Electronic Colloquium on Computational Complexity, Report No. 56 (2014)

Factors of Sparse Polynomials are Sparse

Zeev Dvir* Rafael Oliveira†

Abstract

We show that if $f(x_1, \ldots, x_n)$ is a polynomial with s monomials and $g(x_1, \ldots, x_n)$ divides f then g has at most $\max(s^{O(\log s \log\log s)}, d^{O(\log d)})$ monomials, where d is a bound on the individual degrees of f. This answers a question of von zur Gathen and Kaltofen (JCSS 1985) who asked whether a quasi-polynomial bound holds in this case. Two immediate applications are a randomized quasi-polynomial time factoring algorithm for sparse polynomials and a deterministic quasi-polynomial time algorithm for sparse divisibility.

1 Introduction

Let $f \in \mathbb{F}[x_1, \ldots, x_n]$ be a multivariate polynomial over a field $\mathbb{F}$. The *sparsity* of f, denoted s_f, is the number of monomials (with non zero coefficients) appearing in f. For example, the sparsity of the polynomial $x_1^3 + x_2x_1^5 + 5x_1 + 6$ is four. If we limit the degree of each variable x_i to be at most d, then the maximum number of monomials in such an f is at most $(d+1)^n$. Polynomials that contain much less monomials are considered 'sparse' polynomials. The sparsity of f is a natural complexity measure one can use for polynomials and was studied in various contexts [GK85, KS01, Zip79, SW05, SSS13]. From a practical perspective, one can store a polynomial with s monomials on a computer using $O(s)$ memory locations (assuming a single field element fits in a single location) and then one would like to perform basic operations on polynomials (evaluation, multiplication, composition, etc..) efficiently in the size of this representation.

One of the most basic (and useful) operations on polynomials is factorization. Suppose $f(x) = g(x) \cdot h(x)$ is a given factorization of a sparse polynomial f. If we are to factor $f(x)$ and store the factors $g(x), h(x)$ in memory (also as a list of monomials) then we must first have an upper bound on the sparsity of these factors. This problem was raised in the seminal paper of von zur Gathen and Kaltofen [GK85] who studied the problem of efficient polynomial factorization in the sparse representation and gave a factoring algorithm whose running time is polynomial in the size of the input and in the size of the output (for which they had no a priori bound). Unlike integer factorization, which is believed to be intractable, the problem of factoring a multivariate polynomial into its irreducible factors is solvable in polynomial time

*Department of Computer Science and Department of Mathematics, Princeton University. Email: zeev.dvir@gmail.com. Research supported by NSF grant CCF-1217416 and by the Sloan fellowship.

†Department of Computer Science, Princeton University. Research supported by NSF grant CCF-1217416 and by the Sloan fellowship. Email: rmo@cs.princeton.edu.

ISSN 1433-8092

[LLL82, Kal85, Kal03]. In fact, something stronger is known: if $f(x)$ is given to us as an arithmetic circuit of size s (see below for more details on arithmetic circuits) then one can efficiently output a list of small circuits (of size $\mathsf{poly}(s)$) for each of the irreducible factors of f [Kal89]. This fundamental algorithmic tool has had many applications in computer science including in coding theory [Sud97, GS06], derandomization [KI04] and cryptography [CR88].

The following simple example shows that there can be a super-polynomial blow-up in the sparsity of a factor of f:

Example 1.1. *Let*

$$f(x) = \prod_{i\in[n]} (x_i^d - 1),$$

$$g(x) = \prod_{i\in[n]} (1 + x_i + \ldots, + x_i^{d-1}),$$

$$h(x) = \prod_{i\in[n]} (x_i - 1).$$

Then $s_f = 2^n$ but $s_g = d^n$. If one takes $n = d$ we get that s_g is not polynomial in s_f.

Abandoning polynomial blowup, we can ask weather one can prove a weaker, but still non trivial bound on the blow up in sparsity. Specifically, it was asked in [GK85, Section 5] if one can show a sub exponential bound of the form $s_g \leq \exp(s_f^\epsilon)$ for some $\epsilon > 0$ or even a quasi-polynomial blow up $s_g \leq \exp(\mathsf{poly}(\log(s_f)))$. In this work we answer this question in the affirmative and show a quasi polynomial bound on the sparsity of g in terms of the sparsity of f.

Theorem 1. *Let $f \in \mathbb{F}[x_1, \ldots, x_n]$ with $\mathbb{F}$ being any field. Let d be an upper bound on the individual degree of f. If $f(x) = g(x) \cdot h(x)$ then*

$$s_g \leq \max(s_f^{O(\log s_f \log\log s_f)}, d^{O(\log d)}).$$

Notice that we can assume, without loss of generality, that the field $\mathbb{F}$ is algebraically closed, since $f(\mathbf{x}) = g(\mathbf{x}) \cdot h(\mathbf{x})$ in $\mathbb{F}[\mathbf{x}]$ implies that $f(\mathbf{x}) = g(\mathbf{x}) \cdot h(\mathbf{x})$ also holds in the ring $\overline{\mathbb{F}}[\mathbf{x}]$.

One could hope to improve this bound to a bound of the form $s_g \leq s_f^{O(\log d)}$ which would match the parameters of Example 1.1. While our proof cannot currently give this bound, there is a specific place in the proof where an improvement could lead to this optimal bound. We elaborate more on this in the proof overview at the end of this introduction.

One immediate application of this result is a quasi-polynomial time randomized algorithm for factoring sparse polynomials. This follows from the results of [GK85] and the fact that the output size is quasi-polynomial. Another simple application, observed by Erich Kaltofen, is to the sparse divisibility problem in which one is given two sparse polynomials f, g so that g divides f and is asked to output a polynomial h so that $h \cdot g = f$. Using our sparsity bound, combined with deterministic sparse interpolation due to [KS01], we are able to give a *deterministic* algorithm for this problem. We sketch the argument in Section 7. It is an

interesting open problem whether our results can be used to give a deterministic factoring algorithm for sparse polynomials.

Before giving a high level overview of the proof of Theorem 1, we briefly discuss related work on bounded depth arithmetic circuits.

1.1 Related work: Bounded depth circuits

An arithmetic circuit is a computation DAG in which every gate performs either addition or multiplication (an addition gate is usually allowed to perform an arbitrary linear combination). Each input gate is a variable in some set $\{x_1, \ldots, x_n\}$ of formal variables and the output is some polynomial in the ring $\mathbb{F}[x_1, \ldots, x_n]$. For a given multivariate polynomial $f(x_1, \ldots, x_n)$, the size of the smallest arithmetic circuit computing f is an important complexity measure and has been studied extensively (see, e.g., [SY10] for a recent survey). The famous P vs NP problem has a seemingly more tractable algebraic variant in the form of the VP vs VNP problem [Val79a, Val79b, SY10] which requires proving superpolynomial lower bounds on the circuit size of some specific 'complete' polynomial in VNP.

The *depth* of a circuit is the length of the longest path from an input to an output. Since we can assume w.l.o.g that addition and multiplication gates are alternating, a depth k circuit is a layered computation graph in which each layer has either addition or multiplication gates. In recent years, bounded depth arithmetic circuits have received a lot of attention [GK98, GR00, SW01, RY09, AV08, FLMS13, GKKS12, KLSS14, KS13] (for a more complete list see [SY10]). Much of the interest in recent years is due to a work of Agrawal and Vinay [AV08] showing that sufficiently strong lower bounds for depth 4 circuits would imply super-polynomial lower bounds for general circuits (which could potentially resolve the VP vs VNP problem). Even more recently [GKKS13] showed that sufficiently strong lower bounds for (non homogeneous) depth 3 circuits (which are simply sums of products of linear terms) would result in lower bounds for general circuits. Sparse polynomials can be thought of as polynomials with small depth 2 circuits (that is a sum of products of variables). As such, they present the 'first line' when studying small depth circuits. In addition, depth 4 circuits (which are as hard as general circuits) are usually described as sums of products of sparse polynomials. In this light, we see that understanding sparse polynomials better could help us gain insights into the depth 4 model.

The problem of factoring in bounded depth was studied previously in [DSY09] who showed that if f has a small depth k circuits, then its factors of the form $x_n - \phi(x_1, \ldots, x_{n-1})$ have small depth $k+3$ circuits when the degree of x_n in f is sufficiently small. This result was used to extend the hardness-randomness tradeoffs of [KI04] to the bounded depth model. Our work shows that factors of polynomials with small depth 2 circuits also have small depth 2 circuits (with a quasi-polynomial blow-up). One can hope to extend this to circuits of any (constant) depth.

Conjecture 1.2. *If f has a depth k circuit of size s then any factor of f has a depth k circuit of size $s^{O(\log^c(s))}$, with c possibly depending on the depth k.*

1.2 Proof overview

The starting point is to view the problem in the reverse direction. Suppose s_g is large. How sparse can f be, when $f = g \cdot h$? The operation of multiplying a given polynomial by h can be described as follows: Let $E_g \subset \mathbb{N}_0^n$ be the set of 'exponents' of g: that is, the set of all $(j_1, \ldots, j_n)$ so that the monomial $\prod x_i^{j_i}$ appears in g. Let E_h be the set of exponents of h. Now, the exponents in f can only come from sums of points in E_g and in E_h. The problem is that some of these sums can occur many times (i.e., come from many different pairs) and so the resulting monomials might cancel each other. If we could, however, show that there are many points in E_f that are obtained in *only one way* as a sum in $E_g + E_h$, then we would get a lower bound on s_f. The way to find such points is to consider the Newton polytope P_g obtained as the convex hull of the points in E_g (over the real numbers) and to observe that each vertex of this polytope must give one such unique sum (this is shown in [Sch00]). In Example 1.1 for instance, the set E_g is the sub lattice $\{0, 1, \ldots, d-1\}^n$ and its polytope is the 'box' $P_g = [0, d-1]^n$. This box has exactly 2^n vertices, and it is easy to check that these (when multiplied by h) give the 2^n monomials of f.

The first idea in our proof is to 'hash' the polynomials f, g and h down to a space of lower dimension (using a substitution of variables) in a way that will impose some structure on the polytope P_g. More precisely, replacing each variable x_i with a product $y_1^{\alpha_{i1}} \cdot \ldots \cdot y_\ell^{\alpha_{i\ell}}$ gives us new polynomials $f', g', h' \in \mathbb{F}[y_1, \ldots, y_\ell]$ in a way that the new sets of exponents are mapped linearly from $\mathbb{N}_0^n$ to $\mathbb{N}_0^\ell$ using the $n \times \ell$ matrix A with entries α_{ij}. If we treat A as a linear map from $\mathbb{R}^n$ to $\mathbb{R}^\ell$ then we have that $E_{f'} \subset A(E_f)$ and similarly for g and h. The reason we do not have set equality is that some elements in E_f might map under A to the same element in $\mathbb{N}_0^\ell$ and so the resulting monomials might cancel each other. Using a substitution trick (developed in Section 4) we can bypass this difficulty and obtain equality $E_{f'} = A(E_f)$ and similarly for g and h. We will elaborate on this substitution trick more at the end of this proof overview. The hope is to carefully pick a matrix A so that the resulting polytope $P_{g'}$ of the 'hashed' polynomial g, has many vertices. Notice that we still have $f' = g' \cdot h'$ and that $s_{f'} \leq s_f$. This means that the number of vertices in $P_{g'}$ is a lower bound on the sparsity of f.

Unfortunately, we have to abandon this direct approach since applying a linear map on a polytope cannot increase the number of vertices it has. We overcome this problem by introducing modular reductions in the exponents of the polynomials. This solves the linearity problem (since the reduction mod p can create new vertices) but introduces another problem as we now explain. For a polynomial f, let $f^{(\mathsf{mod}\ p)}$ denote the polynomial obtained from f by reducing all the exponents of f modulo p (and then summing monomials with the same exponent). In algebraic terminology, we are reducing the polynomial $f(x_1, \ldots, x_n)$ modulo the ideal generated by the set $\{x_i^p - 1 \mid i \in [n]\}$. Let $\tilde{f} = (f')^{(\mathsf{mod}\ p)}$ and similarly for $\tilde{g}, \tilde{h}$, where f', g', h' are obtained using some substitution given by an integer matrix A as before (so they are polynomials in $y_1, \ldots, y_\ell$). Notice that now we are applying a linear map on the exponents but then taking it mod p. The nice thing about this operation is that, if s_g is sufficiently larger than p^ℓ, we can pick A so that the set of exponents in $\tilde{g}$ is exactly $\{0, 1, \ldots, p-1\}^\ell$ (this will happen for a random matrix A with high probability). This looks useful since now we know that $P_{\tilde{g}}$ has at least 2^ℓ vertices. The problem is that the equality $\tilde{f} = \tilde{g} \cdot \tilde{h}$ is false, since the

multiplication of $\tilde{g}$ and $\tilde{h}$ might incur another reduction mod p. The correct equality would be $\tilde{f} = (\tilde{g} \cdot \tilde{h})^{(\mathsf{mod}\ p)}$. The crucial observation is that, since $\tilde{g}$ and $\tilde{h}$ are already reduced mod p, the final reduction only reduces from $\{0, 1, \ldots, 2p-2\}$ to $\{0, 1, \ldots, p-1\}$. This means that every monomial in $\tilde{f}$ can come from at most 2^ℓ monomials in $\tilde{g} \cdot \tilde{h}$. Here as well, one needs to be careful since some of the resulting monomials might occur several times and could cancel each other (this is solved using the same substitution trick mentioned above).

At first glance, this loss of 2^ℓ seems to completely destroy our hopes since this was exactly the number of vertices we had in $P_{\tilde{g}}$. The last ingredient in the proof is that, when working mod p, we can actually, construct a structure in $\{0, 1, \ldots, p-1\}^\ell$ that has much more vertices than 2^ℓ and so that we can still 'force' $E_{\tilde{g}}$ to be this structure. This structure turns out to be a hyperplane mod p. That is, we can (using a slight variant of the above substitution) force the set $E_{\tilde{g}}$ into any hyperplane mod p we wish. More precisely, let $H \subset \mathbb{F}_p^\ell$ be a hyperplane (that is the set of solution to a single linear equation over the finite field $\mathbb{F}_p$). Suppose we can 'place' the exponents of $\tilde{g}$ in the set H. If we could show that the number of vertices of the convex hull of the points of H (which is now viewed also as a subset of $\mathbb{R}^\ell$) is much larger than 2^ℓ then we could use the above argument (loosing a factor of 2^ℓ but remaining with 'something').

To see why we might hope for such a weird phenomena to happen, consider the case $\ell = 2$. The simplest hyperplane in $\mathbb{F}_p^2$ is the line $x = y$ which, when viewed in $\mathbb{N}_0^2$ gives the convex hull $\{(x, x) \mid 0 \leq x \leq p-1\}$ which has only 2 vertices. However, if we take a different hyperplane, say $y = 2x$ in $\mathbb{F}_p^2$, we will get a set in $\mathbb{N}_0^2$ containing two 'lines' with $(p-1)/2$ points in each. The convex hull will now be a polytope with 4 vertices. More generally, we show that, under certain conditions on p, there exists a hyperplane in $\mathbb{F}_p^\ell$ so that its convex hull has at least $(2.01)^\ell$ vertices (for the exact parameters see Theorem 5.10), which allows us to handle the loss of 2^ℓ in the modular reduction. The hyperplane we construct uses coefficients that satisfy certain divisibility conditions (as integers). We then count the vertices of the polytope by considering restrictions of the polytope to certain subsets of the vertices (fixing the other variables to either 0 or $p-1$). We show that many of these restrictions (which are also polytopes) contain 'interior' vertices, which are vertices that have all of their coordinates in the set $\{1, 2, \ldots, p-2\}$. These interior vertices are also vertices of the 'big' polytope, but they can only occur in *one* of the restrictions (since vertices coming from other restrictions will have at least one coordinate set to 0 or $p-1$).

Putting things together, we choose $p > d$ so that s_g is roughly p^ℓ. We then hash the polynomials f, g, h to $\tilde{f}, \tilde{g}, \tilde{h}$ using a variable substitution and reduction modulo p so that $s_{\tilde{f}} = s_f$, $E_{\tilde{g}}$ is the hyperplane H we constructed above and so that $\tilde{f} = (\tilde{g} \cdot \tilde{h})^{(\mathsf{mod}\ p)}$. Then, by our lower bound on the number of vertices in $P_{\tilde{g}}$, we get that $s_f = s_{\tilde{f}} > (2.01)^\ell / 2^\ell \geq (1.005)^\ell$ and so s_g is at most $s_f^{O(\log(p))}$. Certain conditions on p (which are needed to construct the special hyperplane H) prevent us from choosing $p \leq O(d)$ and obtaining a dependency of $s_g \leq s_f^{O(\log(d))}$, which would be optimal by Example 1.1. It is not unlikely that a more clever hyperplane construction could result in an optimal bound using this approach.

The 'substitution trick': In the overview above we hinted at a problem that can occur when reducing the exponents of a polynomial modulo a prime p. To be more precise, consider the

polynomial $f(x,y) = x^2y^2 - x^5y^2 + 1$. Using our notations we have $E_f = \{(2,2),(5,2),(0,0)\}$. Reducing the exponents modulo $p = 3$ we get that the two monomials x^2y^2 and x^5y^2 cancel each other and so $E_{f^{\mathsf{mod}\ p}} \neq E_f^{(\mathsf{mod}\ p)}$. To solve this we can introduce new 'randomness' into f as follows: For each $(a,b) \in \mathbb{F}^2$ consider the polynomial

$$f_{a,b}(x,y) = f(ax,by) = a^2b^2x^2y^2 + a^5b^2x^5y^2 + 1.$$

Notice that $E_{f_{a,b}} = E_f$. Now, when we reduce $f_{a,b}$ modulo $p = 3$ we get the polynomial

$$f_{a,b}(x,y)^{\mathsf{mod}\ 3} = (a^2b^2 - a^5b^2)x^2y^2 + 1.$$

The main observation is that the coefficient of x^2y^2 in $f_{a,b}$ is a non zero polynomial in a,b and so will not vanish for almost all a,b. Working over an infinite field allows us to choose a,b so that there will be no cancellations at all (even if there are many monomials that we worry about) in the reduction modulo p. The same idea is used to control other types of cancellations that arise in the 'hashing' part of our argument.

1.3 Organization

We begin in Section 2 with some general preliminaries. In Section 3 we discuss some basic (known) properties of the Newton polytope that will be used in the proof. The actual proof is divided into three sections. Section 4 contains the 'hashing' part of the proof and shows how to hash the polynomials f,g,h to new polynomials having a nice structure on their sets of exponents. Section 5 contains the construction of the special hyperplane in $\mathbb{F}_p^\ell$ whose associated polytope in $\mathbb{R}^\ell$ contains many vertices. The proof of Theorem 1, which results from combining these two parts, is given in Section 6. In Section 7 we outline the deterministic sparse divisibility algorithm that follows from our main theorem.

2 Preliminaries

In this section, we establish the notation that will be used throughout the paper and some technical background that we will need to develop the proof of our main theorem.

2.1 Notations

From this point on, we will use boldface for vectors, and regular font for scalars. Thus, we will denote the vector $(x_1,\ldots,x_n)$ by $\mathbf{x}$, and for a set $T \subseteq \{1,\ldots,n\}$ we denote by $\mathbf{x}_T$ the vector $(x_i)_{i\in T}$. If we want to multiply the vector $\mathbf{x}$ by a scalar z we will denote this product by $z\mathbf{x}$. The dot product between two vectors is defined as $\mathbf{x}\cdot\mathbf{y} = \sum_{i=1}^{n} x_iy_i$, and, more generally, we define $\mathbf{x}_T \cdot \mathbf{y}_T = \sum_{i\in T} x_iy_i$.

Let $\mathbb{N}_0$ be the set of natural numbers including zero, that is, $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$. In addition, from now on we will take $\log x = \log_2 x$, $\ln x = \log_e x$, $\exp(x) = e^x$, where e is Euler's number, and we will define the expectation and the variance of a random variable X to be $\mathbb{E}[X]$ and $\mathbb{V}[X]$, respectively.

From now on, we will always write an equality that happens in $\mathbb{F}_p$ with a "$\stackrel{p}{=}$" sign and an equation over $\mathbb{R}$ with an "=" sign. For example, the equation $ax+b \equiv y \pmod p$, which we will write as $ax+b \stackrel{p}{=} y$, will be the equation of a line in $\mathbb{F}_p^2$ (on the variables x, y) and the equation $ax + b = y$ will be the equation of a line over $\mathbb{R}^2$. Moreover, whenever we use the expression $(a \pmod p)$, we are referring to the only integer r such that $a = pq + r$ and $0 \leq r < p$.

If $\mathbf{e} \in \mathbb{N}_0^n$ is a vector of natural numbers and $\mathbf{x} = (x_1, \ldots, x_n)$ is a vector of formal variables, we define $\mathbf{x}^{\mathbf{e}} = \prod_{i=1}^{n} x_i^{e_i}$. That is, $\mathbf{x}^{\mathbf{e}}$ is the monomial corresponding to the product of the variables $\prod_{i=1}^{n} x_i^{e_i}$, where each variable is raised to the proper power. In addition, we define $(\mathbf{e} \pmod p) \in \mathbb{F}_p^n$ to be the vector defined by $((e_1 \pmod p), (e_2 \pmod p), \ldots, (e_n \pmod p))$. That is, we restrict each coordinate of $\mathbf{e}$ modulo p. Moreover, if $S \subset \mathbb{N}_0^n$, we define

$$S^{(\mathsf{mod}\ p)} = \{(\mathbf{e} \pmod p) \mid \mathbf{e} \in S\}.$$

Notice that $|S^{(\mathsf{mod}\ p)}| \leq |S|$, and this inequality can be strict, since some vectors of S can map to the same vector in $S^{(\mathsf{mod}\ p)}$.

2.2 Number theoretic estimates

We will need the following number theoretic estimate on the product of primes.

Lemma 2.1. *Let $t \in \mathbb{N}_0$ be such that $t \geq 50$. Let*

$$Q_t = 3 \cdot \prod_{i=1}^{t} p_i,$$

where $2 < 3 < p_1 < p_2 < \ldots < p_t$ are the first $t + 2$ prime numbers. Then,

$$t^{\frac{4t}{5}} < Q_t < t^{\frac{11t}{5}}. \tag{1}$$

The proof, which uses known estimates from the literature, is given for completeness in appendix A.

2.3 Hashing

We will need the following lemma saying that we can always hash a set of $\mathbb{F}_p^n$ of size roughly $p^{4\ell}$ to all the points in $\mathbb{F}_p^\ell$.

Lemma 2.2. *Let $S \subseteq \mathbb{F}_p^n$ be a set of points such that $|S| = s = p^{c\ell}$, for some real constant $c > 3$. Then, there exists a surjective linear map $L : \mathbb{F}_p^n \to \mathbb{F}_p^\ell$ such that $L(S) = \mathbb{F}_p^\ell$.*

The proof of this lemma, which uses a standard pairwise independence argument, is given in Appendix B for completeness. The following is an immediate corollary.

Corollary 2.3. *Let $S \subseteq \mathbb{F}_p^n$ be a set of points such that $|S| = s = p^{c\ell}$, for some constant $c > 3$. Moreover, let $\mathcal{H} \subset \mathbb{F}_p^l$ be a hyperplane of $\mathbb{F}_p^\ell$ passing through the origin. Then, there exists a homogeneous linear map $L : \mathbb{F}_p^n \to \mathcal{H}$ such that $L(S) = \mathcal{H}$.*

Proof. By Lemma 2.2, we have that there exists a surjective linear map $P_1 : \mathbb{F}_p^n \to \mathbb{F}_p^\ell$ such that $P_1(S) = \mathbb{F}_p^\ell$. Thus, by projecting P_1 onto the first $\ell - 1$ coordinates, we obtain the surjective linear map $P : \mathbb{F}_p^n \to \mathbb{F}_p^{\ell-1}$, where $P(S) = \mathbb{F}_p^{\ell-1}$. Let $\varphi : \mathbb{F}_p^{\ell-1} \to \mathcal{H}$ be the bijective linear map from $\mathbb{F}_p^{\ell-1}$ to $\mathcal{H}$. Then, the map $L : \mathbb{F}_p^n \to \mathcal{H}$ defined by $L = \varphi \circ P$ is a surjective linear map and $L(S) = \mathcal{H}$. □

3 Polynomials and the Newton Polytopes

In this section we discuss properties of the Newton polytope that will play an important role in our proof. The main result from this section we will need later on is Corollary 3.17, which says that, if $f = g \cdot h$ are polynomials then the sparsity of f is at least $\max\{|V(P_g)|, |V(P_h)|\}$, where P_g and P_h are the Newton polytopes of g and h respectively. Readers familiar with this basic result can skip to the next section.

We begin with some basic properties of polytopes and then move on to discuss the Newton polytope. All of the results below are taken from appendix K of [Sch00] and from the book [Zie95].

3.1 General polytopes

We start by defining the convex span of points in $\mathbb{R}^n$, polytopes in $\mathbb{R}^n$ and the convex hull of a set $S \subset \mathbb{R}^n$.

Definition 3.1 (Convex Span and Polytope)**.** *The* convex span *of points $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$, denoted by $CS(\mathbf{v}_1, \ldots, \mathbf{v}_k)$, is the set defined by*

$$CS(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \left\{ \sum_{i=1}^{k} \lambda_i \mathbf{v}_i \mid \lambda_i \in \mathbb{R}^+ \text{ and } \sum_{i=1}^{k} \lambda_i = 1 \right\}.$$

That is, $CS(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ is the set of all convex combinations of the points $\mathbf{v}_i$, $1 \leq i \leq k$. A set $P \subset \mathbb{R}^n$ is a Polytope *if there exists a finite set of points $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ such that*

$$P = CS(\mathbf{v}_1, \ldots, \mathbf{v}_k).$$

With this definition, we can define convex sets, and the convex hull of a set.

Definition 3.2. *A* convex set *in $\mathbb{R}^n$ is a set $C \subset \mathbb{R}^n$ such that $CS(\mathbf{x}, \mathbf{y}) \subset C$ for all $\mathbf{x}, \mathbf{y} \in C$.*

Definition 3.3. *The* convex hull *of a set $S \subset \mathbb{R}^n$, denoted by $\mathsf{conv}(S)$, is the intersection of all convex subsets of $\mathbb{R}^n$ containing S.*

From these definitions, one can easily obtain the following corollary:

Corollary 3.4. *For every finite set S, we have that $\mathsf{conv}(S) = CS(S)$.*

Definition 3.5 (Supporting Hyperplane)**.** *A* supporting hyperplane *H of a convex set C in $\mathbb{R}^n$ is a hyperplane defined by $\mathbf{h} \cdot \mathbf{x} = a$, where $\mathbf{h} \in \mathbb{R}^n \setminus \{0\}$ and $a \in \mathbb{R}$ such that H intersects the closure of C in $\mathbb{R}^n$ and $\mathbf{h} \cdot \mathbf{x} \leq a$ for all $\mathbf{x} \in C$.*

With this definition, we are now able to define a face of a polytope.

Definition 3.6 (Face of a Polytope)**.** *Let P be a polytope. A* face *of P is the intersection of P with a supporting hyperplane. That is, if $\mathbf{h} \cdot \mathbf{x} = a$ is a supporting hyperplane of P, we have that the set*

$$F = P \cap \{\mathbf{x} \mid \mathbf{h} \cdot \mathbf{x} = a\}$$

is a face of P.

Observation 3.7. *Notice that by definition 3.6, a face F of polytope P is a convex set such that if $\mathbf{a}, \mathbf{b} \in P$ and if there exists $\lambda \in (0, 1)$ such that $\lambda\mathbf{a} + (1 - \lambda)\mathbf{b} \in F$ then $\mathbf{a}, \mathbf{b} \in F$.*

Definition 3.8. *Faces of a polytope P of dimension 0 are called vertices of P. For a polytope P, let $V(P)$ be the set of vertices of P.*

Observation 3.9. *Notice that based on Observation 3.7 and definition 3.8, we have that a vertex $\mathbf{a} \in V(P)$ is a point that cannot be written as $\lambda\mathbf{u} + (1 - \lambda)\mathbf{v}$, for any $\mathbf{u}, \mathbf{v} \in P \setminus \{\mathbf{a}\}$ and $\lambda \in [0, 1]$.*

The following lemma, from [Zie95, Propositions 2.2 and 2.3], captures some essential properties of vertices and faces of a polytope.

Lemma 3.10 (Properties of Vertices and Faces)**.** *Let $P = CS(\mathbf{v}_1, \ldots, \mathbf{v}_k) \subset \mathbb{R}^n$ be a polytope and let F be a face of P.*

(i) F is a polytope, with $V(F) = F \cap V(P)$.

(ii) Every intersection of faces of P is a face of P.

(iii) The faces of F are exactly the faces of P that are contained in F.

(iv) $P = CS(V(P))$.

(v) $V(P) \subseteq \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$.

Now that we defined polytopes and some of its properties, we will define an important operation on polytopes which is intrinsically related to polynomials, as we will see in section 3.

Definition 3.11 (Minkowski sum)**.** *Given two polytopes P_1 and P_2 in $\mathbb{R}^n$, we define their* Minkowski sum $P_1 + P_2$ *to be the set of points given by:*

$$P_1 + P_2 = \{\mathbf{v}_1 + \mathbf{v}_2 \mid \mathbf{v}_1 \in P_1 \text{ and } \mathbf{v}_2 \in P_2\}.$$

The following theorem tells us about some important properties of the Minkowski sum of two polytopes. For a proof of the theorem, see theorem 3 in [Sch00, Appendix k].

Theorem 3.12 (Vertices in Minkowski Sum, [Sch00])**.** *Let P_1, P_2 be polytopes in $\mathbb{R}^n$ and $P_1 + P_2$ be their Minkowski sum. Then*

(i) $P_1 + P_2$ is a polytope.

(ii) Every vertex $\mathbf{v} \in V(P_1 + P_2)$ can be expressed as a sum $\mathbf{x}_1 + \mathbf{x}_2$, $\mathbf{x}_i \in P_i$ $(i = 1, 2)$ in only one way. Further, such $\mathbf{x}_i \in V(P_i)$, that is, the $\mathbf{x}_i$ are vertices of P_i.

(iii) For every vertex $\mathbf{v}_1 \in V(P_1)$, there exists a vertex $\mathbf{v}_2 \in V(P_2)$ such that $\mathbf{v}_1 + \mathbf{v}_2 \in V(P_1 + P_2)$.

As a corollary of Theorem 3.12, we have that given two polytopes P_1 and P_2, the number of vertices of the polytope defined by the Minkowski sum $P_1 + P_2$ is always no less than the maximum number of vertices of the polytopes P_1 and P_2. More precisely, we have:

Corollary 3.13. *Let P_1, P_2 be polytopes in $\mathbb{R}^n$ and $P_1 + P_2$ be their Minkowski sum. Then,*

$$|V(P_1 + P_2)| \geq \max\{|V(P_1)|, |V(P_2)|\}.$$

Proof. Let $\Phi : V(P_1 + P_2) \to V(P_1)$ be a map defined as follows: if $\mathbf{v} \in V(P_1 + P_2)$ is such that $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_i \in V(P_i)$ $(i = 1, 2)$, then $\Phi(\mathbf{v}) = \mathbf{v}_1$. By item (ii) of Theorem 3.12, the map Φ is well defined. By item (iii) of Theorem 3.12, Φ is surjective, and hence we have $|V(P_1 + P_2)| \geq |V(P_1)|$. By symmetry, we have that $|V(P_1 + P_2)| \geq |V(P_2)|$. □

3.2 The Newton polytope

We will begin with the following definition, which will be used throughout the paper.

Definition 3.14. *Let $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ be a polynomial such that*

$$f(\mathbf{x}) = \sum_{i=1}^{s} a_i \mathbf{x}^{\mathbf{e}_i},$$

where $a_i \neq 0$, for $1 \leq i \leq s$. For each monomial $m_i = \mathbf{x}^{\mathbf{e}_i}$ of f, define the exponent vector of m_i as $\mathbf{e}_i = (e_{i1}, e_{i2}, \ldots, e_{in}) \in \mathbb{N}^n$, for $1 \leq i \leq s$. Then, the set of monomials of f and the set of exponents of f, denoted by M_f and E_f, respectively, are defined as follows:

$$M_f = \{m_i \mid 1 \leq i \leq s\} \text{ and } E_f = \{\mathbf{e}_i \mid 1 \leq i \leq s\}.$$

Notice from the definition above that the set of exponents E_f is a subset of $\mathbb{N}^n$. From now on, we will refer to the sparsity of a polynomial f as s_f or as $\|f\|_0$. Hence, in the definition above we have that

$$\|f\|_0 = s_f = s = |M_f| = |E_f|.$$

Notation $\|f\|_0$ will be used in sections 4 and 6, whereas notation s_f is very suitable to state our theorems and lemmas throughout the paper.

Definition 3.15 (Newton Polytope)**.** *The* Newton Polytope *associated to the polynomial $f(\mathbf{x})$, which we denote by P_f, is defined by:*

$$P_f = CS(E_f).$$

That is, P_f is the convex span of the exponent vectors $\mathbf{e} \in E_f$.

Given polynomials $f(\mathbf{x}), g(\mathbf{x}), h(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ such that $f(\mathbf{x}) = g(\mathbf{x})h(\mathbf{x})$, it is a classical fact that $P_f = P_g + P_h$. For completeness, a proof of this fact is presented below:

Proposition 3.16 (Minkowski sum and polynomial multiplication)**.** *Let $f(\mathbf{x}), g(\mathbf{x}), h(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ be polynomials such that $f(\mathbf{x}) = g(\mathbf{x})h(\mathbf{x})$. Then*

$$P_f = P_g + P_h.$$

Proof. Since every monomial $\mathbf{x}^{\mathbf{e}} \in M_f$ comes from a multiplication from a monomial $\mathbf{x}^{\mathbf{e}_g} \in M_g$ by a monomial $\mathbf{x}^{\mathbf{e}_h} \in M_h$, we have that $E_f \subseteq E_g + E_h$, which implies

$$P_f = CS(E_f) \subseteq CS(E_g + E_h) \subseteq P_g + P_h.$$

Now, for the inclusion $P_f \supseteq P_g + P_h$, notice that by Theorem 3.12 we have that for each $\mathbf{e} \in V(P_g + P_h)$, there exists only one $\mathbf{e}_g \in P_g$ and one $\mathbf{e}_h \in P_h$ such that $\mathbf{e} = \mathbf{e}_g + \mathbf{e}_h$. Moreover, since $\mathbf{e}_g \in V(P_g) \subseteq E_g, \mathbf{e}_h \in V(P_h) \subseteq E_h$, where we know that $V(P_t) \subseteq E_t$, $(t = g, h)$ by Lemma 3.10, we have that the coefficient $c_{\mathbf{e}}$ of the monomial $\mathbf{x}^{\mathbf{e}} \in M_f$ will be nonzero, since $c_{\mathbf{e}} = c_{\mathbf{e}_g} c_{\mathbf{e}_h}$, where $c_{\mathbf{e}_g} \in C_g$ and $c_{\mathbf{e}_h} \in C_h$ are the coefficients of $\mathbf{x}^{\mathbf{e}_g}$ and $\mathbf{x}^{\mathbf{e}_h}$, respectively. Therefore, we have that $\mathbf{e} \in P_f$, which implies that $V(P_g + P_h) \subset P_f$, which in turn implies that

$$P_g + P_h = CS(V(P_g + P_h)) \subseteq P_f.$$

□

With this fact in mind and by Corollary 3.13 we obtain the following corollary:

Corollary 3.17. *Let $f(\mathbf{x}), g(\mathbf{x}), h(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ be such that $f(\mathbf{x}) = g(\mathbf{x})h(\mathbf{x})$. Then,*

$$s_f \geq |V(P_f)| \geq \max\{|V(P_g)|, |V(P_h)|\}.$$

Proof. Notice that $|V(P_f)| \geq \max\{|V(P_g)|, |V(P_h)|\}$ follows directly from Corollary 3.13 and the fact that $P_f = P_g + P_h$. To see that $s_f \geq |V(P_f)|$ holds, observe that $V(P_f) \subseteq E_f \Rightarrow |V(P_f)| \leq |E_f| = s_f$. □

4 Polynomial Substitutions

The goal of this section is to prove Lemma 4.13 which comprises the heart of the 'hashing' step of our proof. This lemma will be proved by combining two other lemmas stated after some preliminaries.

Throughout this section, we will assume that the field $\mathbb{F}$ is algebraically closed. We will need this assumption because we apply Corollary 4.4 repeatedly in the proof of the lemmas in this section.

4.1 Preliminaries

Definition 4.1 (Restrictions)**.** *Let $f(\mathbf{x}), g(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ be polynomials such that*

$$f(\mathbf{x}) = \sum_{i=1}^{s} a_{\mathbf{e}_i} \mathbf{x}^{\mathbf{e}_i}, \quad g(\mathbf{x}) = \sum_{j=1}^{s'} b_{\mathbf{e}_j} \mathbf{x}^{\mathbf{e}_j}$$

where $a_{\mathbf{e}_i}, b_{\mathbf{e}_j} \neq 0$, for $1 \leq i \leq s$, $1 \leq j \leq s'$. If $S \subseteq E_f$, we define $f|_S(\mathbf{x})$ as the following polynomial:

$$f|_S(\mathbf{x}) = \sum_{\mathbf{e} \in S} a_{\mathbf{e}} \mathbf{x}^{\mathbf{e}}.$$

If $U \subseteq E_f \times E_g$, we define $(f \cdot g)|_U(\mathbf{x})$ as the following polynomial:

$$(f \cdot g)|_U(\mathbf{x}) = \sum_{(\mathbf{e}, \mathbf{e}') \in U} a_{\mathbf{e}} b_{\mathbf{e}'} \mathbf{x}^{\mathbf{e}+\mathbf{e}'}.$$

Observation 4.2 (Properties of Restrictions)**.** *The following are some properties of the restriction operation:*

(i) If $\emptyset \neq S \subseteq E_f$, then $f|_S(\mathbf{x})$ is a nonzero polynomial.

(ii) If $S, T \subseteq E_f$ are such that $S \cap T = \emptyset$, then

$$f|_S(\mathbf{x}) + f|_T(\mathbf{x}) = f|_{S \cup T}(\mathbf{x})$$

(iii) Let $\mathbf{b} \in E_{f \cdot g}$, where $(f \cdot g)(\mathbf{x}) = \sum_{i=1}^{r} c_{\mathbf{e}_i} \mathbf{x}^{\mathbf{e}_i}$. If $U \subseteq E_f \times E_g$ is such that

$$U = \{(\mathbf{e}, \mathbf{e}') \in E_f \times E_g \mid \mathbf{e} + \mathbf{e}' = \mathbf{b}\}$$

then $(f \cdot g)|_U(\mathbf{x}) = c_{\mathbf{b}} \mathbf{x}^{\mathbf{b}}$.

Lemma 4.3 (Schwartz-Zippel-DeMillo-Lipton, [Sch80, Zip79, DL78])**.** *Let $f \in \mathbb{F}[\mathbf{x}]$ be a nonzero polynomial such that $\deg(f) \leq d$. If $|\mathbb{F}| > d$, there exists a point $\mathbf{a} \in \mathbb{F}^n$ such that $f(\mathbf{a}) \neq 0$.*

A very useful corollary of Lemma 4.3 is obtained below.

Corollary 4.4. *Let $\mathbb{F}$ be an algebraically closed field and $f_1, \ldots, f_r \in \mathbb{F}[\mathbf{x}]$ be nonzero polynomials such that $\deg f_k \leq d,\ 1 \leq k \leq r$. Then, there exists a point $\mathbf{a} \in \mathbb{F}^n$ such that $f_k(\mathbf{a}) \neq 0$ for all $1 \leq k \leq r$.*

Proof. Let $f(\mathbf{x}) = \prod_{k=1}^{r} f_k(\mathbf{x})$. This polynomial is clearly nonzero, since all f_k's are nonzero. Notice that $\deg(f_k) \leq d$ implies $\deg(f) = \sum_{k=1}^{r} \deg(f_k) \leq rd$. Hence, by Lemma 4.3 and the fact that $\mathbb{F}$ is algebraically closed (and thus has infinitely many elements), there exists a point $\mathbf{a} \in \mathbb{F}^n$ such that

$$0 \neq f(\mathbf{a}) = \prod_{k=1}^{r} f_k(\mathbf{a}) \Rightarrow f_k(\mathbf{a}) \neq 0,\ \forall 1 \leq k \leq r.$$

□

Definition 4.5. *Let $D \in \mathbb{N}_0^{m \times n}$ be a matrix of natural numbers and let $E \subset \mathbb{Z}^n$ be any set of integer vectors. We define the set $D(E)$ as*

$$D(E) = \{D\mathbf{e} \mid \mathbf{e} \in E\}.$$

Observation 4.6. *Since $D(E)$ is the image of E under a linear map, we have $|E| \geq |D(E)|$.*

Definition 4.7. *Let $\mathbf{x} = (x_1, \ldots, x_n)$ be a vector of formal variables, R be any commutative ring and let $f \in R[\mathbf{x}]$ be a polynomial given by*

$$f(\mathbf{x}) = \sum_{k=1}^{s} c_k \mathbf{x}^{\mathbf{e}_k}.$$

Let $p \in \mathbb{N}$ be a prime number and define $f^{(\mathsf{mod}\ p)} \in R[\mathbf{x}]$ as the following polynomial

$$f^{(\mathsf{mod}\ p)}(\mathbf{x}) = \sum_{k=1}^{s} c_k \mathbf{x}^{(\mathbf{e}_k \mod p)}.$$

Remark 4.8. *Notice that $E_{f^{(\mathsf{mod}\ p)}} \subseteq E_f^{(\mathsf{mod}\ p)}$ and sometimes the inclusion is strict, since there may be some cancellations in the transformation from $f \mapsto f^{(\mathsf{mod}\ p)}$. For instance, for the polynomial $f(x) = x^3 - 1$ and $p = 3$, we have that $f^{(\mathsf{mod}\ 3)}(\mathbf{x}) = 1 - 1 = 0$ and thus $E_{f^{(\mathsf{mod}\ 3)}} = \emptyset \subset \{0, 3\} = E_f^{(\mathsf{mod}\ 3)}$.*

Observation 4.9. *Let $I = (x_1^p - 1, x_2^p - 1, \ldots, x_n^p - 1)$ be the ideal generated by polynomials $x_i^p - 1$ and $R[\mathbf{x}]$ be the quotient ring $\mathbb{F}[\mathbf{x}]/I$. Notice that $f(\mathbf{x}) \equiv f^{(\mathsf{mod}\ p)}(\mathbf{x})$ and $g(\mathbf{x}) \equiv g^{(\mathsf{mod}\ p)}(\mathbf{x})$ in the ring $R[\mathbf{x}]$. Hence, we have that*

$$f(\mathbf{x}) \cdot g(\mathbf{x}) \equiv (f(\mathbf{x}) \cdot g(\mathbf{x}))^{(\mathsf{mod}\ p)} \equiv \left(f^{(\mathsf{mod}\ p)}(\mathbf{x}) \cdot g^{(\mathsf{mod}\ p)}(\mathbf{x})\right)^{(\mathsf{mod}\ p)}$$

in $R[\mathbf{x}]$*. Since* $(f(\mathbf{x}) \cdot g(\mathbf{x}))^{(\mathsf{mod}\ p)}$ *and* $\left(f^{(\mathsf{mod}\ p)}(\mathbf{x}) \cdot g^{(\mathsf{mod}\ p)}(\mathbf{x})\right)^{(\mathsf{mod}\ p)}$ *are both polynomials of individual degree at most* $p-1$ *in each variable that are equivalent in* $R[\mathbf{x}]$*, we must have that*

$$(f(\mathbf{x}) \cdot g(\mathbf{x}))^{(\mathsf{mod}\ p)} = \left(f^{(\mathsf{mod}\ p)}(\mathbf{x}) \cdot g^{(\mathsf{mod}\ p)}(\mathbf{x})\right)^{(\mathsf{mod}\ p)}$$

as polynomials in $\mathbb{F}[\mathbf{x}]$*.*

4.2 Lemmas on polynomial substitutions

Our first lemma shows that there exists a polynomial homomorphism whose action on the set of exponents of a given polynomial is the same as applying a given linear map (with integer coordinates) on those exponents.

Lemma 4.10. *Let* $\mathbf{x} = (x_1, \ldots, x_n)$ *and* $\mathbf{y} = (y_1, \ldots, y_m)$ *be formal variables,* $D \in \mathbb{N}_0^{m \times n}$ *be a matrix with column vectors* $\mathbf{d}_k$*, for* $1 \leq k \leq n$*. Let* $g \in \mathbb{F}[\mathbf{x}]$*. Then, there exists a homomorphism* $\mu^D : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{y}]$ *such that*

$$E_{\mu^D(g)} = D(E_g).$$

Proof. Let $\mathbf{v} = (v_1, \ldots, v_n)$ be formal variables and $\psi^D : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{v}][\mathbf{y}]$ be the ring homomorphism defined by

$$\psi^D(x_k) = v_k \cdot \prod_{i=1}^{m} y_i^{d_{ik}} = v_k \mathbf{y}^{\mathbf{d}_k},\ (1 \leq k \leq n).$$

In addition, for each $\mathbf{a} \in \mathbb{F}^n$, let $\psi_{\mathbf{a}}^D : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{y}]$ be the homomorphism defined by

$$\psi_{\mathbf{a}}^D(x_k) = a_k \mathbf{y}^{\mathbf{d}_k},\ (1 \leq k \leq n).$$

For each $\mathbf{b} \in D(E_g)$, let $S_{\mathbf{b}} = \{\mathbf{e} \in E_g \mid D\mathbf{e} = \mathbf{b}\}$. Because $\bigcup_{\mathbf{b} \in D(E_g)} S_{\mathbf{b}} = E_g$ and $S_{\mathbf{b}} \cap S_{\mathbf{b}'} = \emptyset$ for every $\mathbf{b}, \mathbf{b}' \in D(E_g)$ such that $\mathbf{b} \neq \mathbf{b}'$, we have that

$$g(\mathbf{x}) = \sum_{\mathbf{b} \in D(E_g)} \sum_{\mathbf{e} \in S_{\mathbf{b}}} a_{\mathbf{e}} \mathbf{x}^{\mathbf{e}}.$$

For $\mathbf{e} \in S_{\mathbf{b}}$ notice that

$$\psi^D(a_{\mathbf{e}} \mathbf{x}^{\mathbf{e}}) = a_{\mathbf{e}} \prod_{k=1}^{n} \psi^D(x_k)^{e_k} = a_{\mathbf{e}} \prod_{k=1}^{n} \left(v_k^{e_k} \mathbf{y}^{\mathbf{d}_k e_k}\right) = a_{\mathbf{e}} \mathbf{v}^{\mathbf{e}} \mathbf{y}^{D\mathbf{e}} = a_{\mathbf{e}} \mathbf{v}^{\mathbf{e}} \mathbf{y}^{\mathbf{b}}.$$

Therefore:

$$\psi^D(g) = \sum_{\mathbf{b} \in D(E_g)} \sum_{\mathbf{e} \in S_{\mathbf{b}}} \psi^D(a_{\mathbf{e}} \mathbf{x}^{\mathbf{e}}) = \sum_{\mathbf{b} \in D(E_g)} \sum_{\mathbf{e} \in S_{\mathbf{b}}} a_{\mathbf{e}} \mathbf{v}^{\mathbf{e}} \mathbf{y}^{\mathbf{b}}$$

$$= \sum_{\mathbf{b} \in D(E_g)} \left(\sum_{\mathbf{e} \in S_{\mathbf{b}}} a_{\mathbf{e}} \mathbf{v}^{\mathbf{e}} \right) \mathbf{y}^{\mathbf{b}} = \sum_{\mathbf{b} \in D(E_g)} g|_{S_{\mathbf{b}}}(\mathbf{v})\ \mathbf{y}^{\mathbf{b}}.$$

Thus, for each $\mathbf{b} \in D(E_g)$, the coefficient of monomial $\mathbf{y}^{\mathbf{b}}$ in $\psi^D(g)$ is the nonzero polynomial $g|_{S_{\mathbf{b}}}(\mathbf{v}) \in \mathbb{F}[\mathbf{v}]$. Let

$$\mathcal{G} = \{g|_{S_{\mathbf{b}}}(\mathbf{v}) \mid \mathbf{b} \in D(E_g)\}.$$

Since each polynomial in $\mathcal{G}$ is nonzero, Corollary 4.4 implies that there exists an assignment $\mathbf{a} \in \mathbb{F}^n$ such that $g|_{S_{\mathbf{b}}}(\mathbf{a}) \neq 0$ for all $g|_{S_{\mathbf{b}}}(\mathbf{v}) \in \mathcal{G}$. This implies

$$E_{\psi^D_{\mathbf{a}}(g)} = D(E_g).$$

Since $\psi^D_{\mathbf{a}}$ is a homomorphism, if we set $\mu^D = \psi^D_{\mathbf{a}}$ we are done. □

The following lemma handles the ($\mathsf{mod}\ p$) reduction and gives a bound on the number of monomials in the product after the reduction.

Lemma 4.11. *Let $\mathbf{x} = (x_1, \ldots, x_m)$ be formal variables, and let $f, g, h \in \mathbb{F}[\mathbf{x}]$ be such that $f = g \cdot h$. Then, there exists a homomorphism $\gamma : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{x}]$ such that:*

$$E_{\gamma(g)^{(\mathsf{mod}\ p)}} = E_g^{(\mathsf{mod}\ p)}, \quad \textit{and} \tag{2}$$

$$\|\gamma(f)^{(\mathsf{mod}\ p)}\|_0 \geq \frac{\|\gamma(g)^{(\mathsf{mod}\ p)} \cdot \gamma(h)^{(\mathsf{mod}\ p)}\|_0}{2^m}. \tag{3}$$

Proof. Let $\nu : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{z}][\mathbf{x}]$ be the homomorphism defined by

$$\nu(x_i) = z_i x_i,\ (1 \leq i \leq m)$$

and let $\nu^I : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{z}]$ be the homomorphism defined by

$$\nu(x_i) = z_i,\ (1 \leq i \leq m).$$

In addition, for $\mathbf{a} \in \mathbb{F}^n$, define $\nu_{\mathbf{a}} : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{x}]$ as the homomorphism defined by

$$\nu_{\mathbf{a}}(x_i) = a_i x_i,\ (1 \leq i \leq m).$$

Notice that for any polynomial $q(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$, we are regarding $\nu(q) \in \mathbb{F}[\mathbf{z}][\mathbf{x}]$ as a polynomial in the variables $\mathbf{x}$ and coefficients in the ring $\mathbb{F}[\mathbf{z}]$. For instance, if $q(\mathbf{x}) = 2x_1^4 + x_1^2 + x_1$, we have that $\nu(q) = 2z_1^4 x_1^4 + z_1^2 x_1^2 + z_1 x_1$ has monomials x_1^4, x_1^2 and x_1, with coefficients $2z_1^4, z_1^2$ and z_1, respectively. Hence, we have that

$$\nu(q)^{(\mathsf{mod}\ 3)} = 2z_1^4 x_1^{(4 \mod 3)} + z_1^2 x_1^{(2 \mod 3)} + z_1 x_1^{(1 \mod 3)} = z_1^2 x_1^2 + (2z_1^4 + z_1) x_1.$$

Notice that we can use the homomorphism ν^I to extract the coefficients of $\nu(q)^{(\mathsf{mod}\ p)}$, since for instance $\nu^I(q)|_{\{1,4\}} = 2z_1^4 + z_1$ is the coefficient of x_1 in $\nu(q)^{(\mathsf{mod}\ p)}$.

For each $\mathbf{b} \in E_g^{(\mathsf{mod}\ p)}$, let $S_{\mathbf{b}} = \{\mathbf{e} \in E_g \mid \mathbf{e} \overset{p}{=} \mathbf{b}\}$. Because $\bigcup_{\mathbf{b} \in E_g^{(\mathsf{mod}\ p)}} S_{\mathbf{b}} = E_g$ and $S_{\mathbf{b}} \cap S_{\mathbf{b}'} = \emptyset$ for every $\mathbf{b}, \mathbf{b}' \in E_g^{(\mathsf{mod}\ p)}$ such that $\mathbf{b} \neq \mathbf{b}'$, we have

$$g(\mathbf{x}) = \sum_{\mathbf{b} \in E_g^{(\mathsf{mod}\ p)}} \sum_{\mathbf{e} \in S_{\mathbf{b}}} a_{\mathbf{e}} \mathbf{x}^{\mathbf{e}}.$$

For $\mathbf{e} \in S_{\mathbf{b}}$ notice that

$$\nu(a_{\mathbf{e}}\mathbf{x}^{\mathbf{e}}) = a_{\mathbf{e}} \prod_{k=1}^{m} \nu(x_k)^{e_k} = a_{\mathbf{e}} \prod_{k=1}^{m} \left(z_k^{e_k} x_k^{e_k}\right) = a_{\mathbf{e}} \mathbf{z}^{\mathbf{e}} \mathbf{x}^{\mathbf{e}}$$
$$\Rightarrow \nu(a_{\mathbf{e}}\mathbf{x}^{\mathbf{e}})^{(\mathsf{mod}\ p)} = a_{\mathbf{e}} \mathbf{z}^{\mathbf{e}} \mathbf{x}^{(\mathbf{e} \mod p)} = a_{\mathbf{e}} \mathbf{z}^{\mathbf{e}} \mathbf{x}^{\mathbf{b}},$$

where the last implication follows from the fact that $\nu(a_{\mathbf{e}}\mathbf{x}^{\mathbf{e}})$ is a polynomial in $\mathbb{F}[\mathbf{z}][\mathbf{x}]$.

Therefore:

$$\nu(g)^{(\mathsf{mod}\ p)} = \sum_{\mathbf{b} \in E_g^{(\mathsf{mod}\ p)}} \sum_{\mathbf{e} \in S_{\mathbf{b}}} \nu(a_{\mathbf{e}}\mathbf{x}^{\mathbf{e}})^{(\mathsf{mod}\ p)} = \sum_{\mathbf{b} \in E_g^{(\mathsf{mod}\ p)}} \sum_{\mathbf{e} \in S_{\mathbf{b}}} a_{\mathbf{e}} \mathbf{z}^{\mathbf{e}} \mathbf{x}^{\mathbf{b}}$$
$$= \sum_{\mathbf{b} \in E_g^{(\mathsf{mod}\ p)}} \left(\sum_{\mathbf{e} \in S_{\mathbf{b}}} a_{\mathbf{e}} \mathbf{z}^{\mathbf{e}} \right) \mathbf{x}^{\mathbf{b}} = \sum_{\mathbf{b} \in E_g^{(\mathsf{mod}\ p)}} g|_{S_{\mathbf{b}}}(\mathbf{z})\ \mathbf{x}^{\mathbf{b}}.$$

Thus, for each $\mathbf{b} \in E_g^{(\mathsf{mod}\ p)}$, the coefficient of the monomial $\mathbf{x}^{\mathbf{b}}$ in $\nu(g)^{(\mathsf{mod}\ p)}$ is the nonzero polynomial $g|_{S_{\mathbf{b}}}(\mathbf{z}) \in \mathbb{F}[\mathbf{z}]$. Let

$$\mathcal{G}_1 = \{g|_{S_{\mathbf{b}}}(\mathbf{z}) \mid \mathbf{b} \in E_g^{(\mathsf{mod}\ p)}\}.$$

Let us now look at the products $\nu(g)^{(\mathsf{mod}\ p)} \cdot \nu(h)^{(\mathsf{mod}\ p)}$ and $\left(\nu(g)^{(\mathsf{mod}\ p)} \cdot \nu(h)^{(\mathsf{mod}\ p)}\right)^{(\mathsf{mod}\ p)}$.

By Observation 4.9,

$$\nu(f)^{(\mathsf{mod}\ p)} = \nu(g \cdot h)^{(\mathsf{mod}\ p)} = \left(\nu(g)^{(\mathsf{mod}\ p)} \cdot \nu(h)^{(\mathsf{mod}\ p)}\right)^{(\mathsf{mod}\ p)}. \tag{4}$$

We will now need to prove the following claim, which states that there are no monomial cancellations when we apply the $(\mathsf{mod}\ p)$ transformation on $\nu(g)^{(\mathsf{mod}\ p)} \cdot \nu(h)^{(\mathsf{mod}\ p)}$ to get the polynomial $\nu(g \cdot h)^{(\mathsf{mod}\ p)}$:

Claim 4.12. *The following equality holds:*

$$E_{\nu(g \cdot h)^{(\mathsf{mod}\ p)}} = E^{(\mathsf{mod}\ p)}_{\nu(g)^{(\mathsf{mod}\ p)} \cdot \nu(h)^{(\mathsf{mod}\ p)}}. \tag{5}$$

Proof. For each $\mathbf{b} \in E_g^{(\mathsf{mod}\ p)}$, let $S_{\mathbf{b}} = \{\mathbf{e} \in E_g \mid \mathbf{e} \overset{p}{=} \mathbf{b}\}$. For each $\mathbf{b}' \in E_h^{(\mathsf{mod}\ p)}$ let $T_{\mathbf{b}'} = \{\mathbf{e}' \in E_h \mid \mathbf{e}' \overset{p}{=} \mathbf{b}'\}$. In addition, let

$$M = \{\mathbf{b} + \mathbf{b}' \mid \mathbf{b} \in E_g^{(\mathsf{mod}\ p)} \text{ and } \mathbf{b}' \in E_h^{(\mathsf{mod}\ p)}\}.$$

For each $\mathbf{c} \in M$, let

$$U_{\mathbf{c}} = \{(\mathbf{b}, \mathbf{b}') \in E_g^{(\mathsf{mod}\ p)} \times E_h^{(\mathsf{mod}\ p)} \mid \mathbf{b}' + \mathbf{b} = \mathbf{c}\},$$

$$\mathcal{U}_{\mathbf{c}} = \bigcup_{(\mathbf{b},\mathbf{b}') \in U_{\mathbf{c}}} S_{\mathbf{b}} \times T_{\mathbf{b}'} \quad \text{and} \quad S(\mathcal{U}_{\mathbf{c}}) = \{\mathbf{e} + \mathbf{e}' \mid (\mathbf{e}, \mathbf{e}') \in \mathcal{U}_{\mathbf{c}}\}$$

and for $\mathbf{d} \in M^{(\mathsf{mod}\ p)}$, let

$$W_{\mathbf{d}} = \{(\mathbf{b}, \mathbf{b}') \in E_g^{(\mathsf{mod}\ p)} \times E_h^{(\mathsf{mod}\ p)} \mid \mathbf{b} + \mathbf{b}' \stackrel{p}{=} \mathbf{d}\}$$

$$\mathcal{W}_{\mathbf{d}} = \bigcup_{(\mathbf{b},\mathbf{b}') \in W_{\mathbf{d}}} S_{\mathbf{b}} \times T_{\mathbf{b}'} = \bigcup_{\mathbf{c} \in M, \mathbf{c} \stackrel{p}{=} \mathbf{d}} \mathcal{U}_{\mathbf{c}} \quad \text{and} \quad S(\mathcal{W}_{\mathbf{d}}) = \{\mathbf{e} + \mathbf{e}' \mid (\mathbf{e}, \mathbf{e}') \in \mathcal{W}_{\mathbf{d}}\} = \bigcup_{\mathbf{c} \in F, \mathbf{c} \stackrel{p}{=} \mathbf{d}} S(\mathcal{U}_{\mathbf{c}}).$$

Then, by Observation 4.2, and since $E_{\nu(g)} = E_g$ and $E_{\nu(h)} = E_h$, we have that

$$\begin{aligned} \nu(g)^{(\mathsf{mod}\ p)} \cdot \nu(h)^{(\mathsf{mod}\ p)} &= \sum_{\mathbf{c} \in M} \left(\nu^I(g) \cdot \nu^I(h)\right)|_{\mathcal{U}_{\mathbf{c}}}\ \mathbf{x}^{\mathbf{c}} \\ &= \sum_{\mathbf{c} \in M} \nu^I(g \cdot h)|_{S(\mathcal{U}_{\mathbf{c}})}\ \mathbf{x}^{\mathbf{c}} = \sum_{\mathbf{c} \in M} \nu^I(f)|_{S(\mathcal{U}_{\mathbf{c}})}\ \mathbf{x}^{\mathbf{c}} \end{aligned}$$

and that

$$\begin{aligned} \nu(g \cdot h)^{(\mathsf{mod}\ p)} &= \left(\nu(g)^{(\mathsf{mod}\ p)} \cdot \nu(h)^{(\mathsf{mod}\ p)}\right)^{(\mathsf{mod}\ p)} = \left(\sum_{\mathbf{c} \in M} \nu^I(f)|_{S(\mathcal{U}_{\mathbf{c}})}\ \mathbf{x}^{\mathbf{c}}\right)^{(\mathsf{mod}\ p)} \\ &= \sum_{\mathbf{d} \in M^{(\mathsf{mod}\ p)}} \left(\sum_{\mathbf{c} \in M, \mathbf{c} \stackrel{p}{=} \mathbf{d}} \nu^I(f)|_{S(\mathcal{U}_{\mathbf{c}})}\right) \mathbf{x}^{\mathbf{d}} \\ &= \sum_{\mathbf{d} \in M^{(\mathsf{mod}\ p)}} \nu^I(f)|_{S(\mathcal{W}_{\mathbf{d}})}\ \mathbf{x}^{\mathbf{d}}. \end{aligned}$$

Notice that for $\mathbf{c} \in M$ such that $\mathbf{c} \stackrel{p}{=} \mathbf{d}$ and the polynomial $\nu^I(f)|_{S(\mathcal{U}_{\mathbf{c}})}$ is a nonzero polynomial, then we must have that $\nu^I(f)|_{S(\mathcal{W}_{\mathbf{d}})}$ is also nonzero, since

$$\nu^I(f)|_{S(\mathcal{W}_{\mathbf{d}})} = \sum_{\mathbf{c} \in M, \mathbf{c} \stackrel{p}{=} \mathbf{d}} \nu^I(f)|_{S(\mathcal{U}_{\mathbf{c}})} \tag{6}$$

and $S(\mathcal{U}_{\mathbf{c}}) \cap S(\mathcal{U}_{\mathbf{c}'}) = \emptyset$ for $\mathbf{c} \neq \mathbf{c}'$ implies that each polynomial $\nu^I(f)|_{S(\mathcal{U}_{\mathbf{c}})}$ on the right hand side of equation (6) has a different set of monomials from the other polynomials $\nu^I(f)|_{S(\mathcal{U}_{\mathbf{c}'})}$. Therefore, each nonzero coefficient of $\nu(g)^{(\mathsf{mod}\ p)} \cdot \nu(h)^{(\mathsf{mod}\ p)}$ corresponds to a nonzero coefficient of $\nu(g \cdot h)^{(\mathsf{mod}\ p)}$ when we apply the $(\mathsf{mod}\ p)$ map. Equivalently, this proves that for each vector $\mathbf{c} \in E_{\nu(g)^{(\mathsf{mod}\ p)} \cdot \nu(h)^{(\mathsf{mod}\ p)}}$, there exists a vector $\mathbf{d} \in E_{\nu(g \cdot h)^{(\mathsf{mod}\ p)}}$ such that $(\mathbf{c} \mod p) = \mathbf{d}$. This implies that equation (5) holds. □

Since $E_g^{(\mathsf{mod}\ p)}, E_h^{(\mathsf{mod}\ p)} \subseteq \{0,\ldots,p-1\}^m$, we have

$$E_{\nu(g)^{(\mathsf{mod}\ p)}\cdot\nu(h)^{(\mathsf{mod}\ p)}} \subseteq \{0,1,\ldots,2p-2\}^m, \tag{7}$$

whereas we know that

$$E_{\nu(g\cdot h)^{(\mathsf{mod}\ p)}} \subseteq \{0,1,\ldots,p-1\}^m.$$

Let $\mathbf{e}$ be an element of $E_{\nu(g\cdot h)^{(\mathsf{mod}\ p)}}$. There can be at most 2^m elements $\mathbf{e}' \in E_{\nu(g)^{(\mathsf{mod}\ p)}\cdot\nu(h)^{(\mathsf{mod}\ p)}}$ such that $\mathbf{e}' \overset{p}{=} \mathbf{e}$, because of (7). That is, these will be the elements of the form $\mathbf{e}' = \mathbf{e} + p\mathbf{v}$, where $\mathbf{v} \in \{0,1\}^n$. Therefore, we have that the coefficient of $\mathbf{x}^{\mathbf{e}}$ in $\nu(g\cdot h)^{(\mathsf{mod}\ p)}$ is the sum of at most 2^m coefficients of $\nu(g)^{(\mathsf{mod}\ p)}\cdot\nu(h)^{(\mathsf{mod}\ p)}$. This fact, together with equality (5), implies that

$$|E_{\nu(g)^{(\mathsf{mod}\ p)}\cdot\nu(h)^{(\mathsf{mod}\ p)}}| \leq 2^m \cdot |E_{\nu(g\cdot h)^{(\mathsf{mod}\ p)}}|. \tag{8}$$

Let

$$\mathcal{G}_2 = \{\nu(f)|_{S(\mathcal{W}_{\mathbf{e}})}(\mathbf{z}) \mid \mathbf{e} \in E_{\nu(g\cdot h)^{(\mathsf{mod}\ p)}}\} \cup \{\nu(f)|_{S(\mathcal{W}_{\mathbf{c}})}(\mathbf{z}) \mid \mathbf{c} \in E_{\nu(g)^{(\mathsf{mod}\ p)}\cdot\nu(h)^{(\mathsf{mod}\ p)}}\}$$

Since $\mathcal{G}_1 \cup \mathcal{G}_2$ is a set of non-zero polynomials in $\mathbf{z}$, by Corollary 4.4 we can find $\mathbf{a} \in \mathbb{F}^n$ no polynomial in $\mathcal{G}_1 \cup \mathcal{G}_2$ will evaluate to zero, and hence we obtain that

$$g|_{S_{\mathbf{b}}}(\mathbf{a}) \neq 0 \ \text{ for all } \ g|_{S_{\mathbf{b}}}(\mathbf{z}) \in \mathcal{G}_1 \Rightarrow E_{\nu_{\mathbf{a}}(g)^{(\mathsf{mod}\ p)}} = E_g^{(\mathsf{mod}\ p)}$$

$$\nu(f)|_{S(\mathcal{W}_{\mathbf{e}})}(\mathbf{a}) \neq 0\ \forall \mathbf{e} \in E_{\nu(g\cdot h)^{(\mathsf{mod}\ p)}} \Rightarrow \|\nu_{\mathbf{a}}(g\cdot h)^{(\mathsf{mod}\ p)}\|_0 = |E_{\nu(g\cdot h)^{(\mathsf{mod}\ p)}}|$$

$$\nu(f)|_{S(\mathcal{W}_{\mathbf{c}})}(\mathbf{a}) \neq 0\ \forall \mathbf{c} \in E_{\nu(g)^{(\mathsf{mod}\ p)}\cdot\nu(h)^{(\mathsf{mod}\ p)}} \Rightarrow |E_{\nu(g)^{(\mathsf{mod}\ p)}\cdot\nu(h)^{(\mathsf{mod}\ p)}}| = \|\nu_{\mathbf{a}}(g)^{(\mathsf{mod}\ p)}\cdot\nu_{\mathbf{a}}(h)^{(\mathsf{mod}\ p)}\|_0$$

Hence,

$$\begin{aligned}
\|\nu_{\mathbf{a}}(g\cdot h)^{(\mathsf{mod}\ p)}\|_0 &= |E_{\nu(g\cdot h)^{(\mathsf{mod}\ p)}}| \\
&\geq \frac{|E_{\nu(g)^{(\mathsf{mod}\ p)}\cdot\nu(h)^{(\mathsf{mod}\ p)}}|}{2^m} && \text{(by equation (8))} \\
&= \frac{\left\|\nu_{\mathbf{a}}(g)^{(\mathsf{mod}\ p)}\cdot\nu_{\mathbf{a}}(h)^{(\mathsf{mod}\ p)}\right\|_0}{2^m}.
\end{aligned}$$

Since $\nu_{\mathbf{a}}$ is a homomorphism, by setting $\gamma = \nu_{\mathbf{a}}$ we are done. □

Combining Lemma 4.10 and 4.11 we obtain our final lemma which combines an application of a linear map on the exponents together with the modular reduction.

Lemma 4.13. *Let $\mathbf{x} = (x_1,\ldots,x_n)$ and $\mathbf{y} = (y_1,\ldots,y_m)$ be formal variables and $f,g,h \in \mathbb{F}[\mathbf{x}]$ be such that $f(\mathbf{x}) = g(\mathbf{x})\cdot h(\mathbf{x})$. Let $D \in \mathbb{N}_0^{m\times n}$ be a matrix with column vectors $\mathbf{d}_k$, for $1 \leq k \leq n$. Then, there exists a transformation $\varphi^D : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{y}]$ such that*

(i) $E_{\varphi^D(g)} = D(E_g)^{(\mathsf{mod}\ p)}$,

(ii) $\|\varphi^D(f)\|_0 \geq \dfrac{\|\varphi^D(g)\cdot\varphi^D(h)\|_0}{2^m}$

(iii) $\|f\|_0 \geq \|\varphi^D(f)\|_0$.

Proof. Let $\varphi^D : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{y}]$ be defined as

$$\varphi^D = (\mathsf{mod}\ p)\circ\gamma\circ\mu^D,$$

where $\mu^D : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{y}]$ is the homomorphism from Lemma 4.10 and $\gamma : \mathbb{F}[\mathbf{y}] \to \mathbb{F}[\mathbf{y}]$ is the homomorphism from Lemma 4.11. That is, if $f \in \mathbb{F}[\mathbf{x}]$ then $\varphi^D(f) = \gamma(\mu^D(f))^{(\mathsf{mod}\ p)}$.

Part (i) holds because

$$\begin{aligned} E_{\gamma(\mu^D(g))^{(\mathsf{mod}\ p)}} &= E^{(\mathsf{mod}\ p)}_{\mu^D(g)} && \text{(by Lemma 4.11)}\\ &= D(E_g)^{(\mathsf{mod}\ p)} && \text{(by Lemma 4.10, since } E_{\mu^D(g)} = D(E_g)\text{)}\end{aligned}$$

To prove part (ii) notice that

$$\begin{aligned}\varphi^D(f) &= \gamma(\mu^D(f))^{(\mathsf{mod}\ p)} = \gamma(\mu^D(g\cdot h))^{(\mathsf{mod}\ p)}\\ &= \gamma(\mu^D(g)\cdot\mu^D(h))^{(\mathsf{mod}\ p)} && \text{(because } \mu^D \text{ is a homomorphism)}\end{aligned}$$

and therefore by Lemma 4.11 applied on polynomials $\mu^D(f), \mu^D(g)$ and $\mu^D(h)$, we have

$$\begin{aligned}\|\varphi^D(f)\|_0 &= \|\gamma(\mu^D(f))^{(\mathsf{mod}\ p)}\|_0\\ &\geq \frac{\|\gamma(\mu^D(g))^{(\mathsf{mod}\ p)}\cdot\gamma(\mu^D(h))^{(\mathsf{mod}\ p)}\|_0}{2^m}\\ &= \frac{\|\varphi^D(g)\cdot\varphi^D(h)\|_0}{2^m}\end{aligned}$$

Notice that part (iii) follows from

$$\begin{aligned}\|f\|_0 = |E_f| &\geq |D(E_f)| && \text{(by Observation 4.6)}\\ &\geq |D(E_f)^{(\mathsf{mod}\ p)}| \geq |E_{\varphi^D(f)}| && \text{(by remark 4.8)}\\ &= \|\varphi^D(f)\|_0.\end{aligned}$$

□

5 Polytopes Defined by a Hyperplane of $\mathbb{F}_p^\ell$

In this section, we will introduce a polytope in $\mathbb{R}^\ell$ defined by the points of a hyperplane of $\mathbb{F}_p^\ell$ embedded in $\mathbb{N}_0^\ell$ and we will give a lower bound on the number of vertices of some polytopes defined by certain special hyperplanes in $\mathbb{F}_p^\ell$, for a prime $p > \ell$. These special polytopes are the key ingredients of the proof of the sparsity bounds that we obtain in section 6. More intuition on our motivation to study these polytopes will be given after the basic definitions.

Throughout this section, we identify $\mathbb{F}_p$ with the set $\{0,\ldots,p-1\} \subset \mathbb{N}_0$. Based on this convention, we define the natural embedding of $\mathbb{F}_p^\ell$ into $\mathbb{N}_0^\ell$ as the map $\phi : \mathbb{F}_p^\ell \to \mathbb{N}_0^\ell$ defined by $\phi(\mathbf{x}) = \mathbf{x}$ (notice that here we use that $\mathbf{x} \in \{0,\ldots,p-1\}^\ell$). We extend the definition of ϕ to sets $S \subset \mathbb{N}_0$ by applying ϕ element wise, that is, $\phi(S) = \{\phi(\mathbf{x}) \mid \mathbf{x} \in S\}$.

We begin with the following basic definitions:

Definition 5.1. *Let $A \subset \mathbb{F}_p^\ell$ be any set. The polytope P_A associated to set A is defined by*

$$P_A = CS(\phi(A)).$$

Observation 5.2. *Notice that, by this definition, we have that $V(P_A) \subset \phi(A) \subset \mathbb{N}_0^\ell$, and therefore all vertices of P_A have integer coordinates.*

Definition 5.3 (Border points and Proper points)**.** *A point $\mathbf{v} \in \mathbb{F}_p^\ell$ is called a* border point *if $\exists j \in \{1,\ldots,\ell\}$ s.t. $\mathbf{v}_j \in \{0, p-1\}$. If no $\mathbf{v}_j \in \{0, p-1\}$, then we say that $\mathbf{v}$ is a* proper point*.*

We start with an initial polytope construction that we will later use as a building block in our final polytope.

Theorem 5.4. *Let $t \in \mathbb{N}$, $p_1, p_2, \ldots, p_t$ be primes such that $3 < p_1 < \ldots < p_t$ and $Q = 3 \cdot \prod_{i=1}^{t} p_i$. In addition, let p be a prime such that $2tQ < p$, $q_i = Q/p_i$, for $1 \le i \le t$ and $\mathbf{h} = (-q_1, -q_2, \ldots, -q_t, 1) \in \mathbb{F}_p^{t+1}$. Let $B \subset \mathbb{F}_p$ be the set of values of $b \in \mathbb{F}_p$ for which the hyperplane*

$$\mathcal{H} = \{\mathbf{x} \in \mathbb{F}_p^{t+1} \mid \mathbf{h} \cdot \mathbf{x} \stackrel{p}{=} b\}$$

is such that $V(P_{\mathcal{H}})$ has at least one proper point. Then, $|B| \ge \dfrac{Q}{3(5/4)^t}$.

Proof. Let $b \in \{0, 1 \ldots, p-1\}$ be such that

$$b = p - 2 - \sum_{i=1}^{t} a_i q_i,$$

where $a_i \in \mathbb{N}_0$, $p_i \nmid a_i$ and $a_i < \dfrac{p}{2q_i t}$, for $1 \le i \le t$. Notice that $0 \le b < p-1$, because

$$\sum_{i=1}^{t} a_i q_i < \sum_{i=1}^{t} \frac{p}{2t} = t \cdot \frac{p}{2t} = \frac{p}{2} < p - 1.$$

Given our choice for the vector $\mathbf{h}$, the equation for the hyperplane $\mathcal{H}$ is

$$x_{t+1} \stackrel{p}{=} b + \sum_{i=1}^{t} q_i x_i. \tag{9}$$

Claim 5.5. *Let $F = \{\mathbf{x} \in P_{\mathcal{H}} \mid x_{t+1} = b + \sum_{i=1}^{t} q_i x_i\}$ be a hyperplane in $\mathbb{R}^{t+1}$. Then, F defines a face of $P_{\mathcal{H}}$.*

Proof. By definition 3.6, to show that F is a face of $P_{\mathcal{H}}$ we need to show that the hyperplane defined by (now equation is taken over $\mathbb{R}$)

$$x_{t+1} = b + \sum_{i=1}^{t} q_i x_i \tag{10}$$

is a supporting hyperplane of $P_{\mathcal{H}}$. This is indeed the case because

$$\mathbf{x} \in \phi(\mathcal{H}) \Rightarrow x_{t+1} - b - \sum_{i=1}^{t} q_i x_i \stackrel{p}{=} 0 \Rightarrow p \mid (x_{t+1} - b - \sum_{i=1}^{t} q_i x_i)$$

and

$$\mathbf{x} \in \phi(\mathcal{H}) \Rightarrow x_{t+1} - b - \sum_{i=1}^{t} q_i x_i < p,$$

where the last inequality follows since $x_{t+1} < p$ and the terms $b, q_i, x_i \geq 0$, for $i \leq i \leq t$.

Therefore, we have that

$$x_{t+1} - b - \sum_{i=1}^{t} q_i x_i \leq 0$$

for all $\mathbf{x} \in \phi(\mathcal{H})$, which implies that this inequality holds for all points $\mathbf{x} \in P_{\mathcal{H}}$. In addition, notice that the point $(a_1, a_2, \ldots, a_t, p-2)$ satisfies equation (10) and it also belongs to $P_{\mathcal{H}}$. Therefore the hyperplane defined by equation (10) intersects $P_{\mathcal{H}}$.

Hence, equation (10) indeed defines a supporting hyperplane, as defined in definition 3.5. □

To finish the proof, we will prove that the proper point $\mathbf{v} = (a_1, a_2, \ldots, a_t, p-2) \in V(F)$, and therefore by Lemma 3.10 must belong to $V(P_{\mathcal{H}})$. First of all, notice that $\mathbf{v} \in \mathcal{H}$ and $\mathbf{v} \in F$, since

$$0 < a_i < \frac{p}{2t} < p, \ \forall i \in \{1, \ldots, t\}$$

$$\sum_{i=1}^{t} q_i v_i = \sum_{i=1}^{t} a_i q_i < \sum_{i=1}^{t} \frac{p}{2t} < \frac{p}{2} \quad \text{and}$$

$$p - 2 = v_{t+1} = b + \sum_{i=1}^{t} q_i v_i = p - 2 - \sum_{i=1}^{t} a_i q_i + \sum_{i=1}^{t} a_i q_i = p - 2.$$

In order to prove that $\mathbf{v} \in V(P_{\mathcal{H}})$, we will need the following claim:

Claim 5.6. *Let $\mathbf{x} \in F \setminus \{\mathbf{v}\}$. Then, $x_{t+1} \leq p - 3$.*

Proof. Notice that to prove this claim, it is enough to prove it for the points $\mathbf{x} \in V(F) \setminus \{\mathbf{v}\}$, since points of F are convex combinations of $V(F) \cup \{\mathbf{v}\}$. Since $V(F) \subseteq V(P_{\mathcal{H}}) \subset \mathbb{N}_0^{t+1}$, we know that all points in $V(F)$ are integer points.

Suppose, for the sake of contradiction, that there was a point $\mathbf{x} \in V(F) \setminus \{\mathbf{v}\}$ such that $x_{t+1} \in \{p-2, p-1\}$. Then, we have two cases to analyze:

Case 1: $x_{t+1} = p-1$.

In this case, by equation (10) we have that

$$p-1 = x_{t+1} = b + \sum_{i=1}^{t} q_i x_i = p-2 - \sum_{i=1}^{t} q_i a_i + \sum_{i=1}^{t} q_i x_i \Rightarrow$$

$$\Rightarrow 1 = \sum_{i=1}^{t} q_i (x_i - a_i) \tag{11}$$

Notice that 3 divides the right hand side of equation (11) (since $3 \mid q_i,\ \forall i$), but it does not divide the left hand side. Hence we reached a contradiction.

Case 2: $x_{t+1} = p-2$.

Again, by equation (10) we have

$$p-2 = x_{t+1} = b + \sum_{i=1}^{t} q_i x_i = p-2 + \sum_{i=1}^{t} q_i (x_i - a_i) \Rightarrow$$

$$\Rightarrow \sum_{i \in S_-} (a_i - x_i) q_i = \sum_{i \in S_+} q_i (x_i - a_i), \tag{12}$$

where $S_- = \{i \mid x_i < a_i,\ 1 \leq i \leq t\}$ and $S_+ = \{i \mid x_i > a_i,\ 1 \leq i \leq t\}$. Notice that $\mathbf{x} \in F \setminus \{\mathbf{v}\}$ implies that $\mathbf{x}$ must have an index $i \in \{1, \dots t\}$ such that $x_i < a_i$, otherwise we will have $x_i \geq a_i, 1 \leq i \leq t$ and $\exists k \in \{1, \dots t\}$ such that $x_k > a_k$ (because $\mathbf{x} \neq \mathbf{v}$). This implies that $p-2 = x_{t+1} = p-2 + \sum_{i=1}^{t} q_i (x_i - a_i) \geq p-2+q_k \geq p-2+3 > p$, which is a contradiction. Hence, we have that $S_- \neq \emptyset$. In addition, notice that $S_+ \neq \emptyset$, otherwise we would have that $x_i \leq a_i$ for all $1 \leq i \leq t$ and $\exists k \in \{1, \dots t\}$ such that $x_k < a_k$ (because $\mathbf{x} \neq \mathbf{v}$), which implies that $p-2 = x_{t+1} = p-2 + \sum_{i=1}^{t} q_i (x_i - a_i) < p-2-q_k < p-2$, which is a contradiction. Therefore, if we let $i \in S_+$, we have that p_i does not divide the right hand side of equation (12), whereas p_i must divide the left hand side of equation (12) (because $i \notin S_-$ and $S_- \neq \emptyset$), since $p_i \mid q_j$ for all $j \neq i$ and $p_i \nmid q_i$.

Since we reached a contradiction in both cases, we have proved Claim 5.6. □

Now, the proof that $\mathbf{v} \in V(F)$ follows from Claim 5.6 and from Observation 3.9 because $\mathbf{v}$ is the only point of F such that $v_{t+1} = p-2$ and therefore it cannot be written as the convex combination of any point of the set $F \setminus \{\mathbf{v}\}$ (because by Claim 5.6 their last coordinate is $< p-2$).

Since we have proved that $\mathbf{v} \in V(F)$ and we showed that it is a proper point, by Lemma 3.10 we have that $\mathbf{v} \in V(P_{\mathcal{H}})$.

To obtain the bound on the size of the set B of good values of b, notice that $2Qt < p \Rightarrow p_i < \frac{p}{2q_i t}$ and therefore for any choice of $(a_1, \ldots, a_t)$ such that $0 < a_i < p_i$ in the proof above shows that $\mathbf{v}$ will be a proper vertex of $P_{\mathcal{H}}$. This is true because

$$0 < a_i < p_i \Rightarrow p_i \nmid a_i \text{ and } 0 < a_i < p_i < \frac{p}{2q_i t}.$$

Hence, we have that

$$|B| \geq \prod_{i=1}^{t}(p_i - 1) = \frac{Q}{3} \cdot \prod_{i=1}^{t} \frac{p_i - 1}{p_i} \geq \frac{Q}{3} \cdot \prod_{i=1}^{t} \frac{4}{5} = \frac{Q}{3} \cdot (4/5)^t$$

□

Our final construction will use the polytope from Theorem 5.4 to construct a polytope in slightly higher dimension with many more vertices. The idea is to construct a polytope in which many restrictions 'look like' the polytope of Theorem 5.4 and so that the proper vertices of these restricted polytopes do not overlap. Before describing the construction we state some useful facts on restrictions.

Definition 5.7 (Variable Fixings)**.** *Let $\mathbb{F}$ be a field and $\mathcal{H} \subseteq \mathbb{F}^m$. Let $T \subseteq \{1, \ldots, m\}$ be a set of indices, $\overline{T} = \{1, \ldots, m\} \setminus T$ and $\mathbf{c} \in \mathbb{F}^{\overline{T}}$. Define the fixing $(T, \mathbf{c})$ as the set*

$$\mathcal{H}|_{(T,\mathbf{c})} = \{\mathbf{x} \in \mathcal{H} \mid \mathbf{x}_{\overline{T}} = \mathbf{c}\}.$$

That is, $\mathcal{H}|_{(T,\mathbf{c})}$ is the set of all points of $\mathcal{H}$ with coordinates $\overline{T}$ fixed to $\mathbf{c}$.

Observation 5.8. *Notice that for $\mathcal{H} = \{\mathbf{x} \in \mathbb{F}_p^m \mid \mathbf{h} \cdot \mathbf{x} \overset{p}{=} b\}$, where $\mathbf{h}$ has more than one nonzero coordinate, the hyperplane*

$$Z_i = \{\mathbf{x} \in \mathbb{R}^m \mid x_i = 0\}$$

is a supporting hyperplane of $P_{\mathcal{H}}$, since $\mathbf{x} \in P_{\mathcal{H}} \Rightarrow 0 \leq x_i \leq p - 1 \ \forall i$, and there exists $\mathbf{x} \in P_{\mathcal{H}}$ such that $\mathbf{x} \in P_{\mathcal{H}} \cap Z_i$. Similarly, the hyperplane $\{\mathbf{x} \in \mathbb{R}^m \mid x_i = p - 1\}$ is also a supporting hyperplane of $P_{\mathcal{H}}$.

The following simple Corollary shows that restricting some of the coordinates to 0 or to $p - 1$ gives a supporting hyperplane for the polytope $P_{\mathcal{H}}$.

Corollary 5.9. *Let $\mathcal{H} = \{\mathbf{x} \in \mathbb{F}_p^m \mid \mathbf{h} \cdot \mathbf{x} \overset{p}{=} b\}$ be a hyperplane of $\mathbb{F}_p^m$, $T \subseteq \{1, 2, \ldots, m\}$ and $\mathbf{c} \in \mathbb{F}_p^{\overline{T}}$ be a vector such that $\mathbf{c} \in \{0, p-1\}^{\overline{T}}$. Then, $P_{\mathcal{H}|_{(T,\mathbf{c})}}$ is a face of $P_{\mathcal{H}}$.*

Proof. For $i \in \{1, 2, \ldots, m\}$, let $L_i = \mathbb{R}^m|_{\overline{\{i\}},c_i} = \{\mathbf{x} \in \mathbb{R}^m \mid x_i = c_i\}$. Notice that

$$\phi(\mathcal{H}|_{(T,\mathbf{c})}) = \phi(\mathcal{H}) \cap \left(\bigcap_{i \in \overline{T}} L_i \right).$$

Therefore, we have that

$$P_{\mathcal{H}|_{(T,\mathbf{c})}} = P_{\mathcal{H}} \cap \left(\bigcap_{i \in \overline{T}} L_i \right) = \bigcap_{i \in \overline{T}} (P_{\mathcal{H}} \cap L_i). \tag{13}$$

Notice that $\mathbf{c} \in \{0, p-1\}^{\overline{T}}$ and Observation 5.8 imply that L_i is a supporting hyperplane of $P_{\mathcal{H}}$ for each $i \in \overline{T}$, which implies that $L_i \cap P_{\mathcal{H}}$ is a face of $P_{\mathcal{H}}$. Hence, Lemma 3.10 and equation (13) imply that $P_{\mathcal{H}|_{(T,\mathbf{c})}}$ is the intersection of faces of $P_{\mathcal{H}}$, and therefore $P_{\mathcal{H}|_{(T,\mathbf{c})}}$ must be a face of $P_{\mathcal{H}}$ as well. □

We now state and prove our main construction.

Theorem 5.10. *Let $\ell, t \in \mathbb{N}$ be such that t divides ℓ. Let $p_1, p_2, \ldots, p_t$ be primes such that $3 < p_1 < \ldots < p_t$, $Q = 3 \cdot \prod_{i=1}^{t} p_i$ and p be a prime such that $2tQ < p$. Then, there exists a hyperplane $\mathcal{H} = \{\mathbf{x} \in \mathbb{F}_p^{\ell+2} \mid \mathbf{h} \cdot \mathbf{x} \stackrel{p}{=} 0\}$, passing through the origin, such that*

$$|V(P_{\mathcal{H}})| \geq 2^{\ell} \cdot \frac{(2\ell)^t}{(5t)^t} \cdot \frac{Q}{3p}.$$

Proof. Let $r = \frac{\ell}{t}$, $q_i = Q/p_i$, for $1 \leq i \leq t$. For each $b \in \mathbb{F}_p$, let $\mathcal{H}_b$ be the hyperplane defined by

$$\mathcal{H}_b = \{\mathbf{x} \in \mathbb{F}_p^{\ell+2} \mid b x_{\ell+2} + x_{\ell+1} - \sum_{i=1}^{t} q_i \sum_{j=1}^{r} x_{(i-1)r+j} \stackrel{p}{=} 0\} \tag{14}$$

Notice that to finish the proof of this theorem, it suffices to prove the following lower bound:

$$\sum_{b \in \mathbb{F}_p} |V(P_{\mathcal{H}_b})| \geq 2^{\ell} \cdot \frac{(2\ell)^t}{(5t)^t} \cdot \frac{Q}{3}.$$

This will imply that one of the hyperplanes $\mathcal{H}_b$ is such that $P_{\mathcal{H}_b}$ has the desired number of vertices.

Let $S_i = \{k \mid k = (i-1)r + j,\ 1 \leq j \leq r\}$, for $1 \leq i \leq t$. That is, S_i is the set of indices of all variables that have coefficient $-q_i$ in $\mathcal{H}_b$. Notice that $|S_i| = r$, for every i.

The following claim shows that many restrictions of $P_{\mathcal{H}_b}$ have a proper vertex.

Claim 5.11. *Let $T \subset \{1, 2, \ldots, \ell+2\}$ be a set of size $t+1$ so that $\{\ell+1\} \subset T$, $\{\ell+2\} \not\subset T$ and $|T \cap S_i| = 1$, for all $1 \leq i \leq t$. Let $\mathbf{c} \in \{0, p-1\}^{\overline{T}}$ be such that $c_{\ell+2} = p-1$. Then, there are at least $\dfrac{Q}{3(5/4)^t}$ values of $b \in \mathbb{F}_p$ for which $P_{\mathcal{H}_b|_{T,\mathbf{c}}}$ has a vertex of the form $(\mathbf{x}_T, \mathbf{c})$, where $1 \leq x_i \leq p-2$ for all $i \in T$.*

Proof. To simplify notations we will treat the elements of a vector $\mathbf{x} \in \mathbb{F}_p^m$ as though they are ordered with the indices in T first, followed by the indices in $\overline{T}$. That is, a vector $\mathbf{x} = (\mathbf{x}_T, \mathbf{x}_{\overline{T}})$. For each $1 \leq i \leq t$, let $\{k_i\} = T \cap S_i$. Let $\alpha = \mathbf{h}_{\overline{T}} \cdot \mathbf{c}$. Then,

$$\mathcal{H}_b|_{(T,\mathbf{c})} = \{(\mathbf{x}_T, \mathbf{c}) \in \mathbb{F}_p^{\ell+2} \mid x_{\ell+1} - b - \sum_{i=1}^{t} q_i x_{k_i} \stackrel{p}{=} \alpha\}.$$

That is, $\mathcal{H}_b|_{(T,\mathbf{c})}$ is the set of all points $\mathbf{x} \in \mathcal{H}_b$ such that $\mathbf{x}_{\overline{T}} = \mathbf{c}$. This definition implies that all points $\mathbf{x} \in P_{\mathcal{H}_b|_{(T,\mathbf{c})}}$ are such that $\mathbf{x}_{\overline{T}} = \mathbf{c}$, since the points of $P_{\mathcal{H}_b|_{(T,\mathbf{c})}}$ are all convex combinations of points in $\phi(\mathcal{H}_b|_{(T,\mathbf{c})})$. Hence, polytope $P_{\mathcal{H}_b|_{(T,\mathbf{c})}}$ is isomorphic to the polytope $P_{\mathcal{F}}$ defined by hyperplane

$$\mathcal{F} = \{\mathbf{x}_T \in \mathbb{F}_p^{\ell+2} \mid x_{\ell+1} - b - \sum_{i=1}^{t} q_i x_{k_i} \stackrel{p}{=} \alpha\}.$$

Hence, a proper vertex of $P_{\mathcal{F}}$ corresponds to a vertex of $P_{\mathcal{H}_b|_{(T,\mathbf{c})}}$ of the form $(\mathbf{x}_T, \mathbf{c})$, where $1 \leq x_i \leq p-2$ for all $i \in T$.

By Theorem 5.4, there are at least $\dfrac{Q}{3(5/4)^t}$ values of $b + \alpha$ for which $P_{\mathcal{F}}$ has at least one proper vertex. Since α is fixed, this implies that there are at least $\dfrac{Q}{3(5/4)^t}$ values for b. By the isomorphism described above, this concludes the claim. □

By Corollary 5.9, $P_{\mathcal{H}_b|_{(T,\mathbf{c})}}$ is a face of $P_{\mathcal{H}_b}$. Hence, Lemma 3.10 implies that any vertex of $P_{\mathcal{H}_b|_{(T,\mathbf{c})}}$ is a vertex of $P_{\mathcal{H}_b}$. Let $\mathbf{x}_{(T,\mathbf{c})}$ be any such vertex of $P_{\mathcal{H}_b|_{(T,\mathbf{c})}}$ as described by Claim 5.11. We will need the following claim:

Claim 5.12. *$\mathbf{x}_{(T,\mathbf{c})} \neq \mathbf{x}_{(T',\mathbf{c}')}$, for any two distinct fixings $(T,\mathbf{c}) \neq (T',\mathbf{c}')$.*

Proof. Let $\mathbf{y} = \mathbf{x}_{(T,\mathbf{c})}$ and $\mathbf{z} = \mathbf{x}_{(T',\mathbf{c}')}$. Thus we need to show that $\mathbf{y} \neq \mathbf{z}$.

Notice that $(T,\mathbf{c}) \neq (T',\mathbf{c}')$ implies that $T \neq T'$ or that $T = T'$ and $\mathbf{c} \neq \mathbf{c}'$. Hence, we have two cases to analyze:

Case 1: $T \neq T'$.

By symmetry, we can assume w.l.o.g. that $T \not\subset T'$.

$$T \not\subset T' \Rightarrow \exists i \in T \setminus T' \Rightarrow y_i \notin \{0, p-1\} \text{ and } z_i \in \{0, p-1\} \Rightarrow$$
$$\Rightarrow y_i \neq z_i \Rightarrow \mathbf{y} \neq \mathbf{z}.$$

Case 2: $T = T'$ and $\mathbf{c} \neq \mathbf{c}'$.

In this case, there exists $i \in \overline{T}$ such that $c_i \neq c'_i$, which implies that $\mathbf{y} \neq \mathbf{z}$. □

Claims 5.11 and 5.12 imply that each pair $(T,\mathbf{c})$ gives us a distinct vertex of $P_{\mathcal{H}_b}$, for at least $\dfrac{Q}{3(5/4)^t}$ polytopes $P_{\mathcal{H}_b}$. Hence, the total number of vertices in the polytopes $P_{\mathcal{H}_b}$, where $b \in \mathbb{F}_p$, is lower bounded by the number of pairs $(T,\mathbf{c})$ multiplied by $\dfrac{Q}{3(5/4)^t}$.

Notice that we can choose the set T in r^t ways, because there are r ways of choosing an index $k_i \in S_i$ for which $S_i \cap T = \{k_i\}$. In addition, $|T| = t$ implies that $|\overline{T}| = l - t + 1$, and therefore are 2^{l-t} possible fixings of the variables in $\mathbf{c} \in \{0,1\}^{\overline{T}}$ for which $c_{\ell+2} = p-1$. Hence, the number of pairs $(T, \mathbf{c})$ is $2^{\ell-t}r^t$.

Hence, we have the following inequality:

$$\sum_{b \in \mathbb{F}_p} |V(P_{\mathcal{H}_b})| \geq 2^{\ell-t}r^t \cdot \frac{Q}{3(5/4)^t} = 2^\ell \cdot \frac{(2\ell)^t}{(5t)^t} \cdot \frac{Q}{3}.$$

This proves the theorem. □

6 Proof of Sparsity Bound

We now restate our main theorem (Theorem 1) and prove it. We did not make an effort to optimize the constants hidden in the big 'O' and it is very possible that the constant resulting from our proof can be substantially improved.

Theorem 6.1. *Let $f, g \in \mathbb{F}[x_1, \ldots, x_n]$ be polynomials such that $g(\mathbf{x})$ divides $f(\mathbf{x})$ and let $d \geq 64$ be an upper bound on the degree of each variable of f, that is, $\deg_i(f) \leq d$ for all $1 \leq i \leq n$. If s_f and s_g are the sparsities of f and g, respectively, then*

$$s_g \leq \max(s_f^{O(\log s_f \log\log s_f)}, d^{O(\log d)}).$$

Proof. Notice that we can assume, without loss of generality, that the field $\mathbb{F}$ is algebraically closed, since $g(\mathbf{x}) \mid f(\mathbf{x})$ in $\mathbb{F}[\mathbf{x}]$ implies that $g(\mathbf{x}) \mid f(\mathbf{x})$ in $\overline{\mathbb{F}}[\mathbf{x}]$.

We start by setting the parameters involved in the proof and calculating the relationships between them.

Setting the parameters: We can assume that $s_g > d^{18000 \log d}$, else we are done. Thus, we need to show that

$$s_g \leq s_f^{O(\log s_f \log\log s_f)}.$$

Let $f = g \cdot h$, $t \in \mathbb{N}_0$ be the integer such that

$$t^{24t^2} \leq s_g < (t+1)^{24(t+1)^2}.$$

Notice that $d^{18000 \log d} < s_g < (t+1)^{24(t+1)^2}$ and $d \geq 64$ implies

$$2^{36 \cdot 18000} < s_g < (t+1)^{24(t+1)^2} \Rightarrow 27000 < (t+1)^2 \log(t+1) \Rightarrow t \geq 60.$$

Let

$$Q = 3 \cdot \prod_{i=1}^{t} p_i,$$

where $2 < 3 < p_1 < p_2 \ldots < p_t$ are the first $t+2$ primes. By the bounds on the primorial function given by Lemma 2.1, we know that

$$t^{4t/5} < Q < t^{(11t/5)}.$$

Let p be a prime such that $2tQ < p < 4tQ$. Hence, we have that

$$t^{4t/5} < Q < p < t^{3/2}Q < t^{(11t/5+3/2)}. \tag{15}$$

which implies

$$\begin{aligned} p^{10t} < t^{10t(11t/5+3/2)} &= t^{22t^2+15t} \\ &< t^{23t^2} && \text{(because } t \geq 50\text{)} \\ &< t^{24t^2} \leq s_g \end{aligned}$$

and

$$p^{35t} > t^{35t\frac{4t}{5}} = t^{28t^2} > \tag{16}$$

$$> (t+1)^{26t^2} > (t+1)^{24(t+1)^2} > s_g \qquad \text{(because } t \geq 50\text{)}. \tag{17}$$

Therefore, we have that

$$p^{10t} < s_g < p^{35t}. \tag{18}$$

Therefore, if we set $\ell = 3t$, we have that

$$p^{c(\ell+2)} < s_g < p^{12\ell}, \quad \text{for} \quad c = 3 + 1/9.$$

Claim 6.2. *For this choice of parameters we have $p > d$.*

Proof. Notice that

$$s_g < p^{35t} < t^{35t(\frac{11t}{5}+3/2)} < t^{77t^2+53t} < t^{80t^2} \quad \text{for } t \geq 50.$$

On the other hand, we have $d^{125\log d} < s_g$. Putting the two inequalities together, we get

$$d^{125\log d} < s_g < t^{80t^2} \Rightarrow 125\log^2 d < 80t^2 \log t < 80t^2 \log^2 t \Rightarrow$$

$$\Rightarrow \log d < \frac{4}{5} t \log t \Rightarrow d < t^{\frac{4t}{5}} < p$$

where the last inequality is true by equation (15). □

Hashing the polynomials: Let P_g be the Newton polytope of the polynomial g and E_g be the set of exponents of g, as they are defined in section 3. Then, we have that $|E_g| = s_g > p^{c(\ell+2)}$.

Given the parameters above, Theorem 5.10 implies that there exists $\mathbf{h} \in \mathbb{F}_p^{\ell+2}$ for which the hyperplane

$$\mathcal{H} = \{\mathbf{x} \in \mathbb{F}_p^{\ell+2} \mid \mathbf{h} \cdot \mathbf{x} \equiv 0 \pmod p\}, \text{ is such that } |V(P_{\mathcal{H}})| \geq 2^{\ell} \cdot \frac{(2\ell)^t}{(5t)^t} \cdot \frac{Q}{3p}. \tag{19}$$

Notice that $d < p$ implies that $E_g = E_g^{(\mathsf{mod}\ p)}$ and therefore we can identify E_g with its image in $\mathbb{F}_p^n$. Since $E_g \subset \mathbb{F}_p^n$ and $p^{c(\ell+2)} < s_g = |E_g|$, for some constant $c > 3$, Corollary 2.3 tells us that there exists a linear transformation $A : \mathbb{F}_p^n \to \mathbb{F}_p^{\ell+2}$ such that $A(E_g) = \mathcal{H}$.

Let $\mathbf{y} = (y_1, \ldots, y_{\ell+2})$ be new variables, and let $D \in \mathbb{N}_0^{(\ell+2)\times n}$ be the $\ell \times n$ integer matrix such that $A(\mathbf{b}) \stackrel{p}{=} D\mathbf{b}$, for all $\mathbf{b} \in \mathbb{F}_p^n$. Notice that

$$D(E_g)^{(\mathsf{mod}\ p)} = A(E_g) = \mathcal{H}.$$

Let $\varphi^D : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{y}]$ be the transformation given by Lemma 4.13. Hence, by Lemma 4.13 we have that

$$\begin{aligned} E_{\varphi^D(g)} &= D(E_g)^{(\mathsf{mod}\ p)} = \mathcal{H} \quad \text{and} \\ \|\varphi^D(f)\|_0 &\geq \frac{\|\varphi(g) \cdot \varphi(h)\|_0}{2^{\ell+2}}. \end{aligned}$$

Therefore, we have that $P_{\varphi(g)} = P_{\mathcal{H}}$, which implies that

$$|V(P_{\varphi(g)})| = |V(P_{\mathcal{H}})| \geq 2^{\ell} \cdot \frac{(2\ell)^t}{(5t)^t} \cdot \frac{Q}{3p}. \tag{20}$$

Finishing up: We now have that

$$\begin{aligned} s_f = \|f(\mathbf{x})\|_0 &\geq \|\varphi^D(f)\|_0 && \text{(by Lemma 4.13)} \\ &\geq \frac{\|\varphi^D(g) \cdot \varphi^D(h)\|_0}{2^{\ell+2}} && \text{(by Lemma 4.13)} \\ &\geq \frac{|V(P_{\varphi(g)})|}{2^{\ell+2}} && \text{(by Corollary 3.17)} \\ &\geq \frac{(2\ell)^t}{(5t)^t} \cdot \frac{Q}{12p} && \text{(by equation (20))} \\ &\geq \frac{(2\ell)^t}{48t(5t)^t} \\ &= \frac{(6t)^t}{48t(5t)^t} = \frac{(6/5)^t}{48t} > \left(\frac{24}{23}\right)^t && \text{(for } t \geq 60\text{).} \end{aligned}$$

Therefore, by using the bound $t^{28t^2} > s_g$ from equation (16), we have

$$t^{28t^2} > s_g \Rightarrow 28t^2 \log t > \log s_g \Rightarrow t > \sqrt{\frac{\log s_g}{15 \log\log s_g}},$$

otherwise

$$\begin{aligned} t \leq \sqrt{\frac{\log s_g}{15 \log\log s_g}} &\Rightarrow 28t^2 \log t \leq 28 \frac{\log s_g}{15 \log\log s_g} \cdot \frac{1}{2} \cdot \log\left(\frac{\log s_g}{15 \log\log s_g}\right) \\ &< \frac{\log s_g}{\log\log s_g} \cdot \log\left(\frac{\log s_g}{\log\log s_g}\right) < \frac{\log s_g}{\log\log s_g} \cdot \log\log s_g = \log s_g. \end{aligned}$$

Hence, we obtain the following bound on s_g, in terms of s_f:

$$\begin{aligned} \frac{\log s_f}{\log(24/23)} > t > \sqrt{\frac{\log s_g}{15 \log\log s_g}} \\ \Rightarrow s_g < 2^{\gamma \log^2 s_f \log\log s_f}, \end{aligned}$$

where γ can be any constant such that $\gamma > \dfrac{900}{\log^2(24/23)}$. □

7 Deterministic sparse divisibility

Suppose we are given two polynomials $f(\mathbf{x})$ and $g(\mathbf{x})$, both given as a list of s_f (resp. s_g) monomials, and are asked whether or not there exists $h(\mathbf{x})$ so that $f(\mathbf{x}) = g(\mathbf{x}) \cdot h(\mathbf{x})$ and, if it exists, output it (also as a list of monomials). Using our main theorem we know that s_h is at most quasi-polynomial in s_f and d (the individual degree of f). We now sketch a way to find h in deterministic $\mathsf{poly}(s_f, s_g, s_h)$ time (which is $\mathsf{quasi\text{-}poly}(s_f, d)$ by our theorem). We can assume that h exists (that is, g divides f) since, if it does not, the final step in the algorithm will discard it (we will check the sparse identity $f = g \cdot h$).

We will use ideas from [KS01] who gave polynomial time identity testing and interpolation algorithms for sparse polynomials. The main tool is the following simple claim.

Claim 7.1. *Let d, n, s be integers and let $p > d$ be a prime. Then, there exists an explicit set $A \subset \{0, 1, \ldots, p-1\}^n$ of size $|A| \leq \mathsf{poly}(p, n, s)$ so that the following holds: For any set $E \subset \{0, 1, \ldots, d\}^n$ of size at most s there exists some $\mathbf{a} \in A$ so that the mapping $\mathbf{e} \mapsto \mathbf{e} \cdot \mathbf{a}$ (inner product over the integers) is one-to-one on the set E. By* explicit *we mean that there is a deterministic algorithm that, given d, s, n and p outputs the set A in polynomial time in d, s, n and p.*

Therefore, for a polynomial h with at most s monomials and individual degrees $\leq d$, there will be some $\mathbf{a} \in A$ so that the univariate polynomial $\hat{h}(T) = h(T^{a_1}, \ldots, T^{a_n})$ has the same coefficients of h. More formally, if

$$h(\mathbf{x}) = \sum_{\mathbf{e} \in E_h} c_{\mathbf{e}} \cdot \mathbf{x}^{\mathbf{e}},$$

Then,

$$\hat{h}(T) = \sum_{\mathbf{e} \in E_h} c_{\mathbf{e}} \cdot T^{\mathbf{e} \cdot \mathbf{a}},$$

where the last sum contains no cancellations (each power of T appears only once). This already gives a deterministic identity testing algorithm for sparse polynomials (compute $\hat{h}$ for all $\mathbf{a} \in A$ and check that all are zero, using a prime $p \leq O(d)$). But it also gives a way to deterministically interpolate a sparse polynomial h, given only black box access. The interpolation algorithm will try all $\mathbf{a}$'s in A and will succeed only for one of them. This is OK since we can discard a wrong answer by performing identity testing on the equality $h = h'$ for our candidate h' (since both are sparse, the identity can be tested in polynomial time deterministically). Hence, suppose we are given a 'good' $\mathbf{a} \in A$ so that the map $\mathbf{e} \mapsto \mathbf{e} \cdot \mathbf{a}$ is one-to-one on E_h. We can thus interpolate the univariate polynomial $\hat{h}(T)$ to obtain the set

$$S_0 = \{(c_{\mathbf{e}}, \mathbf{e} \cdot \mathbf{a}) \,|\, \mathbf{e} \in E_h\}.$$

That is, we already have the coefficients of h, alas without the matching exponents. To discover the exponents we do the following. Take γ to be some field element of order greater than d and consider the restricted polynomial $H_1(T) = h(\gamma T^{a_1}, T^{a_2}, \ldots, T^{a_n})$. Each coefficient of H_1 will be of the form $c_{\mathbf{e}} \gamma^{e_1} T^{\mathbf{e} \cdot \mathbf{a}}$. Hence, we can interpolate H_1 and obtain, using our previous knowledge of S_0, the new set

$$S_1 = \{(c_{\mathbf{e}}, \mathbf{e} \cdot \mathbf{a}, e_1) \,|\, \mathbf{e} \in E_h\}.$$

Continuing in this fashion we can recover the rest of the exponents.

This interpolation algorithm can be used to recover $h(\mathbf{x}) = f(\mathbf{x})/g(\mathbf{x})$ given the polynomials f and g in a similar way. To begin, we can compute $\hat{h}(T)$ by computing $\hat{f}(T)$ and dividing it by $\hat{g}(T)$. However, we should be careful to apply Claim 7.1 on the union $E = E_h \cup E_g$ (which is also quasi-polynomial) so that $\hat{g}$ does not vanish for the 'good' $\mathbf{a} \in A$. Similarly, we can calculate H_1 etc. and recover h from them.

Acknowledgments

The authors would like to thank Joachim von zur Gathen, Erich Kaltofen, Amir Shpilka and Ilya Volkovich for helpful conversations and comments.

References

[AV08] M. Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. In *Proceedings of the 49th Annual FOCS*, pages 67–75, 2008.

[CR88] Benny Chor and Ronald L. Rivest. A knapsack-type public key cryptosystem based on arithmetic in finite fields. *IEEE Transactions on Information Theory*, 34(5):901–909, 1988.

[DL78] R. A. DeMillo and R. J. Lipton. A probabilistic remark on algebraic program testing. *Inf. Process. Lett.*, 7(4):193–195, 1978.

[DSY09] Z. Dvir, A. Shpilka, and A. Yehudayoff. Hardness-randomness tradeoffs for bounded depth arithmetic circuits. *SIAM J. on Computing*, 39(4):1279–1293, 2009.

[FLMS13] Hervé Fournier, Nutan Limaye, Guillaume Malod, and Srikanth Srinivasan. Lower bounds for depth 4 formulas computing iterated matrix multiplication. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:100, 2013.

[GK85] J. Von Zur Gathen and E. Kaltofen. Factoring sparse multivariate polynomials. *Journal of Computer and System Sciences*, 31(2):265–287, 1985.

[GK98] D. Grigoriev and M. Karpinski. An exponential lower bound for depth 3 arithmetic circuits. In *Proceedings of the 30th Annual STOC*, pages 577–582, 1998.

[GKKS12] A. Gupta, P. Kamath, N. Kayal, and R. Saptharishi. An exponential lower bound for homogeneous depth four arithmetic circuits with bounded bottom fanin. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:98, 2012.

[GKKS13] A. Gupta, P. Kamath, N. Kayal, and R. Saptharishi. Arithmetic circuits: A chasm at depth three. *Electronic Colloquium on Computational Complexity (ECCC)*, 20(26), 2013.

[GR00] D. Grigoriev and A. A. Razborov. Exponential complexity lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. *Applicable Algebra in Engineering, Communication and Computing*, 10(6):465–487, 2000.

[GS06] V. Guruswami and M. Sudan. Improved decoding of reed-solomon and algebraic-geometry codes. *IEEE Trans. Inf. Theor.*, 45(6):1757–1767, September 2006.

[Kal85] E. Kaltofen. Polynomial-time reductions from multivariate to bi- and univariate integral polynomial factorization. *SIAM J. on computing*, 14(2):469–489, 1985.

[Kal89] E. Kaltofen. Factorization of polynomials given by straight-line programs. In S. Micali, editor, *Randomness in Computation*, volume 5 of *Advances in Computing Research*, pages 375–412. 1989.

[Kal03] E. Kaltofen. Polynomial factorization: a success story. In *ISSAC*, pages 3–4, 2003.

[KI04] V. Kabanets and R. Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Computational Complexity*, 13(1-2):1–46, 2004.

[KLSS14] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. An exponential lower bound for homogeneous depth four arithmetic formulas. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:5, 2014.

[KS01] A. Klivans and D. Spielman. Randomness efficient identity testing of multivariate polynomials. In *Proceedings of the 33rd Annual STOC*, pages 216–223, 2001.

[KS13] Mrinal Kumar and Shubhangi Saraf. Superpolynomial lower bounds for general homogeneous depth 4 arithmetic circuits. *CoRR*, abs/1312.5978, 2013.

[LLL82] A. K. Lenstra, H. W. Lenstra, and L. Lovász. Factoring polynomials with rational coefficients. *Mathematische Annalen*, 261(4):515–534, 1982.

[RS62] J. B. Rosser and L. Schoenfeld. Approximate formulas for some functions of prime numbers. *Illinois Journal of Mathematics*, 6(1):64–94, 03 1962.

[RY09] R. Raz and A. Yehudayoff. Lower bounds and separations for constant depth multilinear circuits. *Computational Complexity*, 18(2):171–207, 2009.

[Sch80] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4):701–717, 1980.

[Sch00] A. Schinzel. *Polynomials with special regard to reducibility*, volume 77. Cambridge University Press, 2000.

[SSS13] Chandan Saha, Ramprasad Saptharishi, and Nitin Saxena. A case of depth-3 identity testing, sparse factorization and duality. *Computational Complexity*, 22(1):39–69, 2013.

[Sud97] Madhu Sudan. Decoding of reed solomon codes beyond the error-correction bound. *Journal of Complexity*, 13(1):180 – 193, 1997.

[SW01] Amir Shpilka and Avi Wigderson. Depth-3 arithmetic circuits over fields of characteristic zero. *Computational Complexity*, 10(1):1–27, 2001.

[SW05] Igor Shparlinski and Arne Winterhof. Noisy interpolation of sparse polynomials in finite fields. *Applicable Algebra in Engineering, Communication and Computing*, 16(5):307–317, 2005.

[SY10] A. Shpilka and A. Yehudayoff. Arithmetic circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science*, 5(3-4):207–388, 2010.

[Vad12] S. P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1-3):1–336, 2012.

[Val79a] L. G. Valiant. Completeness classes in algebra. In *Proceedings of the 11th Annual STOC*, pages 249–261, 1979.

[Val79b] L. G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8(2):189–201, 1979.

[Zie95] Günter M Ziegler. *Lectures on polytopes*, volume 152. Springer, 1995.

[Zip79] R. Zippel. Probabilistic algorithms for sparse polynomials. In *EUROSAM*, pages 216–226, 1979.

A Proof of Lemma 2.1:

We begin with some definitions.

Definition A.1 (Primorial Function)**.** *The* primorial function $\vartheta : \mathbb{R} \to \mathbb{N}_0$ *is defined as follows:*

$$\vartheta(x) = \prod_{\substack{p \leq x \\ p \text{ is prime}}} p$$

That is, $\vartheta(x)$ *is the product of all prime numbers that are less than or equal to* x*.*

Definition A.2 (Prime Function)**.** *The* prime function $\pi : \mathbb{R} \to \mathbb{N}_0$ *is defined as follows:*

$$\pi(x) = \sum_{\substack{p \leq x \\ p \text{ is prime}}} 1$$

That is, $\pi(x)$ *is the number of primes that are less than or equal to* x*.*

In [RS62, Eq. (3.4), (3.5), (3.15) and (3.16)], Rosser and Schoenfeld gave the following bounds on the primorial and the prime functions:

Lemma A.3 (Bounds on ϑ and π, [RS62])**.** *The following bounds on* $\vartheta(x)$ *and* $\pi(x)$ *hold, when* $x \geq 41$*:*

$$\exp\left(x\left(1 - \frac{1}{\ln x}\right)\right) < \vartheta(x) < \exp\left(x\left(1 + \frac{1}{2\ln x}\right)\right)$$
$$\frac{x}{\ln x} < \pi(x) < \frac{x}{\ln x - \frac{3}{2}}.$$

Proof of Lemma 2.1. Notice that the following inequalities hold, when $t \geq 50$:

$$\frac{t\ln t}{\ln t + \ln\ln t - \frac{3}{2}} < t + 2 < \frac{2t\ln t}{\ln(2t\ln t)} \tag{21}$$

Hence, by (21) and the bounds on π given in Lemma A.3, for $t \geq 50$, we have

$$\pi(t\ln t) < \frac{t\ln t}{\ln t + \ln\ln t - \frac{3}{2}} < t + 2$$

and

$$\pi(2t\ln t) > \frac{2t\ln t}{\ln(2t\ln t)} > t + 2.$$

Thus, by the above inequalities and the bounds on ϑ in Lemma A.3, we obtain

$$\exp\left(t\ln t\left(1 - \frac{1}{\ln(t\ln t)}\right)\right) < \vartheta(t\ln t) < Q_t < \vartheta(2t\ln t) < \exp\left(2t\ln t\left(1 + \frac{1}{2\ln(2t\ln t)}\right)\right).$$

Because $\ln(t \ln t) > 5$, for $t \geq 50$, we have

$$t^{\frac{4t}{5}} = \exp\left(\frac{4t \ln t}{5}\right) = \exp\left(t \ln t(1 - 1/5)\right) < \exp\left(t \ln t \left(1 - \frac{1}{\ln(t \ln t)}\right)\right)$$

and

$$\exp\left(2t \ln t \left(1 + \frac{1}{2 \ln(2t \ln t)}\right)\right) < \exp\left(2t \ln t \left(1 + \frac{1}{10}\right)\right) = \exp\left(\frac{11t \ln t}{5}\right) < t^{\frac{11t}{5}}.$$

□

B Proof of Lemma 2.2

Before proving the lemma we prove an auxiliary claim. Recall that a family F of maps from D to T is pairwise independent if, for any pair $(a, b) \in D \times D$ with $a \neq b$, we have that $(f(a), f(b))$ is distributed uniformly over $T \times T$ when f is chosen uniformly in F.

Claim B.1. *For every prime p and every $n \in \mathbb{N}_0$ such that $n > 0$, there exists a family of pairwise independent affine maps*

$$\mathcal{F} = \{\psi : \mathbb{F}_p^n \to \mathbb{F}_p^n\}$$

such that $|\mathcal{F}| = p^{2n}$.

Proof. Let $q = p^n$. It is well known (see e.g. [Vad12, Proposition 3.24]) that the family of functions

$$\mathcal{F}' = \{\varphi_{a,b} : \mathbb{F}_q \to \mathbb{F}_q \mid \varphi_{a,b}(x) = ax + b\}_{a,b \in \mathbb{F}_q},$$

is pairwise independent over any finite field $\mathbb{F}_q$. Through the natural isomorphism between $\mathbb{F}_q$ and $\mathbb{F}_p^n$, we have that each affine map $\varphi_{a,b} \in \mathcal{F}'$ corresponds to an affine map $\psi_{a,b} : \mathbb{F}_p^n \to \mathbb{F}_p^n$. Thus, the family of affine maps given by

$$\mathcal{F} = \{\psi_{a,b} : \mathbb{F}_p^n \to \mathbb{F}_p^n \mid \psi_{a,b}(\mathbf{x}) \text{ corresponds to } \varphi_{a,b}(\mathbf{x}) \in \mathcal{F}'\}$$

is pairwise independent. Moreover, notice that the size of $\mathcal{F}$ is $|\mathcal{F}| = |\mathcal{F}'| = q^2 = p^{2n}$. □

Proof of Lemma 2.2. Let $A : \mathbb{F}_p^n \to \mathbb{F}_p^n$ be a random affine map from the pairwise independent family $\mathcal{F}$ given by Claim B.1 and let $\pi : \mathbb{F}_p^n \to \mathbb{F}_p^\ell$ be the projection operator, where we project onto the first ℓ coordinates of the input. Notice that $|\mathcal{F}| = p^{2n}$. Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k \in \mathbb{F}_p^\ell$ be distinct points of $\mathbb{F}_p^\ell$. For each $j \in \{1, \ldots, k\}$, let $M_j(A) = |\{\mathbf{x} \in S \mid \pi(A(\mathbf{x})) = \mathbf{a}_j\}|$. Then, we have that

$$\begin{aligned}
\mathbb{E}_A[M_j(A)] &= \sum_{\mathbf{x} \in S} \mathbf{Pr}_A[\pi(A(\mathbf{x})) = \mathbf{a}_j] \\
&= \sum_{\mathbf{x} \in S} \frac{1}{p^\ell} = \frac{|S|}{p^\ell} = \frac{s}{p^\ell} = p^{(c-1)\ell}
\end{aligned}$$

Now, let us bound the variance of each $M_j(A)$. Notice that we can write

$$M_j(A) = \sum_{\mathbf{x}\in S} 1_{\{\pi(A(\mathbf{x}))=\mathbf{a}_j\}}, \text{ where } 1_{\{\pi(A(\mathbf{x}))=\mathbf{a}_j\}} = \begin{cases} 1, \text{ if } \pi(A(\mathbf{x})) = \mathbf{a}_j \\ 0, \text{ otherwise.} \end{cases}$$

By pairwise independence of the indicator variables $1_{\{\pi(A(\mathbf{x}))=\mathbf{a}_j\}}$, since the family of maps $\mathcal{F}$ is pairwise independent, we have that

$$\mathbb{V}_A[M_j(A)] = \sum_{\mathbf{x}\in S} \mathbb{V}_A[1_{\{\pi(A(\mathbf{x}))=\mathbf{a}_j\}}].$$

Since $1_{\{\pi(A(\mathbf{x}))=\mathbf{a}_j\}} = \begin{cases} 1, \text{ with probability } 1/p^\ell \\ 0, \text{ with probability } 1-1/p^\ell, \end{cases}$ we have that

$$\mathbb{V}_A[1_{\{\pi(A(\mathbf{x}))=\mathbf{a}_j\}}] = \frac{1}{p^\ell}(1-\frac{1}{p^\ell}) < \frac{1}{p^\ell}.$$

Hence, we conclude that

$$\mathbb{V}_A[M_j(A)] < \frac{|S|}{p^\ell} = \frac{s}{p^\ell}.$$

Thus, by Chebyshev's inequality, we obtain:

$$\mathbf{Pr}_A[|M_j(A) - \mathbb{E}_A[M_j(A)]| \geq t] \leq \frac{\mathbb{V}_A[M_j(A)]}{t^2} < \frac{s}{t^2 p^\ell} = \frac{p^{(c-1)\ell}}{t^2}. \tag{22}$$

Setting $t = p^{(c-3/2)\ell}$, we have that $\mathbf{Pr}_A[|M_j(A) - \mathbb{E}_A[M_j(A)]| \geq t] < \frac{1}{p^{(c-2)\ell}}$. Therefore, we have that the probability that $M_j(A) > 0$ for all $j \in [k]$ can be lower bounded by:

$$\begin{aligned}
\mathbf{Pr}_A[M_j(A) > 0, \forall j \in [k]] &= 1 - \mathbf{Pr}_A[\exists j \in [k] \text{ s.t. } M_j(A) = 0] \\
&\geq 1 - \sum_{i=1}^{k} \mathbf{Pr}_A[M_j(A) = 0] && \text{(by the union bound)} \\
&\geq 1 - \sum_{i=1}^{k} \mathbf{Pr}_A[|M_j(A) - \mathbb{E}_A[M_j(A)]| \geq p^{(c-3/2)\ell}] && \text{(since } \mathbb{E}_A[M_j(A)] = p^{(c-1)\ell}) \\
&> 1 - \sum_{i=1}^{k} \frac{1}{p^{(c-2)\ell}} = 1 - \frac{k}{p^{(c-2)\ell}}.
\end{aligned}$$

Hence, if we set $k = p^\ell$ (and therefore the points $\mathbf{a}_j$ will correspond to all points of $\mathbb{F}_p^\ell$) we have that $\mathbf{Pr}_A[M_j(A) > 0, \forall j \in [k]] > 1 - \frac{p^\ell}{p^{(c-2)\ell}} = 1 - \frac{1}{p^{(c-3)\ell}} > 0$, since $c > 3$. This shows that there exists a map $L' : \mathbb{F}_p^n \to \mathbb{F}_p^\ell$ defined by $L'(\mathbf{x}) = \pi(A(\mathbf{x}))$ that is surjective. Since maps A and π are affine, we have that L' is also affine. Hence, there exist a matrix $Q \in \mathbb{F}_p^{\ell\times n}$ and a vector $\mathbf{b} \in \mathbb{F}_p^\ell$ such that $L'(\mathbf{x}) = Q\mathbf{x} + \mathbf{b}$, for all $\mathbf{x} \in \mathbb{F}_p^\ell$. Now, notice that L' is surjective implies that the map $L : \mathbb{F}_p^n \to \mathbb{F}_p^\ell$ defined by $L(\mathbf{x}) = Q\mathbf{x} = L'(\mathbf{x}) - \mathbf{b}$ is also surjective. Since the map L is clearly linear, this proves the lemma. □

ECCC ISSN 1433-8092
http://eccc.hpi-web.de