

[1] The bound on ω may depend on the choice of field, although it is known to not change under field extension [BCS13, Corollary 15.18]. All of the upper bounds on ω we discuss hold over any field.

A key operation on tensors in the study of fast matrix multiplication is the Kronecker product, defined as follows.

Definition 3.1.1. For tensors $f = \sum a_{ijk}x_iy_iz_i$ and $g = \sum b_{ijk}x_iy_iz_i$, their Kronecker product is the tensor

$$f \boxtimes g := \sum_{i,i',j,j',k,k'} a_{ijk}b_{i'j'k'}x_{ii'}y_{jj'}z_{kk'}$$

Note that this is exactly the 3-dimensional analogue of the Kronecker product operation on matrices. A key property of $\langle n,n,n\rangle$ is that it self-reproducing under the Kronecker product: $\langle n,m,p\rangle \boxtimes \langle n',m',p'\rangle = \langle nn',mm',pp'\rangle$. Furthermore, it is not difficult to see that tensor rank is submultiplicative under $\boxtimes$. From these facts, it follows that upper bounds on ω can be obtained from the rank of a *single* matrix multiplication tensor. For example, Strassen showed that $\mathbf{R}(\langle 2,2,2\rangle) \leq 7$ [S+69], which implies that $\omega \leq \log_2 7 \approx 2.81$.

Approaches to ω proceed via a "reduction" to an auxilliary tensor of known rank. The most general notion of reduction is that of a *degeneration* from one tensor to another.

Definition 3.1.2. We say that f is a degeneration of g, and write $f \leq g$, if there exists matrices A_i, B_i, C_i such that $\lim_{i\to\infty} g(A_i(x), B_i(y), C_i(z)) = f(x,y,z)$.

In the special case that the degeneration is obtained via monomial matrices (matrices with at most one nonzero entry per row and column), this is called a *combinatorial degeneration*. We call the even more restricted case when the matrices are zero off the diagonal and have entries in $\{0,1\}$ on the diagonal a *restriction*[2]. This simply corresponds to setting some of the variables in g to zero. Note that if $X = X_1 \sqcup X_2 \sqcup X_3$ is a tripartite 3-graph, and $T = \sum_{(a,b,c)\in E(X)} c_{abc}x_ay_Bz_c$ with $c_{abc} \neq 0$ is a tensor whose support is X, then restrictions of T correspond exactly to induced subhypergraphs of X. In the best current approaches, only restrictions are used.

We may then define the border rank of a tensor as follows.

Definition 3.1.3. f has border rank at most r if $f \leq \sum_{i=1}^{r} x_iy_iz_i$.

Remark 3.1.4. For any tensor f, we may define its asymptotic rank

$$\underset{\sim}{\mathbf{R}}(f) = \lim_{k\to\infty} \mathbf{R}(f^{\boxtimes k})^{1/k}$$

By Fekete's lemma, this limit exists. It is not hard to see that we can then alternatively define $\omega = \underset{\sim}{\mathbf{R}}(\langle 2,2,2\rangle)$.. Thus the conjecture $\omega = 2$ can be equivalently expressed as $4 = \underset{\sim}{\mathbf{R}}(\langle 2,2,2\rangle)$. Much more generally, it has been conjectured that for *any* tensor $f \in \mathbb{C}^n \otimes \mathbb{C}^n \otimes \mathbb{C}^n$, $\underset{\sim}{\mathbf{R}}(f) \leq n$ (the inequality can be strict for trivial reasons). Note that this conjecture is trivially true for tensors of border rank at most n. However, we are unaware of any example of a tensor that has border rank greater than n, but that has minimal asymptotic rank. Is it possible that the *opposite* conjecture is true, namely, if f does not have minimal border rank, then $\underset{\sim}{\mathbf{R}}(f)$ is not minimal? In particular, if $\underline{\mathbf{R}}(T)$ is not minimal, can $\underline{\mathbf{R}}(T^{\otimes 2})$ be minimal?

The workhorse in the current record upper bounds on ω is the *asymptotic sum inequality* of Schöenhage:

Theorem 3.1.5 ([Sch81]). *If* $\mathbf{R}(\bigoplus_{i=1}^{p}\langle k_i,m_i,n_i\rangle) \leq r$ *with* $r > p$ *then* $\omega < \tau$, *where* τ *is the solution to* $\sum_{i=1}^{p}(k_im_in_i)^{\tau/3} = r$.

[2] Usually these are called combinatorial restrictions.

To understand this theorem, consider for simplicity the case when $k_i = m_i = n_i = n$. In this case the theorem says that $\mathbf{R}(\bigoplus\langle n,n,n\rangle) \leq r$ implies that $\omega \leq \log_n r/p$. Now, observe that tensor rank is subadditive under $\oplus$. If it was *additive*, however, this bound would be immediate, as we would have $\mathbf{R}(\oplus_{i=1}^p\langle n,n,n\rangle) \leq r \implies \mathbf{R}(\langle n,n,n\rangle) \leq r/p$, so $\omega \leq \log_n(r/p)$. Thus, one can remember Theorem 3.1.5 as saying that a bound on the rank of a direct sum of matrix multiplications gives the bound on ω that one would get if tensor rank was additive.

In order to apply Theorem 3.1.5, one needs a bound on the rank of a direct sum of matrix multiplications. To obtain this, we start with an auxiliary tensor for which, by some algebraic argument, we have a rank bound. Then, we aim to find a degeneration of this auxiliary tensor to a "large" direct sum of matrix multiplication tensors. As reasoning about degenerations in full generality is very difficult in general (it is as hard as tensor rank!), in practice one only searches for combinatorial restrictions to direct sums of matrix multiplications.

How can we find families of "combinatorially interesting" tensors of low rank? One approach is to take Kronecker powers of a small starting tensor for which we have a rank bound. This general approach leads to Strassen's laser method. In the approach of the next section, one considers the families of tensors corresponding to multiplication in the group algebra of a finite group.

With these notions established, bounds on ω work as follows. We start with an auxiliary tensor T_0 of known rank. Then we hope to find a degeneration of T_0 to a direct sum of many large matrix multiplication tensors. Because rank does not increase under degeneration, this gives an upper bound on the rank of a direct sum of matrix multiplications. At this point we apply the asymptotic sum inequality.

If T_0 was itself a direct sum of matrix multiplication tensors, we could directly apply the asymptotic sum inequality.

In the best bounds since [CW87], T_0 is taken to be a Kronecker power of the *Coppersmith–Winograd tensor*

$$T_q := x_0y_0z_{q+1} + x_0y_{q+1}z_0 + x_{q+1}y_0z_0 + \sum_{i=1}^{q} x_0y_iz_i + x_iy_0z_i + x_iy_iz_0.$$

Coppersmith and Winograd showed that T_q has border rank $q+2$ (which is minimal), and hence $\underline{\mathbf{R}}(T_q^{\otimes k}) = (q+2)^k$. A key fact about T_q is that it is a (non-direct) sum of matrix multiplication tensors: we have

$$T_q = T_{011} + T_{101} + T_{110} + T_{200} + T_{020} + T_{002}$$

where $T_{011} \cong \langle q,1,1\rangle$, $T_{101} \cong \langle 1,q,1\rangle, T_{110} \cong \langle 1,1,q\rangle, T_{200}, T_{020}, T_{002} \cong \langle 1,1,1\rangle$. Strassen's laser method [Str86] makes use of this partitioning of T_q into blocks, each block isomorphic to a matrix multiplication. In particular, the large direct sum of matrix multiplications in a power ultimately comes from the fact that the "outer" structure contains an asymptotically large diagonal tensor. By cleverly zeroing out variables in $T_q^{\otimes k}$ and applying the asymptotic sum inequality, one can obtain the bounds of 2.372 [AW21]. This process is called *Strassen's laser method*.

3.2 A group–theoretic approach

In [CU03] a group-theoretic approach to bounding ω was proposed. Given any finite group G and three subsets of G satisfying a certain condition (the *triple product property*), this approach yields an upper bound on ω by reducing an instance of matrix multiplication to multiplication in the group algebra of G. This approach can capture the previously discussed Coppersmith–Winograd family of algorithms, which includes the current record bound. While some barriers are known, for instance that certain generalizations of the constructions in [CKSU05] cannot achieve $\omega = 2$ (see [BCC$^+$17a]), the possibility that one could show $\omega = 2$ using a suitable family of groups remains wide open.

For a finite group G, we let Irr(G) denote the set of irreducible complex representations of G. For $X \subseteq G$, let $Q(X) = \{xx'^{-1} : x, x' \in X\}$ be the quotient set of X.

Definition 3.2.1. We say that $S, T, U \subseteq G$ satisfy the *triple product property* (or TPP for short) if for all $s \in Q(S), t \in Q(T), u \in Q(U)$,

$$stu = 1 \implies s = t = u = 1.$$

Theorem 3.2.2. *[CU03, Theorem 4.1] If S, T, U satisfy the TPP in G, then*

$$(|S||T||U|)^{\omega/3} \leq \sum_i d_i^\omega$$

where $d_i \in \mathbb{N}$ are the dimensions of the irreducible representations of G.

The idea behind this theorem is as follows. First, consider the multiplication tensor for $\mathbb{C}[G]$, defined by

$$\sum_{a,b,c \in G : abc=1} x_a y_b z_c.$$

By the Artin–Wedderburn theorem, $\mathbb{C}[G]$ is isomorphic to a direct sum of matrix algebras. Using this fact, one can show that $\mathbf{R}(T_G) \leq \mathbf{R}(\bigoplus \langle d_i, d_i, d_i \rangle)$, where the d_i's are the dimensions of the matrix algebras. On the other hand, suppose that S, T, U satisfy the TPP. Then by zeroing-out all variables other than $x_{st^{-1}}, y_{tu^{-1}}, z_{us^{-1}}$, the TPP implies that $T_G \geq \langle |S||T|, |T||U|, |U||S| \rangle$, and hence $\mathbf{R}(T_G) \geq \mathbf{R}(\langle |S||T|, |T||U|, |U||S| \rangle)$. Therefore we obtain the bound $\mathbf{R}(\langle |S||T|, |T||U|, |U||S| \rangle \leq \mathbf{R}(\bigoplus \langle d_i, d_i, d_i \rangle)$. By applying a variant of the asymptotic sum inequality, one concludes Theorem 3.2.2.

The triple product property corresponds to finding a combinatorial restriction of T_G to a single matrix multiplication. However just as with the approach of Schöenhage, one can ask for a restriction of T_G to a direct sum of matrix multiplication tensors, and then apply the asymptotic sum inequality. Doing so one arrives at the more general notion of the simultaneous triple product property.

Definition 3.2.3. A collection of triples of subsets S_i, T_i, U_i of a group G satisfy the simultaneous triple product property (STPP) if

1. For each i, the sets S_i, T_i, U_i satisfy the triple product property
2. Setting $S_i = A_i B_i^{-1}, T_j = B_j C_j^{-1}, U_k = C_k A_k^{-1}$,

$$s_i t_j u_k = 1 \iff i = j = k$$

for all $s_i \in S_i, t_j \in T_j, u_k \in U_k$.

By the asymptotic sum inequality, we have the following:

Proposition 3.2.4. *If $S_i, T_i, U_i \subseteq G$ satisfy the STPP, then*

$$\sum_i (|S_i||T_i||U_i|)^{\omega/3} \leq \sum d_i^\omega.$$

Using Theorem 3.2.2 it was shown in [CKSU05] that $\omega < 2.41$. In fact, this framework is powerful enough to capture the Coppersmith-Winograd family of algorithms, which includes the bound of $\omega < 2.37286$ [AW21]. These are STPP constructions in $\mathbb{Z}_m^n$, where m is fixed.

Remark 3.2.5. A zeroing out of T_G to $\langle n, n, n \rangle$ corresponds to finding $\{a_{ij}\}_{i,j \in [n]}$, $\{b_{ij}\}_{i,j \in [n]}$, $\{c_{ij}\}_{i,j \in [n]} \subseteq G$ such that $a_{ij} b_{kl} c_{mp} = 1$ if and only if $j = k, l = m, p = i$. Given sets S, T, U satisfying the TPP, such elements are obtained from the product set $A = ST^{-1}, B = TU^{-1}, C = US^{-1}$. However, the TPP a-priori might not give the most general zeroing outs to matrix multiplication tensors. It turns out that this is the case: any zeroing out yields a TPP. Similarly, STPP's are in bijection with zeroing outs to direct sums of matrix multiplication tensors.

Remark 3.2.6. One may be able to prove a better bound on ω in a particular group using simultaneous triple product property constructions rather than TPP constructions. For instance, a TPP construction in an abelian group is easily shown to give no bound better than $\omega \leq 3$, while the current best bounds can be obtained via STPP construction in abelian groups. However, given a STPP construction obtaining a bound on ω, there is a family of TPP constructions inside different groups yielding the same bound on ω [CKSU05, Theorem 7.1]. Thus if one is optimizing over all groups, STPP constructions are no more general than TPP constructions.

A useful lower bound on the upper bound on ω obtained via Theorem 3.2.2 is the *pseudo-exponent* of a group. This is the bound on ω one would obtain if $d_i = 1$ for all i, i.e., if G was abelian.

Definition 3.2.7. If S, T, U satisfy the TPP in G, we say that the pseudo-exponent of G is at most $3 \log_{|S||T||U|} |G|$.

The conditions of the STPP imply that they satisfy the "packing bound" $\sum_i |S_i||T_i| \leq |G|$, and similarly for the other pairs of sets. A simple necessary condition for obtaining $\omega = 2$ via STPP constructions it that they asymptotically meet this bound (see [BCC$^+$17a]).

Definition 3.2.8. We say that a family of STPP constructions S_i, T_i, U_i meets the packing bound $\sum_i |S_i||T_i|, \sum_i |S_i||T_i|, \sum_i |S_i||T_i| \geq |G_i|^{1-o(1)}$.

3.3 Barriers

In this section we discuss two barriers for the group–theoretic approach. These are similar at a high level. In both cases, we show that M_n has a particular structure (in the first case this is a large induced matching, and in the second case this is a large independent set). Then, we identify properties of groups that are incompatible with this structure.

3.3.1 Slice rank and multiplicative matchings

In [BCC$^+$17a], it was shown that the current approach via STPP constructions in abelian groups of bounded exponent cannot yield $\omega = 2$:

Theorem 3.3.1. *For every $\ell \in \mathbb{N}$, there is an $\varepsilon_\ell > 0$ such that no STPP construction in any abelian group of exponent at most ℓ can yield a bound better than $\omega \leq 2 + \varepsilon_\ell$ via Proposition 3.2.4.*

This was subsequently extended to nilpotent groups satisfying a mild additional condition in [BCC⁺17b]. The key to these results is the notion of the *slice rank* of a tensor. This notion was introduced (indirectly) in the solution to the *cap set problem* in additive combinatorics, which we now discuss. It turns out that the proof of Theorem 3.3.1 and the proof of the cap-set problem are very similar.

A foundational problem in additive combinatorics is determining $r_3(n)$, the size of the largest subset of $[n]$ containing no nontrivial 3-term arithmetic progression, i.e., three numbers $x, x + y, x + 2y$ with $y \neq 0$. A lower bound of Behrend [Beh46] shows that $r_3(n) \geq n/e^{O(\sqrt{\log n})}$, and a long line of work culminating in the breakthrough of [KM23] shows that $r_3(n) \leq n/e^{O(\log^\beta n)}$ for some constant $\beta > 0$. The *cap set problem* is the analogous question in the vector space $\mathbb{F}_3^n$. Using an argument similar to one of Roth [Rot53] in the integer setting, Meshulam showed that $r_3(\mathbb{F}_3^n) \leq 3^n/n$ [Mes95]. Unlike in the integer setting however, only very weak lower bounds of the form c^n for some $c < 3$ were known [Ede04]. Then in a breakthrough of [EG17], it was shown that $r_3(\mathbb{F}_3^n) \leq 2.77^n$.

We will be interested in the following generalization of a 3AP-free set.

Definition 3.3.2. A *3-matching* in a finite group G is a triple of subsets $\{a_i\}_{i=1}^m, \{b_i\}_{i=1}^m, \{c_i\}_{i=1}^m$ of G such that $a_i b_j c_k = 1 \iff i = j = k$.

Let $M(G)$ denote the largest size of a 3-matching in G. The solution to the cap set problem can be adapted to show the following:

Theorem 3.3.3. $M(\mathbb{Z}_p^n) < (p - c_p)^n$*, for some* $c_p > 0$.

More generally, it was shown in [Saw18] that for any group H, $M(H^n) \leq (\delta \cdot |H|)^n$ for some $0 < \delta < 1$ depending on H. Note that this is equivalent to saying that the hypergraph X_{H^n} contains no induced matching of size $(\delta|H|)^n$.

The group–theoretic approach works by finding disjoint induced matrix multiplications inside of X_G. The key fact connecting the cap-set problem and matrix multiplication is the following.

Proposition 3.3.4. *M_n contains an induced matching of size $n^{2-o(1)}$.*

Proof. This follows from viewing M_n as the "edge-triangle" incidence hypergraph of the complete tripartite graph $K_{n,n,n}$, where vertices in M_n correspond to edges in $K_{n,n,n}$, and three vertices in M_n are adjacent if and only if these corresponding edges in $K_{n,n,n}$ form a triangle. From this perspective, an induced matching in M_n corresponds to a tripartite graph on at most $3n^2$ vertices where each edge belongs to a unique triangle. Moreover, the number of edges in this induced matching equals the number of edges in the corresponding graph. But determining the maximum number of edges in a tripartite graph with $3n$ vertices with the property that each edge is contained in a unique triangle is the Rusza-Szemerédi problem[3], and the best-known lower bound for this problem using Behrend's construction shows that this is at least $n^{2-o(1)}$ [Zha23, Corollary 2.5.2]. For completeness we recall this construction. Let $A \subseteq \mathbb{Z}_n$ be 3AP-free. Assume that n is odd (this does not affect the asymptotic claim). Define the tripartite graph with parts X, Y, Z equal to

[3] This was independently noted in [AB23]. The Rusza-Szemeredi problem is usually stated for arbitrary graphs which are not necessarily tripartite, but a standard probabilistic argument reduces the problem to the tripartite case.

$\mathbb{Z}_n$, and where (x, y) is an edge if $x - y \in A$, (y, z) is an edge when $z - y \in A$, and (x, z) is an edge if $(z - x)/2 \in A$. ■

Combining Theorem 3.3.3 and Proposition 3.3.4, we conclude that there cannot be many "large" M_n's inside of X_G, as that would imply that X_G contains a large matching. By quantifying "harge" we obtain Theorem 3.3.1.

A key idea in the proof of Theorem 3.3.3 is that of the *slice rank* of a tensor.

Definition 3.3.5. Let $F : \mathbb{F}^n \times \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ be a 3-tensor. The slice rank of F is the smallest number r for which there exist bilinear forms B_i and linear forms ℓ_i such that we can write

$$F(x, y, z) = \sum_{i=1}^{a} B_i(x, y)\ell_i(z) + \sum_{i=a+1}^{b} B_i(x, z)\ell_i(y) + \sum_{i=b+1}^{r} B_i(y, z)\ell_i(x)$$

An important fact is that the "diagonal" tensor has maximal slice rank:

Proposition 3.3.6. *The slice rank of $\sum_{i=1}^{n} x_i y_i z_i$ is n over any field.*

By combining these facts with an upper bound on the slice rank of $X_{\mathbb{Z}_p}$ in characteristic p, along with the fact that slice rank does not increase under changes of variables, one obtains Theorem 3.3.3.

3.3.2 Quasirandomness

In [BCG+22], it was shown that groups with strong "quasirandomness" properties cannot be used to prove bounds on ω. Specifically, the following was shown. Let $n(G)$ denote the minimal dimension of an irreducible representation of G of dimension greater than one.

Theorem 3.3.7. *If subsets S, T, and U satisfy the triple product property in a finite nonabelian group G, then*

$$|S|\,|T|\,|U| \leq \frac{|G|^{3/2}}{n(G)^{1/2}} + |G|.$$

This implies for example that if $n(G) \geq |G|^{\delta}$ for some fixed $\delta > 0$, then one cannot meet the packing bound (Definition 3.2.8) in G, and hence cannot obtain $\omega = 2$. As an example, this rules out groups of Lie type of bounded rank, such as $\mathrm{PSL}(2, p)$.

The proof of Theorem 3.3.7 builds off of Gowers's work on product-free subsets of quasirandom groups [Gow08]. A subset X of group is product free if there is no solution to the equation $ab = c$ with $a, b, c \in X$. For example, in $\mathbb{Z}_n$ the set $\{n/3, n/3 + 1, \ldots, 2n/3\}$ is "product" (sum) free. Babai and Sós asked if all groups behave similarly to $\mathbb{Z}_n$ in that they have product-free sets of size $|G|^{1-o(1)}$. Gowers gave a negative answer to this question, showing that $\mathrm{PSL}_2(p)$ has no product-free set of size greater than $|G|^{7/8}$.

A lack of product-free subsets roughly implies Theorem 3.3.7 due to the simple fact that if S, T, U are subsets of G of size $|G|^{1/2-o(1)}$ satisfying the triple product property, then there are subsets $A, B, C \subseteq G$ each of size $|G|^{1-o(1)}$ that are product-free. For example, one can take $A = ST_0^{-1}, B = T_1U^{-1}, C = US^{-1}$, where $T_0 \cap T_1 = \emptyset$.

Chapter 4

Matrix Groups within the Group–Theoretic Approach

4.1 Introduction

Previous work on the group-theoretic approach has focused mainly on families of very *non-simple* groups, i.e., groups built up from simple groups through repeated group extension. For example, the best bounds known on ω can be obtained using a semidirect product of the symmetric groups with direct products of abelian groups. At the same time, several barrier results have been shown for these kinds of constructions (see, for example, [BCC$^+$17a, BCC$^+$17c, Saw18]). But this has left the simple groups—in some sense the opposite end of the spectrum of finite groups—largely unexplored.

In this chapter we address this gap in knowledge by studying *finite groups of Lie type*[1] in the framework of [CU03]. This is an important class of groups that contains all of the finite simple groups except alternating or cyclic groups and finitely many sporadic groups. Some good examples to keep in mind are the classical matrix groups such as the group $SL(n, q)$ of determinant 1 matrices over the finite field $\mathbb{F}_q$ or the group of $n \times n$ orthogonal matrices over $\mathbb{F}_q$ with respect to a quadratic form.

4.1.1 Results

We start by showing that triple product property constructions (see Definition 3.2.1) in groups of Lie type cannot prove any bound on ω better than $2 + \varepsilon$ for some absolute constant $\varepsilon > 0$ (Corollary 4.2.4). This resolves a question asked in [CU03]. Our proof combines a representation-theoretic argument with known bounds on the dimension and number of irreducible representations in groups of Lie type [LS74, FG12]. More broadly, we identify the second-smallest dimension of

[1] Specifically, by a group of Lie type we mean any one of the (possibly twisted) Chevalley groups, including the Suzuki and Ree groups, or the quotient of such a group by its center. A Chevalley group is the fixed points of a Steinberg endomorphism in a semisimple algebraic group over a finite field (see Definition 21.6 and Table 22.1 in [MT11]). Among others, this list includes $SL(n, q)$, $SU(n, q)$, $SO(2n + 1, q)$, $Sp(2n, q)$, $SO^+(2n, q)$, and $SO^-(2n, q)$. Obtaining simple groups can require taking the quotient by the center, but that does not change our conclusions, such as Corollary 4.2.4.

an irreducible representation as a key parameter of a group that determines its viability for the approach of [CU03]: Theorem 4.2.2 shows that groups where this quantity is large cannot yield good bounds on ω. For example, any family of groups for which the second-smallest dimension of an irreducible representation grows as a power of the size of the group cannot yield $\omega = 2$. It had been known since [CU03] that the largest dimension played a key role in the quality of the bound, but small dimensions were not previously understood to be relevant.

This first barrier builds on Gowers' theorem on product-free sets in quasirandom groups [Gow08]. We note that whereas Gowers' result involves the minimum dimension of a *nontrivial* representation, the additional structure of our problem allows us to consider the second-smallest dimension of an irreducible representation (in other words, we can skip any other representations of dimension 1). This gives us lower bounds in groups where Gowers' result does not apply, such as $SL(n, q)$. It is interesting that while triple product property constructions in abelian groups cannot yield nontrivial bounds on ω, this barrier shows that *highly nonabelian* groups also have significant limitations.

Next we show in Theorem 4.2.6 that subgroups with large normalizers cannot be used in a triple product property construction to obtain $\omega = 2$. This barrier is particularly effective in the setting of matrix groups. For example, one cannot obtain $\omega = 2$ via a triple product property construction using three subgroups inside $GL(n, q)$ for varying q and any fixed n, or even inside products of such groups.

Our first barrier result rules out obtaining exponent 2 from finite groups of Lie type, but still leaves open the possibility that such groups could serve as building blocks in efficient algorithms for matrix multiplication. For example, the *direct product* of such groups escapes the barrier entirely, since the second-smallest dimension of an irreducible representation of a direct product equals the second-smallest dimension among the irreducible representations of the factors. Similarly, our normalizer barrier suggests that constructions should aim to use subgroups that are *self-normalizing*. We therefore view our barriers as giving us useful information about what a possible construction using finite groups of Lie type must look like, if it is to give $\omega = 2$.

In the second part of the paper, we give constructions that naturally use a direct product in a critical way (Theorem 4.3.9), and constructions that use self-normalizing subgroups (Theorem 4.3.10). It is important to note that we know of *no* constructions in *finite* groups of Lie type that even meet a certain "packing bound" (Definition 3.2.8), a prerequisite for obtaining $\omega = 2$. This remains an important challenge. In lieu of such constructions, we direct our efforts at obtaining constructions in *continuous* Lie groups, which seems easier and mathematically cleaner to work in, and where we can ask direct analogues of the main questions. Here we have constructions that meet the packing bound. Moreover, we give examples that use direct products and self-normalizing subgroups, desiderata by the barrier results.

Our constructions do not achieve the Lie analogue of beating exponent 3, and we suggest improving them and finding other examples as key challenges highlighted by this work.

4.1.2 Outline

In the next section we review the group-theoretic approach to bounding ω. In Section 4.2 we give our barriers. We then introduce the Lie exponent in Section 4.3 and give our constructions. We conclude with some questions in Section 4.4.

4.2 Barriers for matrix groups

In this section we explain the two barriers mentioned in the introduction. Each of them is based on an idea that is particularly relevant for matrix groups, although we formulate the bounds in greater generality.

4.2.1 A representation-theoretic barrier

We begin by proving our representation-theoretic barrier, which we then apply to groups of Lie type. Our proof of Theorem 4.2.2 follows the Fourier-analytic proof of Gowers' theorem on mixing in quasirandom groups (see, for example, [Bre14, Lemma 2.2]). Our barrier is a function of the second-smallest dimension of an irreducible representation of G. Because we use this parameter frequently, we introduce notation for it:

Definition 4.2.1. For a finite nonabelian group G, let

$$n(G) \coloneqq \min_{\pi \in \mathrm{Irr}(G)\colon \dim \pi > 1} \dim \pi$$

be the smallest dimension of an irreducible representation of G of dimension greater than 1.

Theorem 4.2.2. *If subsets S, T, and U satisfy the triple product property in a finite nonabelian group G, then*

$$|S|\,|T|\,|U| \leq \frac{|G|^{3/2}}{n(G)^{1/2}} + |G|.$$

Proof. Let 1_X denote the indicator function of a subset $X \subseteq G$. For brevity we will write $\pi(X) \coloneqq \sum_{x \in X} \pi(x)$ for $\pi \in \mathrm{Irr}(G)$ and $X \subseteq G$, and $d_\pi \coloneqq \dim \pi$.

Suppose that S, T, U satisfy the triple product property. Equivalently, the value at the identity in the 6-fold convolution $1_S * 1_{S^{-1}} * 1_T * 1_{T^{-1}} * 1_U * 1_{U^{-1}}$ equals $|S|\,|T|\,|U|$. The Fourier inversion formula says that a function $f\colon G \to \mathbb{C}$ can be reconstructed as

$$f(g) = \frac{1}{|G|} \sum_{\pi \in \mathrm{Irr}(G)} d_\pi \left\langle \sum_{h \in G} f(h)\pi(h), \pi(g) \right\rangle,$$

where $\langle X, Y \rangle = Tr(X\overline{Y}^\top)$. Applying this formula to $f = 1_S * 1_{S^{-1}} * 1_T * 1_{T^{-1}} * 1_U * 1_{U^{-1}}$ and $g = 1$ yields

$$|G|\,|S|\,|T|\,|U| = \sum_{\pi \in \mathrm{Irr}(G)} d_\pi Tr(\pi(S)\pi(S^{-1})\pi(T)\pi(T^{-1})\pi(U)\pi(U^{-1})).$$

When $d_\pi = 1$,

$$\pi(S)\pi(S^{-1}) = \sum_{s \in S} \pi(s) \sum_{s \in S} \overline{\pi(s)} = |\pi(S)|^2,$$

which is a nonnegative real number, and $\pi(S)\pi(S^{-1}) = |S|^2$ if π is the trivial representation. Thus,

$$\begin{aligned}&|G|\,|S|\,|T|\,|U| \geq (|S|\,|T|\,|U|)^2 + \\ &\sum_{\pi\colon\, d_\pi > 1} d_\pi Tr(\pi(S)\pi(S^{-1})\pi(T)\pi(T^{-1})\pi(U)\pi(U^{-1})) \\ &= (|S|\,|T|\,|U|)^2 + \sum_{\pi\colon\, d_\pi > 1} d_\pi Tr(\pi(S^{-1})\pi(T)\pi(T^{-1})\pi(U)\pi(U^{-1})\pi(S)).\end{aligned}$$

By the Cauchy–Schwarz inequality,

$$|G|\,|S|\,|T|\,|U| \geq (|S|\,|T|\,|U|)^2 - \sum_{\pi\colon\, d_\pi > 1} d_\pi \|\pi(S^{-1}T)\| \cdot \|\pi(T^{-1}U)\| \cdot \|\pi(U^{-1}S)\|,$$

where $\|\cdot\|$ denotes the Frobenius norm of a matrix (i.e., $\|M\|^2 = Tr(M\overline{M}^\top)$).

Fourier inversion implies a nonabelian version of Parseval's identity, which states that for any function $f\colon G \to \mathbb{C}$,

$$\sum_{g\in G} |f(g)|^2 = \frac{1}{|G|} \sum_{\pi \in \mathrm{Irr}(G)} d_\pi \left\| \sum_{g \in G} f(g)\pi(g) \right\|^2.$$

Applying this formula with $f = 1_{S^{-1}} * 1_T$, which is equal to the indicator function of $S^{-1}T$ by the triple product property, we obtain

$$|S|\,|T|\,|G| = \sum_{\pi \in \mathrm{Irr}(G)} d_\pi \|\pi(S^{-1}T)\|^2,$$

and thus for each $\pi \in \mathrm{Irr}(G)$ with $d_\pi > 1$,

$$\|\pi(S^{-1}T)\| \leq \sqrt{|S|\,|T|\,|G|/n(G)}.$$

Using this bound and the Cauchy–Schwarz inequality, we find that

$$\begin{aligned}|G|\,|S|\,|T|\,|U| \geq\ & (|S|\,|T|\,|U|)^2 - \\ & \sqrt{|S|\,|T|\,|G|/n(G)} \sum_{\pi\colon\, d_\pi > 1} d_\pi \|\pi(T^{-1}U)\| \cdot \|\pi(U^{-1}S)\| \\ & \geq (|S|\,|T|\,|U|)^2 \\ & - \sqrt{|S|\,|T|\,|G|/n(G)} \sqrt{\sum_{\pi\colon\, d_\pi > 1} d_\pi \|\pi(T^{-1}U)\|^2} \sqrt{\sum_{\pi\colon\, d_\pi > 1} d_\pi \|\pi(U^{-1}S)\|^2},\end{aligned}$$

and so by Parseval's identity,

$$\begin{aligned}|G|\,|S|\,|T|\,|U| &\geq (|S|\,|T|\,|U|)^2 - \sqrt{|S|\,|T|\,|G|/n(G)}\sqrt{|G|\,|T|\,|U|}\sqrt{|G|\,|S|\,|U|} \\ &= (|S|\,|T|\,|U|)^2 - |S|\,|T|\,|U|\,|G|^{3/2}/n(G)^{1/2}.\end{aligned}$$

We conclude that

$$|S|\,|T|\,|U| \leq \frac{|G|^{3/2}}{n(G)^{1/2}} + |G|,$$

as desired. ■

We immediately obtain the following corollary:

Corollary 4.2.3. *No sequence $G_1, G_2, \dots$ of finite groups satisfying $n(G_i) \geq \Omega(|G_i|^\delta)$ with $\delta > 0$ can meet the packing bound.*

Corollary 4.2.4. *There exists a constant $\varepsilon > 0$ such that no triple product property construction in a group of Lie type can yield an upper bound on ω better than $2 + \varepsilon$.*

This corollary is more subtle than the previous one, since it does not simply amount to a failure to meeting the packing bound.

Proof. First, we deal with the case of groups of Lie type of bounded rank. Such groups G satisfy $n(G) \geq \Omega(|G|^\delta)$ for some constant $\delta > 0$, as one can check from the bounds given in [LS74], and this condition suffices by Corollary 4.2.3.

Now let G be a group of Lie type of rank r and dimension d over $\mathbb{F}_q$. Then $|G| = \Theta(q^d)$ (see, for example, [MT11, Table 24.1]), and the lower bound $n(G) \geq \Omega(q^r)$ holds by [LS74]. Hence by Theorem 4.2.2,

$$|S|\,|T|\,|U| \leq |G|^{3/2}/\sqrt{n(G)} + |G| = O(q^{3d/2-r/2}).$$

By [FG12, Theorem 1.1], there are $O(q^r)$ conjugacy classes in G. Let $d_1, \dots, d_m$ be the dimensions of the irreducible representations of G, where m is the number of conjugacy classes of G. We have $\sum_i d_i^2 = |G| = \Theta(q^d)$, and

$$\frac{\sum_i d_i^\omega}{m} \geq \left(\frac{\sum_i d_i^2}{m}\right)^{\omega/2}$$

since $x \mapsto x^{\omega/2}$ is a convex function. Hence $\sum_i d_i^\omega \geq \Omega(q^{r+\omega(d-r)/2})$, and therefore Theorem 3.2.2 cannot yield an upper bound on ω better than

$$\omega \leq 3\left(\frac{r + \log_q C}{r}\right)$$

for some absolute constant $C > 0$. If r is large enough, then this bound cannot approach 2, and the case of bounded r was dealt with above. ■

Another consequence of Theorem 4.2.2 is a slightly sharper estimate for how close $|S|\,|T|\,|U|$ can come to $|G|^{3/2}$ when S, T, and U satisfy the triple product property in G. It follows from [CU03, Lemma 3.1] that $|S|\,|T|\,|U| < |G|^{3/2}$, but this inequality does not rule out the possibility that $|S|$, $|T|$, and $|U|$ might be a large as $\lfloor |G|^{1/2} - 1 \rfloor$. The following corollary shows that this cannot happen when $|G|$ is sufficiently large.

Corollary 4.2.5. *If subsets S, T, and U satisfy the triple product property in a finite group G, then $|S|\,|T|\,|U| \leq |G|^{3/2}/\sqrt{2} + |G|$.*

Proof. If G is abelian, then $|S|\,|T|\,|U| \leq |G|$ by [CU03, Lemma 3.1]. Othewise $n(G) \geq 2$ and the conclusion follows from Theorem 4.2.2. ■

4.2.2 A barrier for subgroups that are not self-normalizing

In contrast to the previous barrier, which follows from properties of the containing group, we now give a barrier in terms of the three subsets used in a triple product property construction. It will apply only to the case of three subgroups, as opposed to arbitrary subsets.

For $X \subseteq G$, let $N(X) = \{g \in G : gXg^{-1} = X\}$ denote the normalizer of X in G, and let $Z(G) = \{g \in G : gh = hg \text{ for all } h \in G\}$ denote the center of G.

Theorem 4.2.6. *Suppose that subgroups H_1, H_2, and H_3 satisfy the triple product property in a finite group G, and let $s_i = |N(H_i)|/|H_i|$. Then*

$$|H_1|\,|H_2|\,|H_3| \leq \frac{|G|^{3/2}}{(s_1 s_2 s_3)^{1/4}}.$$

Proof. The main observation in this proof is that $|H_1|\,|N(H_1) \cap H_2|\,|H_3| \leq |G|$ (and the analogous inequality for any permutation of H_1, H_2, and H_3). To prove this inequality, we will show that the map

$$(h_1, h_2, h_3) \mapsto h_1 h_2 h_3$$

is injective on $H_1 \times (N(H_1) \cap H_2) \times H_3$. If not, then there exist $(h_1, h_2, h_3) \neq (h_1', h_2', h_3')$ for which

$$h_1 h_2 h_3 = h_1' h_2' h_3',$$

which implies that

$$h_2'^{-1} h_1'^{-1} h_1 h_2 h_3 h_3'^{-1} = 1.$$

However, $h_2'^{-1}(h_1'^{-1}h_1)h_2'$ is another element $h_1'' \in H_1$ (not equal to 1 if $h_1' \neq h_1$), since h_2' is in the normalizer of H_1. We thus have $h_1''(h_2'^{-1}h_2)(h_3 h_3'^{-1}) = 1$ with not all three factors equal to 1, which contradicts the triple product property for H_1, H_2, and H_3.

Now this inequality implies that

$$|G| \geq |H_1|\,|N(H_1) \cap H_2|\,|H_3| = |H_1| \frac{|N(H_1)|\,|H_2|}{|N(H_1)H_2|} |H_3| \geq |H_1| \frac{|N(H_1)|\,|H_2|}{|G|} |H_3|.$$

The inequality in the theorem statement follows by repeating this argument with H_2, H_3 and then H_3, H_1 in place of H_1 and H_2 and taking the product. ■

The following corollary shows that triple product property constructions using subgroups of groups G satisfying $|Z(G)| = \Omega(|G|^\delta)$ with $\delta > 0$ cannot meet the packing bound. For example, this shows that triples of subgroups in $GL(n, q)$ with fixed n cannot meet the packing bound.[2]

Corollary 4.2.7. *If subgroups H_1, H_2, and H_3 satisfy the triple product property in a finite group G, then*

$$|H_1|\,|H_2|\,|H_3| \leq \frac{|G|^{3/2}}{|Z(G)|^{1/2}}.$$

[2] More generally, arbitrary subsets cannot meet the packing bound, because intersecting random translates of the subsets with $SL(n, q)$ would give subsets of $SL(n, q)$ meeting the packing bound in expectation, and we have seen that this is impossible since $SL(n, q)$ a group of Lie type of bounded rank when n is fixed.

Proof. Because $H_1 \cap Z(G)$, $H_2 \cap Z(G)$, and $H_3 \cap Z(G)$ satisfy the triple product property in the abelian group $Z(G)$,

$$|Z(G)| \geq |H_1 \cap Z(G)|\,|H_2 \cap Z(G)|\,|H_3 \cap Z(G)|$$

by [CU03, Lemma 3.1]. Combining this inequality with $Z(G) \subseteq N(H_i)$ shows that

$$\begin{aligned} |Z(G)| &\geq |H_1 \cap Z(G)|\,|H_2 \cap Z(G)|\,|H_3 \cap Z(G)| \\ &= |H_1|\,|H_2|\,|H_3|\,|Z(G)|^3/(|H_1Z(G)|\,|H_2Z(G)|\,|H_3Z(G)|) \\ &\geq |H_1|\,|H_2|\,|H_3|\,|Z(G)|^3/(|H_1N(H_1)|\,|H_2N(H_2)|\,|H_3N(H_3)|) \\ &= |H_1|\,|H_2|\,|H_3|\,|Z(G)|^3/(|N(H_1)|\,|N(H_2)|\,|N(H_3)|), \end{aligned}$$

and therefore $|N(H_1)|\,|N(H_2)|\,|N(H_3)|/(|H_1|\,|H_2|\,|H_3|) \geq |Z(G)|^2$. The conclusion now follows by Theorem 4.2.6. ■

4.3 Constructions in Lie groups

In this section, we study triple product property constructions in Lie groups (i.e., groups that are also smooth manifolds). All Lie groups will be assumed to be positive-dimensional.

Definition 4.3.1. The *Lie exponent* $\omega(G)$ of a Lie group G of rank $r(G)$ is the infimum of the quantity

$$\frac{r(G)}{(\dim M_1 + \dim M_2 + \dim M_3)/3 - (\dim G - r(G))/2}$$

over all submanifolds M_1, M_2, and M_3 of G satisfying the triple product property and $(\dim M_1 + \dim M_2 + \dim M_3)/3 > (\dim G - r(G))/2$. (Recall that the infimum of the empty set is $+\infty$.) The Lie exponent of a family of groups is the infimum of $\omega(G)$ over G in the family.

In this definition, the rank $r(G)$ is the real dimension of a Cartan subalgebra of the Lie algebra.[3] We primarily have in mind semisimple Lie groups, or more generally reductive groups, and it is unclear how relevant the Lie exponent is for other groups. Note that if G is abelian, then $r(G) = \dim G$.

Definition 4.3.1 is motivated by the following analogy with the finite field setting. The finite groups of Lie type fall into families of Chevalley groups defined over $\mathbb{F}_q$ as q varies, with the families corresponding to the classification of simple Lie groups (as well as some complications such as twisting). For example, $SL(n, q)$ is analogous to $SL(n, \mathbb{R})$ or $SL(n, \mathbb{C})$. Suppose we have triple product property constructions with subsets of sizes $q^{m_1+o(1)}$, $q^{m_2+o(1)}$, and $q^{m_3+o(1)}$ in such a family of simple groups G_q as $q \to \infty$. This is a finite analogue of having submanifolds of dimensions m_1, m_2, and m_3. It follows from [LMT13, Theorem 1.3] that the largest irreducible representation of G_q has dimension $q^{(d-r)/2+o(1)}$ as $q \to \infty$, where d and r are the dimension and rank of the corresponding Lie group.[4] If the dimensions of the irreducible representations of G_q

[3] Note that this differs from the usual convention for complex Lie groups of using the complex dimension. For example, this is why Table 4.1 shows that $r(GL(n, \mathbb{C})) = 2n$.

[4] To deduce this result from [LMT13, Theorem 1.3], note that the Steinberg representation has dimension $q^{(d-r)/2}$. See also [LMT13, Theorems 5.1–5.3] for some classical groups that are not quite simple.

are $d_1, \dots, d_k$, then by Theorem 3.2.2,

$$\begin{aligned} q^{(m_1+m_2+m_3+o(1))\omega/3} &\le \sum_i d_i^\omega \\ &\le \sum_i d_i^2 \max_j d_j^{\omega-2} \\ &= |G| \max_j d_j^{\omega-2} \\ &= q^{d+(\omega-2)(d-r)/2+o(1)}, \end{aligned}$$

and taking the limit as $q \to \infty$ shows that

$$\omega \le \frac{r}{(m_1+m_2+m_3)/3 - (d-r)/2}$$

if the denominator is positive. In other words, Definition 4.3.1 is exactly the bound on ω one would get if the construction had an analogue in the corresponding finite groups of Lie type. We note that we know of no general reason why such an analogue should exist; indeed, we do not know of any finite analogues of the Lie group constructions given later in this section. We pose the following question:

Question 4.3.2. Is it true that for every Lie group G, the exponent of matrix multiplication is at most $\omega(G)$?

By Proposition 4.3.3 below, the answer must be yes if $\omega = 2$. A direct proof would be of considerable interest, with semisimple Lie groups being the most plausible case for a proof. Even without such a proof, we view $\omega(G)$ as a model for what groups can do in the continuous setting, which allows for geometric constructions that may not work over finite fields.

Proposition 4.3.3. *Every Lie group G has $\omega(G) > 2$.*

Proof. If M_1, M_2, and M_3 satisfy the triple product property in G, then the map $(m_1, m_2) \mapsto m_1^{-1}m_2$ from $M_1 \times M_2$ to G is injective, and so $\dim M_1 + \dim M_2 \le \dim G$. Similarly, $\dim M_2 + \dim M_3 \le \dim G$ and $\dim M_1 + \dim M_3 \le \dim G$. Averaging these inequalities shows that $(\dim M_1 + \dim M_2 + \dim M_3)/3 \le (\dim G)/2$, and therefore the bound we obtain for $\omega(G)$ is at least

$$\frac{r(G)}{(\dim G)/2 - (\dim G - r(G))/2} = 2.$$

Equality could hold only if $\dim M_i = (\dim G)/2$ for all i. In that case, every continuous, injective function from $M_1 \times M_2$ to G must be open by invariance of domain, and therefore has an open image. In particular, for $m_1' \in M_1$ and $m_2' \in M_2$ consider the map sending $(m_1, m_2) \in M_1 \times M_2$ to $m_1' m_1^{-1} m_2 m_2'^{-1}$. Its image contains a neighborhood of 1, but by the triple product property it intersects the quotient set $Q(M_3) = M_3 M_3^{-1}$ only at 1, which is impossible since $Q(M_3)$ contains a submanifold $M_3 m_3^{-1}$ (for fixed $m_3 \in M_3$) of dimension $(\dim G)/2$ that contains 1. ■

Note that in the notation of Definition 4.3.1, if $\dim M_1 + \dim M_2 + \dim M_3 \le \dim G$, then they cannot prove any better upper bound for $\omega(G)$ than 3. This conclusion is immediate if $r(G) = \dim G$, and one can check that it follows from $r(G) \le \dim G$. In fact, the best upper bound we know on the Lie exponent is 3, which holds for abelian groups:

Proposition 4.3.4. *If G is abelian, then $\omega(G) = 3$.*

Proof. Let $H_1 = G$ and $H_2 = H_3 = \{1\}$. Then H_1, H_2, and H_3 satisfy the triple product property in G. Since $r(G) = \dim G$ and $\dim H_1 + \dim H_2 + \dim H_3 = \dim G$, it follows that $\omega(G) \leq 3$.

For the other direction, note that if G is abelian, then the product map $M_1 \times M_2 \times M_3 \to G$ must be injective or else the triple product property fails. If the map is injective, then $\dim M_1 + \dim M_2 + \dim M_3 \leq \dim G$ and hence $\omega(G) \geq 3$. ∎

There is also a Lie analogue of the packing bound from Definition 3.2.8.

Definition 4.3.5. We say a sequence $G_1, G_2, \dots$ of Lie groups *meets the packing bound* if there exist submanifolds $M_{1,i}$, $M_{2,i}$, and $M_{3,i}$ of G_i satisfying the triple product property such that

$$\lim_{i\to\infty} \frac{\dim G_i}{(\dim M_{1,i} + \dim M_{2,i} + \dim M_{2,i})/3} = 2.$$

Proposition 4.3.6. *If Lie groups $G_1, G_2, \dots$ have $\lim_{i\to\infty} \omega(G_i) = 2$, then they achieve the packing bound.*

Proof. It suffices to show that for M_1, M_2, and M_3 satisfying the triple product property in a group G with $(\dim M_1 + \dim M_2 + \dim M_3)/3 > (\dim G - r(G))/2$,

$$\frac{r(G)}{(\dim M_1 + \dim M_2 + \dim M_3)/3 - (\dim G - r(G))/2} \geq \frac{\dim G}{(\dim M_1 + \dim M_2 + \dim M_3)/3}.$$

This assertion follows from the inequality $(\dim M_1 + \dim M_2 + \dim M_3)/3 \leq (\dim G)/2$ used in the proof of Proposition 4.3.3. ∎

It was shown in [CU03, Theorem 6.1] that the Lie groups $SL(n, \mathbb{R})$ meet the packing bound, by taking M_1, M_2, and M_3 to be the groups of upper unitriangular, lower unitriangular, and orthogonal matrices. In this construction, the group is an algebraic group over $\mathbb{R}$, as are the subgroups M_i. In particular, they are all linear algebraic groups over $\mathbb{R}$, i.e., subgroups of $GL(n, \mathbb{R})$ defined by polynomial equations. Algebraic groups are a little more general than linear algebraic groups; they are to algebraic varieties as Lie groups are to manifolds.

However, algebraic varieties over $\mathbb{C}$ (or any algebraically closed field) cannot help:

Theorem 4.3.7. *Let G be an algebraic group over an algebraically closed field, and let V_1, V_2, and V_3 be subvarieties of G that satisfy the triple product property. Then*

$$\dim V_1 + \dim V_2 + \dim V_3 \leq \dim G.$$

As a consequence, subvarieties of an algebraic groups over $\mathbb{C}$ cannot be used to meet the packing bound or obtain a better Lie exponent than 3.

Proof. Let v_i' be any element of V_i, and define $\varphi\colon V_1 \times V_2 \times V_3 \to G$ by

$$\varphi(v_1, v_2, v_3) = v_1 v_1'^{-1} v_2 v_2'^{-1} v_3 v_3'^{-1}.$$

By the triple product property, the only solution of $\varphi(v_1, v_2, v_3) = 1$ is (v_1', v_2', v_3'), and so the solution set is zero-dimensional since we are working over an algebraically closed field. However, the fiber dimension theorem [Har92, Theorem 17.24] says the dimension of the solution set must be at least $\dim V_1 + \dim V_2 + \dim V_3 - \dim G$, which yields the desired inequality. ∎

The intuitive difference between $\mathbb{R}$ and $\mathbb{C}$ here is that a variety can have fewer points over $\mathbb{R}$ than one might expect by counting degrees of freedom. For example, the equation $x_1^2+\cdots+x_n^2 = 0$ defines an $(n-1)$-dimensional variety, which has plenty of points over $\mathbb{C}$, but over $\mathbb{R}$ it consists of just a single point. Fields that are not algebraically closed may lead to an anomalously low number of solutions, and we cannot conclude that a variety is zero-dimensional just because it has only one real point. What Theorem 4.3.7 indicates is that to obtain strong examples, we must either use constructions that are not defined by polynomial equations, or choose equations that have fewer solutions over $\mathbb{R}$ than they do over $\mathbb{C}$. (Note that there are many possibilities that are not defined by polynomial equations. For example, constructions that use complex conjugation generally do not define complex subvarieties.)

Theorem 4.3.7 does rule out one superficially attractive possibility, namely obtaining Lie exponent 2 via subvarieties of algebraic groups over $\overline{\mathbb{F}}_q$ and then transitioning to finite groups by using finite subfields of $\overline{\mathbb{F}}_q$ with sizes tending to infinity.

We now give several new constructions over $\mathbb{R}$ with parameters that improve upon the previously known construction from [CU03]. Our constructions make use of the following observation, which relaxes the triple product property for subgroups.

Lemma 4.3.8. *Suppose that H_1, H_2, and H_3 are Lie subgroups of a Lie group G and K is a compact subgroup of G such that the equation $h_1h_2h_3 = 1$ with $h_i \in H_i$ implies that $h_1, h_2, h_3 \in K$. Then the Lie exponent of G is at most*

$$\frac{r(G)}{(\dim H_1 + \dim H_2 + \dim H_3 - 2\dim K)/3 - (\dim G - r(G))/2}.$$

We will refer to this situation as the *K-triple product property*. More generally, the same holds for submanifolds M_i such that $q_1q_2q_3 = 1$ with $q_i \in Q(M_i)$ implies $q_1, q_2, q_3 \in K$, but we will need it only for subgroups.

Proof. By the slice theorem [Aud04, Theorem I.2.1], there exist submanifolds H_i' of H_i such that $\dim H_i' = \dim H_i - \dim K$ and no two elements $h, h' \in H_i'$ satisfy $hh'^{-1} \in K$ unless $h = h'$. Then the submanifolds H_1, H_2', and H_3' of G satisfy the triple property property and yield the asserted bound. (Note that the first submanifold is H_1, not H_1'.) ■

4.3.1 Asymptotic Lie exponent 3

As mentioned above and shown in [CU03, Theorem 6.1], the upper unitriangular, lower unitriangular, and special orthogonal groups satisfy the triple product property in $SL(n,\mathbb{R})$. Since these subgroups have dimension $n(n-1)/2$ inside a group of dimension n^2-1, the groups $SL(n,\mathbb{R})$ meet the packing bound.

However, the rank of $SL(n,\mathbb{R})$ is $n-1$, and so this example does not yield any Lie exponent bound (the denominator in Definition 4.3.1 would vanish). In this section we modify the construction to get a bound on the Lie exponent approaching 3 for powers of $SL(n,\mathbb{R})$.

Theorem 4.3.9. *For $m > 1$ and $n > 1$, the Lie exponent of $SL(n,\mathbb{R})^m$ is at most*

$$\frac{3m(n-1)}{m(n-1)-n}.$$

In the proof we will denote the i, j entry of a matrix M by $M_{i,j}$. To avoid ambiguity we will use superscripts to index sequences of matrices, with parentheses around the superscripts to distinguish them from exponents. Recall also that a unitriangular matrix is a triangular matrix with diagonal entries equal to 1.

Proof. Let $H_1 = SO(n, \mathbb{R})^m$, let

$$H_2 = \{(A^{(1)}, \dots, A^{(m)}) \in SL(n, \mathbb{R})^m : \text{each } A^{(i)} \text{ is upper unitriangular}\},$$

and let H_3 equal

$$\left\{(B^{(1)}, \dots, B^{(m)}) \in SL(n, \mathbb{R})^m : \text{each } B^{(i)} \text{ is lower triangular and } \prod_{i=1}^{m} B^{(i)}_{j,j} = 1 \text{ for all } j\right\},$$

These subgroups have dimensions $mn(n-1)/2$, $mn(n-1)/2$, and $m(n(n+1)/2-1)-n$, respectively. We claim that they satisfy the K-triple product property in $SL(n.\mathbb{R})^m$, where

$$K = \{(C^{(1)}, \dots, C^{(m)}) \in SL(n, \mathbb{R})^m : \text{each } C^{(i)} \text{ is diagonal with } \pm 1 \text{ entries}\}.$$

Since $\dim SL(n, \mathbb{R})^m = m(n^2-1)$ and the rank of $SL(n, \mathbb{R})^m$ is $m(n-1)$, while $\dim K = 0$, the claimed bound on the Lie exponent will follow by Lemma 4.3.8.

Let $M = (M^{(1)}, \dots, M^{(m)}) \in H_1$, $A = (A^{(1)}, \dots, A^{(m)}) \in H_2$, $B = (B^{(1)}, \dots, B^{(m)}) \in H_3$. We will show that if $MA = B$, then for each i, the matrix $M^{(i)}$ is diagonal with ± 1 diagonal entries, in which case the same follows for $A^{(i)}$ and $B^{(i)}$. We will prove by induction on j that $M^{(i)} e_j = \pm e_j$ for all i, where $e_1, \dots, e_n$ are the standard basis vectors.

Let the jth column of $A^{(i)}$ be $A^{(i)}_j$. For any i, $A^{(i)}_1 = e_1$, and so $M^{(i)} e_1 = B^{(i)}_1$. Since $M^{(i)}$ is orthogonal, $M^{(i)} e_1$ and thus also $B^{(i)}_1$ must be unit vectors. In particular, $|B^{(i)}_{1,1}| \le 1$ for all i. But since $\prod_{i=1}^m B^{(i)}_{1,1} = 1$ this forces $B^{(i)}_{1,1} = \pm 1$, which proves the claim for $j = 1$.

Now suppose $M^{(i)} e_j = \pm e_j$ for all $j < k$ and all i. For any i, $A^{(i)}_k = e_k + \sum_{j<k} A^{(i)}_{j,k} e_j$ and $B^{(i)}_k = B^{(i)}_{k,k} e_k + \sum_{j>k} B^{(i)}_{j,k} e_j$. From the induction hypothesis we deduce that

$$\begin{aligned} M^{(i)} e_k &= M^{(i)}\big(A^{(i)}_k - \sum_{j<k} A^{(i)}_{j,k} e_j\big) \\ &= B^{(i)}_{k,k} e_k + \sum_{j>k} B^{(i)}_{j,k} e_j - \sum_{j<k} A^{(i)}_{j,k} (\pm e_j). \end{aligned}$$

Since $M^{(i)}$ is orthogonal, $M^{(i)} e_k$ is a unit vector. Because $\prod_i B^{(i)}_{k,k} = 1$, it follows as above that $B^{(i)}_{k,k} = \pm 1$ for all i. Hence $M^{(i)} e_k = \pm e_k$ for all i, which proves the claim. ∎

4.3.2 Conjugates of rotation groups

Theorem 4.3.10. *There are three conjugates of* $\mathrm{O}(n, \mathbb{R})$ *inside of* $GL(n, \mathbb{R})$ *satisfying* K*-triple product property, where* K *is the subgroup of diagonal matrices with* ± 1 *entries on the diagonal.*

This construction meets the packing bound. In particular, it evades the normalizer barrier since the normalizer of $\mathrm{O}(n,\mathbb{R})$ in $GL(n,\mathbb{R})$ is $\mathbb{R}^\times \cdot \mathrm{O}(n,\mathbb{R})$. However, it does not prove a bound for $\omega(GL(n,\mathbb{R}))$, because each of these subgroups has dimension $n(n-1)/2$, which equals $(\dim GL(n,\mathbb{R}) - n)/2$, and $r(GL(n,\mathbb{R})) = n$. Note that the center of $GL(n,\mathbb{R})$ plays no role, and we could just as well have stated the theorem for conjugates of $SO(n,\mathbb{R})$ inside $SL(n,\mathbb{R})$. We use the slightly more general formulation since it is convenient to allow determinant -1 in the proof by induction.

Proof. Let $G = GL(n,\mathbb{R})$ and $H = \mathrm{O}(n,\mathbb{R})$. To specify the conjugates of H, we take $D_1 = diag(x_1,\dots,x_n)$ with $x_1 > x_2 > \cdots > x_n > 0$ and $D_2 = diag(y_1,\dots,y_n)$ with $0 < y_1 < y_2 < \cdots < y_n$. Then we will show that H, $H_1 \coloneqq D_1 H D_1^{-1}$, and $H_2 \coloneqq D_2 H D_2^{-1}$ satisfy the K-triple product property in G, where K is the group of diagonal ± 1 matrices.

In particular, we will show that for every $h_1 \in H_1$ and $h_2 \in H_2$, if $h_1^\top h_1 = h_2^\top h_2$, then $h_1, h_2 \in K$. The conclusion then follows, since if $h h_1 h_2^{-1} = I$ with $h \in H$, then $h_1 h_2^{-1} \in H$, meaning $(h_1 h_2^{-1})^\top (h_1 h_2^{-1}) = I$, and hence $h_1^\top h_1 = h_2^\top h_2$.

Suppose that $h_1 = D_1 M_1 D_1^{-1}$ and $h_2 = D_2 M_2 D_2^{-1}$, where $M_1, M_2 \in H$, and consider

$$h_1^\top h_1 = (D_1^{-1} M_1^\top D_1)(D_1 M_1 D_1^{-1}).$$

If $(a_1, a_2, \dots, a_n)$ is the first column of M_1, then the first column of $D_1 M_1 D_1^{-1}$ is

$$(a_1, x_2 x_1^{-1} a_2, \dots, x_n x_1^{-1} a_n)$$

as is the first row of $D_1^{-1} M_1^\top D_1$, so the upper-left entry of their product $h_1^\top h_1$ is

$$a_1^2 + (x_2/x_1)^2 a_2^2 + (x_3/x_1)^2 a_3^2 + \cdots + (x_n/x_1)^2 a_n^2.$$

Now, since $a_1^2 + a_2^2 + \cdots + a_n^2 = 1$, we can substitute for a_1^2 to obtain

$$1 + ((x_2/x_1)^2 - 1)a_2^2 + ((x_3/x_1)^2 - 1)a_3^2 + \cdots + ((x_n/x_1)^2 - 1)a_n^2.$$

Because $x_i > x_1$ for all $i > 1$, this quantity is at most 1, with equality exactly when $a_2^2 = a_3^2 = \cdots = a_n^2 = 0$. By an identical argument, if $(b_1, b_2, \dots, b_n)$ is the first column of M_2, then the upper-left entry of $h_2^\top h_2$ is

$$1 + ((y_2/y_1)^2 - 1)b_2^2 + ((y_3/y_1)^2 - 1)b_3^2 + \cdots + ((y_n/y_1)^2 - 1)b_n^2,$$

which is at least 1, with equality exactly when $b_2^2 = b_3^2 = \cdots = b_n^2 = 0$. So if $h_1^\top h_1 = h_2^\top h_2$, then in particular their upper-left entries are equal, and we conclude that $a_1^2 = b_1^2 = 1$, while $a_i = b_i = 0$ for all $i > 1$.

Now h_1 has the form

$$\left(\begin{array}{c|ccc} \pm 1 & 0 & \dots & 0 \\ \hline 0 & & & \\ \vdots & & h_1' & \\ 0 & & & \end{array}\right),$$

Table 4.1: Parameters for the real, complex, and quaternionic versions of Theorem 4.3.10.

(skew) field	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{H}$
$\dim G$	n^2	$2n^2$	$4n^2$
$r(G)$	n	$2n$	$4n$
$\dim H$	$n(n-1)/2$	n^2	$n(2n+1)$
$\dim K$	0	n	$3n$
Meets packing bound as $n \to \infty$	yes	yes	yes
Lie exponent upper bound	∞	6	4

where h_1' is an element of $D_1' \, \mathrm{O}(n-1, \mathbb{R}) D_1'^{-1}$ with $D_1' = diag(x_2, \ldots, x_n)$, and h_2 has the form

$$\left(\begin{array}{c|ccc} \pm 1 & 0 & \ldots & 0 \\ \hline 0 & & & \\ \vdots & & h_2' & \\ 0 & & & \end{array}\right),$$

where h_2' is an element of $D_2' \, \mathrm{O}(n-1, \mathbb{R}) D_2'^{-1}$ with $D_2' = diag(y_2, \ldots y_n)$. Finally, since $h_1 h_2^{-1} h = 1$ we find h also has the same block-diagonal form, and so $h_1'(h_2')^{-1} \in \mathrm{O}(n-1, \mathbb{R})$, and then we are done by induction on n. ■

Remark 4.3.11. This theorem holds when we replace G with $GL(n, \mathbb{C})$ (resp., $GL(n, \mathbb{H})$), H with $\mathrm{U}(n, \mathbb{C})$ (resp., $Sp(n)$), and K with the group of diagonal matrices with unit complex (resp., quaternionic) numbers on the diagonal. This follows from a similar argument, where one replaces transpose with conjugate transpose and uses the positivity of the complex/quaternionic norm. The corresponding dimensions are shown in Table 4.1.

4.4 Open problems

The most important challenge highlighted by this chapter is to find a construction proving that the Lie exponent of a family of Lie groups approaches 2, or to prove that such a construction cannot exist. Many questions can be asked along the way, including whether there is a Lie group with Lie exponent less than 3, and whether the Lie exponent of $SL(n, \mathbb{R})$ is even finite.

It follows from [Saw18] that a triple product property construction inside $SL(2, q)^m$ can't give $\omega = 2$ for fixed q and growing m, and Theorem 4.2.2 shows that $SL(n, q)^m$ can't give $\omega = 2$ for fixed m and growing q. The proofs of these two facts are quite different: one uses the polynomial method, and the other is Fourier analytic. Is there a common generalization of these two facts that would rule out obtaining $\omega = 2$ with m and q both growing?

Together with the fact that abelian groups cannot yield exponent less than 3, Theorem 4.2.2 implies that the alternating groups are the only simple groups left that could yield $\omega = 2$ via a triple product property construction. The representation-theoretic argument fails in this case, since A_n has an irreducible representation of dimension $n-1$ but $|A_n| = n!/2$. Can an alternate argument rule out these groups?

4.5 Comparing barriers: quasirandomness vs. slice rank

In the previous sections we saw two barriers to the group–theoretic approach: the slice rank barrier and the quasirandomness barrier. The slice-rank barrier shows that any tensor with low slice rank cannot be used successfully in the asympotic sum inequality, and the quasirandomness barrier rules out using quasirandom hypergraphs. In this section we explore interactions between these two barriers.

A multiplicative 3-matching in a group G is a triple of sets $\{a_i\}, \{b_i\}, \{c_i\} \subseteq G$ such that $a_i b_j c_k = 1$ if and only if $i = j = k$. Here we record the fact that $\mathrm{PSL}(2, p)$ has no multiplicative 3-matching of size greater than $O(p^{8/3})$, yet the slice rank of its group algebra's multiplication tensor is at least $\Omega(p^3)$ over any field. This gives a negative answer to a conjecture of Petrov.

4.5.1 Introduction

A *multiplicative 3-matching* in a finite group G, hereon abbreviated to a 3-matching, is a triple of subsets $\{a_i\}_{i=1}^m, \{b_i\}_{i=1}^m, \{c_i\}_{i=1}^m$ of G such that $a_i b_j c_k = 1 \iff i = j = k$. Let $M(G)$ denote the largest size[5] of a 3-matching in G. This quantity is of interest in additive combinatorics, as finite groups provide a model setting for understanding 3-term arithmetic-progression-free sets in the integers. It also has connections to algorithms for fast matrix multiplication [CKSU05].

The polynomial method of Croot, Lev, and Pach [CLP17] and Ellenberg and Gijswijt [EG17], and its formulation in terms of *slice rank* due to Tao [Tao16], is a powerful tool for establishing upper bounds on $M(G)$. Remarkably, it gives an asymptotically tight bound on $M(G)$ in the case of $\mathbb{F}_p^n$ with $p > 2$ a fixed prime. Specifically, it shows that for a certain $c_p < p$ we have $M(\mathbb{F}_p^n) \leq c_p^n$, and it is also known that $M(\mathbb{F}_p^n) \geq c_p^{(1-o(1))n}$ [KSS16]. This raises the following question, previously asked in [Pet16]: is the slice rank bound on $M(G)$ always tight?

We now make these notions formal. We assume throughout that k is an algebraically closed field[6]. Following [TS16, Lemma 1(iv)], the slice rank of a trilinear form $T : U_1 \times U_2 \times U_3 \to k$ is

$$\mathrm{SR}(T) = \max_{\substack{V_i \leq U_i: \\ T(V_1, V_2, V_3) = 0}} \mathrm{Codim}(V_1) + \mathrm{Codim}(V_2) + \mathrm{Codim}(V_3).$$

This can be thought of as an analogue of the "codimension of the kernel" definition of matrix rank for trilinear forms. To use slice rank to prove bounds on $M(G)$, we consider the multiplication tensor $T_{k[G]} : (k^{|G|})^{\times 3} \to k$ of the group algebra $k[G]$, defined as $T_{k[G]} = \sum_{g,h \in G} x_g y_h z_{gh}$. The key fact is that $\mathrm{SR}(T_{k[G]})$ is at least $M(G)$ for any k [Tao16, Lemma 1]; this is an analogue of the fact that the rank of a matrix is at least the size of the largest identity-submatrix it contains. Hence upper bounds on $\mathrm{SR}(T_{k[G]})$ imply upper bounds on $M(G)$. Motivated by this, we make the following definition.

Definition 4.5.1. $\mathrm{SR}(G) = \min_k \mathrm{SR}(T_{k[G]})$.

Here the minimum is taken over all algebraically closed fields; crucially, the characteristic of k can be arbitrary. In the case that $G = \mathbb{F}_p^n$ for example, $\mathrm{SR}(T_{\mathbb{C}[\mathbb{F}_p^n]}) = p^n$ [BCC+17b, Corollary

[5] By size we mean the parameter m.

[6] This is without loss of generality for us, since slice rank can only decrease under field extensions and we will prove a slice rank lower bound.

b.17] but $\mathrm{SR}(T_{\mathbb{F}_p[\mathbb{F}_p^n]}) \leq c_p^n$. This leads to the following conjecture, which is slightly weaker than one appearing in [Pet16]:[7]

Conjecture 4.5.2. $\mathrm{SR}(G) \leq M(G) \cdot |G|^{o(1)}$.

In this work we note that $\mathrm{PSL}(2, p)$ has no large 3-matchings but has high slice rank over any field, so Conjecture 4.5.2 is false. Both of these facts are in large part due to the lower bound of [LS74] on the dimensions of nontrivial irreducible representations of $\mathrm{PSL}(2, p)$. The fact that $\mathrm{PSL}(2, p)$ has no large 3-matching follows almost immediately from Gowers's result on quasirandom groups [Gow08]. We now give a quick proof of this.

Proposition 4.5.3. $M(\mathrm{PSL}(2, p)) \leq O(p^{8/3})$.

Proof. A triple of subsets $A, B, C \subseteq G$ is called product-free if $abc \neq 1$ for all $a \in A, b \in B, c \in C$. If A, B, C is a 3-matching of size m, then there is a product-free triple of sets in G of size $m' := \lfloor m/3 \rfloor$ consisting of $\{a_i\}_{i=1}^{m'}, \{b_i\}_{i=m'+1}^{2m'}, \{c_i\}_{i=2m'+1}^{3m'}$. In [Gow08] it is shown that $\mathrm{PSL}(2, p)$ does not contain product-free subsets larger than $O(p^{8/3})$, so the proposition follows. ■

In the next section we the following slice rank lower bound.

Theorem 4.5.4. $\mathrm{SR}(\mathrm{PSL}(2, p)) \geq \Omega(p^3)$.

4.5.2 Proof of Theorem 4.5.4

If A is a finite dimensional algebra over k, we let $T_A \in A^* \otimes A^* \otimes A$ denote its multiplication tensor. If $e_1, \ldots, e_n$ is a basis of A with dual basis $e_1^*, \ldots, e_n^*$, this tensor is given in coordinates by $\sum_{1 \leq i,j \leq n} e_i^* \otimes e_j^* \otimes (e_i \cdot e_j)$. We can view this as a trilinear form, for instance by linearly mapping $e_i^* \otimes e_j^* \otimes e_k$ to the monomial $x_i y_j z_k$, and define its slice rank as in the introduction. We write $\mathrm{SR}(A)$ for the slice rank of the multiplication tensor of A.

Now we recall some basic facts about slice rank, namely that it is nonincreasing under linear transformations, and that the slice rank of an algebra is nonincreasing under quotients.

Lemma 4.5.5. *[TS16, Lemma 3] Let* $T = \sum c_{ijk} u_i \otimes v_j \otimes w_k \in U \otimes V \otimes W$*, and let* $A \in \mathrm{Hom}(U, U'), B \in \mathrm{Hom}(V, V'), C \in \mathrm{Hom}(W, W')$*. Then* $\mathrm{SR}(\sum c_{ijk} A(u_i) \otimes B(v_j) \otimes C(w_k)) \leq \mathrm{SR}(T)$.

Lemma 4.5.6. *If I is a two-sided ideal of A, then* $\mathrm{SR}(A/I) \leq \mathrm{SR}(A)$.

Proof. Let $\varphi : A \to A/I$ be the quotient map. Let $e_1, \ldots, e_n$ be a basis for A. Since φ is an onto linear map, there exists $S \subseteq [n]$ so that $\{\varphi(e_i)\}_{i \in S}$ is a basis of A/I. Let $P : A \to A/I$ be given by $P(e_i) = \varphi(e_i)$ for $i \in S$, and $P(e_i) = 0$ if $i \notin S$. Similarly define $P' : A^* \to (A/I)^*$ by $P'(e_i^*) = P(e_i)^*$. Applying P' to the first two factors of T_A and P to the third, we obtain $T_{A/I}$. By Lemma 4.5.5 this proves the claim. ■

For an algebra A, we denote by $\mathrm{Irr}(A)$ the set of non-isomorphic irreducible representations of A over k. Recall that in the ordinary (i.e., characteristic 0) representation theory of finite groups, representations are completely reducible, and in particular the group algebra $k[G]$ is isomorphic

[7] The conjecture of [Pet16] asked if $M(G)$ is roughly the sum of codimensions of subspaces multiplying to 0 in $k[G]$, whereas Conjecture 4.5.2 is equivalent to asking if $M(G)$ is roughly the sum of codimensions of subspaces whose product merely vanishes on the coefficient of 1 in $k[G]$.

to a direct sum of matrix algebras. While this is false when the characteristic of the field divides the order of the group, we still have the following.

Definition 4.5.7. The radical of A, denoted $\mathrm{J}(A)$, is the two-sided ideal of all elements of A which act by 0 on all irreducible representations of A.

Lemma 4.5.8. *[EGH+11, Theorem 2.12] Let A be a finite-dimensional algebra. Then*

$$A/\mathrm{J}(A) \cong \bigoplus_{V \in \mathrm{Irr}(A)} \mathrm{End}(V).$$

We will use the following fact, which says that the slice rank of direct sums of matrix multiplication tensors is maximal.

Lemma 4.5.9. *[BCC+17b, Proposition B.6] For any field k, $\mathrm{SR}(\bigoplus_{i=1}^{m} \mathrm{End}(k^{d_i})) = \sum_{i=1}^{m} d_i^2$.*

The proof of Theorem 4.5.4 will go as follows. By Lemma 4.5.8, $k[G]/\mathrm{J}(k[G])$ is a direct sum of matrix algebras, one for each irreducible representation of $k[G]$. So by Lemma 4.5.9, $k[G]/\mathrm{J}(k[G])$ has full slice rank, and by Lemma 4.5.6 this is a lower bound on the slice rank of $k[G]$. So if we can show that there are many sufficiently large irreps of G, we conclude that $\mathrm{SR}(k[G])$ is large. To give an example of when this fails dramatically, when G is any p-group, the only irrep of G when k has characteristic p is the trivial one [S+77, Corollary of Proposition 26], so this argument only says that $\mathrm{SR}(k[G]) \geq 1$.

Proof of Theorem 4.5.4. Let $G = \mathrm{PSL}(2,p)$ and let k be a field of characteristic ℓ. First, if $\ell = 0$ or ℓ is coprime to $|G| = (p-1)p(p+1)/2$, then $k[G]$ is semisimple and so by Lemma 4.5.9 the slice rank of $k[G]$ equals $|G| = \Omega(p^3)$. Next, if $\ell = p$, then the irreps of $SL(2,p)$ are given by the action of G on homogeneous polynomials in two variables of degree up to $p-1$ with coefficients in k [Alp93, p. 15]; since the center of $SL(2,p)$ acts trivially on even degree polynomials, these are also irreps of $\mathrm{PSL}(2,p)$ (in fact, all of them). By Lemma 4.5.8 the dimension of $k[G]/\mathrm{J}(k[G])$ is then $\sum_{i=0}^{(p-1)/2}(2i+1)^2 \geq \Omega(p^3)$, so the claim holds.

So suppose $\ell \neq p$ divides $|G| = (p-1)p(p+1)/2$. By [Alp93, I.3 Theorem 2], the number of irreducible representations of G equals the number of conjugacy classes having order coprime to ℓ. Next we show that there are $\Omega(p)$ such conjugacy classes of G. This follows from the more general bound of [HM22, Theorem 6.1]; here we sketch a proof for the special case of $\mathrm{PSL}(2,p)$. See [FH13, p. 71] for a reference on conjugacy classes of $SL(2,p)$, which we adapt to $\mathrm{PSL}(2,p)$. Most elements in G are either conjugate to an element of the *split torus*, a cyclic subgroup of order $(p-1)/2$, or the *non-split torus*, a cyclic subgroup of order $(p+1)/2$. The number of non-conjugate elements in the split torus is at least $(p-3)/4$, and the number of non-conjugate elements in the non-split torus is at least $(p-5)/4$ (with the exact values depending on $p \bmod 4$). Because the orders of these tori are coprime, all conjugacy classes of elements in at least one of the subgroups have order coprime to ℓ. So there are at least $(p-5)/4$ such conjugacy classes.

Finally, since the minimum dimension of a nontrivial irrep of G is at least $(p-1)/2$ in characteristic $\ell \neq p$ [LS74], we conclude by Lemma 4.5.8 that $\dim k[G]/\mathrm{J}(k[G]) \geq \Omega(p^3)$, and thus by Lemma 4.5.9 $\mathrm{SR}(k[G]) \geq \Omega(p^3)$. ∎

One might wonder if all sufficiently quasirandom groups (groups with no small nontrivial irreps over $\mathbb{C}$) have high slice rank. Here is a conjecture towards this question.

Conjecture 4.5.10. *For a fixed $\varepsilon > 0$, let G be a group of order n that is n^{ε}-quasirandom. Then for all fields k, we have the uniform bound of* $\dim k[G]/\mathtt{J}(k[G]) \geq \Omega(n)$.

Chapter 5

Towards Further Barriers To $\omega = 2$

In Chapter 3, we saw that the best upper bounds on ω obtained since 1987 (with possible exception to [WXXZ23]) can be understood as solutions to the following hypergraph packing problem. Let M_n be the *matrix multiplication hypergraph*, the tripartite 3-uniform hypergraph with parts $\{(i,j) : i,j \in [n]\}$, and where $\{(i,j),(k,l),(m,n)\} \in E(X) \iff j = k, l = m, n = i$. Given an abelian group G, let X_G be the tripartite 3-uniform hypergraph with vertex sets $X_1 \sqcup X_2 \sqcup X_3 = G$, and where $(x_1, x_2, x_3) \in X_1 \times X_2 \times X_3$ is a hyperedge exactly when $x_1 + x_2 + x_3 = 0$. Suppose that X_G contains k disjoint copies of $M_{n,n,n}$ as an induced subhypergraph. Then by the asymptotic sum inequality (Theorem 3.1.5),

$$\omega < \log_n(|G|/k).$$

So in order to prove good upper bounds on ω, we aim to find many large disjoint induced matrix multiplication hypergraphs inside X_G. Equivalently, we are searching for simultaneous triple product property constructions inside of G.

In the current fastest algorithms, G is taken to be a group of bounded exponent, specifically, $\mathbb{Z}_7^n$. At the same time, in the last section we saw how ideas related to the cap-set problem show that groups of bounded exponent cannot yield $\omega = 2$ within this approach. In this section we explore the viability of general abelian groups within this framework. It turns out that this reduces to understanding the case of $\mathbb{Z}_n$. In this case, the induced matching barrier says nothing, since $X_{\mathbb{Z}_n}$ contains an induced matching of size $|G|^{1-o(1)}$, which follows from the existence of large 3AP-free subsets of $\mathbb{Z}_n$. In the absence of this, there is no known barrier to obtaining $\omega = 2$ using cyclic groups within this framework. A conjecture of [CKSU05] would imply that this is the case.

In this section we explore problems in combinatorial geometry and additive combinatorics that could potentially rule out using any abelian group in the approach of [CKSU05]. Before we introduce one such problem, recall the *corners problem* of determining the largest subset of $\Delta_n := \{(x,y,z) \in [n]^3 : x + y + z = n\}$ containing no three points of the form

$$(x, y, z + \delta), (x, y + \delta, z), (x + \delta, y, z)$$

where $\delta \neq 0$. Using a density–increment argument, Ajtai and Szemerédi [AS74] showed an upper bound on this quantity of $o(n^2)$ and a roughly matching lower bound of $n^{2-o(1)}$. More quantitatively, Shkredov showed the upper bound of $n^2/(\log\log n)^c$ where $c \approx 0.0137$, and Green

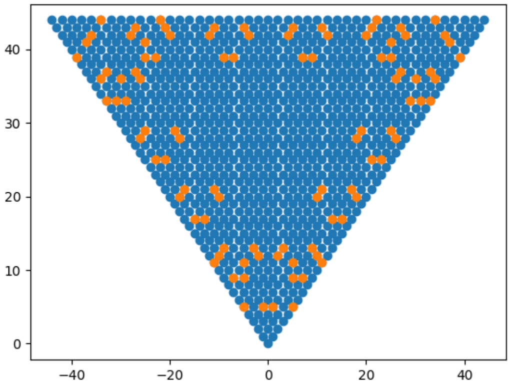

Figure 5.1: The orange points form a skew-corner free subset of Δ_{45} of size 90.

showed the lower bound of $n^2/e^{(c'+o(1))\sqrt{\log n}}$ with $c' \approx 1.822$. One problem we investigate is the following strengthening of the condition of the corners problem, for which we know comparatively very little.

Question 5.0.1. What is the size of the largest $S \subseteq \Delta_n$ which, for any permutation of the coordinates, does not contain three points of the form

$$(x, y, z + \delta), (x, y + \delta, z), (x + \delta, y', z')$$

with $\delta \neq 0$?

In words, S should have the property that for any three points in S forming an equilateral triangle with sides parallel to one of the three "sides" of Δ_n, if any two points belong to this triangle, then S cannot contain any points on the line parallel to these two points and passing through the third must. See Question 5.0.1 for a nontrivial example of such a set.

It follows immediately from the upper bound of the corners problem that this quantity is at most $o(n^2)$. On the other hand, the best lower bound we know is $\Omega(n)$, which is trivially obtained by taking all points with one coordinate fixed.

We show that if the answer to this question is $O(n)$, then one cannot obtain $\omega = 2$ via the aformentioned approach to ω. This would rule out the "two families conjecture" of [CKSU05, Conjecture 4.7]. More precisely, we show that if the answer to this question is $o(n \cdot f(n))$ where $e^{\Omega(\log^* n)} \leq f(n) \leq e^{O(\sqrt{\log n})}$ is a yet-to-be-determined function arising from the quantitative bound in the triangle removal lemma [Zha23, Remark 2.3.4], one cannot obtain $\omega = 2$ via this approach.

More broadly, we show that constructions within the group–theoretic approach of [CU03, CKSU05] imply nontrivial solutions to (possibly nonabelian analogues of) Question 5.0.1. In fact, they yield solutions to significantly more restricted problems; Question 5.0.1 is one of the weakest problems we identify that could resolve the conjecture of [CKSU05] but which we have not been able to rule out. These problems seem intermediate to understanding the best bound on ω provable within the group–theoretic approach.

5.1 Triangle removal and the group-theoretic approach

In [SV$^+$09] an alternative proof of Green's arithmetic removal lemma [Gre05] was given using the directed graph removal lemma of Alon and Shapira [AS03]. Because their proof technique was combinatorial, it had the benefit of generalizing to nonabelian groups. Specifically, they gave a simple proof of the following:

Theorem 5.1.1. *Let G be a finite group of order N. Let $A_1, \dots, A_m, m \geq 2$, be sets of elements of G and let g be an arbitrary element of G. If the equation $x_1 x_2 \cdots x_m = g$ has $o(N^{m-1})$ solutions with $x_i \in A_i$, then there are subsets $A_i' \subseteq A_i$ with $|A_i \setminus A_i'| = o(N)$ such that there is no solution of the equation $x_1 x_2 \cdots x_m = g$ with $x_i \in A'$.*

The best quantiative bounds in this theorem come from the best bounds for the directed cycle removal lemma. The best bounds for this problem, due to Fox [Fox11], in turn imply that if there are at most δN^{m-1} solutions to $x_1 \cdots x_m = g$, one can remove subsets of A_i of size εN and eliminate all solutions, so long as $\varepsilon \leq 1/e^{\Omega(\log^*(1/\delta))}$, where $\log^*$ is the iterated logarithm. We will be interested in the case when G is abelian and $m = 3$, in which the dependence between ε and δ implicit in Theorem 5.1.1 can be bounded by the bounds of the undirected triangle removal lemma. It is known that we cannot have $\varepsilon < e^{-C\sqrt{\log \delta^{-1}}}$ for some large C [Zha23, Remark 2.3.4]. In the case when $G = \mathbb{F}_p^n$, Fox and Lovasz [FL17] showed, using ideas from the resolution of the cap–set problem, that one only needs $\varepsilon < \delta^{O_p(1)}$.

Theorem 5.1.1 implies the following.

Corollary 5.1.2. *If X_i, Y_i, Z_i satisfy the STPP in a group G of order n, then at least one of $\sum |X_i||Y_i|, \sum |X_i||Z_i|, \sum |Y_i||Z_i|$ is at most $o(n)$.*

Proof. Let $A_1 = \sqcup_i X_i Y_i^{-1}, A_2 = \sqcup_i Y_i Z_i^{-1}, A_3 = \sqcup_i Z_i X_i^{-1}$. By definition of the STPP, the equation $x_1 x_2 x_3 = I$ with $x_i \in A_i$ has $\sum_i |X_i||Y_i||Z_i|$ solutions. By the packing bound, $\sum_i |X_i||Y_i|, \sum_i |Y_i||Z_i|, \sum_i |Z_i||X_i| \leq n$, so by Hölder's inequality there are at most $n^{3/2} = o(n^2)$ solutions to $a_1 a_2 a_3 = I$.

Now suppose that $B_j \subseteq A_j$ satisfy $|B_j|/|A_j| > 0.9999$; we will show that there is a solution to $b_1 b_2 b_3 = I$. For more than a 0.99 fraction of the values of i we must have $|B_1 \cap X_i Y_i^{-1}|/|X_i Y_i^{-1}| > 0.99$ (because $0.99 \cdot 1 + 0.01 \cdot 0.99 = 0.9999$) and similarly for the other sets. Hence by the pigeonhole principle there is some i for which $|B_1 \cap X_i Y_i^{-1}|/|X_i Y_i^{-1}| > 0.99$, $|B_2 \cap Y_i Z_i^{-1}|/|Y_i Z_i^{-1}| > 0.99$, $|B_3 \cap Z_i X_i^{-1}|/|Y_i Z_i^{-1}| > 0.99$.

Now consider the tripartite graph with parts X_i, Y_i, Z_i, where (x, y) is an edge between X_i and Y_i if $xy^{-1} \in B_1 \cap X_i Y_i^{-1}$, (y, z) is an edge between Y_I, Z_i when $yz^{-1} \in B_2 \cap Y_i Z_i^{-1}$, and (z, x) is an edge when $zx^{-1} \in B_3 \cap Z_i X_i^{-1}$. Note that the existence of a triangle in this graph implies that there is a solution to $b_1 b_2 b_3 = I$. First, note that at least $0.9|X_i|$ vertices in X_i have at least $0.9|Y_i|$ neighbors in Y_i. If this were not the case, there would be at most $0.9|X_i||Y_i| + 0.1 \cdot 0.9 \cdot |X_i||Y_i| \leq 0.99|X_i||Y_i|$ edges between X_i and Y_i, and hence $|B_1 \cap X_i Y_i^{-1}|/|X_i Y_i^{-1}| \leq 0.99$, a contradiction. Similarly, at least $0.9|X_i|$ vertices in X_i have at least $0.9|Z_i|$ neighbors in Z_i. Hence at least $0.8|X_i|$ vertices in X_i have $0.9|Y_i|$ neighbors in Y_i and $0.9|Z_i|$ neighbors in Z_i. Pick any such vertex $x_0 \in X_i$. There must be an edge between a neighbor of x_0 in Y_i and a neighbor of x_0 in Z_i, since if not, there would be at most $|Y_i||Z_i| - 0.9^2|Y_i||Z_i| = 0.19|Y_i||Z_i|$ edges between Y_i and Z_i. Thus we have found our triangle.

By Theorem 5.1.1, we can delete subsets of A_i of size $o(n)$ to eliminate all solutions to $x_1x_2x_3 = I$. On the other hand, any three subsets of the A_i's of density 0.9999 contains some such solution. Hence we must have $|A_i| = o(n)$ for some i. ■

Remark 5.1.3. For $G = \mathbb{Z}_q^n$ with q a prime power we can use the bounds of [FL17] and improve the bound of $o(q^n)$ to $q^{n(1-\Theta(1/\log q))} \leq (q/C)^n$ for an absolute constant $C > 1$. While the bound of [FL17] is only stated for $\mathbb{Z}_p^n$, it extends to $\mathbb{Z}_q^n$ by the same argument by using [BCC$^+$17a, Theorem A'].

One can interpret this as saying that the best upper bound on the rank of a direct sum of matrix multiplication tensors provable via the group–theoretic approach is superlinear. We remark the only important property of the matrix multiplication hypergraph for this result was that it satisfies a very weak "regularity" condition. Specifically, considerations similar to those of Corollary 5.1.2 easily show the following:

Theorem 5.1.4. *Fix $\varepsilon > 0$. Let G be a group of order n. Let $X = \sqcup_{i=1}^3 A_i$ be a tripartite hypergraph with $o(n^2)$ triangles such that for any $Y_i \subseteq A_i$ with $|Y_i|/n \geq 1 - \varepsilon$, there exists a $y_i \in Y_i$ such that $(y_1, y_2, y_3) \in E(X)$. Then if X is an induced subhypergraph of X_G, $|A_i| \leq o(n)$ for $i = 1, 2, 3$.*

5.2 An extremal hypergraph problem and matrix multiplication

We begin with the observation that the matrix multiplication hypergraph is an extremal solution to a certain forbidden hypergraph problem.

Proposition 5.2.1. *Let X be a linear tripartite hypergraph with parts of size N such that any two vertices from different parts are incident to at most one common vertex in the third part. Then the number of triangles in X is at most $N3/2$. Furthermore, an extremal example is the matrix multiplication hypergraph $M_{N^{1/2}}$.*

The hypergraphs satisfying the condition of Proposition 5.2.1 can be alternatively characterized as the linear hypergraphs that do not contain (not necessarily induced) copies of the hypergraphs in Figure 5.2.[1]

Proof. We restrict our attention to one of the parts X_1 of X and let d_v be the number of triangles that vertex v in X_1 is contained in. Each $v \in X_1$ is contained in d_v triangles, where the vertices of these triangles belonging to X_2 and X_3 are distinct (as X is linear). Additionally, no pair of such vertices in X_2 and X_3 can be contained in a triangle incident to another vertex $u \in X_1$, so there are $2\binom{d_v}{2}$ pairs of vertices in X_2 and X_3 that are contained in no common triangle. Let (x_2, x_3) be some such pair of vertices. Observe that furthermore, for all $u \neq v \in X_1$, the set of vertices in X_2 and X_3 incident to the set of triangles containing u cannot also contain both x_2 and x_3. For if this happened, there would be triangles (v, x_2, x_3'), (v, x_2', x_3), (u, x_2, x_3''), (u, x_2'', x_3), and then x_2 and x_3 violate the constraint. The total number of triangles equals $m := \sum_{v\in X_1} d_v$, and by the

[1] Here colors represent the parts, so there are really 6 forbidden subhypergraphs for different permutations of the colors.

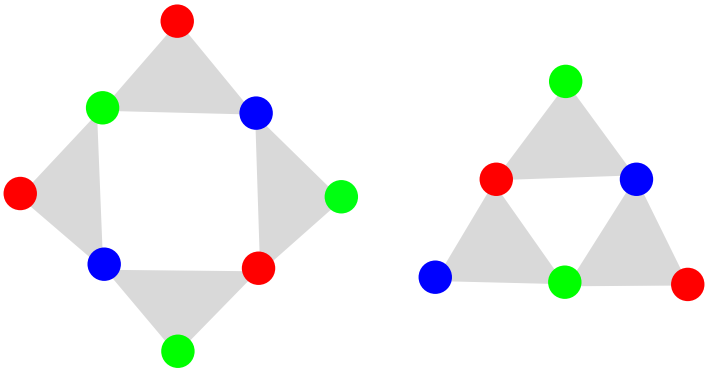

Figure 5.2: The two forbidden hypergraphs in Proposition 5.2.1, up to permutations of the 3 parts.

prior observations it follows that $\sum 2\binom{d_v}{2} + m \leq N^2$. So $\sum d_v(d_v - 1) + m = \sum d_v^2 \leq N^2$. The conclusion follows from Cauchy–Schwarz.

To see that $M_{N^{1/2}}$ is extremal, first note that it contains $N^{3/2}$ triangles, has parts of size N, and is linear. To see that it satisfies the second condition, let (i, j) be a vertex in the first part, and (k, l) in the second part. Then (i, j) is contained in a common triangle with exactly the vertices in the third part of the form $(*, i)$, and (k, l) is incident to exactly the vertices in the third part of the form $(l, *)$. Hence (l, i) is the unique neighbor of both. The same argument shows the claim for vertices in any two parts. ■

Remark 5.2.2. The matrix multiplication hypergraph is not the unique extremal example for this problem. For instance, another example when $N = 4$ (viewed as a Latin square) is

$$\begin{bmatrix} 1 & 2 & & \\ 3 & & & 4 \\ & & 1 & 2 \\ & 4 & 3 & \end{bmatrix}$$

which is easily seen to be non-isomorphic to the 2 by 2 matrix multiplication hypergraph,

$$\begin{bmatrix} 1 & 2 & & \\ 3 & 4 & & \\ & & 1 & 2 \\ & & 3 & 4 \end{bmatrix}.$$

5.3 Equilateral trapezoid-free subsets of groups

By asking what is the densest subgraph of X_G satisfying the conditions of Proposition 5.2.1, we are led to the following problem.

Definition 5.3.1. Let $A, B, C \subseteq G$. We call (A, B, C) equilateral trapezoid-free if for any fixed a', b', c', the following systems of equations in the variables $a \in A, b \in B, c \in C$ each have at most one solution:

$$\begin{aligned} 0 &= a'bc = ab'c \\ 0 &= a'bc = abc' \\ 0 &= ab'c = abc'. \end{aligned}$$

Let $\operatorname{val}(G)$ be the maximum number of solutions to $abc = I$ over all trapezoid-free triples (A, B, C).

The relevance of $\operatorname{val}(G)$ to ω is due to the following.

Proposition 5.3.2. *Suppose that $X_G \geq \bigoplus_i M_{n_i,m_i,p_i}$. Then,* $\operatorname{val}(G) \geq \sum n_i m_i p_i$*.*

Proof. If S_i, T_i, U_i satisfy the STPP, then X_G contains disjoint induced subgraphs $M_{|S_i|,|T_i|,|U_i|}$. By translating the conditions of Proposition 5.2.1 to groups, these satisfy the constraints of Problem 5.4.6 and contain $|S_i||T_i||U_i|$ triangles each. ■

To start, we have the following trivial bounds.

Proposition 5.3.3. *For any group G, $|G| \leq \operatorname{val}(G) \leq |G|^{3/2}$.*

Proof. The lower bound is obtained by taking $A = \{1\}, B = G, C = G$. The upper bound follows from Proposition 5.2.1. ■

The following super multiplicative behavior of val is easily checked.

Proposition 5.3.4. *If (A, B, C) is trapezoid-free triple in G, and (A', B', C') is a trapezoid-free triple in H, then $(A \times A', B \times B', C \times C')$ is a trapezoid-free triple in $G \times H$.*

It is also easily seen that being trapezoid-free is preserved by cyclic permutations of the three sets.

Proposition 5.3.5. *If (A, B, C) is trapezoid-free, then so is (B, C, A), and it achieves the same value.*

By an application of Theorem 5.1.4 combined with the observation that near-extremal solutions to Proposition 5.2.1 are highly "regular", we have the following weak improvement to the trivial upper bound of $|G|^{3/2}$.

Proposition 5.3.6. *For any group G,* $\operatorname{val}(G) \leq |G|^{3/2-o(1)}$*.*

Proof. Suppose for contradiction that there exists $\varepsilon_0 > 0$ such that $\operatorname{val}(G) > \varepsilon_0 |G|^{3/2}$, and let $A_0, B_0, C_0 \subseteq G$ witness $\operatorname{val}(G) = \varepsilon_0 |G|^{3/2}$. It is convenient to first consider the triple $(A, B, C) := (A_0 \times B_0 \times C_0, B_0 \times C_0 \times A_0, C_0 \times A_0 \times B_0)$, which is equilateral-trapezoid free inside of $H := G^3$ by Proposition 5.3.4 and Proposition 5.3.5, and witnesses $\operatorname{val}(H) \geq \varepsilon |H|^{3/2}$ where $\varepsilon := \varepsilon_0^3$. Let $|H| = N$. Let X be the tripartite hypergraph with parts A, B, C and where there is a triangle between all triples (a, b, c) where $abc = I$. Let $n := |A| = |B| = |C|$. By Proposition 5.2.1 we must have $n \geq \varepsilon^{2/3} N$. Note that the number of triangles in X equals $\varepsilon N^{3/2} \geq \varepsilon n^{3/2}$. In what follows, we define the degree of a vertex in X to be the number of triangles containing it.

Let Y be the random variable that is uniformly distributed over the multiset of vertex degrees from one part of X, say A. Then $\mathbf{E}[Y] \geq \varepsilon n^{1/2}$ and $\mathbf{E}[Y^2] \leq n$ (this second inequality follows from Cauchy–Schwarz as in the proof of Proposition 5.2.1). By the Payley-Zygmund inequality, for any $\theta > 0$, $\mathbf{P}(Y > \theta \cdot \varepsilon n^{1/2}) \geq (1-\theta^2)\varepsilon^2$. Taking $\theta = 1/2$, we conclude that at least $p \cdot n := 3n\varepsilon^2/4$ vertices in A have degree at least $\varepsilon n^{1/2}/2$. This holds for B and C as well.

Now let S, T, and U be any subsets of A, B, C of size at least $n(1-p/\lambda)$; we'll pick $\lambda \in \mathbb{N}$ later. Then the number of triangles incident to any one of these sets, say S, is at least

$$np(1-\lambda^{-1}) \cdot \varepsilon n^{1/2}/2 = (3/8)n^{3/2}\varepsilon^3(1-\lambda^{-1}),$$

and the number of triangles incident to $[n] \setminus T$ or $[n] \setminus U$, sets of size at most np/λ, is at most

$$(n^2 \cdot np/\lambda)^{1/2} = (3^{1/2}/2)n^{3/2}\varepsilon\lambda^{-1/2}$$

by Cauchy–Schwarz. It follows that the number of triangles with one vertex in each of S, T, U is at least

$$(3/8)n^{3/2}\varepsilon^3(1-\lambda^{-1}) - 2 \cdot (3^{1/2}/2)n^{3/2}\varepsilon\lambda^{-1/2}$$

which is greater than 1 for $\lambda \gg \varepsilon^{-4}$. In summary, between any three subsets of A, B, C size roughly $n(1-\varepsilon^6)$, there is a triangle.

Recall that $n \geq \varepsilon^{2/3}N$. Since X has at most $N^{3/2} \leq o(N^2)$ triangles, by Theorem 5.1.4 we can remove $o(N) = o(n)$ vertices to remove all triangles. But by what we have just shown, after deleting this few vertices some triangle will remain, a contradiction. ■

If we use the strongest-known quantitative bound for Theorem 5.1.1 in Proposition 5.3.6, we find that $\operatorname{val}(G) \leq |G|^{3/2}/\log^* |G|$. Using the polynomial bound for the removal lemma in vector spaces [FL17], this can be improved to $\operatorname{val}(\mathbb{Z}_q^n) \leq q^{3/2(1-1/\log q)n}$.

The following theorems show the importance of understanding $\operatorname{val}(\mathbb{Z}_m)$.

Theorem 5.3.7. *Suppose that one can achieve $\omega = 2$ via STPP constructions in the family of groups $\mathbb{Z}_q^n$ where q is a prime power. Then there exists a constant $c > 0$ such that $\operatorname{val}(\mathbb{Z}_m) \geq \Omega(m^{1+c})$.*

Proof. By Corollary 5.1.2 and remark 5.1.3, any STPP construction satisfies $\sum |X_i||Y_i| \leq (q/C)^n$ (without loss of generality) where C is an absolute constant. By Hölder's inequality, $\sum(|X_i||Y_i||Z_i|)^{2/3} \leq q^{2n/3}(q/C)^{n/3} = (q/C^{1/3})^n$. If we can obtain $\omega < 3 - \delta$ then

$$\begin{aligned} q^n < \sum(|X_i||Y_i||Z_i|)^{2/3 \cdot \delta + (1-\delta)} &= \sum(|X_i||Y_i||Z_i|)^{2/3 \cdot \delta}(|X_i||Y_i||Z_i|)^{1-\delta} \\ &\leq \left(\sum(|X_i||Y_i||Z_i|)^{2/3}\right)^{\delta}\left(\sum |X_i||Y_i||Z_i|\right)^{1-\delta} \\ &\leq (q/C^{1/3})^{\delta n}\operatorname{val}(G)^{1-\delta} \end{aligned}$$

so $\operatorname{val}(G) > q^n(C^{\delta/3(1-\delta)})^n$. By choosing δ sufficiently close to 1, $\operatorname{val}(G) > q^n 4^n$. By taking k-fold products of the sets defining the STPP constructions, we find that $\operatorname{val}(\mathbb{Z}_q^{kn}) > (4q)^{kn}$ for all k. Let $N = kn$.

Next consider the embedding $\varphi : \mathbb{Z}_q^N \to \mathbb{Z}_{(3q)^N}$ defined by $\varphi(x_1, \ldots, x_N) = x_1 + x_2 3q + \cdots + x_n(3q)^{N-1}$. Note that

$$a_1 + a_2 + a_3 \neq a_4 + a_5 + a_6 \implies \varphi(a_1) + \varphi(a_2) + \varphi(a_3) \neq \varphi(a_4) + \varphi(a_5) + \varphi(a_6).$$

Hence the image of an STPP under φ is an STPP inside of $\mathbb{Z}_{(3q)^N}$, so $\operatorname{val}(\mathbb{Z}_{(3q)^N}) > (4q)^N$. Because this holds for some fixed q and all $N = kn$, the theorem follows. ■

Although we expect that Theorem 5.3.7 should hold for arbitrary abelian groups, we do not know how to remove the constraint that the modulus is a prime power. This is due to the fact that the slice rank argument yields better bounds on the size of matchings for prime power modului than for general moduli (compare Theorem A and A' in [BCC$^+$17a]). These weaker bounds for non-prime power moduli are not known to be tight as far as we are aware.

Next we show that sufficiently strong *simultaneous double product property constructions* [CKSU05], which are known to achieve $\omega < 2.48$, imply strong lower bounds on $\mathrm{val}(\mathbb{Z}_m)$.

Definition 5.3.8. We say that sets $(A_i, B_i)_{i=1}^n$ satisfy the simultaneous double product property (or SDPP for short) if

1. For all i, $aa'^{-1} = bb'^{-1}$ only has the solution $a = a', b = b'$ for $a, a' \in A_i, b, b' \in B_i$,
2. $a_i(a'_j)^{-1} b_j (b'_k)^{-1} = 1$ implies $i = k$, where $a_i \in A_i, a'_j \in A_j, b_j \in B_j, b'_k \in B_k$.

In [CKSU05] it was conjectured that one can achieve $\omega = 2$ using SDPP constructions in abelian groups. This amounts to the following:

Conjecture 5.3.9. *[CKSU05, Conjecture 4.7] For arbitrarily large n, there exists an abelian group G of order $n^{2-o(1)}$ and n pairs of sets A_i, B_i where $|A_i||B_i| > n^{2-o(1)}$ satisfying the SDPP.*

We thank Chris Umans for informing us of the fact that if Conjecture 5.3.9 is true, then it is true in cyclic groups. The following theorem was motivated by this fact.

Theorem 5.3.10. *If Conjecture 5.3.9 is true, then for any $\varepsilon > 0$, $\mathrm{val}(\mathbb{Z}_m) \geq O(m^{4/3-\varepsilon})$.*

Proof. We begin by recalling how to turn an SDPP construction into an STPP construction [CKSU05, Section 6.2]. Let $\Delta_n = \{(a, b, c) \in \mathbb{Z}^3_{\geq 0} : a + b + c = n - 1\}$, and let $S \subseteq \Delta_n$ be corner-free of size $n^{2-o(1)}$. For all $v = (v_1, v_2, v_3) \in S$, define the following subsets of G^3:

$$\begin{aligned} A_v &= A_{v_1} \times \{1\} \times B_{v_3} \\ B_v &= B_{b_1} \times A_{v_2} \times \{1\} \\ C_v &= \{1\} \times B_{v_2} \times A_{v_3} \end{aligned}$$

It can be verified that the sets $(A_v, B_v, C_v)_{v \in S}$ satisfy the STPP. Hence Conjecture 5.3.9 yields an STPP with $n^{2-o(1)}$ triples of sets of size $n^{2-o(1)}$, inside a group of size $n^{6-o(1)}$.

Now consider the map from $G^3 \cong \mathbb{Z}_{m_1} \times \cdots \times \mathbb{Z}_{m_k}$ to $G' := \mathbb{Z}_{\prod_i 3m_i}$ where $m_1 \leq \cdots \leq m_k$ sending $(x_1, \ldots, x_k)$ to $x_1 + (3m_1)x_2 + (3m_1)(3m_2)x_3 + \cdots$. First, the image of an STPP construction under this map is still an STPP. This shows that $\mathrm{val}(G') > n^{2-o(1)} \cdot n^{3(2-o(1))} = n^{8-o(1)}$. Second, for all fixed $c > 0$ and $\ell \in \mathbb{N}$, G^3 cannot contain a subgroup of size $|G^3|^c$ generated by elements of order at most ℓ, by the slice–rank barrier [BCC$^+$17a, Proposition 4.2]. Hence the number of m_i's which are at most ℓ is at most $\log_2(|G^3|^c)$. The number of m_i's which are greater than ℓ is trivially less than $\log_\ell |G^3|$. So,

$$|G'| = \prod_{m_i \leq \ell} 3m_i \prod_{m_i > \ell} 3m_i \leq 3^{\log_2(|G^3|^c) + \log_\ell |G^3|} \cdot |G^3|.$$

By taking $c \to 0$ and $\ell \to \infty$, we find that $|G'| \leq n^{6+o(1)}$, and the claimed bound follows. ∎

Note that here there is no restriction on the abelian groups in consideration, unlike there was in the previous theorem.

5.4 The value of $\mathbb{Z}_n$

In light of Theorems 5.3.7 and 5.3.10, we now turn our attention to the problem of determining $\mathrm{val}(\mathbb{Z}_n)$. Our weakest conjecture in this direction is the following.

Conjecture 5.4.1. *For all $\varepsilon > 0$, $\mathrm{val}(\mathbb{Z}_n) \leq O(n^{1+\varepsilon})$.*

While the quantity $\mathrm{val}(\mathbb{Z}_n)$ may seem opaque from Definition 5.3.1, it can easily be visualized. This is done by considering the natural notion of a trapezoid-free subset of the plane:

Definition 5.4.2. Let $A, B, C \subseteq [n]$. We call (A, B, C) trapezoid free if for any fixed a', b', c', the following systems of equations in the variables $a \in A, b \in B, c \in C$ each have at most one solution:

$$\begin{aligned} n &= a' + b + c = a + b' + c \\ n &= a' + b + c = a + b + c' \\ n &= a + b' + c = a + b + c'. \end{aligned}$$

Let $\mathrm{val}(n)$ be the maximum number of solutions to $a + b + c = n$ over all trapezoid-free triples (A, B, C).

The following shows that it suffices to study $\mathrm{val}(n)$.

Proposition 5.4.3. $3 \cdot \mathrm{val}(3n) \geq \mathrm{val}(\mathbb{Z}_n) \geq \mathrm{val}(n/3)$.

Proof. Suppose that $\mathrm{val}(n)$ is witnessed by sets A, B, C. For $N > n$, $A + (N - n), B, C$ then witness $\mathrm{val}(N) \geq \mathrm{val}(n)$. If we take $N = 3n$, we have that $A + 2n \subseteq [N]$ and $B, C \subseteq [N/3]$, so $a + b + c \leq 5N/3 < 2N$. Since $a + 2n + b + c = 0 \bmod N \iff a + 2n + b + c = N$, this implies that the sets $A + 2n, B, C \bmod N$ are trapezoid-free.

In the other direction, suppose $\mathrm{val}(\mathbb{Z}_n)$ is witnessed by $A, B, C \bmod n$. There are at least $\mathrm{val}(\mathbb{Z}_n)/3$ solutions to one of $a + b + c = n, a + b + c = 2n, a + b + c = 3n$; let N be the right-hand side of the most frequently satisfied equation. Every solution to $a + b + c = N$ is a solution to $a + b + c = 0 \bmod n$, and so $A, B, C \subseteq [N]$ must be trapezoid-free. ■

We may visualize trapezoid-free sets as follows. Draw Δ_n in the plane as a triangular grid of points. Sets A, B, C correspond to collections of lines parallel to the sides of Δ_n, and a solution $a + b + c = n$ corresponds to a points in Δ_n contained in one line in each of these three directions. Let $S \subseteq \Delta_n$ be the collection of points contained in three lines. A violation of a constraint of Definition 5.4.2 corresponds to either a subset of 3 points in S forming an equilateral triangle with sides parallel to the sides of Δ_n, or a subset of 4 points with sides parallel to the sides of Δ_n forming an equilateral trapezoid. Equivalently, we are deleting lines parallel to the sides of Δ_n to eliminate all of these configurations, while leaving as many points as possible. This explains the name "trapezoid-free". See Figure 5.3.

A moment's reflection shows that the proof of the $n^{3/2}$ upper bound of Proposition 5.3.3 actually held for a (possibly) much weaker version of Definition 3.2.1 where one only requires that the *expected* number of solutions of *one* of the three systems of two equations was at most 1. We begin by noting that this upper bound is essentially best-possible for this weakened problem in $\mathbb{Z}_n$. In other words, one cannot hope to prove the above conjecture via an "asymmetric" averaging argument.

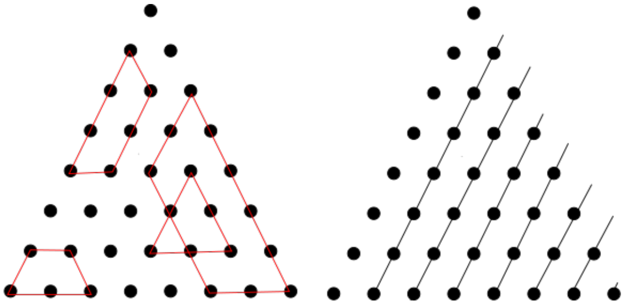

Figure 5.3: Left: some forbidden trapezoids and triangles in Δ_n. Right: a trapezoid-free subset of Δ_n of size n is obtained by deleting all lines but one along one direction.

Proposition 5.4.4. *There exist $A, B, C \subseteq \mathbb{Z}_n$ such that*

$$\mathop{\mathbf{E}}_{a' \in A, b' \in B} [\#\{(a,b,c) : 0 = a' + b + c = a + b' + c\}] \le 1$$

and there are $n^{3/2-o(1)}$ solutions to the equation $a + b + c = 0$ with $a \in A, b \in B, c \in C$.

Proof. It is convenient to work in $\mathbb{Z}$. Let $r(A, B, c)$ denote the number of representations of $-c$ as $a+b$. First note that the proposition is equivalent to the statement that $\sum_{c \in C} r(A, B, c)^2 \le |A||B|$ and $\sum_{c \in C} r(A, B, c) = n^{3/2-o(1)}$.

Let $S \subseteq [n]$ be 3AP-free and of size $n^{1-o(1)}$. Consider the sets:

$$A = B = [3n^2, 4n^2] \cup \bigcup_{x \in S} [xn, xn + n/2]$$

$$C = \{2xn + y : x \in S, y \in [n]\}$$

By definition, for any $x \in S$ and $y \in [n]$, $2xn + y = c \in C$. If we have any representation $c = a + b$, then $a, b < 3n^2$. So we have $a = x_1 n + y_1, b = x_2 n + y_1$ with $x_1, x_2 \in S$ and $1 \le y_1, y_2 \le n$. So $(x_1 + x_2)n + (y_1 + y_2) = 2xn + y$, and then we are forced to have $x_1 + x_2 = 2x$ and $y_1 + y_2 = y$. But because S is 3AP-free, we must have $x_1 = x_2 = x$. Hence $r(A, B, c)$ is exactly the number of solutions to $y = y_1 + y_2$ with $y_1, y_2 \in [n]$, which is $\Omega(n)$ for $\Omega(n)$ choices of $y \in [n]$. Hence $\sum_{c \in C} r(A, B, c) = \Theta(|S|n^2) = n^{3-o(1)} \approx (4n^2)^{3/2}$. Also, we have that $\sum_{c \in C} r(A, B, c)^2 = n^{4-o(1)} < |A||B| = \Theta(n^4)$. ■

Can one find a construction achieving $n^{3/2-o(1)}$ for the averaging version of Definition 5.3.1 that involves all three systems of equations? That is:

Question 5.4.5. What is the maximum over all $A, B, C \subseteq \mathbb{Z}_n$ satisfying

$$\mathop{\mathbf{E}}_{a' \in A, b' \in B} [\#\{(a,b,c) : 0 = a' + b + c = a + b' + c\}] \le 1,$$
$$\mathop{\mathbf{E}}_{a' \in A, c' \in C} [\#\{(a,b,c) : 0 = a' + b + c = a + b + c'\}] \le 1,$$
$$\mathop{\mathbf{E}}_{b' \in B, c' \in C} [\#\{(a,b,c) : 0 = a + b' + c = a + b + c'\}] \le 1,$$

of the number of solutions to $a + b + c = 0$?

A relaxation of a trapezoid-free triple of $[n]$ that still seems very stringent is that of a *triforce-free* triple, defined as follows.

Problem 5.4.6. Let $A, B, C \subseteq [n]$. We say that (A, B, C) is triforce-free if there is no solution to

$$a + b + c' = a + b' + c = a' + b + c = n$$

with $a \neq a', b \neq b', c \neq c'$. We write $trival(n)$ for the maximum over all such A, B, C of the number of solutions to $a + b + c = n$.

This condition just says that $\{(a, b, c) \subseteq A \times B \times C : a+b+c = n\} \cap \Delta_n$ is corner-free. Every trapezoid-free triple of sets is therefore triforce-free, so $trival(n) \geq \mathrm{val}(n)$. Note that (A, B, C) is triforce-free if the subhypergraph of the hypergraph with parts $[n]$ and triangles between any triples summing to 0 induced by A, B, C does not contain the triforce hypergraph (the second hypergraph in Figure 5.2).

We expect that being triforce-free is much stronger than being corner-free, however. Here is a weaker notion than triforce-free. We thank Ryan O'Donnell for suggesting this definition.

Definition 5.4.7. We call $S \subseteq \Delta_n$ skew-corner free if for $(x, y, z), (x, y', z') \in S$, it holds that $(x+y-y', y'', z'') \notin S$ for all y'', z'', and this remains true after any permutation of the coordinates of S.

Pictorially, this says that for any two points lying on an axis-aligned line, the parallel line passing through the third point that would form a corner with these two points must contain no points. Since S is corner-free, we automatically have that $|S| \leq n^2/(\log \log n)^c$ by [Shk06]. Yet we conjecture that $|S| \leq n^{2+\varepsilon}$ for all $\varepsilon > 0$. Note that to rule out Conjecture 5.3.9, we would only need to show that $|S| \leq n^{4/3-0.01}$.

The best lower bound that we know is $\Omega(n)$; n is obtained trivially by taking one line on the side of Δ_n, and it is not hard to improve this to $3n/2$. We have found examples exceeding these bounds with computer search (see **??**).

If we weaken Definition 5.4.7 by dropping the requirement that the condition holds for all permutations of coordinates, we are led to the following problem.

Question 5.4.8. What is the size of the largest subset of $[n]^2$ containing no configuration $(x, y), (x, y + d), (x + d, y')$ with $d \neq 0$?

We then have the following nontrivial lower bound for this relaxed problem, due to a MathOverflow answer of Fedor Petrov [hp].

Proposition 5.4.9. *There is a subset of* $[n]^2$ *of size* $n \log n/\sqrt{\log \log n}$ *with no three-point configuration* $(x, y), (x, y + d), (x + d, y')$ *with* $d \neq 0$.

Proof. $A \subseteq [n]$ is called *primitive* if for all $a \neq a' \in A, a \nmid a'$. It is easily seen that if A is primitive then the set of points $(a, ka) \subseteq [n]^2$ for all $k \leq n/a$ avoids the forbidden configurations. This gives a subset of size $n \sum_{a \in A} 1/a$. It was observed by Pillai that there exists a $c > 0$ and a primitive set A where $\sum_{a \in A} 1/a > c \log n/(\log \log n)^{1/2}$ [ESS67]. We note that this is best-possible, matching (up to the constant) a lower bound on $\sum_{a \in A} 1/a$ for primitive A due to Behrend [Beh35]. ■

This construction breaks when considering one nontrivial permutations of the coordinates in Definition 5.4.7. This corresponds to the following problem:

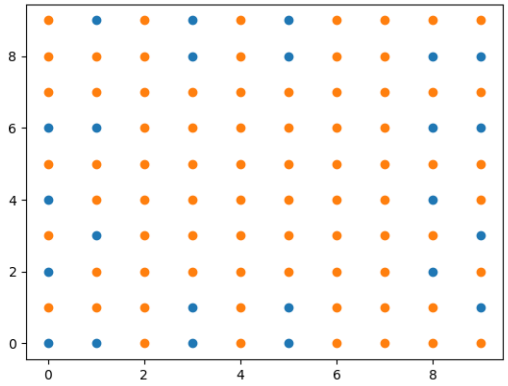

Figure 5.4: 24 points in a 10×10 grid avoiding the configurations of Proposition 5.4.9, but not of Question 5.4.10.

Question 5.4.10. What is the largest subset of $[n]^2$ with no three point configurations $(x, y), (x, y + d), (x + d, y')$ or $(x, y), (x + d, y), (x', y + d)$, with $d \neq 0$?

As far as we know, it is possible that the answer to this is $O(n)$.

Chapter 6

New High-Dimensional Expanders from Matrix Groups

In 1989, Babai, Kantor, and Lubotzky made a conjecture that significantly guided research on expander graphs:

Conjecture 6.0.1. *([BKL89].) There are constants $k \in \mathbb{N}$ and $\lambda < 1$ such that for every nonabelian finite simple group G, there is a symmetric set $S \subseteq G$ of $2k$ generators such that the Cayley graph $\mathrm{Cay}(G, S)$ is a λ-spectral expander graph.*

(Here we say that a graph is a λ-spectral expander if all the eigenvalues of its random walk matrix, excluding the largest, are at most λ.)

Notable achievements toward the conjecture include: Kassabov's proof [Kas07] for the alternating groups; work of Lubotzky and Nikolov [KLN06] proving the conjecture for non-Suzuki groups of Lie type (the Chevalley groups and their twisted versions); and, the Breuillard–Green–Tao [BGT11] proof for the Suzuki groups. In light of the Classification of Finite Simple Groups [Asc04], these completed the proof of Conjecture 6.0.1. An immediate consequence is that for every nonabelian simple group G, there is a $2k$-regular λ-spectral expander $\mathfrak{K}$ such that G acts transitively on the vertices of $\mathfrak{K}$.

Having expander graphs with such nontrivial symmetry properties (or even stronger ones) has played an important role in applications to computer science. For example, motivated by the search for locally testable codes (see [KS08]), Kaufman and Wigderson [KW16] made substantial progress on finding so-called "highly symmetric" LDPC codes of constant rate and relative distance ("good") using expanding Cayley graphs of nonabelian groups; at the same time, they showed that highly symmetric LDPC codes arising from abelian — or even solvable — groups cannot work. Later, notable work of Kaufman and Lubotzky [KL12] (see also [Bec16]) positively resolved the problem, giving explicit, highly symmetric, good LDPC codes; the main tool was the use of explicit *edge-transitive* (not just vertex-transitive), highly expanding (indeed, Ramanujan) Cayley graphs of $\mathrm{PSL}_2(\mathbb{F}_q)$ (for $q = 4093$). In turn, the existence of these highly-symmetric expanders arose from the construction of Ramanujan *high-dimensional expanders* (HDXs) [Bal00, CSZ03, Li04, LSV05b, LSV05a, Sar04] from Bruhat–Tits buildings, relying on the Lafforgue's work [Laf02] on the Langlands correspondence.

High-dimensional expanders — defined, say, as simplicial complexes where the 1-skeleton of every link is a λ-spectral expander — have been crucial in many new works in theoretical computer science, either through inspiration, their spectral analysis, or their direct construction. Example applications include results in analysis of Boolean functions [DDFH18], computational geometry [FGL+12], inapproximability [AJT19, DFHT21], list-decoding [AJQ+20, DHK+21], Markov chain mixing [AL20], property testing [DK17, DD19, KM20, KO20], and quantum codes [EKZ20, KT21]; particularly notable examples including the resolution of the Mihail–Vazirani Conjecture on the bases-exchange walk for matroids [ALOV19] and the construction of locally testable codes of constant rate, distance, and locality [DEL+21, PK21].

6.0.1 Our goal

In this chapter, we investigate a problem similar to Conjecture 6.0.1 for high-dimensional expanders. Namely, for nonabelian finite simple groups G, we seek:

1. bounded-degree λ-spectral HDXs whose top-dimensional faces are acted on transitively by G,
2. with λ arbitrarily close to 0, as opposed to merely bounded away from 1.

(Recall that existence of highly symmetric good LDPC codes was resolved by obtaining *one*-dimensional HDXs — i.e., expander graphs — with both properties.) The aforementioned HDXs built from Bruhat–Tits buildings [Li04, LSV05b, LSV05a, Sar04] have property (2) above, and the work of Kaufman and Lubotzky [KL12] also verified property (1) for $G = \mathrm{PSL}_3(\mathbb{F})$ (for $\mathrm{char}\mathbb{F}$ sufficiently large). Later, Kaufman and Oppenheim [KO18] gave a new (and elementary) construction of HDX families of any dimension d satisfying both (1) and (2) with $G = \mathrm{PSL}_{d+1}(\mathbb{F})$. These two constructions are the only previous examples of bounded-degree λ-spectral HDXs of which we are aware. To quote the final remark from [LSV05b]: "*Of course one hopes eventually to define and construct Ramanujan complexes as quotients of the Bruhat–Tits buildings of other simple groups as well.*"

Results. We give strongly explicit constructions of d-dimensional HDX families satisfying properties (1) and (2) above, for any rank-d *Chevalley group* G (except for "G_2") over any field $\mathbb{F}$ of characteristic exceeding 3.[1] Informally, Chevalley groups (also known as the untwisted groups of Lie type) are the finite-field analogues of continuous Lie groups. These groups are specified by two pieces of data: a *root system* Φ, consisting of a set of vectors in $\mathbb{R}^d$ with certain symmetry properties, and a finite field $\mathbb{F}$. Our work gives a general recipe that produces HDX families from Chevalley groups with Φ and the characteristic of $\mathbb{F}$ being fixed, and with $|\mathbb{F}|$ growing. Our approach generalizes that of [KO18], which corresponds to the case $\Phi = A_d$. As with their work, our construction incidentally gives new families of strongly explicit Δ-degree-bounded λ-spectral expander graphs, with $\lambda \to 0$ as $\Delta \to \infty$.

6.0.2 Our approach

As in [KO18], we associate to G a *coset complex*, a kind of d-dimensional simplicial complex determined by G and a choice of subgroups $H_1, \dots, H_{d+1}$ of G. A few challenges arise in

[1] In fact, we can show that our construction works for characteristic 3 when one excludes the case of G_2. But for simplicity of presentation we will just assume the characteristic exceeds 3.

generalizing the construction of [KO18] to Chevalley groups G of type other than A_d. One immediate question is: what is a "good" choice of $H_1, \ldots, H_{d+1}$? We give one such choice, which has an elegant description in terms of the root system Φ associated with G: the H_i's are certain unipotent subgroups of G (these are essentially groups of upper unitriangular matrices), obtained from a set of fundamental roots of Φ. While all of our constructions can be realized with matrices (see the examples in the next section), it is more convenient in our analysis to work with a set of generators and relations of G known as the *Steinberg presentation*. In particular, the *Chevalley commutator formula* gives us workable descriptions of the subgroups H_i, and the links of our complexes.

As in [KO18], we apply the *trickling down* theorem of [Opp18] (originating in work of Garland [Gar73]) to show that these coset complexes have expanding links. This theorem says that under a mild connectivity condition, it suffices to show that the links of the $(d-2)$-dimensional faces are good expander graphs. The connectivity condition will follow from some calculations using the properties of Chevalley groups and root systems. In their case of $\Phi = A_d$, Kaufman–Oppenheim establish expansion of links by appealing to a general result of Ershov–Jaikin-Zapirain [EJ10] on expansion in certain groups of nilpotency class two. Unfortunately, to handle root systems Φ that are not "simply-laced", one would need an analogous result for groups of nilpotency class three (and higher, when $\Phi = G_2$). Related results were given in [EJK17] (see its Sec. 10.3), but these are not strong enough for our setting. An alternative, and much simpler, proof of expansion of the Kaufman–Oppenheim complexes was given by Harsha and Saptharishi [HS19]; their proof was quite specific to the $\Phi = A_d$ case, but we were much inspired its elementary nature.

We prove expansion by observing that, when $\Phi \neq G_2$, the squares of the links of the $(d-2)$-dimensional faces are Cayley graphs of abelian groups. This allows us to express their eigenvalues as character sums, which we bound with an elementary argument that ultimately boils down to the Schwarz–Zippel lemma. (In the G_2 case, the squared links are not abelian Cayley graphs, but we discuss some approach that might be used ito show their expansion.)

6.0.3 Example constructions

In this section we explicitly give the easiest new HDX family implied by our work. We start by recalling the basic construction of [KO18], arising from the group $G = SL_3(\mathbb{F})$.[2] Let $\mathbb{F} = \mathbb{F}_p[x]/(f)$ where p is prime and $f \in \mathbb{F}_p[x]$ is irreducible of degree m. Now define the

[2] In [KO18] they work over the ring $\mathbb{F}_p[x]/(x^m)$ rather than the field $\mathbb{F}_{p^m}$, but this does not materially change their result, and we prefer to work with the field. Also, regarding the distinction between $SL_3(\mathbb{F})$ and $\mathrm{PSL}_3(\mathbb{F})$, see Footnote 3.

following three subgroups of G:

$$H_1 = \left\{ \begin{bmatrix} 1 & \ell_1 & Q \\ & 1 & \ell_2 \\ & & 1 \end{bmatrix} : \deg(\ell_1), \deg(\ell_2) \leq 1, \deg(Q) \leq 2 \right\},$$
$$H_2 = \left\{ \begin{bmatrix} 1 & \ell_1 & \\ & 1 & \\ \ell_2 & Q & 1 \end{bmatrix} : \deg(\ell_1), \deg(\ell_2) \leq 1, \deg(Q) \leq 2 \right\},$$
$$H_3 = \left\{ \begin{bmatrix} 1 & & \\ Q & 1 & \ell_1 \\ \ell_2 & & 1 \end{bmatrix} : \deg(\ell_1), \deg(\ell_2) \leq 1, \deg(Q) \leq 2 \right\}.$$

Let $\mathfrak{K}(p, m)$ be the 2-dimensional simplicial complex whose vertices are the cosets of these subgroups inside $SL_3(\mathbb{F})$, and where a "triangle" (2-dimensional face) is added between a triple of cosets g_1H_1, g_2H_2, g_3H_3 whenever $g_1H_1 \cap g_2H_2 \cap g_3H_3 \neq \emptyset$. Edges are included between any two cosets contained in a common triangle. This is an example of a *coset complex*, a well-studied construction from the theory of algebraic groups.

In [KO18] it was shown that for any fixed $\lambda > 0$, and for sufficiently large p, the complex $\mathfrak{K}(p, m)$ is a bounded-degree λ-spectral HDX. (Here "bounded-degree" means that each vertex of $\mathfrak{K}(p, m)$ is contained in a number of triangles depending only on p.) Moreover, $SL_3(\mathbb{F})$ acts simply transitively on the set of triangles in $\mathfrak{K}(p, m)$. In a similar manner, Kaufman and Oppenheim show how a d-dimensional HDX family can be associated to $SL_{d+1}(\mathbb{F})$.

Following this, the most basic new construction provided by our work is as follows. Again, we form a coset complex, but this time we will consider cosets of subgroups of the 4×4 *symplectic group*, $\mathrm{Sp}_4(\mathbb{F})$,[3] defined by

$$\mathrm{Sp}_4(\mathbb{F}) = \left\{ A \in \mathbb{F}^{4\times 4} : A \begin{bmatrix} 0 & I_{2\times 2} \\ -I_{2\times 2} & 0 \end{bmatrix} A^{\intercal} = \begin{bmatrix} 0 & I_{2\times 2} \\ -I_{2\times 2} & 0 \end{bmatrix} \right\}.$$

The vertices of our coset complex $\mathfrak{K}_{\mathrm{Sp}_4}(p, m)$ will be the cosets of the following subgroups of

[3] Technically, this group is not simple; it only becomes the simple group $\mathrm{PSp}_4(\mathbb{F})$ upon identifying the matrices A and $-A$. This is an example of the (very minor) distinction between "universal" and "adjoint" Chevalley groups that is explained in Definition 6.1.24.

$\mathrm{Sp}_4(\mathbb{F})$:

$$H_1 = \left\{ \begin{bmatrix} 1 & \ell_1 & C & \ell_1\ell_2 + Q \\ & 1 & Q & \ell_2 \\ & & 1 & \\ & & -\ell_1 & 1 \end{bmatrix} : \deg(\ell_1), \deg(\ell_2) \leq 1, \deg(Q) \leq 2, \deg(C) \leq 3 \right\},$$

$$H_2 = \left\{ \begin{bmatrix} 1 & \ell_1 & & \\ & 1 & & \\ \ell_2 & Q & 1 & \\ \ell_1\ell_2 + Q & C & -\ell_1 & 1 \end{bmatrix} : \deg(\ell_1), \deg(\ell_2) \leq 1, \deg(Q) \leq 2, \deg(C) \leq 3 \right\},$$

$$H_3 = \left\{ \begin{bmatrix} 1 & & & \\ & 1 & & \ell_1 \\ \ell_2 & & 1 & \\ & & & 1 \end{bmatrix} : \deg(\ell_1), \deg(\ell_2) \leq 1 \right\}.$$

The triangles in $\mathfrak{K}_{\mathrm{Sp}_4}(p, m)$ are again added between triples of cosets whenever they have a nontrivial intersection. Our work shows that for any $\lambda > 0$, provided $p \geq 2^{\frac{(1+\lambda)^2}{\lambda^2}}$, the 2-dimensional complexes $(\mathfrak{K}_{\mathrm{Sp}_4}(p, m))_m$ form a (strongly explicit) λ-spectral HDX family of size $\Theta(p^{10m-4})$ in which each vertex participates in at most p^{22} triangles. Moreover, the group $\mathrm{Sp}_4(\mathbb{F}_{p^m})$ acts on $\mathfrak{K}_{\mathrm{Sp}_4}(p, m)$, with the action being transitive on triangles. Finally, we remark that the underlying skeleton of $\mathfrak{K}_{\mathrm{Sp}_4}(p, m)$ is a (strongly explicit) λ-spectral expander graph of degree at most p^{11} and with $\Theta(p^{10m-4})$ vertices. Since this graph is tripartite, its smallest eigenvalue is at least $-1/2$, and it is therefore also a *two*-sided $1/2$-spectral expander.

6.0.4 Outline

In Section 6.1 we give an overview of high-dimensional spectral expansion and coset complexes. We then briefly discuss Chevalley groups and root systems, making explicit all facts about Chevalley groups that we will need.

In Section 6.2 we give the choice of subgroups used in our coset complex construction. We show in Corollary 6.2.19 that these have the connectivity properties needed to apply the trickling down Theorem 6.1.2. In Section 6.2.3 we show that the links of the $(d-2)$-dimensional faces in these complexes are good expander graphs. By Fact 6.1.9, this conveniently reduces to studying the expansion of vertex links in three different 2-dimensional complexes, two of which are the examples in Section 6.0.3.

We conclude with further questions. It is interesting to ask if our analogue of Conjecture 6.0.1 has an affirmative answer when G is a more "combinatorial" group; for example, the symmetric/alternating group. We also leave open the case of the Chevalley group based on root system G_2; we conjecture it has the desired expansion properties, and suggest an approach to proving this.

6.1 Preliminaries

Let $\mathbb{N} = \{0, 1, 2, \ldots\}$, $\mathbb{Z}_+ = \{1, 2, 3, \ldots\}$. We identify elements in a finite field $\mathbb{F}$ of size p^m (where p is prime) with polynomials in $\mathbb{F}_p[x]/(f)$ for some irreducible polynomial $f \in \mathbb{F}_p[x]$ of degree m. When we write $\deg(t)$ for $t \in \mathbb{F}$ we mean the degree of the corresponding polynomial in the quotient ring. If g and h are elements of a group, we use the notation $[g, h] = g^{-1}h^{-1}gh$ for their commutator.

6.1.1 High-dimensional spectral expansion

In this section we recall the notion of spectral HDX families. Let $\mathfrak{K}(0)$ be a finite set. A *simplicial complex* $\mathfrak{K}$ with vertex set $\mathfrak{K}(0)$ is a collection of subsets of $\mathfrak{K}(0)$ satisfying the following conditions:

1. $\{v\} \in \mathfrak{K}$ for all $v \in \mathfrak{K}(0)$;
2. If $\sigma \in \mathfrak{K}$, then $\tau \in \mathfrak{K}$ for all $\tau \subseteq \sigma$.

Said differently, $\mathfrak{K}$ is a downward-closed hypergraph on the set $\mathfrak{K}(0)$. For $i = -1, 0, 1, 2, \dots$, we denote by $\mathfrak{K}(i)$ the set of subsets of size $i + 1$ in $\mathfrak{K}$. An element of $\mathfrak{K}(i)$ is called an *i-dimensional face*. $\mathfrak{K}$ is said to be *pure* if all maximal faces are d-dimensional for some d; in this case we say that d is the *dimension* of $\mathfrak{K}$, denoted $\dim \mathfrak{K}$. (In this work, all simplicial complexes will be pure.) Note that a 1-dimensional simplicial complex can be identified with an ordinary graph. We say that $\mathfrak{K}$ is of *Δ-bounded degree* if every vertex participates in at most Δ maximal faces; and, we say that $\mathfrak{K}$ is *k-partite* if there is a partition of $\mathfrak{K}(0)$ into k parts such that each face has intersection size at most 1 with each part. (Pure $(d + 1)$-partite complexes are sometimes called *balanced*, or *numbered*.)

The *link* of a face $\sigma \in \mathfrak{K}$ is the simplicial complex $\mathrm{Link}_\sigma(\mathfrak{K}) = \{\tau \setminus \sigma : \tau \in \mathfrak{K}, \sigma \subseteq \tau\}$. In particular, the link of the (-1)-dimensional face $\emptyset$ is $\mathfrak{K}$. For a pure d-dimensional complex $\mathfrak{K}$, we define the *1-skeleton* of $\mathfrak{K}$ to be the *multigraph* on vertex set $\mathfrak{K}(0)$ in which $j, k \in \mathfrak{K}(0)$ are connected by a number of edges equal to the number of d-dimensional faces containing $\{j, k\}$. We will say that $\mathfrak{K}$ is *connected* if its 1-skeleton is a connected (multi)graph. Finally, for a face σ, we introduce the notation K_σ for the 1-skeleton of $\mathrm{Link}_\sigma(\mathfrak{K})$, and we will write $\lambda_2(K_\sigma)$ for the second largest eigenvalue of the standard random walk matrix of K_σ. (This refers to the walk on the vertices of K_σ in which a random out-edge is taken at each step.)

By now, the most common definition of expansion for HDXs is probably the following:

Definition 6.1.1. ([Opp18, KO18].) A d-dimensional pure simplicial complex $\mathfrak{K}$ is a *λ-spectral HDX* (also known as *λ-link* or *λ-local-spectral* HDX) if $\lambda_2(K_\sigma) \le \lambda$ for all faces σ of dimension at most $d - 2$.

(Note that the $d = 1$ case yields the usual notion of a λ-expander graph, one in which the second eigenvalue of the random walk matrix is at most λ.) The trickling down theorem [Opp18] essentially shows that a d-dimensional complex is an HDX provided the links of its $(d - 2)$-dimensional faces are λ-expander graphs for $\lambda < \frac{1}{d}$:

Theorem 6.1.2. ([Opp18].) *Let $\mathfrak{K}$ be a d-dimensional pure simplicial complex in which $\mathrm{Link}_\sigma(\mathfrak{K})$ is connected for all $\sigma \in \mathfrak{K}(i)$, $i \le d - 2$. Further suppose that $\lambda_2(K_\sigma) \le \gamma \le \frac{1}{d}$ for all $\sigma \in \mathfrak{K}(d - 2)$. Then $\mathfrak{K}$ is a $\left(\frac{\gamma}{1-(d-1)\gamma}\right)$-spectral HDX.*

(We remark that in the case $d = 2$, if $\mathfrak{K}$ is a Cayley graph then the conclusion of this theorem can be improved by a factor of $2/\sqrt{3}$; see [Żuk03].)

The objects we seek are (highly symmetric versions of) the following:

Definition 6.1.3. A *d-dimensional, Δ-bounded degree, λ-spectral HDX family* is a sequence $(\mathfrak{K}_n)_{n\in\mathbb{N}}$ of pure d-dimensional, Δ-bounded degree complexes, with $\mathfrak{K}_n$ having some $n' = \Theta(n)$ vertices, such that $\mathfrak{K}_n$ is a λ-spectral HDX for sufficiently large n. We also say the family is *explicit* if there is a $\mathrm{poly}(n)$-time algorithm for computing the description of $\mathfrak{K}_n$, and *strongly explicit* if there is a $\mathrm{polylog}(n)$-time algorithm. (See the proof of Theorem 6.2.6, item 1 for more details.)

6.1.2 Coset complexes

The following notion has been studied since at least the 1950 PhD thesis of Lannér [Lan50]:

Definition 6.1.4. Let G be a finite group and let $\mathcal{H} = (H_1, \dots, H_{d+1})$ be a sequence of subgroups. The associated *coset complex* $\mathcal{CC}(G;\mathcal{H})$ is the pure d-dimensional, $(d+1)$-partite simplicial complex with vertices being the cosets $\bigsqcup_i G/H_i$, and with maximal faces $\{gH_1, \dots, gH_{d+1} : g \in G\}$. Equivalently, a set of cosets forms a face if all cosets have an element in common.

Some well-studied instances of coset complexes are *Coxeter complexes* and *Tits buildings* [Bjö84].

Definition 6.1.5. The ith part of the $(d+1)$-partite coset complex $\mathcal{CC}(G;\mathcal{H})$ is the coset G/H_i, and the *type* of a face σ refers to the subset of parts $[d+1]$ to which its vertices belong.

The group G naturally acts on $\mathcal{CC}(G;\mathcal{H})$ by left-multiplication, and it is easy to see the following:

Fact 6.1.6. The action of G on the $(d+1)$-partite complex $\mathcal{CC}(G;\mathcal{H})$ is *type-preserving* (it does not change the type of any face), and transitive on the maximal faces. Moreover, the action is simply transitive if $H_1 \cap H_2 \cap \cdots \cap H_{d+1} = \{1\}$.

(In fact, Lannér [Lan50] showed that whenever there is a G-action on some $(d+1)$-partite complex that is type-preserving and transitive on maximal faces, then the complex must be of the form $\mathcal{CC}(G;\mathcal{H})$ for some subgroups $H_1, \dots, H_{d+1}$.)

We can also easily understand the connectivity and link structure of coset complexes, as the following facts show.

Definition 6.1.7. Given $\mathcal{H}$ and $T \subseteq [d+1]$ we write $H_T = \bigcap_{i\in T} H_i$, with the convention that $H_\emptyset = \langle H_1, \dots, H_{d+1}\rangle$, the subgroup of G generated by $\mathcal{H}$.

The following facts are easy to prove:

Fact 6.1.8. ([AH93].) $\mathcal{CC}(G;\mathcal{H})$ is connected if and only if $H_\emptyset = G$.

Fact 6.1.9. ([Gar79, p. 13], [HS19].) Let σ be a face in $\mathcal{CC}(G;\mathcal{H})$ of type $T \neq \emptyset$. Then the link of F is isomorphic to the coset complex $\mathcal{CC}(H_T; (H_{T\cup\{i\}} : i \notin T))$.

Note that Fact 6.1.9 says that, up to isomorphism, the link of a face only depends on its type. This will help us apply Theorem 6.1.2, as we will only have to consider a small number of cases. Finally we quote another easy-to-prove fact from Kaufman and Oppenheim, which we can use to pass between the (very slightly different) different universal and adjoint Chevalley groups:

Fact 6.1.10. ([KO18, essentially Prop. 2.12].) Let $\overline{\mathfrak{K}} = \mathcal{CC}(\overline{G};\overline{\mathcal{H}})$ be a coset complex with $\overline{\mathcal{H}} = (\overline{H}_1, \dots, \overline{H}_{d+1})$, suppose $Z \lhd \overline{G}$ is a normal subgroup (e.g., if Z is the center of $\overline{G}$), and suppose that $Z \cap \overline{H}_i = \{1\}$ for all $i \in [d+1]$. Then for $G = \overline{G}/Z$ and $\mathcal{H} = (H_1, \dots, H_{d+1})$, where $H_i = \overline{H}_i Z/Z$, the coset complex $\mathfrak{K} = \mathcal{CC}(G;\mathcal{H})$ is "covered" by $\overline{\mathfrak{K}}$, and the following property holds: every link in $\mathfrak{K}$ of type $T \neq \emptyset$ is isomorphic to every link of type T in $\overline{\mathfrak{K}}$

Remark 6.1.11. A consequence of Fact 6.1.10 is that if $\overline{\mathfrak{K}}$ is a Δ-bounded degree, λ-spectral HDX, then so too is $\mathfrak{K}$; moreover, provided $H_1 Z \cap H_2 Z \cap \cdots \cap H_{d+1} Z = Z$, the group G acts simply transitively on the maximal faces of $\mathfrak{K}$. We note that our complexes satisfy this condition in Observation 6.2.17.

6.1.3 Root systems

Killing and Cartan [Car94] classified simple Lie algebras over $\mathbb{C}$ via root systems:

Definition 6.1.12. A *(reduced) root system* of rank d is a finite set Φ of nonzero vectors spanning a d-dimensional real vector space such that for each $\alpha \in \Phi$:

- Φ is closed under w_α, where w_α is the reflection through the hyperplane orthogonal to α;
- $w_\alpha(\beta) - \beta$ is an integer multiple of α for all $\beta \in \Phi$;
- for $\lambda \in \mathbb{R}$ we have $\lambda\alpha \in \Phi$ (if and) only if $\lambda \in \{\pm 1\}$.

The root system Φ is *irreducible* if it cannot be written as $\Phi_1 \sqcup \Phi_2$ with Φ_1, Φ_2 nonempty and lying in orthogonal subspaces. Root system Φ' is said to be *isomorphic* to Φ if there is bijection between them that preserves inner products up to a fixed positive scalar multiple.

Figure 6.1 shows the three non-isomorphic rank-2 root systems (all of which are irreducible). The irreducible root systems have been completely classified:

Notation 6.1.13. Up to isomorphism, the irreducible root systems are classified as the families A_d $(d \geq 1)$, B_d $(d \geq 2)$, C_d $(d \geq 3)$, D_d $(d \geq 4)$, and the exceptional systems G_2, F_4, E_6, E_7, E_8. In all cases, the subscript gives the dimension of the root system. For explicit descriptions of these root systems, see e.g. [Car89, Sec. 3.6].

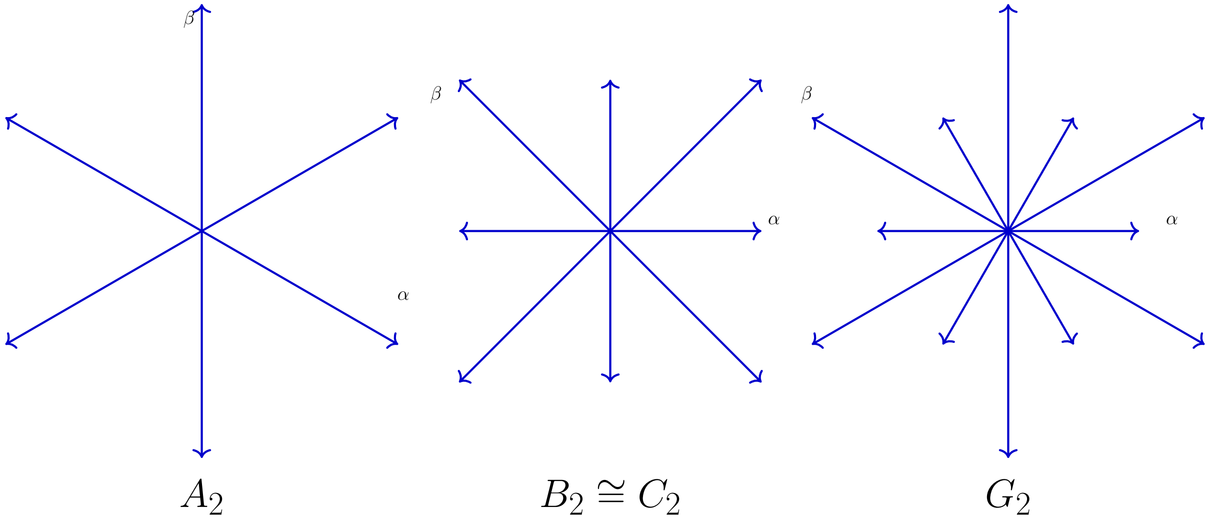

Figure 6.1: The rank 2 (irreducible) root systems, with a simple set $\{\alpha, \beta\}$ shown.

Remark 6.1.14. The restriction of a root system to a subspace is also a root system. Thus if Φ is a root system containing roots α, β, and $\alpha + \beta$, then the restriction of Φ to the subspace spanned by α and β must be (isomorphic to) A_2, B_2, or G_2. In fact, since G_2 is the only irreducible root system containing vectors at an angle of 30° (see, e.g., [Car89, Sec. 3.6]), an irreducible root system containing G_2 as a subsystem must in fact be isomorphic to G_2.

Let us now record a handy fact involving the inner product $\alpha \cdot \beta$ of two roots:

Fact 6.1.15. ([Hum72, p. 45, Lem. 9.4].) Let α, β be roots. If $\alpha \cdot \beta < 0$ then $\alpha + \beta \in \Phi \cup \{0\}$, and if $\alpha \cdot \beta > 0$ then $\alpha - \beta \in \Phi \cup \{0\}$.

This fact can be used to prove another simple result (which is surely well known, though we could not find a reference):

Fact 6.1.16. Let Φ be an irreducible root system of rank at least 2, and let $\alpha \in \Phi$. Then α is the sum of two other roots.

Proof. We claim there must exist a root $\beta \neq \pm\alpha$ with $\alpha \cdot \beta \neq 0$. Otherwise, every root is either orthogonal to α or parallel to α, meaning Φ is either irreducible or of rank 1. We may assume $\alpha \cdot \beta > 0$, by replacing β by the root $-\beta$, if necessary. Thus Fact 6.1.15 tells us that $\alpha - \beta \in \Phi$. But now $\alpha - \beta$ and β are roots summing to α. ■

We now discuss "simple" subsets of roots:

Definition 6.1.17. Let Φ be a root system spanning $\mathbb{R}^d$. A set of roots $\Pi = \{\alpha_1, \dots, \alpha_d\} \subseteq \Phi$ is called *simple* (or a *base*) if it is a basis for $\mathbb{R}^d$, and every root $\gamma \in \Phi$ may be expressed as

$$\gamma = n_1\alpha_1 + \cdots + n_d\alpha_d$$

either with $n_1, \dots, n_d \in \mathbb{N}$ or with $-n_1, \dots, -n_d \in \mathbb{N}$. (Since Π is a basis, there is a unique such expression.) In the former case, γ is called a *positive root*; in the latter case, a *negative root*. One also defines the *height* of γ (with respect to Π, or more generally a set of linearly independent roots whose span contains γ), denoted $\mathrm{ht}_\Pi(\gamma)$, to be $\sum_{i=1}^d |n_i|$.

(In Figure 6.1, each root system has labeled a simple set $\{\alpha, \beta\}$.)

In a certain sense, up to symmetries there is a unique choice of simple roots for a given root system:

Fact 6.1.18. ([Car89, Prop. 2.1.2, Cor 2.2.5].) Every root system has a set of simple roots. Further, for any two simple sets, there is a unique reflection w_α mapping one to the other.

Definition 6.1.19. For any subset $\Psi \subseteq \Phi$, we write $\Psi^+ = \Phi \cap \{\sum_{\alpha \in \Psi} n_\alpha \alpha : n_\alpha \in \mathbb{N}\}$, and $\Psi^- = -\Psi^+$.

Fact 6.1.20. Let $\Psi \subseteq \Phi$ be a set of linearly independent roots. Then there is set of simple roots Π of Φ where $\Psi \subseteq \Pi^+$.

Proof. We can always find a hyperplane H not containing any root, and where all of Ψ is contained on one side of H. Then by [Hal03, Thm. 8.16], there is a set of simple roots Π such that the roots in Φ on this side of H are positive with respect to Π. ■

The following fact is very similar to a standard one about root systems, but it is usually only stated when $\{\alpha, \dots, \alpha_m\}$ form a simple set (see, e.g., [Car89, Lem. 3.6.2]):

Fact 6.1.21. Let $A = \{\alpha_1, \dots, \alpha_\ell\} \subseteq \Phi$ be *any* set of roots, and suppose that $\gamma = \sum_{i=1}^\ell n_i\alpha_i \in \Phi$ for $n_1, \dots, n_\ell \in \mathbb{N}$. Then we may express $\gamma = \sum_{j=1}^m \alpha_{i_j}$ for certain $i_1, \dots, i_m \in [\ell]$ in such a way that all the prefix-sums $\sum_{j=1}^k \alpha_{i_j}$ $(1 \leq k \leq \ell)$ are in Φ.

Proof. By induction, it suffices to show that if γ is not already in A, then there exists $i_0 \in [\ell]$ with $n_{i_0} > 0$ such that $\gamma - \alpha_{i_0} \in \Phi$. To do this, note that $0 < \gamma \cdot \gamma = \sum_{i=1}^{\ell} n_i(\gamma \cdot \alpha_i)$, and since the n_i's are nonnegative we must have $\gamma \cdot \alpha_{i_0} > 0$ for (at least) one i_0. By Fact 6.1.15 we conclude that $\gamma - \alpha_{i_0} \in \Phi \cup \{0\}$, and the case $\gamma - \alpha_{i_0} = 0$ (i.e., $\gamma = \alpha_{i_0}$) is impossible because γ is assumed not already in A. ■

Finally, we need the following known fact [Hil16]:

Fact 6.1.22. Let Φ be an irreducible root system with simple roots $\Pi = \{\alpha_1, \dots, \alpha_d\}$. Then $\sum_{i=1}^{d} \alpha_i \in \Phi$.

6.1.4 Chevalley groups

We may now define the Chevalley groups, via the *Steinberg presentation* (see, e.g., [Car89, Thm. 12.1.1]).

Definition 6.1.23. Corresponding to any irreducible root system Φ of rank at least 2, and any finite field $\mathbb{F}$, there is an associated *universal (or simply connected) Chevalley group*, denoted $\overline{\mathrm{G}}(\Phi, \mathbb{F})$. Abstractly, it is generated by symbols $x_\alpha(t)$ for $\alpha \in \Phi$ and $t \in \mathbb{F}$, subject to the relations

$$\begin{aligned} x_\alpha(t)x_\alpha(u) &= x_\alpha(t+u) \\ [x_\alpha(t), x_\beta(u)] &= \prod_{i,j>0} x_{i\alpha+j\beta}(C_{ij}^{\alpha,\beta} t^i u^j) \qquad (\text{for } \alpha+\beta \neq 0) \\ h_\alpha(t)h_\alpha(u) &= h_\alpha(tu) \quad (\text{for } tu \neq 0), \\ \text{where} \qquad h_\alpha(t) &= n_\alpha(t)n_\alpha(-1) \\ \text{and} \qquad n_\alpha(t) &= x_\alpha(t)x_{-\alpha}(-t^{-1})x_\alpha(t). \end{aligned}$$

The second relation above is the *Chevalley commutator formula*, and it is elaborated upon in Theorem 6.1.27 below.

Definition 6.1.24. Let $\mathrm{Z}(\Phi, \mathbb{F})$ denote the center of $\overline{\mathrm{G}}(\Phi, \mathbb{F})$. The *adjoint Chevalley group*, which we denote by $\mathrm{G}(\Phi, \mathbb{F})$, is the quotient $\overline{\mathrm{G}}(\Phi, \mathbb{F})/\mathrm{Z}(\Phi, \mathbb{F})$. In all cases, $\mathrm{Z}(\Phi, \mathbb{F})$ is a constant-sized subgroup (of size $d+1$ for $\Phi = A_d$, and of size at most 4 otherwise).[4] It is generated by certain products $\prod_{\alpha \in \Pi} h_\alpha(t_\alpha)$ (i.e., diagonal matrices in the matrix realizations), where Π is a simple set of roots and the t_α's are roots of unity in $\mathbb{F}$.

Remark 6.1.25. The Classification of Finite Simple Groups [Asc04] states that as $\mathbb{F}$ ranges over all finite fields, the adjoint Chevalley groups (excluding $\mathrm{G}(A_1, \mathbb{F}_2)$, $\mathrm{G}(A_1, \mathbb{F}_3)$, $\mathrm{G}(B_2, \mathbb{F}_2)$, $\mathrm{G}(G_2, \mathbb{F}_2)$, but including the "twisted" versions, which we do not discuss in this work) constitute the finite simple groups, together with the cyclic, alternating, and sporadic simple groups.

Although, strictly speaking, it is the adjoint Chevalley groups that are the simple ones, it is more convenient to work with the very slightly larger universal Chevalley groups. If one wants to precisely fulfill the goal concerning simple (adjoint) Chevalley groups described in Section 6.0.1, one may use do so by appealing to Remark 6.1.11. But henceforth we work exclusively with the universal Chevalley groups, and we will drop the adjective "universal".

[4] Specifically, it is isomorphic to $\mathbb{Z}_{d+1}$ when $\Phi = A_d$, to $\mathbb{Z}_2$ when $\Phi \in \{B_d, C_d, E_7\}$, to $\mathbb{Z}_4$ or $\mathbb{Z}_2 \times \mathbb{Z}_2$ when $\Phi = D_d$ (for odd, even d respectively), to $\mathbb{Z}_3$ when $\Phi = E_7$, and is trivial otherwise. [Ste16, Sec. 3.3].

Although we have defined the Chevalley groups abstractly, we have [Ste16, Sec. 3.3] the isomorphisms with classical groups shown in Table 6.1, for the "classical" root systems of types A, B, C, and D.

Type of Φ	$\mathrm{G}(\Phi,\cdot)$	$\overline{\mathrm{G}}(\Phi,\cdot)$
A_d	PSL_{d+1}	SL_{d+1}
B_d	SO_{2d+1}	Spin_{2d+1}
C_d	PSp_{2d}	Sp_{2d}
$D_{2\ell}$	$\mathrm{PSO}_{4\ell}$	$\mathrm{Spin}_{4\ell}$
$D_{2\ell+1}$	$\mathrm{PSO}_{4\ell+2}$	$\mathrm{Spin}_{4\ell+2}$

Table 6.1: The Chevalley groups corresponding to classical root systems.

Identifications of the root elements $x_\alpha(t)$ of the Chevalley groups of classical type as elements of the corresponding matrix groups can be found in [Car89, Sec. 11.3], and matrix realizations for the exceptional Chevalley groups can be found in [HRT01].

As we discuss in Section 6.1.5, we have

$$n := |\overline{\mathrm{G}}(\Phi,\mathbb{F})| = |\mathbb{F}|^{\Theta(1)} = \exp(\Theta(m))$$

for fixed p and Φ; indeed, an exact formula for $|\overline{\mathrm{G}}(\Phi,\mathbb{F})|$ is known, and one can compute within $\overline{\mathrm{G}}(\Phi,\mathbb{F})$ (and $\mathrm{G}(\Phi,\mathbb{F})$) in $\mathrm{poly}(m) = \mathrm{polylog}(n)$ time (see Section 6.1.5).

Remark 6.1.26. From the first relation of Definition 6.1.23, it follows that the subgroup $\langle x_\alpha(r) : r \in \mathbb{F}\rangle$ of $\overline{\mathrm{G}}(\Phi,\mathbb{F})$ is isomorphic to the additive group of $\mathbb{F}$. This subgroup is called the *root subgroup* associated to α.

The second relation in Definition 6.1.23 will be used to give explicit descriptions of the links in our constructions, so we elaborate on it here.

Theorem 6.1.27. *The* Chevalley commutator formula *asserts that within* $\overline{\mathrm{G}}(\Phi,\mathbb{F})$ *and* $\mathrm{G}(\Phi,\mathbb{F})$*, if* $\alpha,\beta\in\Phi$ *with* $\alpha+\beta\neq 0$*, and* $t,u\in\mathbb{F}$*, then*

$$[x_\alpha(t), x_\beta(u)] = \prod_{\substack{i,j\in\mathbb{Z}_+\\ i\alpha+j\beta\in\Phi}} x_{i\alpha+j\beta}(C_{ij}^{\alpha,\beta} t^i u^j)$$

for certain structure constants $C_{ij}^{\alpha,\beta}\in\{\pm1,\pm2,\pm3\}$ *that can be found in, e.g., [Car89, Sec. 5.2]. Here the product above is taken in order of increasing* $i+j$*.*[5] *In addition, the structure constants only depend on the set* $\{(i,j) : i\alpha+j\beta\in\Phi\}$.

Remark 6.1.28. In particular, the commutator formula implies that if $\alpha+\beta\notin\Phi\cup\{0\}$, then $[x_\alpha(r), x_\beta(r)] = 1$.

Remark 6.1.29. The constants $C_{ij}^{\alpha,\beta}$ are determined uniquely by Φ up to signs. Different signs can arise from different choices of a *Chevalley basis*. The resulting groups are isomorphic, however. See [Car89, Prop. 4.2.2] and the preceding discussion.

Although not strictly necessary for our work, we give explicit structure constants in the following description of the commutator formula for root systems of rank 2:

[5] Ties may be broken arbitrarily, as it turns out that elements with equal $i+j$ commute.

Proposition 6.1.30. ([Hum95, Sec. 33.3–33.5].) *Let Φ be one of A_2, B_2, or G_2 and let $t, u \in \mathbb{F}$. Then:*[6]

- *If $\Phi = A_2$ with positive roots $\alpha, \beta, \alpha+\beta$, then*

$$[x_\alpha(t), x_\beta(u)] = x_{\alpha+\beta}(tu).$$

- *If $\Phi = B_2$ with positive roots $\alpha, \beta, \alpha+\beta, 2\alpha+\beta$, then*

$$\begin{aligned}[x_\beta(t), x_\alpha(u)] &= x_{\alpha+\beta}(tu)x_{2\alpha+\beta}(t^2u)\\ [x_{\alpha+\beta}(t), x_\alpha(u)] &= x_{2\alpha+\beta}(2tu).\end{aligned}$$

- *If $\Phi = G_2$ with positive roots $\alpha, \beta, \alpha+\beta, 2\alpha+\beta, 3\alpha+\beta, 3\alpha+2\beta$, then*

$$\begin{aligned}[x_\beta(t), x_\alpha(u)] &= x_{\alpha+\beta}(tu)x_{2\alpha+\beta}(tu^2)x_{3\alpha+\beta}(tu^3)x_{3\alpha+2\beta}(-t^2u^3)\\ [x_{\alpha+\beta}(t), x_\alpha(u)] &= x_{2\alpha+\beta}(2tu)x_{3\alpha+\beta}(3tu^2)x_{3\alpha+2\beta}(-3t^2u)\\ [x_{2\alpha+\beta}(t), x_\alpha(u)] &= x_{3\alpha+\beta}(3tu)\\ [x_{3\alpha+\beta}(t), x_\beta(u)] &= x_{3\alpha+2\beta}(tu)\\ [x_{2\alpha+\beta}(t), x_{\alpha+\beta}(u)] &= x_{3\alpha+2\beta}(-3tu).\end{aligned}$$

Finally, we will require two more key facts:

Proposition 6.1.31. ([Ste16, Lem. 17]. *In the Chevalley group $\overline{\mathrm{G}}(\Phi, \mathbb{F})$, suppose $S \subseteq \Phi$ is a set of roots with the following two properties: (i) $\alpha, \beta \in S$ and $\alpha+\beta \in \Phi$ implies $\alpha+\beta \in S$; (ii) $\alpha \in S$ implies $-\alpha \notin S$. Then each element of the subgroup $\langle x_\alpha(t) : \alpha \in S, t \in \mathbb{F}\rangle$ can be expressed uniquely as $\prod_{\alpha \in S} x_\alpha(t_\alpha)$ for some $t_\alpha \in \mathbb{F}$, where the product is taken in some fixed order (and this is true for any fixed ordering of S for the product).*

Proposition 6.1.32. *Let Π be a set of simple roots, and define the two subgroups $U^\pm = \langle x_\alpha(t) : \alpha \in \Pi^\pm\rangle$. Then $U^+ \cap U^- = \{1\}$.*

Proof. By [Ste16, Lem. 18, Cor. 3], $\overline{\mathrm{G}}(\Phi, \mathbb{F})$ can be realized as a group of matrices over $\mathbb{F}$ where the subgroup U^+ is upper-unitriangular and U^- is lower-unitriangular. The proposition follows. ■

6.1.5 Computation within the Chevalley groups

Given field $\mathbb{F} = \mathbb{F}_q = \mathbb{F}_{p^m}$ and root system Φ of rank d, let us treat Φ and p as fixed, and $m \to \infty$ as an asymptotically growing parameter. Here we recap the known facts that the Chevalley group $\overline{\mathrm{G}}(\Phi, \mathbb{F})$ has order $n = \exp(\Theta(m))$ and that one can compute within $\overline{G}$ in deterministic $\mathrm{poly}(m) = \mathrm{polylog}(n)$ time. (The same is true for the adjoint Chevalley group $\mathrm{G}(\Phi, \mathbb{F})$.)

First, field arithmetic is efficient, thanks to Shoup:

[6] For G_2 we fix the signs implemented in the GAP [GAP21] package Unipot [HH18], which we used for calculations in Remark 6.2.24.

Theorem 6.1.33. ([Sho90].) *For a fixed prime p, there is an deterministic $\mathrm{poly}(m)$-time algorithm for finding an irreducible $f \in \mathbb{F}_p[x]$ of degree d, and thereby "constructing" the field $\mathbb{F} = \mathbb{F}_q = \mathbb{F}_{p^m}$. The elements of $\mathbb{F}$ are encoded by bit-strings of length $\Theta(m)$, and field operations may be computed in deterministic $\mathrm{poly}(m)$ time — this includes computing all kth roots of unity in $\mathrm{poly}(k, m)$ time.*

Next, we note that there is an easy-to-compute formula for the order of a given Chevalley group:

Theorem 6.1.34. *For Φ of rank d, the order of the group $\overline{\mathrm{G}}(\Phi, \mathbb{F})$ is of the form $q^{\Theta(d^2)} = p^{\Theta(d^2 m)}$, where the constant hidden in the $\Theta(\cdot)$ depends only on Φ. Moreover there is a precise formula for $|\overline{\mathrm{G}}(\Phi, \mathbb{F})|$ that can easily be computed in $\mathrm{poly}(d, \log p, m)$ time; see, e.g. [Ste16, Thm. 25]. (All of this is also true of $\mathrm{G}(\Phi, \mathbb{F})$.)*

Finally, we appeal to the work of Cohen, Murray, and Taylor [CMT04] to show that one can efficiently construct and compute within Chevalley groups:

Theorem 6.1.35. ([CMT04], see especially Sec. 8.1.) *For Φ of rank d and $\mathbb{F} = \mathbb{F}_{p^m}$, there is a canonical representation ("Bruhat normal form") for each element of $\overline{\mathrm{G}}(\Phi, \mathbb{F})$, encoded by a bit-string of length $\mathrm{poly}(d, \log q, m)$. One can pass between this form, a natural matrix representation, and an expression in the Steinberg presentation — and also compute group products and inverses — via deterministic $\mathrm{poly}(d, \log q, m)$-time algorithms. (Since kth roots of unity can also be computed efficiently (Theorem 6.1.33), the $O(d)$-size center Z of $\overline{\mathrm{G}}(\Phi, \mathbb{F})$ can also be constructed efficiently, and hence this whole theorem is also true for $\mathrm{G}(\Phi, \mathbb{F})$.)*

6.2 The construction

For the rest of this chapter we fix a field $\mathbb{F}$ of size p^m where $p > 3$, an irreducible root system Φ of rank at least 2, and a set of simple roots $\Pi = \{\alpha_1, \dots, \alpha_d\} \subseteq \Phi$. With this in mind, $x_\alpha(t)$ refers to the corresponding root element of $\overline{\mathrm{G}}(\Phi, \mathbb{F})$.

Definition 6.2.1. For $S \subseteq \Phi$ and $d \in \mathbb{N}$, let $X_{S,d} = \langle x_\alpha(t) : \alpha \in S, t \in \mathbb{F}, \deg(t) \le d \rangle$. For shorthands we write $X_S = X_{S,1}$ and also $X_{\alpha,d} = X_{\{\alpha\},d}$.

Definition 6.2.2. Recalling $\Pi = \{\alpha_1, \dots, \alpha_d\}$, we define $\mathcal{S}$ to be the following particular set of roots:

$$\mathcal{S} = \Pi \cup \{-(\alpha_1 + \cdots + \alpha_d)\}\,. \tag{6.1}$$

(The last of these is a root by by Fact 6.1.22.)

Remark 6.2.3. Since Π is a basis, it follows that every subset of $\mathcal{S}$ of cardinality d is linearly independent.

Definition 6.2.4. For each $\alpha \in \mathcal{S}$, we introduce the following subgroup of $\overline{\mathrm{G}}(\Phi, \mathbb{F})$:

$$H_\alpha = X_{\mathcal{S} \setminus \{\alpha\}}.$$

Finally, we can introduce our coset complex:

Definition 6.2.5. $\mathfrak{K} = \mathfrak{K}_m := \mathcal{CC}(\overline{\mathrm{G}}(\Phi, \mathbb{F}); (H_\alpha)_{\alpha \in \mathcal{S}})$.

Theorem 6.2.6. *For $d = \mathrm{rank}(\Phi)$, it holds that $\mathfrak{K}$ is a d-dimensional pure simplicial complex, where:*

1. $|\mathfrak{K}(0)| = p^{\Theta(m)}$, *where the constant hidden by* $\Theta(\cdot)$ *depends only on* Φ; *moreover, the family that arises as* $m \to \infty$ *is strongly explicit.*
2. *Every vertex participates in at most* $\Delta = \Delta(\Phi, p) = p^{\Theta(1)}$ *maximal faces, where the* $\Theta(\cdot)$ *constant (independent of* m*) depends only on* Φ *(indeed, it is* $\Theta(d^2)$*).*
3. *If* $p > 3$,[7] *then* $\mathrm{Link}_\sigma(\mathfrak{K})$ *is connected for all* $\sigma \in \mathfrak{K}(i)$, $i \leq d-2$.
4. *If* $\Phi \neq G_2$ *and* $p > 2$, *then for all* $\sigma \in \mathfrak{K}(d-2)$ *it holds that* K_σ *is a* p^2*-regular bipartite graph with* $\lambda_2(K_\sigma) \leq \sqrt{2/p}$.
5. $\overline{\mathrm{G}}(\Phi, \mathbb{F})$ *acts simply transitively on the maximal faces of* $\mathfrak{K}$ *(and this is also true if one constructs* $\mathfrak{K}$ *from* $\mathrm{G}(\Phi, \mathbb{F})$ *rather than* $\overline{\mathrm{G}}(\Phi, \mathbb{F})$*).*

By Theorem 6.1.2, we conclude our final goal:

Corollary 6.2.7. *Fixing* $\Phi \neq G_2$ *of rank* $d \geq 2$, $p > 3$ *prime, and taking* $m \to \infty$, *the sequence* $(\mathfrak{K}_m)$ *forms a strongly explicit* d*-dimensional,* Δ*-bounded degree (*$\Delta = p^{\Theta(d^2)}$*),* λ*-spectral HDX family, where*

$$\lambda \leq \frac{1}{\sqrt{p/2} - d + 1}.$$

(Hence for large p*, we have* $\lambda \sim 1/\Delta^{\Theta(1/d^2)}$*.) Moreover, the universal Chevalley group* $\overline{\mathrm{G}}(\Phi, \mathbb{F})$ *acts simply transitively on* $\mathfrak{K}_m$*'s maximal faces.*

We add that the results all remain true if uses the (simple) adjoint Chevalley groups $\mathrm{G}(\Phi, \mathbb{F})$ in place of $\overline{\mathrm{G}}(\Phi, \mathbb{F})$.

6.2.1 Global connectivity of the coset complex

The main goal of this section is to show that the subgroups H_α for $\alpha \in \mathcal{S}$ generate $\overline{\mathrm{G}}(\Phi, \mathbb{F})$. By Fact 6.1.8, this is necessary to ensure that the 1-skeleton of $\mathfrak{K}$ is connected.

Theorem 6.2.8. *Let* $S \subseteq \Phi$ *be a subset of* $\mathrm{rank}(\Phi) + 1$ *roots where* $S^+ = \Phi$. *Then* $X_S = \overline{\mathrm{G}}(\Phi, \mathbb{F})$.

The particular set of roots $\mathcal{S}$ we selected in Equation (6.1) has the desired property, as the following shows:

Proposition 6.2.9. *For* $\mathcal{S}$ *as in* Equation (6.1) *we have* $\mathcal{S}^+ = \Phi$.

Proof. We have $\mathcal{S}^+ \supseteq \Pi^+$, and so $\mathcal{S}^+$ certainly contains all positive roots in Φ (recall Definition 6.1.17). It remains to show that $\mathcal{S}^+$ contains each negative root $\gamma \in \Phi$. Writing $\gamma = -n_1\alpha_1 - \cdots - n_d\alpha_d$, it follows that we can reexpress it as

$$\gamma = r(-(\alpha_1 + \cdots + \alpha_d)) + r_1\alpha_1 + \cdots + r_d\alpha_d$$

for a sufficiently large positive integer r, and positive integers $r_1, \ldots, r_d$. Thus indeed $\gamma \in \mathcal{S}^+$. ∎

Example 6.2.10. A_d is the set of vectors $\{e_i - e_j, i \neq j\} \subseteq \mathbb{R}^d$. A set of simple roots is given by $\Pi = \{e_i - e_{i+1} : i \in [d]\}$; in this case $-\sum_{\alpha \in \Pi} \alpha = e_d - e_1$. It is straightforward to check that $S = \{e_i - e_{i+1} : i \in [d]\} \cup \{e_d - e_1\} \subseteq A_d$ satisfies the hypothesis of Theorem 6.2.8. This is the set of roots implicitly used in [KO18].

[7] Recall Footnote 1.

Remark 6.2.11. There are other choices of S besides our $\mathcal{S}$ from Equation (6.1) that satisfy the condition of Theorem 6.2.8. These can be used to obtain slightly different constructions. For example, referring to Figure 6.1 one see that in B_2 one can take $S = \{\alpha, \beta, -\beta - 2\alpha\}$, or in G_2 one can take $S = \{\alpha, \alpha + \beta, -2\alpha - \beta\}$.

We will require the following (presumably known) fact:

Lemma 6.2.12. *For $i, j, d_1, d_2 \in \mathbb{N}$ with $\mathrm{char}(\mathbb{F}) > \max(i, j)$, write $d = id_1 + jd_2$. Then*

$$\mathbb{F}[x]^{\leq d} = \mathrm{span}\{f^i g^j : f \in \mathbb{F}[x]^{\leq d_1},\ g \in \mathbb{F}[x]^{\leq d_2}\}.$$

where $\mathbb{F}[x]^{\leq k}$ represents the polynomials of degree at most k.

Proof. It suffices to establish that x^e is in the span, for any $e \leq d$. Express $e = a_1 + \cdots + a_i + b_1 + \cdots + b_j$, with each a being a natural number at most d_1 and each b being a natural number at most d_2. Now note that the monomial

$$x_{a_1} x_{a_2} \cdots x_{a_i} x_{b_1} x_{b_2} \cdots x_{b_j} \tag{6.2}$$

becomes equal to x^e if each indeterminate x_c is substituted with x^c. Next, we use the identity

$$x_{a_1} x_2 \cdots x_{a_i} = \frac{1}{i!} \sum_{s \in \{0,1\}^i} (-1)^{|s|+i} \left(\sum_{\ell=1}^{i} s_\ell x_{a_\ell} \right)^i,$$

with the constant $\frac{1}{i!}$ being sensible in the field $\mathbb{F}$ since $\mathrm{char}(\mathbb{F}) > i$. (This is the "higher order polarization identity", or Ryser's formula applied to the matrix where every row is $\begin{bmatrix} x_{a_1} & x_{a_2} & \cdots & x_{a_i} \end{bmatrix}$.) Multiplying this against the analogous identity with the b's (and using $\mathrm{char}(\mathbb{F}) > j$), we get that (6.2) can be expressed as a linear combination of multivariate polynomials $F^i G^j$, where each F is a linear combination of x_c's with $c \leq d_1$ and each G is a linear combination of x_c's with $c \leq d_2$. Now substituting $x_c = x^c$ yields the desired univariate expression for x^e. ■

A key goal now is to establish the below Lemma 6.2.13. We remark that several times it will use Lemma 6.2.12; in each application we will have "i" and "j" at most 3, less than $\mathrm{char}(\mathbb{F}) = p > 3$ as required.

Lemma 6.2.13. *Fix roots $\beta \neq -\alpha$ and any $d_1, d_2 \in \mathbb{N}$. Then*

$$\langle X_{\alpha,d_1}, X_{\beta,d_2} \rangle = \langle X_{i\alpha+j\beta, id_1+jd_2} : i, j \in \mathbb{N}, i\alpha + j\beta \in \Phi \rangle.$$

Proof. The inclusion $\subseteq$ is immediate by taking $(i, j) \in \{(1, 0), (0, 1)\}$, so it suffices to prove the reverse inclusion $\supseteq$. The case $\beta = \alpha$ is trivial, so we may assume that α, β span some 2-dimensional subspace H. Let $R \coloneqq \{i\alpha + j\beta \in \Phi : i, j \in \mathbb{N}\}$, a subset of the 2-dimensional root system $\Phi' = \Phi \cap H$. If $R = \{\alpha, \beta\}$ only then the lemma is immediate. Otherwise, R must also contain $\alpha + \beta$ (using Fact 6.1.21) and hence Φ' is isomorphic to A_2, B_2, or G_2 as explained in Remark 6.1.14. This allows us to classify the possibilities for R; with the assistance of Figure 6.1, we see there are four cases, namely $R = \{\alpha, \beta\} \cup R'$ for R' equal to...

1. $\{\alpha+\beta\}$, 2. $\{\alpha+\beta, 2\alpha+\beta\}$, 3. $\{\alpha+\beta, 2\alpha+\beta, \alpha+2\beta\}$, 4. $\{\alpha+\beta, 2\alpha+\beta, 3\alpha+\beta, 3\alpha+2\beta\}$.

In each case, we need to show for every $\gamma = i\alpha + j\beta \in R'$ that $x_\gamma(w) \in \langle X_{\alpha,d_1}, X_{\beta,d_2}\rangle$ for all $w \in \mathbb{F}$ of degree at most $d = id_1 + jd_2$. By virtue of Lemma 6.2.12 (and using $\mathrm{char}(\mathbb{F}) > 3 \geq i, j$), it suffices to show this for w's that are linear combinations of field elements of the form $t^i u^j$, where t has degree d_1 and u has degree d_2. Further, since $x_\gamma(r+s) = x_\gamma(r)x_\gamma(s)$, it suffices to handle w of the form ct^iu^j for arbitrary $c \in \mathbb{F}_p$. Finally, it suffices to handle just one specific $c \neq 0$, because if $x_\gamma(ct^iu^j)$ is in $\langle X_{\alpha,d_1}, X_{\beta,d_2}\rangle$ then so too is its kth power $x_\gamma(ct^iu^j)^k = x_\gamma(kct^iu^j)$, and kc varies over all $\mathbb{F}_p$ as k varies in $\mathbb{N}$. We will always use a c which is the product of structure constants $C^{\alpha',\beta'}_{i',j'}$, and such are never 0 in $\mathbb{F}_p$ because $1 \leq |C^{\alpha',\beta'}_{i',j'}| \leq 3 < p$.

Summarizing, for fixed $t, u \in \mathbb{F}$ of degree at most d_1, d_2 (respectively), it suffices to show the following in Cases 1–4: For each $\gamma = i\alpha + j\beta \in R'$ we have $x_\gamma(ct^iu^j) \in \langle x_\alpha(t), x_\beta(u)\rangle$ for some product of structure constants c.

Case 1: $R' = \{\alpha + \beta\}$. This case arises when $\Phi' = A_2$ and $\angle(\alpha,\beta) = 120°$, or when $\Phi' = B_2$ and α, β are short roots with $\angle(\alpha,\beta) = 90°$, or when $\Phi' = G_2$ and α,β are short roots with $\angle(\alpha,\beta) = 60°$. We handle $\gamma = \alpha + \beta$ via the commutator formula $[x_\alpha(t), x_\beta(u)] = x_{\alpha+\beta}(C^{\alpha,\beta}_{1,1}tu)$.

Case 2: $R' = \{\alpha + \beta, 2\alpha + \beta\}$, which arises for $\Phi' = B_2$. We first treat the root $\gamma = 2\alpha + \beta$. By the commutator formula we have

$$[[x_\alpha(t), x_\beta(u)], x_\alpha(t)] = [x_{\alpha+\beta}(C^{\alpha,\beta}_{1,1}tu)x_{2\alpha+\beta}(C^{\alpha,\beta}_{2,1}t^2u), x_\alpha(t)].$$

In this latter commutator we can delete $x_{2\alpha+\beta}(C^{\alpha,\beta}_{2,1}t^2u)$ because it commutes with the other two elements. (This is since no root is a nontrivial $\mathbb{N}$-linear combination involving $2\alpha + \beta$.) Thus

$$[[x_\alpha(t), x_\beta(u)], x_\alpha(t)] = [x_{\alpha+\beta}(C^{\alpha,\beta}_{1,1}tu), x_\alpha(t)] = x_{2\alpha+\beta}(C^{\alpha+\beta,\alpha}_{1,1}C^{\alpha,\beta}_{1,1}t^2u). \tag{6.3}$$

Thus $\gamma = 2\alpha + \beta$ is handled. As for $\gamma = \alpha + \beta$, the commutator formula gives

$$\begin{aligned}[x_\alpha(t), x_\beta(u)] \cdot x_{2\alpha+\beta}(-C^{\alpha,\beta}_{2,1}t^2u) &= x_{\alpha+\beta}(C^{\alpha,\beta}_{1,1}tu)x_{2\alpha+\beta}(C^{\alpha,\beta}_{2,1}t^2u) \cdot x_{2\alpha+\beta}(-C^{\alpha,\beta}_{2,1}t^2u)\\ &= x_{\alpha+\beta}(C^{\alpha,\beta}_{1,1}tu),\end{aligned}$$

and so $\gamma = \alpha + \beta$ is also handled (since we already know $x_{2\alpha+\beta}(-C^{\alpha,\beta}_{2,1}t^2u)$ is in $\langle x_\alpha(t), x_\beta(u)\rangle$ via Equation (6.3)).

Case 3: $R' = \{\alpha + \beta, 2\alpha + \beta, \alpha + 2\beta\}$. This case only arises for $\Phi' = G_2$. We start by treating $\gamma = 2\alpha + \beta$. We have

$$[[x_\alpha(t), x_\beta(u)], x_\alpha(t)] = [x_{\alpha+\beta}(C^{\alpha,\beta}_{1,1}tu)y, x_\alpha(t)] \quad \text{for } y = x_{2\alpha+\beta}(C^{\alpha,\beta}_{2,1}t^2u)x_{\alpha+2\beta}(C^{\alpha,\beta}_{3,1}tu^2),$$

and similar to Case 2 we can delete y from this commutator as it commutes with the other two elements (by virtue of the height of $2\alpha + \beta$ and $\alpha + 2\beta$). Hence

$$[[x_\alpha(t), x_\beta(u)], x_\alpha(t)] = [x_{\alpha+\beta}(C^{\alpha,\beta}_{1,1}tu), x_\alpha(t)] = x_{2\alpha+\beta}(C^{\alpha,\beta}_{1,1}C^{\alpha+\beta,\alpha}_{1,1}t^2u)$$

and we've handled $\gamma = 2\alpha + \beta$. The case of $\gamma = \alpha = 2\beta$ is similar. Finally the treatment of $\gamma = \alpha + \beta$ is similar to Case 2; it follows from

$$[x_\alpha(t), x_\beta(u)]x_{\alpha+2\beta}(-C^{\alpha,\beta}_{3,1}tu^2)x_{2\alpha+\beta}(C^{\alpha,\beta}_{2,1}t^2u) = x_{\alpha+\beta}(C^{\alpha,\beta}_{1,1}tu).$$

Case 4: $R' = \{\alpha+\beta, 2\alpha+\beta, 3\alpha+\beta, 3\alpha+2\beta\}$. This case only arises for $\Phi' = G_2$. To reduce clutter in this case, we will sometimes abbreviate $x_{i\alpha+j\beta}(ct^iu^j)$ to $x_{i\alpha+j\beta}$. We start with

$$[x_\alpha(t), x_\beta(u)] = x_{\alpha+\beta} \cdot x_{2\alpha+\beta} \cdot x_{3\alpha+\beta} \cdot x_{3\alpha+2\beta}, \tag{6.4}$$

which implies

$$[[x_\alpha(t), x_\beta(u)], x_\beta(u)] = [x_{\alpha+\beta} \cdot x_{2\alpha+\beta} \cdot x_{3\alpha+\beta}, x_\beta],$$

where we deleted the $x_{3\alpha+2\beta}$ element since it commutes with everything else. Now since x_β commutes with $x_{\alpha+\beta}$ and $x_{2\alpha+\beta}$, we get

$$[x_{\alpha+\beta} \cdot x_{2\alpha+\beta} \cdot x_{3\alpha+\beta}, x_\beta] = [x_{3\alpha+\beta}, x_\beta] = x_{3\alpha+2\beta} = x_{3\alpha+2\beta}(C_{1,1}^{3\alpha+\beta,\beta} C_{3,1}^{\alpha,\beta} t^3u^2),$$

where in the last step we explicitly wrote in the argument to $x_{3\alpha+2\beta}$ that arises. Thus we have handled $\gamma = 3\alpha + 2\beta$. Taking care of $\gamma = 3\alpha + \beta$ is somewhat more tedious. Considerations similar to the above lead us to

$$[[x_\alpha(t), x_\beta(u)], x_\alpha(t)] = [x_{\alpha+\beta} \cdot x_{2\alpha+\beta}, x_\alpha],$$

which in turn equals

$$x_{2\alpha+\beta}(-C_{2,1}^{\alpha,\beta}t^2u) \cdot [x_{\alpha+\beta}, x_\alpha] \cdot x_{2\alpha+\beta}(C_{2,1}^{\alpha,\beta}t^2u) \cdot [x_{2\alpha+\beta}, x_\alpha], \tag{6.5}$$

where we explicitly wrote in the arguments to $x_{2\alpha+\beta}$ that arise. We now observe that when the commutator rule is twice applied in the above, the resulting elements are $x_{2\alpha+\beta} \cdot x_{3\alpha+2\beta} \cdot x_{3\alpha+\beta}$ (first commutator) and $x_{3\alpha+\beta}$ (second commutator), and these all commute with the $x_{2\alpha+\beta}(\pm C_{2,1}^{\alpha,\beta}t^2u)$ in Equation (6.5). Thus said $x_{2\alpha+\beta}(\pm C_{2,1}^{\alpha,\beta}t^2u)$ cancel out, and we end up deducing that

$$\begin{aligned}[][x_\alpha(t), x_\beta(u)], x_\alpha(t)] &= x_{2\alpha+\beta}(C_{1,1}^{\alpha+\beta,\alpha} C_{1,1}^{\alpha,\beta} t^2u) \\ &\quad \cdot x_{3\alpha+2\beta}(C_{2,1}^{\alpha+\beta,\alpha} C_{1,1}^{\alpha,\beta} t^3u^2) \cdot x_{3\alpha+\beta}((C_{1,2}^{\alpha+\beta,\alpha} C_{1,1}^{\alpha,\beta} + C_{1,1}^{2\alpha+\beta,\alpha} C_{2,1}^{\alpha,\beta})t^3u). \end{aligned} \tag{6.6}$$

Finally, we take one more commutator with $x_\alpha(t)$. The latter two elements in the above commute with $x_\alpha(t)$ and thus may be deleted; we are left with

$$\begin{aligned}[][[x_\alpha(t), x_\beta(u)], x_\alpha(t)], x_\alpha(t)] &= [x_{2\alpha+\beta}(C_{1,1}^{\alpha+\beta,\alpha} C_{1,1}^{\alpha,\beta} t^2u), x_{\alpha(t)}] \\ &= x_{3\alpha+\beta}(C_{1,1}^{2\alpha+\beta,\alpha} C_{1,1}^{\alpha+\beta,\alpha} C_{1,1}^{\alpha,\beta} t^3u). \end{aligned} \tag{6.7}$$

Thus we have handled $\gamma = 3\alpha + \beta$. Since $\gamma = 3\alpha + 2\beta$ has also been treated, we get $\gamma = 2\alpha + \beta$ from Equation (6.6), and then $\gamma = \alpha + \beta$ from Equation (6.4). ■

We may now complete our goal for this section:

Proof of Theorem 6.2.8. We first show that $X_\alpha = X_{\alpha,1} \subseteq X_S$ for all $\alpha \in \Phi$. Since we are assuming $S^+ = \Phi$, we can write $\alpha = \sum_{\beta \in S} n_\beta \beta$ with $n_\beta \in \mathbb{N}$. Then by Fact 6.1.21 we can write $\alpha = p_{i_1} + p_{i_2} + \cdots + p_{i_\ell}$ with $p_{i_j} \in S$ and where all prefix sums are roots. Clearly we may assume that $p_j \neq -(p_1 + \cdots + p_{j-1})$ does not occur for any j, as otherwise the first j terms could be excised from the expression for α. Then by Lemma 6.2.13 it follows that $X_{p_{i_1}+p_{i_2}} \subseteq \langle X_{p_{i_1}}, X_{p_{i_2}} \rangle$, $X_{p_{i_1}+p_{i_2}+p_{i_3}} \subseteq \langle X_{p_{i_1}}, X_{p_{i_2}}, X_{p_{i_3}} \rangle$, and so on, eventually yielding $X_\alpha \subseteq \langle X_\beta : \beta \in \Phi \rangle$.

Now suppose by induction on $i \geq 0$ that $X_{\alpha,2^i} \subseteq X_S$ for all $\alpha \in \Phi$. By Fact 6.1.16, for any root $\gamma \in \Phi$ we can write $\gamma = \alpha + \beta$ for some $\alpha, \beta \in \Phi$, and it follows from Lemma 6.2.13 that $X_{\gamma,2^i+2^i} \subseteq \langle X_{\alpha,2^i}, X_{\beta,2^i} \rangle$. Thus indeed $X_{\gamma,2^{i+1}} \subseteq X_S$, completing the induction. ■

6.2.2 Structure of the links

In this section we describe the structure of the subgroups X_T where $T \subseteq \mathcal{S}$. This will be used to show that the links of all faces of $\mathfrak{K}$ are connected.

We will first need a "graded" version of Proposition 6.1.31.

Proposition 6.2.14. *Fix any ordering $\prec$ of the roots Φ, and let $\Psi \subseteq \Phi$ be linearly independent. Then the elements of X_Ψ are in 1-1 correspondence with expressions of the form $\prod_{\gamma \in \Psi^+} x_\gamma(t_\gamma)$ with $x_\gamma(t_\gamma) \in X_{\gamma,\mathrm{ht}_\Psi(\gamma)}$ (and the product taken in order $\prec$).*

Proof. We first prove that every expression of the given form is indeed in X_Ψ. Precisely, we show by induction on h that X_Ψ contains all subgroups $X_{\gamma,h}$ with $h = \mathrm{ht}(\gamma)$. The base case of $h = 1$ is immediate. For general h, take any $\gamma \in \Psi^+$ with height h and write $\gamma = \alpha + \beta$ with $\alpha, \beta \in \Psi^+$ of height smaller than h. (This is possible by Fact 6.1.21.) Now it follows from Lemma 6.2.13 that $X_{\gamma,\mathrm{ht}(\gamma)} = X_{\gamma,\mathrm{ht}(\alpha)+\mathrm{ht}(\beta)} \subseteq \langle X_{\alpha,\mathrm{ht}(\alpha)}, X_{\beta,\mathrm{ht}(\beta)} \rangle$, and this is in X_Ψ by induction.

We next show that every element in X_Ψ has a unique expression of the given form. In fact, it suffices to show existence, since uniqueness follows from Proposition 6.1.31 (note that Ψ^+ satisfies its hypotheses). Let us say that an expression of the form

$$x_{\gamma_1}(u_1)x_{\gamma_2}(u_2)\cdots x_{\gamma_m}(u_m) \tag{6.8}$$

with $\gamma_i \in \Psi^+$ is *well-bounded* if each u_i has degree at most $\mathrm{ht}(\gamma_i)$. The desired existence result is that every $z \in X_\Psi$ has a well-bounded expression as above, where $\gamma_1, \dots, \gamma_m$ list the elements of Ψ^+ in the order $\prec$. (We remark that it doesn't matter whether we are allowing consecutive duplicate γ_i's in this list, since $x_\gamma(u)x_\gamma(u') = x_\gamma(u+u')$ and this preserves well-boundedness.)

To show this existence, it actually suffices to repeat the existence proof in Proposition 6.1.31. At a high level, this works because that proof ultimately only uses the commutator formula, and applications of the commutator formula preserve well-boundedness. That is, starting from an arbitrary $z \in X_\Psi$, by definition we may express z as in Equation (6.8) with each $\gamma_i \in \Psi$ and each u_i of degree at most 1. This is well-bounded. Then an application of the commutator formula switches some consecutive $x_\gamma(u)x_{\gamma'}(u')$ to $x_{\gamma'}(u')x_\gamma(u)[x_\gamma(u), x_{\gamma'}(u')]$, and this commutator is the product of elements of the form $x_{i\gamma+j\gamma'}(C_{i,j}^{\gamma,\gamma'} u^i (u')^j)$. But this product is indeed well-bounded, presuming the former expression was well-bounded.

For completeness, we sketch why the existence result in Proposition 6.1.31 only relies on the commutator formula. We prefer to first follow the existence result in [Car89, Thm. 5.3.3], which assumes that the order $\prec$ is consistent with heights (meaning $\mathrm{ht}_\Psi(\alpha) \leq \mathrm{ht}_\Psi(\beta)$ implies $\alpha \prec \beta$). Under this assumption, we may repeatedly reorder consecutive products $x_\gamma(u)x_{\gamma'}(u')$ whenever $\gamma' \prec \gamma$, as described above. Notice that the new products of elements of the form $x_{i\gamma+j\gamma'}(C_{i,j}^{\gamma,\gamma'} u^i (u')^j)$ that arise are have $\mathrm{ht}(i\gamma + j\gamma') > \mathrm{ht}(\gamma), \mathrm{ht}(\gamma')$. Because of this, and the height-respecting property of $\prec$, this process must eventually terminate with a (well-bounded) expression like Equation (6.8) where the roots γ_i are in the order $\prec$ (and any missing root $\gamma \in \Phi^+$ can be inserted via $x_\gamma(0)$).

It remains to treat the case that the root order $\prec$ does *not* necessarily respect heights. For this we appeal to [Ste16, Lem. 18], the associated component of the proof of Proposition 6.1.31. It says that it suffices to check — when γ *is* a height-respecting order, and $\Psi^+ = \{\gamma_1 \prec \gamma_2 \prec \cdots \prec \gamma_m\}$

— that each subgroup of the form

$$B_i := X_{\gamma_i,\mathrm{ht}(\gamma_i)} \cdot X_{\gamma_{i+1},\mathrm{ht}(\gamma_{i+1})} \cdots X_{\gamma_r,\mathrm{ht}(\gamma_m)}$$

is normal in X_Ψ. To see this, take a generic well-bounded expression

$$y = x_{\gamma_i}(t_i)x_{\gamma_{i+1}}(t_{i+1}) \cdots x_{\gamma_m}(t_m)$$

in B_i and consider conjugating it by an arbitrary well-bounded expression w as in Equation (6.8). We have $w^{-1}yw = y[y, w]$, and expanding the commutator yields a well-bounded expression consisting only of $x_\gamma(v)$'s where $\mathrm{ht}(\gamma) \geq \mathrm{ht}(\gamma_i)$. Now as in the previous argument, this may be further rearranged into a well-bounded expression in B_i, showing that B_i is closed under conjugation and hence normal. ■

We have the following immediate consequence:

Corollary 6.2.15. *Let Ψ be a set of linearly independent roots. Then* $|X_\Psi| = \prod_{\alpha\in\Psi^+} p^{\mathrm{ht}_\Psi(\alpha)+1}$.

Importantly, $|X_\Psi|$ can be bounded independently of m (where recall $|\mathbb{F}| = p^m$). This will imply that a vertex in $\mathfrak{K}$ belongs to just $p^{O(1)}$ faces where the $O(1)$ does not depend on m.

The proceeding normal form result also helps us show the following:

Proposition 6.2.16. *Let Ψ and Ψ' be sets of linearly independent roots. Then $X_\Psi \cap X_{\Psi'} = X_{\Psi\cap\Psi'}$.*

Proof. By Fact 6.1.20 we may choose a set Π of simple roots with $\Psi \subseteq \Pi^+$. We apply Proposition 6.2.14 to any $g \in X_\Psi$ and $h \in X_{\Psi'}$, writing them as $g = \prod_{\alpha\in\Psi^+} x_\alpha(t_\alpha)$ and $h = \prod_{\alpha\in\Psi'^+} x_\alpha(u_\alpha) = U \cdot L$, where we have ordered h as a product U of root elements in Π^+ times a product L of root elements in Π^-. Now supposing $g = h$, we get $U^{-1}g = L$. But by Proposition 6.1.32, the only way this equality can hold is if $L = 1$. Hence we have $\prod_{\alpha\in\Psi^+} x_\alpha(t_\alpha) = \prod_{\alpha\in\Psi'^+} x_\alpha(u_\alpha)$, where on both sides α is ranging in Π^+; hence by uniqueness of these expressions (assuming the products are taken in the same order), equality holds just when $t_\alpha = u_\alpha$ for all α. So the elements of $X_\Psi \cap X_{\Psi'}$ are exactly the elements of the form

$$\prod_{\alpha\in\Psi^+\cap\Psi'^+} x_\alpha(f_\alpha)$$

where $\deg(f_\alpha) \leq \min(\mathrm{ht}_\Psi(\alpha), \mathrm{ht}_{\Psi'}(\alpha))$. But note that $\Psi^+ \cap \Psi'^+ = (\Psi \cap \Psi')^+$ and $\mathrm{ht}_\Psi(\alpha) = \mathrm{ht}_{\Psi'}(\alpha) = \mathrm{ht}_{\Psi\cap\Psi'}(\alpha)$ for $\alpha \in (\Psi \cap \Psi')^+$ due to linear independence. So any such an element belongs to $X_{\Psi\cap\Psi'}$ (using Proposition 6.2.14 again), which proves the proposition. ■

Observation 6.2.17. In fact, $Z \cdot X_\Psi \cap Z \cdot X_{\Psi'} = Z \cdot X_{\Psi\cap\Psi'}$, where Z denotes the center of $\overline{\mathrm{G}}(\Phi, \mathbb{F})$. The proof proceeds in the same fashion: Under the matrix identification of Proposition 6.1.32, Z consists of diagonal matrices. Thus if $D_1U^{-1}g = D_2L$ with D_1 and D_2 diagonal, L lower-unitriangular and $U^{-1}g$ upper-unitriangular, we must have $D_1 = D_2$ and $L = U^{-1}g = 1$. This implies $\prod_{\alpha\in\Psi^+} x_\alpha(t_\alpha) = \prod_{\alpha\in\Psi'^+} x_\alpha(u_\alpha)$, and the rest of the proof follows as before.

Combining Proposition 6.2.16 with Fact 6.1.9 lets us understand the structure of the links in $\mathfrak{K}$:

Theorem 6.2.18. *Let $\sigma \in \mathfrak{K}$ be a face of type $T \subsetneq \mathcal{S}$. Then the link of F is isomorphic to the coset complex $\mathcal{CC}(X_{\mathcal{S}\setminus T}; (X_{\mathcal{S}\setminus T\setminus\{\alpha\}} : \alpha \in \mathcal{S}\setminus T))$.*

Proof. For $T = \emptyset$, this is the combination of Theorem 6.2.8 and Proposition 6.2.9. Otherwise, by virtue of Fact 6.1.9 it suffices to show that for any $U \subseteq \mathcal{S}$,

$$H_U = \bigcap_{\alpha\in U} H_\alpha = \bigcap_{\alpha\in U} X_{\mathcal{S}\setminus\{\alpha\}} = X_{\mathcal{S}\setminus U}.$$

But this follows from Proposition 6.2.16 after recalling (Remark 6.2.3) that $\mathcal{S}\setminus\{\alpha\}$ is linearly independent for any α. ■

Finally, whenever $|T| \le d-1$ the sets $\mathcal{S}\setminus T\setminus\{\alpha\}$ are nonempty, and so we may therefore conclude using Fact 6.1.8:

Corollary 6.2.19. *For all $\sigma \in \mathfrak{K}(i)$ with $i \le d-2$, K_σ is connected.*

Remark 6.2.20. The fact that $\mathfrak{K}$ and all of its links of dimension at most $d-2$ are connected is equivalent to saying that $\mathfrak{K}$ is *strongly gallery connected* [KO18, Rem. 2.1].

6.2.3 Expansion of links

Definition 6.2.21. For $\alpha, \beta \in \Phi$ with $\alpha \neq -\beta$ we write $\mathcal{CC}(\alpha;\beta) = \mathcal{CC}(X_{\{\alpha,\beta\}}; (X_\alpha, X_\beta))$.

It follows from Theorem 6.2.18 that the link of every $(d-2)$-dimensional face in our complex $\mathfrak{K}$ is isomorphic to $\mathcal{CC}(\alpha;\beta)$ for distinct $\alpha, \beta \in \mathcal{S}$. The main goal of this section is to show that the bipartite skeleton graphs of these $\mathcal{CC}(\alpha;\beta)$ are good expanders. (For this we will not even need to recall our specific choice of $\mathcal{S}$.) Combined with Theorem 6.1.2 and the connectivity result Corollary 6.2.19, it follows that all links of $\mathfrak{K}$ are good expanders.

We begin with a simple observation:

Proposition 6.2.22. *For $\alpha \neq -\beta$, the (skeleton of) $\mathcal{CC}(\alpha;\beta)$ is a p^2-regular bipartite (multi)graph.*

Proof. From Fact 6.1.9, the link of a vertex in $X_{\alpha,\beta}/X_\beta$ is isomorphic to $\mathcal{CC}(X_\beta; X_\alpha \cap X_\beta) = \mathcal{CC}(X_\beta; 1)$, where we used Proposition 6.2.16. But this is equivalent to saying the neighborhood of a vertex in the skeleton is a set of size $|X_\beta| = p^2$ (recalling Corollary 6.2.15). The same consideration holds for vertices in $X_{\alpha,\beta}/X_\alpha$. ■

The key idea we will use in understanding the expansion of the links $\mathcal{CC}(\alpha;\beta)$ will be to look at the graph-theoretic *square*, $\mathcal{CC}(\alpha;\beta)^2$, of (the skeleton of) $\mathcal{CC}(\alpha;\beta)$. Since $\mathcal{CC}(\alpha;\beta)$ is connected and bipartite, we know that its random walk matrix has isolated "trivial" eigenvalues of ± 1, and all other eigenvalues are between $\pm\lambda_2(\mathcal{CC}(\alpha;\beta))$. Thus if we exclude from $\mathcal{CC}(\alpha;\beta)^2$ the "trivial" eigenvalue 1, its maximum eigenvalue will be $\lambda_2(\mathcal{CC}(\alpha;\beta))^2$, the square of what we wish to bound. In fact, since $\mathcal{CC}(\alpha;\beta)$ is bipartite, $\mathcal{CC}(\alpha;\beta)^2$ will have two disconnected components corresponding to the two parts of $\mathcal{CC}(\alpha;\beta)$. It is a simple and well-known linear algebra fact that these two components have the same eigenvalues (possibly up to some eigenvalues of 0). Hence it suffices for us to bound the eigenvalues of $\mathcal{CC}(\alpha;\beta)^2$ on only *one* of the two sides, $X_{\alpha,\beta}/X_\alpha$ or $X_{\alpha,\beta}/X_\beta$.

As we will now show, whenever $\Phi \neq G_2$, at least one of these two sides is an abelian Cayley graph. (Interestingly, we do not know that both sides are.) Thus we can understand the eigenvalues by elementary methods. We discuss a potential approach to handling the G_2 case in **??**.

Theorem 6.2.23. *Let $\alpha, \beta \in \Phi \neq G_2$, with $\alpha \neq -\beta$. Then the nontrivial eigenvalues of $\mathcal{CC}(\alpha;\beta)^2$ are at most $2/p$; hence $\lambda_2(K_\sigma) \leq \sqrt{2/p}$ for every $\sigma \in \mathfrak{K}(d-2)$.*

Proof. When it is relevant, we will follow the convention of calling the shorter of the two roots α and the longer β. Then, with foresight toward Case 3 below, we choose to study the $X_{\alpha,\beta}/X_\alpha$ side of $\mathcal{CC}(\alpha;\beta)^2$.

By virtue of Proposition 6.2.14, we can describe coset representatives for $X_{\alpha,\beta}/X_\alpha$ fairly simply; fixing an ordering for the roots in which α is last, we can take as coset representatives precisely those elements of the form

$$g = \prod\{x_{i\alpha+j\beta}(t_{ij}) : (i,j) \in \mathbb{N}\times\mathbb{N}\setminus\{(1,0)\},\ i\alpha+j\beta \in \Phi,\ \deg(t_{ij}) \leq i+j\}. \tag{6.9}$$

Moreover, the p^2 neighbors (counted with multiplicity) of vertex gX_α in the squared (multi)graph $\mathcal{CC}(\alpha;\beta)^2$ are the following cosets:

$$(g\cdot x_\alpha(f_0)\cdot x_\beta(f_1))X_\alpha, \quad \text{for } f_0, f_1 \in \mathbb{F} \text{ of degree at most } 1.$$

Via the commutator formula one sees that the associated coset representatives are

$$\begin{aligned} g\cdot x_\alpha(f_0)\cdot x_\beta(f_1)\cdot x_\alpha(-f_0) &= g\cdot x_\beta(f_1)\cdot[x_\beta(f_1), x_\alpha(-f_0)] \\ &= g\cdot x_\beta(f_1)\cdot \prod_{\substack{i,j\in\mathbb{Z}_+ \\ i\alpha+j\beta\in\Phi}} x_{i\alpha+j\beta}(C^{\beta,\alpha}_{ij}(-f_0)^i f_1^j). \end{aligned} \tag{6.10}$$

By Remark 6.1.14, either $\alpha+\beta \notin \Phi$, or the root subsystem of Φ spanned by α and β is one of A_2, B_2, or G_2. We will skip the case when α and β span G_2, as it only arises when $\Phi = G_2$. Now as in the proof of Lemma 6.2.13, we will do case analysis on the possible sets $R = \{i,j : i\alpha+j\beta\in\Phi\}$.

Case 1: $R = \{\alpha,\beta\}$**.** If $\alpha+\beta\notin\Phi$, then X_α and X_β commute by Theorem 6.1.27, and it is easy to check that $\mathcal{CC}(\alpha;\beta)$ is in fact the complete p^2-regular bipartite graph; hence the nontrivial eigenvalues of $\mathcal{CC}(\alpha;\beta)^2$ are all 0.

Case 2: $\{\alpha,\beta,\alpha+\beta\}$**.** It was shown in [KO18], and alternatively in [HS19, Corollary 5.6], that $\lambda_2(\mathcal{CC}(\alpha;\beta)) = \sqrt{1/p}$; equivalently, the nontrivial eigenvalues of $\mathcal{CC}(\alpha;\beta)$ are at most $1/p$. Here we give a different proof of this fact, the strategy of which will be generalized in Case 3.

From Equations (6.9) and (6.10) we have that a typical coset representative $g = x_\beta(t_{01})\cdot x_{\alpha+\beta}(t_{11})$ is connected in $\mathcal{CC}(\alpha;\beta)^2$ to the following coset representatives, for $f_0, f_1 \in \mathbb{F}$ of degree at most 1:

$$x_\beta(t_{01})\cdot x_{\alpha+\beta}(t_{11})\cdot x_\beta(f_1)\cdot x_{\alpha+\beta}(-C^{\beta,\alpha}_{11}f_0f_1) = x_\beta(t_{01}+f_1)\cdot x_{\alpha+\beta}(t_{11}-C^{\beta,\alpha}_{11}f_0f_1).$$

Reparameterizing with $f_2 = -C^{\beta,\alpha}_{11}f_0$ (and recalling $C^{\beta,\alpha}_{11}\neq 0$), it is evident that $\mathcal{CC}(\alpha;\beta)^2$ is an abelian Cayley group, wherein each vertex is a pair (ℓ,q) with ℓ linear and q quadratic, hence $(\ell,q)\cong\mathbb{F}_p^5$, and with edges involve adding a pair (f_1, f_1f_2) for f_1, f_2 linear. With x denoting the field indeterminate, we can write $f_1 = a + bx$ and $f_2 = c + dx$; then the $X_{\alpha,\beta}/X_\alpha$ side of our graph $\mathcal{CC}(\alpha;\beta)^2$ may be identified as an abelian Cayley graph on $\mathbb{F}_p^5$ with symmetric generating set

$$\{(a, b, ac, ad + bc, bd) : a, b, c, d \in \mathbb{F}_p^4\}.$$

Then it is well known that the eigenvalues of this graph are given by the exponential sums

$$\begin{aligned} &\mathop{\mathbf{E}}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d}\sim\mathbb{F}_p}\big[\mathrm{Exp}_p(r_1\boldsymbol{a} + r_2\boldsymbol{b} + r_3\boldsymbol{ac} + r_4(\boldsymbol{ad} + \boldsymbol{bc}) + r_5\boldsymbol{bd})\big] \\ &= \mathop{\mathbf{E}}_{\boldsymbol{c},\boldsymbol{d}}\Big[\mathop{\mathbf{E}}_{\boldsymbol{a}}\big[\mathrm{Exp}_p(\boldsymbol{a}\cdot h(\boldsymbol{c},\boldsymbol{d}))\big]\mathop{\mathbf{E}}_{\boldsymbol{b}}\big[\mathrm{Exp}_p(\boldsymbol{b}\cdot h'(\boldsymbol{c},\boldsymbol{d}))\big]\Big] \end{aligned} \tag{6.11}$$

for $r_1, \dots, r_5 \in \mathbb{F}_p$, where $\mathrm{Exp}_p(z) = e^{2\pi iz/p}$, and

$$h(c, d) = r_1 + r_3c + r_4d, \qquad h'(c, d) = r_2 + r_4c + r_5d.$$

Notice whenever the outcome $\boldsymbol{c}, \boldsymbol{d}$ has $h(\boldsymbol{c}, \boldsymbol{d}) \neq 0$, the quantity $\mathbf{E}_{\boldsymbol{a}}\big[\mathrm{Exp}_p(\boldsymbol{a}\cdot h(\boldsymbol{c},\boldsymbol{d}))\big]$ inside Equation (6.11) becomes 0. On the other hand, if $h(\boldsymbol{c}, \boldsymbol{d}) = 0$ then this quantity is 1. Similar considerations hold for h', and we conclude that the eigenvalue in Equation (6.11) is precisely

$$\mathop{\mathbf{P}}_{\boldsymbol{c},\boldsymbol{d}}[h(\boldsymbol{c},\boldsymbol{d}) = h'(\boldsymbol{c},\boldsymbol{d}) = 0].$$

Of course if $r_1 = \cdots = r_5 = 0$ then h, h' are formally 0 and the above is the trivial eigenvalue of 1. But otherwise, at least one of h, h' is nonzero — say, h — and, being an affine linear polynomial over $\mathbb{F}_p$, it has $\mathbf{P}_{\boldsymbol{c},\boldsymbol{d}}[h(\boldsymbol{c},\boldsymbol{d})] \leq 1/d$. This shows that indeed the nontrivial eigenvalues of $\mathcal{CC}(\alpha;\beta)^2$ are at most $1/p$.

Case 3: $R = \{\alpha, \beta, \alpha + \beta, 2\alpha + \beta\}$. As mentioned earlier, here we have named the shorter root α and the longer root β. From Equations (6.9) and (6.10) we have that a typical coset representative $g = x_\beta(t_{01}) \cdot x_{\alpha+\beta}(t_{11}) \cdot x_{2\alpha+\beta}(t_{21})$ is connected in $\mathcal{CC}(\alpha;\beta)^2$ to the following coset representatives, for $f_0, f_1 \in \mathbb{F}$ of degree at most 1:

$$\begin{aligned} &x_\beta(t_{01}) \cdot x_{\alpha+\beta}(t_{11}) \cdot x_{2\alpha+\beta}(t_{21}) \cdot x_\beta(f_1) \cdot x_{\alpha+\beta}(-C_{11}^{\beta,\alpha} f_0 f_1) \cdot x_{2\alpha+\beta}(C_{12}^{\beta,\alpha} f_0^2 f_1) \\ &\qquad\qquad = x_\beta(t_{01} + f_1) \cdot x_{\alpha+\beta}(t_{11} - C_{11}^{\beta,\alpha} f_0 f_1)) \cdot x_{2\alpha+\beta}(t_{21} + C_{12}^{\beta,\alpha} f_0^2 f_1). \end{aligned}$$

(We remark that had we looked at the $X_{\alpha,\beta}/X_\beta$ side of $\mathcal{CC}(\alpha;\beta)^2$, we would not have gotten all of the commutativity in the above calculation.) Reparameterizing again with $f_2 = -C_{11}^{\beta,\alpha} f_0$, this is

$$x_\beta(t_{01} + f_1) \cdot x_{\alpha+\beta}(t_{11} + f_1 f_2) \cdot x_{2\alpha+\beta}(t_{21} + C f_1 f_2^2)$$

for some constant $C \neq 0$ in $\mathbb{F}_p$. Similar to Case 2, we see that this is an abelian Cayley graph on $\mathbb{F}_p^9$ with symmetric generating set

$$\{(a, b, ac, ad + bc, bd, Cac^2, C(bc^2 + 2acd), C(2bcd + ad^2), Cbd^2) : a, b, c, d \in \mathbb{F}_p\}.$$

As before, the eigenvalues of this graph are given by

$$\begin{aligned}
\mathop{\mathbf{E}}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d}\in\mathbb{F}_p}\Big[&\mathrm{Exp}_p(r_1\boldsymbol{a}+r_2\boldsymbol{b}+r_3\boldsymbol{ac}+r_4(\boldsymbol{ad}+\boldsymbol{bc})+r_5\boldsymbol{bd}+r_6C\boldsymbol{ac}^2+ \\
&\qquad r_7C(\boldsymbol{bc}^2+2\boldsymbol{acd})+r_8C(2\boldsymbol{bcd}+\boldsymbol{ad}^2)+r_9C\boldsymbol{bd}^2)\Big] \\
&= \mathop{\mathbf{E}}_{\boldsymbol{c},\boldsymbol{d}}\Big[\mathop{\mathbf{E}}_{\boldsymbol{a}}\big[\mathrm{Exp}_p(\boldsymbol{a}\cdot h(\boldsymbol{c},\boldsymbol{d}))\big]\mathop{\mathbf{E}}_{\boldsymbol{b}}\big[\mathrm{Exp}_p(\boldsymbol{b}\cdot h'(\boldsymbol{c},\boldsymbol{d}))\big]\Big], \qquad (6.12)
\end{aligned}$$

for all $r_1,\dots,r_9\in\mathbb{F}_p$, where

$$\begin{aligned}
h(c,d) &= r_1+r_3c+r_4d+Cr_6c^2+2Cr_7cd+Cr_8d^2, \\
h'(c,d) &= r_2+r_4c+r_5d+Cr_7c^2+2Cr_8cd+Cr_9d^2.
\end{aligned}$$

The argument is now the same as in Case 2, except we reason that if h is nonzero, then $\mathbf{P}_{\boldsymbol{c},\boldsymbol{d}}[h(\boldsymbol{c},\boldsymbol{d})]\le 2/p$ by Schwarz–Zippel, since now h is quadratic. ■

Remark 6.2.24. When $\Phi = G_2$ two other graphs can arise as $\mathcal{CC}(\alpha;\beta)$. The squares of these graphs are not Cayley graphs of abelian groups, and so the previous approach fails. For completeness we now give explicit description of the squared graphs $\mathcal{CC}(\alpha;\beta)^2$ restricted to the vertices on the side $X_{\alpha,\beta}/X_\alpha$.

From Figure 6.1 we see that if $\alpha,\beta\in G_2$ and $\alpha+\beta\in G_2$ then $\angle(\alpha,\beta)\in\{60^\circ,120^\circ,150^\circ\}$. If $\angle(\alpha,\beta)=60^\circ$ or if $\angle(\alpha,\beta)=120^\circ$ and α and β are long roots, the analysis is the same as in Case 2 of the previous proof. There are two remaining cases: (I) α and β are simple roots and $\angle(\alpha,\beta)=150^\circ$ as in Figure 6.1; (II) $\angle(\alpha,\beta)=120^\circ$ and α and β are short roots.

Case I: $\angle(\alpha,\beta)=150^\circ$. From Equation (6.9), a typical coset representative in $X_{\alpha,\beta}/X_\alpha$ is

$$g = x_\beta(t_{01})\cdot x_{\alpha+\beta}(t_{11})\cdot x_{2\alpha+\beta}(t_{21})\cdot x_{3\alpha+\beta}(t_{31})\cdot x_{3\alpha+2\beta}(t_{32})$$

with $\deg(t_{ij})\le i+j$. By Equation (6.10), the neighbors of this coset representative in $\mathcal{CC}(\alpha;\beta)^2$ are parameterized by

$$g\cdot x_\beta(f_1-t_{01})\cdot[x_\beta(f_1-t_{01}),x_\alpha(-f_0)]$$

for all f_0,f_1 of degree at most 1. Using Proposition 6.1.30, one can show that this is the multigraph with vertices $(t_{01},t_{11},t_{21},t_{31},t_{32})$ whose neighbors are parameterized by

$$(f_1+t_{01},-f_0f_1+t_{11},f_0^2f_1+t_{21},-f_0^3f_1+t_{31},-f_1(t_{31}+3t_{21}f_0+f_0^3f_1)+t_{32}).$$

Case II: $\angle(\alpha,\beta)=120^\circ$. A typical coset representative in $X_{\alpha,\beta}/X_\alpha$ is

$$g = x_\beta(t_{01})\cdot x_{\alpha+\beta}(t_{11})\cdot x_{2\alpha+\beta}(t_{21})\cdot x_{\alpha+2\beta}(t_{12}),$$

where $\deg(t_{ij})\le i+j$. The neighbors of this coset representative are parameterized by

$$g\cdot x_\beta(f_1-t_{01})\cdot[x_\beta(f_1-t_{01}),x_\alpha(-f_0)]$$

for all f_0,f_1 of degree at most 1. Using Proposition 6.1.30, one can show that this is the multigraph with vertices $(t_{01},t_{11},t_{21},t_{12})$ whose neighbors are parameterized by

$$(f_1+t_{01},-2f_0f_1+t_{11},3f_0^2f_1+t_{21},3f_1(t_{11}+f_0f_1)+t_{12}).$$

6.3 Concluding

Finally we can prove Theorem 6.2.6.

Proof of Theorem 6.2.6.

1. By Theorem 6.1.34, we have $|\overline{\mathrm{G}}(\Phi, \mathbb{F})| = p^{\Theta(m)}$. By Corollary 6.2.15, the subgroups H_α have size at most $p^{O(1)}$. (Here the $\Theta(\cdot)$ and $O(\cdot)$ depend only on Φ.) Hence there are $p^{\Theta(m)}$ total cosets, and the claim that $|\mathfrak{K}(0)| = p^{\Theta(m)}$ follows. Note that as m increases by 1, the size of the complex grows by a constant factor $p^{O(1)}$; thus we have the linear growth in size needed for a strongly explicit family, and the *exact* number of vertices n can be computed efficiently in $\mathrm{poly}(m) = \mathrm{polylog}(n)$ time (by Theorem 6.1.34). The resulting family is strongly explicit thanks to Theorem 6.1.35: one can and construct all the group elements in $\overline{\mathrm{G}}(\Phi, \mathbb{F})$ efficiently, one can identify the vertices (cosets) explicitly and naively by listing all their elements (recall each H_α has constant size), and one can compute the complex's adjacency structure (e.g., list all maximal faces to which a given vertex belongs) thanks to the efficient ($\mathrm{poly}(m) = \mathrm{polylog}(n)$ time) group arithmetic from Theorem 6.1.35.
2. Again, by Corollary 6.2.15 the subgroups H_α have size at most $p^{O(1)}$. The number of maximal faces containing a vertex is therefore at most $p^{O(1)\cdot(d+1)} = p^{O(1)}$.
3. This is Corollary 6.2.19.
4. This is Theorem 6.2.23.
5. From Proposition 6.2.16, $\bigcap_{\alpha\in\mathcal{S}} H_\alpha = \{1\}$. The claim then follows from Fact 6.1.6. In addition, by Observation 6.2.17 we have that $\bigcap_{\alpha\in\mathcal{S}} ZH_\alpha = Z$, and so in the quotient group $\mathrm{G}(\Phi, \mathbb{F})$ the subgroups H_iZ/Z intersect trivially. Hence we also get a simply-transitive action for the adjoint Chevalley groups. ■

6.4 Expansion of the G_2 complex

As mentioned in Remark 6.2.24, the above methods do not work to show that the 2-dimensional complexes obtained from our construction for $\overline{\mathrm{G}}(G_2, \mathbb{F})$ yield HDX families, either in the case $\mathcal{S} = \{\alpha, \beta, -\alpha-\beta\}$ (as we selected in Definition 6.2.2) or in the alternative case $\mathcal{S} = \{\alpha, \alpha+\beta, -2\alpha-\beta\}$ mentioned in Remark 6.2.11. The issue is that the resulting link graphs are not Cayley graphs of abelian groups. We note that this latter case with $\mathcal{S} = \{\alpha, \alpha+\beta, -2\alpha-\beta\}$ is particularly appealing, as all vertex links are isomorphic (i.e., the 1-skeleton is a *graph of constant link*). Additionally, to the best of our knowledge none of the links in this case arise in previous HDX constructions. This is in contrast to our other constructions, which always contain some links isomorphic to those studied in [KO18].

One approach to prove the expansion of these links is to count the number of closed walks of some fixed length k in one side of their square, which (by the trace method) equals the sum of the kth powers of the eigenvalues of their adjacency matrices. By Remark 6.2.24, the number of length-k paths starting and ending at a fixed vertex on the side $X_{\alpha,\beta}/X_\alpha$ equals the number of solutions to the following systems of equations, corresponding to the first and second cases in

Remark 6.2.24, respectively:

$$0 = \sum_{i=1}^{k} g_i = \sum_{i=1}^{k} f_i g_i = \sum_{i=1}^{k} f_i^2 g_i = \sum_{i=1}^{k} f_i^3 g_i = \sum_{i=1}^{k} -g_i(f_i^3 g_i + \sum_{j=1}^{i-1}(f_j^3 g_j + 3f_j^2 g_j f_i)),$$
$$0 = \sum_{i=1}^{k} g_i = \sum_{i=1}^{k} f_i g_i = \sum_{i=1}^{k} f_i^2 g_i = \sum_{i=1}^{k} g_i(f_i g_i - 2\sum_{j=1}^{i-1} f_j g_j).$$

Here f_i and g_i are linear polynomials in $\mathbb{F}_p[x]$. The graphs corresponding to the first and second systems have p^n vertices, where $n = 20$ and $n = 13$, respectively. Therefore if for some particular k one could bound the number of solutions to either of these by, say, $p^{4k-n} + p^{3.99k}$, expansion of the corresponding complexes would follow. To show this it would suffice to show that the varieties defined over $\mathbb{C}$ by these systems are irreducible and of dimension at most $4k - 20$ in the first case, or $4k - 13$ in the second case, for some k. This seems potentially tractable for a computer algebra system.

In this section, we give a proof of the expansion of our second construction obtained from the group $G_2(q)$. The method we use is very powerful and can be applied to all of our other constructions as well.

Definition 6.4.1. The *Schmidt rank* $r(f)$ of $f \in \mathbb{F}[X]_d$ is the minimum r such that $f = \sum_{i=1}^{r} g_i h_i$ where g_i and h_i have degree less than d. The Schmidt rank of a collection of polynomials is defined as the minimum Schmidt rank of any nontrivial linear combination of polynomials in the family.

Theorem 6.4.2. *[KZ20, Proposition 3.1] Let $\mathcal{F} = \{f_1, \dots, f_\ell\}$ be polynomials over $\mathbb{F}_p$ in s variables of degree at most $d < p$. Then for all t, there exists $c = c(d, \ell, t)$, independent of p and s and depending polynomially on ℓ and t, such that if $r(\mathcal{F}) > c$, then for all $y \in \mathbb{F}_p^\ell$,*

$$|\#\{x \in \mathbb{F}_p^\ell : f_i(x) = y_i\} - p^{s-\ell}| \le p^{s-\ell-t}.$$

This theorem belongs to line of work beginning with that of Green and Tao [GT07], who gave a non-quantitative bound on the function c. Subsequent work by Kaufman and Lovett gave Ackermann-type bounds. The fact that one can take c to be polynomial is due to a result of Milićević [Mil19]; work of Janzer also gave a slightly weaker result with dependence on $|\mathbb{F}|$. These theorems are typically stated for a single polynomial, but using straightforward Fourier analysis the above result can be established.

Now, recall that to showing the expansion of one of our G_2 constructions could be reduced to counting the solutions to

$$\sum_{i=1}^{k} g_i = \sum_{i=1}^{k} f_i g_i = \sum_{i=1}^{k} f_i^2 g_i = \sum_{i=1}^{k} g_i(f_i g_i - 2\sum_{j=1}^{i-1} f_j g_j) = 0 \tag{6.13}$$

where $f_i, g_i \in \mathbb{F}_p[x]_{\le 1}$. These correspond to the number of closed walks from one vertex to itself in the square of the bipartite link graph. We would like to show that this is at most $p^{4k-4} + o(p^{4k-4})$. To do so, we will show that this system of polynomials has Schmidt rank $\Omega(k)$. By applying Theorem 6.4.2 we will conclude that the number of closed walks is at most $p^{4k-4} + p^{4k-4-k^\delta}$ for some $\delta > 0$. We next show that it suffices to show that the "unlifted" system of equations where f_i, g_i are field elements has high Schmidt rank.

Definition 6.4.3. Let f be homogeneous of degree d. For $j \leq d$, let $L_j(f) \in \mathbb{F}[x_{ij} : 1 \leq i \leq n, 0 \leq j \leq 1]$ be the coefficient of t^j in $f(x_{1,0} + x_{1,1}t, \ldots, x_{n,0} + x_{n,1}t) \in \mathbb{F}[X, t]$.

Note that $L_j(f)$ has the same degree as f. Also note that L_j is a linear map on the space of polynomials, and acts on monomials via $L_j(\prod_{i=1}^n x_i^{\alpha_i}) = \sum_{\beta<\alpha,|\beta|=j} \prod \binom{\alpha_i}{\beta_i} x_{i,0}^{\alpha_i-\beta_i} x_{i,1}^{\beta_i}$. It follows from this formula that L_j is invertible in characteristic bigger than $\binom{d}{j}$; explicitly, $(L_j)^{-1}x_{i,j} = \binom{d}{j}^{-1} x_i$.

Definition 6.4.4. Given a set of homogeneous polynomials $\mathscr{F} = \{f_1, \ldots, f_m\} \subseteq \mathbb{F}[x_1, \ldots, x_n]$, let $L(\mathscr{F}) = \{L_j(f_i) : j \leq \deg(f_i)\}$.

Note that solutions to a lifted system correspond to solutions to the original system where the variables are interpreted as polynomials of degree at most 1.

Lemma 6.4.5. *Let $\mathscr{F} = \{f_1, \ldots, f_m\}$. If $\operatorname{char}(\mathbb{F}) > 2^d$, then $r(L(\mathscr{F})) \geq d^{-1} \cdot r(\mathscr{F})$.*

Proof. Suppose that $r(L(\mathscr{F})) = r$. If $r = 0$ the claim is trivial so assume $r > 0$. By definition, there exists $g = \sum_{1\leq i\leq m, 0\leq j\leq \deg(f_i)} \lambda_{ij} L_j(f_i) \in \mathit{span}(L(\mathscr{F}))$ nonzero and with $r(g) = r$, so $g = \sum_{i=1}^r P_i Q_i$ where $\deg(P_i)$ and $\deg(Q_i)$ are at most $\deg(g)$. Pick any j such that $\sum_i \lambda_{ij} L_j(f_i) \neq 0$. Substitute $x_{i,j} := x_{i,j} t^j$ for all i, j. Note that the coefficient of t^j equals $\sum_i \lambda_{ij} L_j(f_i)$. Since $|\mathbb{F}| \geq d \geq t$, we can interpolate this coefficient and find that $r(\sum_i \lambda_{ij} f_{ij}) \leq rd$. Now by substituting $x_{ij} := \binom{d}{j}^{-1} x_i$ in this decomposition, we find that $r(\sum_i \lambda_{ij} \binom{\deg f_i}{j}^{-1} f_i) \leq rd$. The linear combination on the left hand side is nontrivial because $r(L(\mathscr{F})) > 0$ implies the elements in $\mathscr{F}$ are linearly independent. We have shown that the rank of a nontrivial linear combination of the elements in $\mathscr{F}$ is at most rd, so the claim follows. ■

By Lemma 6.4.5, it suffices to prove a lower bound on $r(\mathscr{F})$. Furthermore, it suffices to do so over $\overline{\mathbb{F}}$. To do so we'll use the following well-known bound on $r(f)$. Let $\operatorname{Sing}(f) = \boldsymbol{V}(\partial_1 f, \ldots, \partial_n f)$ denote the singular locus of f.

Proposition 6.4.6. *[LZ22, Claim 2.2]* $2 \cdot r(f) \geq \operatorname{Codim}(\operatorname{Sing}(f))$.

Theorem 6.4.7. *For $p > 3$ and sufficiently large k, the Schmidt rank of the system Equation (6.13) with $f_i, g_i \in \mathbb{F}_p$ is at least $k/6$.*

Proof. Let $\mathscr{F}$ be as in Equation (6.13). By the previous proposition, it suffices to show that over $\overline{\mathbb{F}}$, $\operatorname{Codim}(\operatorname{Sing}(f))$ is large for any nonzero f in the span of $\mathscr{F}$. It is not difficult to show this for any linear combination not involving the last two equations, so we only consider linear combinations involving the last two equations.

Let $f = \sum_{i=1}^k x_i^2 y_i, g = \sum_{i=1}^k y_i(x_i y_i - 2\sum_{j=1}^{i-1} x_j y_j)$. Let $h = c_1 f + c_2 g$ be a nontrivial linear combination of f and g. Consider the variety defined by $\{\partial_{x_i} g = 0\}_{i=1}^k$. We claim that this has codimension k. To see this, consider the ordering on monomials given by $m(x^\alpha y^\beta) = |\beta|$. Note that the leading monomials of the defining equations are given by $y_1^2, y_2^2, \ldots$; since these are mutually coprime, by Buchberger's first criterion these equations form a Gröbner basis [CLO13, Chapter 2, Theorem 6]. Since the largest subset S of variables with the property that no leading monomial depends only on the variables in S is given by $S = \{x_1, \ldots, x_k\}$, it follows that the codimension of this variety is k. Therefore by Proposition 6.4.6, $r(f) \geq k/2$. ■

Corollary 6.4.8. *Let $X_{p,k}$ denote the number of solutions to Equation (6.13) with $f_i, g_i \in \mathbb{F}_p[X]_{\leq 1}$. Then for $p > 3$ and sufficiently large k,*

$$|X_{p,k} - p^{4k-13}| \leq p^{4k-13-p^{\delta}}.$$

Proof. By Theorem 6.4.7, and Lemma 6.4.5 we conclude that $r(\mathscr{F}) \geq k/6$. The theorem then follows by Theorem 6.4.2. ■

Bibliography

[AB23] Josh Alman and Jarosław Błasiok. Matrix multiplication and number on the forehead communication. *arXiv preprint arXiv:2302.11476*, 2023. 54

[ACDM18] V. Arvind, A. Chatterjee, R. Datta, and P. Mukhopadhyay. Fast Exact Algorithms Using Hadamard Product of Polynomials. *ArXiv e-prints*, July 2018. 13

[ADH$^+$08] Noga Alon, Phuong Dao, Iman Hajirasouliha, Fereydoun Hormozdiari, and S Cenk Sahinalp. Biomolecular network motif counting and discovery by color coding. *Bioinformatics*, 24(13):i241–i249, 2008. 5, 8, 13

[AFS09] Omid Amini, Fedor V Fomin, and Saket Saurabh. Counting subgraphs via homomorphisms. In *International Colloquium on Automata, Languages, and Programming*, pages 71–82. Springer, 2009. 32

[AG07] Noga Alon and Shai Gutner. Balanced families of perfect hash functions and their applications. In *International Colloquium on Automata, Languages, and Programming*, pages 435–446. Springer, 2007. 11, 14, 30, 36

[AG09] Noga Alon and Shai Gutner. Balanced hashing, color coding and approximate counting. In *International Workshop on Parameterized and Exact Computation*, pages 1–16. Springer, 2009. 15, 32, 38

[AGV18] Nima Anari, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials, entropy, and a deterministic approximation algorithm for counting bases of matroids. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 35–46. IEEE, 2018. 6, 7, 8

[AH93] Herbert Abels and Stephan Holz. Higher generation by subgroups. *Journal of Algebra*, 160(2):310–341, 1993. 93

[AH20] Tigran Ananyan and Melvin Hochster. Small subalgebras of polynomial rings and stillman's conjecture. *Journal of the American Mathematical Society*, 33(1):291–309, 2020. 3

[AJQ$^+$20] Vedat Alev, Fernando Jeronimo, Dylan Quintana, Shashank Srivastava, and Madhur Tulsiani. List decoding of direct sum codes. In *Proceedings of the 14th annual Symposium on Discrete Algorithms (SODA)*, pages 1412–1425, 2020. 88

[AJT19] Vedat Levi Alev, Fernando Granha Jeronimo, and Madhur Tulsiani. Approximating constraint satisfaction problems on high-dimensional expanders. In *Proceedings of the 60th annual Symposium on Foundations of Computer Science (FOCS)*, pages

180–201, 2019. 88

[AL20] Vedat Levi Alev and Lap Chi Lau. Improved analysis of higher order random walks and applications. In *Proceedings of the 52nd annual Symposium on Theory of Computing (STOC)*, pages 1198–1211, 2020. 88

[ALGV19] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials ii: High-dimensional walks and an fpras for counting bases of a matroid. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 1–12, 2019. 1, 7

[Alo03] Noga Alon. Problems and results in extremal combinatorics–I. *Discrete Mathematics*, 273(1-3):31–53, 2003. 16, 24

[ALOV19] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials II: high-dimensional walks and an FPRAS for counting bases of a matroid. In *Proceedings of the 51st annual Symposium on Theory of Computing (STOC)*, pages 1–12, 2019. 88

[Alp93] Jonathan L Alperin. *Local representation theory: Modular representations as an introduction to the local representation theory of finite groups*. Cambridge University Press, 1993. 72

[ALSV13] Noga Alon, Troy Lee, Adi Shraibman, and Santosh Vempala. The approximate rank of a matrix and its algorithmic applications: approximate rank. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 675–684. ACM, 2013. 16

[AOGSS17] Nima Anari, Shayan Oveis Gharan, Amin Saberi, and Mohit Singh. Nash social welfare, matrix permanent, and stable polynomials. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017. 14

[AR02] Vikraman Arvind and Venkatesh Raman. Approximation algorithms for some parameterized counting problems. In *International Symposium on Algorithms and Computation*, pages 453–464. Springer, 2002. 14

[AS74] Miklós Ajtai and Endre Szemerédi. Sets of lattice points that form no squares. *Stud. Sci. Math. Hungar*, 9(1975):9–11, 1974. 75

[AS03] Noga Alon and Asaf Shapira. Testing subgraphs in directed graphs. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 700–709, 2003. 77

[Asc04] Michael Aschbacher. The status of the classification of the finite simple groups. *Notices of the American Mathematical Society*, 51(7):736–740, 2004. 87, 96

[Aud04] Michèle Audin. *Torus actions on symplectic manifolds*, volume 93 of *Progress in Mathematics*. Birkhäuser Verlag, Basel, revised edition, 2004. 66

[AW21] Josh Alman and Virginia Vassilevska Williams. A refined laser method and faster matrix multiplication. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, (SODA 2021)*, pages 522–539. SIAM, 2021. 51, 53

[AYZ95] Noga Alon, Raphael Yuster, and Uri Zwick. Color-coding. *Journal of the ACM (JACM)*, 42(4):844–856, 1995. 11, 14, 32

[Bab16] László Babai. Graph isomorphism in quasipolynomial time. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 684–697. ACM, 2016. 37

[Bal00] Cristina Ballantine. Ramanujan type buildings. *Canadian Journal of Mathematics*, 52(6):1121–1148, 2000. 87

[Bar95] Alexander I Barvinok. New algorithms for lineark-matroid intersection and matroid k-parity problems. *Mathematical Programming*, 69(1-3):449–470, 1995. 40

[Bar96] Alexander I Barvinok. Two algorithmic results for the traveling salesman problem. *Mathematics of Operations Research*, 21(1):65–84, 1996. 18

[Bax93] Eric T Bax. Inclusion and exclusion algorithm for the Hamiltonian path problem. *Information Processing Letters*, 47(4):203–207, 1993. 18

[BBDS12] Guillaume Blin, Paola Bonizzoni, Riccardo Dondi, and Florian Sikora. On the parameterized complexity of the repetition free longest common subsequence problem. *Information Processing Letters*, 112(7):272–276, 2012. 14

[BCC$^+$17a] Jonah Blasiak, Thomas Church, Henry Cohn, Joshua A. Grochow, Eric Naslund, William F. Sawin, and Chris Umans. On cap sets and the group-theoretic approach to matrix multiplication. *Discrete Anal.*, pages Paper No. 3, 1–27, 2017. 1, 52, 53, 57, 78, 82

[BCC$^+$17b] Jonah Blasiak, Thomas Church, Henry Cohn, Joshua A Grochow, and Chris Umans. Which groups are amenable to proving exponent two for matrix multiplication? *arXiv preprint arXiv:1712.02302*, 2017. 54, 70, 72

[BCC$^+$17c] Jonah Blasiak, Thomas Church, Henry Cohn, Joshua A. Grochow, and Chris Umans. Which groups are amenable to proving exponent two for matrix multiplication? *preprint*, 2017. arXiv1712.02302. 57

[BCG$^+$22] Jonah Blasiak, Henry Cohn, Josh Grochow, Kevin Pratt, and Chris Umans. Matrix multiplication via matrix groups. *Preprint*, 2022. 9, 55

[BCRL79] Dario Bini, Milvio Capovani, Francesco Romani, and Grazia Lotti. O (n2. 7799) complexity for n× n approximate matrix multiplication. *Information processing letters*, 8(5):234–235, 1979. 3

[BCS13] Peter Bürgisser, Michael Clausen, and Mohammad A Shokrollahi. *Algebraic complexity theory*, volume 315. Springer Science & Business Media, 2013. 49

[BCZ17] Markus Bläser, Matthias Christandl, and Jeroen Zuiddam. The border support rank of two-by-two matrix multiplication is seven. *arXiv preprint arXiv:1705.09652*, 2017. 16

[BDH18] Cornelius Brand, Holger Dell, and Thore Husfeldt. Extensor-coding. In *Symposium on Theory of Computing*. ACM, 2018. 14, 29, 37

[BDYW11] Boaz Barak, Zeev Dvir, Amir Yehudayoff, and Avi Wigderson. Rank bounds for design matrices with applications to combinatorial geometry and locally correctable codes. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 519–528. ACM, 2011. 16

[Bec16] Oren Becker. Symmetric unique neighbor expanders and good LDPC codes. *Discrete Applied Mathematics. The Journal of Combinatorial Algorithms, Informatics and Computational Sciences*, 211:211–216, 2016. 87

[Beh35] Felix Behrend. On sequences of numbers not divisible one by another. *Journal of the London Mathematical Society*, 1(1):42–44, 1935. 85

[Beh46] Felix A Behrend. On sets of integers which contain no three terms in arithmetical progression. *Proceedings of the National Academy of Sciences*, 32(12):331–332, 1946. 54

[BGT11] Emmanuel Breuillard, Ben Green, and Terence Tao. Suzuki groups as expanders. *Groups, Geometry, and Dynamics*, 5(2):281–299, 2011. 87

[BH06] Andreas Björklund and Thore Husfeldt. Inclusion–exclusion algorithms for counting set partitions. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 575–582. IEEE, 2006. 13, 14, 18

[BHK09] Andreas Björklund, Thore Husfeldt, and Mikko Koivisto. Set partitioning via inclusion-exclusion. *SIAM Journal on Computing*, 39(2):546–563, 2009. 18

[BIP19] Peter Bürgisser, Christian Ikenmeyer, and Greta Panova. No occurrence obstructions in geometric complexity theory. *Journal of the American Mathematical Society*, 32(1):163–193, 2019. 11

[Bjö84] Anders Björner. Some combinatorial and algebraic properties of Coxeter complexes and Tits buildings. *Advances in Mathematics*, 52(3):173–212, 1984. 93

[Bjö10a] Andreas Björklund. Determinant sums for undirected hamiltonicity. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 173–182. IEEE, 2010. 11, 14, 16, 17, 29

[Bjö10b] Andreas Björklund. Determinant sums for undirected hamiltonicity. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 173–182, 2010. 41

[BKL89] László Babai, William Kantor, and Alexander Lubotsky. Small-diameter Cayley graphs for finite simple groups. *European Journal of Combinatorics*, 10(6):507–522, 1989. 87

[BL06] Yonatan Bilu and Nathan Linial. Lifts, discrepancy and nearly optimal spectral gap. *Combinatorica*, 26(5):495–519, 2006. 6

[BP21] Cornelius Brand and Kevin Pratt. Parameterized applications of symbolic differentiation of (totally) multilinear polynomials. In *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, page 38:1–38:19. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021. 9, 40, 41

[Bra] Cornelius Brand. Patching colors with tensors. In *27th Annual European Symposium on Algorithms, ESA 2019, September 09-11, 2019, Munich, Germany*. 41

[Bre14] Emmanuel Breuillard. A brief introduction to approximate groups. In *Thin groups and superstrong approximation*, volume 61 of *Math. Sci. Res. Inst. Publ.*, pages 23–50. Cambridge Univ. Press, Cambridge, 2014. 59

[Car94] Élie Cartan. *Sur la structure des groupes de transformations finis et continus*. PhD thesis, Université de Paris, 1894. 94

[Car89] Roger Carter. *Simple groups of Lie type*. John Wiley & Sons, 1989. 94, 95, 96, 97, 104

[CCC15] Enrico Carlini, Maria Virginia Catalisano, and Luca Chiantini. Progress on the symmetric Strassen conjecture. *Journal of Pure and Applied Algebra*, 219(8):3149–3157, 2015. 33

[CCG11] Enrico Carlini, Maria Virginia Catalisano, and Anthony V Geramita. The solution to Waring's problem for monomials. *arXiv preprint arXiv:1110.0745*, 2011. 21

[CFK$^+$15] Marek Cygan, Fedor V. Fomin, Lukasz Kowalik, Daniel Lokshtanov, Daniel Marx, Marcin Pilipczuk, Michal Pilipczuk, and Saket Saurabh. *Parameterized Algorithms*. Springer, 2015. 41

[CGLM08] Pierre Comon, Gene Golub, Lek-Heng Lim, and Bernard Mourrain. Symmetric tensors and symmetric tensor rank. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1254–1279, 2008. 18

[CHI$^+$18] Luca Chiantini, Jonathan D Hauenstein, Christian Ikenmeyer, Joseph M Landsberg, and Giorgio Ottaviani. Polynomials and the exponent of matrix multiplication. *Bulletin of the London Mathematical Society*, 50(3):369–389, 2018. 11

[CKSU05] Henry Cohn, Robert Kleinberg, Balázs Szegedy, and Christopher Umans. Group-theoretic algorithms for matrix multiplication. In *Proceedings of the 46th Annual Symposium on Foundations of Computer Science (FOCS 2005)*, pages 379–388. IEEE Computer Society, 2005. 4, 52, 53, 70, 75, 76, 82

[CLO13] David Cox, John Little, and Donal OShea. *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*. Springer Science & Business Media, 2013. 112

[CLP17] Ernie Croot, Vsevolod F Lev, and Péter Pál Pach. Progression-free sets in are exponentially small. *Annals of Mathematics*, pages 331–337, 2017. 70

[CM23] Alex Cohen and Guy Moshkovitz. Partition and analytic rank are equivalent over large fields. *Duke Mathematical Journal*, 1(1):1–38, 2023. 1

[CMT04] Arjeh Cohen, Scott Murray, and Don Taylor. Computing in groups of Lie type. *Mathematics of Computation*, 73(247):1477–1498, 2004. 99

[Con98] Aldo Conca. Straightening law and powers of determinantal ideals of Hankel matrices. *Advances in Mathematics*, 138(2):263–292, 1998. 40, 45, 46

[CSZ03] Donald I. Cartwright, Patrick Solé, and Andrzej Żuk. Ramanujan geometries of type $\tilde{A}_n$. *Discrete Mathematics*, 269(1-3):35–43, 2003. 87

[CU03] Henry Cohn and Christopher Umans. A group-theoretic approach to fast matrix multiplication. In *Proceedings of the 44th Annual Symposium on Foundations of Computer Science (FOCS 2003)*, pages 438–449. IEEE Computer Society, 2003. 4, 9, 49, 52, 57, 58, 61, 63, 65, 66, 76

[CU13] Henry Cohn and Christopher Umans. Fast matrix multiplication using coherent configurations. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 1074–1086. Society for Industrial and Applied Mathematics, 2013. 16

[CW87] Don Coppersmith and Shmuel Winograd. Matrix multiplication via arithmetic progressions. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 1–6, 1987. 4, 49, 51

[DD19] Yotam Dikstein and Irit Dinur. Agreement testing theorems on layered set systems. In *Proceedings of the 60th annual Symposium on Foundations of Computer Science (FOCS)*, pages 1495–1524, 2019. 88

[DDFH18] Yotam Dikstein, Irit Dinur, Yuval Filmus, and Prahladh Harsha. Boolean function analysis on high-dimensional expanders. In *Proceedings of the 22nd annual International Conference on Randomization and Computation (RANDOM)*, volume 116, pages Art. No. 38, 20, 2018. 88

[DEL$^+$21] Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. Locally testable codes with constant rate, distance, and locality, 2021. Announced at `https://www.youtube.com/watch?v=pjc6GCRFnpg`. 88

[DEL$^+$22] Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. Locally testable codes with constant rate, distance, and locality. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 357–374, 2022. 1, 6, 8

[DFHT21] Irit Dinur, Yuval Filmus, Prahladh Harsha, and Madhur Tulsiani. Explicit SoS lower bounds from high-dimensional expanders. In *Proceedings of the 12th Innovations in Theoretical Computer Science Conference (ITCS)*, volume 185, pages 38:1–38:16, 2021. 88

[DHK$^+$21] Irit Dinur, Prahladh Harsha, Tali Kaufman, Inbal Livni Navon, and Amnon Ta-Shma. List-decoding with double samplers. *SIAM Journal on Computing*, 50(2):301–349, 2021. 88

[Din07] Irit Dinur. The pcp theorem by gap amplification. *Journal of the ACM (JACM)*, 54(3):12–es, 2007. 6

[DK17] Irit Dinur and Tali Kaufman. High dimensional expanders imply agreement expanders. In *Proceedings of the 58th annual Symposium on Foundations of Computer Science (FOCS)*, pages 974–985, 2017. 88

[Ede04] Yves Edel. Extensions of generalized product caps. *Designs, Codes and Cryptography*, 31:5–14, 2004. 54

[Edm67] Jack Edmonds. Systems of distinct representatives and linear algebra. *J. Res. Nat. Bur. Standards Sect. B 71*, (4):241–245, 1967. 41

[EG17] Jordan S Ellenberg and Dion Gijswijt. On large subsets of with no three-term arithmetic progression. *Annals of Mathematics*, pages 339–343, 2017. 54, 70

[EGH$^+$11] Pavel I Etingof, Oleg Golberg, Sebastian Hensel, Tiankai Liu, Alex Schwendner, Dmitry Vaintrob, and Elena Yudovina. *Introduction to representation theory*, volume 59. American Mathematical Soc., 2011. 72

[EGOW18] Klim Efremenko, Ankit Garg, Rafael Oliveira, and Avi Wigderson. Barriers for rank methods in arithmetic complexity. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018. 11, 19

[EJ10] Mikhail Ershov and Andrei Jaikin-Zapirain. Property (T) for noncommutative universal lattices. *Inventiones Mathematicae*, 179(2):303–347, 2010. 89

[EJK17] Mikhail Ershov, Andrei Jaikin-Zapirain, and Martin Kassabov. Property (T) for groups graded by root systems. *Memoirs of the American Mathematical Society*, 249(1186):v+135, 2017. 89

[EKZ20] Shai Evra, Tali Kaufman, and Gilles Zémor. Decodable quantum LDPC codes beyond the square root distance barrier using high dimensional expanders. In *Proceedings of the 61st annual Symposium on Foundations of Computer Science (FOCS)*, pages 218–227, 2020. 88

[ESS67] P Erdös, A Sárközy, and E Szemerédi. On a theorem of behrend. *Journal of the Australian Mathematical Society*, 7(1):9–16, 1967. 85

[FG12] Jason Fulman and Robert Guralnick. Bounds on the number and sizes of conjugacy classes in finite Chevalley groups with applications to derangements. *Trans. Amer. Math. Soc.*, 364(6):3023–3070, 2012. 57, 61

[FGL$^+$12] Jacob Fox, Mikhail Gromov, Vincent Lafforgue, Assaf Naor, and János Pach. Overlap properties of geometric expanders. *Journal für die Reine und Angewandte Mathematik. (Crelle's Journal)*, 671:49–83, 2012. 88

[FH13] William Fulton and Joe Harris. *Representation theory: a first course*, volume 129. Springer Science & Business Media, 2013. 72

[FL17] Jacob Fox and LászlóMiklós Lovász. A tight bound for green's arithmetic triangle removal lemma in vector spaces. In *Proceedings of the twenty-eighth annual acm-siam symposium on discrete algorithms*, pages 1612–1617. SIAM, 2017. 77, 78, 81

[FLPS16] Fedor V. Fomin, Daniel Lokshtanov, Fahad Panolan, and Saket Saurabh. Efficient computation of representative families with applications in parameterized and exact algorithms. *J. ACM*, 63(4):29:1–29:60, 2016. 40

[FLR$^+$12] Fedor V Fomin, Daniel Lokshtanov, Venkatesh Raman, Saket Saurabh, and BV Raghavendra Rao. Faster algorithms for finding and counting subgraphs. *Journal of Computer and System Sciences*, 78(3):698–706, 2012. 12, 37

[Fox11] Jacob Fox. A new proof of the graph removal lemma. *Annals of Mathematics*, pages 561–579, 2011. 77

[Fri08] Joel Friedman. *A proof of Alon's second eigenvalue conjecture and related problems*. American Mathematical Soc., 2008. 5

[GAP21] The GAP Group. *GAP – Groups, Algorithms, and Programming, Version 4.11.1*, 2021. https://www.gap-system.org. 98

[Gar73] Howard Garland. p-adic curvature and the cohomology of discrete subgroups of p-adic groups. *Annals of Mathematics. Second Series*, 97:375–423, 1973. 6, 89

[Gar79] Peter Garst. *Cohen–Macaulay complexes and group actions*. PhD thesis, Unviersity of Wisconsin–Madison, 1979. 93

[Gat14] Andreas Gathmann. Algebraic geometry. 2014. 23, 24

[GGOW19] Ankit Garg, Leonid Gurvits, Rafael Oliveira, and Avi Wigderson. Operator scaling: theory and applications. *Foundations of Computational Mathematics*, pages 1–68, 2019. 41

[GK23] Roy Gotlib and Tali Kaufman. No where to go but high: A perspective on high dimensional expanders. *arXiv preprint arXiv:2304.10106*, 2023. 7

[Gly13] David G Glynn. Permanent formulae from the Veronesean. *Designs, codes and cryptography*, 68(1-3):39–47, 2013. 19

[Gow08] W. T. Gowers. Quasirandom groups. *Combin. Probab. Comput.*, 17(3):363–387, 2008. 55, 58, 71

[Gre05] Ben Green. A szemerédi-type regularity lemma in abelian groups, with applications. *Geometric & Functional Analysis GAFA*, 15(2):340–376, 2005. 77

[GRWZ18a] Gregory Gutin, Felix Reidl, Magnus Wahlström, and Meirav Zehavi. Designing deterministic polynomial-space algorithms by color-coding multivariate polynomials. *Journal of Computer and System Sciences*, 95:69–85, 2018. 9, 32, 36

[GRWZ18b] Gregory Z. Gutin, Felix Reidl, Magnus Wahlström, and Meirav Zehavi. Designing deterministic polynomial-space algorithms by color-coding multivariate polynomials. *J. Comput. Syst. Sci.*, 95:69–85, 2018. 41

[GS06] Oded Goldreich and Madhu Sudan. Locally testable codes and pcps of almost-linear length. *Journal of the ACM (JACM)*, 53(4):558–655, 2006. 8

[GS13] Sylvain Guillemot and Florian Sikora. Finding and counting vertex-colored subtrees. *Algorithmica*, 65(4):828–844, 2013. 14

[GT07] Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the gowers norms. *arXiv preprint arXiv:0711.3191*, 2007. 9, 111

[Gur03] Leonid Gurvits. Classical deterministic complexity of Edmonds' problem and quantum entanglement. In *Proceedings of the Thirty-fifth Annual ACM Symposium on Theory of Computing*, STOC '03, pages 10–19, New York, NY, USA, 2003. ACM. 19, 25, 41

[Gur04] Leonid Gurvits. Combinatorial and algorithmic aspects of hyperbolic polynomials. *arXiv preprint math/0404474*, 2004. 14

[Gur06] Leonid Gurvits. Hyperbolic polynomials approach to Van der Waerden/Schrijver-

Valiant like conjectures: sharper bounds, simpler proofs and algorithmic applications. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 417–426. ACM, 2006. 14

[GW11] William T Gowers and Julia Wolf. Linear forms and higher-degree uniformity for functions on. *Geometric and Functional Analysis*, 21(1):36–69, 2011. 3

[Hal03] Brian Hall. *Lie groups, Lie algebras, and representations: an elementary introduction*. Springer, 2003. 95

[Har92] Joe Harris. *Algebraic geometry: a first course*, volume 133 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1992. 65

[HH18] Sergei Haller and Max Horn. Unipot — a system for computing with elements of unipotent subgroups of Chevalley groups, 2018. `https://www.gap-system.org/Packages/unipot.html`. 98

[Hil16] David Hill. Prove that the sum of all simple roots is a root. Mathematics Stack Exchange, 2016. `https://math.stackexchange.com/q/1760142` (version: 2016-04-26). 96

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin (new series) of the American Mathematical Society*, 43(4):439–561, 2006. 5, 6, 7

[HM22] Nguyen Ngoc Hung and Attila Maróti. p-regular conjugacy classes and p-rational irreducible characters. *Journal of Algebra*, 607:387–425, 2022. 72

[hp] Fedor Petrov (https://mathoverflow.net/users/4312/fedor petrov). A variant of the corners problem. MathOverflow. URL:https://mathoverflow.net/q/451594 (version: 2023-07-27). 85

[HRT01] Robert Howlett, Leanne Rylands, and Donald Taylor. Matrix generators for exceptional groups of Lie type. *Journal of Symbolic Computation*, 31(4):429–445, 2001. 97

[HS19] Prahladh Harsha and Ramprasad Saptharishi. A note on the elementary HDX construction of Kaufman–Oppenheim. Technical Report 1912.11225, arXiv, 2019. 89, 93, 107

[Hum72] James Humphreys. *Introduction to Lie algebras and representation theory*. Springer–Verlag, 1972. 95

[Hum95] James Humphreys. *Linear algebraic groups*. Springer–Verlag, 1995. 98

[HWZ08] Falk Hüffner, Sebastian Wernicke, and Thomas Zichner. Algorithm engineering for color-coding with applications to signaling pathway detection. *Algorithmica*, 52(2):114–132, 2008. 11, 14, 32

[IK99] Anthony Iarrobino and Vassil Kanev. *Power sums, Gorenstein algebras, and determinantal loci*. Springer Science & Business Media, 1999. 11, 15, 18, 19, 27

[Jel13] Joachim Jelisiejew. An upper bound for the waring rank of a form. *arXiv preprint arXiv:1305.6957*, 2013. 26

[Kar82] Richard M Karp. Dynamic programming meets the principle of inclusion and exclusion. *Operations Research Letters*, 1(2):49–51, 1982. 18

[Kas07] Martin Kassabov. Symmetric groups and expander graphs. *Inventiones Mathematicae*, 170(2):327–354, 2007. 87

[KGK77] Samuel Kohn, Allan Gottlieb, and Meryle Kohn. A generating function approach to the traveling salesman problem. In *Proceedings of the 1977 annual conference*, pages 294–300. ACM, 1977. 13, 18

[KI04] Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Computational Complexity*, 13(1-2):1–46, 2004. 41

[KL12] Tali Kaufman and Alexander Lubotzky. Edge transitive Ramanujan graphs and highly symmetric LDPC good codes. In *Proceedings of the 44th annual Symposium on Theory of Computing (STOC)*, pages 359–366, 2012. 87, 88

[KL14] Tali Kaufman and Alexander Lubotzky. High dimensional expanders and property testing. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 501–506, 2014. 7

[KLN06] Martin Kassabov, Alexander Lubotzky, and Nikolay Nikolov. Finite simple groups as expanders. *Proceedings of the National Academy of Sciences of the United States of America*, 103(16):6116–6119, 2006. 87

[KM20] Tali Kaufman and David Mass. Local-to-global agreement expansion via the variance method. In *Proceedings of the 11th annual Innovations in Theoretical Computer Science Conference (ITCS)*, pages 74:1–74:14, 2020. 88

[KM23] Zander Kelley and Raghu Meka. Strong bounds for 3-progressions. *arXiv preprint arXiv:2302.05537*, 2023. 54

[KO18] Tali Kaufman and Izhar Oppenheim. Construction of new local spectral high dimensional expanders. In *Proceedings of the 50th Annual Symposium on Theory of Computing (STOC)*, pages 773–786, 2018. 9, 88, 89, 90, 92, 93, 100, 106, 107, 110

[KO20] Tali Kaufman and Izhar Oppenheim. High order random walks: beyond spectral gap. *Combinatorica*, 40(2):245–281, 2020. 88

[Kou08] Ioannis Koutis. Faster algebraic algorithms for path and packing problems. In *International Colloquium on Automata, Languages, and Programming*, pages 575–586. Springer, 2008. 11, 14, 17, 29, 30

[KS08] Tali Kaufman and Madhu Sudan. Algebraic property testing: the role of invariance. In *Proceedings of the 40th annual Symposium on Theory of Computation (STOC)*, pages 403–412. ACM, New York, 2008. 87

[KSS16] Robert Kleinberg, Will Sawin, and David E Speyer. The growth rate of tri-colored sum-free sets. *arXiv preprint arXiv:1607.00047*, 2016. 70

[KT21] Tali Kaufman and Ran Tessler. New cosystolic expanders from tensors imply explicit quantum LDPC codes with $\omega(\sqrt{n}\log^k n)$ distance. In *Proceedings of the 53rd annual Symposium on Theory of Computing (STOC)*, pages 1317–1329, 2021.

88

[KW09] Ioannis Koutis and Ryan Williams. Limits and applications of group algebras for parameterized problems. In *International Colloquium on Automata, Languages, and Programming*, pages 653–664. Springer, 2009. 8, 13, 14, 15

[KW15] Ioannis Koutis and Ryan Williams. Algebraic fingerprints for faster algorithms. *Communications of the ACM*, 59(1):98–105, 2015. 14

[KW16] Tali Kaufman and Avi Wigderson. Symmetric LDPC codes and local testing. *Combinatorica*, 36(1):91–120, 2016. 87

[KZ20] David Kazhdan and Tamar Ziegler. Properties of high rank subvarieties of affine spaces. *Geometric and Functional Analysis*, 30(4):1063–1096, 2020. 111

[Laf02] Laurent Lafforgue. Chtoucas de Drinfeld et correspondance de Langlands. *Inventiones Mathematicae*, 147(1):1–241, 2002. 87

[Lan50] Folke Lannér. On complexes with transitive groups of automorphisms. *Communications du Séminaire Mathématique de l'Université de Lund*, 11:71, 1950. 93

[Lan12] Joseph M Landsberg. Tensors: geometry and applications. *Representation theory*, 381:402, 2012. 11, 19, 25, 48

[Lan17] J. M. Landsberg. *Geometry and Complexity Theory*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2017. 11, 19

[Lee16] Hwangrae Lee. Power sum decompositions of elementary symmetric polynomials. *Linear Algebra and its Applications*, 492:89–97, 2016. 12, 13, 15, 19, 22, 31, 35, 38

[Li04] Wen-Ching Winnie Li. Ramanujan hypergraphs. *Geometric and Functional Analysis*, 14(2):380–399, 2004. 87, 88

[LMT13] Michael Larsen, Gunter Malle, and Pham Huu Tiep. The largest irreducible representations of simple groups. *Proc. Lond. Math. Soc. (3)*, 106(1):65–96, 2013. 63

[LPS88] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988. 5

[LS74] Vicente Landazuri and Gary M. Seitz. On the minimal degrees of projective representations of the finite Chevalley groups. *J. Algebra*, 32:418–443, 1974. 57, 61, 71, 72

[LSV05a] Alexander Lubotzky, Beth Samuels, and Uzi Vishne. Explicit constructions of Ramanujan complexes of type $\tilde{A}_d$. *European Journal of Combinatorics*, 26(6):965–993, 2005. 87, 88

[LSV05b] Alexander Lubotzky, Beth Samuels, and Uzi Vishne. Ramanujan complexes of type $\tilde{A}_d$. *Israel Journal of Mathematics*, 149:267–299, 2005. Probability in mathematics. 87, 88

[LZ22] Amichai Lampert and Tamar Ziegler. Schmidt rank and algebraic closure. *arXiv preprint arXiv:2205.05329*, 2022. 112

[Mac94] Francis Sowerby Macaulay. *The algebraic theory of modular systems*, volume 19. Cambridge University Press, 1994. 27

[Mar09] Dániel Marx. A parameterized view on matroid optimization problems. *Theor. Comput. Sci.*, 410(44):4471–4479, 2009. 40

[Mat79] Rudolf Mathon. A note on the graph isomorphism counting problem. *Information Processing Letters*, 8(3):131–136, 1979. 37

[Mes95] Roy Meshulam. On subsets of finite abelian groups with no 3-term arithmetic progressions. *Journal of Combinatorial Theory, Series A*, 71(1):168–172, 1995. 54

[Mil19] Luka Milićević. Polynomial bound for partition rank in terms of analytic rank. *Geometric and Functional Analysis*, 29:1503–1530, 2019. 9, 111

[Mon85] Burkhard Monien. How to find long paths efficiently. In *North-Holland Mathematics Studies*, volume 109, pages 239–254. Elsevier, 1985. 41

[MT11] Gunter Malle and Donna Testerman. *Linear algebraic groups and finite groups of Lie type*, volume 133 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2011. 57, 61

[MV97] Meena Mahajan and V Vinay. A combinatorial algorithm for the determinant. In *In Proceedings of the 8th Annual ACM-SIAM Symposium on Discrete Algorithms*. Citeseer, 1997. 41

[Ned09] Jesper Nederlof. Fast polynomial-space algorithms using möbius inversion: Improving on steiner tree and related problems. In *International Colloquium on Automata, Languages, and Programming*, pages 713–725. Springer, 2009. 13

[OP22] Ryan O'Donnell and Kevin Pratt. High-dimensional expanders from chevalley groups. In *37th Computational Complexity Conference (CCC 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022. 9

[Opp18] Izhar Oppenheim. Local spectral expansion approach to high dimensional expanders part I: Descent of spectral gaps. *Discrete & Computational Geometry*, 59(2):293–330, 2018. 7, 89, 92

[Pet16] Fedor Petrov. Combinatorial results implied by many zero divisors in a group ring. *arXiv preprint arXiv:1606.03256*, 2016. 70, 71

[PK21] Pavel Panteleev and Gleb Kalachev. Asymptotically good quantum and locally testable classical LDPC codes. Technical Report 2111.03654, arXiv, 2021. 1, 6, 7, 88

[Pra19] Kevin Pratt. Waring rank, parameterized and exact algorithms. In *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, MD, USA, November 9-12, 2019*, 2019. 5, 9

[Pra22] Kevin Pratt. A note on slice rank and matchings in groups. *arXiv preprint arXiv:2210.05488*, 2022. 9

[Raz13] Ran Raz. Tensor-rank and lower bounds for arithmetic formulas. *Journal of the ACM (JACM)*, 60(6):1–15, 2013. 1, 2

[RBG01] Lajos Rónyai, László Babai, and Murali Ganapathy. On the number of zero-patterns of a sequence of polynomials. *Journal of the American Mathematical Society*, 14(3):717–735, 2001. 39

[Rei08] Omer Reingold. Undirected connectivity in log-space. *Journal of the ACM (JACM)*, 55(4):1–24, 2008. 6

[Rot53] Klaus F Roth. On certain sets of integers. *J. London Math. Soc*, 28(104-109):3, 1953. 54

[RS11] Kristian Ranestad and Frank-Olaf Schreyer. On the rank of a symmetric form. *Journal of Algebra*, 346(1):340–342, 2011. 2, 16, 19, 21, 22, 23, 28, 33

[Rys64] H.J. Ryser. *Combinatorial Mathematics*. Carus Mathematical Monographs. Cambridge University Press, 1964. 13, 18

[S^+69] Volker Strassen et al. Gaussian elimination is not optimal. *Numerische mathematik*, 13(4):354–356, 1969. 1, 2, 3, 49, 50

[S^+77] Jean-Pierre Serre et al. *Linear representations of finite groups*, volume 42. Springer, 1977. 72

[Sar04] Alireza Sarveniazi. *Ramanujan Hypergraph Based on Bruhat–Tits Building*. PhD thesis, University of Göttingen, 2004. 87, 88

[Saw18] Will Sawin. Bounds for matchings in nonabelian groups. *Electron. J. Combin.*, 25(4):Paper No. 4.23, 1–21, 2018. 54, 57, 69

[Sch81] Arnold Schönhage. Partial and total matrix multiplication. *SIAM Journal on Computing*, 10(3):434–455, 1981. 4, 49, 50

[Sch85] Wolfgang M Schmidt. The density of integer points on homogeneous varieties. *Acta Mathematica*, 154(3-4):243–296, 1985. 3

[Sha15] Masoumeh Sepideh Shafiei. Apolarity for determinants and permanents of generic matrices. *Journal of Commutative Algebra*, 7(1):89–123, 2015. 42

[Shi17] Yaroslav Shitov. A counterexample to Strassen's direct sum conjecture. *arXiv preprint arXiv:1712.08660*, 2017. 33

[Shi18] Yaroslav Shitov. A counterexample to comon's conjecture. *SIAM Journal on Applied Algebra and Geometry*, 2(3):428–443, 2018. 2

[Shk06] Ilya D Shkredov. On a generalization of szemerédi's theorem. *Proceedings of the London Mathematical Society*, 93(3):723–760, 2006. 85

[Sho90] Victor Shoup. New algorithms for finding irreducible polynomials over finite fields. *Mathematics of Computation*, 54(189):435–447, 1990. 99

[SS96] Michael Sipser and Daniel A Spielman. Expander codes. *IEEE transactions on Information Theory*, 42(6):1710–1722, 1996. 6

[Ste16] Robert Steinberg. *Lectures on Chevalley groups*, volume 66 of *University Lecture Series*. American Mathematical Society, 2016. Notes prepared by John Faulkner and Robert Wilson, revised and corrected edition of the 1968 original. 96, 97, 98, 99, 104

[Str86] Volker Strassen. The asymptotic spectrum of tensors and the exponent of matrix multiplication. In *27th Annual Symposium on Foundations of Computer Science (sfcs 1986)*, pages 49–54. IEEE, 1986. 3, 4, 49, 51

[SV$^+$09] Oriol Serra, Lluís Vena, et al. A combinatorial proof of the removal lemma for groups. *Journal of Combinatorial Theory, Series A*, 116(4):971–978, 2009. 77

[Syl52] J.J. Sylvester. On the principles of the calculus of forms. *Cambridge and Dublin Mathematical Journal*, 7:52–97, 1852. 2, 19

[Tao16] Terence Tao. Symmetric formulation of the Croot-Lev-Pach-Ellenberg-Gijswijt capset bound, 2016. Available at https://terrytao.wordpress.com. 70

[Tei14] Zach Teitler. Geometric lower bounds for generalized ranks. *arXiv preprint arXiv:1406.5145*, 2014. 20, 25

[TS16] Terence Tao and Will Sawin. Notes on the "slice rank" of tensors, 2016. Available at https://terrytao.wordpress.com/2016/08/24/notes-on-the-slice-rank-of-tensors/. 70, 71

[Tsu19] Dekel Tsur. Faster deterministic parameterized algorithm for k-Path. *Theoretical Computer Science*, 790:96–104, 2019. 40, 41

[VW09] Virginia Vassilevska and Ryan Williams. Finding, minimizing, and counting weighted subgraphs. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 455–464. ACM, 2009. 12

[WGE16] Michael Walter, David Gross, and Jens Eisert. Multi-partite entanglement. *arXiv preprint arXiv:1612.02437*, 2016. 16

[Wil09a] Ryan Williams. Finding paths of length k in $O^*(2^k)$ time. *Information Processing Letters*, 109(6):315–318, 2009. 11, 14, 17, 29, 30

[Wil09b] Ryan Williams. Finding paths of length k in $O^*(2^k)$ time. *Inf. Process. Lett.*, 109(6):315–318, 2009. 41

[WXXZ23] Virginia Vassilevska Williams, Yinzhan Xu, Zixuan Xu, and Renfei Zhou. New bounds for matrix multiplication: from alpha to omega. *arXiv preprint arXiv:2307.07970*, 2023. 3, 4, 49, 75

[Zha23] Yufei Zhao. *Graph Theory and Additive Combinatorics: Exploring Structure and Randomness*. Cambridge University Press, 2023. 54, 76, 77

[Żuk03] Andrzej Żuk. Property (T) and Kazhdan constants for discrete groups. *Geometric and Functional Analysis*, 13(3):643–670, 2003. 6, 92