Waring Rank, Parameterized and Exact Algorithms

Kevin Pratt
Computer Science Department
Carnegie Mellon University
Pittsburgh, United States
kpratt@andrew.cmu.edu

***Abstract*—We show that the Waring rank (symmetric tensor rank) of a certain family of polynomials has intimate connections to the areas of parameterized and exact algorithms, generalizing some well-known methods and providing a concrete approach to obtain faster approximate counting and deterministic decision algorithms.**

To illustrate the amenability and utility of this approach, we give an algorithm for approximately counting subgraphs of bounded treewidth, improving on earlier work of Alon, Dao, Hajirasouliha, Hormozdiari, and Sahinalp. Along the way we give an exact answer to an open problem of Koutis and Williams and sharpen a lower bound on the size of perfectly balanced hash families given by Alon and Gutner.

***Keywords*-Parameterized complexity; Algebraic computation; Approximation algorithms;**

I. Introduction

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

$$f = \ell_1^d + \cdots + \ell_r^d, \tag{1}$$

for some linear forms $\ell_1, \ldots, \ell_r \in \mathcal{S}_1^n$. 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 [1, Introduction]. It has enjoyed a recent resurgence of popularity within algebraic geometry [1], [2] and has connections in computer science to the limiting exponent of matrix multiplication ω [3], the Mulmuley-Sohoni Geometric Complexity Theory program [4], and several other areas in algebraic complexity [5], [6]. This paper adds *parameterized algorithms* to this list, showing that several methods in this area (color-coding methods [7]–[9], the group-algebra/determinant sum approach [10]–[12], and inclusion-exclusion methods) fundamentally result from rank upper bounds for a specific family of polynomials. 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 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 paper, which has been largely unexploited, is that in general this is a question about Waring rank.

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

Question 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]

Question 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 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, Lokshtanov, Raman, Saurabh, and Rao [13] 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 [14].

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,

$$f_G \coloneqq \operatorname{tr}(A_G^d) = \sum_{\substack{\text{closed walks} \\ (v_{i_1}, v_{i_2}, \ldots, v_{i_d}) \in G}} x_{i_1} \cdots x_{i_d} \in \mathcal{S}_d^n. \tag{2}$$

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

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

2575-8454/19/$31.00 ©2019 IEEE
DOI 10.1109/FOCS.2019.00053

$g = \sum_\alpha b_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ are both elements of $\mathcal{S}_d^n$,

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

The algorithm is based on two easy observations:

Observation 4. *The number of simple cycles of length d in G equals $e_{n,d}(\partial \mathbf{x}) f_G$.*

Observation 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 $\mathbf{R}(e_{n,d}) = A^0(n,d)$ evaluations of f_G. It was shown in [15] that

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

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,

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

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

$$\frac{1}{2^{d-1}} \sum_{\substack{S \subset [n] \\ |S| \leq \lfloor d/2 \rfloor}} (-1)^{|S|} \binom{n - \lfloor d/2 \rfloor - |S| - 1}{\lfloor d/2 \rfloor - |S|}. \quad (3)$$
$$f_G(\delta_{S,1}, \ldots, \delta_{S,n}).$$

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.

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 an open problem asked by Koutis and Williams [16] in a completely black-box way.[2] Moreover, it follows from a special case of our Theorem 6 that *any* algorithm must make $\mathbf{R}(e_{n,d}) \geq \Omega(\binom{n}{\leq \lfloor d/2 \rfloor})$ [15] queries to compute $e_{n,d}(\partial \mathbf{x}) f$ in the black-box setting:

[2]An alternate solution to this problem was given contemporaneously in [17].

Theorem 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 $\mathbf{R}(g)$, 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 3:

Theorem 7. *Let $f \in \mathbb{R}_{\geq 0}[x_1, \ldots, x_n]_d$ 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 $\{\pm 1\}^n$.

The algorithm and the proof behind Theorem 7 are simple and can be found in Section IV. Applying this theorem to to the graph polynomial f_G, an algorithm for approximately counting simple cycles of length d is immediate. More generally, we have the following:

Theorem 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.

After the writing of this paper, a $O(4^d n^{O(1)})$ time deterministic algorithm for this problem was given in [18]. The previous fastest algorithm ran in time $5.44^d n^{\mathrm{tw}(H)+O(1)} \varepsilon^{-2}$-time algorithm of Alon, Dao, Hajirasouliha, Hormozdiari, and Sahinalp [19], improving on a $5.44^{d \log \log d} n^{\mathrm{tw}(H)+O(1)} \varepsilon^{-2}$-time algorithm of Alon and Gutner [8]. The first parameterized algorithm for a variant of this problem was given by Arvind and Raman [20] 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, Dell, and Husfeldt [21] runs in time $4^d n^{\mathrm{pw}(H)+O(1)} \varepsilon^{-2}$. We stress that this application is only a motivating example – Theorem 7 is extremely general and can be applied to approximately count set partitions and packings [22], dominating sets [16], repetition-free longest common subsequences [23], and functional motifs in biological networks [24].

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 7 (and hence Theorem 8) from 4.075 to 2. Specifically, we raise the following question:

Question 9. Is $A^\varepsilon(n, d) \leq 2^d \cdot \mathrm{poly}(n, \varepsilon^{-1})$?

Prior to this work it was believed [25] 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 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. [26]. In particular, a result of Anari, Oveis Gharan, Saberi, and Singh [27] 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 paper shows that the general (i.e., *unstable*) case raises interesting questions as well.

A. 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 [10], [11], [28]:

Problem 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 11 with one-sided error (Proposition 22 (a)). This suggests studying upper bounds on $A(n,d)$ (Question 1) as an approach to solve Problem 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 [7] 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, Wernicke, and Zichner [9] follows from an upper bound on $A(n,d)$ of $O(4.32^d \log n)$ (Remark 60). The group-algebra/determinant sum approach of [10]–[12] reduces to answering a generalization of Question 1 (see Definition 48) in the case that the underlying field is not $\mathbb{C}$ but of characteristic 2. (In Theorem 52 we give the essentially optimal upper bound of $2^d - 1$ for this variant, which in turn can be used to recover [10]–[12]). Prior to this work, no connection of this precision between these methods was known.

Question 1 provides insight into lower bounds on previous methods as well. For example, the bounds on $\mathbf{R}(e_{n,d})$ given in [15] directly yield asymptotically sharper lower bounds than those given by Alon and Gutner [29, Theorem 1] on the size of *perfectly balanced hash families* used by exact-counting color-coding algorithms (Theorem 74). This improvement is ultimately a consequence of Bézout's theorem in algebraic geometry. Question 1 and a classical lower bound on Waring rank (Theorem 16) explain why *disjointness matrices* arose in the context of lower bounds on color coding [29] and the group-algebra approach [16]: they are the partial derivatives matrices of the elementary symmetric polynomials.

Our main answers to Question 1 are the following. By our Theorems 28, 41 and 59, it follows that

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

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 1 as an approach to obtain algorithms faster than $2^d\mathrm{poly}(n)$ for Problem 11; moreover, we show in Theorem 24 that there is also a lower bound of 2^{d-1} on the number of queries needed to solve Problem 11 with one-sided error.

It is easily seen by Observation 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 68 that if $A^+(33700,4) \le 10$ then $A^+(n,d) \le O(3.9999^d \log n)$. It follows from our Theorem 28 and Theorem 59 that

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

and from our Corollary 36 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

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

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.

B. 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 [1, Appendix A]). Except for in Section IV, we assume that arithmetic operations can be performed with infinite precision and at unit cost.

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

Problem 12. Fix $g \in \mathcal{S}_d^n$. Given black-box access to $f \in \mathcal{S}_d^n$,

a) Compute $g(\partial \mathbf{x})f$.
b) Compute a $(1 \pm \varepsilon)$ approximation of $g(\partial \mathbf{x})f$ (assuming $f, g \in \mathbb{R}_{\geq 0}[x_1, \ldots, x_n]$).
c) Determine if $\operatorname{supp}(f) \cap \operatorname{supp}(g) = \emptyset$.

The fundamental connection between Waring rank and Problem 12 (a) is given by our Theorem 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 24. 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 Problems 12 (a) to 12 (c). A related notion of support rank for tensors has previously appeared in the context of ω and quantum communication complexity [30]–[32], 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 [33], [34].

In Section III we study $A(n, d)$ and its variants. We start in Section III-A by proving negative results, showing that $A(n, d) \geq 2^{d-1}$ (Theorem 28), and that for sufficiently large n, $A(n, 2) = 3$ (Proposition 33) and $A(n, 3) \geq 5$ (Corollary 31). Using bounds on the *ε-rank* of the identity matrix [35], we show in Corollary 36 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 68 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.

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 [36]. On this note, we show in Proposition 30 that Question 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 III is focused on general upper bounds on $A(n, d)$ and its variants. Proposition 38 will give a simple explanation as to why *determinant sums* (as in the title of [12]) can be computed in a parameterized way: for all $d \times n$ matrices A and B, the Waring rank of

$$\sum_{\substack{\alpha \in \{0,1\}^n \\ |\alpha| = d}} \det(A_\alpha B_\alpha) x_1^{\alpha_1} \cdots x_n^{\alpha_n} \tag{4}$$

is at most $\mathbf{R}(\det_d)$. A special case of this example is used in Theorem 41 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 43 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 49 that $A_{\mathsf{k}}(n, d) \geq 2^{d-1}$. Theorem 52 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 (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 [10], [11] and the basis of [12] 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 [10], [11] is given by Theorem 54.

In Section III-C 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 67). This method also allows us to recursively bound $A^{\varepsilon}(n, d)$ for fixed d (Theorem 57). 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 59.

In Section IV we give applications of the previous section. We start by giving the proof Theorem 7, which is then used to prove Theorem 8. We end with an improved lower bound on the size of perfectly-balanced hash families in Theorem 74.

We conclude by giving several standalone problems.

II. Preliminaries and Methods

We use multi-index notation: for $f \in \mathcal{S}_d^n$, 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 $\mathbb{N}_d^n := \{\alpha \in \mathbb{N}^n : |\alpha| = d\}$, and similarly $\{0,1\}_d^n := \{\alpha \in \{0,1\}^n : |\alpha| = d\}$. 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 $\ell = \sum_{i=1}^n a_i x_i \in \mathcal{S}_1^n$, 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 $\det_d$, $\mathrm{per}_d \in \mathsf{k}[x_{ij} : i, j \in [d]]_d$ denote the degree-d determinant and permanent polynomials, respectively. Recall that the permanent is defined by

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

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 paper. The second shows that Waring rank is always defined (i.e., finite).

Theorem 13. *Let $f \in \mathcal{S}^n_d$ and let $j \geq d$.*

a) [1, Lemma 1.15(i)] For all $\ell_1, \ldots, \ell_r \in \mathcal{S}^n_1$,

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

b) [37, Lemma 3.5] For all $g \in \mathcal{S}^n_d$, $f(\partial \mathbf{x}) g = g(\partial \mathbf{x}) f$.

Theorem 14. *[1, Corollary 1.16]* $\mathbf{R}(f) \leq \dim \mathcal{S}^n_d = \binom{n+d-1}{d}$.

Importantly, Theorems 13 (a) and 13 (b) imply that $g(\partial \mathbf{x}) f$ can be computed with $\mathbf{R}(g)$ queries in Problem 12 (a), as noted in Observation 5. We will show in the next subsection that this is optimal, even if we are allowed to query f adaptively.

Example 15. The following Waring decomposition of $e_{n,d}$ is easily seen by inclusion-exclusion:

$$d! \cdot e_{n,d} = \sum_{\substack{\alpha \in \{0,1\}^n \\ |\alpha| \leq d}} (-1)^{|\alpha|+d} \binom{n - |\alpha|}{d - |\alpha|} \left(\sum_{i=1}^{n} \alpha_i x_i \right)^d. \tag{5}$$

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

$$\mathrm{P}_A \coloneqq (A_{1,1}x_1 + \cdots + A_{1,n}x_n) \cdots (A_{n,1}x_1 + \cdots + A_{n,n}x_n).$$

It is easily seen that the coefficient of $x_1 \cdots x_n$ in P_A equals the permanent of A. In other words, $\mathrm{per}(A) = e_{n,n}(\partial \mathbf{x}) \mathrm{P}_A$. It follows directly from Theorem 13 and Equation (5) that

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

which is Ryser's formula for computing the permanent [38]. As another example, applying Theorem 13 and Equation (5) to the closed-walk generating polynomial Equation (2), 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 [39] and rediscovered several times thereafter [40], [41]. As a third example, let $S_1, \ldots, S_m \subseteq [k \cdot r]$, where $|S_i| = r$ for all i. Note that that the coefficient of $x_1 \cdots x_{kr}$ in $\mathrm{Part}_{S_1,\ldots,S_m} \coloneqq \left(\sum_{i=1}^{m} \prod_{j \in S_i} x_j \right)^k$ equals the number of ordered partitions of $[kr]$ into k of the sets S_i. Therefore the number of such partitions equals

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

which was given in [22], [42]. 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 [43].

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 (5). This immediately raises the question: what is $\mathbf{R}(x_1 \cdots x_n)$? This was only answered recently in [36], where a lower bound on the *degree* of a form's *apolar subscheme* was used to show that $\mathbf{R}(x_1 \cdots x_n) = 2^{n-1}$.[3] This lower bound shows that the above algorithms are, in a restricted sense, optimal. Similar observations have been made in [44], [45].

Although the Waring decomposition of Equation (5) is essentially optimal in the case when $n = d$, it is far from optimal in general. Indeed, Equation (5) only shows that $\mathbf{R}(e_{n,d}) \leq \binom{n}{\leq d}$, whereas it was shown in [15] that for d odd, $\mathbf{R}(e_{n,d}) = \binom{n}{\leq \lfloor d/2 \rfloor}$, and for d even,

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

A. 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 [46]. 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 [5, Section 6.2.2].

[3] 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.

Theorem 16. *[1, pg. 11] For all $g \in \mathcal{S}_d^n$ and integers $u, v \geq 0$ such that $u + v = d$,*

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

Remark 17. 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 16 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 [2, Section 3.2] that the rank for *almost all* $g \in \mathcal{S}_d^n$ 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 [5, Section 10.1] and [6].

Example 18. It is a classical fact from linear algebra that for $g \in \mathcal{S}_2^n$, $\mathbf{R}(g) = \operatorname{rank}(Cat_g(1, 1))$. 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 $g_j^\perp := \ker Cat_g(j, d-j)$ 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.

Lemma 19. *[47, Theorem 4.2] Let $\ell_1, \ldots, \ell_r \in \mathcal{S}_1^n$ be pairwise linearly independent. Then for all $g \in \mathcal{S}_d^n$, $g \in \operatorname{span}\{\ell_1^d, \ldots, \ell_r^d\}$ if and only if $I(\{\ell_1^*, \ldots, \ell_r^*\})_d \subset g_d^\perp$.*

A complete answer to the complexity of Problem 12 (a) is now in hand.

Theorem 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 $\mathbf{R}(g)$, assuming unit-cost arithmetic operations.*

Proof: The upper bound is immediate from Theorem 13 (b). To prove the lower bound we first show the following: for any pairwise linearly independent points $v_1, \ldots, v_m \in \mathbb{C}^n$ where $m < \mathbf{R}(g)$, there exists a $p \in \mathcal{S}_d^n$ such that $p \in I(\{v_1, \ldots, v_m\})$ but $g(\partial \mathbf{x}) p \neq 0$. If this were not the case, there exist pairwise linearly independent points $v_1, \ldots, v_m$ such that $I(\{v_1, \ldots, v_m\})_d \subset g_d^\perp$. 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 $v_1, \ldots, v_m$, 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 $g(\partial \mathbf{x}) f \neq g(\partial \mathbf{x})(p + f)$. ∎

B. Support Rank, Nonnegative Support Rank, and ε-Support Rank

We now introduce variants of Waring rank of algorithmic relevance.

Definition 20. The support rank and nonnegative support rank of $f \in \mathcal{S}_d^n$ are given by

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

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

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

Note that condition in the definition of $\mathbf{R}_{\operatorname{supp}}^\varepsilon$ 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 21. 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 22. *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 $\mathbf{R}_{\operatorname{supp}}(g)$ queries.*

b) For a fixed $g \in \mathcal{S}_d^n$ 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 decides if $\operatorname{supp}(g) \cap \operatorname{supp}(f)$ using $\mathbf{R}_{\operatorname{supp}}^+(g)$ 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 $\mathbf{R}_{\operatorname{supp}}^\varepsilon(g)$ queries.

Proof:

a. Let $U \subseteq \mathbb{C}$, where $|U| \geq d/\delta$. Let $a_1, \ldots, a_n$ be indeterminates. Note that $g(\partial \mathbf{x}) f(a_1 x_1, \ldots, a_n x_n)$ is not identically zero in $\mathbb{C}[a_1, \ldots, a_n]$ if and only if $\operatorname{supp}(f) \cap \operatorname{supp}(g) \neq \emptyset$. Then by choosing $a_1, \ldots, a_n$ uniformly at random from U, $g(\partial \mathbf{x}) f(a_1 x_1, \ldots, a_n x_n)$ will evaluate to zero whenever $\operatorname{supp}(f) \cap \operatorname{supp}(g) = \emptyset$, and whenever $\operatorname{supp}(f) \cap \operatorname{supp}(g) \neq \emptyset$ this does not evaluate to zero with

probability at least $1-\delta$ by the Schwartz-Zippel lemma. By Theorem 13, $g(\partial\mathbf{x})f(a_1x_1,\dots,a_nx_n)$ can be computed using $\mathbf{R}(g)$ 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 $\operatorname{supp}(f)\cap\operatorname{supp}(g)\neq\emptyset$. The result follows from Theorem 13.

c. This is immediate from Theorem 13.

■

It follows from a variation of the proof of Theorem 6 that Proposition 22 (a) is optimal for monomials:

Proposition 23. *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 $\mathbf{R}_{\operatorname{supp}}(x^\alpha)=\prod_{i=1}^n(1+\alpha_i)/\min_{i\in[n]}(1+\alpha_i)$ queries.*

Proof: The upper bound follows from Theorem 13 (b); in fact, this shows that we can compute $\partial^\alpha f$ exactly using $\mathbf{R}(x^\alpha)$ queries.

For the lower bound, given any $f\in\mathcal{S}_d^n$ where $\alpha\in\operatorname{supp}(f)$, suppose a support intersection certification algorithm queries f at pairwise linearly independent points $v_1,\dots,v_m$, where $m<\mathbf{R}(x^\alpha)$. Then by the Apolarity Lemma, there exists a $p\in\mathcal{S}_d^n$ such that $p\in I(\{v_1,\dots,v_m\})$ but $\partial^\alpha p\neq 0$ (see the proof of Theorem 6). Note that the condition that $\partial^\alpha p\neq 0$ is equivalent to saying that $\alpha\in\operatorname{supp}(p)$. Therefore there exists some $\lambda\in\mathbb{C}$ such that $\alpha\notin\operatorname{supp}(f+\lambda p)$. 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 $\mathbf{R}(x^\alpha)$ given in [48]. ■

Theorem 24. *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 $f(x_1,\dots,x_d,0,\dots,0)$. 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 $\mathbf{R}(x_1\cdots x_d)=2^{d-1}$ [36], this contradicts Proposition 23. ■

III. Support Ranks of Elementary Symmetric Polynomials

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

Problem 25. Determine $A(n,d):=\mathbf{R}_{\operatorname{supp}}(e_{n,d})$, $A^+(n,d):=\mathbf{R}^+_{\operatorname{supp}}(e_{n,d})$ and $A^\varepsilon(n,d):=\mathbf{R}^\varepsilon_{\operatorname{supp}}(e_{n,d})$.

Obviously $A(n,d)\le A^+(n,d)\le A^\varepsilon(n,d)$, and for all n, $A(n,1)=1$. It follows from [36] that $A^\varepsilon(n,n)=2^{n-1}$ and from [15] 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 25 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 27 (a) that $A(n,d),A^+(n,d)$, and $A^\varepsilon(n,d)$ are nondecreasing in n, in Proposition 38 that $A^+(\mathbb{N},d)$ is finite for each d, and in Corollary 36 that $A^\varepsilon(\mathbb{N},d)$ is infinite for $\varepsilon<1/2$ and $d>1$.

For notational convenience, we define

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

Remark 26. Our upper bounds to Problem 25 will be obtained by the following general method. We start with some $f\in\mathcal{S}_d^m$ whose rank is known. We then find $L_1,\dots,L_m\in\mathcal{S}_1^n$, where $n\gg m$, so that $f(L_1,\dots,L_m)\in\mathfrak{E}(n,d)$. This will show that

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

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 60 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 29.

A. 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 27. *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 $\mathbf{R}(f')\le\mathbf{R}(f)$ 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 $A(n,d)\le\mathbf{R}(\partial_{n+1}f)\le\mathbf{R}(f)$, where the final inequality follows from Theorem 13 (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 28. *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 [36] that $\mathbf{R}(x_1 \cdots x_d) = 2^{d-1}$, and therefore $A(d,d) = 2^{d-1}$. The theorem is then immediate from Proposition 27 (a). ∎

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

Proposition 29. *$A(n,d) \leq r$ 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 $\mathbf{R}(f) \leq r$ 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 $\mathbf{R}(f) \leq r$. 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

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

It is immediate that $\mathbf{R}(f') \leq \mathbf{R}(f)$. 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 30. *$A(n,d) \leq r$ 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 $v_j \coloneqq ((\ell_1{}^*)_j, (\ell_2{}^*)_j, \ldots, (\ell_r{}^*)_j)$ 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 29 we conclude that $A(n,d) \leq r$. ∎

Corollary 31. $5 \leq A(8,3) \leq A(\mathbb{N},3)$.

Proof: Suppose for contradiction that $A(8,3) = 4$. By Proposition 30, 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 $\mathbb{CP}^3$ [49, Lemma 11.1], a contradiction. ∎

Remark 32. 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 [49, Exercise 11.10.b]).

The $d = 2$ case of Problem 25 is solved using linear algebra.

Proposition 33. $A(\mathbb{N},2) = 3$.

Proof: It suffices by Example 18 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 34. *[35, 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 35. *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 18 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 34.

To show (b), let $m \coloneqq O(\log n/\varepsilon^2)$. 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 36. *For all $0 < \varepsilon < 1/2$ and $d \geq 2$, $A^\varepsilon(\mathbb{N},d) = \infty$.*

Proof: Fix $0 < \varepsilon < 1/2$. By Proposition 35 (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 27, for all n

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

a contradiction. ∎

B. Upper Bounds via the Determinant

The relevance of the determinant to Problem 25 is immediate from Proposition 29. 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 29, we conclude that $A(n,d) \leq \mathbf{R}(\det_d)$. We now make this more explicit.

Definition 37. Let $d \leq n$. For $A, B \in \mathbb{C}^{d \times n}$, let

$$g_{A,B} \coloneqq \sum_{\alpha \in \{0,1\}^n_d} \det_d(A_\alpha B_\alpha) x^\alpha. \tag{6}$$

Proposition 38. *For all $A, B \in \mathbb{C}^{d \times n}$,*

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

Furthermore, $A^+(\mathbb{N},d) \leq \mathbf{R}(\det_d) \leq (5/6)^{\lfloor d/3 \rfloor} 2^{d-1} d!$ 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 $\mathbf{R}(\det_d) \leq (5/6)^{\lfloor d/3 \rfloor} 2^{d-1} d!$ [47, Example 1.14].

Note that by taking A and B to have positive minors[4], $g_{A,B} \in \mathfrak{E}^+(n,d)$. This shows that $A^+(\mathbb{N},d) \leq \mathbf{R}(\det_d)$. Since Proposition 27 (a) shows that $(A^+(n,d))_n$ is nondecreasing, it follows that the limit $A^+(\mathbb{N},d)$ exists. ∎

Remark 39. The asymptotically best known lower bound on $\mathbf{R}(\det_d)$ is $\binom{d}{\lfloor d/2 \rfloor}^2$, which follows from the method of partial derivatives [44] [2, Theorem 9.3.2.1]. Therefore one cannot hope to improve the upper bound given by Proposition 38 exponentially beyond 4^d by finding a better upper bound on the Waring rank of the determinant.

Definition 40. Let $h_d \in \mathcal{S}^{2d-1}_d$ 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 41.

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

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

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,

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

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 $A^+(\mathbb{N},d) \leq \mathbf{R}(h_d)$. Since h_d is a degree-d polynomial in $2d-1$ variables, by the dimension bound of Theorem 14 we have that $\mathbf{R}(h_d) \leq \binom{3d-2}{d}$, and therefore $A^+(\mathbb{N},d) \leq \binom{3d-2}{d}$. The theorem follows from Stirling's approximation. ∎

Remark 42. The above theorem can be slightly improved by using the state-of-the-art bound [50] 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 $n, d \geq 3$, which shows that

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

It follows from Remark 17 that the lower bound on $\mathbf{R}(h_d)$ 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 43. *For all integers $d, u, v > 0$ such that $u + v = d$,*

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

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 $f \coloneqq g_{A,A}$. Furthermore we assume that $u \leq v$; this is without loss of generality as $Cat_f(u,v) = Cat_f(v,u)^T$. We will then show that $\operatorname{rank}(Cat_f(u,v)) \leq m \coloneqq \binom{2v+u}{u}$. 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 $\mathrm{supp}(\alpha) \subseteq [2d-1]$.

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 $\prod_{i\neq j \in I\cup J}(a_i - a_j)^2$ otherwise. Hence the row indexed by I is a multiple of $\prod_{i\neq j\in I}(a_i-a_j)^2$, and similarly the column indexed by J is a multiple of $\prod_{i\neq j\in J}(a_i-a_j)^2$. Therefore $M = D_1 Q D_2$ 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 $I = \{i_1,\ldots,i_u\}$, $J = \{j_1,\ldots,j_v\}$. We now claim that there exist $g_1, h_1, \ldots, g_m, h_m$ with $g_i \in \mathcal{S}^u$, $h_i \in \mathcal{S}^v$, such that

$$Q_{IJ} = \sum_{k=1}^{m} g_k(a_{i_1},\ldots,a_{i_u})h_k(a_{j_1},\ldots,a_{j_v}). \tag{7}$$

To see this, view Q_{IJ} as a polynomial in the variables $a_{i_1},\ldots,a_{i_u}$ with coefficients in $\mathbb{C}[a_{j_1},\ldots,a_{j_v}]$. 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 (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 $\mathbb{C}[a_{j_1},\ldots,a_{j_v}]$. 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

$$Q = \sum_{k=1}^{m} (g_k(a_{i_1},\ldots,a_{i_u}))_I^T (h_k(a_{j_1},\ldots,a_{j_v}))_J,$$

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

Remark 44. Numerical evidence suggests that equality holds in Theorem 43 when $u = \lfloor d/2 \rfloor$. This would imply that $\mathbf{R}(h_d) = \Omega(2.59^d)$.

C. $A(n,d)$ in Positive Characteristic and Abelian Group Algebras

We briefly introduce a generalization of Waring rank to $\mathcal{S}_d^n(\mathsf{k}) \coloneqq \mathsf{k}[x_1,\ldots,x_n]_d$, where k is a field of arbitrary characteristic. This notion has been studied extensively as early as 1916 [51], 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 [1]. Assume k is algebraically closed unless stated otherwise.

Definition 45. For $\ell = \sum_{i=1}^n a_i x_i \in \mathcal{S}_1^n(\mathsf{k})$, let

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

Note that $\ell^{[d]}$ is just ℓ^d without any multinomial coefficients. We remark that the projectivization of the set $\{\ell^{[d]} : \ell \in \mathcal{S}_1^n(\mathsf{k})\}$ is the classical *Veronese variety* in algebraic geometry [1, Corollary A.10].

Definition 46. For $f \in \mathcal{S}_d^n(\mathsf{k})$, let $\mathbf{R}^\nu(f)$ be the minimum r such that there exist linear forms $\ell_1,\ldots,\ell_r$ with

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

and let

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

The next proposition shows that the $d = j$ case of Theorem 13 (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 47. *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 48. Let $A_{\mathsf{k}}(n,d) \coloneqq \mathbf{R}^\nu_{\mathrm{supp}}(e_{n,d})$.

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

Theorem 49. *For all $n \geq d$, $A_{\mathsf{k}}(n,d) \geq 2^{d-1}$.*

Proof: It follows from an argument identical to that of Proposition 27 (a) that $A(d,d) \leq A(n,d)$ for all $n \geq d$. As it was shown in [36] that $\mathbf{R}^\nu(x_1\cdots x_d) \geq 2^{d-1}$, the conclusion follows. ∎

Definition 50. Given $A \in \mathsf{k}^{d\times n}$, let

$$g_A \coloneqq \sum_{\alpha\in\mathbb{N}_d^n} \mathrm{per}_d(A_\alpha)x^\alpha. \tag{8}$$

Lemma 51. *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 $L_i \coloneqq \sum_{j=1}^n A_{ij}y_j \in \mathsf{k}[y_1,\ldots,y_n]$. 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 (5)) that this coefficient equals

$$\sum_{\alpha \in \{0,1\}^d} (-1)^{|\alpha|+d} (\sum_{i=1}^{n} x_i \sum_{j=1}^{d} \alpha_j A_{i,j})^{[d]}. \tag{9}$$

■

Theorem 52. *If* k *is infinite and* $\mathrm{char}(\mathsf{k}) = 2$, $A_{\mathsf{k}}(n,d) \leq 2^d - 1$.

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

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

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 51. ■

Remark 53. Theorem 52 gives the following $2^d \mathrm{poly}(n)$-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

$$\sum_{\alpha \in \{0,1\}^d} f(a_1 \sum_{j=1}^{d} \alpha_j A_{1,j}, \ldots, a_n \sum_{j=1}^{d} \alpha_j A_{n,j}). \tag{10}$$

It follows from Proposition 47, Theorem 52, and the Schwartz-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 [10] 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 [11] 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 [12] for detecting Hamiltonian cycles reduces to computing Equation (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 [12] and shows that [10], [11] were in fact also computing "determinant sums". This connection was made earlier in [21].

The algorithms of [10], [11] were presented in terms of a property of abelian group algebras. The following theorem elucidates the connection between support rank and this property.

Theorem 54. *Let* G *be an abelian group, and let* $y_1, \ldots, y_n \in \mathsf{k}[G]$. *For* $\alpha \in \mathbb{N}^n$, *let* $f_\alpha \coloneqq \prod_{i=1}^n y_i^{\alpha_i}$. *Define*

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

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

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 $\alpha \in \mathbb{N}_d^n$, $f_\alpha(\mathrm{Id}_G) \neq 0$ if and only if $\alpha \in T$. Note that $f_\alpha(\mathrm{Id}_G) \neq 0$ 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 $\mathrm{tr}(D \cdot \rho(f_\alpha)) \neq 0$ if and only if $\alpha \in T$. Note that

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

Let $M_i \coloneqq D^{1/n} \Lambda_i$. 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 $P \coloneqq \sum_{i=1}^n \ell_i^{[d]}$ 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 54 allows to to recover the approach of [10], [11] 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 $y_i \coloneqq \mathrm{Id}_G + v_i \in \mathsf{k}[G]$ for all i in the statement of Theorem 54. The key fact used in [10], [11] 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 [10], [11] then follow by using the decomposition given by Theorem 54. 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*.

D. 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.

1) A Recursive Bound on $A^\varepsilon(n,d)$ for Fixed d: We will first need the following tool introduced in [8].

Definition 55. 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 [8]; we reproduce the proof for completeness. Here $(n)_k \coloneqq n(n-1)\cdots(n-k+1)$ denotes the falling factorial.

Lemma 56. *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 $p \coloneqq \frac{(l)_k}{l^k}$ and $M \coloneqq \lceil \frac{8(k\log n+1)}{p(\delta-1)^2} \rceil$. 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 $2e^{-(\delta-1)^2 pM/8}$. 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

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

■

Theorem 57. *Suppose $f \in \mathfrak{E}^{\varepsilon_0}(n_0,d)$ where $0 < \varepsilon_0 < 1$. Then for all $\varepsilon_0 < \varepsilon < 1$ and all $n \geq d$,*

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

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

Proof: If $n \leq n_0$ the theorem follows from Proposition 27 (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,

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

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, $\mathbf{R}(f') \leq M \cdot \mathbf{R}(f)$, and the theorem follows by the bound on M given by Lemma 56. ■

Remark 58. 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 63.

Theorem 59. *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 57,

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

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 $\mathbf{R}(e_{n_0,d})$ given in [15],

$$\begin{aligned} A^\varepsilon(n,d) &\leq O\left(\binom{n_0}{\lfloor d/2 \rfloor} \frac{n_0^d}{(n_0)_d} \varepsilon^{-2} d^2 \log n\right) \\ &= O\left(\frac{\lceil cd \rceil^d}{\lfloor d/2 \rfloor!} \frac{\lceil d(c-1) \rceil!}{\lceil d(c-1/2) \rceil!} \varepsilon^{-2} d^2 \log n\right). \end{aligned}$$

Applying Stirling's inequality,

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

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 60. If we take $f = x_1 x_2 \cdots x_d$ and use the upper bound on $\mathbf{R}(x_1 \cdots x_d)$ given by Equation (5), it follows from Theorem 57 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 Equation (2), 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 [29], 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 [7]. 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 (5).

An improvement to color-coding was made in [9] 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 57 with $f = e_{n_0,d}$ and using the suboptimal bound on $\mathbf{R}(e_{n_0,d})$ given by Equation (5),

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

In fact, the choice of $n_0 = \lceil 1.3d\rceil$ is optimal if we are using the rank bound of Equation (5); this follows from the same calculation done in [52, Section 8]. The algorithm resulting from this bound was virtually described in [52], [53].

2) A Recursive Bound on $A^+(n,d)$ for all n and d:

Definition 61. For $g \in \mathcal{S}_d^n$ and $s,t \in \mathbb{N}$, let

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

In other 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 62. *For all $g \in \mathcal{S}_d^n$, $\mathbf{R}(g^{\circledast(s,t)}) \le s((d+1)\mathbf{R}(g))^t$.*

Proof: By subadditivity of Waring rank, $\mathbf{R}(g^{\circledast(s,t)}) \le s\mathbf{R}(g^{\circledast(1,t)})$. Now letting $r = \mathbf{R}(g)$, 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 $\mathbf{R}(\prod_{i=1}^t x_i^d) \le (d+1)^t$ (which follows from e.g. Equation (5)), it follows that $\mathbf{R}(g^{\circledast(s,t)}) \le s\mathbf{R}(g^{\circledast(1,t)}) \le s((d+1)\mathbf{R}(g))^t$. ∎

Remark 63. 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 [54] for more. It was recently shown in [55] that the tensor version of this conjecture is false; if the polynomial version is also false, the upper bound of Lemma 62 may not be optimal.

Definition 64. 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 65. For $n\ge d$, $n_0\ge d_0$, and $d_0 \mid d$, let

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

Proposition 66. *There exists an (n,d,n_0,d_0)-perfect splitter of size $\sigma(n,d,n_0,d_0)$.*

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

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

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

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

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

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

■

Theorem 67. *Let $f \in \mathfrak{E}^+(n_0, d_0)$. Then for all integers n, d where $n \geq d$,*

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

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 $L_{i,j,k} \coloneqq \sum_{m \in \pi_i^{-1}(j,k)} x_m$. 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

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

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 66,

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

Now suppose that $d_0 \nmid d$. By Proposition 27 (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 $s((d_0+1)\mathbf{R}(f))^{\lceil d/d_0 \rceil}$ by a reduction to the case when $d_0 \mid d$. ■

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

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

recovering Theorem 57 in the case of nonnegative support rank.

Example 68. Theorem 67 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 $A^+(\mathbb{N}, 4) \leq 10$. Then we have that for all $n_0 \geq 4$ and all n, d,

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

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

In contrast, the best upper bound we know on $A^+(\mathbb{N}, 4)$ is 79, which follows from Remark 42. When used in Theorem 67 this only shows that $A^+(n,d) \leq O(6.706^d \log n)$.

IV. Applications

We now recall and prove Theorem 7.

Theorem 7. *Let $f \in \mathbb{R}_{\geq 0}[x_1, \ldots, x_n]_d$ 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 $\{\pm 1\}^n$.

Proof: Set $n_0 \coloneqq \lceil 1.55d \rceil$, $p \coloneqq (n_0)_d / n_0^d$, and $M \coloneqq \lceil 3\varepsilon^{-2}/p \rceil$. 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 $L_{\pi,i} \coloneqq \sum_{j \in \pi^{-1}(i)} x_j$. The algorithm will compute and return

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

By Theorem 13 (a) and the upper bound on $e_{n_0,d}$ given in [15], for d odd this equals

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

and for d even equals

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

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

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

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

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

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

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

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

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

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 69. In order to derandomize Theorem 7, it would suffice to give a near-optimal construction of a $(1+\varepsilon)$-balanced $(n, d, 1.55d)$-splitter, as first defined in [8]. We note that such a construction was given for ("unbalanced") $(n, k, \alpha k)$-splitters for all $\alpha \ge 1$ in [52]. Furthermore, note that for any *fixed* values of n and d, Theorem 7 can be made deterministic by taking $\mathcal{F}$ to be a $(1+\varepsilon)$-balanced $(n, d, 1.55d)$-splitter of optimal size.

A. Approximately Counting Subgraphs of Bounded Treewidth

We now give an application of Theorem 7. First, we recall the notion of treewidth:

Definition 70. 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 71. For graphs G, H, where $|G| = n$ and $|H| = d$, let

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

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 [13], [21].

Lemma 72. *[21, 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 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.

Proof: We first construct a formula C computing $P_{H,G}$ using Lemma 72. 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 $\log f(1, 1, \ldots, 1) = \log(|\mathrm{Hom}(H, G)|) \le d \log n$.

Next note that $e_{n,d}(\partial \mathbf{x}) P_{H,G}$ equals the number of injective homomorphisms from H to G. Using Theorem 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 [56] and the quasi-polynomial time graph isomorphism algorithm of [57].

The total time taken is

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

■

B. Lower Bounds on Perfectly Balanced Hash Families

In this section we show how the bounds on $\mathbf{R}(e_{n,d})$ given in [15] imply lower bounds on the size of perfectly balanced hash families.

Definition 73. [29, 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 74. *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 $L_{\pi,i} \coloneqq \sum_{j \in \pi^{-1}(i)} x_j$. Consider the polynomial

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

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

V. Open Problems

Question 75. 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 43 shows that this is at most 2.6^d.

Question 76. 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 77. 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 29 and the fact that $\mathbf{R}(\mathrm{per}_d) \leq 4^{d-1}$ [2] that $A(\mathbb{N}, d) \leq 4^{d-1}$.

Question 78. Do all (g, ε)-support intersection certification algorithms require $\mathbf{R}_{\mathrm{supp}}(g)$ queries? Proposition 23 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 6 shows that this is true when $\varepsilon = 0$.

Remark 79. Theorem 67 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.

Acknowledgment

I am very grateful to Ryan O'Donnell for numerous comments and suggestions, as well as feedback on an earlier draft of this paper. I would also like to thank Ryan Williams and an anonymous reviewer for comments on an earlier draft.

References

[1] A. Iarrobino and V. Kanev, *Power sums, Gorenstein algebras, and determinantal loci*. Springer Science & Business Media, 1999.

[2] J. M. Landsberg, "Tensors: geometry and applications," *Representation theory*, vol. 381, p. 402, 2012.

[3] L. Chiantini, J. D. Hauenstein, C. Ikenmeyer, J. M. Landsberg, and G. Ottaviani, "Polynomials and the exponent of matrix multiplication," *Bulletin of the London Mathematical Society*, vol. 50, no. 3, pp. 369–389, 2018.

[4] P. Bürgisser, C. Ikenmeyer, and G. Panova, "No occurrence obstructions in geometric complexity theory," *Journal of the American Mathematical Society*, vol. 32, no. 1, pp. 163–193, 2019.

[5] J. M. Landsberg, *Geometry and Complexity Theory*, ser. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2017.

[6] K. Efremenko, A. Garg, R. Oliveira, and A. 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.

[7] N. Alon, R. Yuster, and U. Zwick, "Color-coding," *Journal of the ACM (JACM)*, vol. 42, no. 4, pp. 844–856, 1995.

[8] N. Alon and S. Gutner, "Balanced families of perfect hash functions and their applications," in *International Colloquium on Automata, Languages, and Programming*. Springer, 2007, pp. 435–446.

[9] F. Hüffner, S. Wernicke, and T. Zichner, "Algorithm engineering for color-coding with applications to signaling pathway detection," *Algorithmica*, vol. 52, no. 2, pp. 114–132, 2008.

[10] I. Koutis, "Faster algebraic algorithms for path and packing problems," in *International Colloquium on Automata, Languages, and Programming*. Springer, 2008, pp. 575–586.

[11] R. Williams, "Finding paths of length k in $O^*(2^k)$ time," *Information Processing Letters*, vol. 109, no. 6, pp. 315–318, 2009.

[12] A. Björklund, "Determinant sums for undirected hamiltonicity," in *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*. IEEE, 2010, pp. 173–182.

[13] F. V. Fomin, D. Lokshtanov, V. Raman, S. Saurabh, and B. R. Rao, "Faster algorithms for finding and counting subgraphs," *Journal of Computer and System Sciences*, vol. 78, no. 3, pp. 698–706, 2012.

[14] V. Vassilevska and R. Williams, "Finding, minimizing, and counting weighted subgraphs," in *Proceedings of the forty-first annual ACM symposium on Theory of computing*. ACM, 2009, pp. 455–464.

[15] H. Lee, "Power sum decompositions of elementary symmetric polynomials," *Linear Algebra and its Applications*, vol. 492, pp. 89–97, 2016.

[16] I. Koutis and R. Williams, "Limits and applications of group algebras for parameterized problems," in *International Colloquium on Automata, Languages, and Programming*. Springer, 2009, pp. 653–664.

[17] V. Arvind, A. Chatterjee, R. Datta, and P. Mukhopadhyay, "Fast Exact Algorithms Using Hadamard Product of Polynomials," *ArXiv e-prints*, Jul. 2018.

[18] A. Björklund, D. Lokshtanov, S. Saurabh, and M. Zehavi, "Approximate Counting of k-Paths: Deterministic and in Polynomial Space," in *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, ser. Leibniz International Proceedings in Informatics (LIPIcs), C. Baier, I. Chatzigiannakis, P. Flocchini, and S. Leonardi, Eds., vol. 132. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019, pp. 24:1–24:15. [Online]. Available: http://drops.dagstuhl.de/opus/volltexte/2019/10600

[19] N. Alon, P. Dao, I. Hajirasouliha, F. Hormozdiari, and S. C. Sahinalp, "Biomolecular network motif counting and discovery by color coding," *Bioinformatics*, vol. 24, no. 13, pp. i241–i249, 2008.

[20] V. Arvind and V. Raman, "Approximation algorithms for some parameterized counting problems," in *International Symposium on Algorithms and Computation*. Springer, 2002, pp. 453–464.

[21] C. Brand, H. Dell, and T. Husfeldt, "Extensor-coding," in *Symposium on Theory of Computing*. ACM, 2018.

[22] A. Björklund and T. Husfeldt, "Inclusion–exclusion algorithms for counting set partitions," in *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2006, pp. 575–582.

[23] G. Blin, P. Bonizzoni, R. Dondi, and F. Sikora, "On the parameterized complexity of the repetition free longest common subsequence problem," *Information Processing Letters*, vol. 112, no. 7, pp. 272–276, 2012.

[24] S. Guillemot and F. Sikora, "Finding and counting vertex-colored subtrees," *Algorithmica*, vol. 65, no. 4, pp. 828–844, 2013.

[25] I. Koutis and R. Williams, "Algebraic fingerprints for faster algorithms," *Communications of the ACM*, vol. 59, no. 1, pp. 98–105, 2015.

[26] L. 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*. ACM, 2006, pp. 417–426.

[27] N. Anari, S. Oveis Gharan, A. Saberi, and M. 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.

[28] L. Gurvits, "Combinatorial and algorithmic aspects of hyperbolic polynomials," *arXiv preprint math/0404474*, 2004.

[29] N. Alon and S. Gutner, "Balanced hashing, color coding and approximate counting," in *International Workshop on Parameterized and Exact Computation*. Springer, 2009, pp. 1–16.

[30] H. Cohn and C. Umans, "Fast matrix multiplication using coherent configurations," in *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2013, pp. 1074–1086.

[31] M. Bläser, M. Christandl, and J. Zuiddam, "The border support rank of two-by-two matrix multiplication is seven," *arXiv preprint arXiv:1705.09652*, 2017.

[32] M. Walter, D. Gross, and J. Eisert, "Multi-partite entanglement," *arXiv preprint arXiv:1612.02437*, 2016.

[33] B. Barak, Z. Dvir, A. Yehudayoff, and A. 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*. ACM, 2011, pp. 519–528.

[34] N. Alon, T. Lee, A. Shraibman, and S. 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*. ACM, 2013, pp. 675–684.

[35] N. Alon, "Problems and results in extremal combinatorics–I," *Discrete Mathematics*, vol. 273, no. 1-3, pp. 31–53, 2003.

[36] K. Ranestad and F.-O. Schreyer, "On the rank of a symmetric form," *Journal of Algebra*, vol. 346, no. 1, pp. 340–342, 2011.

[37] P. Comon, G. Golub, L.-H. Lim, and B. Mourrain, "Symmetric tensors and symmetric tensor rank," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1254–1279, 2008.

[38] H. Ryser, *Combinatorial Mathematics*, ser. Carus Mathematical Monographs. Cambridge University Press, 1964.

[39] S. Kohn, A. Gottlieb, and M. Kohn, "A generating function approach to the traveling salesman problem," in *Proceedings of the 1977 annual conference*. ACM, 1977, pp. 294–300.

[40] R. M. Karp, "Dynamic programming meets the principle of inclusion and exclusion," *Operations Research Letters*, vol. 1, no. 2, pp. 49–51, 1982.

[41] E. T. Bax, "Inclusion and exclusion algorithm for the Hamiltonian path problem," *Information Processing Letters*, vol. 47, no. 4, pp. 203–207, 1993.

[42] A. Björklund, T. Husfeldt, and M. Koivisto, "Set partitioning via inclusion-exclusion," *SIAM Journal on Computing*, vol. 39, no. 2, pp. 546–563, 2009.

[43] A. I. Barvinok, "Two algorithmic results for the traveling salesman problem," *Mathematics of Operations Research*, vol. 21, no. 1, pp. 65–84, 1996.

[44] L. Gurvits, "Ryser (or polarization) formula for the permanent is essentially optimal: the Waring rank approach," *Los Alamos Technical report, LA-UR-08-06583*, 2008.

[45] D. G. Glynn, "Permanent formulae from the Veronesean," *Designs, codes and cryptography*, vol. 68, no. 1-3, pp. 39–47, 2013.

[46] J. Sylvester, "On the principles of the calculus of forms," *Cambridge and Dublin Mathematical Journal*, vol. 7, pp. 52–97, 1852.

[47] Z. Teitler, "Geometric lower bounds for generalized ranks," *arXiv preprint arXiv:1406.5145*, 2014.

[48] E. Carlini, M. V. Catalisano, and A. V. Geramita, "The solution to Waring's problem for monomials," *arXiv preprint arXiv:1110.0745*, 2011.

[49] A. Gathmann, "Algebraic geometry," 2014.

[50] J. Jelisiejew, "An upper bound for the waring rank of a form," *arXiv preprint arXiv:1305.6957*, 2013.

[51] F. S. Macaulay, *The algebraic theory of modular systems*. Cambridge University Press, 1994, vol. 19.

[52] G. Gutin, F. Reidl, M. Wahlström, and M. Zehavi, "Designing deterministic polynomial-space algorithms by color-coding multivariate polynomials," *Journal of Computer and System Sciences*, vol. 95, pp. 69–85, 2018.

[53] O. Amini, F. V. Fomin, and S. Saurabh, "Counting subgraphs via homomorphisms," in *International Colloquium on Automata, Languages, and Programming*. Springer, 2009, pp. 71–82.

[54] E. Carlini, M. V. Catalisano, and L. Chiantini, "Progress on the symmetric Strassen conjecture," *Journal of Pure and Applied Algebra*, vol. 219, no. 8, pp. 3149–3157, 2015.

[55] Y. Shitov, "A counterexample to Strassen's direct sum conjecture," *arXiv preprint arXiv:1712.08660*, 2017.

[56] R. Mathon, "A note on the graph isomorphism counting problem," *Information Processing Letters*, vol. 8, no. 3, pp. 131–136, 1979.

[57] L. Babai, "Graph isomorphism in quasipolynomial time," in *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. ACM, 2016, pp. 684–697.