On Multilinear Forms: Bias, Correlation, and Tensor Rank

Abhishek Bhrushundi* Prahladh Harsha† Pooya Hatami‡ Swastik Kopparty§
Mrinal Kumar¶

Abstract

In this paper, we prove new relations between the bias of multilinear forms, the correlation between multilinear forms and lower degree polynomials, and the rank of tensors over $\mathbb{F}_2 = \{0,1\}$. We show the following results for multilinear forms and tensors.

Correlation bounds. We show that a random d-linear form has exponentially low correlation with low-degree polynomials. More precisely, for $d \ll 2^{o(k)}$, we show that a random d-linear form $f(X_1, X_2, \ldots, X_d) : \left(\mathbb{F}_2^k\right)^d \to \mathbb{F}_2$ has correlation $2^{-k(1-o(1))}$ with any polynomial of degree at most $d/2$.

This result is proved by giving near-optimal bounds on the bias of a random d-linear form, which is in turn proved by giving near-optimal bounds on the probability that a sum of t random d-dimensional rank-1 tensors is identically zero.

Tensor-rank vs Bias. We show that if a d-dimensional tensor has small rank, then the bias of the associated d-linear form is large. More precisely, given any d-dimensional tensor

$$T : \underbrace{[k] \times \ldots [k]}_{d \text{ times}} \to \mathbb{F}_2$$

of rank at most t, the bias of the associated d-linear form

$$f_T(X_1, \ldots, X_d) := \sum_{(i_1,\ldots,i_d) \in [k]^d} T(i_1, i_2, \ldots, i_d) X_{1,i_1} \cdot X_{1,i_2} \cdots X_{d,i_d}$$

is at least $\left(1 - \frac{1}{2^{d-1}}\right)^t$.

The above bias vs tensor-rank connection suggests a natural approach to proving nontrivial tensor-rank lower bounds for $d = 3$. In particular, we use this approach to give a new proof that the finite field multiplication tensor has tensor rank at least $3.52k$, which is the best known rank lower bound for any explicit tensor in three dimensions over $\mathbb{F}_2$.

*Dept. of Computer Science, Rutgers University, U.S.A. abhishek.bhr@gmail.com.

†Tata Institute of Fundamental Research, India. prahladh@tifr.res.in. This work was done when the author was visiting Rutgers University/DIMACS, USA and Weizmann Institute of Science, Israel. This work was partially supported by the DIMACS/Simons Collaboration in Cryptography through NSF grant #CNS-1523467 and the Israel-India ISF-UGC grant.

‡Dept. of Computer Science, University of Texas at Austin, U.S.A. pooyahat@gmail.com. Part of this work was done when the author was a postdoc at DIMACS. Supported by a Simons Investigator Award (#409864, David Zuckerman)

§Dept. of Computer Science & Dept. of Mathematics, Rutgers University, U.S.A. Research supported in part by NSF grants CCF-1253886 and CCF-1540634. swastik.kopparty@gmail.com.

¶Center for Mathematical Sciences and Applications, Harvard University, U.S.A. mrinalkumar08@gmail.com.

1 Introduction

This work is motivated by two fundamental questions regarding "explicit constructions" in complexity theory: finding functions uncorrelated with low degree polynomials, and finding tensors with high tensor rank.

Functions uncorrelated with low degree polynomials. The first question is that of finding an explicit function uncorrelated with low degree polynomials. More concretely, we seek functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$ such that for every polynomial $P(X_1, \ldots, X_n) \in \mathbb{F}_2[X_1, \ldots, X_n]$ of degree at most ℓ (assume $\ell \approx n^{0.1}$ say),

$$\Pr_{x \in \mathbb{F}_2^n}[f(x) = P(x)] \leq \frac{1}{2} + \varepsilon_n.$$

It is well known (and easy to prove) that a random function f has this property with ε_n super-polynomially small (and even exponentially small); the challenge is to find an explicit function f.

A solution to this problem will have immediate applications in Boolean circuit complexity. It will give hard-on-average problems for $AC^0(\oplus)$, and via the Nisan-Wigderson hardness vs. randomness technique [NW94], it will give pseudorandom generators against $AC^0(\oplus)$ (improving upon analogous results for AC^0 from the late 1980s). The original motivation for an explicit function with small ε_n came from the seminal work of Razborov [Raz87] and Smolensky [Smo87] who showed that any function computable by a sub-exponential sized $AC^0(\oplus)$ circuit satisfies $\varepsilon_n = \Omega(1)$ and furthermore that the MOD_3 has $\varepsilon_n = O(1)$. The Nisan-Wigderson paradigm [NW94] of pseudorandom generator construction requires explicit functions with exponentially small ε_n. The current best known constructions of explicit functions [Raz87, Smo87, BK12, VW08] that cannot be approximated by low-degree polynomials come in two flavors, (a) polynomially small ε_n (in fact, $O(1/\sqrt{n})$) for large degree bounds (d as large as $n^{0.1}$) or (b) exponentially small ε_n for small degree bounds ($d << \log n$). However, we do not know of any explicit function f that exhibits exponentially small ε_n against low-degree polynomials of polynomially large (or even super-logarithmically large) degree polynomials. For a nice survey on correlation with low degree polynomials, see [Vio09].

Tensors with high rank. The second question is that of finding an explicit tensor of high tensor rank. Tensors are a high-dimensional generalization of (2-dimensional) matrices. Just as a matrix of size k over a field $\mathbb{F}$ is given by a map $M : [k]^2 \to \mathbb{F}$, a tensor T of dimension d and size k is given by a map $T : [k]^d \to \mathbb{F}$. A tensor T is said to be of rank one if there exist vectors $u_1, u_2, \ldots, u_d \in \mathbb{F}_2^k$ such that $T = u_1 \otimes u_2 \otimes \cdots \otimes u_d$ or equivalently, for all $(i_1, \ldots, i_d) \in [k]^d$, we have $T(i_1, \ldots, i_d) = u_{1,i_1} \cdot u_{2,i_2} \cdots u_{d,i_d}$. A tensor T is said to be of tensor-rank at most t if it can be written as the sum of t rank one tensors. We seek tensors with tensor-rank as high as possible.

It is well known (and easy to prove) that a random tensor T has tensor rank t as large as $\Omega(k^{d-1}/d)$. The challenge is to find an explicit such T with tensor rank larger than $k^{\lfloor \frac{d}{2} \rfloor}$. A substantial improvement on this lower bound for any explicit tensor will have immediate applications in arithmetic circuit complexity; for $d = 3$, it will give improved arithmetic circuit lower bounds [Str73], and for large d it will give superpolynomial arithmetic formula lower bounds [Raz13, CKSV16]. For general *odd* d, a lower bound of $2k^{\lfloor d/2 \rfloor} + k - O(d \log k)$ was shown

for an explicit tensor by Alexeev et al. [AFT11], while for *even* d, no lower bounds better than the trivial bound $k^{\lfloor \frac{d}{2} \rfloor}$ are known for any explicit tensor.

Unlike matrix rank, we do not have a good understanding of tensor-rank even for 3-dimensional tensors. For instance, it is known that for a given 3-dimensional tensor T over the rationals, the problem of deciding if the rank of T is at most k is NP-hard [Hås90]. In the case of dimension three, the tensor-rank of very specific tensors like the matrix multiplication tensor [Blä99, Shp03], the finite field multiplication tensor [CC88, STV92] and the polynomial multiplication tensor [BD80, Kam05] has been studied in prior works. For this case, the current best lower bound known for any explicit tensor over $\mathbb{F}_2$ is a lower bound of $3.52k$ for the finite field multiplication tensor due to Chudnovsky and Chudnovsky [CC88, STV92], which builds on the lower bound result of Brown and Dobkin [BD80] for the polynomial multiplication tensor. For general fields, the best known lower bound for any explicit tensor is $2.5k - o(k)$ for the matrix multiplication tensor due to Bläser [Blä99].

Also relevant to this discussion is a recent result of Effremenko et al. [EGOW18], who showed that a fairly general class of lower bound techniques called *rank methods* are not strong enough to give lower bounds on tensor rank stronger than $2^d \cdot k^{\lfloor d/2 \rfloor}$. In a nutshell, not only can we not prove good tensor rank lower bounds, we do not even have techniques, which 'in principle' could be useful for such lower bounds!

1.1 Our results

We make contributions to both the above questions by studying *multilinear forms* and their *bias*. A d-linear form is a map $f : (\mathbb{F}_2^k)^d \to \mathbb{F}_2$ which is linear in each of its arguments. The *bias* of a d-linear form is defined as follows.

$$\mathrm{bias}(f) := \left| \mathbb{E}_{x_1,\ldots,x_d \in \mathbb{F}_2^k} [(-1)^{f(x_1,\ldots,x_k)}] \right| .$$

This measures the difference between the probability of output 1 and output 0. Similarly, the correlation of a d-linear form f with another function g is defined as $\mathrm{Corr}(f,g) := \mathrm{bias}(f-g)$, which measures the difference between the probabilities (on a random input) that f and g agree and disagree.

A d-linear form f can naturally be viewed as a polynomial of degree d in $n = kd$ variables. We can then ask, for some $\ell \ll d$, is there a d-linear form f such that the correlation of f with every degree ℓ polynomial in $\mathbb{F}_2[X_1,\ldots,X_n]$ is small? Knowing the existence of a d-linear f that achieves this small correlation property gives a significantly reduced search space for finding an explicit f with small correlation with lower degree polynomials. Our first result gives a positive answer to this question for a large range of ℓ and d.

Theorem A (informal). *Let $d \ll o(n/\log n)$ and let $k = \frac{n}{d}$. Let $\ell < d/2$. Then with high probability, for a uniformly random d-linear form $f : (\mathbb{F}_2^k)^d \to \mathbb{F}_2$, we have that for all polynomial $P(X_1,\ldots,X_n) \in \mathbb{F}_2[X_1,\ldots,X_n]$ of degree at most ℓ:*

$$\mathrm{Corr}(f,P) \leq 2^{-k(1-o(1))} = 2^{-\frac{n}{d}(1-o(1))}.$$

Moreover, for every d-linear form, there is a degree 0 polynomial P (namely the constant 0 polynomial) such that

$$\mathrm{Corr}(f,P) \geq \Omega(2^{-k}).$$

For d small enough ($\tilde{O}(\log n)$), the above theorem actually holds with $\ell = d - 1$.

An important step towards proving Theorem A is a precise understanding of the distribution of the *bias* of a random d-linear form. Along the way, we give tight upper bounds on the probability that the sum of t random *rank-1* d-dimensional tensors equals 0.

Previously, a beautiful result of Ben-Eliezer, Lovett and Hod [BHL12] showed that for all $d < \alpha n$, there are polynomials $f(X_1, \ldots, X_n)$ of degree d whose correlation with polynomials of degree $\ell = d - 1$ is $2^{-\tilde{\Omega}(n/d)}$. The results are incomparable; the f in [BHL12] need not come from a d-linear form, and for this more general setting the bound $2^{-\Omega(n/d)}$ might not be tight, but on the positive side [BHL12] can handle larger d while proving correlation bounds against polynomials with degree as large as $d - 1$.

A d-linear form f can also be naturally viewed as a d-dimensional tensor. Indeed, f can be completely specified by the tensor T of values $f(e_{i_1}, e_{i_2}, \ldots, e_{i_d})$, as the i_j vary in $[k]$. We can then ask, are there natural properties of the d-linear form f which would imply that the tensor rank of T is high?

We show that having low bias, which is a simple measure of pseudorandomness for d-linear forms, already implies something nontrivial about the tensor rank. We prove a lower bound on the tensor rank in terms of the bias of the form.

Theorem B. *Let $f : (\mathbb{F}_2^k)^d \to \mathbb{F}_2$ be a d-linear form. Let T be its associated tensor, and let t be the rank of T. Then*

$$\mathrm{bias}(f) \geq \left(1 - \frac{1}{2^{d-1}}\right)^t.$$

In particular, if $\mathrm{bias}(f) = 2^{-(1-o(1))k}$, then

$$t \geq k \cdot \log_2 \frac{2^{d-1}}{2^{d-1} - 1}.$$

Moreover, for every t there is a tensor T with tensor rank t such that the following is true.

$$\mathrm{bias}(f) \leq \left(1 - \frac{1}{2^{d-1}}\right)^t + \frac{d}{2^k}.$$

This lower bound on tensor rank in terms of bias is almost optimal for any fixed d. It implies that any explicit d-linear form with low bias (such d-linear forms are easy to construct) automatically must have tensor rank $(1 + \Omega(1)) \cdot k$. Purely from the point of view of proving tensor rank lower bounds for explicit tensors, these results are only interesting in the case of $d = 3$ (for larger d the implied tensor rank lower bounds fail to beat trivial explicit tensor rank lower bounds).

For $d = 3$, this gives a natural and clean route to proving nontrivial tensor rank lower bounds for explicit tensors. In particular, trilinear forms with nearly minimal bias of of $2^{-(1-o(1))k}$ must have tensor rank at least $2.409k$ (which happens to be tight). A finer analysis of our arguments shows that trilinear forms with *exactly* minimal bias of $\approx 2 \cdot 2^{-k}$, such as the finite field multiplication tensor, have tensor rank $\geq 3.52k$, thus matching the best known explicit tensor rank lower bound for 3-dimensional tensors [BD80, CC88, STV92]. It also immediately implies that the matrix multiplication tensor has tensor rank $\geq 1.8k$, which is nontrivial (but still far from the best known bound of $3k$ [Shp03, Blä99]).

1.2 Methods

Underlying our main results, Theorem A and Theorem B, are two related combinatorial bounds involving rank-t d-linear forms. We now state these bounds for the special case of $d = 3$. For $i \in [t]$, let $x_i, y_i, z_i \in \mathbb{F}_2^k$. Let $P_i(u, v, w)$ be the trilinear form defined as

$$P_i(u, v, w) = \langle u, x_i \rangle \cdot \langle v, y_i \rangle \cdot \langle w, z_i \rangle.$$

Now, consider the trilinear form $P(u, v, w)$ given by

$$P(u, v, w) = \sum_{i=1}^{t} P_i(u, v, w).$$

Then, we have the following.

1. If x_i, y_i, z_i are picked uniformly at random from $\mathbb{F}_2^k$, then the probability that P is identically 0 is very small. Concretely,
$$\Pr_{x_i, y_i, z_i}[P \equiv 0]$$
is about 2^{-kt}, provided $t \ll k^2$. This bound is essentially optimal.

2. For arbitrary x_i, y_i, z_i, the bias of P is large. Concretely,
$$\min_{x_i, y_i, z_i} [\text{bias}(P)] \geq (3/4)^t.$$
This bound is also essentially optimal.

We now give an outline of the proofs of Theorem A and Theorem B.

The proof of Theorem A follows the high-level outline of [BHL12]. We first use the method of moments to show that for a fixed n-variate polynomial P of degree ℓ, the correlation of a random d-linear f with P is small with extremely high probability. Then, by a union bound over all P, we conclude that a random f is uncorrelated with all P with quite high probability.

Implementing this approach gives rise to some natural and interesting questions about rank-1 tensors. How many rank-1 tensors can lie in a given low dimensional linear space of tensors? Given a collection of t random rank-1 tensors, what is the probability that the dimension of the space spanned by them is small? What is the probability that the sum of t random rank-1 tensors equals 0? We investigate these questions using linear-algebraic ideas, and obtain near-optimal answers for all of them.

For example, the $d = 3$ case requires us to study the probability that

$$\sum_{i=1}^{t} x_i \otimes y_i \otimes z_i = 0.$$

By some simple manipulations, this reduces to bounding the probability that the linear space of matrices

$$span\{x_i \otimes y_i : i \in [t]\}$$

has dimension $\leq t - r$. We bound this by studying the probability that $x_i \otimes y_i$ lies in the linear space

$$span\{x_j \otimes y_j : j \in [i-1]\}.$$

This final probability is bounded using the following general theorem.

Lemma. *For any linear space $U \subseteq \mathbb{F}_2^{k^2}$ of dimension $u \ll k^2$, the probability that $x \otimes y \in U$ is at most $\tilde{O}\left(\frac{2^{u/k}}{2^k}\right)$.*

The proof of this lemma is hands on, and uses basic linear algebra and some elementary analytic inequalities. The key is to take an echelon form basis for U. We use this basis to understand which $\tilde{x} \in \mathbb{F}_2^k$ are "important"; i.e., they have the property that $\tilde{x} \otimes y \in U$ with noticeable probability for a random y.

The above lemma is essentially tight: with $U = V \otimes \mathbb{F}_2^k$ and $\mathbb{F}_2^k \otimes V$ being tight examples. The sets of the important $\tilde{x}$ in these two examples look very different. Because of this, our final proof involves proving tight upper bounds on an analytic maximization problem that has multiple very different global maxima.

For Theorem B, which gives a relationship between tensor rank and bias, the proof proceeds in the contrapositive. We show that any d-linear form whose underlying tensor has low rank must have high bias. Let us illustrate the underlying ideas in the case of $d = 3$. Here, we are given the 3-linear form P, defined as

$$P(u, v, w) = \sum_{i=1}^{t} \langle x_i, u \rangle \cdot \langle y_i, v \rangle \cdot \langle z_i, w \rangle.$$

We want to show that this has high bias if t is small. The key claim that we show is the following.

Lemma. *Let $y_1, \ldots, y_t, z_1, \ldots, z_t \in \mathbb{F}_2^t$. For at least $(3/4)^t$ fraction of the pairs $(v, w) \in \mathbb{F}_2^t$, we have that for all $i \in [t]$:*

$$\langle v, y_i \rangle \cdot \langle w, z_i \rangle = 0.$$

For any fixed i, the set of (v, w) satisfying the above is the union of two codimension 1 hyperplanes in $\mathbb{F}_2^{2t}$, and thus a random (v, w) satisfies it with probability $3/4$. The above lemma shows that the probability of all these events happening together is at least as large as it would have been had they been independent.

2 Preliminaries

Unless otherwise stated, we always work over the field $\mathbb{F}_2$. We use capital X, Y, Z etc. to denote formal variables or sets of formal variables, and small letters x, y, z to denote instantiations of these formal variables.

For integers $n, d \geq 0$, denote by $\mathrm{Poly}(n, d)$ the set of all degree $\leq d$ multilinear polynomials in $\mathbb{F}_2[X]$, where $X = \{X_1, ..., X_n\}$ is a variable set. Note that every $f \in \mathrm{Poly}(n, d)$ naturally corresponds to a unique map $f : \mathbb{F}_2^n \to \mathbb{F}_2$.

2.1 Bias and Correlation

Two fundamental notions used in this paper are those of bias and correlation, which we now define.

Definition 2.1 (Bias). *Bias of a function $f : \mathbb{F}_2^n \to \{0,1\}$ is defined as*

$$\mathrm{bias}(f) := \left| \mathbb{E}_{x \in \mathbb{F}_2^n} (-1)^{f(x)} \right| .$$

The bias of an $\mathbb{F}_2$-valued function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is defined as $\mathrm{bias}(f) := \mathrm{bias}(\iota(f))$*, where ι is the standard map from $\mathbb{F}_2$ to $\{0,1\}$.*

Definition 2.2 (Correlation). *We define the correlation between two functions $f, g : \mathbb{F}_2^n \to \mathbb{F}_2$, by*

$$\mathrm{Corr}(f,g) := \mathrm{bias}(f-g) .$$

Given a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$, we will be interested in its maximum correlation with low degree polynomials. Towards this we define

$$\mathrm{Corr}(f,d) := \max_{g \in \mathrm{Poly}(n,d)} \mathrm{Corr}(f,g) .$$

More generally, given a class $\mathcal{C}$ of functions, we define

$$\mathrm{Corr}(f,\mathcal{C}) := \max_{g \in \mathcal{C}} \mathrm{Corr}(f,g) .$$

2.2 Tensors and d-linear forms

Tensors are generalizations of matrices to higher dimensions.

Definition 2.3 (Tensors and Tensor rank). *Let k and d be natural numbers. A d dimensional tensor T of size k over a field $\mathbb{F}$ is a map $T : [k]^d \to \mathbb{F}$. T is said to be of rank one if there exist d vectors $u_1, u_2, \ldots, u_d : [k] \to \mathbb{F}$ such that for every $(i_1, i_2, \ldots, i_d) \in [k]^d$, $T(i_1, i_2, \ldots, i_d) = \prod_{j=1}^d u_j(i_j)$. The rank of T is the minimum t such that T can be written as a sum of t rank one tensors.*

Every matrix can be naturally associated with a bilinear polynomial, and in some cases, one can study the properties of this bilinear polynomial as a proxy of studying various properties of the matrix itself. This paradigm also generalizes to tensors, as the following definition indicates.

Definition 2.4 (Tensors as Multilinear Forms). *Let $T : [k]^d \to \mathbb{F}$ be a d dimensional tensor. Then, the* set-multilinear *polynomial associated with T is the polynomial f_T in variables $\left\{ X_{i,j} : i \in [d], j \in [k] \right\}$ over $\mathbb{F}$ defined as follows.*

$$f_T(X_{1,1}, X_{1,2}, \ldots, X_{d,k}) = \sum_{(i_1, i_2, \ldots, i_d) \in [k]^d} T(i_1, i_2, \ldots, i_d) \cdot \prod_{j=1}^{d} X_{j,i_j}.$$

Given the above association between d-dimensional tensors and d-linear forms, we will use the terms tensor and d-linear form interchangeably.

2.3 Some explicit tensors

We now define some explicit tensors which we use at various places in this paper. We start with the trace function.

2.3.1 Trace tensor

Definition 2.5. Trace : $\mathbb{F}_{2^k} \to \mathbb{F}_2$ *is the* $\mathbb{F}_2$*-linear map defined as follows.*

$$\text{Trace}(\alpha) = \alpha + \alpha^2 + \ldots + \alpha^{2^{k-1}}.$$

The Trace map will be useful for us as we define the candidate hard tensor for our lower bounds.

Definition 2.6. *Let* $Tr : \mathbb{F}_2^{k \times k \times k} \to \mathbb{F}_2$ *be the function defined as follows.*

$$Tr(X, Y, Z) := \text{Trace}(XYZ),$$

where XYZ *denotes multiplication over the larger field* $\mathbb{F}_{2^k}$ *when* $X = (X_1, X_2, \ldots, X_k), Y = (Y_1, Y_2, \ldots, Y_k), Z = (Z_1, Z_2, \ldots Z_k)$ *are viewed as encodings of elements in* $\mathbb{F}_{2^k}$*.*

Since Trace is an $\mathbb{F}_2$-linear map, the function $Tr(X, Y, Z)$ can be viewed as a 3-linear polynomial in the variables $X = (X_1, X_2, \ldots, X_k), Y = (Y_1, Y_2, \ldots, Y_k), Z = (Z_1, Z_2, \ldots Z_k)$. For the rest of this paper, when we say $Tr(X, Y, Z)$, we refer to this natural 3-linear polynomial and the three dimensional tensor associated with it. Up to a change of basis, this is the finite field multiplication tensor, which was analyzed by Chudnovsky-Chudnovsky [CC88] and Shparlinksi-Tsfasman-Vladut [STV92]. It is also worth noting that these papers also proved a surprising and beautiful $O(k)$ *upper bound* on the tensor rank of this tensor.

2.3.2 Matrix multiplication tensor

Definition 2.7. *The tensor corresponding to the product of two* $n \times n$ *matrices is defined as*

$$M_n(X, Y, Z) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} X_{i,j} Y_{j,k} Z_{i,k}.$$

Here, $X = \{X_{i,j} : i, j \in [n]\}, Y = \{Y_{i,j} : i, j \in [n]\}, Z = \{Z_{i,j} : i, j \in [n]\}$.

Note that $M_n(X, Y, Z)$ is the trace of the matrix product $X \cdot Y \cdot Z^T$. In other words, $M_n(X, Y, Z^T) = \text{Trace}(X \cdot Y \cdot Z)$. Note this is the matrix trace and is different from the trace function considered in the previous section where we viewed X, Y, Z as elements of the large field.

3 Correlation of random d-linear forms

In this section, we study the correlation of random d-linear forms with lower degree polynomials. Our main result in this section is the following theorem, which states that a random d-linear form is uncorrelated with degree ℓ polynomials under certain conditions.

Theorem 3.1. *Let* ℓ, d, n *be integers such that* $d \mid n$, $d = o(\frac{n}{\log n})$ *and* $\ell < d/2$*. Set* $k = n/d$*.*

Pick a uniformly random d*-linear form* $f : (\mathbb{F}_2^k)^d \to \mathbb{F}_2$*. Then, with probability* $1 - o(1)$*,* f *has the following property. For all polynomials* $P(X_1, \ldots, X_n) \in \mathbb{F}_2[X_1, \ldots, X_n]$ *with degree at most* ℓ*, we have,*

$$\text{Corr}(f, P) < 2^{-(1-o(1))n/d}.$$

Along the way, we develop several tools to understand the bias of random d-linear forms. For example, we show that a random d-linear form is unbiased with extremely high probability.

Theorem 3.2. *Let $\varepsilon > 0$ be fixed. Let d, k be integers with $d < 2^{\varepsilon k/5}$, and consider a uniformly random d-linear form $f : (\mathbb{F}_2^k)^d \to \mathbb{F}_2$. Then,*

$$\Pr[\text{bias}(f) \geq 2^{-(1-\varepsilon)k}] \leq 2^{-\Omega(\varepsilon^2 k^d)}.$$

Remark 3.3. Note that any d-linear form $f(X_1, \ldots, X_d)$ vanishes if any one of the block of variables $X_1, \ldots, X_d$ is zero. Hence, the bias of any d-linear form (or equivalently its correlation with the constant 0 polynomial) is at least $2^{-k} = 2^{-n/d}$. Theorem 3.2 states that it is extremely unlikely for a random d-linear form to have even slightly more bias while Theorem 3.1 states that it is extremely unlikely for a random d-linear form to have slightly better correlation with any degree ℓ polynomial.

The key ingredient in the proofs of the above theorems is the following theorem on the distribution of the sum of random rank-1 tensors.

Theorem 3.4. *Let $\varepsilon > 0$ be a constant. Let d, k, t be integers with $d < 2^{\varepsilon k/5}$, and $t < \frac{\varepsilon}{5} k^{d-1}$. Let $\{x^{(i,j)}\}_{i \in [t], j \in [d]}$ be picked independently and uniformly distributed in $\mathbb{F}_2^k$.Then,*

$$\Pr\left[\sum_{i=1}^{t} \bigotimes_{j=1}^{d} x^{(i,j)} = 0\right] \leq 2^{-(1-\varepsilon/2)\cdot kt}.$$

Remark 3.5. If any block of vectors (say wlog. $\{x^{(i,1)}\}_{i \in [t]}$, the first block of vectors) are all $\bar{0}$ (this happens with probability 2^{-kt}), then the d-dimensional linear form $\sum_{i=1}^{t} \bigotimes_{j=1}^{d} x^{(i,j)} = 0$. The above theorem states that the probability of the d-linear form vanishing is not significantly larger.

In turn, the proof of the above theorem is based on the following lemma, which gives an upper bound on the probability that a random rank-1 tensor lies in a fixed low dimensional subspace.

Lemma 3.6. *Let k, d be integers and U be a subspace of $(\mathbb{F}_2^k)^{\otimes d}$ of dimension u. Let $x_1, \ldots, x_d \in \mathbb{F}_2^k$ be picked independently and uniformly at random, and let $T = \otimes_{i=1}^{d} x_i$. Then,*

$$\Pr[T \in U] \leq \frac{d}{2^k} + \frac{2^{u/k^{d-1}}}{2^k}.$$

Remark 3.7. Let $U = V \otimes (\mathbb{F}_2^k)^{\otimes (d-1)}$ where V is a u/k^{d-1}-dimensional subspace of $\mathbb{F}_2^k$. Note, $\dim(U) = u$. Clearly, $\Pr[\otimes_{i=1}^{d} x_i \in U] = \Pr[x_1 \in V] = 2^{u/k^{d-1}}/2^k$. The above lemma states that the probability is not significantly larger than this for any other U.

In the next subsection, we show how Theorem 3.1 and Theorem 3.2 follow from Theorem 3.4. After that, we prove Theorem 3.4 by studying the distribution of the dimension of a collection of random rank 1 tensors.

3.1 Proofs of Theorem 3.1 and Theorem 3.2

We first prove Theorem 3.2.

Proof of Theorem 3.2. We want to bound $\Pr_f[\text{bias}(f) \geq 2^{-(1-\varepsilon)k}]$. We shall do so by bounding the t^{th} moment of $\text{bias}(f)$ for a suitable choice of t and applying Markov's inequality.

Let $T : [k]^d \to \mathbb{F}_2$ denote the tensor associated with f. Thus $T(i_1, \ldots, i_d)$ are all independent and uniformly distributed in $\mathbb{F}_2$.

We now compute the t^{th} moment of f.

$$\begin{aligned}
\mathbb{E}_f[(\text{bias}(f))^t] &= \mathbb{E}_f\left[\left(\mathbb{E}_{x^{(1)},\ldots,x^{(d)}\sim\mathbb{F}_2^k}\left[(-1)^{f\left(x^{(1)},\ldots,x^{(d)}\right)}\right]\right)^t\right] \\
&= \mathbb{E}_f\left[\prod_{i\in[t]}\left(\mathbb{E}_{x^{(i,1)},\ldots,x^{(i,d)}\sim\mathbb{F}_2^k}\left[(-1)^{f\left(x^{(i,1)},\ldots,x^{(i,d)}\right)}\right]\right)\right] \\
&= \mathbb{E}_{\{x^{(i,j)}\}_{i\in[t],j\in[d]}}\left[\mathbb{E}_f\left[(-1)^{\sum_{i=1}^t f\left(x^{(i,1)},\ldots,x^{(i,d)}\right)}\right]\right] \\
&= \mathbb{E}_{\{x^{(i,j)}\}_{i\in[t],j\in[d]}}\left[\prod_{(\ell_1,\ldots,\ell_d)\in[k]^d}\left(\mathbb{E}_{T(\ell_1,\ldots,\ell_d)\sim\mathbb{F}_2}\left[(-1)^{T(\ell_1,\ldots,\ell_d)\cdot\left(\sum_{i=1}^t\prod_{j=1}^d x_{\ell_j}^{(i,j)}\right)}\right]\right)\right] \\
&= \mathbb{E}_{\{x^{(i,j)}\}_{i\in[t],j\in[d]}}\left[\prod_{(\ell_1,\ldots,\ell_d)\in[k]^d}\mathbb{1}_{\sum_{i=1}^t\prod_{j=1}^d x_{\ell_j}^{(i,j)}=0}\right] \\
&= \mathbb{E}_{\{x^{(i,j)}\}_{i\in[t],j\in[d]}}\left[\mathbb{1}_{\forall(\ell_1,\ldots,\ell_d)\in[k]^d,\ \sum_{i=1}^t\prod_{j=1}^d x_{\ell_j}^{(i,j)}=0}\right] \\
&= \Pr_{\{x^{(i,j)}\}_{i\in[t],j\in[d]}}\left[\forall(\ell_1,\ldots,\ell_d)\in[k]^d,\ \sum_{i=1}^t\prod_{j=1}^d x_{\ell_j}^{(i,j)}=0\right] \\
&= \Pr_{\{x^{(i,j)}\}_{i\in[t],j\in[d]}}\left[\sum_{i=1}^t\bigotimes_{j=1}^d x^{(i,j)}=0\right].
\end{aligned}$$

Setting $t = \frac{\varepsilon}{10}k^{d-1}$, Theorem 3.4 tells us that

$$\mathbb{E}_f[(\text{bias}(f))^t] = 2^{-(1-\varepsilon/2)kt}.$$

Using Markov's inequality,

$$\Pr_f\left[\text{bias}(f) \geq 2^{-(1-\varepsilon)k}\right] \leq \frac{2^{-(1-\varepsilon/2))kt}}{2^{-(1-\varepsilon)kt}} \leq 2^{-\varepsilon kt/2} \leq 2^{-\Omega(\varepsilon^2 k^d)}$$

as claimed.

□

We now use a similar argument to prove Theorem 3.1.

Proof of Theorem 3.1. Fix an arbitrary $\varepsilon > 0$. Let $\mathcal{C}$ denote the space of degree $\leq \ell$ polynomials in $\mathbb{F}_2[X_1, \ldots, X_n]$. We want to show that with high probability over the choice of f, we have that for every $P \in \mathcal{C}$, $\mathrm{Corr}(f, P) \leq 2^{-(1-\varepsilon)k}$.

Fix $P \in \mathcal{C}$ and consider the t^{th} moment of $\mathrm{bias}(f - P)$. Imitating the proof of Theorem 3.2, we get

$$\begin{aligned}
\mathbb{E}_f[(\mathrm{bias}(f-P))^t] &= \mathbb{E}_{\{x^{(i,j)}\}_{i\in[t],j\in[d]}}\left[(-1)^{\sum_{i=1}^t P\left(x^{(i,1)},\ldots,x^{(i,d)}\right)} \cdot \mathbb{1}_{\forall(\ell_1,\ldots,\ell_d)\in[k]^d,\ \sum_{i=1}^t\prod_{j=1}^d x_{\ell_j}^{(i,j)}=0}\right] \\
&\leq \mathbb{E}_{\{x^{(i,j)}\}_{i\in[t],j\in[d]}}\left[\mathbb{1}_{\forall(\ell_1,\ldots,\ell_d)\in[k]^d,\ \sum_{i=1}^t\prod_{j=1}^d x_{\ell_j}^{(i,j)}=0}\right] \\
&= \Pr\left[\sum_{i=1}^t \bigotimes_{j=1}^d x^{(i,j)} = 0\right].
\end{aligned}$$

Now we will apply Theorem 3.4. Observe that since $d = o(n/\log n)$, we have,

$$d < 2^{\varepsilon k/5}.$$

As in the proof of Theorem 3.2, we set $t = \frac{\varepsilon}{10}k^{d-1}$, invoke Theorem 3.4 and apply Markov's inequality to get,

$$\Pr_f\left[\mathrm{bias}(f-P) \geq 2^{-(1-\varepsilon)k}\right] \leq 2^{-\varepsilon^2 k^d/20}.$$

Now $\mathrm{bias}(f - P) = \mathrm{Corr}(f, P)$. Thus, by a union bound over all $P \in \mathcal{C}$, we have the following.

$$\Pr_f\left[\mathrm{Corr}(f,\mathcal{C}) \geq 2^{-(1-\varepsilon)k}\right] \leq |\mathcal{C}| \cdot 2^{-\varepsilon^2 k^d/20}. \tag{1}$$

It remains to estimate $|\mathcal{C}|$. We show below that $|\mathcal{C}| = o(k^d)$. The proof of this lemma works for any other C as long as C satisfies $|\mathcal{C}| = o(k^d)$. Note that $|\mathcal{C}| = 2^{\binom{n}{\leq \ell}}$. Let δ denote d/n.

$$\begin{aligned}
\binom{n}{\leq \ell} \leq \binom{n}{\leq d/2} &\leq \left(\frac{2en}{d}\right)^{d/2} \leq \left(\frac{2e}{\delta}\right)^{\delta n/2} \\
&= o\left(\left(\frac{1}{\delta}\right)^{\delta n}\right) \qquad [\text{Since } \delta = o(1)] \\
&= o(k^d).
\end{aligned}$$

Combining this with Equation (1), we get,

$$\Pr_f\left[\mathrm{Corr}(f,\mathcal{C}) \geq 2^{-(1-\varepsilon)k}\right] \leq 2^{o(k^d)} \cdot 2^{-\varepsilon^2 k^d/20}.$$

Since this holds for every $\varepsilon > 0$, we get the desired result. □

3.2 Random rank-1 tensors

In this subsection, we first prove Lemma 3.6 on the probability that a random rank-1 tensor lies in a fixed low-dimensional subspace. We then give a corollary of this lemma which bounds the probability that a collection of random rank-1 tensors spans a very low dimensional subspace. This corollary will be used in the proof of Theorem 3.4.

Proof of Lemma 3.6. Define

$$f_{d,k}(u) = \left(1 - (1 - \frac{1}{2^k})^{d-1}\right) + (1 - \frac{1}{2^k})^{d-1} \cdot \frac{2^{u/k^{d-1}}}{2^k}.$$

We will prove, by induction on d, the following stronger bound.

$$\Pr[T \in U] \leq f_{d,k}(u).$$

The fact that this implies the lemma, follows from the observations that $1 - \frac{d-1}{2^k} \leq (1 - \frac{1}{2^k})^{d-1}$ and that $(1 - \frac{1}{2^k})^{d-1} \leq 1$.

Base case. The $d = 1$ case is trivial (using the observation that $f_{1,k}(u) = \frac{2^u}{2^k}$). We now show the statement holds for larger d.

Induction step. Let $k' = k^{d-1}$. We will view $(\mathbb{F}_2^k)^{\otimes d}$ as $\mathbb{F}_2^k \otimes \mathbb{F}_2^{k'}$. Every element v of $(\mathbb{F}_2^k)^{\otimes d}$ can thus be written as a tuple $(v_1, \ldots, v_k)$, where each v_i is an element of $\mathbb{F}_2^{k'}$ (thus the k^d coordinates are partitioned into k blocks of coordinates, with each block having k' coordinates). We let $\pi_i : (\mathbb{F}_2^k)^{\otimes d} \to \mathbb{F}_2^{k'}$ be the ith projection map, mapping v to v_i.

With this convention, we take a basis for U in *row echelon form*. Concretely, this gives us a basis $\mathcal{B}$ for U, such that $\mathcal{B}$ is a disjoint union of $\mathcal{B}_1, \ldots, \mathcal{B}_k$ ($\mathcal{B}_j$ is the set of basis vectors pivoted in the j'th block of coordinates), such that,

- for all $v \in \mathcal{B}_j$ and $i < j$, $\pi_i(v) = 0$,
- the vectors $\pi_j(v) \in \mathbb{F}_2^{k'}$, as v varies in $\mathcal{B}_j$, are linearly independent.

Define $U_j = \text{span}\{\pi_j(v) \mid v \in \mathcal{B}_j)\}$. Thus we have $\dim(U_j) = |\mathcal{B}_j|$ and

$$\sum_{j=1}^{k} \dim(U_j) = \dim(U).$$

For $i > j$, we define a linear map $\psi_{ij} : U_j \to \mathbb{F}_2^{k'}$ by defining ψ_{ij} on a basis for U_j:

$$\psi_{ij}(\pi_j(v)) = \pi_i(v), \ \forall v \in B_j.$$

Then we have the following basic claim (which follows immediately from the above echelon form representation of U).

Claim 3.8. *Let $v \in (\mathbb{F}_2^k)^{\otimes d}$. Then $v \in U$ only if there exists $(u_1, \ldots, u_k) \in \prod_{i=1}^k U_i$ such that for each $i \in [k]$ we have*

$$\pi_i(v) = u_i + \sum_{j<i} \psi_{ij}(u_j).$$

To simplify notation, we will denote x_1 by y and $\otimes_{i=2}^{d} x_i$ by z. We want to find an upper bound on $\Pr[y \otimes z \in U]$.

Claim 3.9. *Let $\tilde{z} \in (\mathbb{F}_2^k)^{\otimes(d-1)}$ and $S = \{i \mid \tilde{z} \in U_i\}$, then,*

$$\Pr_{y \in \mathbb{F}_2^k}[y \otimes \tilde{z} \in U] \leq \frac{2^{|S|}}{2^k}.$$

Proof. For fixed $\tilde{z}$, given the random variable $v = y \otimes \tilde{z}$, we define random variables $u_1, u_2, ..., u_k$ by: $u_i := \pi_i(v) - \sum_{j<i} \psi_{ij}(u_j)$. Note that $\pi_i(v) = \pi_i(y \otimes \tilde{z}) = y_i\tilde{z}$. Also note that u_i is only a function of $y_1, \ldots, y_i$. By Claim 3.8, $v \in U$ only if for all i, $u_i \in U_i$.

$$\begin{aligned}
\Pr_{y \in \mathbb{F}_2^k}[y \otimes \tilde{z} \in U] &\leq \Pr_{y}[\forall i \leq k,\ u_i \in U_i] \\
&= \prod_{i=1}^{k} \Pr\left[u_i \in U_i \mid u_1 \in U_1, \ldots, u_{i-1} \in U_{i-1}\right] \\
&= \prod_{i=1}^{k} \mathbb{E}_{u_1 \in U_1, \ldots, u_{i-1} \in U_{i-1}} \left[\Pr_{u_i}\left[u_i \in U_i \mid u_1, \ldots, u_{i-1}\right]\right] \\
&= \prod_{i=1}^{k} \mathbb{E}_{u_1 \in U_1, \ldots, u_{i-1} \in U_{i-1}} \left[Pr_{u_i}\left[\pi_i(v) - \sum_{j<i} \psi_{ij}(u_j) \in U_i \;\middle|\; u_1, \ldots, u_{i-1}\right]\right] \\
&= \prod_{i=1}^{k} \mathbb{E}_{u_1 \in U_1, \ldots, u_{i-1} \in U_{i-1}} \left[Pr_{y_i}\left[y_i\tilde{z} - \sum_{j<i} \psi_{ij}(u_j) \in U_i \;\middle|\; u_1, \ldots, u_{i-1}\right]\right] \\
&\leq \prod_{i \notin S} \left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^{k-|S|},
\end{aligned}$$

where the last inequality follows since for every $i \notin S$ and every vector w, at most one of w and $w + \tilde{z}$ can lie in U_i (as $\tilde{z} \notin U_i$).

□

For $S \subseteq [k]$, let

$$U_S = \bigcap_{i \in S} U_i\,.$$

Then,

$$\begin{aligned}
\Pr_{y,z}[y \otimes z \in U] &\leq \mathbb{E}_z \left[\frac{2^{\sum_{i=1}^{k} 1_{U_i}(z)}}{2^k} \right] \qquad \text{[Follows from the above claim]} \\
&= \frac{1}{2^k} \mathbb{E}_z \left[\prod_{i=1}^{k} 2^{1_{U_i}(z)} \right] \\
&= \frac{1}{2^k} \mathbb{E}_z \left[\prod_{i=1}^{k} (1 + 1_{U_i}(z)) \right] \\
&= \frac{1}{2^k} \mathbb{E}_z \left[\sum_{S \subseteq [k]} 1_{U_S}(z) \right] \\
&= \frac{1}{2^k} \sum_{S \subseteq [k]} \Pr_z[z \in U_S].
\end{aligned}$$

Now, observe that for each $i \in S$, we have $\Pr[z \in U_S] \leq \Pr[z \in U_i]$. Thus if we sort the U_i so that $\dim(U_1) \geq \dim(U_2) \geq \ldots \geq \dim(U_k)$, then we have the following sequence of inequalities.

$$\begin{aligned}
\Pr_{y,z}[y \otimes z \in U] &\leq \frac{1}{2^k} \left(1 + \sum_{i \in [k]} \sum_{S \subseteq [i], i \in S} \Pr_z[z \in U_S] \right) \\
&\leq \frac{1}{2^k} \left(1 + \sum_{i \in [k]} 2^{i-1} \Pr_z[z \in U_i] \right) \\
&\leq \frac{1}{2^k} \left(1 + \sum_{i \in [k]} 2^{i-1} f_{d-1,k}(\dim(U_i)) \right),
\end{aligned}$$

where the last step follows from the induction hypothesis. To find an upper bound for this last expression, we let $a_i = \dim(U_i)$. We have the constraints

$$\sum_i a_i = u,$$

$$k' \geq a_1 \geq a_2 \geq \ldots \geq a_k \geq 0,$$

where $k' = k^{d-1}$, and we want to maximize an expression of the form

$$\sum_{i=1}^{k} 2^{i-1} (\alpha + \beta 2^{a_i / k^{d-2}}) = \alpha \cdot (2^k - 1) + \beta \cdot \left(\sum_{i=1}^{k} 2^{i-1+a_i/k^{d-2}} \right).$$

where $\alpha, \beta > 0$.

It is worth noting what happens in the two examples $U = V \otimes \mathbb{F}_2^{k'}$ and $U = \mathbb{F}_2^k \otimes W$, where $V \subseteq \mathbb{F}_2^k$ and $W \subseteq \mathbb{F}_2^{k'}$ are subspaces of the appropriate dimension. In the first case, $a_1 = a_2 = \ldots = a_{u/k'} = k'$ and the remaining a_i are 0. In the second case, all the $a_i = u/k$. Both are global maxima

of the expression we want to maximize! The existence of these very different maxima makes this maximization problem somewhat tricky.

In Theorem 3.10 we prove a tight upper bound for this function. For every $i \in [k]$, let $b_i = a_i/k^{d-2}$, and let $\tilde{u} = u/k^{d-2}$. Then, $b_1, b_2, \ldots, b_k$ and $\tilde{u}$ satisfy the constraints in the hypothesis of Theorem 3.10, and Theorem 3.10 tells us that a global maxima is achieved when all the a_i are equal to $\dim(U)/k$. Thus,

$$\begin{aligned}
\Pr_{y,z}[y \otimes z \in U] &\leq \frac{1}{2^k}\left(1 + \sum_{i\in[k]} 2^{i-1} f_{d-1,k}(u/k)\right) \\
&= \frac{1}{2^k}\left(1 + (2^k - 1) f_{d-1,k}(u/k)\right) \\
&= \left(\frac{1}{2^k} + (1 - \frac{1}{2^k}) f_{d-1,k}(u/k)\right) \\
&= f_{d,k}(u).
\end{aligned}$$

This completes the induction step. □

Theorem 3.10. *Let k be a positive integer, and let $\tilde{u} \in [0, k^2]$ be a real number. Suppose $b_1, b_2, \ldots, b_k$ are real numbers satisfying the following constraints.*

$$k \geq b_1 \geq b_2 \ldots \geq b_k \geq 0, \tag{2}$$

$$\sum_{i=1}^{k} b_i = \tilde{u}. \tag{3}$$

Then,

$$\sum_{i=1}^{k} 2^{i-1} 2^{b_i} \leq \sum_{i=1}^{k} 2^{i-1} 2^{\tilde{u}/k} = (2^k - 1) 2^{\tilde{u}/k}.$$

Theorem 3.10 is proved in the appendix.

We now use the previous lemma to prove a corollary about the dimension of the span of several random rank 1 tensors.

Corollary 3.11. *Let d, k, t be integers. For each $i \in [t]$ and $j \in [d]$, pick $x^{(i,j)} \in \mathbb{F}_2^k$ uniformly at random. For $i \in [t]$, let T_i be the rank-1 tensor $\otimes_{j=1}^{d} x^{(i,j)}$. Then, for every $0 \leq r \leq t$,*

$$\Pr[\dim(\mathrm{span}(\{T_1, \ldots, T_t\})) = r] \leq \binom{t}{r} \left(\frac{d + 2^{t/k^{d-1}}}{2^k}\right)^{t-r}.$$

Proof. Let us reveal $T_1, \ldots, T_t$ one at a time. For $0 \leq i \leq t$, let $V_i = \mathrm{span}(\{T_1, \ldots, T_{i-1}, T_i\})$. Thus we have $0 = \dim(V_0) \leq \dim(V_1) \leq \ldots \dim(V_t)$. We want to estimate the probability that $\dim(V_t) = r$. Let E_i denote the event that $T_i \in V_{i-1}$. For $I \subseteq [t]$, let E_I denote the event $\bigcap_{i\in I} E_i$. In terms of these events, we can bound $\Pr[\dim(V_t) = r]$ as follows.

$$\begin{aligned}
\Pr[\dim(V_t) = r] &\leq \Pr[\exists I \subseteq [t], |I| = t - r \text{ such that } E_I \text{ occurs}] \\
&\leq \sum_{I\subseteq[t],|I|=t-r} \Pr[E_I].
\end{aligned}$$

We conclude the proof by bounding $\Pr[E_I]$. Fix $I \subseteq [t]$ with $|I| = t - r$. Let $I = \{i_1, \ldots, i_{t-r}\}$ with $i_1 < i_2 < \ldots < i_{t-r}$.

$$\Pr[E_I] = \prod_{j=1}^{t-r} \Pr[E_{i_j} | \bigcap_{\ell<j} E_{i_\ell}].$$

Lemma 3.6 implies the following.

$$\Pr[E_i | T_1, \ldots, T_{i-1}] \leq \frac{d + 2^{\dim(V_{i-1})/k^{d-1}}}{2^k}.$$

For any given $j \in [t-r]$, the events $E_{i_1}, \ldots, E_{i_{j-1}}$ are all determined by $T_1, \ldots, T_{i_j-1}$ (since E_{i_ℓ} depends on $T_1, \ldots, T_{i_\ell}$, and $i_{j-1} \leq i_j - 1$). Thus, for each $j \in [t-r]$, we have,

$$\Pr[E_{i_j} | \bigcap_{\ell<j} E_{i_\ell}] \leq \frac{d + 2^{t/k^{d-1}}}{2^k}.$$

Here we used the fact that $\dim(V_{i_j-1}) \leq t$. Using this in our previous bound, we conclude that

$$\Pr[E_I] \leq \left(\frac{d + 2^{t/k^{d-1}}}{2^k}\right)^{t-r},$$

and thus,

$$\Pr[\dim(V_t) = r] \leq \binom{t}{r} \cdot \left(\frac{d + 2^{t/k^{d-1}}}{2^k}\right)^{t-r}. \qquad \square$$

3.3 Proof of Theorem 3.4

We now use Corollary 3.11 to prove Theorem 3.4.

Proof of Theorem 3.4. The equation

$$\sum_{i=1}^{t} \bigotimes_{j=1}^{d} x^{(i,j)} = 0 \tag{4}$$

implies that

$$\forall \ell \in [k], \sum_{i=1}^{t} x_\ell^{(i,1)} \cdot \bigotimes_{j=2}^{d} x^{(i,j)} = 0. \tag{5}$$

Let T_i denote $\bigotimes_{j=2}^{d} x^{(i,j)}$ for $i \in [t]$ and $\mathcal{T} = \mathrm{span}(\{T_1, \ldots, T_t\})$. Then we have,

$$\begin{aligned}
\Pr[\{x^{(i,j)}\}_{i\in[t],j\in[d]} \text{ satisfy (4)}] &\leq \Pr[\{x^{(i,j)}\}_{i\in[t],j\in[d]} \text{ satisfy (5)}] \\
&= \sum_{r=0}^{t} \Pr\left[\{x^{(i,j)}\}_{i\in[t],j\in[d]} \text{ satisfy (5)} \middle| \dim(\mathcal{T}) = r\right] \Pr\left[\dim(\mathcal{T}) = r\right] \\
&= \sum_{r=0}^{t} \left(\prod_{\ell\in[k]} \Pr\left[\sum_{i=1}^{t} x_\ell^{(i,1)} \cdot T_i = 0 \middle| \dim(\mathcal{T}) = r\right]\right) \cdot \Pr\left[\dim(\mathcal{T}) = r\right] \\
&\leq \sum_{r=0}^{t} \left(\frac{1}{2^r}\right)^k \cdot \Pr\left[\dim(\mathcal{T}) = r\right]. \qquad (6)
\end{aligned}$$

Here, the equality in the third step follows from the fact that $\{x_\ell^{(i,1)}\}_{i\in[t],\ell\in[k]}$ are independently and uniformly distributed in $\mathbb{F}_2$.
By the given distribution of $T_1,\ldots,T_t$ in $(\mathbb{F}_2^k)^{\otimes(d-1)}$, Corollary 3.11 says that

$$\Pr\left[\dim(\mathcal{T})=r\right] \leq \binom{t}{r}\left(\frac{d-1+2^{\frac{t}{k^{d-2}}}}{2^k}\right)^{t-r}.$$

Plugging this bound back into (6) gives

$$\begin{aligned}
\Pr[\{x^{(i,j)}\}_{i\in[t],j\in[d]} \text{ satisfy (4)}] &\leq \sum_{r=0}^{t}\binom{t}{r}\frac{1}{2^{rk}}\left(\frac{d-1+2^{\frac{t}{k^{d-2}}}}{2^k}\right)^{t-r}\\
&\leq \sum_{r=0}^{t}\binom{t}{r}\left(\frac{1}{2^k}\right)^{r}\left(\frac{d-1+2^{\frac{t}{k^{d-2}}}}{2^k}\right)^{t-r}\\
&= \left(\frac{1}{2^k}+\frac{d-1+2^{\frac{t}{k^{d-2}}}}{2^k}\right)^{t}\\
&\leq \left(\frac{d+2^{\frac{t}{k^{d-2}}}}{2^k}\right)^{t}.
\end{aligned}$$

Now, since $d<2^{\varepsilon k/5}$ and $t<\varepsilon k^{d-1}/5$, we have

$$d+2^{\frac{t}{k^{d-2}}} < 2\cdot 2^{\varepsilon k/5} < 2^{\varepsilon k/2},$$

we conclude that

$$\Pr[\sum_{i=1}^{t}\bigotimes_{j=1}^{d}x^{(i,j)}=0] < 2^{-(1-\varepsilon/2)kt}.$$

This completes the proof. □

3.4 Explicit d-linear forms with small correlations with $(d-1)$-linear forms

In this section, we dwell a bit on the question of constructing explicit d-linear forms which have small correlation with lower degree multilinear polynomials. In particular, we present an explicit d-linear form that has exponentially small correlation with any lower degree multilinear form. Define $f:(\mathbb{F}_2^k)^d\to\mathbb{F}_2$ as

$$f(x_1,\ldots,x_d)=\langle x_1\cdot x_2\cdots x_{d-1},x_d\rangle,$$

where $\cdot$ denotes multiplication over the bigger field $\mathbb{F}_{2^k}$. It is easy to see that f is a d-linear form.

Ideally, we would like to show that the map f defined above has small correlation with *any* polynomial of degree at most $d-1$. But, we do not know how to show this. In the rest of this section, we show that f has small correlation with any polynomial of degree $d-1$ which respects the partition of the inputs to f. We now prove the following lemma.

Lemma 3.12. *The function $f = \langle x_1 \cdot x_2 \cdots x_{d-1}, x_d \rangle$ has correlation at most $(d-1)2^{-k}$ with any degree $\leq d-1$ multilinear form.*

Proof. Let g be a $(d-1)$-linear form over $(\mathbb{F}_2^k)^d$. A similar proof as below works for any d'-linear form g' for $d' < d-1$ also. We want to understand

$$\mathrm{Corr}(f,g) = \mathrm{bias}(f-g).$$

Since g is a $(d-1)$-form, it is of the form $g(x_1,...,x_d) = \sum_{i=1}^{d} g_i(x_{[d]\setminus\{i\}})$. Since g_i is a $(d-1)$-linear form in the variables $x_{[d]\setminus\{i\}}$, for each $i \in [d-1]$ there exists an $\mathbb{F}_2^k$-valued linear form $v_i = v_i(x_{[d-1]\setminus\{i\}})$ such that $g_i(x_{[d]\setminus\{i\}}) = \langle v_i(x_{[d-1]\setminus\{i\}}), x_d \rangle = \langle v_i, x_d \rangle$. In particular

$$\mathrm{Corr}(f,g) = \mathrm{bias}(f-g) = \mathrm{bias}\left((f - \sum_{i=1}^{d-1} g_i) - g_d\right) = \Pr_{x_1,\ldots x_{d-1}}\left[x_1 \cdot x_2 \cdots x_{d-1} - \sum_{i=1}^{d-1} v_i = \overline{0}\right]. \quad (7)$$

This is because g_d does not depend on x_d and for any fixing of $x_1,...,x_{d-1} \in \mathbb{F}_2^k$, $f-g$ is an affine form in the variable x_d that is biased if $x_1 \cdot x_2 \cdots x_{d-1} - \sum_{i=1}^{d-1} v_i = \overline{0}$ (in which case the bias is 1) and is otherwise an unbiased function. We will prove

$$\Pr_{x_1,\ldots x_{d-1} \in \mathbb{F}_2^k}\left[x_1 \cdot x_2 \cdot \cdots \cdot x_{d-1} - \sum_{i=1}^{d-1} g_i = \overline{0}\right] \leq \Pr_{x_1,\ldots,x_{d-1} \in \mathbb{F}_2^k}[x_1 \cdot x_2 \cdot \cdots \cdot x_{d-1} = \overline{0}], \quad (8)$$

by repeatedly applying the following fact.

Fact 3.13. *Let $h : \mathbb{F}_2^k \to \mathbb{F}_2^k$ be a linear map. Then for every $\overline{a} \in \mathbb{F}_2^k$,*

$$\Pr_{x \in \mathbb{F}_2^k}[h(x) = \overline{a}] \leq \Pr_{x \in \mathbb{F}_2^k}[h(x) = \overline{0}].$$

Note that applying this fact we have

$$\begin{aligned}\Pr_{x_1,\ldots x_{d-1}}\left[x_1 \cdot x_2 \cdots x_{d-1} - \sum_{i=1}^{d-1} v_i = \overline{0}\right] &= \Pr_{x_1,\ldots x_{d-1}}\left[x_1 \cdot x_2 \cdots x_{d-1} - \sum_{i=1}^{d-2} v_i = v_{d-1}\right] \\ &\leq \Pr_{x_1,\ldots x_{d-1}}\left[x_1 \cdot x_2 \cdots x_{d-1} - \sum_{i=1}^{d-2} v_i = \overline{0}\right],\end{aligned}$$

since for every fixing of $x_1,...,x_{d-2} \in \mathbb{F}_2^k$, $x_1 \cdot x_2 \cdots x_{d-1} - \sum_{i=1}^{d-2} v_i$ is a $\mathbb{F}_2^k$-linear form over x_{d-1}. Applying Fact 3.13 similarly for coordinates $i = 1,...,d-2$, we get Eq. (8). Finally, by a simple union bound we can bound $\Pr_{x_1,\ldots,x_{d-1} \in \mathbb{F}_2^k}[x_1 \cdot x_2 \cdots x_{d-1} = \overline{0}] = 1 - (1-2^{-k})^{d-1} \leq (d-1) \cdot 2^{-k}$. Combining this with Eq. (8) and Eq. (7) finishes our proof. □

4 High-rank tensors from unbiased polynomials

It is well-known that the bias of a bilinear form corresponding to a matrix $M \in \mathbb{F}_2^{k \times k}$ is tightly related to its rank $\mathrm{rank}(M)$ (more precisely, $\mathrm{bias}(M) = 2^{-\mathrm{rank}(M)}$). In this section, we explore a similar connection for higher dimensional tensors. We then use this to (re)prove some existing tensor rank lower bounds (e.g., for the trace tensor and the matrix multiplication tensor)

4.1 Small Bias implies large tensor rank

We begin with the main theorem of this section which shows tensors with small bias have large rank.

Theorem 4.1 (Small bias implies large rank)**.** *Let $P \in \mathbb{F}_2^{k \times k \cdots \times k}$ be any d-dimensional tensor of rank $\leq t$. Then*

$$\text{bias}(P) \geq \left(1 - \frac{2}{2^d}\right)^t.$$

An important ingredient of our proof will be the following lemma.

Lemma 4.2. *Let d be a natural number. Let $M_1, M_2, \ldots, M_t \in \mathbb{F}_2^{k \times k \cdots \times k}$ be d-dimensional tensors of rank at most 1. Then,*

$$\Pr_{x_1, x_2, \ldots, x_d \in \mathbb{F}_2^k} [\forall i \in [t], M_i(x_1, x_2, \ldots, x_d) = 0] \geq \left(1 - \frac{1}{2^d}\right)^t. \tag{9}$$

Proof. Our proof is by induction on d.

Base Case. The base case when $d = 1$ trivially follows since if there are t linear forms $u_1, u_2, \ldots, u_t$ over $\mathbb{F}_2$, then the maximum number r of independent linear forms among them is at most t. We hence have,

$$\Pr_{x \in \mathbb{F}_2^k} [\forall i \in [t], u_i(x) = 0] = (1/2)^r \geq (1/2)^t. \tag{10}$$

Induction Step. Before proving the general inductive step from $d - 1$ to d, we first show the $d = 2$ case as a warm up as it illustrates the main idea and then do the general case.

For this case, we have $k \times k$ matrices $M_1, M_2, \ldots, M_t$ of rank one over $\mathbb{F}_2$, and the goal is to show that

$$\Pr_{y, z \in \mathbb{F}_2^k} [\forall i \in [t], \langle y, M_i z \rangle = 0] \geq (3/4)^t. \tag{11}$$

The proof involves several steps of manipulation of the probability of interest. For a set $S \subseteq [t]$,

denote by $M_S := \sum_{i \in S} M_i$.

$$\begin{aligned}
\Pr_{y,z \in \mathbb{F}_2^k}[\forall i \in [t], \langle y, M_i z \rangle = 0] &= \mathbb{E}_{y,z \in \mathbb{F}_2^k}\left[\prod_{i=1}^{t}\left(\frac{1+(-1)^{\langle y, M_i z\rangle}}{2}\right)\right] \\
&= \mathbb{E}_{y,z \in \mathbb{F}_2^k}\left[\frac{1}{2^t}\cdot\sum_{S \subseteq [t]}(-1)^{\langle y, M_S z\rangle}\right] \\
&= \mathbb{E}_{y,z \in \mathbb{F}_2^k}\left[\mathbb{E}_{S \subseteq [t]}\left[(-1)^{\langle y, M_S z\rangle}\right]\right] \\
&= \mathbb{E}_{S \subseteq [t]}\left[\mathbb{E}_{y,z \in \mathbb{F}_2^k}\left[(-1)^{\langle y, M_S z\rangle}\right]\right] \\
&= \mathbb{E}_{S \subseteq [t]}\left[\mathbb{E}_{z \in \mathbb{F}_2^k}\left[1_{M_S z = \overline{0}}\right]\right] \\
&= \mathbb{E}_{S \subseteq [t]}\left[\Pr_{z \in \mathbb{F}_2^k}\left[M_S z = \overline{0}\right]\right] \\
&= \mathbb{E}_{S \subseteq [t]}\left[2^{-\mathrm{rank}(M_S)}\right] \\
&\geq \mathbb{E}_{S \subseteq [t]}\left[2^{-|S|}\right] = \frac{1}{2^t}\cdot\left(1+\frac{1}{2}\right)^t = \left(\frac{3}{4}\right)^t .
\end{aligned}$$

Now, for the general inductive step, we assume that the lemma is true up to dimension $d-1$, and prove it for d dimensions. For every $i \in [t]$, we denote by u_i as the linear form in $\mathbb{F}_2^k$ and M_i' as the $d-1$ dimensional tensor of rank 1 in $\mathbb{F}_2^{k \times k \times k \cdots \times k}$ such that

$$M_i(x_1, x_2, \ldots, x_d) = u_i(x_1) \cdot M_i'(x_2, x_3, \ldots, x_d) .$$

And, once again, for every $S \subseteq [t]$, M_S denotes the tensor $\sum_{j \in S} M_j$, which has rank at most $|S|$. We proceed via a sequence of inequalities as in the case of $d = 2$ above.

$$\begin{aligned}
\Pr_{x_1,x_2,\ldots,x_d \in \mathbb{F}_2^k}[\forall i \in [t], M_i(x_1, x_2, \ldots, x_d) = 0] &= \mathbb{E}_{x_1,x_2,\ldots,x_d \in \mathbb{F}_2^k}\left[\prod_{i=1}^{t}\left(\frac{1+(-1)^{M_i(x_1,x_2,\ldots,x_d)}}{2}\right)\right] \\
&= \mathbb{E}_{x_1,x_2,\ldots,x_d \in \mathbb{F}_2^k}\left[\frac{1}{2^t}\cdot\sum_{S \subseteq [t]}(-1)^{M_S(x_1,x_2,\ldots,x_d)}\right] \\
&= \mathbb{E}_{x_1,x_2,\ldots,x_d \in \mathbb{F}_2^k}\left[\mathbb{E}_{S \subseteq [t]}\left[(-1)^{M_S(x_1,x_2,\ldots,x_d)}\right]\right] \\
&= \mathbb{E}_{S \subseteq [t]}\left[\mathbb{E}_{x_1,x_2,\ldots,x_d \in \mathbb{F}_2^k}\left[(-1)^{M_S(x_1,x_2,\ldots,x_d)}\right]\right] .
\end{aligned}$$

Now, observe that for every $S \subseteq [t]$,

$$\mathbb{E}_{x_1,x_2,\ldots,x_d \in \mathbb{F}_2^k}\left[(-1)^{M_S(x_1,x_2,\ldots,x_d)}\right] \geq \Pr_{x_2,x_3,\ldots,x_d}\left[\forall j \in S, M_j'(x_2, x_3, \ldots, x_d) = 0\right] .$$

Moreover, from the induction hypothesis, we get that for all $S \subseteq [t]$,

$$\Pr_{x_2,x_3,\ldots,x_d}\left[\forall j \in S, M_j'(x_2, x_3, \ldots, x_d) = 0\right] \geq \left(1 - \frac{1}{2^{d-1}}\right)^{|S|} .$$

Plugging this back in the calculations, we get

$$\begin{aligned}\Pr_{x_1,x_2,\ldots,x_d\in\mathbb{F}_2^k}[\forall i\in[t], M_i(x_1,x_2,\ldots,x_d)=0] &\geq \mathbb{E}_{S\subseteq[t]}\left[\left(1-\frac{1}{2^{d-1}}\right)^{|S|}\right]\\ &\geq \frac{1}{2^t}\cdot\left(1+1-\frac{1}{2^{d-1}}\right)^t=\left(1-\frac{1}{2^d}\right)^t.\end{aligned}$$

□

We now complete the proof of Theorem 4.1.

Proof of Theorem 4.1. Since P has rank $\leq t$, then there is a collection of linear forms $u_1,u_2,\ldots,u_t$ and tensors $M_1,M_2,\ldots,M_t$ of rank at most 1 in $d-1$ dimensions such that

$$P(X_1,X_2,\ldots,X_d)=\sum_{i=1}^{t}u_i(X_1)\cdot M_i(X_2,X_3,\ldots,X_d).$$

Now, observe that

$$\begin{aligned}\mathrm{bias}(P) &= \left|\mathbb{E}_{x_1,x_2,\ldots,x_d\in\mathbb{F}_2^k}\left[(-1)^{P(x_1,x_2,\ldots,x_d)}\right]\right|\\ &= \Pr_{x_2,x_3,\ldots,x_d\in\mathbb{F}_2^k}[P(X_1,x_2,x_3,\ldots,x_d)\equiv 0]\\ &\geq \Pr_{x_2,x_3,\ldots,x_d\in\mathbb{F}_2^k}[\forall i\in[t],\ M_i(x_2,x_3,\ldots,x_d)=0]\\ &\geq \left(1-\frac{1}{2^{d-1}}\right)^t \qquad \text{[By Lemma 4.2]}.\end{aligned}$$

□

We now accompany the above theorem with an almost matching upper bound on the bias of random high rank tensors. It is known that a random high rank tensor has low bias. The following lemma gives a precise quantitative version of this observation (the idea for the proof was suggested to us by Shubhangi Saraf).

Lemma 4.3. *For $i\in[t]$ and $j\in[d]$, let $u_{i,j}\in\mathbb{F}_2^k$ be a uniformly random vector. Consider the random rank-t d-linear form $p:(\mathbb{F}_2^k)^d\to\mathbb{F}_2$ given by*

$$p(x_1,x_2,\ldots,x_d)=\sum_{i=1}^{t}\prod_{j=1}^{d}\langle x_j,u_{i,j}\rangle.$$

Then

$$\mathbb{E}[\mathrm{bias}(p)]\leq d\cdot 2^{-k}+\left(1-\frac{2}{2^d}\right)^t$$

Proof. We have

$$\begin{aligned}
\mathbb{E}_p[\text{bias}(p)] &= \mathbb{E}_p \, \mathbb{E}_{x_1,x_2,\ldots,x_d \in \mathbb{F}_2^k} \left[(-1)^{\sum_{i=1}^t \prod_{j=1}^d \langle x_j, u_{i,j} \rangle} \right] \\
&= \mathbb{E}_{x_1,x_2,\ldots,x_d \in \mathbb{F}_2^k} \, \mathbb{E}_p \left[(-1)^{\sum_{i=1}^t \prod_{j=1}^d \langle x_j, u_{i,j} \rangle} \right] \\
&= \Pr_{x_1,\ldots,x_d} \left[\exists i, x_i = \bar{0} \right] + \Pr_{x_1,\ldots,x_d} \left[\forall i, x_i \neq \bar{0} \right] \cdot \mathbb{E}_{x_1,x_2,\ldots,x_d \in \mathbb{F}_2^k \setminus \{\bar{0}\}} \left[\prod_{i=1}^t \left(\mathbb{E}_{u_{i,1},u_{i,2},\ldots,u_{i,d}} (-1)^{\prod_{j=1}^d \langle x_j, u_{i,j} \rangle} \right) \right] \\
&= 1 - \left(1 - \frac{1}{2^k}\right)^d + \left(1 - \frac{1}{2^k}\right)^d \cdot \mathbb{E}_{x_1,x_2,\ldots,x_d \in \mathbb{F}_2^k \setminus \{\bar{0}\}} \left[\prod_{i=1}^t \left(\Pr_{u_{i,1},\ldots,u_{i,d-1}} \left[\exists j \in [d-1], \langle x_j, u_{i,j} \rangle = 0 \right] \right) \right] \\
&= 1 - \left(1 - \frac{1}{2^k}\right)^d + \left(1 - \frac{1}{2^k}\right)^d \cdot \left(1 - \frac{1}{2^{d-1}}\right)^t \\
&\leq d \cdot 2^{-k} + \left(1 - \frac{2}{2^d}\right)^t .
\end{aligned}$$

□

The following special cases of Theorem 4.1, for $d = 2$ and $d = 3$ will be useful for us, on our way to proving lower bounds on the rank of three dimensional tensors.

Corollary 4.4. *Let $P \in \mathbb{F}_2^{k \times k}$ be a matrix of rank $\leq t \leq k$. Then,* $\text{bias}(P) \geq 2^{-t}$.

Corollary 4.5. *Let $P \in \mathbb{F}_2^{k \times k \times k}$ be a 3-dimensional tensor of rank $\leq t$. Then,* $\text{bias}(P) \geq \left(\frac{3}{4}\right)^t$.

In the subsequent two sections, we will observe that some well-known explicit tensors in three dimensions have very low bias, and then use the above corollaries to conclude that these tensors have large rank.

4.2 A 3.52k Tensor Rank Lower Bound for $\text{Trace}(XYZ)$

In this section, we use the bias-vs-tensor-rank connection explored in the previous section to construct explicit 3-dimensional tensors with large tensor rank. Corollary 4.5 suggests the following natural approach to construct tensors of large rank: find a 3-linear form with as small a bias as possible. What is the least bias of a 3-linear form? Let $P(X, Y, Z) = \sum_{i=1}^k \langle Y, M_i Z \rangle X_i$ be an arbitrary 3-linear form. Clearly, $\text{bias}(P) \geq \Pr_{y,z}[\forall i \in [k], \ \langle y, M_i z \rangle = 0] \geq \Pr_{y,z}[y = \bar{0} \text{ or } z = \bar{0}] = 2/2^k - 1/2^{2k}$. The $\text{Trace}(XYZ)$ is a function with bias exactly $2/2^k - 1/2^{2k}$ (see Lemma 4.6). In the rest of this section, we prove an upper bound on the bias of this function. To this end, we first show that the bias of $Tr(X, Y, Z)$ is small. This will immediately via Corollary 4.5 give a very simple proof that $\text{Trace}(XYZ)$ tensor has rank at least $2.409k$. We remark that a much stronger rank lower-bound of $3.52k$ is known due to Chudnovsky and Chudnovsky [CC88, STV92] and indeed we do a more careful analysis of our ideas to get a new proof of the $3.52k$ lower bound (here too the only property of Tr that is used is that it is of very low bias).

Lemma 4.6.

$$\text{bias}(Tr(X, Y, Z)) = 2 \cdot 2^{-k} - 2^{-2k}.$$

Proof. The trace function satisfies the simple property that for every non-zero $\alpha \in \mathbb{F}_{2^k}$, the linear function $\text{Trace}(\alpha X)$ is unbiased. Hence,

$$\text{bias}(Tr(X, Y, Z)) = \Pr_{x,y \in \mathbb{F}_2^k} [x \cdot y = 0] = 2 \cdot 2^{-k} - 2^{-2k}.$$

□

The above lemma coupled with Corollary 4.5 immediately gives the following lower bound on tensor rank of $Tr(X, Y, Z)$.

Corollary 4.7. $\text{rank}(Tr(X, Y, Z)) \geq (\log_{4/3} 2) \cdot k \geq 2.409k$.

We now strengthen this bound to show a $3.52k$ lower bound on the rank of $Tr(X, Y, Z)$. As we alluded to in earlier discussion, this matches the best known lower bound on the tensor rank of *any* explicit tensor in three dimensions. The proof follows from a more careful use of the ideas already present in the proof of Corollary 4.7. We will need the following well-known rate-distance MRRW tradeoff for linear codes.

Theorem 4.8 ([MRR+77])**.** *Let S be a subspace of dimension at least k of $\mathbb{F}_2^t$, such that every non-zero vector in S has weight at least k. Then, $t \geq 3.52k$.*[1]

Theorem 4.9. *The rank of the tensor $Tr(X, Y, Z)$ is at least $3.52k$.*

Proof. Let the tensor rank of $Tr(X, Y, Z)$ be t. Then there exists t vectors $a_1, a_2, \ldots, a_t \in \mathbb{F}_2^k$ and t rank-1 matrices $M_1, M_2, \ldots, M_t$ such that

$$Tr(X, Y, Z) = \sum_{i=1}^{t} \langle a_i, X \rangle \cdot \langle Y, M_i Z \rangle \ . \tag{12}$$

Let A be the $k \times t$ matrix such that for every $i \in [t]$, the i^{th} column of A equals a_i. Let K be the kernel of A. Clearly, $\dim(K) \geq t - k$. In fact, $\dim(K) = t - k$. To see this, observe that if $\dim(K) \geq t - k + 1$, then by the rank-nullity theorem, $\text{rank}(A) \leq k - 1$. Thus, there is a non-zero $x \in F_2^k$ denoted by x_0 such that for every $i \in [t]$, $\langle a_i, x_0 \rangle = 0$. Thus, $Tr(x_0, Y, Z) \equiv 0$ for a non-zero x_0, which is a contradiction.

From proof of Corollary 4.5, we know that

$$\text{bias}(Tr(X, Y, Z)) = \Pr_{y,z \in \mathbb{F}_2^k}[Tr(X, y, z) = 0] \, .$$

So far we were proving a lower bound on $\Pr_{y,z \in \mathbb{F}_2^k}[Tr(X, y, z) = 0]$ by proving a lower bound on $\Pr_{y,z \in \mathbb{F}_2^k}[\forall i \in [t], \langle y, M_i z \rangle = 0]$. Clearly, this seems to be somewhat lossy since even for a choice of y and z in $\mathbb{F}_2^k$ such that $\langle y, M_i z \rangle \neq 0$ for some $i \in [t]$, it is conceivable that $Tr(X, y, z)$ is identically zero. For this proof, we try to be a bit more careful about this. Note that for every $u \in K \subset \mathbb{F}_2^t$,

$$\sum_{i=1}^{t} u_i \cdot \langle a_i, X \rangle \equiv 0 \, .$$

[1] The MRRW bound for binary codes states that any family of codes with fractional distance δ satisfies $R(\delta) \leq h_2\left(\frac{1}{2} - \sqrt{\delta(1-\delta)}\right)$ where $h_2(x) = x \log_2(1/x) + (1-x)\log_2(1/1-x)$ is the binary entropy function. The above mentioned bound can be obtained from this (see [BD80] for details).

Thus, we have,

$$\begin{aligned}\Pr_{y,z\in\mathbb{F}_2^k}[Tr(X,y,z)=0] &= \sum_{u\in K}\Pr_{y,z\in\mathbb{F}_2^k}[\forall i\in[t],\langle y,M_iz\rangle=u_i]\\ &= \sum_{u\in K}\mathbb{E}_{y,z}\left[\prod_{i\in[t]}\left(\frac{1+(-1)^{\langle y,M_iz\rangle+u_i}}{2}\right)\right]\\ &= \sum_{u\in K}\mathbb{E}_{y,z}\left[\mathbb{E}_{S\subseteq[t]}(-1)^{\langle y,M_Sz\rangle}\cdot(-1)^{\langle u,1_S\rangle}\right].\end{aligned}$$

Here, for every $S\subseteq[t]$, 1_S is the characteristic vector of S in t dimensions, and $M_S=\sum_{i\in S}M_i$. Simplifying further, we get,

$$\Pr_{y,z\in\mathbb{F}_2^k}[Tr(X,y,z)=0]=\mathbb{E}_{S\subseteq[t]}\left[\left(\mathbb{E}_{y,z}(-1)^{\langle y,M_Sz\rangle}\right)\cdot\left(\sum_{u\in K}(-1)^{\langle u,1_S\rangle}\right)\right].$$

Now, we observe that the term $\left(\sum_{u\in K}(-1)^{\langle u,1_S\rangle}\right)=|K|$ if and only if $1_S\in K^\perp$, otherwise it equals zero. Also, from Corollary 4.4, we know that $\left(\mathbb{E}_{y,z}(-1)^{\langle y,M_Sz\rangle}\right)=2^{-\mathrm{rank}M_S}$ is at at least $\max\{2^{-k},2^{-|S|}\}$. Plugging these into the inequality above, we have the following inequality.

$$\begin{aligned}\Pr_{y,z\in\mathbb{F}_2^k}[Tr(X,y,z)=0] &\geq \frac{|K|}{2^t}\cdot\sum_{v\in K^\perp}\max\{2^{-k},2^{-|v|}\} && [\text{Here, } |v| \text{ is the Hamming weight of } v]\\ &\geq \mathbb{E}_{v\in K^\perp}\max\{2^{-k},2^{-|v|}\} && [\text{ Since } |K|\cdot\left|K^\perp\right|=2^t]\end{aligned}$$

Recall that the dimension of $K^\perp$ equals k. Now,

$$\mathbb{E}_{v\in K^\perp}\max\{2^{-k},2^{-|v|}\}=2^{-k}+\mathbb{E}_{v\in K^\perp\setminus\{0^k\}}\max\{2^{-k},2^{-|v|}\}.$$

From Lemma 4.6, we know that the bias of $Tr(X,Y,Z)$ is at most $2\cdot2^{-k}-2^{-2k}$. Thus, it must be the case that $\mathbb{E}_{v\in K^\perp\setminus\{0^k\}}\max\{2^{-k},2^{-|v|}\}\leq(1-2^{-k})\cdot2^{-k}$. But this is possible only if all the vectors in $K^\perp\setminus\{0^k\}$ have weight at least k. In this case, the space $K^\perp$ is a linear subspace of $\mathbb{F}_2^t$ of dimension k such that every non-zero vector in it has Hamming weight at least k. From Theorem 4.8, we get that $t\geq3.52k$. This completes the proof. □

4.3 Lower Bound on the Rank of Matrix Multiplication Tensor

In this section, we obtain a lower bound on the rank of the matrix multiplication tensor by proving an upper bound on its bias. Even though better bounds are known for this tensor, our proof is a fairly straightforward application of our techniques, and we believe this is instructive.

Our main technical observation in this section is the following lemma which gives an upper bound on the bias of $M_n(\overline{X},\overline{Y},\overline{Z})$ as each of the variables take values in $\mathbb{F}_2$.

Lemma 4.10. *The bias of $M_n(\overline{X},\overline{Y},\overline{Z})$ is at most $n\cdot2^{-\frac{3n^2}{4}}$.*

Before proceeding with the proof, we note that Theorem 4.10 and Corollary 4.5 immediately imply a non-trivial lower bound on the tensor rank of M_n.

Theorem 4.11. *The tensor rank of M_n is at least $\frac{3n^2}{4\log_2(4/3)} \geq 1.8n^2$.*

We now prove Lemma 4.10.

Proof of Lemma 4.10. We observe that for any two fixed matrices x, y, the 3-linear form M_n reduces to a linear form in z which is non-zero iff the product of the two matrices x and y is non-zero. Furthermore, given a matrix y, the probability (over x) that the product matrix $x \cdot y$ is zero is exactly $2^{-n\cdot\mathrm{rank}(y)}$. Combining these observations, we have

$$\begin{aligned}
\mathrm{bias}(M_n) &= \Pr_{x,y}\left[x \cdot y = 0_{n\times n}\right] \\
&= \mathbb{E}_y\left[2^{-n\cdot\mathrm{rank}(y)}\right] \\
&= \sum_{r=0}^{n} \Pr_y\left[\mathrm{rank}(y) = r\right] \cdot 2^{-nr} \,.
\end{aligned}$$

To complete the proof, we rely on the following claim, whose proof we defer to the end of this section.

Claim 4.12. *For every $r \in \{0, 1, \ldots, n\}$, the following inequality is true.*

$$\Pr_y\left[\mathrm{rank}(y) = r\right] \leq 2^{-(n-r)^2} \,.$$

From the claim above, we get

$$\begin{aligned}
\mathrm{bias}(M_n) &\leq \sum_{r=0}^{n} 2^{-(n-r)^2 - nr} \\
&\leq \sum_{r=0}^{n} 2^{-n^2 - r^2 + nr} \\
&\leq 2^{-n^2} \sum_{r=0}^{n} 2^{r(n-r)} \\
&\leq 2^{-n^2} n \cdot 2^{n^2/4} \\
&\leq n \cdot 2^{-3n^2/4} \,.
\end{aligned}$$

□

For completeness, we now provide a proof of Claim 4.12. We remark that the following tighter bound is known (see [Kol98, Theorem 3.2.1]).

$$\begin{aligned}
\Pr_y\left[\mathrm{rank}(y) = r\right] &= 2^{-(n-r)^2} \cdot \prod_{i=n-r+1}^{n}\left(1 - \frac{1}{2^i}\right) \cdot \left(\sum_{0\leq i_1 \leq \ldots i_{n-r} \leq r} \frac{1}{2^{i_1+\ldots+i_{n-r}}}\right) \\
&\leq 2^{-(n-r)^2} \cdot \prod_{i=n-r+1}^{n}\left(1 - \frac{1}{2^i}\right) \cdot \prod_{i=1}^{n-r}\left(1 - \frac{1}{2^i}\right)^{-1} \,.
\end{aligned}$$

However, the weaker bound given in the claim suffices for our purposes.

Proof of Claim 4.12. The goal is to upper bound the probability that a uniformly random $n \times n$ matrix y over $\mathbb{F}_2$ has rank equal to r. This probability is upper bounded by the probability that the rows of y are contained within a subspace of dimension r of $\mathbb{F}_2^n$. For any fixed subspace S of dimension equal to r, this event happens with a probability equal to $2^{-n(n-r)}$. The number of subspaces of $\mathbb{F}_2^n$ of dimension equal to r is given by the Gaussian binomial coefficient $\begin{bmatrix} n \\ r \end{bmatrix}_2 = \prod_{i=0}^{r-1} \frac{(2^n - 2^i)}{(2^r - 2^i)} \leq \frac{2^{nr}}{2^{r^2}}$. Thus, by a union bound, we get the following.

$$\Pr_y [\text{rank}(y) = r] \leq \frac{2^{nr}}{2^{r^2}} \cdot 2^{-n(n-r)} = 2^{-(n-r)^2}. \qquad \square$$

Acknowledgements

We would like to thank Suryateja Gavva for helpful discussions. We would like to thank Shubhangi Saraf for suggesting the idea for the proof of Lemma 4.3.

References

[AFT11] BORIS ALEXEEV, MICHAEL A. FORBES, and JACOB TSIMERMAN. *Tensor rank: Some lower and upper bounds*. In *Proc. 26th IEEE Conf. on Comput. Complexity*, pages 283–291. 2011. eccc:2011/TR11-010, doi:10.1109/CCC.2011.28. 3

[BD80] MARK R. BROWN and DAVID P. DOBKIN. *An improved lower bound on polynomial multiplication*. IEEE Trans. Computers, C-29(5):337–340, 1980. doi:10.1109/TC.1980.1675583. 3, 4, 23

[BHL12] IDO BEN-ELIEZER, RANI HOD, and SHACHAR LOVETT. *Random low-degree polynomials are hard to approximate*. Comput. Complexity, 21(1):63–81, 2012. (Preliminary version in *13th RANDOM*, 2009). eccc:2008/TR08-080, doi:10.1007/s00037-011-0020-6. 4, 5

[BK12] ELI BEN-SASSON and SWASTIK KOPPARTY. *Affine dispersers from subspace polynomials*. SIAM J. Comput., 41(4):880–914, 2012. (Preliminary version in *41st STOC*, 2009). eccc:2010/TR10-044, doi:10.1137/110826254. 2

[Blä99] MARKUS BLÄSER. *A $\frac{5}{2}n^2$-lower bound for the rank of $n \times n$ matrix multiplication over arbitrary fields*. In *Proc. 40th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 45–50. 1999. doi:10.1109/SFFCS.1999.814576. 3, 4

[CC88] DAVID V. CHUDNOVSKY and GREGORY V. CHUDNOVSKY. *Algebraic complexities and algebraic curves over finite fields*. J. Complexity, 4(4):285–316, 1988. doi:10.1016/0885-064X(88)90012-X. 3, 4, 8, 22

[CKSV16] SURYAJITH CHILLARA, MRINAL KUMAR, RAMPRASAD SAPTHARISHI, and V. VINAY. *The chasm at depth four, and tensor rank : Old results, new insights*, 2016. arXiv:1606.04200, eccc:2016/TR16-096. 2

[EGOW18] KLIM EFREMENKO, ANKIT GARG, RAFAEL MENDES DE OLIVEIRA, and AVI WIGDERSON. *Barriers for rank methods in arithmetic complexity*. In ANNA KARLIN, ed., *Proc. 9th Innovations in Theor. Comput. Sci. (ITCS)*, volume 94 of *LIPIcs*, pages 1:1–1:19. Schloss Dagstuhl, 2018. arXiv:1710.09502, eccc:2017/TR17-162, doi:10.4230/LIPIcs.ITCS.2018.1. 3

[Hås90] JOHAN HÅSTAD. *Tensor rank is np-complete*. J. Algorithms, 11(4):644–654, 1990. (Preliminary version in *16th ICALP*, 1989). doi:10.1016/0196-6774(90)90014-6. 3

[Kam05] MICHAEL KAMINSKI. *A lower bound on the complexity of polynomial multiplication over finite fields*. SIAM J. Comput., 34(4):960–992, 2005. (Preliminary version in *22nd STACS*, 2005). doi:10.1137/S0097539704442118. 3

[Kol98] VALENTIN K. KOLCHIN. *Random Graphs*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1998. doi:10.1017/CBO9780511721342. 25

[MRR+77] ROBERT J. MCELIECE, EUGENE R. RODEMICH, HOWARD RUMSEY, LLOYD, and R. WELCH. *New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities*. IEEE Trans. Inform. Theory, 23(2):157–166, 1977. doi:10.1109/TIT.1977.1055688. 23

[NW94] NOAM NISAN and AVI WIGDERSON. *Hardness vs. randomness*. J. Comput. Syst. Sci., 49(2):149–167, October 1994. (Preliminary version in *29th FOCS*, 1988). doi:10.1016/S0022-0000(05)80043-1. 2

[Raz87] ALEXANDER A. RAZBOROV. Нжние оценки размера схем ограниченной глубины в полном базисе, содержащем функцию логического сложения *(Russian) [Lower bounds on the size of bounded depth circuits over a complete basis with logical addition]*. Mathematicheskie Zametki, 41(4):598–607, 1987. (English translation in *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987). doi:10.1007/BF01137685. 2

[Raz13] RAN RAZ. *Tensor-rank and lower bounds for arithmetic formulas*. J. ACM, 60(6):40:1–40:15, 2013. (Preliminary version in *42nd STOC*, 2010). eccc:2010/TR10-002, doi:10.1145/2535928. 2

[Shp03] AMIR SHPILKA. *Lower bounds for matrix product*. SIAM J. Comput., 32(5):1185–1200, 2003. (Preliminary version in *42nd FOCS*, 2001). arXiv:cs/0201001, eccc:2001/TR01-060, doi:10.1137/S0097539702405954. 3, 4

[Smo87] ROMAN SMOLENSKY. *Algebraic methods in the theory of lower bounds for Boolean circuit complexity*. In *Proc. 19th ACM Symp. on Theory of Computing (STOC)*, pages 77–82. 1987. doi:10.1145/28395.28404. 2

[Str73] VOLKER STRASSEN. *Die berechnungskomplexität von elementarsymmetrischen funktionen und von interpolationskoeffizienten (German) [The computational complexity of elementary symmetric functions and interpolation coefficients)*. Numerische Mathematik, 20(3):238–251, June 1973. doi:10.1007/BF01436566. 2

[STV92] IGOR E. SHPARLINSKI, MICHAEL A. TSFASMAN, and SERGE G. VLADUT. *Curves with many points and multiplication in finite fields*. In HENNING STICHTENOTH and MICHAEL A. TSFASMAN, eds., *Proc. Int. Workshop on Coding Theory and Algebraic Geometry*, volume 1518 of *LNM*, pages 145–169. Springer, 1992. doi:10.1007/BFb0087999. 3, 4, 8, 22

[Vio09] EMANUELE VIOLA. *On the power of small-depth computation*. Foundations and Trends in Theoretical Computer Science, 5(1):1–72, 2009. doi:10.1561/0400000033. 2

[VW08] EMANUELE VIOLA and AVI WIGDERSON. *Norms, XOR lemmas, and lower bounds for polynomials and protocols*. Theory Comput., 4(1):137–168, 2008. (Preliminary version in *22nd CCC*, 2007). doi:10.4086/toc.2008.v004a007. 2

A A maximization problem

In this section, we prove Theorem 3.10. We start with restating it here.

Theorem A.1 (Restatement of Theorem 3.10)**.** *Let k be a positive integer, and let $u \in [0, k^2]$ be a real number. Suppose $b_1, b_2, \ldots, b_k$ are real numbers satisfying the following constraints.*

$$k \geq b_1 \geq b_2 \ldots \geq b_k \geq 0, \tag{13}$$

$$\sum_{i=1}^{k} b_i = u. \tag{14}$$

Then,

$$\sum_{i=1}^{k} 2^{i-1} 2^{b_i} \leq \sum_{i=1}^{k} 2^{i-1} 2^{u/k} = (2^k - 1) 2^{u/k}.$$

Proof. Let $\mathcal{P}$ denote the convex polytope defined as follows.

$$\mathcal{P} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid k \geq x_1 \geq \ldots \geq x_k \geq 0 \text{ and } \sum_i x_i = u\}.$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be the function:

$$\sum_{i=1}^{k} 2^{i-1} 2^{x_i}.$$

Observe that $\mathcal{P}$ is bounded and nonempty, and f is a convex function. Thus the maximum M of f on $\mathcal{P}$ is achieved at an extreme point. Since $\mathcal{P}$ is defined by $k+1$ inequalities and 1 equality, extreme points satisfy the 1 equality and make at least $k-1$ of the inequalties tight. Thus any extreme point $(y_1, y_2, \ldots, y_k)$ of $\mathcal{P}$ satisfies, for some integers $a, b, c \geq 0$ with $a + b + c = k$, and some $\ell \in (0, k)$, the following equalities.

$$y_1 = y_2 = \ldots = y_a = k,$$

$$y_{a+1} = y_{a+2} = \ldots = y_{a+b} = \ell,$$

$$y_{a+b+1} = y_{a+b+2} = \ldots = y_k = 0.$$

$$ak + b\ell = u.$$

At such an extreme point $(y_1, \ldots, y_k)$, the value of f can be expressed in terms of a, b, c, ℓ as

$$\begin{aligned} f(y_1, \ldots, y_k) &= 2^k(2^a - 1) + 2^\ell 2^a (2^b - 1) + 2^0 2^{a+b}(2^c - 1) \\ &= 2^a(2^k - 2^\ell) + 2^{a+b}(2^\ell - 1). \end{aligned}$$

The following lemma then completes the proof of the theorem.

Lemma A.2. *Let k be a positive integer, and let $u \in [0, k^2]$ be a real number. Let $a, b, c, \ell \in [0, k]$ be real numbers with:*

$$a + b + c = k,$$

$$ak + b\ell = u.$$

Then

$$2^a(2^k - 2^\ell) + 2^{k-c}(2^\ell - 1) \leq (2^k - 1)2^{u/k}.$$

Proof. Let $\alpha, \beta, \gamma, \lambda, \eta \in [0,1]$ be given by

$$a = \alpha k, b = \beta k, c = \gamma k, \ell = \lambda k, u = \eta k^2.$$

Then, we have

$$\alpha + \beta + \gamma = 1, \tag{15}$$
$$\alpha + \beta\lambda = \eta. \tag{16}$$

Let $Z = 2^k$. Then we want to show that whenever $\alpha, \beta, \gamma, \lambda, \eta$ are as above, we have

$$Z^\alpha(Z - Z^\lambda) + Z^{1-\gamma}(Z^\lambda - 1) \leq (Z-1)Z^\eta.$$

Eliminating α, γ from Equation (15) and Equation (16), we have $\gamma = (1-\eta) - \beta(1-\lambda)$, and $\alpha = \eta - \beta\lambda$. Substituting this in, we want to show that

$$Z^{\eta-\beta\lambda}(Z - Z^\lambda) + Z^{\eta+\beta(1-\lambda)}(Z^\lambda - 1) \leq (Z-1)Z^\eta.$$

Dividing throughout by Z^η, we want to show that

$$Z^{-\beta\lambda}(Z - Z^\lambda) + Z^{\beta(1-\lambda)}(Z^\lambda - 1) \leq Z - 1.$$

Rewriting, this is the same as

$$Z - Z^\lambda + Z^{\beta+\lambda} - Z^\beta \leq (Z-1)Z^{\beta\lambda},$$

which is equivalent to

$$(Z^\beta - 1)(Z^\lambda - 1) \leq (Z^{\beta\lambda} - 1)(Z - 1).$$

This follows from Lemma B.2. □

This completes the proof. □

B Numerical Inequalities

In this section we list some numerical inequalites that are used in the previous section.

Lemma B.1. *For all real $r \geq 1$, the function $f : [1, \infty) \to \mathbb{R}$ given by*

$$f(x) = \frac{x^r - 1}{x - 1}$$

is increasing in x.

Proof. We show that $f'(x) \geq 0$ for all $x \geq 1$. Compute

$$f'(x) = \frac{(rx^{r-1}) \cdot (x-1) - (x^r - 1) \cdot 1}{(x-1)^2}$$

Define

$$g(x) = r(x^r - x^{r-1}) - (x^r - 1) = (r-1)x^r - rx^{r-1} + 1.$$

The positivity of $f'(x)$ would follow if we can show:

$$g(x) \geq 0$$

for all $x \geq 1$. We prove this by first observing that $g(1) = 0$, and then showing that for all $x \geq 1$, we have $g'(x) \geq 0$. Indeed,

$$\begin{aligned} g'(x) &= r(r-1)x^{r-1} - r(r-1)x^{r-2} \\ &= r(r-1)(x^{r-2})(x-1) \\ &\geq 0. \end{aligned}$$

This completes the proof that $g(x) \geq 0$ for all $x \geq 1$, and thus the proof that $f'(x) \geq 0$ for all $x \geq 1$. □

Lemma B.2. *For all real $z \geq 1$ and all real $\beta, \lambda \in [0,1]$, we have:*

$$(z^{\lambda} - 1)(z^{\beta} - 1) \leq (z^{\beta\lambda} - 1)(z - 1).$$

Proof. If either $\lambda = 0$ or $z = 1$, the inequality trivially holds (with equality). Now suppose $\lambda \neq 0$ and $y \neq 1$. Set $r = 1/\lambda$ and $x = z^{\lambda}$ and $y = z^{\beta\lambda}$. Then, $1 \leq y \leq x$ and $r \geq 1$. Then, the inequality we want to prove can be written as follows.

$$(x-1)(y^r - 1) \leq (y-1)(x^r - 1),$$

i.e.,

$$\frac{y^r - 1}{y - 1} \leq \frac{x^r - 1}{x - 1}.$$

This follows from Lemma B.1, completing the proof. □