Electronic Colloquium on Computational Complexity, Report No. 25 (2016)

The Fourier structure of low degree polynomials

Shachar Lovett*
Computer Science and Engineering
University of California, San Diego
slovett@ucsd.edu

February 26, 2016

Abstract

We study the structure of the Fourier coefficients of low degree multivariate polynomials over finite fields. We consider three properties: (i) the number of nonzero Fourier coefficients; (ii) the sum of the absolute value of the Fourier coefficients; and (iii) the size of the linear subspace spanned by the nonzero Fourier coefficients. For quadratic polynomials, tight relations are known between all three quantities. In this work, we extend this relation to higher degree polynomials. Specifically, for degree d polynomials, we show that the three quantities are equivalent up to factors exponential in d.

1 Introduction

Low degree polynomials play an important role in mathematics and computer science, as well as Fourier analysis. In this paper, we study the structure of the Fourier coefficients of low degree multivariate polynomials over finite fields. Let us first state our main result, before providing background and motivation. Let $\mathbb{F}$ be a finite field, $e : \mathbb{F} \to \mathbb{C}$ a nontrivial additive character, and let $f : \mathbb{F}^n \to \mathbb{F}$ be a polynomial of total degree d. We show that the following three quantities are equal, up to a factor of 16^d:

- The number of nonzero Fourier coefficients of $e(f)$.
- The sum of the absolute values of the Fourier coefficients of $e(f)$ (namely, it's L_1 Fourier norm).
- The size of the linear space spanned by the nonzero Fourier coefficients of $e(f)$.

*Research supported by NSF CAREER award 1350481.

ISSN 1433-8092

Rank of polynomials. Various notions of rank for polynomials have been considered in the literature, motivated by different applications.

The *rank* of a polynomial is the minimal number of lower degree polynomials required to compute it. Here and throughout, by degree we mean total degree. Let $\mathbb{F}$ denote a finite field.

Definition 1.1 (Rank)**.** *Let $f : \mathbb{F}^n \to \mathbb{F}$ be a polynomial of degree d. The* rank *of f is the minimial $r \geq 1$, such that there exist r polynomials $g_1, \ldots, g_r : \mathbb{F}^n \to \mathbb{F}$ of degree $\leq d-1$, and a function $\Gamma : \mathbb{F}^r \to \mathbb{F}$ such that*

$$f(x) = \Gamma(g_1(x), \ldots, g_r(x)).$$

The property that a polynomial has low rank is equivalent to the existence of a lower degree polynomial which approximates it non-trivially [6, 9, 13]. This is a key ingredient in higher-order Fourier analysis, a theory introduced by Gowers [8] which generalizes classical Fourier analysis, and which found applications in several domains, including number theory, coding theory, property testing and complexity theory [1–5, 7, 11, 12, 15–19, 21].

A current caveat of this theory is that the bounds it produces have horrible dependency on the degree d (Ackerman-type bounds). Thus, it makes sense to consider more restricted notions of rank, which may apply for polynomials of higher degree. One exception is the work of Haramaty and Shpilka [10] on the structure of cubic and quartic polynomials, which achieves quasi-polynomial bounds on the underlying parameters.

A more refined notion of rank, termed *linear rank*, was introduced by Tsang et al. [20], motivated by a potential approach towards the log-rank conjecture in communication complexity, for some special families of functions (XOR functions).

Definition 1.2 (Linear rank)**.** *Let $f : \mathbb{F}^n \to \mathbb{F}$ be a polynomial of degree d. The* linear rank *of f is the minimial $r \geq 1$, such that there exist r linear functions $\ell_1, \ldots, \ell_r : \mathbb{F}^n \to \mathbb{F}$ and $r+1$ polynomials $g_0, g_1, \ldots, g_r : \mathbb{F}^n \to \mathbb{F}$ of degree $\leq d-1$, such that*

$$f(x) = g_0(x) + \ell_1(x)g_1(x) + \ldots + \ell_r(x)g_r(x).$$

Tsang et al. [20] proved a relation between the linear rank of a polynomial and its Fourier coefficients, for $\mathbb{F} = \mathbb{F}_2$. The Fourier coefficents of $f : \mathbb{F}_2^n \to \mathbb{F}_2$ are given by

$$\widehat{(-1)^f}(\gamma) = \mathbb{E}_{x \in \mathbb{F}_2^n} \left[(-1)^{f(x) - \langle x, \gamma \rangle} \right],$$

where $\gamma \in \mathbb{F}_2^n$. The L_1 *spectral norm* of f is

$$\|\widehat{(-1)^f}\|_1 = \sum_{\gamma \in \mathbb{F}_2^n} |\widehat{(-1)^f}(\gamma)|.$$

We shorthand here $\|\widehat{f}\|_1 := \|\widehat{(-1)^f}\|_1$.

Theorem 1.3 ([20])**.** *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a polynomial of degree d. The linear rank of f is at most $O(2^{d^2/2} \log^{d-2} \|\widehat{f}\|_1)$.*

An even more refined notion of structure is the number of variables that a polynomial depends on, possibly after a change of basis. This notion makes sense for any function, not necessarily a low degree polynomial.

Definition 1.4 (Linear dimension). *Let $f : \mathbb{F}^n \to \mathbb{F}$. The* linear dimension *of f is the minimal $r \geq 1$, such that f depends on r linear functions of the inputs. Equivalently, there exists an invertible linear change of basis $\varphi : \mathbb{F}^n \to \mathbb{F}^n$ such that for $g(x) = f(\varphi(x))$ it holds that*

$$g(x) = g(x_1, \ldots, x_r).$$

A corollary of Theorem 1.3 is that low degree polynomials with bounded spectral norm have low linear dimension.

Theorem 1.5 ([20]). *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a polynomial of degree d. The linear dimension of f is at most $O(2^{d^3/2} \log^{d^2} \|\widehat{f}\|_1)$.*

Our contribution. Our main theorem improves upon both Theorem 1.3 and Theorem 1.5. We prove tight relations between the L_1 spectral norm of polynomials and their linear dimension. We also extend the theorem for any finite field, not just $\mathbb{F}_2$. Let $\mathbb{F}$ be a finite field. An additive character $e : \mathbb{F} \to \mathbb{C}$ is a nonzero function which satisfies $e(x + y) = e(x)x(y)$. It is trivial if $e = 1$ and nontrivial otherwise. For example, if $\mathbb{F} = \mathbb{F}_p$ is a prime finite field, the characters are $e_a(x) = \exp(2\pi i a x/p)$ for $a = 0, \ldots, p - 1$.

Theorem 1.6 (Main theorem). *Let $\mathbb{F}$ be a finite field, $e : \mathbb{F} \to \mathbb{C}$ a nontrivial additive character. Let $f : \mathbb{F}^n \to \mathbb{F}$ be a polynomial of degree $d \geq 2$. The linear dimension of f is at most $16^d \log_{|\mathbb{F}|} \|\widehat{e(f)}\|_1$.*

Note that for $d = 1$ we have $\|\widehat{e(f)}\|_1 = 1$ and f has linear dimension 1. We observe that the bound in Theorem 1.6 is essentially tight over small fields. For simplicity of exposition, we consider $\mathbb{F}_2$.

Example 1.7. *Fix $s \in \mathbb{N}$. We construct a sequence of polynomials $f_d : \mathbb{F}_2^{n_d} \to \mathbb{F}_2$, where f_d has degree $2d$, log spectral norm $\approx ds$, and linear dimension $\approx 2^d s$.*

Let $n_1 = 2s$, $f_1(x) = x_1x_2 + x_3x_4 + \ldots + x_{2s-1}x_{2s}$. Then f_1 has linear dimension $2s$, all its 2^{2s} Fourier coefficients equal to $\pm 2^{-s}$, and hence $\|\widehat{e(f_1)}\|_1 = 2^s$. Define inductively $n_d = 2n_{d-1} + 2s$, $f_d : \mathbb{F}_2^{n_d} \to \mathbb{F}_2$ as follows:

$$f_d(x', x'', x''') = f_1(x')f_{d-1}(x'') + (1 - f_1(x'))f_{d-1}(x''')$$

where $x' \in \mathbb{F}_2^{2s}$ and $x'', x''' \in \mathbb{F}_2^{n_{d-1}}$ are disjoint variables. One can verify inductively that f_d has linear dimension $n_d \geq 2^d s$ and that

$$\|\widehat{(-1)^{f_d}}\|_1 \leq 2(2^s + 1)^d + 1.$$

We state an immediate corollary of Theorem 1.6, relating the sparsity, norm and dimensionality of the Fourier coefficients of low degree polynomials.

Corollary 1.8. *Let $\mathbb{F}$ be a finite field, $e : \mathbb{F} \to \mathbb{C}$ a nontrivial additive character. Let $f : \mathbb{F}^n \to \mathbb{F}$ be a polynomial of degree $d \geq 2$. Let $s = \|\widehat{e(f)}\|_1$. Then*

(i) $e(f)$ has at most s^{16^d} nonzero Fourier coefficients.

(ii) They are supported on an affine subspace of dimension $16^d \log_{|\mathbb{F}|} s$.

(iii) Let $e' : \mathbb{F} \to \mathbb{C}$ be any other additive character. Then $\|\widehat{e'(f)}\|_1 \leq s^{16^d}$.

Relations to communication complexity The motivation for introducing the notion of linear rank in [20] is related to the log-rank conjecture [14], a famous open problem in communication complexity, asking about the relating between the rank of a matrix and the best deterministic protocol for computing it. In [20], they focus on so-called "XOR functions", an interesting sub-case of the log rank conjecture.

Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$. It defines the following communication problem: there are two players, Alice and Bob. Alice is given $x \in \mathbb{F}_2^n$ and Bob is given $y \in \mathbb{F}_2^n$ as inputs. Their goal is to compute $f(x \oplus y)$ while communicating as few bits as possible. The associated matrix with this problem is the $2^n \times 2^n$ matrix $M_{x,y} = f(x \oplus y)$. The log rank conjecture speculates that up to polynomial factors, log of the rank of M is an upper bound on the deterministic communication complexity of the problem. In the context of XOR functions, it turns out that having an efficient protocol would follow from a large subspace on which f is constant. Thus, we get the following, possibly simpler, problem: if f has only s nonzero Fourier coefficients, is it always true that there exists a subspace $V \subset \mathbb{F}_2^n$ of co-dimension $(\log s)^{O(1)}$ such that $f|_V$ is constant?

The work of [20] focused on the case where f is additionally assumed to have a low degree as a polynomial over $\mathbb{F}_2$. If it's degree is d, Theorem 1.5 provides such a subspace of co-dimension $O(2^{d^2/2}(\log s)^{d-2})$ on which f is constant, where s can be taken to be the L_1 spectral norm of f (which is an upper bound on the Fourier sparsity of f). This gives a classical deterministic protocol which sends $O(2^{d^2/2}(\log s)^{d-2})$ many bits. Subsequently, Zhang [22] gave an improved quantum protocol which uses only $O(2^d \log s)$ quantum bits. This still leaves open the question of finding an improved deterministic protocol.

Here, we note that such a protocol follows as a direct application of our main result. By Theorem 1.6, if the L_1 spectral norm of f is s (or even better, if the Fourier sparsity of f is s), then f depends on at most $16^d \log s$ many inputs (after an appropriate change of basis). So, after this change of basis (which is known in advance to both players), each player can simply send the relevant bits of their input to the other player. This protocol is a classical deterministic one-round protocol which sends $O(16^d \log s)$ bits.

1.1 Proof overview

Let $f : \mathbb{F}^n \to \mathbb{F}$ be a polynomial of degree d. Let f_d be the homogeneous part of f of degree d. The main idea is to bound the linear dimension of f_d, use this to remove f_d from f and reduce to a polynomial of degree $d-1$, and continue inductively. In order to isolate f_d we apply derivatives to f and obtain the derivative polynomial of f, which is a symmetric set-multilinear polynomial (also known as a symmetric tensor) which symmetrizes f_d. In order to study it, we develop a general theory for the linear dimension of tensors and symmetric tensors.

A tensor $T : (\mathbb{F}^n)^d \to \mathbb{F}$ is a multi-linear map. We say that $T(x^1, \ldots, x^d)$ has linear dimension r if it only depends on at most r inputs out of each of $x^1, \ldots, x^r$, possibly after some change of basis. We show that linear dimension behaves well under restrictions. Let $T|_{x^i=a}$ be the restricted tensor obtained by fixing $x^i = a$ for some $i \in [d], a \in \mathbb{F}^n$. Clearly, if T has linear dimension d then also all of $T|_{x^i=a}$ have linear dimension at most d. We show (Theorem 4.2) that the inverse relation also holds: if all of $T|_{x^i=a}$ have linear dimension $\leq d$, then T has linear dimension $\leq 4d$. We also prove (Theorem 4.4) a version specialized for symmetric tensors, such as the ones obtained by the derivatives of a polynomial.

The proofs of Theorem 4.2 and Theorem 4.4 follow from a general theorem about linear spaces of functions of low linear dimension (we call such functions "linear-juntas"). We show (Theorem 3.5) that any such linear space must be very structured, in a way that explains why all the functions in it have low linear dimension.

Paper organization. We start with some preliminaries in Section 2. We then develop a theory of subspaces of linear juntas in Section 3. We apply these to the study of the linear dimension of tensors in Section 4. We apply these to the study of the Fourier structure of polynomials in Section 5.

Acknowledgements. We thank the organizers of the workshop "log rank conjecture" held in the National University of Singapore in January 2016 where this work started. We specifically thank Shengyu Zhang, from whom we learned about this problem, and who gave us helpful suggestions on an earlier version of this manuscript.

2 Preliminaries

Polynomials. Let $\mathbb{F}$ be a field, A a finite dimensional linear space over $\mathbb{F}$. A polynomial $f : \mathbb{F}^n \to A$ is any function of the form

$$f(x) = \sum_{I \in \mathbb{N}^n} f_I \prod_{i=1}^{n} x_i^{I_i},$$

where $x = (x_1, \ldots, x_n) \in \mathbb{F}^n$, $f_I \in A$ and only a finite number of f_I are nonzero. We denote by $\mathrm{Poly}(\mathbb{F}^n, A)$ the set of all such polynomials, which is a $\mathbb{F}$-linear space. Note that if $\mathbb{F}$ is a finite field, then $\mathrm{Poly}(\mathbb{F}^n, A)$ includes all functions $f : \mathbb{F}^n \to A$.

The (total) degree of a polynomial is $\deg(f) = \max_{f_I \neq 0} \sum_i I_i$. We denote by $\mathrm{Poly}_d(\mathbb{F}^n, A)$ the linear space of polynomials of degree at most d.

Fourier analysis. Let p be prime, $q = p^k$, $\mathbb{F} = \mathbb{F}_q$ be a finite field. Let $\mathrm{Tr} : \mathbb{F}_q \to \mathbb{F}_p$ be the trace map. The additive characters $e : \mathbb{F}^n \to \mathbb{C}$ are nonzero functions which satisfy $e(x + y) = e(x)e(y)$. They are given by

$$e_\gamma(x) = \omega_p^{\mathrm{Tr}(\langle \gamma,x\rangle)} \qquad \gamma \in \mathbb{F}^n.$$

where $\omega_p = \exp(2\pi i/p)$ is a primitive p-th root of unity and $\langle x, \gamma\rangle = \sum x_i\gamma_i$. The trivial character is $e_0 \equiv 1$; the rest are called nontrivial. The Fourier coefficients of $f : \mathbb{F}^n \to \mathbb{C}$ are given by

$$\widehat{f}(\gamma) = \mathbb{E}_{x\in\mathbb{F}^n}\left[f(x)\overline{e_\gamma(x)}\right].$$

The Fourier inversion formula is

$$f(x) = \sum_{\gamma\in\mathbb{F}^n} \widehat{f}(\gamma)e_\gamma(x).$$

Parseval's identity is

$$\mathbb{E}_{x\in\mathbb{F}^n}|f(x)|^2 = \sum_{\gamma\in\mathbb{F}^n} |\widehat{f}(\gamma)|^2.$$

The L_1 spectral norm of f is

$$\|\widehat{f}\|_1 = \sum_{\gamma\in\mathbb{F}^n} |\widehat{f}(\gamma)|.$$

We state two simple claims regarding the behaviour of the L_1 spectral norm under restriction and multiplication.

Claim 2.1. *Let $f : \mathbb{F}^n \to \mathbb{C}$. For some $m < n$ let $g : \mathbb{F}^m \to \mathbb{C}$ be obtained by fixing $n - m$ inputs of f, say $g(x_1, \dots, x_m) = f(x_1, \dots, x_m, c_{m+1}, \dots, c_n)$ for some $c_i \in \mathbb{F}$. Then $\|\widehat{g}\|_1 \le \|\widehat{f}\|_1$.*

Claim 2.2. *Let $f, g : \mathbb{F}^n \to \mathbb{C}$ and define $h(x) = f(x)g(x)$. Then $\|\widehat{h}\|_1 \le \|\widehat{f}\|_1\|\widehat{g}\|_1$.*

3 Linear spaces of linear juntas

Let $\mathbb{F}$ be a field, $f : \mathbb{F}^n \to \mathbb{F}$ be a function. A well studied notion is that of a *junta*: a function is said to be a d-junta if it depends on at most d of its inputs. Here, we define the notion of a linear junta. A function is a d-linear junta if it is a d-junta, after applying some change of basis.

Definition 3.1 (Linear junta). *A function $f : \mathbb{F}^n \to \mathbb{F}$ is a d-linear junta if it depends on at most d inputs, possibly after a change of basis. Equivalently, if there exist d linear functions $\ell_1, \dots, \ell_d : \mathbb{F}^n \to \mathbb{F}$ such that $f(x)$ is determined by $\ell_1(x), \dots, \ell_d(x)$. The* linear dimension *of f is the smallest such d, which we denote by LinDim(f).*

Our interest here will be in a collection of linear juntas which form a linear space. For technical reasons, we restricted our attentions to polynomials. We note that this is a restriction only when $\mathbb{F}$ is an infinite field.

Definition 3.2 (Linear space of linear juntas)**.** *A collection of functions $\Lambda \subset Poly(\mathbb{F}^n, \mathbb{F})$ is said to be a* linear space of d-linear juntas *if*

(i) Each function $f \in \Lambda$ is a d-linear junta.

(ii) Λ is a linear subspace of $Poly(\mathbb{F}^n, \mathbb{F})$. That is, if $f, g \in \Lambda$ then also $\alpha f + \beta g \in \Lambda$ for all $\alpha, \beta \in \mathbb{F}$.

It will be instructive to consider a couple of examples for subspaces of linear d-juntas.

Example 3.3. *Let $\Lambda = \{f(x) = f(x_1, \ldots, x_d) : f \in Poly(\mathbb{F}^d, \mathbb{F})\}$ be the space of all functions that depends on the first d inputs. It is a linear space of d-linear juntas.*

Example 3.4. *Let $\Lambda = \{f(x) = \langle a, x \rangle : a \in \mathbb{F}^n\}$ where $\langle a, x \rangle = \sum a_i x_i$ be the space of all linear functions from $\mathbb{F}^n$ to $\mathbb{F}$. It is a linear space of 1-linear juntas.*

Our main theorem in this section is that these examples are more or less exhaustive. For any linear space of d-linear juntas, after an appropriate change of basis, any fixing of the first $2d$ inputs results in all functions becoming linear functions. For the proof we would need to study functions with multiple outputs, and extend the previous definitions to such functions.

Let $A \cong \mathbb{F}^m$ be a linear subspace over $\mathbb{F}$. Typically we would not care about the dimension of A. Two functions $f, f' : \mathbb{F}^n \to A$ are said to be isomorphic, denoted $f \equiv f'$, if they are equal up to a change of basis in their inputs. That is, $f \equiv f'$ if there exists an invertible linear map $\varphi : \mathbb{F}^n \to \mathbb{F}^n$ such that $f'(x) = f(\varphi(x))$. A function $f : \mathbb{F}^n \to A$ is a d-junta if it depends on at most d of its inputs, and it is a d-linear junta if it is isomorphic to a d-junta. Let $\mathrm{Poly}(\mathbb{F}^n, A)$ denote the $\mathbb{F}$-linear subspace of all functions from $\mathbb{F}^n$ to A. Given two linear spaces $\Lambda, \Lambda' \subset \mathrm{Poly}(\mathbb{F}^n, A)$, we say they are isomorphic, denoted $\Lambda \equiv \Lambda'$, if they are equal up to a change of input basis. That is, if there exists an invertible linear map $\varphi : \mathbb{F}^n \to \mathbb{F}^n$ such that $\Lambda' = \{f(\varphi(x)) : f \in \Lambda\}$. A function $f : \mathbb{F}^n \to A$ is a *linear function* if $f(x) = \sum a_i x_i$ for some $a_i \in A$.

Theorem 3.5. *Let $\mathbb{F}$ be a field, A a linear subspace over $\mathbb{F}$. Let $\Lambda \subset Poly(\mathbb{F}^n, A)$ be a linear subspace of d-linear juntas. Then there exists an isomorphic subspace $\Lambda' \equiv \Lambda$ such that, for any fixing of the first $2d$ inputs, the functions in Λ' become linear functions. That is,*

$$\{f(c_1, \ldots, c_{2d}, x_{2d+1}, \ldots, x_n) : \mathbb{F}^{n-2d} \to A : f \in \Lambda', c_1, \ldots, c_{2d} \in \mathbb{F}\}$$

are all linear functions from $\mathbb{F}^{n-2d}$ to A.

3.1 Proof of Theorem 3.5

We prove Theorem 3.5 in this section. We use the following notation: for $x \in \mathbb{F}^n$ and $S \subset [n]$ let $x_S = (x_i : i \in S)$. For $c \in \mathbb{F}^S$ we denote by $f|_{x_S=c}$ the function f restricted to inputs with $x_S = c$. We denote $[r] = \{1, \dots, r\}$.

Claim 3.6. *The linear dimension of $f : \mathbb{F}^n \to A$ is equal to the co-dimension of the subspace $\mathcal{O}_f \subset \mathbb{F}^n$ given by the invariant shifts of f,*

$$\mathcal{O}_f := \{v \in \mathbb{F}^n : f(x) = f(x+v) \ \forall x \in \mathbb{F}^n\}.$$

Proof. Assume $\mathrm{LinDim}(f) = d$. As a change of basis does not change the dimension of $\mathcal{O}_f$, we may assume that $f(x) = f(x_1, \dots, x_d)$. Thus $\{0\}^d \times \mathbb{F}^{n-d} \subset \mathcal{O}_f$. Moreover, if $v \notin \{0\}^d \times \mathbb{F}^{n-d}$ then by the minimality of d, $f(x) \neq f(x+v)$ for some $x \in \mathbb{F}^n$. Thus $\dim(\mathcal{O}_f) = n - d$. □

Lemma 3.7. *Fix $n \geq 2d$. Let $f, g \in Poly(\mathbb{F}^n, A)$ be functions such that*

(i) $LinDim(f) = d$. Moreover, $f(x) = f(x_1, \dots, x_d)$.

(ii) $LinDim(g), LinDim(f+g) \leq d$.

Define $\tilde{g} : \mathbb{F}^{n-d} \to Poly(\mathbb{F}^d, A)$ by

$$\tilde{g}(x_{d+1}, \dots, x_n) = \left((c_1, \dots, c_d) \to g(c_1, \dots, c_d, x_{d+1}, \dots, x_n) : c \in \mathbb{F}^d\right).$$

Then we can decompose $\tilde{g} = \tilde{g}_{lin} + \tilde{g}_{rest}$, where

(a) $\tilde{g}_{lin} : \mathbb{F}^{n-d} \to Poly(\mathbb{F}^d, A)$ is a linear function.

(b) $LinDim(\tilde{g}_{rest}) \leq d/2$.

We note that the proof of Lemma 3.7 is inspired by the proof of Lemma 3.7 in the arxiv version of [10] (the fact that the lemmas numbering match is an interesting coincidence).

Proof. The conditions and conclusions of the lemma do not change if we apply an invertible change of basis to $x_1, \dots, x_d$ or to $x_{d+1}, \dots, x_n$. Thus, by applying an appropriate change of basis, we may assume that $g(x) = g(x_{s+1}, \dots, x_{s+d})$ for some $0 \leq s \leq d$ (as we only assume that $\mathrm{LinDim}(g) \leq d$, it may be the case that g does not depend on all of these inputs; still, for simplicity of exposition, we list all the d inputs). As non of $f, g, f+g$ depend on $x_{s+d+1}, \dots, x_n$, we may set all of them to zero; to simplify notations, simply assume from now that $n = s + d$.

Decompose $x \in \mathbb{F}^n$ as $x = (x', x'', x''')$ where $x' \in \mathbb{F}^s, x'' \in \mathbb{F}^{d-s}, x''' \in \mathbb{F}^s$. Note that $f(x) = f(x', x'')$ and $g(x) = g(x'', x''')$. Expand $g(x)$ as

$$g(x) = \sum_I g_I(x', x'') \prod_{j=1}^{s} (x'''_j)^{I_j},$$

where $I \in \{0, \ldots, |\mathbb{F}|-1\}^s$ and $g_I : \mathbb{F}^d \to A$. Equivalently, $\tilde{g} : \mathbb{F}^s \to \mathrm{Poly}(\mathbb{F}^d, A)$ is given by

$$\tilde{g}(x''') = \left(c \to \sum_I g_I(c) \prod_{j=1}^{s} (x'''_j)^{I_j} : c \in \mathbb{F}^d \right).$$

Define $\tilde{g}_{\mathrm{lin}} : \mathbb{F}^s \to \mathrm{Poly}(\mathbb{F}^d, A)$ as the terms in $\tilde{g}$ which are linear in x'''. Let $e_1, \ldots, e_s \in \mathbb{F}^s$ denote the standard basis vectors (where $(e_i)_j = 1_{i=j}$). Define

$$\tilde{g}_{\mathrm{lin}}(x''') := \left(c \to \sum_{i \in [s]} g_{e_i}(c) x'''_i : c \in \mathbb{F}^d \right).$$

By definition, $\tilde{g}_{\mathrm{lin}}$ is a linear function from $\mathbb{F}^s$ to $\mathrm{Poly}(\mathbb{F}^d, A)$. We need to show that $\tilde{g}_{\mathrm{rest}} = \tilde{g} - \tilde{g}_{\mathrm{lin}}$ satisfies $\mathrm{LinDim}(\tilde{g}_{\mathrm{rest}}) \le d/2$.

First, note that $\tilde{g}(x''')$ depends only on the s variables $x'''_1, \ldots, x'''_s$, hence clearly $\mathrm{LinDim}(\tilde{g}) \le s$ and also $\mathrm{LinDim}(\tilde{g}_{\mathrm{rest}}) \le s$, and the lemma follows if $s \le d/2$. So, assume from now on that $s > d/2$. By assumption, $f + g$ has linear rank $\le d$. By Claim 3.6, $\dim(\mathcal{O}_{f+g}) \ge n - d = s$. Thus, if we set

$$U := \{u \in \mathcal{O}_{f+g} : u'' = 0\}$$

then $r := \dim(U) \ge \dim(\mathcal{O}_{f+g}) - (d-s) = 2s - d$. For any $u \in U$ we have the identity

$$f(x'+u', x'') - f(x', x'') + g(x'', x'''+u''') - g(x'', x''') = (f+g)(x+u) - (f+g)(x) = 0. \quad (1)$$

We first argue that if $u \in U$ is nonzero, then $u''' \neq 0$. Indeed, if $u''' = 0$ then $u' \neq 0$ and Equation (1) implies that

$$f(x'+u', x'') - f(x', x'') = 0$$

which implies that $\mathrm{LinDim}(f) \le d-1$, contradicting our assumption. As U is a linear space, this implies that

$$\dim\{u''' : u \in U\} = \dim U = r.$$

For simplicity of exposition, apply a change of basis to x''' so that $\{u''' : u \in U\}$ is spanned by the first r standard basis vectors in $\mathbb{F}^s$, namely by $e_1, \ldots, e_r$. We will next show that $\tilde{g}_{\mathrm{rest}}$ does not depend on $x'''_1, \ldots, x'''_r$. Thus, it depends on at most $s - r \le s - (2s-d) = d - s \le d/2$ inputs, which implies that $\mathrm{LinDim}(\tilde{g}_{\mathrm{rest}}) \le d/2$ as claimed.

So, fix $i \in [r]$, where our goal is to show that $\tilde{g}_{\mathrm{rest}}$ does not depend on x'''_i. Let $u_i \in U$ be such that $u'''_i = e_i$. By Equation (1)

$$g(x'', x''' + e_i) - g(x'', x''') = f(x', x'') - f(x' + u'_i, x'')$$

where the right hand side is independent of x'''. On the other hand we have

$$g(x'', x''' + e_i) - g(x'', x''') = \sum_I g_I(x'') \cdot I_i \cdot (x'''_i)^{I_i - 1} \cdot \prod_{j \in [s], j \neq i} (x'''_j)^{I_j}.$$

This implies that we must have $g_I(x'') = 0$ whenever $I_i \ge 1$, except for possibly $I = e_i$. However, as we already account for $g_{e_i}(x'')$ in $\tilde{g}_{\mathrm{lin}}$, we conclude that $\tilde{g}_{\mathrm{rest}}$ is independent of x'''_i. □

Lemma 3.8. *Let $\Lambda \subset Poly(\mathbb{F}^n, A)$ be a linear space of d-linear juntas. Assume furthermore that some $f \in \Lambda$ has $LinDim(f) = d$ and $f(x) = f(x_1, \ldots, x_d)$. For each $g \in \Lambda$ define $\tilde{g}, \tilde{g}_{lin}, \tilde{g}_{rest}$ as in Lemma 3.7. Then the set*

$$\{\tilde{g}_{rest} : g \in \Lambda\} \subset Poly(\mathbb{F}^{n-d}, Poly(\mathbb{F}^d, A))$$

is a linear space.

Proof. This follows directly from the construction of $\tilde{g}_{\text{rest}}$ in Lemma 3.7 and the linearity of Λ. Let $x = (\overline{x}, \overline{\overline{x}})$ with $\overline{x} \in \mathbb{F}^d, \overline{\overline{x}} \in \mathbb{F}^{n-d}$. For any $g^1, g^2 \in \Lambda$ let $g^3 = \alpha g^1 + \beta g^2 \in \Lambda$. Expand

$$g^k(x) = \sum_I g_I^k(\overline{x}) \prod_{j=1}^{d} (\overline{\overline{x}}_j)^{I_j} \qquad \forall k \in \{1, 2, 3\}.$$

Then $g_I^3 = \alpha g_I^1 + \beta g_I^2$ and hence by construction, $\alpha \tilde{g}_{\text{rest}}^1 + \beta \tilde{g}_{\text{rest}}^2 = \tilde{g}_{\text{rest}}^3$ is in the set. □

We are now ready to prove Theorem 3.5.

Proof of Theorem 3.5. The proof is by induction on d. If $d = 1$ then there is nothing to prove, so assume $d > 1$. We may assume without loss of generality that there exists $f \in \Lambda$ such that $\text{LinDim}(f) = d$. By applying a change of basis on all functions in Λ, we may assume that $f(x) = f(x_1, \ldots, x_d)$. Applying Lemma 3.7 to any $g \in \Lambda$, we conclude that we can decompose $\tilde{g} = \tilde{g}_{\text{lin}} + \tilde{g}_{\text{rest}}$, where $\Lambda' = \{\tilde{g}_{\text{rest}} : g \in \Lambda\} \subset \text{Poly}(\mathbb{F}^{n-d}, A')$ is a linear subspace of $d/2$ linear juntas. Applying induction, we may change basis for $\mathbb{F}^{n-d}$ so that after the change of basis, then functions

$$\{f'(c_{d+1}, \ldots, c_{2d}, x_{2d+1}, \ldots, x_n) : f' \in \Lambda', c_{d+1}, \ldots, c_{2d} \in \mathbb{F}\}$$

are all linear functions from $\mathbb{F}^{n-2d}$ to $\text{Poly}(\mathbb{F}^d, A)$. Recalling that $f'(c_{d+1}, \ldots, c_{2d}, x_{d+1}, \ldots, x_{n-d})(c_1, \ldots, c_d) = f(c_1, \ldots, c_{2d}, x_{2d+1}, \ldots, x_n)$ for any $f \in \Lambda$, we conclude that (after an appropriate change of basis), all functions in Λ become linear after any fixing of the first $2d$ inputs. □

4 Linear dimension of tensors

Fix $k \geq 1$, a field $\mathbb{F}$ and a linear space A over $\mathbb{F}$. An order k tensor is a multi-linear map $T : (\mathbb{F}^n)^k \to A$ given by

$$T(x^1, \ldots, x^k) = \sum_{I \in [n]^k} T_I \prod_{i=1}^{k} x_{i,I_i},$$

where $x^i = (x_{i,1}, \ldots, x_{i,n}) \in \mathbb{F}^n$ for $i \in [k]$ and $T_I \in A$. Two tensors are said to be isomorphic if they are equal, up to a change of basis for each $x^1, \ldots, x^k$. That is, if $\varphi_1, \ldots, \varphi_k : \mathbb{F}^n \to \mathbb{F}^n$ are invertible linear transformations, then T is isomorphic to T' defined as $T'(x^1, \ldots, x^k) =$

$T(\varphi_1(x^1), \ldots, \varphi_k(x^k))$. We denote this $T \equiv T'$. Given an order k tensor T, let $T|_{x^i=a}$ for $i \in [k], a \in \mathbb{F}^n$ be the order $k-1$ tensor given by fixing $x^i = a$. That is

$$T|_{x^i=a}(x^1, \ldots, x^{i-1}, x^{i+1}, x^k) = \sum_{I \in [n]^k} T_I \cdot a_{i,I_i} \prod_{j \in [k], j \neq i} x_{j,I_j}.$$

The *linear dimension* of a tensor T is the minimal d, such that T depends on at most d linear functions of each of $x^1, \ldots, x^k$.

Definition 4.1 (Linear dimension of tensors)**.** *The linear dimension of an order k tensor T, denoted LinDim(T), is the minimal $d \geq 1$ such that the following holds. There exists an order k tensor T', where $T' \equiv T$, such that*

$$T'(x^1, \ldots, x^k) = \sum_{I \in [d]^k} T'_I \prod_{i=1}^{k} x_{i,I_i}.$$

It is obvious that if $\mathrm{LinDim}(T) = d$ then $\mathrm{LinDim}(T|_{x^i=a}) \leq d$ for all $i \in [k], a \in \mathbb{F}^n$. Our main theorem in this section is the inverse relation: if all restrictions of tensor have low linear dimension, then so does the tensor. This in fact is false if $k = 2$ and say $A = \mathbb{F}$, as any restriction of order 2 tensor is a linear function, and hence has linear dimension 1. However, we show it does hold whenever $k \geq 3$. In fact, it is sufficient if $\mathrm{LinDim}(T|_{x^i=a}) \leq d$ for all $a \in \mathbb{F}^n$ and two indices $i \in [k]$.

Theorem 4.2. *Let $k \geq 3$, $\mathbb{F}$ a field, A a linear space over $\mathbb{F}$. Let $T : (\mathbb{F}^n)^k \to A$ be a tensor such that*

$$\mathit{LinDim}(T|_{x^i=c}) \leq d \qquad \forall i \in \{1, 2\}, c \in \mathbb{F}^n.$$

Then LinDim(T) $\leq 4d$.

Proof. In order to prove Theorem 4.2, it suffices to prove that for every $i \in [k]$, there exists a linear transformation $\varphi_i : \mathbb{F}^n \to \mathbb{F}^n$ such that the tensor $T'(x^1, \ldots, x^k) = T(x^1, \ldots, x^{i-1}, \varphi_i(x^i), x^{i+1}, \ldots, x^k)$ depends on only the first $4d$ variables from x^i. That is, $T'_I = 0$ if $I_i \notin [4d]$. Fix $j \in \{1, 2\} \setminus \{i\}$. In the proof below, we would only use the assumption that $\mathrm{LinDim}(T|_{x_j=a}) \leq d$ for all $a \in \mathbb{F}^n$. To simplify the presentation, fix $j = 1, i = 2$.

For every $a \in \mathbb{F}^n$ define the $k-1$ dimensional tensor $T_a = T|_{x_1=a}$. By assumption, $\mathrm{LinDim}(T_a) \leq d$. Define a function $f_a : \mathbb{F}^{2n} \to \mathrm{Poly}((\mathbb{F}^n)^{k-3}, \mathbb{F})$ as follows. Identify $\mathbb{F}^{2n} \cong (\mathbb{F}_n)^2$ and let

$$f_a(y, z) = ((c_4, \ldots, c_k) \to T(a, y, z, c_4, \ldots, c_k) : c_4, \ldots, c_k \in \mathbb{F}^n).$$

We claim that $\mathrm{LinDim}(f_a) \leq 2d$. To see that, note that as T_a has linear dimension $\leq d$, we can apply a change of basis to each x^i, so that afterwards T_a depends only on the first d inputs of each x^i. If we apply the change of basis of x^2, x^3 to y, z, we get that f_a depends only on $y_1, \ldots, y_d, z_1, \ldots, z_d$. Hence, $\mathrm{LinDim}(f_a) \leq 2d$. Next, observe that $\{f_a : a \in \mathbb{F}^n\}$ is a linear space of functions, since $\alpha f_a + \beta f_b = f_{\alpha a + \beta b}$ for all $a, b \in \mathbb{F}^n, \alpha, \beta \in \mathbb{F}$. We may thus apply Theorem 3.5.

Thus, there is a linear subspace $V \subset \mathbb{F}^{2n}$ of co-dimension $4d$ such that, for any coset $V + w$ we have that $(f_a)|_{V+w}$ is a linear function of y, z. Let V_1, V_2 be the projections of V to the first and last n variables, respectively. As $V \subset V_1 \times V_2$, we get that the same holds for any coset of $V_1 \times V_2$. By applying an additional change of basis to both y and z, we may assume that $V_1 = V_2 = \{0\}^{4d} \times \mathbb{F}^{n-4d}$. That is, after applying this change of basis to y and to z, we have that f_a becomes linear in y, z whenever we fix $y_1, \dots, y_{4d}, z_1, \dots, z_{4d}$.

Hence, the same property also holds for T_a. That is, after applying the same change of basis to x^2, x^3, for every $c_2, c_3 \in \mathbb{F}^{4d}$ we have

$$T_a(x^2, \dots, x^k)|_{x^2_{[4d]}=c_2, x^3_{[4d]}=c_3} = \sum_{i=4d+1}^{n} x_{2,i} F'_{a,i,c_2,c_3}(x^4, \dots, x^k) + \sum_{i=4d+1}^{n} x_{3,i} F''_{a,i,c_2,c_3}(x^4, \dots, x^k),$$

where $F'_{a,i,c_2,c_3}, F''_{a,i,c_2,c_3}$ are some functions on $x^4, \dots, x^k$. However, as T_a is a tensor, any monomial in it must depend on exactly one variable from each of $x^2, \dots, x^k$. Thus, we must have $F'_{a,i,c_2,c_3} = F''_{a,i,c_2,c_3} = 0$ for all a, i, c_2, c_3. This implies that T_a depends from x^2, x^3 only on the first $4d$ inputs of each, that is

$$T_a(x^2, \dots, x^k) = \sum_{i \in [4d]} \sum_{j \in [4d]} x_{2,i} x_{3,j} T_{a,i,j}(x^4, \dots, x^k),$$

where $T_{a,i,j}$ are some order $k-3$ tensors. As this holds for all $a \in \mathbb{F}^n$, we have that

$$T(x^1, x^2, \dots, x^k) = \sum_{i \in [4d]} \sum_{j \in [4d]} x_{2,i} x_{3,j} \tilde{T}_{i,j}(x^1, x^4, \dots, x^k),$$

where $\tilde{T}_{i,j}$ are some order $k-2$ tensors. This concludes the proof: we showed that, after applying an appropriate change of basis to x^2, T depends only on the first $4d$ variables in x^2. □

4.1 Symmetric tensors

An order k tensor T is said to be *symmetric* if T_I depends only on the multi-set of I. As we will see later, symmetric tensors arise naturally in the study of polynomials. We extend some of the definitions from general tensors to symmetric ones.

First, note that any restriction $T|_{x^i=a}$ is a symmetric tensor of order $k-1$, and it's the same tensor for all $i \in [k]$. Two symmetric order k tensors are isomorphic if they are equal up to the same change of basis in each variable. That is, $T \equiv T'$ if $T'(x^1, \dots, x^k) = T(\varphi(x^1), \dots, \varphi(x^k))$ for some invertible linear transformation $\varphi : \mathbb{F}^n \to \mathbb{F}^n$. The *symmetric linear dimension* of a symmetric tensor is defined analogous to the definition of the linear dimension of a tensor, except that we require to apply the same change of basis to all inputs.

Definition 4.3 (Symmetric linear dimension of symmetric tensors)**.** *The symmetric linear dimension of an order k symmetric tensor T, denoted $\mathit{LinDim}_{\mathit{Sym}}(T)$, is the minimal $d \geq 1$*

such that the following holds. There exists an order k symmetric tensor T', where $T' \equiv T$, such that

$$T'(x^1, \dots, x^k) = \sum_{I \in [d]^k} T'_I \prod_{i=1}^{k} x_{i,I_i}.$$

We state a variant of Theorem 4.2 for symmetric tensors.

Theorem 4.4. *Let $k \geq 3$, $\mathbb{F}$ a field, A a linear space over $\mathbb{F}$. Let $T : (\mathbb{F}^n)^k \to A$ be a symmetric tensor such that*

$$\mathit{LinDim}_{\mathit{Sym}}(T|_{x^1=c}) \leq d \qquad \forall c \in \mathbb{F}^n.$$

Then $\mathit{LinDim}_{\mathit{Sym}}(T) \leq 8d$.

Proof. The proof is nearly identical to the proof of Theorem 4.2, so we only highlight the differences. As T is symmetric we have that $\mathrm{LinDim}_{\mathrm{Sym}}(T|_{x^i=c}) \leq d$ for all $i \in [k], c \in \mathbb{F}^n$. Defining f_a as in the proof of Theorem 4.2, we still have $\mathrm{LinDim}(f_a) \leq 2d$ for all $a \in \mathbb{F}^n$. Thus, there exist subspaces $V_1, V_2 \subset \mathbb{F}^n$ of co-dimension $4d$ each, such that f_a becomes linear when restricted to any coset of $V_1 \times V_2$. The only difference comes now: we are only allowed to apply the same change of basis to V_1 and V_2. Thus, it might be that after this common change of basis, we would need to fix $y_1, \dots, y_{8d}, z_1, \dots, z_{8d}$ so that f_a would become linear. The remainder of the proof is unchanged. □

5 Fourier structure of polynomials

Let $\mathbb{F}$ be a finite field and let $e : \mathbb{F} \to \mathbb{C}$ be a nontrivial additive character. Let $f \in \mathrm{Poly}_d(\mathbb{F}^n, \mathbb{F})$ be an n-variate degree-d polynomial over $\mathbb{F}$. Its monomials of degree k, for $k \leq d$, are indexed by multi-sets $\{i_1, \dots, i_k\} \subset [n]$, where each index i to appear in a multi-set at most $|\mathbb{F}| - 1$ times. We denote this set by $[n]_{k,\mathbb{F}}$. As the field is fixed throughout, we shorthand $[n]_k = [n]_{k,\mathbb{F}}$. We have

$$f(x) = \sum_{k=0}^{d} \sum_{I \in [n]_k} f_I \prod_{i \in I} x_i.$$

Theorem 5.1. *Let $f \in \mathit{Poly}_d(\mathbb{F}^n, \mathbb{F})$ for $d \geq 2$. Then $\mathit{LinDim}(f) \leq 2^{4d} \log_{|\mathbb{F}|} \|\widehat{e(f)}\|_1$.*

We prove Theorem 5.1 in this section by induction on d. In order to reduce degrees, we apply derivatives.

Definition 5.2 (Derivative)**.** *The directional derivative of $f : \mathbb{F}^n \to \mathbb{F}$ in direction $h \in \mathbb{F}^n$ is $\Delta_h f : \mathbb{F}^n \to \mathbb{F}$ given by*

$$\Delta_h f(x) = f(x+h) - f(x).$$

Note that if $\deg(f) = d$ then $\deg(\Delta_h f) \leq d - 1$. The derivative polynomial of f is $Df : (\mathbb{F}^n)^d \to \mathbb{F}$ given by

$$Df(y^1, \dots, y^d) = \Delta_{y^1} \dots \Delta_{y^d} f(x) = \Delta_{y^1} \dots \Delta_{y^d} f(0).$$

It depends only on the monomials of f of maximal degree d, and is given by the symmetric order d tensor

$$Df(y^1, \dots, y^d) = \sum_{I \in [n]_d} f_I \sum_{\sigma \in S_d} \prod_{j=1}^{d} y_{\sigma(j), i_j}$$

where S_d is the group of all permutations on $[d]$.

Claim 5.3. $\|\widehat{e(Df)}\|_1 \leq \|\widehat{e(f)}\|_1^{2^d}$.

Proof. Claim 2.2 implies that

$$\|\widehat{e(\Delta_y f(x))}\|_1 = \|\widehat{e(f(x+y))e(-f(x))}\|_1 \leq \|\widehat{e(f(x+y))}\|_1 \|\widehat{e(-f(x))}\|_1 = \|\widehat{e(f)}\|_1^2.$$

The claim follows by applying this iteratively d times. □

This motivates studying the linear dimension of Df which is a symmetric tensor of order d. For a tensor $T : (\mathbb{F}^n)^d \to \mathbb{F}$, we define its Fourier coefficients by identifying it with a function $F : \mathbb{F}^{nd} \to \mathbb{F}$ in the obvious way. In particular, $\|\widehat{e(T)}\|_1 = \|\widehat{e(F)}\|_1$.

Lemma 5.4. *Let $T : (\mathbb{F}^n)^d \to \mathbb{F}$ be a symmetric tensor for $d \geq 2$. Then $LinDim_{Sym}(T) \leq 8^{d-2} \log_{|\mathbb{F}|} \|\widehat{e(T)}\|_1$.*

Proof. We prove the lemma by induction on d. The base case $d = 2$ follows from basic linear algebra.

We have $T(x, y) = xMy$ for some symmetric $n \times n$ matrix M. Assume that M has rank r. By applying a change of basis to x and to y (not necessarily the same one), we may assume that M is the $r \times r$ identity matrix. That is, $T(x, y) = \sum_{i=1}^{r} x_i y_i$. One can then verify that $e(T)$ has exactly $|\mathbb{F}|^{2r}$ nonzero Fourier coefficients, supported on $x_1, \dots, x_r, y_1, \dots, y_r$, each of which equal in absolute value to $|\mathbb{F}|^{-r}$. Thus $\|\widehat{e(T)}\|_1 = |\mathbb{F}|^r$ and $\mathrm{LinDim}_{\mathrm{Sym}}(T) \leq 2r$ (recall that T is a symmetric tensor, hence linear dimension is only defined up to a simultaneous change of basis to x, y).

In fact, a more careful analysis shows that $\mathrm{LinDim}_{\mathrm{Sym}}(T) = r$. If $\mathrm{char}(F) \neq 2$ then there exists a simultaneous change of basis to x, y such that $T(x, y) = \sum_{i=1}^{r} a_i x_i y_i$ for some nonzero $a_i \in \mathbb{F}$. If $\mathrm{char}(\mathbb{F}) = 2$ then r is even and there exists a simultaneous change of basis to x, y such that $T(x, y) = \sum_{i=1}^{r/2} a_i (x_{2i-1} y_{2i} + x_{2i} y_{2i-1})$. In either case we get $\|\widehat{e(T)}\|_1 = |\mathbb{F}|^r$ and $\mathrm{LinDim}_{\mathrm{Sym}}(T) = r$.

So, assume $d \geq 3$. For any $a \in \mathbb{F}^n$ let T_a be the order $d - 1$ tensor given by restricting a variable to a, that is

$$T_a(x^1, \dots, x^{d-1}) = T(x^1, \dots, x^{d-1}, a).$$

By Claim 2.1 we know that $\|\widehat{e(T_a)}\|_1 \leq \|\widehat{e(T)}\|_1$. By the inductive hypothesis of the lemma, we have $\mathrm{LinDim}_{\mathsf{Sym}}(T_a) \leq 8^{d-3} \log_{|\mathbb{F}|} \|\widehat{e(T_a)}\|_1$. By Theorem 4.4, this implies that $\mathrm{LinDim}_{\mathsf{Sym}}(T) \leq 8^{d-2} \log_{|\mathbb{F}|} \|\widehat{e(T)}\|_1$. □

Corollary 5.5. *Let $f \in Poly_d(\mathbb{F}^n, \mathbb{F})$ for $d \geq 2$. Let $r = 2^{4d-6} \log_{|\mathbb{F}|} \|\widehat{e(f)}\|_1$. Then there exists an invertible change of basis, after which all the monomials of degree d of f are supported on the first r variables.*

Proof. Apply Lemma 5.4 to $T = Df$. Since $\mathrm{LinDim}_{\mathsf{Sym}}(Df) \leq 8^{d-2} \log_{|\mathbb{F}|}(\|\widehat{e(f)}\|_1^{2^d}) = 2^{4d-6} \log_{|\mathbb{F}|} \|\widehat{e(f)}\|_1 = r$, we have that there exists a change a basis, so that all the monomials in $T(y^1, \ldots, y^d)$ depend on $\{y_{i,j} : i \in [d], j \in [r]\}$. By definition of Df, this implies that all the degree d monomials in f are supported on $x_1, \ldots, x_r$. □

For $x \in \mathbb{F}^n$ let $x = (x', x'')$ with $x' \in \mathbb{F}^r$ and $x'' \in \mathbb{F}^{n-r}$. Define

$$f_0(x) = f(x', 0), \quad g(x) := f(x) - f_0(x).$$

By Corollary 5.5, $\deg(g) \leq d - 1$. By Claim 2.1 $\|\widehat{e(f_0)}\|_1 \leq \|\widehat{e(f)}\|_1$ and by Claim 2.2, $\|\widehat{e(g)}\|_1 \leq \|\widehat{e(f)}\|_1^2$. We may thus apply the inductive hypothesis to g, and deduce that

$$\mathrm{LinDim}(g) \leq 2^{4(d-1)} \log_{|\mathbb{F}|} \|\widehat{e(g)}\|_1 \leq 2^{4d-3} \log_{|\mathbb{F}|} \|\widehat{e(f)}\|_1.$$

We may thus conclude that

$$\mathrm{LinDim}(f) \leq \mathrm{LinDim}(f_0) + \mathrm{LinDim}(g) \leq (2^{4d-6} + 2^{4d-3}) \log_{|\mathbb{F}|} s \leq 2^{4d} \log_{|\mathbb{F}|} \|\widehat{e(f)}\|_1.$$

References

[1] V. Bergelson, T. Tao, and T. Ziegler. An inverse theorem for the uniformity seminorms associated with the action of $\mathbb{F}_p^\infty$. *Geom. Funct. Anal.*, 19(6):1539–1596, 2010.

[2] A. Bhattacharyya. Polynomial decompositions in polynomial time. In *Algorithms-ESA 2014*, pages 125–136. Springer, 2014.

[3] A. Bhattacharyya and A. Bhowmick. Using higher-order Fourier analysis over general fields. *arXiv preprint arXiv:1505.00619*, 2015.

[4] A. Bhattacharyya, E. Fischer, H. Hatami, P. Hatami, and S. Lovett. Every locally characterized affine-invariant property is testable. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 429–436. ACM, 2013.

[5] A. Bhattacharyya, P. Hatami, and M. Tulsiani. Algorithmic regularity for polynomials and applications. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1870–1889. SIAM, 2015.

[6] A. Bhowmick and S. Lovett. Bias vs structure of polynomials in large fields, and applications in effective algebraic geometry and coding theory. *arXiv preprint arXiv:1506.02047*, 2015.

[7] A. Bhowmick and S. Lovett. The list decoding radius of Reed-Muller codes over small fields. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 277–285. ACM, 2015.

[8] W. Gowers. A new proof of Szemerédi's theorem. *Geometric and Functional Analysis GAFA*, 11(3):465–588, 2001.

[9] B. Green and T. Tao. The distribution of polynomials over finite fields, with applications to the Gowers norms. *Contrib. Discrete Math*, 4(2):1–36, 2009.

[10] E. Haramaty and A. Shpilka. On the structure of cubic and quartic polynomials. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 331–340. ACM, 2010.

[11] H. Hatami, P. Hatami, and S. Lovett. General systems of linear forms: equidistribution and true complexity. *arXiv preprint arXiv:1403.7703*, 2014.

[12] H. Hatami and S. Lovett. Estimating the distance from testable affine-invariant properties. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 237–242. IEEE, 2013.

[13] T. Kaufman and S. Lovett. Worst case to average case reductions for polynomials. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*, pages 166–175. IEEE, 2008.

[14] L. Lovász and M. Saks. Lattices, mobius functions and communications complexity. 1988.

[15] S. Lovett. Holes in generalized Reed–Muller codes. *Information Theory, IEEE Transactions on*, 56(6):2583–2586, 2010.

[16] A. Samorodnitsky. Low-degree tests at large distances. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 506–515. ACM, 2007.

[17] T. Tao and T. Ziegler. The inverse conjecture for the Gowers norm over finite fields via the correspondence principle. *Analysis and PDE*, 3:1–20, 2010.

[18] T. Tao and T. Ziegler. The inverse conjecture for the Gowers norm over finite fields in low characteristic. *ArXiv e-prints*, Jan. 2011.

[19] T. Tao and T. Ziegler. The inverse conjecture for the Gowers norm over finite fields in low characteristic. *Annals of Combinatorics*, 16(1):121–188, 2012.

[20] H. Y. Tsang, C. H. Wong, N. Xie, and S. Zhang. Fourier sparsity, spectral norm, and the log-rank conjecture. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 658–667. IEEE, 2013.

[21] M. Tulsiani and J. Wolf. Quadratic Goldreich–Levin theorems. *SIAM Journal on Computing*, 43(2):730–766, 2014.

[22] S. Zhang. Efficient quantum protocols for XOR functions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1878–1885. SIAM, 2014.