

Fourier Sparsity of GF(2) Polynomials

Hing Yin Tsang*

Ning Xie†

Shengyu Zhang‡

Abstract

We study a conjecture called “linear rank conjecture” recently raised in (Tsang *et al.*, FOCS’13), which asserts that if many linear constraints are required to lower the degree of a GF(2) polynomial, then the Fourier sparsity (i.e. number of non-zero Fourier coefficients) of the polynomial must be large. We notice that the conjecture implies a surprising phenomenon that if the highest degree monomials of a GF(2) polynomial satisfy a certain condition, then the Fourier sparsity of the polynomial is large regardless of the monomials of lower degrees – whose number is generally much larger than that of the highest degree monomials. We develop a new technique for proving lower bound on the Fourier sparsity of GF(2) polynomials, and apply it to certain special classes of polynomials to showcase the above phenomenon.

1 Introduction

The study of *communication complexity*, introduced by Yao [Yao79] in 1979, aims at investigating the minimum amount of information exchange required for computing functions whose inputs are distributed among multiple parties [KN97]. In the standard two-party setting, Alice holds an input x , Bob holds an input y , and they wish to compute a function F on (x, y) by as little communication as possible. Perhaps the most important open problem in communication complexity is the so-called *Log-rank Conjecture* proposed by Lovász and Saks [LS88], which states that the *deterministic communication complexity* of any $F : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$, $D^{\text{CC}}(F)$, is upper bounded by a polynomial of the logarithm of the rank the communication matrix $M_F = [F(x, y)]_{x, y}$, where the rank is taken over the reals. Although a lot of effort has been devoted to the conjecture in the past two decades, very little progress has been achieved and the best upper bound known to date is $D^{\text{CC}}(F) = O\left(\sqrt{\text{rank}(M_F)} \log(\text{rank}(M_F))\right)$, due to Lovett [Lov14a]. Note that there is still an exponential gap between this and the best known lower bound, which is $D^{\text{CC}}(F) = \Omega((\log \text{rank}(M_F))^{\log_3 6})$ due to Kushilevitz (unpublished, cf. [NW95]). For an overview of recent developments in this direction, see [Lov14b].

An interesting special class of functions computable by two parties is the so-called *XOR functions*. Specifically, F is an XOR function if there exists an $f : \{0, 1\}^n \rightarrow \{0, 1\}$ such that for all x and y , $F(x, y) = f(x \oplus y)$, where \oplus is the bit-wise XOR. Denote such F by $f \circ \oplus$. Besides

*University of Chicago, Chicago, IL 60637, USA. Email: hytsang@uchicago.edu.

†Florida International University, Miami, FL 33199, USA. Email: nxie@cis.fiu.edu. Research supported in part by NSF grant 1423034.

‡The Chinese University of Hong Kong, Shatin, NT, Hong Kong. Email: syzhang@cse.cuhk.edu.hk. Research supported in part by RGC of Hong Kong (Project no. CUHK419413).

including important examples such as Equality and Hamming Distance, XOR functions are particularly interesting for studying the Log-rank Conjecture due to its intimate connection with the analysis of Boolean functions. Specifically, if F is an XOR function, then the rank of M_F is just the Fourier sparsity of f (i.e., the number of non-zero Fourier coefficients of f) [BC99]. Therefore proving the Log-rank conjecture for XOR functions can be achieved by demonstrating short *parity decision tree* protocols¹ computing Fourier sparse Boolean functions, and this problem attracted a lot of attention [ZS09, LZ10, MO09, TWXZ13, STV14] during the past years.

Recently, by viewing Boolean functions as \mathbb{F}_2 -polynomials, a new communication protocol based on \mathbb{F}_2 -degree reduction was proposed in [TWXZ13] for XOR functions: suppose $f(x \oplus y)$ is a degree- d polynomial and r_d is the minimum number of variables (up to an invertible linear transformation) restricting of which reduces f 's degree to at most $d - 1$, then Alice and Bob both apply the optimal linear map to their inputs and send each other r_d bits of their respective inputs. Repeating this process at most $d - 1$ times, the restricted function of f becomes a constant function hence they successfully compute $f(x \oplus y)$. Of course, such a protocol is efficient only if the numbers r_d, r_{d-1}, \dots, r_1 , of the restricted variables that they need to exchange, are not large. Studying these quantities, namely *linear ranks* of polynomials, is one the central objectives of this paper.

Definition 1 (linear rank of a polynomial). *Let f be a degree- d polynomial, V be a subspace in $\{0, 1\}^n$ and $H = a + V$ be any affine shift of V . Denote by $f|_H$ the restriction of f on H . Then the linear rank of f , denoted $\text{lin-rank}(f)$, is the minimum co-dimension of any subspace H such that the degree of $f|_H$ is strictly less than d ; that is,*

$$\text{lin-rank}(f) = \min_{\deg_2(f|_H) < \deg_2(f)} \text{co-dim}(H).$$

In other words, $\text{lin-rank}(f)$ is the minimum number of linear functions one needs to fix in order to lower the degree of f . Consider, for example, the degree-3 polynomial $f(x_1, \dots, x_{3n}) = (x_1 + \dots + x_n)(x_{n+1} + \dots + x_{2n})(x_{2n+1} + \dots + x_{3n})$. In the original basis, one needs to fix at least n variables to lower the degree of f . However, fixing one linear function $x_1 + \dots + x_n = 0$ is enough to lower its degree. Therefore $\text{lin-rank}(f) = 1$.

For a Boolean function f , let $\text{spar}(f)$ denote the Fourier sparsity of f and $D_{\oplus}(f)$ denote the parity decision tree complexity of f . As restrictions do not increase $\text{spar}(f)$ (cf. Lemma 5) and $\deg_2(f) \leq \log \text{spar}(f)$ for every f , the following *linear rank conjecture*—if true—would readily implies the Log-rank Conjecture for XOR functions.

Conjecture 1 (Linear rank conjecture [TWXZ13]). *For any $f : \{0, 1\}^n \rightarrow \{0, 1\}$, the linear rank of f is upper bounded by polylogarithmic of the Fourier sparsity of f : $\text{lin-rank}(f) = O(\log^c(\text{spar}(f)))$ for some $c = O(1)$. Equivalently, if $\text{lin-rank}(f) = r$, then $\text{spar}(f) = 2^{r^{\Omega(1)}}$.*

Although it is still open whether the linear rank conjecture is equivalent to the Log-rank Conjecture for XOR functions, it is worthwhile to note that it is equivalent to the stronger statement that $D_{\oplus}(f) = \text{polylog}(\text{spar}(f))$ for any Boolean function f .

¹Recall that a *parity decision tree* T for a function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ generalizes an ordinary decision tree in the sense that each internal node of T is now associated with a linear function $\ell(x)$, instead of a single bit, of the input, and T branches according to the parity of $\ell(x)$.

1.1 Large Fourier sparsity determined by highest degree monomials only

Before further discussing the linear rank conjecture, let us first state a lemma of [TWXZ13] (Lemma 19) in a slightly stronger form and give an alternative simple proof (another simple proof used polynomial derivatives [CT13]). The lemma says that, once the linear subspace V in Definition 1 is identified, it does not matter which affine shift is used in the definition of linear rank: all affine subspaces of V are equally good. More specifically, if f restricted to $a + V$ has degree at most $d - 1$ (where $d = \deg_2(f)$), then f restricted to any other $a' + V$ also has degree at most $d - 1$. This can be seen by the following argument. Call a monomial in f a *maxonomial* if it is of the maximal degree (i.e., degree d). Apply a linear map to $\{0, 1\}^n$ so that $V = \{x : x_1 = \dots = x_r = 0\}$, where $r = \text{co-dim}(V)$. Then $f|_{a+V}$ becomes a polynomial of degree at most $d - 1$ if and only if every maxonomial of f (under the new basis) contains at least one variable in the set $\{x_1, \dots, x_r\}$. Moreover, when this happens it does not matter whether x_i ($i \leq r$) is restricted to 0 or 1, the degree of the maxonomial always decreases, thus $\deg_2(f|_{a'+V}) \leq d - 1$ for all $a' \in \{0, 1\}^n$.

The above fact also reveals that the linear rank r of any polynomial $f(x)$ is determined by the maxonomials in $f(x)$ *only*. Fourier sparsity in general, on the other hand, should depend on all GF(2) monomials, not only those with the highest degree. However, the linear rank conjecture claims that if the maxonomials in $f(x)$ make the linear rank large, then no matter how the lower-degree monomials behave, the Fourier sparsity is large. Therefore, for the effect of forcing the Fourier sparsity of GF(2) polynomial to be large, there exists a surprising fact (assuming the linear rank conjecture) that can be summarized by paraphrasing a famous quote from *Animal Farm*: “All monomials are equal, but some monomials are more equal than others”.

In retrospect, this phenomenon is known for some extremal cases. When $\deg_2(f) = 2$, the lower degree terms form a linear function χ_α , adding which only shifts Fourier spectrum by α and thus does not affect the Fourier sparsity. When $\deg_2(f) = n$, the Fourier sparsity is at least $2^{\deg_2(f)} - 1 = 2^n - 1$, which is again determined by the (unique) maxonomial. But for general $2 < d < n$, maxonomials by themselves do not necessarily determine large Fourier sparsity. For instance, if there is only one maxonomial $x_1 \dots x_d$, then the Fourier sparsity can be as small as 2^d (when, say, the lower degree part is $x_1 + \dots + x_n$), and as large as 2^{n-d} (when, say, the lower degree part is a bent function² over x_{d+1}, \dots, x_n). Despite this uncertainty, we will show that when the maxonomials form certain patterns, the Fourier sparsity is guaranteed to be large, regardless of the lower degree terms (whose number can be much larger than that of maxonomials). One sufficient condition for the pattern is that the linear rank, which depends on maxonomials only, is large. And we will showcase some specific classes of good patterns.

Therefore, apart from leading directly to a proof of the Log-rank Conjecture for XOR functions, studying the linear rank conjecture is interesting in its own right, due to its close connection to the Fourier analysis of Boolean functions in the GF(2) polynomial representation.

1.2 Our work

We study the linear rank conjecture and in particular investigate how could the maxonomials of a \mathbb{F}_2 -polynomial possibly determine by themselves the Fourier sparsity of the polynomial. We develop a new technique which is able to show that, under certain circumstances, the Fourier sparsity is large for all possible settings of lower degree monomials. It is hoped that this new framework of studying

²A Boolean function $f : \{0, 1\}^m \rightarrow \{-1, 1\}$ is *bent* if its Fourier coefficients satisfy that $|\hat{f}(\alpha)| = 2^{-m/2}$ for all $\alpha \in \{0, 1\}^m$.

the Fourier coefficients based on GF(2) monomials may be further extended and generalized to yield more structural results on the analysis of Boolean functions, such as sparsity, granularity and Fourier mass distribution.

For general degree- d polynomials, we investigate the linear rank and Fourier sparsity for several special cases. Since the maxonomials of a polynomial are the main concern of the conjecture, it is convenient to borrow the terminology of hypergraphs to define these maxonomials. For example, the complete d -uniform maxonomials corresponds to the degree- d polynomial who has all $\binom{n}{d}$ maxonomials.

1.2.1 Linear rank of polynomials with complete d -uniform maxonomials

We determine the exact values of the linear ranks of degree- d polynomials with all $\binom{n}{d}$ maxonomials. Specifically, let $f = \sum_{S \subset [n], |S|=d} \prod_{i \in S} x_i + f'$, where f' is an arbitrary polynomial of degree at most $d - 1$, we show that for such an f ,

$$\text{lin-rank}(f) = \begin{cases} \lfloor \frac{n}{2} \rfloor - \frac{d}{2} + 1 & \text{if } d \text{ is even,} \\ 1 & \text{if } d \text{ is odd.} \end{cases}$$

The proof exploits the symmetry of maxonomials and goes through a careful induction on n and d . In particular we prove a “step-function” type behaviour of the linear rank (for fixed d and with respect to n), by showing both upper and lower bounds for the number of linear functions one needs to fix in order to decrease the degree of the polynomial.

1.2.2 Fourier sparsity of polynomials with complete d -uniform maxonomials

If the linear rank conjecture is true, then for any polynomial with complete d -uniform maxonomials (d is even), the Fourier sparsity must be $2^{n^{\Omega(1)}}$ regardless of the lower degree monomials. We are only able to verify this for a small (but infinite) set of d 's: for any d that is a power of 2, if $f : \{0, 1\}^n \rightarrow \{0, 1\}$ is a degree- d polynomial with complete d -uniform maxonomials, then

$$\text{spar}(f) \geq 2^{d \cdot \lfloor n/d \rfloor} - 1 = \Omega(2^n).$$

We prove this sparsity lower bound by developing a new technique to be discussed more later. Zhang and Shi [ZS09] proved that any symmetric boolean function has Fourier sparsity $2^{\Omega(n)}$, unless it is constant, the parity function over n bits or its negation. However, as the polynomials considered there are symmetric, their result requires the degree- d' monomials to be either empty or complete d' -uniform, for every $d' \leq d$. On the contrary, our lower bound applies to a broader class of functions as it holds for all possible choices of lower degree monomials, as long as the highest-degree monomials are symmetric.

1.2.3 Other results

We further demonstrate the power of our technique by applying it to several other special forms of sparse maxonomials. In particular, we show lower bounds on the Fourier sparsity of polynomials whose maxonomials are pairwise disjoint or have certain “regular” overlaps.

Gopalan *et al.* [GOS⁺11] studied the *granularity* of a function’s Fourier spectrum, which is the smallest integer k such that all Fourier coefficients of the function can be expressed as integer

multiples of $1/2^k$. They showed that for any Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$, $\text{gran}(f) \leq \log \text{spar}(f)$. On the other hand, by Parseval’s identity, $\log \text{spar}(f) \leq 2\text{gran}(f)$. The granularity of a linear functions is 1 and the maximum granularity of any n -variate quadratic polynomial is $n/2$. It thus natural to conjecture that, for any n -variate low-degree polynomial $f(x)$, although $\text{spar}(f)$ can be as large as 2^n , the granularity of $f(x)$ is always bounded away from n . We are able to apply our technique to show the following *upper bound* on the granularity of low-degree polynomials: for any degree- d polynomial f , $\text{gran}(f) \leq n - \lceil \frac{n}{d} \rceil + 1$. It is easy to see this bound is tight as it is attained by the “generalized inner product function”: $f(x) = x_1x_2 \cdots x_d + \cdots + x_{(k-1)d+1}x_{(k-1)d+2} \cdots x_{kd}$, where $n = kd$.

1.2.4 Techniques

The main challenge in proving sparsity lower bounds based on *only* the maxonomials of a polynomial is how to isolate the effect of *all* lower degree monomials. To the best of our knowledge, there is no prior method or result of this kind. Our method is to first apply the standard procedure to transform a degree- d polynomial f into a Fourier polynomial, and then define a “weight function” $w_f(T)$ on each set $T \subseteq [n]$ such that the Fourier coefficient of f at any set S can be written as $\sum_{T \supseteq S} w_f(T)$. This implies that the weight function at $[n]$ is the most important term as it contributes to all the Fourier coefficients of f . Another nice property of the weight function is that for any T , $2^{|T|}w_f(T)$ can be expressed as a sum of alternating terms in which the k^{th} term is $(-2)^k N_k(T)$, where $N_k(T)$ is the number of ways to cover T with (the supports of) exactly k monomials of $f(x)$. Therefore, the problem of computing the Fourier coefficients of an \mathbb{F}_2 -polynomial is now reduced to a combinatorial problem of counting the numbers of covers of all subsets of $[n]$ using various numbers of sets from the set family defined by the monomials of the polynomial. Moreover, the parity of $2^{|T|}w_f(T)$ is likely to be determined by the numbers of smaller covers due to the factor $(-2)^k$ in each term of the sum. Using the notion of “granularity” introduced in [GOS⁺11], our strategy for showing sparsity lower bound is to argue that $w_f([n])$ is the single one with the highest granularity among all weight function values. Note that if $n = kd$ and we can cover $[n]$ with (the supports of) maxonomials of $f(x)$ only, then these covers would be the minimum covers as they require only $k = n/d$ sets while any cover involving lower monomials is of size at least $k + 1$. Hence to prove that $w_f([n])$ has the highest possible granularity, it suffices to show that the number of k -covers of $[n]$ is odd, as we did for the several sparsity lower bounds.

1.3 Organization of the paper

Section 2 contains notations and preliminaries that will be used throughout the paper. In Section 3 we compute exactly the linear rank of polynomials with complete d -uniform maxonomials. The basic machinery for proving sparsity lower bounds are described in Section 4, and we then use this in Section 5 to prove the linear rank conjecture for complete d -uniform polynomials when d is a power of 2. In Section 6, we apply our technique to study the sparsity of several more special polynomials and prove an upper bound on the granularity of low-degree polynomials.

2 Preliminaries

All logarithms in this paper are base 2. For two n -bit vectors $\alpha, \beta \in \{0, 1\}^n$, define their inner product as $\alpha \cdot \beta = \langle \alpha, \beta \rangle = \sum_{i=1}^n \alpha_i \beta_i \pmod 2$ and for simplicity we write $\alpha + \beta$ for $\alpha \oplus \beta$. We often

use f to denote a real function defined on $\{0, 1\}^n$. In most occurrences f is a Boolean function, whose range can be represented by either $\{0, 1\}$ or $\{+1, -1\}$. For $f : \{0, 1\}^n \rightarrow \{0, 1\}$, we use $f^\pm = 1 - 2f$ to denote the equivalent Boolean function with range converted to $\{+1, -1\}$.

2.1 GF(2) polynomials

If $S \subseteq [n]$ is a set of (indices of) variables, then the monomial x_S is the product of variables in S : $x_S = \prod_{i \in S} x_i$. The *degree* of this monomial is the cardinality of S , and S is called the *support* of the monomial. We say a set T *meets* a monomial x_S if $T \cap S \neq \emptyset$.

Every Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ can be uniquely expressed as a multilinear polynomial over \mathbb{F}_2 : $p_f(x_1, \dots, x_n) = \sum_{S \subseteq \mathcal{F}} x_S$ where \mathcal{F} is a collection of subsets of $[n]$ (here additions are performed modulo 2). The *degree* of f , denoted $\deg_2(f)$, is the maximum degree of its monomials. In this paper, whenever there is no risk of confusion, we use f and multilinear polynomial representation of p_f interchangeably.

2.2 Fourier analysis

For any real function $f : \{0, 1\}^n \rightarrow \mathbb{R}$, the Fourier coefficients are defined by $\hat{f}(\alpha) = 2^{-n} \sum_x f(x) \chi_\alpha(x)$, where $\chi_\alpha(x) = (-1)^{\alpha \cdot x}$. The function f can be written as $f(x) = \sum_\alpha \hat{f}(\alpha) \chi_\alpha(x)$. The Fourier sparsity of f , denoted by $\|\hat{f}\|_0$, is the number of nonzero Fourier coefficients of f . The Fourier coefficients of $f : \{0, 1\}^n \rightarrow \{0, 1\}$ and f^\pm are related by $\widehat{f^\pm}(\alpha) = \delta_{\alpha, 0^n} - 2\hat{f}(\alpha)$, where $\delta_{x,y}$ is the Kronecker delta function. Therefore we have

$$\|\hat{f}\|_0 - 1 \leq \|\widehat{f^\pm}\|_0 \leq \|\hat{f}\|_0 + 1. \quad (1)$$

Sometimes we employ the one-to-one mapping between vectors in $\{0, 1\}^n$ and subsets of $[n]$: $x \leftrightarrow \{i \in [n] : x_i = 1\}$, and use the subsets of $[n]$ to index the Fourier coefficients.

For any function $f : \{0, 1\}^n \rightarrow \mathbb{R}$, Parseval's Identity says that $\sum_\alpha \hat{f}^2(\alpha) = \mathbf{E}_x[f(x)^2]$. When the range of f is $\{0, 1\}$, then $\sum_\alpha \hat{f}^2(\alpha) = \mathbf{E}_x[f(x)]$. We sometimes use \hat{f} to denote the vector of $\{\hat{f}(\alpha) : \alpha \in \{0, 1\}^n\}$.

2.3 Granularity and sparsity of Fourier spectrum

Definition 2 (Granularity [GOS⁺11]). *Let $r \in \mathbb{Q}$, the granularity of r , denoted $\mathbf{gran}(r)$, is the smallest nonnegative integer k such that $2^k r \in \mathbb{Z}$, and $\mathbf{gran}(r) = \infty$ if no such k exists. The Fourier granularity of a Boolean function f , denoted $\mathbf{gran}(f)$, is the maximum granularity over all the Fourier coefficients of f ; i.e., $\mathbf{gran}(f) = \max_{\alpha \in \{0, 1\}^n} (\mathbf{gran}(\hat{f}(\alpha)))$.*

Clearly, $\mathbf{gran}(-x) = \mathbf{gran}(x)$ for any $x \in \mathbb{Q}$. An easy but useful fact is that $\mathbf{gran}(x + y) \leq \max(\mathbf{gran}(x), \mathbf{gran}(y))$ for all $x, y \in \mathbb{Q}$. More generally, $\mathbf{gran}(\sum_{i=1}^k x_i) \leq \max_{1 \leq i \leq k} \mathbf{gran}(x_i)$, where $x_i \in \mathbb{Q}$ for every $1 \leq i \leq k$.

Fact 2. *Let $f^\pm, g^\pm : \{0, 1\}^n \rightarrow \{-1, 1\}$ be two Boolean functions. Let $h = f \oplus g$. Then $|\mathbf{gran}(f^\pm) - \mathbf{gran}(g^\pm)| \leq \mathbf{gran}(h^\pm) \leq \mathbf{gran}(f^\pm) + \mathbf{gran}(g^\pm)$.*

Proof. Since the Fourier spectrum of h^\pm is given by the convolution formula

$$\widehat{h^\pm}(\alpha) = \sum_{\beta \in \{0, 1\}^n} \widehat{f^\pm}(\beta) \widehat{g^\pm}(\alpha + \beta),$$

the upper bound on $\text{gran}(h^\pm)$ follows directly from the definition of granularity. Now suppose $\text{gran}(f^\pm) \geq \text{gran}(g^\pm)$, then applying the granularity upper bound on XOR of two functions we just show on $g \oplus h$, which is f , gives the desired lower bound. \square

Gopalan *et al.* [GOS⁺11] showed that, if a Boolean function has only a small number of non-zero Fourier coefficients, then all these non-zero Fourier coefficients have small granularities.

Lemma 3 ([GOS⁺11]). *Suppose $f^\pm : \{0, 1\}^n \rightarrow \{-1, 1\}$ is s -sparse with $s > 0$, then all the Fourier coefficients of f^\pm have granularity at most $\lceil \log s \rceil - 1$.*

The following claim shows that the logarithm of the sparsity and granularity of a Boolean function are in fact equivalent up to a constant factor.

Proposition 4. *Let $f^\pm : \{0, 1\}^n \rightarrow \{-1, 1\}$ be a Boolean function, then*

$$\text{gran}(f^\pm) + 1 \leq \log \text{spar}(f^\pm) \leq 2\text{gran}(f^\pm).$$

Proof. Suppose that $\text{gran}(f^\pm) = k$. Then for any $\alpha \in \{0, 1\}^n$, if $\hat{f}^\pm(\alpha) \neq 0$, then $|\hat{f}^\pm(\alpha)| \geq 1/2^k$. By Parseval's identity, we have $\text{spar}(f^\pm) \leq 2^{2k}$, or $\lceil \log(\text{spar}(f^\pm)) \rceil \leq 2k$. Combining with Lemma 3 gives the desired result. \square

Note that both bounds in Proposition 4 are tight: for the first inequality, consider the n -variate degree- n polynomial $f(x) = x_1x_2 \cdots x_n$, which satisfies $\text{spar}(f^\pm) = 2^n$ and $\text{gran}(f^\pm) = n - 1$; for the second inequality, consider for any even integer n and the Inner Product function on n variables $f(x) = x_1x_2 + x_3x_4 + \cdots + x_{n-1}x_n$, then f^\pm has sparsity 2^n and granularity $n/2$.

2.4 Linear maps and restrictions

Sometimes we need to rotate the input space: For an *invertible* linear map L on $\{0, 1\}^n$, define Lf by $Lf(x) = (f \circ L)(x) = f(Lx)$.

For a function $f : \{0, 1\}^n \rightarrow \mathbb{R}$, define two subfunctions f_0 and f_1 , both on $\{0, 1\}^{n-1}$: $f_b(x_2, \dots, x_n) = f(b, x_2, \dots, x_n)$. It is easy to see that for any $\alpha \in \{0, 1\}^{n-1}$, $\hat{f}_b(\alpha) = \hat{f}(0\alpha) + (-1)^b \hat{f}(1\alpha)$, thus

$$\|\hat{f}_b\|_0 \leq \|\hat{f}\|_0 \text{ and } \|\hat{f}_b\|_1 \leq \|\hat{f}\|_1. \quad (2)$$

where $\|\hat{f}\|_p = (\sum_\alpha |\hat{f}(\alpha)|^p)^{1/p}$ and $\|\hat{f}\|_0 = |\{\alpha : \hat{f}(\alpha) \neq 0\}|$. The notion of subfunctions can be generalized to restrictions with respect to a general direction. Suppose $f : \{0, 1\}^n \rightarrow \mathbb{R}$ and $S \subseteq \{0, 1\}^n$ is a subset of the domain. Then the restriction of f on S , denoted by $f|_S$ is the function from S to \mathbb{R} defined naturally by $f|_S(x) = f(x)$, $\forall x \in S$. In this paper, we are concerned with restrictions on affine subspaces.

Lemma 5. *Let $f : \{0, 1\}^n \rightarrow \mathbb{R}$ and $H = a + V$ be an affine subspace, then one can (recursively) define the spectrum $\widehat{f|_H}$ of the restricted function $f|_H$ such that*

1. *If $\text{co-dim}(H) = 1$, then $\widehat{f|_H}$ is the collection of $\hat{f}(\alpha) + (-1)^b \hat{f}(\alpha + \beta)$ for all unordered pair $(\alpha, \alpha + \beta)$, where β is the unique non-zero vector orthogonal to V , and $b = 0$ if $a \in V$ and $b = 1$ otherwise.*
2. *$\|\widehat{f|_H}\|_p \leq \|\hat{f}\|_p$, for any $p \in [0, 1]$. In particular, restriction does not increase the Fourier sparsity of a function.*

It is worth noticing that, for any Boolean function, its \mathbb{F}_2 -degree, Fourier sparsity and granularity are all invariant under invertible linear maps.

Fact 6. *Let f be an \mathbb{F}_2 -polynomial. Then for any invertible linear map L , $\deg_2(f) = \deg_2(f \circ L)$.*

Fact 7. *Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ be a Boolean function and L an invertible linear map. Then $\widehat{f \circ L}(\alpha) = \widehat{f}((L^T)^{-1}\alpha)$. In particular, $\text{spar}(f) = \text{spar}(f \circ L)$ and $\text{gran}(f) = \text{gran}(f \circ L)$.*

3 Linear rank of complete d -uniform maxonomials

We now compute the exact value of the linear rank of a degree d polynomial whose set of maxonomials consists of all $\binom{n}{d}$ degree- d monomials, and give explicit linear constraints restriction of which reduces the degree of such a polynomial.

Define $\mathcal{C}_{d,n}(x) = \sum_{I \subseteq [n], |I|=d} \prod_{i \in I} x_i$, the summation of all degree- d monomials over variables $x_1, \dots, x_n \in \mathbb{F}_2$. The subscript n is dropped when it is clear from the context. We use the equivalence relation \equiv_d for polynomials with the same maxonomials, *i.e.* $p \equiv_d q$ if both p and q have \mathbb{F}_2 -degree d and $p + q$ has \mathbb{F}_2 -degree strictly less than d . It is clear that if $p \equiv_d q$, then $\text{lin-rank}(p) = \text{lin-rank}(q)$.

Theorem 8. *Let $n \geq d \geq 0$ be integers. Then the following hold:*

1. *If d is odd, then $\text{lin-rank}(\mathcal{C}_{d,n}) = 1$.*
2. *If d is even, then $\text{lin-rank}(\mathcal{C}_{d,n}) = \lfloor \frac{n}{2} \rfloor - \frac{d}{2} + 1$, *i.e.**

$$\text{lin-rank}(\mathcal{C}_{d,n}) = \begin{cases} \frac{n-d}{2} + 1 & \text{if } n \text{ is even,} \\ \frac{n-d-1}{2} + 1 & \text{if } n \text{ is odd.} \end{cases}$$

Proof. The first item follows simply by the factorization $\mathcal{C}_{d,n} \equiv_d \mathcal{C}_{1,n} \mathcal{C}_{d-1,n}$. Indeed, when we multiply $\mathcal{C}_{1,n} = \sum_{i \in [n]} x_i$ and $\mathcal{C}_{d-1,n} = \sum_{|I|=d-1} x_I$, for $i \notin I$, $x_i x_I = x_{I \cup \{i\}}$, and each J with $|J| = d$ comes from d many (i, I) . For each $i \in I$, $x_i x_I = x_I$, and each resulting x_I with $|I| = d-1$ comes from $d-1$ many $i \in I$. Thus

$$\begin{aligned} \mathcal{C}_{1,n} \mathcal{C}_{d-1,n} &= d \left(\sum_{|J|=d} x_J \right) + (d-1) \left(\sum_{|I|=d-1} x_I \right) = d \mathcal{C}_{d,n} + (d-1) \mathcal{C}_{d-1,n} \\ &= \mathcal{C}_{d,n}, \end{aligned}$$

for all odd d .

Now we consider the second item in the statement and assume from now on that d is even and $d \leq n$. The second item follows from the following two claims.

Claim 9. *If $\text{lin-rank}(\mathcal{C}_{d,n+1}) = \text{lin-rank}(\mathcal{C}_{d,n})$, then $\text{lin-rank}(\mathcal{C}_{d,n+2}) > \text{lin-rank}(\mathcal{C}_{d,n+1})$.*

Claim 10. $\text{lin-rank}(\mathcal{C}_{d,n+2}) \leq \text{lin-rank}(\mathcal{C}_{d,n}) + 1$.

Let us first show Theorem 8 assuming these two claims. We prove by induction on the number of variables that for all $k \geq d/2$,

$$\text{lin-rank}(\mathcal{C}_{d,2k}) = \text{lin-rank}(\mathcal{C}_{d,2k+1}) = k - \frac{d}{2} + 1. \quad (3)$$

which is just a restatement of the second item of Theorem 8.

Base case $k = d/2$. We have

$$\mathcal{C}_d(x_1, \dots, x_{2k}) = \mathcal{C}_d(x_1, \dots, x_d) = \mathcal{C}_{d-1}(x_1, \dots, x_{d-1}) \cdot x_d, \quad (4)$$

so $\text{lin-rank}(\mathcal{C}_{d,2k}) = 1$. For $n = 2k + 1$, note that

$$\begin{aligned} \mathcal{C}_d(x_1, \dots, x_{2k+1}) &= \mathcal{C}_d(x_1, \dots, x_{d+1}) \\ &= \mathcal{C}_{d-1}(x_1, \dots, x_{d-1})(x_d + x_{d+1}) + \mathcal{C}_{d-2}(x_1, \dots, x_{d-1})x_dx_{d+1}, \end{aligned} \quad (5)$$

Putting restriction $x_d = x_{d+1}$ makes the first summand vanish and decreases the degree of the second summand, hence $\text{lin-rank}(\mathcal{C}_{d,2k+1}) = 1$.

General k . Now we assume that Eq. (3) holds for k and will prove the case for $k + 1$. The following sequence of inequalities hold.

$$k - \frac{d}{2} + 1 < \text{lin-rank}(\mathcal{C}_{d,2(k+1)}) \leq \text{lin-rank}(\mathcal{C}_{d,2(k+1)+1}) \leq k - \frac{d}{2} + 2,$$

where the first inequality follows by Claim 9; the second follows by the facts that $\mathcal{C}_{d,n-1}$ can be obtained from $\mathcal{C}_{d,n}$ by restricting $x_n = 0$ and restriction does not increase lin-rank ; and the last inequality follows by Claim 10. Therefore Eq. (3) also holds for $k + 1$. \square

Now it remains to prove the two claims. We start with Claim 10, which is simpler.

Proof of Claim 10. We first observe the following identity:

$$\begin{aligned} \mathcal{C}_d(x_1, \dots, x_{n+2}) &= \mathcal{C}_d(x_1, \dots, x_n) + \mathcal{C}_{d-1}(x_1, \dots, x_n)(x_{n+1} + x_{n+2}) + \mathcal{C}_{d-2}(x_1, \dots, x_n)x_{n+1}x_{n+2} \\ &\equiv_d \mathcal{C}_d(x_1, \dots, x_n) + \mathcal{C}_{d-1}(x_1, \dots, x_n, x_{n+1})(x_{n+1} + x_{n+2}). \end{aligned} \quad (6)$$

Therefore the restriction $x_{n+2} = x_{n+1}$ reduces $\mathcal{C}_d(x_1, \dots, x_{n+2})$ to

$$\mathcal{C}_d(x_1, \dots, x_{n+2})|_{x_{n+1}=x_{n+2}} \equiv_d \mathcal{C}_d(x_1, \dots, x_n).$$

Since each restriction can reduce lin-rank by at most 1, we have

$$\text{lin-rank}(\mathcal{C}_{d,n+2}) - 1 \leq \text{lin-rank}(\mathcal{C}_{d,n+2}|_{x_{n+2}=x_{n+1}}) = \text{lin-rank}(\mathcal{C}_{d,n}),$$

as desired. \square

Proof of Claim 9. For the sake of contradiction, assume that

$$\text{lin-rank}(\mathcal{C}_{d,n+2}) = \text{lin-rank}(\mathcal{C}_{d,n+1}) = \text{lin-rank}(\mathcal{C}_{d,n}) = r.$$

Fix an optimal set of linear restrictions for $\text{lin-rank}(\mathcal{C}_{d,n+2})$. Without loss of generality, we can assume it contains a restriction of the form $x_{n+2} = \ell(x_1, \dots, x_{n+1}) = \ell(x)$ for some linear form ℓ . It is clear that such restriction will reduce the lin-rank by exactly 1. So we have

$$\text{lin-rank}(\mathcal{C}_{d,n+2}|_{x_{n+2}=\ell(x)}) \leq \text{lin-rank}(\mathcal{C}_{d,n+2}) - 1 = r - 1. \quad (7)$$

But by the expansion

$$\mathcal{C}_d(x_1, \dots, x_{m+1}) = \mathcal{C}_d(x_1, \dots, x_m) + \mathcal{C}_{d-1}(x_1, \dots, x_m)x_{m+1},$$

we have

$$\begin{aligned}\mathcal{C}_d(x_1, \dots, x_{n+2})|_{x_{n+2}=\ell(x)} &= \mathcal{C}_d(x_1, \dots, x_{n+1}) + \mathcal{C}_{d-1}(x_1, \dots, x_{n+1})\ell(x) \\ &= \mathcal{C}_d(x_1, \dots, x_n) + \mathcal{C}_{d-1}(x_1, \dots, x_n)x_{n+1} + \mathcal{C}_{d-1}(x_1, \dots, x_{n+1})\ell(x).\end{aligned}\quad (8)$$

Now, consider to further restrict $x_{n+1} = x_1 + x_2 + \dots + x_n = \mathcal{C}_1(x_1, \dots, x_n)$. By the fact that $\mathcal{C}_{d-1}(x_1, \dots, x_m) \equiv_d \mathcal{C}_{d-2}(x_1, \dots, x_m)\mathcal{C}_1(x_1, \dots, x_m)$ for every even $d \geq 4$, the second term on the right of Eq.(8) is \equiv_d -equivalent to

$$\begin{aligned}&\mathcal{C}_{d-2}(x_1, \dots, x_n)\mathcal{C}_1(x_1, \dots, x_n)x_{n+1}|_{x_{n+1}=\mathcal{C}_1(x_1, \dots, x_n)} \\ &= \mathcal{C}_{d-2}(x_1, \dots, x_n)\mathcal{C}_1^2(x_1, \dots, x_n) \\ &= \mathcal{C}_{d-2}(x_1, \dots, x_n)\mathcal{C}_1(x_1, \dots, x_n) \equiv_d 0,\end{aligned}$$

and the last term becomes

$$\mathcal{C}_{d-2}(x_1, \dots, x_{n+1})\mathcal{C}_1(x_1, \dots, x_{n+1})\ell(x)|_{x_{n+1}=\mathcal{C}_1(x_1, \dots, x_n)} = 0.$$

Plugging these two back to Eq.(8),

$$\mathcal{C}_{d,n+2}|_{x_{n+2}=\ell(x), x_{n+1}=x_1+\dots+x_n} \equiv_d \mathcal{C}_{d,n}.$$

As restriction does not increase linear rank, we have from Eq.(7) that

$$r = \text{lin-rank}(\mathcal{C}_{d,n}) = \text{lin-rank}(\mathcal{C}_{d,n+2}|_{x_{n+2}=\ell(x), x_{n+1}=x_1+\dots+x_n}) \leq \text{lin-rank}(\mathcal{C}_{d,n+2}|_{x_{n+2}=\ell(x)}) \leq r - 1,$$

which is a contradiction. \square

As a simple application of Theorem 8, for any symmetric function f , let r_1, r_0 be the largest and smallest integers such that $f(x)$ is constant or parity on $\{x \in \{0, 1\}^n : r_0 \leq |x| \leq n - r_1\}$. The quantity $r := r_0 + r_1$ turns out to be an important complexity measure for symmetric functions. For example, the randomized and quantum communication complexity of symmetric XOR functions is characterized by this r [ZS09, LLZ11, LZ13], and $\log \|\hat{f}\|_1 = \Theta(r \log(n/r))$ for all symmetric functions f [AFH12].

Here we relate this measure to the \mathbb{F}_2 -degree of f . It is clear that we can fix $x_1 = x_2 = \dots = x_{r_0} = 1$ and $x_n = x_{n-1} = \dots = x_{n-r_1+1} = 0$ to reduce the degree of f to at most 1. We therefore have the following corollary.

Corollary 11. *Let f be a symmetric function with even \mathbb{F}_2 -degree d , then*

1. $\lfloor \frac{n}{2} \rfloor - \frac{d}{2} + 1 \leq r_0 + r_1$.
2. $\log \|\hat{f}\|_1 = \Omega(n/\log n)$, if $d = (1 - \Omega(1))n$.

3.1 An explicit form of linear restrictions for complete d -uniform monomials

The proof of Theorem 8 can be used to find a linear transformation which explicitly show the restrictions for $\mathcal{C}_{d,n}$. Indeed, starting from either Eq. (4) or Eq. (5) and recursively applying

Eq. (6), gives, when $n = d + 2k$ is even,

$$\begin{aligned}
& \mathcal{C}_d(x_1, \dots, x_n) \\
& \equiv_d \mathcal{C}_d(x_1, \dots, x_{n-2}) + (x_{n-1} + x_n)\mathcal{C}_{d-1}(x_1, \dots, x_{n-1}) \\
& \equiv_d \mathcal{C}_d(x_1, \dots, x_{n-4}) + (x_{n-3} + x_{n-2})\mathcal{C}_{d-1}(x_1, \dots, x_{n-3}) + (x_{n-1} + x_n)\mathcal{C}_{d-1}(x_1, \dots, x_{n-1}) \\
& \equiv_d \dots \\
& \equiv_d \mathcal{C}_d(x_1, \dots, x_d) + (x_{d+1} + x_{d+2})\mathcal{C}_{d-1}(x_1, \dots, x_{d+1}) + \dots + (x_{n-1} + x_n)\mathcal{C}_{d-1}(x_1, \dots, x_{n-1}) \\
& = x_d\mathcal{C}_{d-1}(x_1, \dots, x_{d-1}) + (x_{d+1} + x_{d+2})\mathcal{C}_{d-1}(x_1, \dots, x_{d+1}) + \dots + (x_{n-1} + x_n)\mathcal{C}_{d-1}(x_1, \dots, x_{n-1}).
\end{aligned}$$

Then in the new basis where $y_1 = x_1, \dots, y_d = x_d, y_{d+1} = x_{d+1}, y_{d+2} = x_{d+1} + x_{d+2}, \dots, y_{n-1} = x_{n-1}, y_n = x_{n-1} + x_n$, we have

$$\begin{aligned}
& \mathcal{C}_d(x_1, \dots, x_n) = \mathcal{C}_d(y_1, \dots, y_d, y_{d+1}, y_{d+1} + y_{d+2}, \dots, y_{n-1}, y_{n-1} + y_n) \\
& \equiv_d y_d\mathcal{C}_{d-1}(y_1, \dots, y_{d-1}) + y_{d+2}\mathcal{C}_{d-1}(y_1, \dots, y_d, y_{d+1}) + y_{d+4}\mathcal{C}_{d-1}(y_1, \dots, y_d, y_{d+1}, y_{d+1} + y_{d+2}, y_{d+3}) \\
& \quad + \dots + y_n\mathcal{C}_{d-1}(y_1, \dots, y_d, y_{d+1}, y_{d+1} + y_{d+2}, y_{d+3}, y_{d+3} + y_{d+4}, \dots, y_{n-3}, y_{n-3} + y_{n-2}, y_{n-1}).
\end{aligned}$$

Hence $\{y_d, y_{d+2}, \dots, y_n\}$ is a set of $k + 1 = \lfloor \frac{n}{2} \rfloor - \frac{d}{2} + 1$ linear restrictions that reduce $\mathcal{C}_{d,n}$'s degree. By Theorem 8, this is the best possible.

Similarly, when $n = d + 2k + 1$ is odd,

$$\begin{aligned}
& \mathcal{C}_d(x_1, \dots, x_n) \\
& \equiv_d \mathcal{C}_d(x_1, \dots, x_{n-2}) + (x_{n-1} + x_n)\mathcal{C}_{d-1}(x_1, \dots, x_{n-1}) \\
& \equiv_d \dots \\
& \equiv_d \mathcal{C}_d(x_1, \dots, x_{d+1}) + (x_{d+2} + x_{d+3})\mathcal{C}_{d-1}(x_1, \dots, x_{d+2}) + \dots + (x_{n-1} + x_n)\mathcal{C}_{d-1}(x_1, \dots, x_{n-1}) \\
& \equiv_d (x_d + x_{d+1})\mathcal{C}_{d-1}(x_1, \dots, x_d) + \dots + (x_{n-1} + x_n)\mathcal{C}_{d-1}(x_1, \dots, x_{n-1}).
\end{aligned}$$

Now if we switch to the basis in which $y_1 = x_1, \dots, y_d = x_d, y_{d+1} = x_d + x_{d+1}, \dots, y_{n-1} = x_{n-1}, y_n = x_{n-1} + x_n$, then

$$\begin{aligned}
& \mathcal{C}_d(x_1, \dots, x_n) = \mathcal{C}_d(y_1, \dots, y_d, y_d + y_{d+1}, \dots, y_{n-1}, y_{n-1} + y_n) \\
& \equiv_d y_{d+1}\mathcal{C}_{d-1}(y_1, \dots, y_d) + y_{d+3}\mathcal{C}_{d-1}(y_1, \dots, y_d, y_d + y_{d+1}, y_{d+2}) + \dots + \\
& \quad y_n\mathcal{C}_{d-1}(y_1, \dots, y_d, y_d + y_{d+1}, \dots, y_{n-3}, y_{n-3} + y_{n-2}, y_{n-1}).
\end{aligned}$$

Consequently, $\{y_{d+1}, y_{d+3}, \dots, y_n\}$ is a set of $k + 1 = \lfloor \frac{n}{2} \rfloor - \frac{d}{2} + 1$ linear restrictions that reduce $\mathcal{C}_{d,n}$'s degree and meet the bound in Theorem 8.

4 Fourier spectra of GF(2) polynomials

In this Section, we present a framework for computing the Fourier spectrum of a GF(2) polynomial based on its monomials. We suspect that such a formalism was known before but we could not track any previous sources.

For a fixed $S \subseteq [n]$, a collection $\{S_1, \dots, S_k\}$ of k (distinct) subsets of $[n]$ form a k -cover of S if $\cup_{i=1}^k S_i = S$. The main result of this section is the following lemma, which shows that the Fourier coefficients of a GF(2) polynomial can be computed by counting the number of k -covers of subsets

of $[n]$ — for different values of k — using the supports of monomials in the GF(2) polynomial as subsets. Of particular importance is the number of k_{\min} -covers of $[n]$, where k_{\min} is minimum number of subsets that are required to cover $[n]$.

For a family $\mathcal{F} = \{S_i\}_{i \in [m]}$ of subsets S_i of the base set $[n]$ and an index set $M \subseteq [m]$, let $S_M := \cup_{k \in M} S_k$, the union of the subsets with indices in M .

Let $f(x_1, \dots, x_n) = \sum_{i=1}^m x_{S_i}$ be the GF(2) polynomial representation of f . Define a *weight* function $w_f : \{0, 1\}^n \rightarrow \mathbb{Q}$ as

$$w_f(T) = \sum_{M \subseteq [m]: S_M = T} c(M), \quad \text{where } c(M) = \frac{(-2)^{|M|}}{2^{|S_M|}}. \quad (9)$$

Equivalently, if we denote $\mathcal{F} = \{S_i\}_{i \in [m]}$ and let $N_k(T)$ be the number of k -covers of T using sets in \mathcal{F} , then

$$w_f(T) = \frac{1}{2^{|T|}} \sum_{k=1}^m (-2)^k N_k(T). \quad (10)$$

Lemma 12. *Let $f(x_1, \dots, x_n) = \sum_{i=1}^m x_{S_i}$ be a GF(2) polynomial, then the Fourier coefficients of f^\pm are given by*

$$\widehat{f^\pm}(S) = (-1)^{|S|} \sum_{T \supseteq S} w_f(T). \quad (11)$$

Proof. For a Boolean variable $x_i \in \{0, 1\}$, let $\tilde{x}_i = (-1)^{x_i} = 1 - 2x_i$ be its $\{+1, -1\}$ representation, with the inverse transformation given by $x_i = (1 - \tilde{x}_i)/2$. Recall that $f^\pm = 1 - 2f$. We next express f^\pm as a multilinear polynomial over \mathbb{R} from which its Fourier coefficients can be readily read out.

Note that x_S corresponds to $1 - 2 \prod_{i \in S} \frac{1 - \tilde{x}_i}{2}$ and $\prod_{i \in S} \tilde{x}_i$ corresponds to \tilde{x}_S , thus

$$f^\pm(\tilde{x}_1, \dots, \tilde{x}_n) = \prod_{i \in [m]} \left(1 - 2 \prod_{j \in S_i} \frac{1 - \tilde{x}_j}{2} \right) \quad (12)$$

Fact 13. *For $x \in \{-1, 1\}$ and integer $k \geq 1$, we have $(1 - x)^k = 2^{k-1}(1 - x)$.*

By Eq.(12), the Fourier polynomial of f^\pm in terms of \tilde{x} is

$$\begin{aligned} f^\pm(\tilde{x}) &= \prod_{i=1}^m \left(1 - \frac{\prod_{j \in S_i} (1 - \tilde{x}_j)}{2^{|S_i|-1}} \right) \\ &= \sum_{k=0}^m (-1)^k \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq m} \frac{\prod_{j_1 \in S_{i_1}} (1 - \tilde{x}_{j_1}) \prod_{j_2 \in S_{i_2}} (1 - \tilde{x}_{j_2}) \cdots \prod_{j_k \in S_{i_k}} (1 - \tilde{x}_{j_k})}{2^{|S_{i_1}| + |S_{i_2}| + \dots + |S_{i_k}| - k}} \\ &= \sum_{k=0}^m (-1)^k \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq m} \frac{\prod_{j \in S_{i_1} \cup \dots \cup S_{i_k}} (1 - \tilde{x}_j)}{2^{|S_{i_1} \cup \dots \cup S_{i_k}| - k}} \quad (\text{by Fact 13}) \\ &= \sum_{M \subseteq [m]} (-1)^{|M|} \frac{\prod_{j \in S_M} (1 - \tilde{x}_j)}{2^{|S_M| - |M|}} \\ &= \sum_{S \subseteq [n]} (-1)^{|S|} \left(\sum_{M \subseteq [m]: S_M \supseteq S} (-1)^{|M|} \cdot \frac{2^{|M|}}{2^{|S_M|}} \right) \tilde{x}_S, \end{aligned}$$

Since the coefficient of \tilde{x}_S in $f^\pm(\tilde{x})$ is just the Fourier coefficient $\widehat{f^\pm}(S)$, this completes the proof of the lemma. \square

The weight function's value at $[n]$, $w_f([n])$, is the a particularly important term, as it contributes to all the Fourier coefficients of f^\pm . In particular, if the granularity of $w_f([n])$ is larger than the granularity of any other $w_f(T)$, then all Fourier coefficients of f^\pm are non-zero. This will be used to lower bound granularity for different functions in the next two sections.

5 Fourier sparsity of polynomials with complete d -uniform maxonomials

This section is devoted to the proof of the following Fourier sparsity lower bound for polynomials whose maxonomials are the complete d -uniform monomials.

Theorem 14. *Let d be a power of 2. For any degree- d polynomial $f \in \mathbb{F}_2[x_1, \dots, x_n]$ whose maxonomials include all $\binom{[n]}{d}$ degree- d monomials, its Fourier sparsity has the following lower bound*

$$\text{spar}(f) \geq 2^{d \cdot \lfloor n/d \rfloor} - 1 = \Omega(2^n),$$

regardless of the lower degree monomials.

Remark 15. In the rest of this section, we fix $k = \lfloor n/d \rfloor$.

First we apply a restriction to set, say the last $n - kd$ variables in f to zero. This leaves us with a function g on $n' = kd$ variables, and by Item 2 of Lemma 5, $\text{spar}(f) \geq \text{spar}(g)$. Furthermore, the maxonomials of g are still complete d -uniform monomials (now over n' variables).

Let \mathcal{F} be the set of the supports of all monomials in g . In particular, \mathcal{F} contains all d -subsets of $[n']$: $\binom{[n']}{d} \subseteq \mathcal{F}$.

Lemma 16. *The granularity of the weight function at $[n']$ (hence the Fourier coefficient of g at $[n']$) is $\text{gran}(w_g([n'])) = n' - k$.*

Proof. By Lemma 12, as the degrees of all monomials in g are at most d , the minimum number of subsets required from \mathcal{F} to cover $[n']$ is k , therefore

$$\widehat{g^\pm}([n']) = (-1)^{n'} w_g([n']) = (-1)^{n'} \sum_{j=k}^m \frac{(-2)^j N_j([n'])}{2^{n'}}. \quad (13)$$

Claim 17. $N_k([n']) \equiv 1 \pmod{2}$.

Proof. Clearly any k -cover of $[n']$ consists of k distinct sets in $\binom{[n']}{d}$, and there are exactly $\frac{\binom{[n']}{d}}{k!}$ such k -covers. Hence we have

$$\begin{aligned} N_k([n']) &= \frac{\binom{[n']}{d, \dots, d}}{k!} = \frac{1}{k} \binom{kd}{d} \cdot \frac{1}{k-1} \binom{(k-1)d}{d} \cdots 1 \cdot \binom{d}{d} \\ &= \binom{kd-1}{d-1} \cdot \binom{(k-1)d-1}{d-1} \cdots \binom{d-1}{d-1}. \end{aligned}$$

Recall the following Lucas' theorem:

Theorem 18 (Lucas' theorem, c.f. [Fin47]). *Let s and t be non-negative integers and p be a prime. Let $s = s_0 + s_1p + \cdots + s_ip^i$ and $t = t_0 + t_1p + \cdots + t_ip^i$, $0 \leq s_j, t_j < p$, be the base- p expansions of s and t respectively, then*

$$\binom{s}{t} \equiv \prod_{j=0}^i \binom{s_j}{t_j} \pmod{p}.$$

In fact, what we need is the a simple corollary of Lucas' theorem (known as Kummer's theorem) for the special case of $p = 2$: the largest integer j such that 2^j divides $\binom{s}{t}$ is equal to the number of carries that occur when s and $s - t$ are added in the binary.

Since d is a power of 2, the binary representation of $d - 1$ is $\underbrace{1 \cdots 1}_{\log d}$ and the binary representation of $jd - 1 - (d - 1) = (j - 1)d$ is $\cdots \underbrace{0 \cdots 0}_{\log d}$, for every $j \geq 1$. Therefore no carry occurs when adding $(j - 1)d$ to $d - 1$ and thus, by Kummer's theorem, $\binom{jd-1}{d-1} \equiv 1 \pmod{2}$ for all $j \geq 1$. It follows that $N_k([n']) \equiv 1 \pmod{2}$. \square

Finally note that the granularity of the $(j - k + 1)^{\text{st}}$ term in Eq. (13) satisfies

$$\text{gran} \left(\frac{(-2)^j N_j([n'])}{2^{n'}} \right) \leq n' - j < n' - k,$$

for all $j > k$, therefore the first term is the unique term in the sum which has the highest granularity $n' - k$. Hence its granularity is also the granularity of the sum in Eq. (13). This completes the proof of Lemma 16. \square

Now we need the following simple observations, which are simple consequences of Lemma 12.

Fact 19. *Let $g : \{0, 1\}^{n'} \rightarrow \{0, 1\}$ be a degree- d polynomial. Then for any $T \subseteq [n']$, the granularity of the weight function of g at T $\text{gran}(w_g(T))$ is at most $|T| - \lceil |T|/d \rceil$.*

Proof. This follows directly from Eq. (10): since every subset in \mathcal{F} is of size at most d , the minimum number of sets to cover T is $\lceil |T|/d \rceil$. \square

As a simple corollary of Fact 19, we have

Corollary 20. *Let $g : \{0, 1\}^{n'} \rightarrow \{0, 1\}$ be a degree d polynomial. Then for any $T \subseteq [n']$, $\text{gran}(w_g(T)) \leq n' - \lceil |T|/d \rceil$, and equality is only possible for $T = [n']$.*

In other words, if the granularity of $w_g([n'])$ is indeed equal to $n' - \lceil |T|/d \rceil$, then that is the unique highest granularity among all weight values.

Now applying Proposition 4 gives $\text{spar}(g) \geq 2^{n'-k}$. To get the stronger lower bound $2^{n'}$ as claimed, let us combine Lemma 16, Corollary 20 and Eq. (10) in Lemma 12, and observe that not only $w_g([n'])$ has the unique highest granularity among all weights $\{w_g(S)\}_{S \subseteq [n']}$, but also it is included in the Fourier coefficient of $\widehat{g^\pm}(S)$ for every $S \subseteq [n']$. We therefore see that for all $S \subseteq [n']$, $\text{gran}(\widehat{g^\pm}(S)) = n' - k > 0$; consequently $\text{spar}(g^\pm) = 2^{n'}$. It follows that

$$\text{spar}(g) \geq \text{spar}(g^\pm) - 1 = 2^{n'} - 1 = 2^{d \lceil n'/d \rceil} - 1,$$

completing the proof of Theorem 14.

6 Fourier sparsity for functions with sparse maxonomials

In the previous two sections, we see cases that when all $\binom{n}{d}$ monomials of the highest degree appear, then the function has large Fourier sparsity, no matter what other lower-degree monomials exist or not. In this section, we will consider the other end of the spectrum when there are only a small number of the maxonomials, and show that the same phenomena can occur in this case as well.

The first example is the class of functions with disjoint maxonomials.

Proposition 21. *Suppose that $f : \{0, 1\}^n \rightarrow \{0, 1\}$ has $\deg_2(f) = d$ where $d|n$. If there are exactly n/d monomials of degree d , and their supports are pairwise disjoint, then $\text{spar}(f) \geq 2^n - 1$, regardless of the lower degree monomials.*

Proof. We apply Lemma 12 to f and note that the smallest number of sets needed to cover $[n]$ is n/d , achieved by the maxonomials. Thus the Fourier coefficient $\widehat{f^\pm}([n])$ equals $\pm \frac{1}{2^{n-n/d}}$ plus some fractions with denominator 2^k for some $k < n - n/d$. Therefore $\text{gran}(\widehat{f^\pm}([n])) = n - n/d$. Now using a similar argument as the last part of the proof for Theorem 14, we see that all Fourier coefficients of f^\pm are non-zero. Thus $\text{spar}(f) \geq \text{spar}(f^\pm) - 1 \geq 2^n - 1$. \square

The second example extends the first class by allowing “regular” overlaps between maxonomials. Assume that $\deg_2(f) = d$ is an odd prime power, and $d^2|n$. Divide $[n]$ into n/d^2 piles of equal size, with each pile identified with a $d \times d$ grid. All maxonomials are linear functions in a pile. More precisely, for the first pile $[d] \times [d]$, for each pair $(a, b) \in \mathbb{F}_d^2$, define univariate polynomial $p_{a,b} \in \mathbb{F}[x]$ by $p_{a,b}(x) = ax + b$. Now define sets

$$S_{a,b} = \{(0, p(0)), (1, p(1)), \dots, (d-1, p(d-1))\}.$$

The first pile thus has d^2 sets inside. Similarly define d^2 sets for each other pile. These sets are supports of the maxonomials. Note that there are $d^2 \cdot n/d^2 = n$ maxonomials, a number much smaller than the possible number of lower degree monomials, which is $\sum_{i=0}^{d-1} \binom{n}{i}$. Yet the next theorem says that this small number of maxonomials determines a large Fourier sparsity, regardless of how the vast majority of other (lower-degree) terms behave.

Theorem 22. *For any function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ with the maxonomials defined as above, $\text{spar}(f) \geq 2^n - 1$, regardless of the lower degree monomials.*

Proof. Clearly the set $[n]$ can be partitioned using supports of n/d maxonomials. We will show that the number of such partitions is d^{n/d^2} , which is an odd number given that d is odd.

Since the piles are disjoint and all maxonomials are defined within each pile, it suffices to show that there are d ways of partitioning each pile into maxonomials. We consider the first pile and the same argument applies to others. Note that for each fixed a , if we vary b over \mathbb{F}_d , then we get d maxonomials that are pairwise disjoint. Since there are d different choices of a , there are at least these d ways to partition the pile into d maxonomials. We next show that there are actually no other partition of the pile using d maxonomials. Indeed, assume that a partition uses d maxonomials and not all these maxonomials have the same a , then there are two maxonomials corresponding to $a_1x + b_1$ and $a_2x + b_2$ and $a_1 \neq a_2$. But now these two “lines” intersect at exactly one point $x = (a_1 - a_2)^{-1}(b_1 - b_2)$, where the existence of $(a_1 - a_2)^{-1}$ uses the assumption that $a_1 \neq a_2$. Note the trivial fact that the union of d maxonomials of degree d is at most d^2 , and it is d^2 only if they are pairwise disjoint. So the existence of intersecting maxonomials in the selected d

maxonomials make them impossible to cover the d^2 points in the pile. This shows that the number of partitions of one pile using d maxonomials is exactly d , and thus the number of covers of $[n]$ using n/d^2 maxonomials is d^{n/d^2} . Now apply a similar argument as the last part of the proof for Theorem 14, we see that $\text{spar}(f) \geq \text{spar}(f^\pm) - 1 \geq 2^n - 1$. \square

It would be nice to also pin down the linear rank of the functions with the maxonomials defined as above. What we are able to say at this moment is an upper bound only.

$$\text{lin-rank}(f) \leq n/d.$$

Indeed, for each pile, we can pick the first column of variables and set them all to 0. This makes all maxonomials vanish, and thus decreases the degree by at least 1.

6.1 Granularity upper bound for low-degree polynomials

Note that there is a gap of factor 2 in characterizing the logarithm of Fourier sparsity of a Boolean function by means of its granularity (cf. Proposition 4). Note also that both lower and upper bounds in Proposition 4 are tight, but one is attained by the AND function (a degree- n polynomial) and the other by any bent function, e.g. the Inner Product function (a degree-2 polynomial). It thus natural to conjecture that, for any low-degree polynomial $f(x)$, although $\text{spar}(f)$ can be as large as 2^n , the granularity of $f(x)$ is always bounded away from n . We now apply our technique developed in Section 4 to prove the following upper bound for the granularity of low-degree polynomials.

Theorem 23. *For any Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$, if $d = \deg_2(f)$ is the \mathbb{F}_2 -degree of f , then $\text{gran}(f^\pm) \leq n - \lceil \frac{n}{d} \rceil$, and consequently, $\text{gran}(f) \leq n - \lceil \frac{n}{d} \rceil + 1$.*

Proof. Suppose $\widehat{f^\pm}(T)$ achieves $\text{gran}(f^\pm)$, i.e., $\widehat{f^\pm}(T) = c/2^{\text{gran}(f^\pm)}$ for some odd integer c . Without loss of generality, we may assume that $T \neq \emptyset$. Actually, if $\widehat{f^\pm}(0)$ is the single Fourier coefficient that achieves $\text{gran}(f^\pm)$, then the sum of the squares of all Fourier coefficients of f^\pm would be a rational number with granularity $2\text{gran}(f^\pm)$ instead of 1, contradicting Parseval's identity.

Now we apply an invertible linear map L such that $(L^T)^{-1}(T) = [n]$. Denote $f \circ L$ by g . By Fact 6, g is also a polynomial of degree d . Moreover, by Fact 7, we have that $\widehat{g^\pm}([n]) = \widehat{f^\pm}(T)$.

Now suppose $g(x) = \sum_{i=1}^m \prod_{j \in S_i} x_j$, where $|S_j| \leq d$ for every $1 \leq j \leq m$. Applying Lemma 12 and notice that, since $|S_j| \leq d$, the minimum number k such that there exists a collection of k subsets from $\{S_j\}_{j \in [m]}$ that cover $[n]$ is $k \geq \lceil \frac{n}{d} \rceil$. Therefore, by Eq. (11),

$$\widehat{g^\pm}([n]) = (-1)^n w_g([n]) = (-1)^n \sum_{j=k}^m \frac{(-2)^j N_j([n])}{2^n}.$$

Note that the granularity of the j^{th} term in the above summation is at most $n - j$ (we only have inequality here as $N_j([n])$ may be an even number), and the granularity of a sum of rational numbers is at most the maximum granularity in the summands:

$$\text{gran} \left(\sum_{j=1}^{\ell} y_j \right) \leq \max_{1 \leq j \leq \ell} \text{gran}(y_j),$$

where $y_j \in \mathbb{Q}$ for $1 \leq j \leq \ell$, we therefore have $\text{gran}(\widehat{g^\pm}([n])) \leq n - k = n - \lceil \frac{n}{d} \rceil$. This finally gives

$$\text{gran}(f^\pm) = \text{gran}(\widehat{f^\pm}(T)) = \text{gran}(\widehat{g^\pm}([n])) \leq n - \lceil \frac{n}{d} \rceil.$$

The upper bound of the granularity of f follows from the easy fact that $\text{gran}(f) \leq \text{gran}(f^\pm) + 1$. \square

Acknowledgements

We are indebted to the anonymous reviewers for their detailed helpful comments.

References

- [AFH12] Anil Ada, Omar Fawzi, and Hamed Hatami. Spectral norm of symmetric functions. In *Proceedings of the 15th International Workshop on Approximation, Randomization, and Combinatorial Optimization*, pages 338–349, 2012.
- [BC99] Anna Bernasconi and Bruno Codenotti. Spectral analysis of boolean functions as a graph eigenvalue problem. *IEEE Transactions on Computers*, 48(3):345–351, 1999.
- [CT13] Gil Cohen and Avishay Tal. Two structural results for low degree polynomials and applications. *ECCC*, TR13-145, 2013.
- [Fin47] Nathan Fine. Binomial coefficients modulo a prime. *American Mathematical Monthly*, 54:589–592, 1947.
- [GOS⁺11] Parikshit Gopalan, Ryan O’Donnell, Rocco Servedio, Amir Shpilka, and Karl Wimmer. Testing Fourier dimensionality and sparsity. *SIAM Journal on Computing*, 40(4):1075–1100, 2011.
- [KN97] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, Cambridge, UK, 1997.
- [LLZ11] Ming Lam Leung, Yang Li, and Shengyu Zhang. Tight bounds on the communication complexity of symmetric XOR functions in one-way and SMP models. In *Proceedings of the 8th Annual Conference on Theory and Applications of Models of Computation*, pages 403–408, 2011.
- [Lov14a] Shachar Lovett. Communication is bounded by root of rank. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 842–846, 2014.
- [Lov14b] Shachar Lovett. Recent advances on the log rank conjecture. In *Bulletin of EATCS*, 2014.
- [LS88] László Lovász and Michael E. Saks. Lattices, Möbius functions and communication complexity. In *Proceedings of the 29th Annual Symposium on Foundations of Computer Science*, pages 81–90, 1988.

- [LZ10] Troy Lee and Shengyu Zhang. Composition theorems in communication complexity. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming*, pages 475–489, 2010.
- [LZ13] Yang Liu and Shengyu Zhang. Quantum and randomized communication complexity of XOR functions in the SMP model. *ECCC*, 20(10), 2013.
- [MO09] Ashley Montanaro and Tobias Osborne. On the communication complexity of XOR functions, 2009. <http://arxiv.org/abs/0909.3392v2>.
- [NW95] Noam Nisan and Avi Wigderson. On rank vs. communication complexity. *Combinatorica*, 15(4):557–565, 1995.
- [STV14] Amir Shpilka, Avishay Tal, and Ben Lee Volk. On the structure of boolean functions with small spectral norm. In *Proceedings of the 5th Innovations in Theoretical Computer Science*, 2014.
- [TWXZ13] Hing Yin Tsang, Chung Hoi Wong, Ning Xie, and Shengyu Zhang. Fourier sparsity, spectral norm, and the log-rank conjecture. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science*, pages 658–667, 2013.
- [Yao79] Andrew Yao. Some complexity questions related to distributive computing. In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing*, pages 209–213, 1979.
- [ZS09] Zhiqiang Zhang and Yaoyun Shi. Communication complexities of symmetric XOR functions. *Quantum Information & Computation*, 9(3):255–263, 2009.