Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 88 (2014)

Near-optimal Upper Bound on Fourier dimension of Boolean Functions in terms of Fourier sparsity

Swagato Sanyal

School of Technology and Computer Science
Tata Institute of Fundamental Research
Mumbai 400005

swagatos@tcs.tifr.res.in

August 9, 2014

Abstract

We prove that the Fourier dimension of any Boolean function with Fourier sparsity s is at most $O\left(\sqrt{s}\log s\right)$. This bound is tight upto a factor of $\log s$ as the Fourier dimension and sparsity of the address function are quadratically related. We obtain the bound by observing that the Fourier dimension of a Boolean function is equivalent to its non-adaptive parity decision tree complexity, and then bounding the latter.

1 Introduction

The study of Boolean functions involves studying various properties of Boolean functions and their inter-relationships. Two such properties, which we investigate in this article, are the Fourier dimension and the Fourier sparsity, which were first studied in the context of property testing by Gopalan *et. al.* [GOS+09]. Given a Boolean function $f : \mathbb{F}_2^n \to \{1, -1\}$ with Fourier expansion

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

Fourier dimension and Fourier sparsity are defined as follows.

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

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

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

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

ISSN 1433-8092

The Fourier sparsity *of f, denoted by* sparsity(f)*, is defined as the size of the support, i.e.,*

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

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

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

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

There are functions (e.g., indicator functions of subspaces) for which the first inequality is tight. For the second inequality, the function known to us having the closest gap between dimension and sparsity is the *address function* $Add_s : \{0,1\}^{\frac{1}{2}\log s + \sqrt{s}} \to \{0,1\}$, defined as

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

In other words, at any input (x, y), $Add_s(x, y)$ is the value of the addresee input bit y_x indexed by the addressing variables x. The address function[1] has sparsity s and dimension at least $\sqrt{s}$. We prove that that this is the tight upper bound for $\dim(f)$ in terms of sparsity(f), upto a factor of $\log s$.[2].

Our main result is the following:

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

Prior to this work, Gavinsky *et. al.* [GKdWS] had proved that the sparsity of any Boolean function with full Fourier dimension n is $\Omega(d\exp(c\sqrt{\log n}))$ for some universal constant c.

Theorem 1.2 is proved using a lemma of Tsang *et. al.* [TWXZ13] bounding the co-dimension of an affine subspace restricted to which the function reduces to a constant, in terms of Fourier sparsity of the function.

Lemma 1.3 (Corollary of [TWXZ13, Lemma 28])**.** *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 co-dimension $O(\sqrt{s})$ such that f is constant on V.*

Proof Idea of Theorem 1.2: We begin by making a simple, but crucial observation 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 upper bounding the Fourier dimension of a Boolean function: exhibiting a shallow non-adaptive parity decision tree of the function.

Towards this end, we first recall the construction of the (adaptive) parity decision tree of Tsang *et. al.* [TWXZ13], which in turn improves on an earlier construction due to Shpilka *et. al.* [STlV14, 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 the original function on different affine subspaces. Then a non-constant leaf is picked arbitrarily, and a small set of linear restrictions is obtained by invoking Lemma 1.3, such that the restricted function at that leaf

[1] To be precise, we should consider the ± 1 valued version of the address function described here, where the 0 and 1 in the range are interpreted as $+1$ and -1 respectively.

[2] This is one of the conjectures proposed in the open problem session at the Simons workshop on Real Analysis in Testing, Learning and Inapproximability.

becomes constant. The next step is observing that if the same function is restricted to all the affine subspaces obtained by setting the same set of parities in all possible ways, the sparsity of each of the corresponding restricted functions is at most half of that of the original function. 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 parity gets identified with at least one other parity. This identification leads to halving of the support. Proceeding in this way, a parity decision tree of depth $\mathrm{O}(\sqrt{s})$ is obtained.

Note that the choice of parities 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), invoke Lemma 1.3, and obtain the restrictions which make it constant. Then we query the same set of parities at every leaf. Then we argue that this leads to a significant reduction of sparsity. Let $\Delta s^{(i)}$ 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 $l^{(i)}$ of the union of the supports of all the leaves falls roughly by $\Delta s^{(i)}/4$. From Lemma 1.3, the number of queries spent in the i-th step is $O(\sqrt{\Delta s^{(i)}})$. Using the *Uncertainty Principle* (Theorem 2.4) one can show that $\Delta s^{(i)} \geq \left(l^{(i)}\right)^2/s$. With all these facts it is easy to show that continuing in this fashion, in a small number of steps and making at most $O(\sqrt{s}\log s)$ queries, the size of the union of the Fourier supports of all the leaves becomes so small that we can query all of them, thereby turning all the leaves into constants. The details of the construction of the non-adaptive parity decision tree, and its analysis, is given in Section 3.

Some Remarks about Lemma 1.3: Lemma 1.3 is not believed to be tight. Tsang *et. al.* investigated this question while studying the *log rank conjecture* in communication complexity for *xor* functions. 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 poly-logarithm of the rank of its communication matrix. Tsang *et. al.* [TWXZ13] suggested a direction towards proving log-rank conjecture for an important class of functions called xor functions. A Boolean function $f(x, y)$ on two n bit inputs is a xor function if there exists a Boolean function F on n bits such that $f(x, y) = F(x \oplus y)$. In particular, the authors propose a protocol for such a f based on a parity decision tree of f and show that the communication complexity of that proposed protocol is polylogarithmic in rank of the communication matrix if the following related conjecture is true.

Conjecture 1.4 ([TWXZ13, Conjecture 27])**.** *There exists a constant $c > 0$ such that for every Boolean function f with Fourier sparsity s, there exists an affine subspace of co-dimension $O\left(\log^c s\right)$ on which f is constant.*

Tsang *et. al.* proved the above conjecture for certain classes of functions, which include functions with constant $\mathbb{F}_2$ degree, and prove Lemma 1.3 for general functions.

We remark 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. If this had not been the case (i.e, our proof actually yielded a super-logarithmic improvement assuming Conjecture 1.4), then this would have refuted the above conjecture since the address function satsifies $\dim = \Theta(\sqrt{\text{sparsity}})$. For further discussion on this topic, please see Section 3.

2 Preliminaries

let $f : \mathbb{F}_2^n \to \{1, -1\}$ be a Boolean function. We think of the range $\{+1, -1\}$ as a subset of $\mathbb{R}$. The inputs to f are n variables $x_1, \ldots, x_n$ which take values in $\mathbb{F}_2$. We identify the additive group in $\mathbb{F}_2$ with the group $\{+1, -1\}$ under real number multiplication, and think of the variables as taking $+1$ and -1 values, where 0 and 1 of $\mathbb{F}_2$ get mapped to $+1$ and -1 respectively. We denote this group isomorphism by $(-1)^{(\cdot)}$, i.e., $(-1)^0$ is 1 and $(-1)^1$ is -1. When the x_i's are ± 1, it is well known that every Boolean function $f(x)$ (where x stands for $x_1, \ldots, x_n$) can be uniquely written as

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \prod_{i \in S} x_i.$$

Thus, when the variables are ± 1, f can be written as a multilinear real polynomial. For every $S \subseteq [n]$, the product $\prod_{i \in S} x_i$ is the logical xor of the bits in S, and $\widehat{f}(S)$ is a real number. These products are exactly the *characters* of $\mathbb{F}_2^n$, which are ± 1 valued versions of the linear forms belonging to the dual vector space $\widehat{\mathbb{F}_2^n}$ of $\mathbb{F}_2^n$. We adopt the following notation in this paper:

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

Here, each $\gamma \in \widehat{\mathbb{F}_2^n}$ is a linear function from $\mathbb{F}_2^n$ to $\mathbb{F}_2$, and $\chi_\gamma(\cdot)$ is $(-1)^{\gamma(\cdot)}$.

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

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

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

Lemma 2.2 (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, $\|\widehat{f}\|_1 \leq \sqrt{s}$.*

Proof.

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

The first inequality follows due to Cauchy-Schwarz inequality while the second equality follows from Parseval's identity. □

For proving our results, we shall use the following version of the Uncertainty Principle. For a proof, the reader is referred to the exercises of chapter 3 of [O'D] where it is given as a hinted exercise.

Theorem 2.4 (Uncertainty Principle)**.** *Let $p : \mathbb{R}^n \to \mathbb{R}$ be a real multilinear 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 stated in the introduction, we need the following theorem due to Tsang *et. al.* [TWXZ13].

Theorem 2.5 ([TWXZ13, Lemma 28])**.** *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 co-dimension $O(A)$ such that f is constant on V.*

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

We end this section by a simple proposition which is crucial to our proofs.

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 $NADT_{\oplus}(f)$), is defined as the minimum integer t such that there exist t linear forms $\gamma_1, \ldots, \gamma_t \in \widehat{\mathbb{F}_2^n}$ such that f is a junta of $\gamma_1, \ldots, \gamma_t$. In other words, on every input, specifying the outputs of the γ_i's specifies the output of f.*

Proposition 2.7. *For a Boolean function f, $NADT_{\oplus}(f) = dim(f)$.*

Proof. If the outputs of a basis of span of $\text{supp}(\widehat{f})$ is specified, then that clearly specifies the outputs of all characters in $\text{supp}(\widehat{f})$, and hence it specifies the output of the function. Thus $\text{NADT}_{\oplus}(f) \leq \dim(f)$.

Now, Let $\text{NADT}_{\oplus}(f) = t$. Let the outputs of $\gamma_1, \ldots, \gamma_t$ specify the output of f. These linear forms are linearly independent as vectors in $\widehat{\mathbb{F}_2^n}$ (else a smaller number of them would decide the output of f). Arbitrarily extend $\gamma_1, \ldots, \gamma_t$ to a basis $\gamma_1, \ldots, \gamma_n$ of $\widehat{\mathbb{F}_2^n}$. For $x = (x_1, \ldots, x_n) \in \mathbb{F}_2^n$, let $L(x) = (\gamma_1(x), \ldots, \gamma_n(x))$. L is easily seen to be an invertible linear transformation from $\mathbb{F}_2^n$ onto itself. Now, $\forall x \in \mathbb{F}_2^n, \forall i = 1, \ldots, n, \gamma_i(x) = (L(x))_i$. Replacing x by $L^{-1}(x)$ we have $\gamma_i(L^{-1}(x)) = x_i$. Now consider the Boolean function $g(x) = f(L^{-1}(x)) = \sum_{\gamma \in \widehat{\mathbb{F}_2^n}} \widehat{f}(\gamma)(-1)^{\gamma(L^{-1}(x))}$. Clearly $\dim(g) = \dim(f)$ as L is a full-rank linear transformation. Also, g is completely specified by the outputs of $\gamma_i(L^{-1}(x))$'s for $i = 1, \ldots, t$. Since $\gamma_i(L^{-1}(x)) = x_i$, we have that g is a junta of $x_1, \ldots, x_t$. Thus all the monomials in $\text{supp}(\widehat{g})$ contain only the variables $x_1, \ldots, x_t$. Thus $\dim(f) = \dim(g) \leq t = \text{NADT}_{\oplus}(f)$.

The proposition follows by combining the two inequalities. ☐

3 Upper Bounding Parity Decision Tree Complexity

In this section, we upper bound the non-adaptive parity decision tree complexity of a Boolean function f with Fourier sparsity at most s. Consider the following procedure, parametrized by a parameter $\tau \in \mathbb{N}$ (that we will set later) that constructs the non-adaptive parity decision tree.

NON-ADAPTIVE-PARITY-DECISION-TREE-PROCEDURE$_\tau(f)$

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

1. Set $\Gamma \leftarrow \emptyset$, $\mathcal{S} \leftarrow \text{supp}(\widehat{f})$ and $\mathcal{F} \leftarrow \{f\}$.
2. While $|\mathcal{S}| > \tau$, do
   (a) Let g be a function in $\mathcal{F}$ with the largest Fourier sparsity. Let $\gamma_1, \ldots, \gamma_{n_g}$ be linear functions and $b_1, \ldots, b_{n_g} \in \mathbb{F}_2$ be such that a largest affine subspace on which g is constant is $\{x \in \mathbb{F}_2^n : \gamma_1(x) = b_1, \ldots, \gamma_{n_g}(x) = b_{n_g}\}$. Query $\gamma_1, \ldots, \gamma_{n_g}$.

(b) Set $\Gamma \leftarrow \Gamma \cup \{\gamma_1, \ldots, \gamma_{n_g}\}$.

(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) $\mathcal{S} \leftarrow \bigcup_{h \in \mathcal{F}} \mathrm{supp}(\widehat{h})$.

3. Query all the parities in $\mathcal{S}$.

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 $\mathcal{S}$ the union of the Fourier supports of functions in $\mathcal{F}$. Let $\Gamma^{(i)}, \mathcal{F}^{(i)}$ and $\mathcal{S}^{(i)}$ denote $\Gamma, \mathcal{F}$ and $\mathcal{S}$ resepectively at the end of the i-th iteration of the while loop. Let $\Gamma^{(0)} = \{\phi\}$, $\mathcal{F}^{(0)} = \{f\}$ and $\mathcal{S}^{(0)} = \mathrm{supp}(\widehat{f})$.

For each i, let $b = (b_\gamma)_{\gamma \in \Gamma^{(i)}} \in \mathcal{F}_2^{|\Gamma^{(i)}|}$ and let V_b be the affine subspace defined by linear constraints $\{\gamma(x) = b_\gamma : \gamma \in \Gamma^{(i)}\}$. In V_b, more than one linear functions of the original space may get identified as same.[3] More specifically, δ_1 and δ_2 get identified as same in V_b if and only if $\delta_1 + \delta_2 \in \mathrm{span}\,\Gamma^{(i)}$, i.e. they belong to the same coset of the supspace $\mathrm{span}\,\Gamma^{(i)}$. Thus, $\mathrm{supp}(\widehat{f})$ gets partitioned into equivalence classes, such that for each class, for every $b \in \mathcal{F}_2^{|\Gamma^{(i)}|}$, the linear functions belonging to that class are identified as same in V_b.

Let $l^{(i)}$ denote the number of cosets of the subspace $\mathrm{span}\,\Gamma^{(i)}$ with which $\mathrm{supp}(\widehat{f})$ has non-empty intersection. For $j = 1, \ldots, l^{(i)}$, let $\beta_j^{(i)}$ be some representative element in $\mathrm{supp}(\widehat{f})$ of the j-th coset of $\mathrm{span}\,\Gamma^{(i)}$ having non-empty intersection with $\mathrm{supp}(\widehat{f})$. For each j, let $\beta_j^{(i)} + \alpha_{j,1}^{(i)}, \ldots, \beta_j^{(i)} + \alpha_{j,k_j}^{(i)}$ be the $k_j^{(i)} (\geq 1)$ elements in $\mathrm{supp}(\widehat{f})$ which are in the same coset of $\mathrm{span}\,\Gamma^{(i)}$ as $\beta_j^{(i)}$. For each i, j, define the polynomials $P_j^{(i)}(x) := \sum_{l=1}^{k_j} \widehat{f}\left(\beta_j^{(i)} + \alpha_{j,l}^{(i)}\right) \chi_{\alpha_{j,l}^{(i)}}(x)$. Note that the polynomials $P_j^{(i)}$, $j = 1, \ldots, l^{(i)}$, are non-zero.

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

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

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

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

Proof. Clearly $|\mathcal{S}^{(i)}| \leq l^{(i)}$. Now, for $j = 1, \ldots, l^{(i)}$, since the polynomial $P_j^{(i)}$ is non-zero, 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 $i = 1, \ldots, l^{(i)}$, $\beta_j^{(i)} \in \mathcal{S}^{(i)}$ which, together with $|\mathcal{S}^{(i)}| \leq l^{(i)}$, implies $|\mathcal{S}^{(i)}| = l^{(i)}$. □

[3] By 'same' we also include their being negations of each other as the smaller subspace is an affine space and not always a vector space.

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.

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

Proof. Consider any function $f|_{V_b} \in \mathcal{F}^{(i)}$. The Fourier decomposition of $f|_{V_b}$ is given by $f|_{V_b} = \sum_{j=1}^{l^{(i)}} P_j^{(i)}(b)\chi_{\beta_j^{(i)}}(x)$. Thus, $|\operatorname{supp}(\widehat{f|_{V_b}})|$ is exactly the number of polynomials $P_j^{(i)}, j = 1, \ldots, l^{(i)}$ 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. For each j, $j = 1, \ldots, l^{(i)}$, by Theorem 2.4, $\Pr_b[P_j^{(i)}(b) \neq 0] \geq \frac{1}{k_j^{(i)}}$ (since each $P_j^{(i)}$ is a non-zero polynomial). Thus,

$$\mathbb{E}_b\left[|\operatorname{supp}(\widehat{f|_{V_b}})|\right] \geq \sum_{j=1}^{l^{(i)}} \frac{1}{k_j^{(i)}} \geq l^{(i)} \cdot \frac{1}{\left(\sum_{j=1}^{l^{(i)}} k_j^{(i)}\right)/l^{(i)}} \qquad \text{[By convexity of } 1/x]$$

$$= \frac{\left(l^{(i)}\right)^2}{s} \qquad \text{[By Observation 3.1].}$$

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

Let $g^{(i)}$ be the function chosen in step 2a of the i-th iteration of NON-ADAPTIVE-PARITY-DECISION-TREE-PROCEDURE. Let $\operatorname{sparsity}(g^{(i)}) = \Delta s^{(i)}$, and $\Delta l^{(i)} = l^{(i-1)} - l^{(i)}$ 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 size of $\mathcal{S}$.

Lemma 3.4. *Assume that* NON-ADAPTIVE-PARITY-DECISION-TREE-PROCEDURE$_\tau(f)$ *be run with $\tau \geq \sqrt{2s}$. Assume that it runs for t iterations. Then for $i = 1, \ldots, t$, $\Delta l^{(i)} \geq \frac{\Delta s^{(i)}}{4}$.*

Proof. Let $\gamma_1, \ldots, \gamma_{n_{g^{(i)}}}$ be the parities queried in iteration i. Hence there is $b = (b_1, \ldots, b_{n_{g^{(i)}}}) \in (\mathbb{F}_2)^{n_{g^{(i)}}}$ such that $g^{(i)}$ is constant on the affine subspace V_b obtained by setting each γ_j to b_j for $j = 1, \ldots, n_{g^{(i)}}$. Since $g^{(i)}$ is constant on V_b, each non-zero parity in it's Fourier support must disappear in V_b. Thus, 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 non-zero parity in $\operatorname{supp}(\widehat{g^{(i)}})$ is matched to some other parity in $\operatorname{supp}(\widehat{g^{(i)}})$. Since $\operatorname{supp}(\widehat{g^{(i)}}) \subseteq \mathcal{S}^{(i-1)}$, it follows that $|\mathcal{S}^{(i)}|$ is at least $\frac{|\operatorname{supp}(\widehat{g^{(i)}})|-1}{2}$ less than $|\mathcal{S}^{(i-1)}|$. By Proposition 3.2 this implies $\Delta l^{(i)} \geq \frac{\Delta s^{(i)}-1}{2}$. Now, $\tau \geq \sqrt{2s}$ implies that for each i, $i = 1, \ldots, t$, $l^{(i-1)} \geq \sqrt{2s}$. From Lemma 3.3, we have that $\Delta s^{(i)} \geq \frac{\left(l^{(i-1)}\right)^2}{s} \geq 2$. Thus $\Delta l^{(i)} \geq \frac{\Delta s^{(i)}-1}{2} \geq \frac{\Delta s^{(i)}}{4}$. □

Now we are ready to prove Theorem 1.2.

Proof of Theorem 1.2. Run NON-ADAPTIVE-PARITY-DECISION-TREE-PROCEDURE with parameter $\tau = \lceil\sqrt{2s}\rceil$. We first prove that the total number of queries made in the *while* loop of the procedure is $O(\sqrt{s}\log s)$. Let the number of queries made in step 2a in i-th iteration of the procedure be $\Delta q^{(i)}$.

By Lemma 1.3, $\Delta q^{(i)} = O(\sqrt{\Delta s(i)})$. By Lemma 3.4, $\Delta l^{(i)} \geq \frac{\Delta s^{(i)}}{4}$. Hence, $\frac{\Delta q^{(i)}}{\Delta l^{(i)}} = \frac{1}{\Omega(\sqrt{\Delta s^{(i)}})}$. From Lemma 3.3 we have $\Delta s^{(i)} \geq \left(l^{(i-1)}\right)^2/s$. hence $\Delta q^{(i)} = \sqrt{s}.O\left(\Delta l^{(i)}/l^{(i-1)}\right)$. Thus the total number of queries made within the *while* loop of the procedure is

$$\sum_{i=1}^{t} \Delta q^{(i)} = \sqrt{s}.\sum_{i=1}^{t} O\left(\frac{\Delta l^{(i)}}{l^{(i-1)}}\right) = \sqrt{s}.\sum_{i=1}^{t} O\left(\frac{\Delta l^{(i)}}{s - \sum_{j=1}^{i-1} \Delta l^{(j)}}\right)$$

$$\leq \sqrt{s}.\sum_{i=1}^{t} O\left(\frac{1}{s - \sum_{j=1}^{i-1} \Delta l^{(j)}} + \frac{1}{s - \sum_{j=1}^{i-1} \Delta l^{(j)} - 1} + \ldots + \frac{1}{s - \sum_{j=1}^{i-1} \Delta l^{(j)} - (\Delta l^{(i)} - 1)}\right)$$

$$\leq \sqrt{s}.\sum_{\ell=1}^{s} O\left(\frac{1}{\ell}\right) = O(\sqrt{s}\log s).$$

Finally, the number of queries made in step 3 of the procedure is $O(\sqrt{s})$ as $\tau = O(\sqrt{s})$. From Proposition 2.7 it follows that $\dim f = O(\sqrt{s}\log s)$. □

Discussion: As said in the Introduction, a natural approach towards disproving Conjecture 1.4 is to assume it to be true, and prove that it implies a $\omega(\sqrt{s})$ upper bound on Fourier dimension. This will refute the conjecture, since, for address function (see Introduction), $\dim(Add_s) = \Theta(\sqrt{\text{sparsity}(Add_s)})$. However, we cannot disprove the conjecture by an analysis of the Non-adaptive-parity-decision-tree-procedure, assuming the conjecture. To see this let us consider the execution of the procedure on the address function. Recall that $Add_s(x, y_1, y_2, \ldots, y_{\sqrt{s}}) = y_x, \quad x \in \{0,1\}^{\frac{1}{2}\log s}, y_i \in \{0,1\}$. 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 $x' \in \{0,1\}^{\frac{1}{2}\log s}$ and $b \in \{0,1\}$. The function takes the value b everywhere in V. Also, if the addressing bits x and the bit $y_{x'}$ are set to other values than x' and b, the restricted functions in the respective affine subspaces are all constants (if x is set to x') or dictators (on addressee bits). This constitutes the first step of Non-adaptive-parity-decision-tree-procedure. Since the size of the union of supports of all restricted functions already drops to $\sqrt{s}$, the subsequent steps are querying different dictators on the addresee bits.

The addres function clearly satisfies Conjecture 1.4, and all the intermediate functions that are given rise to by Non-adaptive-parity-decision-tree-procedure are dictators, which also trivially satisfy the conjecture. Thus this rules out the possibility of refuting Conjecture 1.4 by analysing Non-adaptive-parity-decision-tree-procedure assuming the conjecture. We notice, however, that if we assume the conjecture, we can improve the upper bound by a factor of $\log s$, to the optimal $O(\sqrt{s})$. The complexity of Non-adaptive-parity-decision-tree-procedure will then be dominated by the complexity of step 3 which is $O(\sqrt{s})$.

4 Acknowledgements

The author is grateful to Avishay Tal for noticing a weakness in an earlier analysis of NON-ADAPTIVE-PARITY-DECISION-TREE-PROCEDURE which proved a $O(s^{2/3})$ upper bound, observing that the analysis can be tightened to obtain $O(\sqrt{s}\log s)$ upper bound, and bringing it to the author's notice. The author would like to thank Arkadev Chattopadhyay and Prahladh Harsha for many helpful discussions. The author is thankful to Prahladh Harsha for his help in improving the presentation of this article significantly.

References

[GKdWS] Dmitry Gavinsky, Naomi Kirshner, Ronald de Wolf, and Alex Samorodnitsky. Private Communication.

[GOS+09] Parikshit Gopalan, Ryan O'Donnell, Rocco A. Servedio, Amir Shpilka, and Karl Wimmer. Testing Fourier dimensionality and sparsity. In *ICALP*, pages 500–512, 2009.

[O'D] Ryan O'Donnell. *Analysis of Boolean Functions.* Cambridge University Press.

[STlV14] Amir Shpilka, Avishay Tal, and Ben lee Volk. On the structure of boolean functions with small spectral norm. In *ITCS*, pages 37–48, 2014.

[TWXZ13] Hing Yin Tsang, Chung Hoi Wong, Ning Xie, and Shengyu Zhang. Fourier sparsity, spectral norm, and the log-rank conjecture. In *FOCS*, pages 658–667, 2013.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de