THEORY OF COMPUTING, Volume 15 (11), 2019, pp. 1–13
www.theoryofcomputing.org

Fourier Sparsity and Dimension

Swagato Sanyal

Received December 29, 2018; Revised August 19, 2019; Published October 27, 2019

Abstract: We prove that the Fourier dimension of any Boolean function with Fourier sparsity s is at most $O(\sqrt{s}\log s)$. This bound is tight up to a factor of $O(\log s)$ since the Fourier dimension and sparsity of the address function are quadratically related. We obtain our result by bounding the non-adaptive parity decision tree complexity, which is known to be equivalent to the Fourier dimension. A consequence of our result is that any XOR function has a protocol of complexity $O(\sqrt{r}\log r)$ in the simultaneous communication model, where r is the rank of its communication matrix.

ACM Classification: F.2.3

AMS Classification: 68Q17

Key words and phrases: Boolean functions, Fourier analysis, Boolean complexity, decision tree, query complexity

1 Introduction

The study of Boolean functions involves studying various complexity measures and their inter-relationships. Two such measures, which we investigate in this article, are the Fourier dimension and the Fourier sparsity. Let $f : \mathbb{F}_2^n \to \{1,-1\}$ be a Boolean function with Fourier expansion

$$f(x) = \sum_{\gamma \in \mathbb{F}_2^n} \widehat{f}(\gamma)\chi_\gamma(x).$$

where $\chi_\gamma(x) := (-1)^{\sum_{i=1}^n \gamma_i x_i}$ are the characters and the $\widehat{f}(\gamma)$'s are real numbers, called the Fourier coefficients of f. The Fourier dimension and Fourier sparsity are defined as follows.

A conference version of this paper appeared in the Proceedings of the 42nd International Colloquium on Automata, Languages, and Programming (ICALP 2015) [12].

DOI: 10.4086/toc.2019.v015a011

Definition 1.1 (Fourier sparsity and dimension)**.** For a Boolean function $f : \mathbb{F}_2^n \to \{1,-1\}$ with Fourier expansion

$$f(x) = \sum_{\gamma \in \mathbb{F}_2^n} \widehat{f}(\gamma)\chi_\gamma(x),$$

the *Fourier support* of f, denoted by $\mathrm{supp}(\widehat{f})$, is defined as

$$\mathrm{supp}(\widehat{f}) := \{\gamma \in \mathbb{F}_2^n : \widehat{f}(\gamma) \neq 0\}.$$

The *Fourier sparsity* of f, denoted by $\mathrm{sparsity}(\widehat{f})$, is defined as the size of the Fourier support of f, i. e.,

$$\mathrm{sparsity}(\widehat{f}) := |\,\mathrm{supp}(\widehat{f})|,$$

while the *Fourier dimension* $\dim(\widehat{f})$ of f is defined as the dimension of the span of $\mathrm{supp}(\widehat{f})$.

Note that the Fourier expansion of f is a multilinear polynomial in the variables $y_i := (-1)^{x_i}$. With respect to this view, Fourier sparsity of a Boolean function f is the number of monomials that appear in the Fourier expansion of f. Fourier dimension, on the other hand, is the smallest number of monomials, or equivalently parity functions, whose values always determine f. It is natural to investigate the power and limitation of polynomials with low Fourier sparsity or Fourier dimension. Fourier sparsity and Fourier dimension were studied by Gopalan *et al.* [4] in the context of property testing. More recently these quantities have been studied in the context of learning [7, 1]. An approximate analog of Fourier dimension has been shown to characterize the randomized one-way communication complexity of an important and well-studied subclass of functions called the XOR functions, over uniformly distributed inputs [8]. A Boolean function $F(x,y)$ on two n-bit inputs is an XOR function if there exists a Boolean function f on n bits such that $F(x,y) = f(x \oplus y)$.

Besides, the study on Fourier sparsity has attracted attention of complexity theorists due to its intimate connection to the *log-rank conjecture* of communication complexity. Fourier dimension is related to a simultaneous communication game. This connection is discussed in more detail later in this section.

The following inequalities easily follow from the definition of Fourier sparsity and dimension.

$$\log_2 \mathrm{sparsity}(\widehat{f}) \leq \dim(\widehat{f}) \leq \mathrm{sparsity}(\widehat{f}). \tag{1.1}$$

There are functions (e.g., indicator functions of subspaces) for which the first inequality is tight (i. e., holds with equality). In this note we examine the tightness of the second inequality. Note that the Fourier transform of a Boolean function is a polynomial with a very special property; it evaluates to 1 or -1 on all inputs from $\{1,-1\}^n$. It is thus natural to expect that there is always a good amount of dependency amongst its monomials. The gap between Fourier sparsity and Fourier dimension is one way of quantifying this dependency.

For the second inequality, the *address function* is one function having asymptotically the closest gap between Fourier dimension and sparsity.[1] Let s be a power of 2. The address function $\mathrm{Add}_s : \{0,1\}^{(1/2)\log s + \sqrt{s}} \to \{1,-1\}$ is defined as

$$\mathrm{Add}_s(x, y_1, y_2, \ldots, y_{\sqrt{s}}) := (-1)^{y_x}, \quad x \in \{0,1\}^{\frac{1}{2}\log s}, y_i \in \{0,1\},$$

[1] Recently Khalyavin, Lobanov and Tarannikov have constructed a function for which the gap between Fourier dimension and Fourier sparsity is closer by a constant factor than in the address function [9].

where x is interpreted as an address in $\{1,\ldots,\sqrt{s}\}$. In other words, on every input (x,y), $\mathrm{Add}_s(x,y)$ is the value of the addressed input bit y_x indexed by the addressing variables x. The address function has Fourier sparsity s and Fourier dimension at least $\sqrt{s}$. To see this, note that

$$\mathrm{Add}_s(x,y) = \sum_{\tilde{x}\in\{0,1\}^{(1/2)\log s}} \mathbf{1}_{x=\tilde{x}} \cdot (-1)^{y_{\tilde{x}}}.$$

Each indicator function $\mathbf{1}_{x=\tilde{x}}$ has Fourier sparsity[2] $\sqrt{s}$. Since the summation is over $\sqrt{s}$ terms, and since there is no cancellation due to the presence of a character corresponding to a fresh variable $y_{\tilde{x}}$ in each term, the Fourier sparsity of the address function is equal to $\sqrt{s}\cdot\sqrt{s} = s$. To see that the dimension is at least $\sqrt{s}$, note that for each $\tilde{x} \in \{0,1\}^{(1/2)\log s}$, the character $(-1)^{y_{\tilde{x}}}$ appears in the Fourier transform. We prove that for any Boolean function, this is the asymptotically highest possible value of $\dim(\widehat{f})$ in terms of $\mathrm{sparsity}(\widehat{f})$, up to a factor of $O(\log s)$. This was presented as an open problem at the Simons workshop on Real Analysis in Testing, Learning and Inapproximability, 2013 by John Wright.

Our main result is the following.

Theorem 1.2. *Let f be a Boolean function with* $\mathrm{sparsity}(\widehat{f}) = s$. *Then,* $\dim(\widehat{f}) = O(\sqrt{s}\log s)$.

In a preliminary version of this article, posted on arXiv and ECCC, we proved an upper bound of $O(s^{2/3})$ on $\dim(\widehat{f})$. Avishay Tal observed that the analysis can be tightened to obtain the near-optimal upper bound of $O(\sqrt{s}\log s)$.

Prior to this work, Gavinsky *et al.* [3] had proved that the dimension of any Boolean function with Fourier sparsity s is $O(s/\log s)$.

Theorem 1.2 is proved using a lemma of Tsang *et al.* [14] bounding the codimension of an affine subspace restricted to which the function is constant, in terms of the Fourier sparsity of the function. The following result is a corollary to [14, Lemma 28].

Lemma 1.3 (Tsang, Wong, Xie, and Zhang)**.** *Let $f : \mathbb{F}_2^n \to \{1,-1\}$ be a Boolean function with Fourier sparsity s. Then there is an affine subspace V of $\mathbb{F}_2^n$ of codimension $O(\sqrt{s})$ such that f is constant on V.*

Proof idea of Theorem 1.2. We begin with a simple but crucial observation made by Gopalan *et al.* [4], that the Fourier dimension of a Boolean function is equivalent to its non-adaptive parity decision tree complexity (see Proposition 2.7). This offers us a potential approach towards an upper bound on the Fourier dimension of a Boolean function: exhibiting a shallow non-adaptive parity decision tree of the function. We recall that the character functions essentially compute the parities of various subsets of the input bits. Thus a parity decision tree can be thought of as querying various character functions. Parity functions, in turn, are linear functions from $\mathbb{F}_2^n$ to $\mathbb{F}_2$; thus affine subspaces can be described by specifying a set of parities that are set to various values.

Towards this end, we first recall the construction of an (adaptive) parity decision tree for a Boolean function f of Fourier sparsity s by Tsang *et al.* [14], which in turn improves on an earlier construction due to Shpilka *et al.* [13, Theorem 1.1]. The broad idea of their construction is as follows. At any point in time, a partial tree is maintained whose leaves are functions which are restrictions of f on different

[2] Note that the indicator functions evaluate to $\{0,1\}$ instead of $\{1,-1\}$. However, the theory of Fourier analysis extends to such functions as well.

affine subspaces. Then a non-constant leaf τ is picked arbitrarily, and a small set of linear restrictions is obtained by invoking Lemma 1.3, such that the restricted function $f\,|_\tau$ at that leaf becomes constant. The next step is observing that if $f\,|_\tau$ is further restricted to all the affine subspaces obtained by setting the same set of parities in all possible ways, then the Fourier sparsity of each of the corresponding restricted functions is at most half of that of $f\,|_\tau$. This is because, in the former restriction, since the function becomes constant, the Fourier coefficients corresponding to non-constant characters must disappear in the restricted space. This can only happen if every non-constant character gets identified with at least one other character. This identification leads to halving of the Fourier support. Note that by Lemma 1.3 the number of queries we have spent to achieve this reduction in Fourier sparsity is

$$O\left(\sqrt{\text{sparsity}(\widehat{f\,|_\tau})}\right).$$

A calculation gives us that proceeding in this way a parity decision tree of depth $O(\sqrt{s})$ is obtained.

Note that the choice of parities restricted at various steps depends on the leaf (function) chosen, and hence on the outcomes of the preceding queries. Thus the constructed tree is an adaptive one. In this article, we make this tree non-adaptive, at the cost of a logarithmic increase in depth. At each step, we choose an appropriate function (leaf) of the partial tree constructed thus far, invoke Lemma 1.3, and obtain restrictions which make it constant. Then we query the same set of parities at every leaf. Note that doing that in each step results in a non-adaptive tree, as the set of parities queried at each step is the same for all the leaves of the current partial tree, and hence independent of the responses to the previous queries. Then we argue that this leads to a significant reduction of Fourier sparsity. Let $s^{(i)}$ be the Fourier sparsity of the function (leaf) chosen at the i-th step. It can be shown that, in the next step, the size $\ell^{(i)}$ of the union of the Fourier supports of all the leaves falls at least by $s^{(i)}/2$. From Lemma 1.3, the number of queries spent in the i-th step is $O(\sqrt{s^{(i)}})$. Using the *Uncertainty Principle* (Theorem 2.4) we show that $s^{(i)} \geq \left(\ell^{(i)}\right)^2/s$ (see Lemma 3.3). We combine all these facts to show that continuing in this fashion, in a small number of steps and making at most $O(\sqrt{s}\log s)$ queries, all the leaves become constant functions. The details of the construction of the non-adaptive parity decision tree, and its analysis, are given in Section 3.

Connections to communication complexity and the log-rank conjecture. The log-rank conjecture is a long-standing and important conjecture in communication complexity. The statement of the conjecture is that the deterministic communication complexity of a Boolean function is asymptotically bounded above by some fixed polylogarithm of the rank (over the real numbers) of its communication matrix. The best known upper bound of deterministic communication complexity of a function in terms of the rank is $O(\sqrt{\text{rank}}\log \text{rank})$ due to Lovett [10].[3] The rank of the communication matrix of an XOR function $F(x,y) := f(x \oplus y)$ is known to be equal to the Fourier sparsity s of f. For the special case of XOR functions the result by Lovett also follows from a result of of Tsang *et al.* [14],[4] which improves on a result by Shpilka *et al.* [13]. A consequence of Theorem 1.2 is that XOR functions admit a protocol

[3]We note here that the related *log approximate-rank conjecture* has recently been refuted by Chattopadhyay *et al.* [2], and the function they use is an XOR function.

[4]For XOR functions, the communication complexity upper bound is in fact $O(\sqrt{\text{rank}})$ which is better than $O(\sqrt{\text{rank}}\log\text{rank})$ by a logarithmic factor. It follows from the parity decision tree complexity upper bound proved by Tsang *et al.* [14].

of complexity $O(\sqrt{\mathrm{rank}}\log\mathrm{rank}) = O(\sqrt{s}\log s)$ in the simultaneous communication model.[5] We note that both the earlier protocols [10, 14] are two-way. The simultaneous protocol is as follows. Alice and Bob a priori agree on a set $\mathcal{S}$ of $O(\sqrt{s}\log s)$ parities that span the Fourier support of $F(x,y) = f(x\oplus y)$. The existence of $\mathcal{S}$ is guaranteed by Theorem 1.2. The character corresponding to each parity P in $\mathcal{S}$ is a product of a character M_x in variables in x and a character M_y in variables in y. Upon receiving inputs x and y, Alice and Bob compute the values of M_x and M_y, respectively, for each $P\in\mathcal{S}$, and send the evaluations to the referee. The referee then can evaluate the characters corresponding to each parity $P\in\mathcal{S}$ on the input $x\oplus y$. Since every other character in the Fourier expansion of F is determined by the characters in $\mathcal{S}$, the referee can compute the value of all those characters. Finally, the referee evaluates $F(x,y)$ from its Fourier expansion and outputs it.

Some remarks about Lemma 1.3. Lemma 1.3 is not believed to be tight. Tsang *et al.* [14] investigated this question while studying the log-rank conjecture for XOR functions. They suggested a direction towards proving the log-rank conjecture for XOR functions. In particular, they proposed a protocol for an XOR function $F(x,y) = f(x\oplus y)$ based on a parity decision tree of f and showed that the communication complexity of the proposed protocol is polylogarithmic in the rank of the communication matrix if the following related conjecture, stated as [14, Conjecture 27], is true.

Conjecture 1.4 (Tsang *et al.*). *There exists a constant $c > 0$ such that for every Boolean function f with Fourier sparsity s, there exists an affine subspace of codimension $O(\log^c s)$ on which f is constant.*

Even before, Montanaro and Osborne [11] had conjectured that the parity decision tree complexity of a function is polylogarithmic in its Fourier sparsity. Tsang *et al.* showed that this seemingly stronger conjecture follows from Conjecture 1.4.

Tsang *et al.* proved the above conjecture for certain classes of functions, which include functions with constant $\mathbb{F}_2$ degree, and proved Lemma 1.3 for general functions. It follows from subsequent result by Hatami, Hosseini and Lovett [6] that Conjecture 1.4 is equivalent to the log-rank conjecture for XOR functions (and also to the conjecture by Montanaro and Osborne).

We remark that the bound of Theorem 1.2 is near-optimal; so any significant improvement to it assuming Conjecture 1.4 would be a contradiction, and hence would serve to refute Conjecture 1.4. We note that with our proof technique and analysis, any improvement to Lemma 1.3 (in particular a positive resolution of Conjecture 1.4), does not yield a better than logarithmic improvement to Theorem 1.2. Our result, thus, does not seem to throw any light on the truth of Conjecture 1.4. For further discussion on this topic, the reader is referred to Section 3.

2 Preliminaries

Let $f:\mathbb{F}_2^n \to \{1,-1\}$ be a Boolean function. For $\gamma = (\gamma_1,\ldots,\gamma_n), x = (x_1,\ldots,x_n) \in \mathbb{F}_2^n$, let $\chi_\gamma(x) := (-1)^{\sum_{i=1}^n \gamma_i x_i}$ be the *character* associated with γ. It is well known that every Boolean function $f(x)$ can be

[5] We thank the anonymous referee for pointing this out.

uniquely written as

$$f(x) = \sum_{\gamma \in \mathbb{F}_2^n} \widehat{f}(\gamma)\chi_\gamma(x). \tag{2.1}$$

The right-hand side of Equation (2.1) is called the *Fourier expansion* of f, and the real coefficients $\widehat{f}(\gamma)$ are called the *Fourier coefficients*.

We review some standard definitions and facts about the Fourier coefficients.

Definition 2.1. Let $f(x) = \sum_{\gamma \in \mathbb{F}_2^n} \widehat{f}(\gamma)\chi_\gamma(x)$ be a Boolean function, and $p \geq 1$. The p-th spectral norm $\widehat{\|f\|}_p$ of f is defined as

$$\widehat{\|f\|}_p = \left(\sum_{\gamma \in \mathbb{F}_2^n} \left| \widehat{f}(\gamma) \right|^p \right)^{1/p}.$$

Lemma 2.2 (A special case of Parseval's identity)**.** *For a Boolean function f, $\widehat{\|f\|}_2 = 1$.*

The 1st spectral norm of a Boolean function can be bounded in terms of sparsity as follows.

Claim 2.3. For a Boolean function f with Fourier sparsity s, $\widehat{\|f\|}_1 \leq \sqrt{s}$.

Proof.

$$\widehat{\|f\|}_1 \leq \widehat{\|f\|}_2 \cdot \sqrt{s} = \sqrt{s}.$$

The first inequality is the Cauchy-Schwarz inequality while the second equality follows from Lemma 2.2. □

In order to prove our results, we shall use the following version of the Uncertainty Principle. See, e.g., [5] for a proof.

Theorem 2.4 (Uncertainty Principle)**.** *Let $p : \mathbb{R}^n \to \mathbb{R}$ be a real multilinear non-zero n-variate polynomial with sparsity s (i. e., it has s monomials with non-zero coefficients). Let U_n denote the uniform distribution on $\{1, -1\}^n$. Then*

$$\Pr_{x \sim U_n}[p(x) \neq 0] \geq \frac{1}{s}.$$

As mentioned in the Introduction, we need the following result by Tsang *et al.* [14, Lemma 28].

Theorem 2.5 (Tsang *et al.*)**.** *Let $f : \mathbb{F}_2^n \to \{1, -1\}$ be such that $\widehat{\|f\|}_1 = A$. Then there is an affine subspace V of $\mathbb{F}_2^n$ of codimension $O(A)$ such that f is constant on V.*

Lemma 1.3 is a simple corollary of this theorem via Claim 2.3.

In our proof we crucially use the observation made by Gopalan *et al.* [4] about the equivalence of the non-adaptive parity decision tree complexity of a function (defined below) and its Fourier dimension. We state it in Proposition 2.7.

Definition 2.6 (Non-adaptive parity decision tree complexity)**.** Let f be a Boolean function. The *non-adaptive parity decision tree complexity* of f (denoted by $\mathrm{NADT}_\oplus(f)$) is defined as the minimum integer t such that there exist $\gamma_1, \ldots, \gamma_t \in \mathbb{F}_2^n$ such that f is determined by the evaluation of the characters $\chi_{\gamma_1}, \ldots, \chi_{\gamma_t}$.

Proposition 2.7 (Gopalan *et al.* [4])**.** *For every Boolean function f, $NADT_\oplus(f) = \dim(\widehat{f})$.*

3 Upper bound on parity decision tree complexity

In this section, we prove an upper bound on the non-adaptive parity decision tree complexity of a Boolean function f in terms of its Fourier sparsity s. For $B \subseteq \mathbb{F}_2^n$, let $f|_B$ denote the restriction of the function f to the set B. We will often identify vectors $\gamma \in \mathbb{F}_2^n$ with the linear function that maps a vector $x \in \mathbb{F}_2^n$ to $\gamma(x) := \sum_{i=1}^n \gamma_i x_i \bmod 2 \in \mathbb{F}_2$. We first formally present a procedure NADT that constructs a non-adaptive parity decision tree of f. We then provide a description of the procedure in words (in particular, the roles of the various variables used) and a formal analysis of the same, leading up to the proof of Theorem 1.2.

NADT(f)

Input: Boolean function $f : \mathbb{F}_2^n \to \{1, -1\}$;

Output: A set Γ of parities, whose evaluations determine the value of f;

1. Set $\Gamma \leftarrow \varnothing$, $\mathcal{S} \leftarrow \operatorname{supp}(\widehat{f})$ and $\mathcal{F} \leftarrow \{f\}$.
2. While $\mathcal{S} \neq \{\emptyset\}$, do
   (a) Let g be a function in $\mathcal{F}$ with the largest Fourier sparsity. Let A be a largest affine subspace on which g is constant (breaking ties arbitrarily), with codimension n_g. Let $\gamma_1, \dots, \gamma_{n_g} \in \mathbb{F}_2^n$ and $b_1, \dots, b_{n_g} \in \mathbb{F}_2$ be such that
   $$A = \{x \in \mathbb{F}_2^n : \gamma_1(x) = b_1, \dots, \gamma_{n_g}(x) = b_{n_g}\}.$$
   (b) Set $\Gamma \leftarrow \Gamma \cup \{\gamma_1, \dots, \gamma_{n_g}\}$.
   (c) For each $b = (b_\gamma)_{\gamma \in \Gamma} \in \mathbb{F}_2^{|\Gamma|}$, let V_b be the affine subspace $\{x \in \mathbb{F}_2^n : \forall \gamma \in \Gamma, \gamma(x) = b_\gamma\}$. Set
   $$\mathcal{F} \leftarrow \bigcup_{b \in \mathbb{F}_2^{|\Gamma|}} \{f|_{V_b}\}.$$
   (d) $\mathcal{S} \leftarrow \bigcup_{h \in \mathcal{F}} \operatorname{supp}(\widehat{h})$.
3. Return Γ.

Notation. After each iteration of the while loop in the procedure, Γ is the set of parities that have been queried so far, $\mathcal{F}$ is the set of all restrictions of f to the affine subspaces obtained by different assignments to parities in Γ, and $\mathcal{S}$ is the union of the Fourier supports of functions in $\mathcal{F}$. Throughout this section, i will stand for the index of an arbitrary iteration of NADT. Let $\Gamma^{(i)}, \mathcal{F}^{(i)}$ and $\mathcal{S}^{(i)}$ denote $\Gamma, \mathcal{F}$ and $\mathcal{S}$ respectively at the end of the i-th iteration of the while loop. Let $\Gamma^{(0)} = \varnothing$, $\mathcal{F}^{(0)} = \{f\}$ and $\mathcal{S}^{(0)} = \operatorname{supp}(\widehat{f})$.

For each i, let $b = (b_\gamma)_{\gamma \in \Gamma^{(i)}} \in \mathbb{F}_2^{|\Gamma^{(i)}|}$ and let V_b be the affine subspace defined by the linear constraints $\{\gamma(x) = b_\gamma : \gamma \in \Gamma^{(i)}\}$. In V_b, more than one linear function of the original space may get identified, namely, the restrictions of these linear functions to the sub-domain subdomain V_b are either the same function or negations of each other. More specifically, δ_1 and δ_2 get identified in V_b if and only if $\delta_1 + \delta_2 \in \operatorname{span} \Gamma^{(i)}$ (i. e., they belong to the same coset of the subspace $\operatorname{span} \Gamma^{(i)}$). This equivalence relation induces a partition on $\operatorname{supp}(\widehat{f})$ into equivalence classes. Note that this partition is determined by

$\Gamma^{(i)}$ alone. Further, for each equivalence class, for every $b \in \mathbb{F}_2^{|\Gamma^{(i)}|}$, the linear functions belonging to that class get identified with one another in V_b.

Let $\ell^{(i)}$ denote the number of such equivalence classes.[6] For all $j \in [\ell^{(i)}]$, let $\beta_j^{(i)}$ be an arbitrarily picked representative element in the j-th equivalence class. Let

$$\beta_j^{(i)} + \alpha_{j,1}^{(i)}, \ldots, \beta_j^{(i)} + \alpha_{j,k_j^{(i)}}^{(i)}$$

be the $k_j^{(i)}$ elements in the j-th equivalence class, where $\alpha_{j,1}^{(i)}, \ldots, \alpha_{j,k_j^{(i)}}^{(i)}$ are in $\operatorname{span}\Gamma^{(i)}$. Now define functions

$$P_j^{(i)}(x) := \sum_{l=1}^{k_j^{(i)}} \widehat{f}\left(\beta_j^{(i)} + \alpha_{j,l}^{(i)}\right) \chi_{\alpha_{j,l}^{(i)}}(x).$$

Note that the functions $P_j^{(i)}$ are non-zero. As we will soon see, it will be helpful to think of the $P_j^{(i)}$'s as multilinear polynomials in variables $y_i = (-1)^{x_i}$.

Given this notation, we can then write the Fourier expansion of f in the following form:

$$f(x) = \sum_{j=1}^{\ell^{(i)}} P_j^{(i)}(x) \chi_{\beta_j^{(i)}}(x).$$

Note that the value of each $P_j^{(i)}$ is fixed once the evaluation of each linear form $\gamma \in \Gamma^{(i)}$ is specified. In other words, each $P_j^{(i)}$ is a constant function on each V_b.

Observation 3.1. $\sum_{j=1}^{\ell^{(i)}} k_j^{(i)} = s.$

Proof. The size of the Fourier support of f is the sum of the sizes of its equivalence classes defined above. □

Proposition 3.2. $|\mathcal{S}^{(i)}| = \ell^{(i)}$.

Proof. Clearly $|\mathcal{S}^{(i)}| \leq \ell^{(i)}$, as each term in the Fourier expansion of $f|_{V_b}$ corresponds to a distinct equivalence class (and this correspondence is independent of b). Now, for $j \in [\ell^{(i)}]$, since $P_j^{(i)}$ is a non-zero function, there exists an assignment b to the parities in $\Gamma^{(i)}$ on which $P_j^{(i)}$ evaluates to a non-zero value. Thus the coefficient of $\beta_j^{(i)}$ is non-zero in the restriction of f to the affine subspace obtained by assigning b to the parities in $\Gamma^{(i)}$. Thus for all $j \in [\ell^{(i)}]$, $\beta_j^{(i)} \in \mathcal{S}^{(i)}$ which, together with $|\mathcal{S}^{(i)}| \leq \ell^{(i)}$, implies $|\mathcal{S}^{(i)}| = \ell^{(i)}$. □

We now argue that after every iteration of the while loop, there exists a function $h \in \mathcal{F}^{(i)}$ which has large Fourier support.

[6] $\ell^{(i)}$ is in fact the number of cosets of the subspace $\operatorname{span}\Gamma^{(i)}$ with which $\operatorname{supp}(\widehat{f})$ has non-empty intersection.

Lemma 3.3. *After the i-th iteration, there exists an $h \in \mathcal{F}^{(i)}$ such that $|\operatorname{supp}(\widehat{h})|$ is at least $\left(\ell^{(i)}\right)^2/s$.*

Proof. Consider any function $f|_{V_b} \in \mathcal{F}^{(i)}$. The Fourier decomposition of $f|_{V_b}$ is given by

$$f|_{V_b} = \sum_{j=1}^{\ell^{(i)}} P_j^{(i)}(b) \cdot \chi_{\beta_j^{(i)}}(x).$$

Thus, $|\operatorname{supp}(\widehat{f|_{V_b}})|$ is exactly the number of functions $P_j^{(i)}, j \in [\ell^{(i)}]$ such that $P_j^{(i)}(b)$ is non-zero. We analyze this quantity as follows. Pick a $b \in \mathbb{F}_2^{|\Gamma^{(i)}|}$ uniformly at random. Now, note that for each $j \in [\ell^{(i)}]$, $P_j^{(i)}$ is a non-zero multilinear polynomial in variables $\{(-1)^{b_\gamma} : \gamma \in \Gamma^{(i)}\}$. Thus by Theorem 2.4,

$$\Pr_b[P_j^{(i)}(b) \neq 0] \geq \frac{1}{k_j^{(i)}}.$$

Thus,

$$\begin{aligned}
\mathbb{E}_b\left[|\operatorname{supp}(\widehat{f|_{V_b}})|\right] &\geq \sum_{j=1}^{\ell^{(i)}} \frac{1}{k_j^{(i)}} \\
&\geq \ell^{(i)} \cdot \frac{1}{\left(\sum_{j=1}^{\ell^{(i)}} k_j^{(i)}\right)/\ell^{(i)}} && \text{[By Jensen's inequality applied to } 1/x\text{]} \\
&= \frac{\left(\ell^{(i)}\right)^2}{s} && \text{[By Observation 3.1].}
\end{aligned}$$

Hence, there exists an $h \in \mathcal{F}^{(i)}$ such that $|\operatorname{supp}(\widehat{h})|$ is at least $\left(\ell^{(i)}\right)^2/s$. □

Let $g^{(i)}$ be the function chosen in step 2a of the i-th iteration of NADT. Let sparsity$(\widehat{g^{(i)}}) = s^{(i)}$, and $\Delta\ell^{(i)} = \ell^{(i-1)} - \ell^{(i)}$ for $i \geq 1$. The next lemma proves that, if a function with large Fourier support is picked in step 2a, then that leads to a large reduction in the size of $\mathcal{S}$.

Lemma 3.4. *Assume that neither f nor $1-f$ is a character, and NADT(f) runs for t iterations. Then for all $i \in [t]$, $\Delta\ell^{(i)} \geq s^{(i)}/2$.*

Proof. Let $\gamma_1, \ldots, \gamma_{n_{g^{(i)}}}$ be the parities queried in iteration i. Hence there is a $b = (b_1, \ldots, b_{n_{g^{(i)}}}) \in (\mathbb{F}_2)^{n_{g^{(i)}}}$ such that $g^{(i)}$ is constant on the affine subspace V_b obtained by setting γ_j to b_j for all $j \in [n_{g^{(i)}}]$. Now, consider the following cases.

Case 1. $s^{(i)} \geq 2$**.** Since $g^{(i)}$ is constant on V_b, each Fourier coefficient of $g^{(i)}|_{V_b}$ must be either 0 or ± 1. This is possible only if for every $\alpha_1 \in \operatorname{supp}(\widehat{g^{(i)}})$ there is another $\alpha_2 \in \operatorname{supp}(\widehat{g^{(i)}})$ such that $\alpha_1 + \alpha_2 \in \operatorname{span}\{\gamma_1, \ldots, \gamma_{n_{g^{(i)}}}\}$, i. e., they get identified. This implies that for every

$$b' = (b')_j \in (\mathbb{F}_2)^{n_{g^{(i)}}},$$

in the affine space $V_{b'}$ obtained by restricting each γ_j to b'_j, every parity in $\mathrm{supp}(\widehat{g^{(i)}})$ is identified with some other parity in $\mathrm{supp}(\widehat{g^{(i)}})$. Since $\mathrm{supp}(\widehat{g^{(i)}}) \subseteq \mathcal{S}^{(i-1)}$, it follows that $|\mathcal{S}^{(i)}|$ is at least $|\,\mathrm{supp}(\widehat{g^{(i)}})|/2$ less than $|\mathcal{S}^{(i-1)}|$. By Proposition 3.2 this implies $\Delta\ell^{(i)} \geq s^{(i)}/2$.

Case 2. $s^{(i)} = 1$ (i. e., $g^{(i)}$ is either a parity function or the complement of a parity function). Since neither f nor $1-f$ is a character, we have that $s^{(1)} = \mathrm{sparsity}(\widehat{f}) \geq 2$. Hence we conclude that $i \geq 2$. Note that for $i \geq 1$, $\emptyset \in \mathcal{S}^{(i)}$. Hence for $i \geq 2$, $\Delta\ell^{(i)} = 1 \geq s^{(i)}/2$.

The lemma follows from the two cases. □

Now we are ready to prove Theorem 1.2.

Proof of Theorem 1.2. If either f or $1-f$ is a character, the theorem follows immediately. Hence assume that neither f nor $1-f$ is a character.

We obtain a non-adaptive parity decision tree for f by running NADT. Assume that the while loop runs for t iterations. Let the number of queries made in step 2a in the i-th iteration of the procedure be $q^{(i)}$. By Lemma 1.3, $q^{(i)} = O(\sqrt{s^{(i)}})$. By Lemma 3.4, $\Delta\ell^{(i)} \geq s^{(i)}/2$. Hence,

$$\frac{q^{(i)}}{\Delta\ell^{(i)}} = \frac{1}{\Omega(\sqrt{s^{(i)}})}.$$

From Lemma 3.3 we have $s^{(i)} \geq \left(\ell^{(i-1)}\right)^2/s$. Hence $q^{(i)} = \sqrt{s} \cdot O\left(\Delta\ell^{(i)}/\ell^{(i-1)}\right)$. Thus the total number of queries made within the while loop of the procedure is

$$\begin{aligned}
\sum_{i=1}^{t} q^{(i)} &= \sqrt{s} \cdot \sum_{i=1}^{t} O\left(\frac{\Delta\ell^{(i)}}{\ell^{(i-1)}}\right) = \sqrt{s} \cdot \sum_{i=1}^{t} O\left(\frac{\Delta\ell^{(i)}}{s - \sum_{j=1}^{i-1} \Delta\ell^{(j)}}\right) \\
&\leq \sqrt{s} \cdot \sum_{i=1}^{t} O\left(\frac{1}{s - \sum_{j=1}^{i-1} \Delta\ell^{(j)}} + \frac{1}{s - \sum_{j=1}^{i-1} \Delta\ell^{(j)} - 1} + \cdots + \frac{1}{s - \sum_{j=1}^{i-1} \Delta\ell^{(j)} - (\Delta\ell^{(i)} - 1)}\right) \\
&\leq \sqrt{s} \cdot O\left(\sum_{k=1}^{s} \frac{1}{k}\right) = O(\sqrt{s}\log s).
\end{aligned}$$

From Proposition 2.7 it follows that $\dim(\widehat{f}) = O(\sqrt{s}\log s)$. □

Discussion. As mentioned in the introduction, a potential approach towards disproving Conjecture 1.4 is to assume it to be true, and prove that it implies a $o(\sqrt{s})$ upper bound on Fourier dimension. This will refute the conjecture, since, for the address function (see Section 1),

$$\dim(\widehat{\mathrm{Add}_s}) = \Theta\left(\sqrt{\mathrm{sparsity}(\widehat{\mathrm{Add}_s})}\right).$$

However, we cannot disprove the conjecture by an analysis of the NADT, assuming the conjecture. To see this let us consider the execution of NADT on the address function. Recall that $\mathrm{Add}_s(x, y_1, y_2, \ldots, y_{\sqrt{s}}) = y_x, x \in \{0,1\}^{(1/2)\log s}, y_i \in \{0,1\}$. One easily sees that a largest affine subspace V on which the function is constant is the one defined by the constraints $x = x'$, $y_{x'} = b$ where $x' \in \{0,1\}^{(1/2)\log s}$ and $b \in \{0,1\}$. The function takes the value b everywhere in V. Also, if the address bits x and the bit $y_{x'}$ are set to other values than x' and b, then the restricted functions in the respective affine subspaces are all constants (if x is set to x') or dictators (on the addressed bits). The subsequent steps query different dictators on the addressed bits.

The address function clearly satisfies Conjecture 1.4, and all the intermediate functions that are given rise to by NADT are dictators, which also trivially satisfy the conjecture. Thus this rules out the possibility of refuting Conjecture 1.4 by analyzing NADT assuming the conjecture. We note, however, that if we assume the conjecture, then we can improve the upper bound by a factor of $O(\log s)$, to the optimal $O(\sqrt{s})$.

Acknowledgements

I thank Avishay Tal for observing that an earlier analysis of NADT which proved a $O(s^{2/3})$ upper bound can be tightened to obtain the $O(\sqrt{s}\log s)$ upper bound, and bringing it to my notice. I thank the anonymous referees for pointing out that Theorem 1.2 implies the existence of simultaneous communication protocols for XOR functions, for suggesting simplifications of some proofs, and for suggestions that helped to improve the presentation. I thank Arkadev Chattopadhyay and Prahladh Harsha for many helpful discussions. I thank Prahladh Harsha for his help in significantly improving the presentation of this article.

References

[1] SRINIVASAN ARUNACHALAM, SOURAV CHAKRABORTY, TROY LEE, MANASWI PARAASHAR, AND RONALD DE WOLF: Two new results about quantum exact learning. In *Proc. 46th Internat. Colloq. on Automata, Languages and Programming (ICALP'19)*, pp. 16:1–16:15. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.ICALP.2019.16, arXiv:1810.00481] 2

[2] ARKADEV CHATTOPADHYAY, NIKHIL S. MANDE, AND SUHAIL SHERIF: The log-approximate-rank conjecture is false. In *Proc. 50th STOC*, pp. 42–53. ACM Press, 2019. [doi:10.1145/3313276.3316353] 4

[3] DMITRY GAVINSKY, NAOMI KIRSHNER, ALEX SAMORODNITSKY, AND RONALD DE WOLF: Private communication, 2014. 3

[4] PARIKSHIT GOPALAN, RYAN O'DONNELL, ROCCO A. SERVEDIO, AMIR SHPILKA, AND KARL WIMMER: Testing Fourier dimensionality and sparsity. *SIAM J. Comput.*, 40(4):1075–1100, 2011. Preliminary version in ICALP'09. [doi:10.1137/100785429] 2, 3, 6

[5] TOM GUR AND OMER TAMUZ: Testing Booleanity and the uncertainty principle. *Chicago J. Theoret. Computer Sci.*, 2013(14), 2013. [doi:10.4086/cjtcs.2013.014, arXiv:1204.0944] 6

[6] HAMED HATAMI, KAAVE HOSSEINI, AND SHACHAR LOVETT: Structure of protocols for XOR functions. *SIAM J. Comput.*, 47(1):208–217, 2018. Preliminary version in FOCS'16. [doi:10.1137/17M1136869] 5

[7] ISHAY HAVIV AND ODED REGEV: The list-decoding size of Fourier-sparse Boolean functions. *ACM Trans. Comput. Theory*, 8(3):10:1–10:14, 2016. Preliminary version in CCC'15. [doi:10.1145/2898439, arXiv:1504.01649] 2

[8] SAMPATH KANNAN, ELCHANAN MOSSEL, SWAGATO SANYAL, AND GRIGORY YAROSLAVTSEV: Linear sketching over $\mathbb{F}_2$. In *Proc. 33rd Computational Complexity Conf. (CCC'18)*, pp. 8:1–8:37. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018. [doi:10.4230/LIPIcs.CCC.2018.8, arXiv:1611.01879] 2

[9] ANDREI VYACHESLAVOVICH KHALYAVIN, MIKHAIL SERGEEVICH LOBANOV, AND YURIY VALERIEVICH TARANNIKOV: On plateaued Boolean functions with the same spectrum support. *Sib. Èlektron. Mat. Izv.*, 13:1346–1368, 2016. [doi:10.17377/semi.2016.13.105] 2

[10] SHACHAR LOVETT: Communication is bounded by root of rank. *J. ACM*, 63(1):1:1–1:9, 2016. Preliminary version in STOC'14. [doi:10.1145/2724704, arXiv:1306.1877] 4, 5

[11] ASHLEY MONTANARO AND TOBIAS OSBORNE: On the communication complexity of XOR functions, 2009. [arXiv:0909.3392] 5

[12] SWAGATO SANYAL: Near-optimal upper bound on Fourier dimension of Boolean functions in terms of Fourier sparsity. In *Proc. 42nd Internat. Colloq. on Automata, Languages and Programming (ICALP'15)*, pp. 1035–1045. Springer, 2015. [doi:10.1007/978-3-662-47672-7_84] 1

[13] AMIR SHPILKA, AVISHAY TAL, AND BEN LEE VOLK: On the structure of Boolean functions with small spectral norm. *Comput. Complexity*, 26(1):229–273, 2017. Preliminary version in ITCS'14. [doi:10.1145/2591796.2591799, arXiv:1304.0371] 3, 4

[14] HING YIN TSANG, CHUNG HOI WONG, NING XIE, AND SHENGYU ZHANG: Fourier sparsity, spectral norm, and the Log-rank Conjecture. In *Proc. 54th FOCS*, pp. 658–667. IEEE Comp. Soc. Press, 2013. [doi:10.1109/FOCS.2013.76, arXiv:1304.1245] 3, 4, 5, 6

AUTHOR

Swagato Sanyal
Assistant professor
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
India
swagato@cs.iitkgp.ac.in
http://cse.iitkgp.ac.in/~swagato/

ABOUT THE AUTHOR

SWAGATO SANYAL graduated from the Tata Institute of Fundamental Research, Mumbai, India in 2017; his advisor was Prahladh Harsha. The subject of his thesis was analysis of Boolean functions and query complexity. The work presented in this paper was done as part of his Ph. D. dissertation.