Fourier sparsity, spectral norm, and the Log-rank conjecture

Hing Yin Tsang* hytsang@cse.cuhk.edu.hk
Chung Hoi Wong* hoiy927@gmail.com
Ning Xie† nxie@cis.fiu.edu
Shengyu Zhang* syzhang@cse.cuhk.edu.hk

***Abstract*—We study Boolean functions with sparse Fourier spectrum or small spectral norm, and show their applications to the Log-rank Conjecture for XOR functions $f(x \oplus y)$ — a fairly large class of functions including well studied ones such as Equality and Hamming Distance. The rank of the communication matrix M_f for such functions is exactly the Fourier sparsity of f. Let $d = \deg_2(f)$ be the $\mathbb{F}_2$-degree of f and $\mathsf{D}^{\mathsf{CC}}(f \circ \oplus)$ stand for the deterministic communication complexity for $f(x \oplus y)$. We show that**

1) $D^{CC}(f \circ \oplus) = O(2^{d^2/2} \log^{d-2} \|\hat{f}\|_1)$. **In particular, the Log-rank conjecture holds for XOR functions with constant $\mathbb{F}_2$-degree.**
2) $D^{CC}(f \circ \oplus) = O(d\|\hat{f}\|_1) = \tilde{O}(\sqrt{\mathtt{rank}(M_f)})$. **This improves the (trivial) linear bound by nearly a quadratic factor.**

We obtain our results through a degree-reduction protocol based on a variant of polynomial rank, and actually conjecture that the communication cost of our protocol is at most $\log^{O(1)} \mathtt{rank}(M_f)$. The above bounds are obtained from different analysis for the number of parity queries required to reduce f's $\mathbb{F}_2$-degree. Our bounds also hold for the parity decision tree complexity of f, a measure that is no less than the communication complexity.

Along the way we also prove several structural results about Boolean functions with small Fourier sparsity $\|\hat{f}\|_0$ or spectral norm $\|\hat{f}\|_1$, which could be of independent interest. For functions f with constant $\mathbb{F}_2$-degree, we show that: 1) f can be written as the summation of quasi-polynomially many indicator functions of subspaces with $\pm$-signs, improving the previous doubly exponential upper bound by Green and Sanders; 2) being sparse in Fourier domain is polynomially equivalent to having a small parity decision tree complexity; and 3) f depends only on $\mathrm{polylog}\|\hat{f}\|_1$ linear functions of input variables. For functions f with small spectral norm, we show that: 1) there is an affine subspace of co-dimension $O(\|\hat{f}\|_1)$ on which $f(x)$ is a constant, and 2) there is a parity decision tree of depth $O(\|\hat{f}\|_1 \log \|\hat{f}\|_0)$ for computing f.

***Keywords*-Fourier analysis, Fourier sparsity, Log-rank conjecture, low-degree polynomials**

I. Introduction

Fourier analysis of Boolean functions. Fourier analysis has been widely used in theoretical computer science to study Boolean functions with applications in PCP, property testing, learning, circuit complexity, coding theory, social choice theory and many more; see [O'D12] for a comprehensive survey. The Fourier coefficients of a Boolean function measure the function's correlations with parity functions; the distribution as well as various norms of Fourier spectrum have been found to be related to many complexity measures of the function. However, another natural measure, Fourier sparsity – *i.e.* the number of non-zero Fourier coefficients – has been much less studied. It seems to be of fundamental interest to understand properties of functions that are Boolean in the function domain and, at the same time, sparse in the Fourier domain. In particular, what Boolean functions have sparse Fourier spectra? Being sparse in the Fourier domain should imply that the function is simple, but in which aspects? Gopalan *et al.* [GOS+11] studied the problem of testing Fourier sparsity and low-dimensionality and revealed several interesting structural results for Boolean functions having or close to having sparse Fourier spectra. In a related setting, Green and Sanders [GS08] showed that Boolean functions with a small spectral norm (*i.e.* the ℓ_1-norm of the Fourier spectrum) can be decomposed into a small number of signed indicator functions of subspaces. However, the number of subspaces in their bound is doubly exponential in the function's spectral norm, making their result hard to apply in many computer science related problems.

The Log-rank Conjecture in communication complexity. In a different vein, Fourier sparsity also naturally arises in the study of Log-rank Conjecture in communication complexity. Communication complexity quantifies the minimum amount of communication needed for computation on inputs distributed to different parties [Yao79], [KN97]. In a standard scenario, two parties Alice and Bob each hold an input x and y, respectively, and they desire to compute a function f on input (x, y) by as little communication as possible. Apart from its own interest as a question about distributed computation, communication complexity has also found numerous applications in proving lower bounds in complexity theory.

Of particular interest are lower bounds of communication complexity, and one of the most widely used methods is based on the rank of the communication matrix $M_f = [f(x, y)]_{x,y}$; see [LS09] for an extensive survey on classical and quantum lower bounds proved by rank and its variations. Since it was shown 30 years ago [MS82]

*The Chinese University of Hong Kong, Shatin, NT, Hong Kong. Work supported by Research Grants Council of the Hong Kong S.A.R. (Project no. CUHK418710, CUHK419011)

† SCIS, Florida International University, Miami, FL 33199, USA. Part of the work was done when the author was at CSAIL, MIT and was supported by NSF awards CCF-1217423 and CCF-1065125.

that $\log \texttt{rank}(M_f)$ is a lower bound of the deterministic communication complexity $\mathsf{D}^{\mathsf{CC}}(f)$, the tightness of the lower bound has long been an important open question. The Log-rank Conjecture, proposed by Lovász and Saks [LS88], asserts that the lower bound is polynomially tight for all total Boolean functions f – namely $\mathsf{D}^{\mathsf{CC}}(f) \le \log^c \texttt{rank}(M_f)$ for some absolute constant c. As one of the most important problems in communication complexity, the conjecture links communication complexity – a combinatorially defined quantity, to matrix rank – a much better understood measure in linear algebra. Should the conjecture hold, understanding the communication complexity is more or less reduced to a usually much easier task of calculating matrix ranks. The conjecture is also known to be equivalent to many other conjectures [LS88], [Lov90], [Val04], [ASTS+03].

Despite its importance, Log-rank Conjecture is also notoriously hard to attack. Nisan and Wigderson [NW95] showed that to prove the conjecture, it is sufficient to show a seemingly weaker statement about the existence of a large monochromatic rectangle. In the same paper, they also exhibited an example f for which $\log \texttt{rank}(M_f) = O(\mathsf{D}^{\mathsf{CC}}(f)^\alpha)$ where $\alpha = \log_3 2 = 0.63...$, later improved by Kushilevitz to $\alpha = \log_6 3 = 0.61...$ (also in [NW95]). The best upper bound for the $\mathsf{D}^{\mathsf{CC}}(f)$ in terms of rank is $\mathsf{D}^{\mathsf{CC}}(f) \le (\log \frac{4}{3})\texttt{rank}(M_f)$ [KL96], [Kot97]. Recently, assuming the Polynomial Freiman-Ruzsa conjecture in additive combinatorics, Ben-Sasson, Lovett and Ron-Zewi gave in [BSLRZ12] a better upper bound $\mathsf{D}(f) \le O(\texttt{rank}(M_f)/\log \texttt{rank}(M_f))$.

Communication complexity of XOR functions. In view of the difficulty of the Log-rank Conjecture in its full generality, Shi and Zhang [ZS10] initiated the study of communication complexity of a special class of functions called *XOR functions*.

Definition 1. *We say $F(x,y) : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is an XOR function if there exists an $f : \{0,1\}^n \to \{0,1\}$ such that for all x and y in $\{0,1\}^n$, $F(x,y) = f(x \oplus y)$, where $\oplus$ is the bit-wise XOR. Denote F by $f \circ \oplus$.*

XOR functions include important examples such as Equality and Hamming Distance, and the communication complexity of XOR functions has recently drawn an increasing amount of attention [ZS09], [ZS10], [LZ10], [MO10], [LLZ11], [SW12], [LZ13]. In general, the additional symmetry in the communication matrix M_F should make Log-rank Conjecture easier for XOR functions. In particular, a very nice feature of XOR functions is that the rank of the communication matrix M_F is exactly the Fourier sparsity of f, the number of nonzero Fourier coefficients of f.

Proposition 1 ([BC99]). *For XOR functions $F(x,y) = f(x \oplus y)$, it holds that $\texttt{rank}(M_F) = \|\hat{f}\|_0$.*

Therefore the Log-rank Conjecture for XOR functions is equivalent to the question that whether every Fourier sparse[1] function f admits an efficient communication protocol to compute $f(x \oplus y)$.

However, the Log-rank conjecture seems still very difficult to study even for this special class of functions. One nice approach proposed in [ZS10] is to first design an efficient *parity decision tree* (PDT) for computing f, and then to simulate it by a communication protocol. Parity decision trees allow to query the parity of any subset of input variables (instead of just one input variable as in usual decision trees). A communication protocol can exchange two bits $\ell(x)$ and $\ell(y)$ (here $\ell(\cdot)$ is an arbitrary linear function) to simulate one query $\ell(x \oplus y)$ in a PDT, thus $\mathsf{D}^{\mathsf{CC}}(f \circ \oplus)$ is at most twice of $\mathsf{D}_\oplus(f)$, the parity decision tree complexity of f. It is therefore sufficient to show that $\mathsf{D}_\oplus(f) \le \log^{O(1)} \|\hat{f}\|_0$ for all f to prove the Log-rank Conjecture for XOR functions. Parity decision tree complexity is an interesting complexity in its own right, with connections to learning [KM93] and other parity complexity measures such as parity certificate complexity and parity block sensitivity [ZS10]. This approach is also appealing for the purpose of understanding Boolean functions with sparse Fourier spectra. It is not hard to see that small $\mathsf{D}_\oplus(f)$ implies Fourier sparsity; now if $\mathsf{D}_\oplus(f) \le \log^{O(1)} \|\hat{f}\|_0$ is true, then functions with small Fourier sparsity also have short parity decision trees. Thus the elusive property of being Fourier sparse is roughly equivalent to the combinatorial and computational property of having small PDT.

Back to the Log-rank conjecture, though upper bounds for $\mathsf{D}_\oplus(f)$ translate to efficient protocols for $\mathsf{D}^{\mathsf{CC}}(f \circ \oplus)$, the task of designing efficient PDT algorithms itself does not seem to be an easy task. To see this, let us examine the effect of parity queries. Each query "$t \cdot x = ?$" basically generates two subfunctions through restriction, and its effect on the Fourier domain can be shown to be $\hat{f}_b(s) = \hat{f}(s) + (-1)^b \hat{f}(s+t)$, where f_b is the subfunction obtained from restricting f on the half space $\{x : t \cdot x = b\}$. Thus the process is like to fold the spectrum of f along the line t, and we hope that the folding has many "collisions" in nonzero Fourier coefficients, namely many $s \in \texttt{supp}(\hat{f})$, with $s+t \in \texttt{supp}(\hat{f})$ as well. In general, small $\mathsf{D}_\oplus(f)$ implies that many Fourier coefficients[2] are "well aligned" with respect to a subspace V with a small co-dimension, so that querying basis of $V^\perp$ makes those Fourier coefficients collide. But the question is—*Where is the subspace*?

Note that $\mathsf{D}_\oplus(f)$ is invariant under change of input basis, thus one tempting way to upper bound $\mathsf{D}_\oplus(f)$ is to first rotate input basis, and then (under the new basis) use the fact that the standard decision tree complexity $\mathsf{D}(f)$ is at

[1] Note that if the Fourier sparsity of f is large, say $2^{n^{\Omega(1)}}$, then Log-rank Conjecture is vacuously true for f, as the communication complexity of any function is at most $O(n)$.

[2] Technically, we mean characters with the corresponding Fourier coefficients being nonzero.

most $O(\deg(f)^3)$ [Mid04], where the $\deg(f)$ is the (Fourier) degree ($\max_{s:\hat{f}(s)\neq 0}|s|$) of f [BdW02]. Thus if $\deg(f) = \log^{O(1)}\|\hat{f}\|_0$, then $\mathsf{D}_\oplus(f) \leq \mathsf{D}(f) \leq \log^{O(1)}\|\hat{f}\|_0$. However, one should also note that this approach cannot handle all the Fourier sparse functions because, as shown in [ZS10], there exist functions f such that $\mathsf{D}_\oplus(f) \leq \log_2 n + 4$ but $\mathsf{D}(f) \geq n/4$, the latter holds even under an arbitrary basis change (*i.e.* $\min_L \mathsf{D}(Lf) \geq n/4$ where $Lf(x) = f(Lx)$).

A. Our approach, ideas, and results

Result 1: Main protocol and general conjecture. In previous studies of parity decision tree, one needs to upper bound the number of queries for *all* possible execution paths. In this paper, we show that it suffices to prove the existence of *one* short path! To put this into context, we need the concept of polynomial rank. View a Boolean function $f : \{0,1\}^n \to \{0,1\}$ as a polynomial in $\mathbb{F}_2[x_1, ..., x_n]$. Call the degree of this polynomial the $\mathbb{F}_2$-degree, denoted as $\deg_2(f)$. The polynomial rank of f is the minimum number r *s.t.* f can be written as a function of r polynomials with lower degree. Here we propose a new concept called *linear rank*: For a polynomial f, its linear rank $\text{lin-rank}(f)$ is the minimum number r *s.t.* f can be written as

$$f = \ell_1 f_1 + \cdots + \ell_r f_r + f_0, \tag{1}$$

where each ℓ_i is a linear function in x and each f_i is a function of $\mathbb{F}_2$-degree at most $\deg_2(f) - 1$. Linear rank is in general larger than polynomial rank, but linear rank better suits our need for designing efficient PDT algorithms. Now we will describe a simple PDT algorithm: query all $\ell_i(x)$ and get answers a_i, and we then face a new function $f' = \sum_{i=1}^r a_i f_i + f_0$. Recurse on this function. Note that from f to f', the $\mathbb{F}_2$-degree is reduced by at least 1, and one can also show that the Fourier sparsity of f' is also at most that of f. Finally, it is known that $\deg_2(f) \leq \log\|\hat{f}\|_0$. Putting these nice properties together, we know that, as long as the linear rank of an arbitrary function f is upper bounded by $\log^{O(1)}\|\hat{f}\|_0$, so is $\mathsf{D}_\oplus(f)$.

Conjecture 2. *For all $f : \{0,1\}^n \to \{0,1\}$, we have $\text{lin-rank}(f) = O(\log^c(\|\hat{f}\|_0))$ for some $c = O(1)$.*

Theorem 3. *If Conjecture 2 is true, then*

1) *All Boolean functions with small $\|\hat{f}\|_0$ have small parity decision tree complexity as well: $\mathsf{D}_\oplus(f) = O(\log^{c+1}(\|\hat{f}\|_0))$.*
2) *The Log-rank Conjecture is true for all XOR functions: $\mathsf{D}^{\mathsf{CC}}(f\circ\oplus) = O(\log^{c+1}(\|\hat{f}\|_0))$.*

Result 2: low degree polynomials. Next we focus on upper bounding the linear rank, starting from small degrees. For degree-2 polynomials, the classic theorem by Dickson implies that $\text{lin-rank}(f) = O(\log\|\hat{f}\|_0)$. For degree-3 polynomials, Haramaty and Shpilka proved in [HS10] that $\text{lin-rank}(f) = O(\log^2(1/\|f\|_{U^3})) = O(\log^2(1/\texttt{bias}(f)))$.[3] By a proper shift, we can make $\texttt{bias}(f) \geq 1/\sqrt{\|\hat{f}\|_0}$ and thus get $\text{lin-rank}(f) = O(\log^2\|\hat{f}\|_0)$. For degree-4 polynomials, however, the bound is exponentially worse even for the polynomial rank [HS10], and there were no results for higher degrees. A natural question is: Can one prove the $\text{lin-rank}(f) = O(\log^{O(1)}\|\hat{f}\|_0)$ for degree-4 polynomials? Further, if it is too challenging to prove $\text{lin-rank}(f) \leq \log^{O(1)}\|\hat{f}\|_0$ for general degree d (which is at most $\log\|\hat{f}\|_0$), can one prove it for constant-degree polynomials (even if the power $O(1)$ is a tower of 2's of height d)? In this paper, we show that this is indeed achievable. Actually, we can even replace the ℓ_0-norm by ℓ_1-norm[4] of $\hat{f}$ in the bound, and the dependence on d is "only" singly exponential.

Lemma 4. *For all Boolean functions f with $\mathbb{F}_2$-degree d, we have $\text{lin-rank}(f) = O(2^{d^2/2}\log^{d-2}\|\hat{f}\|_1)$.*

The lemma immediately implies the following two results.

Theorem 5. *If f is a Boolean function of constant $\mathbb{F}_2$-degree, then $\mathsf{D}^{\mathsf{CC}}(f\circ\oplus) \leq \log^{O(1)}(\texttt{rank}(M_{f\circ\oplus}))$.*

Recursively expanding Eq.(1) and applying the bound on ranks in Lemma 4 gives that

Corollary 6. *Every Boolean function f of $\mathbb{F}_2$-degree d depends only on $O(2^{d^3/2}\log^{d^2}\|\hat{f}\|_1)$ linear functions of input variables.*

Another corollary is the following. Green and Sanders proved that any $f : \{0,1\}^n \to \{0,1\}$ can be written as $f = \sum_{i=1}^T \pm\mathbb{1}_{V_i}$, where $T = 2^{2^{O(\|\hat{f}\|_1^4)}}$ and each $\mathbb{1}_{V_i}$ is the indicator function of the subspace V_i. For constant degree polynomials, we can improve their doubly-exponential bound to quasi-polynomial.

Corollary 7. *If $f : \{0,1\}^n \to \{0,1\}$ has constant $\mathbb{F}_2$-degree, then $f = \sum_{i=1}^T \pm\mathbb{1}_{V_i}$ where $T = 2^{\log^{O(1)}\|\hat{f}\|_1}$ and each $\mathbb{1}_{V_i}$ is the indicator function of the subspace V_i.*

The proof of Lemma 4 follows the general approach laid out in the Main protocol, *i.e.*, a rank-based degree-reduction process, with several additional twists. First, to find a "good" affine subspace restricted on which f becomes a lower degree polynomial, we recursively apply the derivatives of f to guide our search. Second, even though our final goal is to reduce the degree of f, we actually achieve this through reducing the spectral norm of f. This is done by studying the effect of restriction on two *non-Boolean* functions. Last, in the induction step, we in fact need to prove a

[3] Recall that for $f : \{0,1\}^n \to \{0,1\}$, the *bias* of f is defined to be $\texttt{bias}(f) = |\mathbf{Pr}(f(x) = 0) - \mathbf{Pr}(f(x) = 1)|$.

[4] Strictly speaking, in view of the corner case of $\|\hat{f}\|_1 = 1$, one should replace $\log(\|\hat{f}\|_1)$ by $\log(\|\hat{f}\|_1 + 1)$. But like in most previous papers, we omit the "+1" term for all $\|\hat{f}\|_1$ in this paper for simplicity of notation.

stronger statement about a chain inequality involving rank, minimum parity 0-certificate complexity $\mathsf{C}^0_{\oplus,\min}$, minimum parity 1-certificate complexity $\mathsf{C}^1_{\oplus,\min}$ and parity decision tree complexity $\mathsf{D}_\oplus$. And the induction is used in a "cyclic" way: we upper bound $\min\{\mathsf{C}^0_{\oplus,\min}, \mathsf{C}^1_{\oplus,\min}\}$ by induction on $\max\{\mathsf{C}^0_{\oplus,\min}, \mathsf{C}^1_{\oplus,\min}\}$, which upper bounds lin-rank. This can then be used to show that $\mathsf{D}_\oplus$ is small, which in turn upper bounds $\max\{\mathsf{C}^0_{\oplus,\min}, \mathsf{C}^1_{\oplus,\min}\}$ to finish the inductive step.

Result 3: functions with small spectral norms. While Theorem 5 handles the low-degree case, the bound deteriorates exponentially with the $\mathbb{F}_2$-degree. Via a different approach, we are able to upper bound $\text{lin-rank}(f)$ by the ℓ_1-norm of $\hat{f}$.

Lemma 8. *For all* $f : \{0,1\}^n \to \{0,1\}$, *we have* $\text{lin-rank}(f) \le O(\|\hat{f}\|_1)$.

In fact we prove a slightly stronger result that there exists an affine subspace of codimension at most $O(\|\hat{f}\|_1)$ on which f is constant. In other words, if a Boolean function has small spectral norm then it is constant on a large affine subspace.

The proof of the lemma uses a greedy algorithm that always makes the two largest Fourier coefficients to collide (with the same sign). Exploiting the property that f is Boolean, one can show that this greedy folding either significantly increases the largest Fourier coefficient, or decreases $\|\hat{f}\|_1$ by a constant.

The lemma immediately implies the following result for general (not necessarily XOR) functions.

Theorem 9. *For all* $f : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}$, *we have* $\mathsf{D}^{\mathsf{CC}}(f) \le 2\mathsf{D}_\oplus(f) = O(\deg_2(f) \cdot \|\hat{f}\|_1)$.

In [Gro97], Grolmusz gave a public-coin randomized protocol with communication cost $O(\|\hat{f}\|_1^2)$. The above theorem gives a *deterministic* protocol, and the bound is better for functions f with $\deg_2(f) = o(\|\hat{f}\|_1)$.

Another implication of Lemma 8 is that the communication complexity of $f \circ \oplus$ is at most the square root of the matrix rank.

Theorem 10. *For all* $f : \{0,1\}^n \to \{0,1\}$,

$$\mathsf{D}^{\mathsf{CC}}(f \circ \oplus) = O(\deg_2(f) \cdot \|\hat{f}\|_1) = \tilde{O}\left(\sqrt{\mathtt{rank}(M_{f\circ\oplus})}\right).$$

The upper bound of $\mathtt{rank}/\log\mathtt{rank}$ in [BSLRZ12] improves the trivial linear bound by a log factor for all Boolean functions, assuming the Polynomial Freiman-Ruzsa conjecture. In comparison, our bound of $\sqrt{\mathtt{rank}}\log\mathtt{rank}$ is only for XOR functions, but it improves the linear bound by a polynomial factor, and it is unconditional.

It is also interesting to note that, for any fixed Boolean function f, at least one of the above two theorems gives a desirable result: either $\|\hat{f}\|_1 \ge \log^k \|\hat{f}\|_0$ where k is a big constant, then Theorem 9 improves Grolmusz's bound almost quadratically (since $\deg_2(f) \le \log \|\hat{f}\|_0$); or $\|\hat{f}\|_1 \le \log^k \|\hat{f}\|_0$, then Theorem 10 confirms the Log-rank conjecture for $f \circ \oplus$!

Result 4: functions with a light Fourier tail. Our last result deals with Boolean functions whose Fourier spectrum has a light tail. We call a function $f : \{0,1\}^n \to \{+1,-1\}$ μ-close to s-sparse in ℓ_2 if $\sum_{i>s} \hat{f}(s_i)^2 \le \mu^2$, where $|\hat{f}(s_1)| \ge \ldots \ge |\hat{f}(s_N)|$. We say two functions $f, g : \{0,1\}^n \to \{+1,-1\}$ are ϵ-close if $\mathbf{Pr}_x[f(x) \ne g(x)] \le \epsilon$.

Theorem 11. *If* $f : \{0,1\}^n \to \{+1,-1\}$ *is* μ*-close to* s*-sparse in* ℓ_2, *where* $\mu \le \frac{\log^{O(1)}\|\hat{f}\|_0}{\sqrt{\|\hat{f}\|_0}}$ *and* $s \le \log^{O(1)} \|\hat{f}\|_0$, *then* $\mathsf{D}_\oplus(f) \le \log^{O(1)} \|\hat{f}\|_0$.

The proof of this theorem uses Chang's lemma about large Fourier coefficients of low-density functions, and a "rounding" lemma from [GOS+11].

B. Related work

The Log-rank Conjecture for XOR functions was shown to be true for symmetric functions [ZS09], linear threshold functions (LTFs), monotone functions [MO10], and AC^0 functions [KS13]. These results fall into two categories. The first one, including symmetric functions and LTFs, is that the rank of the communication matrix (*i.e.* the Fourier sparsity) is so large, that the Log-rank conjecture trivially holds. The second one, including monotone functions and AC^0 functions, is that even the Fourier degree is small, thus the standard decision tree complexity $\mathsf{D}(f)$ is already upper bounded by the poly-logarithmic of the matrix rank. But as we mentioned, there are functions that have small Fourier sparsity *and* high Fourier degree (even after basis rotation). These hardcore cases have not been studied before this work.

In [HS10], Haramaty and Shpilka proved that $\text{lin-rank}(f) = O(\log^2(1/\|f\|_{U^3})) = O(\log^2(1/\mathtt{bias}(f)))$ for degree-3 polynomials. For degree-4 polynomials, however, the bound is exponentially worse even for the polynomial rank [HS10], and there were no results for higher degrees. In comparison, our Lemma 4 gives a polylog upper bound for $\text{lin-rank}(f)$ of all constant degree functions f, but the polylog is in $\|\hat{f}\|_1$ rather than in $\mathtt{bias}(f)$ or Gower's norm ([Gow98], [Gow01], [AKK+05]).

Though Boolean functions with a sparse Fourier spectrum seem to be a very interesting class of functions to study, not many properties are known. It is shown in [GOS+11] that the Fourier coefficients of a Fourier sparse function have large "granularity" and functions that are very close to Fourier sparse can be transformed into one through a "rounding off" procedure. Furthermore, they proved that one can use $2\log\|\hat{f}\|_0$ random linear functions to partition the character space so that, with high probability, each bucket contains at most one nonzero Fourier coefficient. This does not help our problem since what we need is exactly the opposite: to group Fourier coefficients into buckets so that a small number of foldings would make many of them to collide (and thus reducing the Fourier sparsity quickly).

Let $A = \mathtt{supp}(\hat{f})$ be the support of f's Fourier spectrum. One way of designing the parity query is to look for a "heavy hitter" t of set $A + A$, *i.e.* t with many $s_1, s_2 \in A$ and $s_1 + s_2 = t$. If such t exists, then querying the linear function $\langle t, x\rangle$ reduces the Fourier sparsity a lot. One natural way to show the existence of a heavy hitter is by proving that $|A+A|$ is small. Turning this around, one may hope to show that if it is large, then the function is not Fourier sparse or has some special properties to be used. The size of $|A+A|$ has been extensively studied in additive combinatorics, but it seems that all related studies are concerned with the low-end case, in which $|A+A| \le k|A|$ for very small (usually constant) k. Thus those results do not apply to our question.

Recall that the standard polynomial rank is the minimum r *s.t.* f can be expressed as a function F of r *lower degree* polynomials $f_1, ..., f_r$. A nice result for polynomial rank is that large bias implies low polynomial rank [GT09], [KL08]: the rank is a function of the bias and degree only, but not of the input size n. This is, however, insufficient for us because a Fourier sparse function may have very small bias. Furthermore, the dependence of the rank on the degree is a very rapidly growing function (faster than a tower of 2's of height d), while our protocol has "only" single exponential dependence of d.

The work of [SV13]. After completing this work independently, the very recent work [SV13] came to our attention, which studies PDT complexity of functions with small spectral norm. The authors show $\mathsf{C}_{\oplus,\min}(f) \le O(\|\hat{f}\|_1^2)$ and $\mathsf{D}_{\oplus}(f) = O(\|\hat{f}\|_1^2 \log \|\hat{f}\|_0)$. In comparison, our Lemma 28 and Theorem 10 are at least quadratically better. The paper [SV13] also studies the size of PDT and approximation of Boolean functions, which are not studied in this paper.

Remark: Due to space constraint, our results on functions with a light Fourier tail as well as many proofs are omitted from this conference version, and can be found in the full version of the paper [TWXZ13].

II. Preliminaries and notation

All logarithms in this paper are base 2. For two n-bit vectors $s, t \in \{0,1\}^n$, define their inner product as $s \cdot t = \langle s, t\rangle = \sum_{i=1}^n s_i t_i \bmod 2$ and for simplicity we write $s + t$ for $s \oplus t$. 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 $\{+1,-1\}$, and we will specify whenever needed. For $f : \{0,1\}^n \to \{0,1\}$, we define $f^{\pm} = 1 - 2f$ to convert the range to $\{+1,-1\}$. For each $b \in \mathtt{range}(f)$, the b-density of f is $\rho_b = |f^{-1}(b)|/2^n$.

Each Boolean function $f : \{0,1\}^n \to \{0,1\}$ can be viewed as a polynomial over $\mathbb{F}_2$, and we use $\deg_2(f)$ to denote the $\mathbb{F}_2$-degree of f. For a Boolean function $f : \{0,1\}^n \to \{0,1\}$ and a direction vector $t \in \{0,1\}^n - \{0^n\}$, its derivative $\Delta_t f$ is defined by $\Delta_t f(x) = f(x) + f(x+t)$. It is easy to check that $\deg_2(\Delta_t f) < \deg_2(f)$ for any non-constant f and any t.

Complexity measures. A parity decision tree (PDT) for a function $f : \{0,1\}^n \to \{0,1\}$ is a tree with each internal node associated with a linear function $\ell(x)$, and each leaf associated with an answer $a \in \{0,1\}$. When we use a parity decision tree to compute a function f, we start from the root and follow a path down to a leaf. At each internal node, we query the associated linear function, and follow the branch according to the answer to the query. When reaching a leaf, we output the associated answer. The parity decision tree computes f if on any input x, we always get the output equal to $f(x)$. The deterministic parity decision tree complexity of f, denoted by $\mathsf{D}_{\oplus}(f)$, is the least number of queries needed on a worst-case input by a PDT that computes f.

For a Boolean function f and an input x, the parity certificate complexity of f on x is

$$\mathsf{C}_{\oplus}(f,x) = \min\{\mathtt{co\text{-}dim}(H) : x \in H, H \text{ is an affine subspace on which } f \text{ is constant}\}.$$

The parity certificate complexity $\mathsf{C}_{\oplus}(f)$ of f is $\max_x \mathsf{C}_{\oplus}(f,x)$. Since for each x and each parity decision tree T, the leaf that x belongs to corresponds to an affine subspace of co-dimension equal to the length of the path from it to the root, we have that $\mathsf{C}_{\oplus}(f) \le \mathsf{D}_{\oplus}(f)$ [ZS10]. We can also study the minimum parity certificate complexities $\mathsf{C}^b_{\oplus,\min}(f) = \min_{x:f(x)=b} \mathsf{C}_{\oplus}(f,x)$ and $\mathsf{C}_{\oplus,\min}(f) = \min_x \mathsf{C}_{\oplus}(f,x)$.

Denote by $\mathsf{D}^{\mathsf{CC}}(F)$ the deterministic communication complexity of F. One way of designing communication protocols is to simulate a decision tree algorithm, and the following is an adapted variant of a well known relation between deterministic communication complexity and decision tree complexity to the setting of XOR functions and parity decision trees.

Fact 12. $\mathsf{D}^{\mathsf{CC}}(f \circ \oplus) \le 2\mathsf{D}_{\oplus}(f)$.

Fourier analysis

For any real function $f : \{0,1\}^n \to \mathbb{R}$, the Fourier coefficients are defined by $\hat{f}(s) = 2^{-n}\sum_x f(x)\chi_s(x)$, where $\chi_s(x) = (-1)^{s \cdot x}$. The function f can be written as $f = \sum_s \hat{f}(s)\chi_s$. The ℓ_p-norm of $\hat{f}$ for any $p > 0$, denoted by $\|\hat{f}\|_p$, is defined as $(\sum_s |\hat{f}(s)|^p)^{1/p}$. The Fourier sparsity of f, denoted by $\|\hat{f}\|_0$, is the number of nonzero Fourier coefficients of f. As a simple consequence of Cauchy-Schwarz inequality we have $\|\hat{f}\|_1 \le \sqrt{\|\hat{f}\|_0}$. Note that $\|\hat{f}\|_1$ can be much smaller than $\|\hat{f}\|_0$. For instance, the AND function has $\|\hat{f}\|_1 \le 3$ but $\|\hat{f}\|_0 = 2^n$. The Fourier coefficients of $f : \{0,1\}^n \to \{0,1\}$ and $f^{\pm}$ are related by

$\widehat{f^{\pm}}(s) = \delta_{s,0^n} - 2\hat{f}(s)$, where $\delta_{x,y}$ is the Kronecker delta function. Therefore we have

$$2\|\hat{f}\|_1 - 1 \le \|\widehat{f^{\pm}}\|_1 \le 2\|\hat{f}\|_1 + 1,$$
$$\text{and} \quad \|\hat{f}\|_0 - 1 \le \|\widehat{f^{\pm}}\|_0 \le \|\hat{f}\|_0 + 1. \tag{2}$$

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

Proposition 13 (Convolution)**.** *For two functions $f, g : \{0,1\}^n \to \mathbb{R}$, the Fourier spectrum of fg is given by the following formula: $\widehat{fg}(s) = \sum_t \hat{f}(t)\hat{g}(s+t)$.*

Using this proposition, one can characterize the Fourier coefficients of Boolean functions as follows.

Proposition 14. *A function $f : \{0,1\}^n \to \mathbb{R}$ has range $\{+1,-1\}$ if and only if*

$$\sum_{t\in\{0,1\}^n} \hat{f}^2(t) = 1, \sum_{t\in\{0,1\}^n} \hat{f}(t)\hat{f}(s+t) = 0, \forall s \in \{0,1\}^n - 0^n.$$

Another fact easily following from the convolution formula is the following.

Lemma 15. *Let $f, g : \{0,1\}^n \to \mathbb{R}$, then $\|\widehat{fg}\|_0 \le \|\hat{f}\|_0\|\hat{g}\|_0$ and $\|\widehat{fg}\|_1 \le \|\hat{f}\|_1\|\hat{g}\|_1$.*

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 $Lf(x) = f(Lx)$. It is not hard to see that $\deg_2(Lf) = \deg_2(f)$, and that $\widehat{Lf}(s) = \hat{f}((L^T)^{-1}s)$. Thus

$$\|\widehat{Lf}\|_1 = \|\hat{f}\|_1 \text{ and } \|\widehat{Lf}\|_0 = \|\hat{f}\|_0. \tag{3}$$

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

$$\|\hat{f}_b\|_0 \le \|\hat{f}\|_0 \text{ and } \|\hat{f}_b\|_1 \le \|\hat{f}\|_1. \tag{4}$$

The concept of subfunctions can be generalized to general directions. Suppose $f : \{0,1\}^n \to \mathbb{R}$ 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 16. *Suppose $f : \{0,1\}^n \to \mathbb{R}$ and $H = a + V$ is an affine subspace, then one can define the spectrum $\widehat{f|_H}$ of the restricted function $f|_H$ such that*

1) *If $\texttt{co-dim}(H) = 1$, then $\widehat{f|_H}$ is the collection of $\hat{f}(s) + (-1)^b\hat{f}(s+t)$ for all unordered pair $(s, s+t)$, where t is the unique non-zero vector orthogonal to V, and $b = 0$ if $a \in V$ and $b = 1$ otherwise. Sometimes we refer to such restriction as a* folding *over t.*
2) *$\|\widehat{f|_H}\|_p \le \|\hat{f}\|_p$, for any $p \in [0,1]$.*
3) *If $\texttt{range}(f) = \{+1,-1\}$, then the following three statements are equivalent: 1) $f|_H(x) = c\chi_s(x)$ for some $s \in \{0,1\}^n$ and $c \in \{+1,-1\}$, 2) $\|\widehat{f|_H}\|_0 = 1$, and 3) $\|\widehat{f|_H}\|_1 = 1$.*

Using the above lemma, it is not hard to prove by induction the following fact, which says that short PDT gives Fourier sparsity.

Proposition 17. *$\forall f : \{0,1\}^n \to \{0,1\}$, $\|\hat{f}\|_0 \le 4^{\mathsf{D}_{\oplus}(f)}$.*

The following theorem [BC99] says that the $\mathbb{F}_2$-degree can be bounded from above by logarithm of Fourier sparsity.

Fact 18 ([BC99])**.** *For all $f : \{0,1\}^n \to \{0,1\}$, it holds that $\deg_2(f) \le \log\|\hat{f}\|_0$.*

III. Linear rank and the Main PDT algorithm

The notion of polynomial rank for a polynomial f has been extensively studied, and is usually defined as the minimum number of lower degree polynomials that f depends on. In particular, degree-2 polynomials are well understood [Dic58] and degree-3 and -4 polynomials are also recently studied [HS10].

Here we propose a new notion of linear rank, which requires a very specific way of composing lower degree polynomial into the original polynomial f.

Definition 2. *The* linear rank *of a polynomial $f \in \mathbb{F}_2[x_1, \ldots, x_n]$, denoted* $\text{lin-rank}(f)$*, is the minimum integer r s.t. f can be expressed as*

$$f = \ell_1 f_1 + \ldots + \ell_r f_r + f_0,$$

where $\deg_2(\ell_i) = 1$ for all $1 \le i \le r$ and $\deg_2(f_i) < \deg_2(f)$ for all $0 \le i \le r$. Sometimes we emphasize the degree by writing the polynomial rank as $\text{lin-rank}_d(f)$ with $d = \deg_2(f)$.

It is not hard to see that linear rank and polynomial rank are of the same order for polynomials with degree at most 3. But for higher degrees, linear rank can be much larger than polynomial rank.

Recall that a parity certificate is an affine subspace H restricted on which f is a constant. The parity certificate complexity is the largest co-dimension of such H. The next lemma says that the linear rank is quite small compared to the parity certificate complexity, even if we merely require f to have a lower $\mathbb{F}_2$-degree (rather than to be a constant) on the affine subspace; and in addition, even if we take the *minimum* co-dimension over all such H.

Lemma 19. *For all non-constant $f : \{0,1\}^n \to \{0,1\}$, the following properties hold.*

1) *There is a subspace V of co-dimension $r = \text{lin-rank}(f)$ s.t. when restricted to each of the 2^r affine subspaces $a+V$, f has $\mathbb{F}_2$-degree at most $\deg_2(f)-1$.*

2) *For all affine subspaces H with* $\texttt{co-dim}(H) < \text{lin-rank}(f)$*,* $\deg_2(f|_H) = \deg_2(f)$*.*

Lemma 19, though seemingly simple, is of fundamental importance to our problem as well as PDT algorithm designing in general. Note that the second part of Lemma 19 says that, if there exists an affine subspace $V+a$ of co-dimension k and a vector $a \in V^\perp$ such that $\deg_2(f|_{V+a}) < \deg_2(f)$, then $\text{lin-rank}(f) \le k$. Therefore Lemma 19 reduces the challenging task of lowering the degree of $f|_{V+a}$ for *all* a to lowering it for just *one* a.

In the next two sections, what we are going to use is the following corollary of it.

Corollary 20. *For all non-constant $f : \{0,1\}^n \to \{0,1\}$, we have* $\text{lin-rank}(f) \le \mathsf{C}_{\oplus,\min}(f)$*.*

A. Main PDT algorithm

Now we describe the main algorithm for computing function f, by reducing the $\mathbb{F}_2$-degree of f.

Main PDT Algorithm
Input: An PDT oracle for x
Output: $f(x)$.

1) **while** $\deg_2(f) \ge 1$ **do**
   a) Take a fixed decomposition $f = \ell_1 f_1 + \cdots + \ell_r f_r + f_0$, where $r = \text{lin-rank}_{\deg_2(f)}(f)$.
   b) for $i = 1$ to r
   c) Query $\ell_i(x)$ and get answer a_i.
   d) Update the function
   $$f := a_1 f_1 + \cdots + a_r f_r + f_0$$
2) Output the constant value of f.

To analyze the query complexity of this algorithm, we need to bound $\text{lin-rank}(f)$. We conjecture that the following is true for all Fourier sparse Boolean functions.

Conjecture 21. *For all Boolean functions $f : \{0,1\}^n \to \{0,1\}$,* $\text{lin-rank}(f) = O(\log^c(\|\hat{f}\|_0))$ *for some $c = O(1)$.*

Call a complexity measure $M(f)$ *downward non-increasing* if $M(f') \le M(f)$ for any f and any subfunction f' of f. As mentioned earlier (Lemma 16), $M(f) = \|\hat{f}\|_0$ and $M(f) = \|\hat{f}\|_1$ are all downward non-increasing complexity measures.

Theorem 22. *The Main PDT algorithm computes $f(x)$ correctly. If* $\text{lin-rank}(f) \le M(f)$ *for some downward non-increasing complexity measure M, then* $\mathsf{D}_\oplus(f) \le \deg_2(f) M(f)$ *and* $\mathsf{D}^{\mathsf{CC}}(f \circ \oplus) \le 2\log\|\hat{f}\|_0 \cdot M(f)$*. In particular, if Conjecture 21 is true, then the Log-rank conjecture holds for all XOR functions.*

The Main PDT algorithm, though simple, crucially uses the fact that restrictions do not increase the Fourier sparsity and uses the $\mathbb{F}_2$-degree as a progress measure to govern the efficiency. Since $\deg_2(f) \le \log(\|\hat{f}\|_0)$, the algorithm finishes in a small number of rounds.

This algorithm also gives a unified way to construct parity decision tree, reducing the task of designing PDT algorithms to showing that the linear rank is small. Indeed, the results in the next two sections are obtained by bounding rank, where sometimes Theorem 22 will be applied with the complexity measure $\|\hat{f}\|_1$.

Note that if the conjecture $\mathsf{D}_\oplus(f) \le \log^c \|\hat{f}\|_0$ is true, then the Main PDT algorithm always gives the optimal query cost up to a polynomial of power $c+1$.

IV. Functions with low $\mathbb{F}_2$-degree

In this section, we will show that the Log-rank conjecture holds for XOR functions with constant $\mathbb{F}_2$-degree. We will actually prove

$$\mathsf{C}_{\oplus,\min}(f) = O(2^{d^2/2} \log^{d-2} \|\hat{f}\|_1),$$

which is stronger than Lemma 4. Theorem 5 then follows from the PDT algorithm and the simulation protocol for PDT (Theorem 22). Corollary 6 is also easily proven by an induction on $\mathbb{F}_2$-degree.

The case for degree 1 (linear functions) is trivial and the case for degree 2 is also simple: applying Dickson's theorem gives $\text{lin-rank}(f) \le \mathsf{C}_{\oplus,\min}(f) = O(\log\|\hat{f}\|_1)$ for all quadratic polynomials f.

For cubic polynomials, one can show the following

Proposition 23. *For all function $f : \{0,1\}^n \to \{0,1\}$ with $\mathbb{F}_2$-degree 3, it holds that* $\text{lin-rank}(f) = O(\log\|\hat{f}\|_1)$ *and thus* $\mathsf{D}_\oplus(f) = O(\log\|\hat{f}\|_1)$*.*

A. Constant-degree polynomials

Now we will bound the $\text{lin-rank}(f)$ and use the Main PDT algorithm to bound the PDT complexity.

Lemma 24. *For all non-constant function $f : \{0,1\}^n \to \{0,1\}$ of $\mathbb{F}_2$-degree d, we have*

$$\text{lin-rank}(f) \le \mathsf{C}_{\oplus,\min}(f) \le \mathsf{D}_\oplus(f) \le O(2^{d^2/2}(\log^{d-2}\|\hat{f}\|_1 + 1)).$$

Proof: We will prove by induction on degree d that

$$\text{lin-rank}(f) \le \mathsf{C}_{\oplus,\min}(f) \le \max_{b\in\{0,1\}} \mathsf{C}^b_{\oplus,\min}(f) \le \mathsf{D}_\oplus(f) \le B_d\big(\|\widehat{f^{\pm}}\|_1\big),$$

where $B_d(m) < 2^{d^2/2}\log^{d-2} m$ are a class of bounded non-decreasing (with respect to both d and argument m) functions to be determined later. The conclusion then follows from Eq.(2).

We use Proposition 23 as the base case for induction. Now suppose that the bound holds for all polynomials of $\mathbb{F}_2$-degree at most $d-1$, and consider a function f of degree $d \ge 4$. We will first prove a bound for $\mathsf{C}_{\oplus,\min}(f)$, which also implies a bound on $\text{lin-rank}(f)$ by Corollary 20.

First, it is not hard to see that there exists a direction $t \in \{0,1\}^n - \{0^n\}$ such that $\Delta_t f$ is non-constant (unless

f is a linear function, in which case the conclusion trivially holds anyway). Fix such a t. Since $\deg_2(\Delta_t f) \le d-1$, by induction hypothesis, it holds that

$$\mathsf{C}^b_{\oplus,\min}(\Delta_t f) \le B_{d-1}(\|\widehat{(\Delta_t f)^\pm}\|_1).$$

Define $f_t(x) = f(x+t)$, then by Lemma 15, we have

$$\|\widehat{(\Delta_t f)^\pm}\|_1 = \|\widehat{f^\pm \cdot f_t^\pm}\|_1 \le \|\widehat{f^\pm}\|_1 \|\widehat{f_t^\pm}\|_1 = \|\widehat{f^\pm}\|_1^2,$$

which implies that

$$\mathsf{C}^b_{\oplus,\min}(\Delta_t f) \le B_{d-1}(\|\widehat{f^\pm}\|_1^2).$$

Since $\Delta_t f$ is non-constant, $C^b_{\oplus,\min}(\Delta_t f) \in [1,n]$ for *both* $b=0$ and $b=1$. For each b, by the definition of $C^b_{\oplus,\min}(\Delta_t f)$, there exists an affine subspace H_b with $\texttt{co-dim}(H_b) \le B_{d-1}(\|\widehat{f^\pm}\|_1^2)$ such that $(\Delta_t f)|_{H_b} = b$, which is equivalent to $f(x) + f(x+t) = b$ for all $x \in H_b$.

Define

$$g_0(x) = \frac{1}{2}\big(f^\pm(x) +_{\mathbb{R}} f^\pm(x+t)\big),$$

and

$$g_1(x) = \frac{1}{2}\big(f^\pm(x) -_{\mathbb{R}} f^\pm(x+t)\big),$$

where to avoid confusion, special notations plus $+_{\mathbb{R}}$ and minus $-_{\mathbb{R}}$ are used to stress that operations are over $\mathbb{R}$.

These two functions have some nice properties. First, it is easy to see from the definition of g_0 and g_1 that $f^\pm = g_0 +_{\mathbb{R}} g_1$. Second, note that g_0 and g_1 are not Boolean functions any more; they take values in $\{-1,0,+1\}$. However, a simple but crucial fact is that they take very special values on the affine subspace H_b: one always takes value 0, and the other always takes value in $\{+1,-1\}$. Actually, it is not hard to verify that

$$g_b|_{H_b} = f^\pm|_{H_b} \quad \text{and} \quad g_{1-b}|_{H_b} = 0.$$

Third, in the Fourier domain, note that

$$\widehat{f_t^\pm}(s) = \mathbf{E}_x[f^\pm(x+t)\chi_s(x+t)\chi_s(t)] = \widehat{f^\pm}(s)\chi_s(t),$$

and thus

$$\widehat{g_b}(s) = \frac{1}{2}\big(\widehat{f^\pm}(s) +_{\mathbb{R}} (-1)^b \chi_s(t)\widehat{f^\pm}(s)\big).$$

Therefore, we have

$$\widehat{g_0}(s) = \begin{cases} \widehat{f^\pm}(s) & s \in t^\perp \\ 0 & s \in \overline{t^\perp} \end{cases}, \quad \widehat{g_1}(s) = \begin{cases} 0 & s \in t^\perp \\ \widehat{f^\pm}(s) & s \in \overline{t^\perp} \end{cases},$$

where $t^\perp = \{s \in \{0,1\}^n : \langle s,t\rangle = 0\}$. Namely $\hat{g}_0$ and $\hat{g}_1$ each takes the Fourier spectrum $f^\pm$ on one of the two hyperplanes defined by the vector t.

This further implies that $\|\widehat{f^\pm}\|_1 = \|\widehat{g_0}\|_1 + \|\widehat{g_1}\|_1$. Thus, either $\|\widehat{g_0}\|_1$ or $\|\widehat{g_1}\|_1$ is at most half of $\|\widehat{f^\pm}\|_1$. Suppose that $\|\widehat{g_b}\|_1 \le \frac{1}{2}\|\widehat{f^\pm}\|_1$. We claim that restricting $f^\pm$ to H_b reduces its spectral norm a lot. Indeed, since $f^\pm|_{H_b} = g_b|_{H_b}$, we have

$$\|\widehat{f^\pm|_{H_b}}\|_1 = \|\widehat{g_b|_{H_b}}\|_1 \le \|\widehat{g_b}\|_1 \le \frac{1}{2}\|\hat{f}\|_1,$$

where the first inequality is because of Lemma 16. To summarize, we have just shown that we can reduce the spectral norm by at least half using at most $B_{d-1}(\|\widehat{f^\pm}\|_1^2)$ linear restrictions.

Now we recursively repeat the above process on the subfunction $f^\pm|_{H_b}$ until finally we find an affine subspace H *s.t.* $\|\widehat{f^\pm|_H}\| \le 1$, at which moment the subfunction is either a constant or linear function, thus at most one more folding would give a constant function. In total it takes at most $B_{d-1}(\|\widehat{f^\pm}\|_1^2)\log\|\widehat{f^\pm}\|_1 + 1$ linear restrictions to get a constant function, which implies that

$$\mathsf{C}_{\oplus,\min}(f) \le B_{d-1}(\|\widehat{f^\pm}\|_1^2)\log\|\widehat{f^\pm}\|_1 + 1.$$

Next we will show that actually the $\max_{b\in\{0,1\}} \mathsf{C}^b_{\oplus,\min}(f)$ is not much larger either:

$$\max_{b\in\{0,1\}} \mathsf{C}^b_{\oplus,\min}(f) \le B_{d-1}(\|\widehat{f^\pm}\|_1^2)\log\|\widehat{f^\pm}\|_1 + B_{d-1}\big(\|\widehat{f^\pm}\|_1\big) + 1. \tag{5}$$

(We need to show this because in the induction step, we picked one g_b with smaller spectral norm and used the induction hypothesis to upper bound $\mathsf{C}^b_{\oplus,\min}(\Delta_t f)$ for a particular b, which could be $\max_{b\in\{0,1\}} \mathsf{C}^b_{\oplus,\min}(\Delta_t f)$.) Note that by the Main PDT algorithm, we know that

$$\mathsf{D}_\oplus(f) \le \text{lin-rank}(f) + \mathsf{D}_\oplus(f'),$$

for a subfunction f' of f with $\deg_2(f') < \deg_2(f)$. Now by Corollary 20, we can use $\mathsf{C}_{\oplus,\min}(f)$ to upper bound $\text{lin-rank}(f)$. For the second part, since $\deg_2(f') < \deg_2(f)$ and $\|\widehat{(f')^\pm}\|_1 \le \|\widehat{f^\pm}\|_1$, applying the induction hypothesis on f' yields the following upper bound on $\mathsf{D}_\oplus(f')$:

$$\mathsf{D}_\oplus(f) \le B_{d-1}\big(\|\widehat{f^\pm}\|_1^2\big)\log\|\widehat{f^\pm}\|_1 + 1 + B_{d-1}\big(\|\widehat{f^\pm}\|_1\big). \tag{6}$$

Eq.(5) thus follows from the simple bound $\mathsf{C}^b_{\oplus,\min}(f) \le \mathsf{C}_\oplus(f) \le \mathsf{D}_\oplus(f)$. Now define the right-hand side of Eq.(6) to be $B_d\big(\|\widehat{f^\pm}\|_1\big)$, and solve the following recursive relation $B_d(m) = B_{d-1}(m^2)\log m + B_{d-1}(m) + 1$, and $B_3(m) = O(\log m + 1)$, we get

$$B_d(m) = (1+o(1))2^{(d-2)(d-3)/2}\log^{d-2} m,$$

as desired. ■

Note that in the above proof, it seems that we lose something by using $\mathsf{C}_{\oplus,\min}$ to upper bound lin-rank. However, it is crucial to consider the affine subspace H_b on which $\Delta_t f$ becomes a *constant* (instead of, say, a polynomial of lower $\mathbb{F}_2$-degree), because otherwise g_b on H_b is not equal to f (actually not even Boolean), and thus we cannot recursively apply the procedure on $f|_{H_b}$. In addition, if $\Delta_t f$ is not constant on H_b, then we cannot guarantee the decrease of the spectral norm due to restriction on H_b.

We have showed that low degree polynomials have small $\mathsf{C}_{\oplus,\min}$ value in terms of the spectral norm. We actually conjecture that the bound can be improved to the following.

Conjecture 25. *There is some absolute constant c s.t. for any non-constant $f : \{0,1\}^n \to \{0,1\}$, $\mathsf{C}_{\oplus,\min}(f) = O(\log^c \|\hat{f}\|_1)$.*

It has the consequence as follows.

Proposition 26. *If Conjecture 25 is true, then for any $f : \{0,1\}^n \to \{0,1\}$, $\text{lin-rank}(f) = O(\log^c \|\hat{f}\|_1)$ and $\mathsf{D}_{\oplus}(f) = O(\deg_2(f)\log^c \|\hat{f}\|_1)$.*

In fact, we are not aware of any counterexample for Conjecture 25 even for $c = 1$.

Lemma 24 also implies the following Corollary, from which Corollary 7 immediately follows.

Corollary 27. *If $f : \{0,1\}^n \to \{0,1\}$ has $\mathbb{F}_2$-degree d, then $f = \sum_{i=1}^{T} \pm \mathbb{1}_{V_i}$, where $T = 2^{2^{d^2/2}\log^{d-2}\|\hat{f}\|_1}$ and each $\mathbb{1}_{V_i}$ is the indicator function of the subspace V_i.*

V. Functions with a small spectral norm

We prove Lemma 28 in this section, which directly implies Lemma 8, Theorem 9 and Theorem 10.

Lemma 28. *For all Boolean function $f : \{0,1\}^n \to \{+1,-1\}$, we have $\mathsf{C}_{\oplus,\min}(f) \le O(\|\hat{f}\|_1)$.*

Proof: Suppose that the nonzero Fourier coefficients are $\{\hat{f}(\alpha) : \alpha \in A\}$, where $A = \mathtt{supp}(\hat{f})$. Denote by $a_1, a_2, ..., a_s$ the sequence of $|\hat{f}(\alpha)|$ in the decreasing order, and the corresponding characters are $\chi_{\alpha_1}, ..., \chi_{\alpha_s}$ in that order (thus $|\hat{f}(\alpha_i)| = a_i$ and $s = \|\hat{f}\|_0$ is the Fourier sparsity of f). For simplicity, we assume $s \ge 4$, as doing so can only add at most a constant to our bound on $\mathsf{C}_{\oplus,\min}(f)$.

Consider the following *greedy folding* process: fold along $\beta = \alpha_1 + \alpha_2$ and select a proper half-space, namely impose a linear restriction $\chi_\beta(x) = b$ for some $b \in \{0,1\}$, *s.t.* the subfunction has its largest Fourier coefficient being $a_1 + a_2$ (in absolute value). This can be done by Lemma 16.

The lemma now follows from the following two claims, whose proofs are omitted in this version.

Claim 29. *After at most $O(\|\hat{f}\|_1)$ greedy foldings, we have $a_1 \ge 1/2$.*

Claim 30. *Each greedy folding increases a_1 and decreases the Fourier ℓ_1-norm by at least $2a_1 = 2\max_s |\hat{f}(s)|$.*

Thus once $a_1 \ge 1/2$, then each greedy folding decreases the Fourier ℓ_1-norm by at least 1. So it takes at most $\|\hat{f}\|_1$ further steps to make the Fourier ℓ_1-norm to be at most 1, in which case at most one more folding makes the function constant. ∎

Lemma 28 implies that $\text{lin-rank}(f) \le O(\|\hat{f}\|_1)$ (Lemma 8) by Corollary 20 (that $\text{lin-rank}(f) \le \mathsf{C}_{\oplus,\min}(f)$).

Note that our Main PDT algorithm can be simply simulated by a protocol in which Alice and Bob send $\ell_i(x)$ and $\ell_i(y)$, respectively. Thus, similar to Fact 12, we have $\mathsf{D}^{\mathsf{CC}}(f) \le 2\mathsf{D}_{\oplus}(f)$ for $f : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}$. Theorem 9 basically follows from this lemma and the fact that subfunctions have smaller spectral norm (Lemma 16).

Lemma 28 also implies Theorem 10, which asserts upper bounds on the deterministic communication complexity of $f \circ \oplus$ as $\mathsf{D}^{\mathsf{CC}}(f \circ \oplus) = O(\deg_2(f) \cdot \|\hat{f}\|_1) = O\left(\sqrt{\mathtt{rank}(M_{f\circ\oplus})}\log \mathtt{rank}(M_{f\circ\oplus})\right)$. To see this, first recall Theorem 22, which states that $\mathsf{D}^{\mathsf{CC}}(f\circ\oplus) \le 2\log\|\hat{f}\|_0 \cdot M(f)$ where $M(f)$ is a downward non-increasing complexity measure. By Lemma 8, we can take M to be $\|\hat{f}\|_1$. Now combining these with Fact 18 (that $\deg_2(f) \le \log\|\hat{f}\|_0$), and the inequality that $\|\hat{f}\|_1 \le \sqrt{\|\hat{f}\|_0}$ yields Theorem 10.

Acknowledgments

We are indebted to Andrej Bogdanov for an enlightening discussion at the early stage of the work, and we would also like to thank Irit Dinur, Elena Grigorescu, Ronitt Rubinfeld, Rocco Servedio, Amir Shpilka, Xiaorui Sun and Ben Lee Volk for valuable comments.

References

[AKK+05] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing Reed Muller codes. *IEEE Transactions on Information Theory*, 51(11):4032–4039, 2005.

[ASTS+03] Andris Ambainis, Leonard Schulman, Amnon Ta-Shma, Umesh Vazirani, and Avi Wigderson. The quantum communication complexity of sampling. *SIAM Journal on Computing*, 32(6):1570–1585, 2003.

[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.

[BdW02] Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288(1):21–43, 2002.

[Bec75] William Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, 102:159–182, 1975.

[Bon70] Aline Bonami. Étude des coefficients Fourier des fonctiones de $L^p(G)$. *Annales de l'institut Fourier*, 20(2):335–402, 1970.

[BSLRZ12] Eli Ben-Sasson, Shachar Lovett, and Noga Ron-Zewi. An additive combinatorics approach relating rank to communication complexity. In *Proceedings of The 53rd Annual IEEE Symposium on Foundations of Computer Science*, pages 177–186, 2012.

[Cha02] Mei-Chu Chang. A polynomial bound in Freiman's theorem. *Duke Mathematical Journal*, 113(3):399–419, 2002.

[Dic58] L. Dickson. *Linear groups with an exposition of the Galois field theory*. Dover, New York, 1958.

[GOS+11] Parikshit Gopalan, Ryan O'Donnell, Rocco Servedio, Amir Shpilka, and Karl Wimme. Testing Fourier dimensionality and sparsity. *SIAM Journal on Computing*, 40(4):1075–1100, 2011.

[Gow98] Timothy Gowers. A new proof of Szemerédi's theorem for arithmetic progressions of length four. *Geometric and Functional Analysis*, 8(3):529–551, 1998.

[Gow01] Timothy Gowers. A new proof of Szemerédi's theorem. *Geometric and Functional Analysis*, 11(3):465–588, 2001.

[Gro97] Vince Grolmusz. On the power of circuits with gates of low L1 norms. *Theoretical Computer Science*, 188(1-2):117–128, 1997.

[GS08] Ben Green and Tom Sanders. Boolean functions with small spectral norm. *Geometric and Functional Analysis*, 18(1):144–162, 2008.

[GT09] Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the Gowers norms. *Contributions to Discrete Mathematics*, 4(2):1–36, 2009.

[HS10] Elad Haramaty and Amir Shpilka. On the structure of cubic and quartic polynomials. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, pages 331–340, 2010.

[IMR12] Russell Impagliazzo, Cristopher Moore, and Alexander Russell. An entropic proof of Chang's inequality. *arXiv:1205.0263*, 2012.

[KL96] Andrew Kotlov and László Lovász. The rank and size of graphs. *Journal of Graph Theory*, 23:185–189, 1996.

[KL08] Tali Kaufman and Shachar Lovett. Worst case to average case reductions for polynomials. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 166–175, 2008.

[KM93] Eyal Kushilevitz and Yishay Mansour. Learning decision trees using the Fourier spectrum. *SIAM Journal on Computing*, 22(6):1331–1348, 1993.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, Cambridge, UK, 1997.

[Kot97] Andrei Kotlov. Rank and chromatic number of a graph. *Journal of Graph Theory*, 26:1–8, 1997.

[KS13] Raghav Kulkarni and Miklos Santha. Query complexity of matroids. In *Proceedings of the 8th International Conference on Algorithms and Complexity*, 2013.

[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.

[Lov90] László Lovász. Communication complexity: A survey. In *B. Korte, L. Lovász, H. Prömel, and A. Schrijver, editors, Paths, flows, and VLSI-layout, Springer-Verlag*, page 235–265, 1990.

[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.

[LS09] Troy Lee and Adi Shraibman. Lower bounds on communication complexity. *Foundations and Trends in Theoretical Computer Science*, 3(4):263–398, 2009.

[LZ10] Troy Lee and Shengyu Zhang. Composition theorems in communication complexity. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming (ICALP)*, 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.

[Mid04] Gatis Midrijanis. Exact quantum query complexity for total Boolean functions, 2004. http://arxiv.org/abs/quant-ph/0403168.

[MO10] Ashley Montanaro and Tobias Osborne. On the communication complexity of XOR functions. *arXiv:*, 0909.3392v2, 2010.

[MS82] Kurt Mehlhorn and Erik M. Schmidt. Las Vegas is better than determinism in VLSI and distributed computing (extended abstract). In *Proceedings of the fourteenth annual ACM symposium on Theory of computing*, pages 330–337, 1982.

[NW95] Noam Nisan and Avi Wigderson. On rank vs. communication complexity. *Combinatorica*, 15(4):557–565, 1995.

[O'D12] Ryan O'Donnell. Lecture notes: Analysis of Boolean functions. 15-859S, Carnegie Mellon University, 2012.

[SV13] Amir Shpilka and Ben Lee Volk. On the structure of Boolean functions with small spectral norm. *ECCC*, TR13-049, 2013.

[SW12] Xiaoming Sun and Chengu Wang. Randomized communication complexity for linear algebra problems over finite fields. In *Proceedings of the 29th International Symposium on Theoretical Aspects of Computer Science*, pages 477–488, 2012.

[TWXZ13] Hing Yin Tsang, Chung Hoi Wong, Ning Xie, and Shengyu Zhang. Fourier sparsity, spectral norm, and the Log-rank conjecture, 2013. http://arxiv.org/abs/1304.1245.

[Val04] Paul Valiant. The log-rank conjecture and low degree polynomials. *Information Processing Letters*, 89(2):99–103, 2004.

[Yao79] Andrew Yao. Some complexity questions related to distributive computing. In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing (STOC)*, pages 209–213, 1979.

[ZS09] Zhiqiang Zhang and Yaoyun Shi. Communication complexities of symmetric XOR functions. *Quantum Information & Computation*, 9(3):255–263, 2009.

[ZS10] Zhiqiang Zhang and Yaoyun Shi. On the parity complexity measures of Boolean functions. *Theoretical Computer Science*, 411(26-28):2612–2618, 2010.