

The List-Decoding Size of Fourier-Sparse Boolean Functions

Ishay Haviv¹ and Oded Regev^{*2}

- 1 School of Computer Science
The Academic College of Tel Aviv-Yaffo, Israel
- 2 Courant Institute of Mathematical Sciences
New York University, USA

Abstract

A function defined on the Boolean hypercube is *k-Fourier-sparse* if it has at most k nonzero Fourier coefficients. For a function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$ and parameters k and d , we prove a strong upper bound on the number of k -Fourier-sparse Boolean functions that disagree with f on at most d inputs. Our bound implies that the number of uniform and independent random samples needed for learning the class of k -Fourier-sparse Boolean functions on n variables exactly is at most $O(n \cdot k \log k)$.

As an application, we prove an upper bound on the query complexity of testing Booleanity of Fourier-sparse functions. Our bound is tight up to a logarithmic factor and quadratically improves on a result due to Gur and Tamuz (Chicago J. Theor. Comput. Sci., 2013).

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases Fourier-sparse functions, list-decoding, learning theory, property testing

Digital Object Identifier 10.4230/LIPIcs.CCC.2015.58

1 Introduction

Functions defined on the Boolean hypercube $\{0, 1\}^n = \mathbb{F}_2^n$ are fundamental objects in theoretical computer science. It is well known that every such function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$ can be represented as a linear combination

$$f = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \chi_S$$

of the 2^n functions $\{\chi_S\}_{S \subseteq [n]}$ defined by $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$. This representation is known as the *Fourier expansion* of the function f , and the numbers $\hat{f}(S)$ are known as its *Fourier coefficients*. The Fourier expansion of functions plays a central role in analysis of Boolean functions and finds applications in numerous areas of theoretical computer science including learning theory, property testing, hardness of approximation, social choice theory, and cryptography. For an in-depth introduction to the topic the reader is referred to the book of O’Donnell [22].

A classical result in learning theory is a general algorithm due to Kushilevitz and Mansour [19], based on results of Linial, Mansour, and Nisan [20] and Goldreich and

* Supported by the Simons Collaboration on Algorithms and Geometry and by the National Science Foundation (NSF) under Grant No. CCF-1320188. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.



Levin [12], which enables to efficiently learn classes of Boolean functions with a “simple” Fourier expansion. A common notion of simplicity of Fourier expansion is its sparsity. A function is said to be k -Fourier-sparse if it has at most k nonzero Fourier coefficients. It follows from [19] that given *query* access to a k -Fourier-sparse Boolean function $f : \mathbb{F}_2^n \rightarrow \{0, 1\}$ it is possible to estimate its Fourier coefficients and to get a good approximation of f in running time polynomial in n and k . Later, it was shown that such running time even allows to reconstruct the function f exactly [13].

In recent years, properties of the Fourier expansion of functions were studied in the property testing framework. We now mention some of those results; since this will not be needed for the sequel, the reader can skip directly to the description of our results in the next section. Gopalan, O’Donnell, Servedio, Shpilka, and Wimmer considered in [13] the problem of testing if a given Boolean function is k -Fourier-sparse or ε -far from any such function. Another problem studied there is that of deciding if a function is k -Fourier-dimensional, that is, the Fourier support, viewed as a subset of \mathbb{F}_2^n , spans a subspace of dimension at most k , or ε -far from satisfying this property. Gopalan et al. [13] established testers for these properties whose query complexities depend only on k and ε . For k -Fourier-sparsity the query complexity was a certain polynomial in k and $1/\varepsilon$ and for k -Fourier-dimensionality it was $O(k \cdot 2^{2k}/\varepsilon)$. They also proved lower bounds of $\Omega(\sqrt{k})$ and $\Omega(2^{k/2})$ respectively. Another parameter associated with Boolean functions is the degree of its representation as a polynomial over \mathbb{F}_2 . The algorithmic task of testing if a function has \mathbb{F}_2 -degree at most d or is ε -far from any such function was considered by Alon et al. [1] and then by Bhattacharyya et al. [6], who proved tight upper and lower bounds of $\Theta(2^d + 1/\varepsilon)$ on the query complexity. Note that all the above properties fall into the class of linear-invariant properties, i.e., properties that are closed under compositions with any invertible linear transformation of the domain. These properties have recently attracted a significant amount of attention in the attempt to characterize efficient testability of them (see [24, 5] for related surveys).

1.1 Our Results

List-decoding size

Our main technical result from which we derive all other results is concerned with the list-decoding size of Fourier-sparse Boolean functions. In general, the list-decoding problem of an error correcting code for a distance parameter d asks to find all the codewords whose Hamming distance from a given word is at most d . Here we consider the (non-linear) binary code of block length 2^n whose codewords represent all the k -Fourier-sparse Boolean functions on n variables.

It is not difficult to show that the total number of such functions is at most $2^{O(nk)}$. Indeed, there are $2^{O(nk)}$ ways to choose the support of \hat{f} , and $2^{O(nk)}$ ways to set those Fourier coefficients which must all be integer multiples of 2^{-n} in $[-1, +1]$. It is also not difficult to show that the distance between any two distinct codewords is at least $2^n/k$. Indeed, it is known that every k -Fourier-sparse Boolean function has \mathbb{F}_2 -degree $d \leq \log_2 k$ (see, e.g., [4, Lemma 3]), and therefore, by the Schwartz-Zippel lemma, every two distinct k -Fourier-sparse Boolean functions disagree on at least $1/k$ fraction of the inputs. As a result, for every function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$ there is at most one codeword of distance smaller than $2^n/(2k)$ from f .

We are not aware of any other known bounds beyond those two naive ones. We address this question in the following theorem.

► **Theorem 1.1.** *For every function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$, the number of k -Fourier-sparse Boolean functions of distance at most d from f is $2^{O(ndk \log k/2^n)}$.*

We observe that for certain choices of k and d the bound given in Theorem 1.1 is tight. For example, let f be the constant zero function, let $k < 2^{0.9n}$ be a power of 2, and take $d = 2^n/k$. Consider all the indicator functions of linear subspaces of \mathbb{F}_2^n of co-dimension $\log_2 k$. Every such function is of distance d from f and is k -Fourier-sparse (see Claim 2.4). The number of such functions is $2^{\Theta(n \log k)} = 2^{\Theta(ndk \log k/2^n)}$.

Learning from samples

As an application of the list-decoding bound, we next consider the problem of learning the class of k -Fourier-sparse Boolean functions on n variables (exactly) from uniform and independent random *samples* (see, e.g., [2, 18] for related work). Let us note already at the outset that all the results mentioned here are *not* efficient: it is not known if there is an algorithm for the problem whose running time is some fixed polynomial in n times an arbitrary function of k . Among other things, such an algorithm would imply a breakthrough on the long-standing open question of learning juntas from samples [7, 21, 25, 18].

The question of recovering a function that is sparse in the Fourier (or other) basis from a few samples is the central question in the area of sparse recovery. It has been intensely investigated for over a decade and, among other things, has applications for compressed sensing and for the data stream model. The best previously known bounds on our question are $O(n \cdot k \log^3 k) \leq O(n^4 \cdot k)$ due to Cheraghchi, Guruswami, and Velingker [11] and $O(n^2 \cdot k \log k) \leq O(n^3 \cdot k)$ due to Bourgain [8], improving on a previous bound of Rudelson and Vershynin [23] (who themselves improved on the work of Candès and Tao [10]). We note in passing that they actually answer a harder question: first, because they handle all functions, not necessarily Boolean-valued; second, because they show that a randomly chosen set of sample locations of the above cardinality is good with high probability simultaneously *for all* k -Fourier-sparse functions (sometimes known as the “deterministic” setting), whereas we only want a random set of sample locations to be good with high probability for any *fixed* k -Fourier-sparse function (the “randomized” setting); finally, because they obtain the recovery result by proving a “restricted isometry property” of the Fourier matrix which among other things implies a recovery algorithm running in time polynomial in 2^n and k .

Using Theorem 1.1, we improve the upper bound on the sample complexity of learning Fourier-sparse Boolean functions.

► **Corollary 1.2.** *The number of uniform and independent random samples required for learning the class of k -Fourier-sparse Boolean functions on n variables is $O(n \cdot k \log k)$.*

We believe that our better bound and its elementary proof shed more light on the problem and might be useful elsewhere. In fact, in a follow-up work [15] we employ the techniques developed here to study the “restricted isometry property” of random submatrices of Fourier (and other) matrices, improving on the aforementioned works [11, 8]. We finally note that a lower bound of $\Omega(k \cdot (n - \log_2 k))$ on the sample complexity can be obtained by considering the problem of learning indicator functions of affine subspaces of \mathbb{F}_2^n of co-dimension $\log_2 k$ (see Theorem 3.7; see, e.g., [3] for the same lower bound in a different setting).

Testing Booleanity

We next consider the problem of testing Booleanity of Fourier-sparse functions, which was introduced and studied by Gur and Tamuz in [14]. In this problem, given access to a

k -Fourier-sparse function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$, one has to decide if f is Boolean, i.e., its image is contained in $\{0, 1\}$, or not. The objective is to distinguish between the two cases with some constant probability using as few queries to f as possible. It was shown in [14] that there exists a (non-adaptive one-sided error) tester for the problem with query complexity $O(k^2)$, and that every tester for the problem has query complexity $\Omega(k)$. Here, we use our result on learning k -Fourier-sparse Boolean functions to improve the upper bound of [14] and prove the following.

► **Theorem 1.3.** *For every k there exists a non-adaptive one-sided error tester that using $O(k \cdot \log^2 k)$ queries to an input k -Fourier-sparse function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$ decides if f is Boolean or not with constant success probability.*

We note that, while the tester established in Theorem 1.3 has an improved query complexity, it is not clear if it is efficient with respect to running time. It can be shown, though, that using the learning algorithm of Fourier-sparse functions that follows from [8, 15] (instead of Corollary 1.2) in our proof of Theorem 1.3, one can obtain an efficient algorithm (running in time polynomial in n and k) with the slightly worse query complexity of $O(k \cdot \log^3 k)$.

Finally, we complement Theorem 1.3 by the following nearly matching lower bound.

► **Theorem 1.4.** *Every non-adaptive one-sided error tester for Booleanity of k -Fourier-sparse functions has query complexity $\Omega(k \cdot \log k)$.*

1.2 Overview of Proofs

1.2.1 The List-Decoding Size of Fourier-Sparse Boolean Functions

In order to prove Theorem 1.1, we have to bound from above the number of k -Fourier-sparse Boolean functions of distance at most d from a general function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$. In the discussion below, let us consider the special case where f is the constant zero function. The general result follows easily.

Here, we have to bound the number of k -Fourier-sparse Boolean functions $g : \mathbb{F}_2^n \rightarrow \{0, 1\}$ of support size at most d . We start by observing using Parseval's theorem that such functions have small spectral norm $\|\hat{g}\|_1 = \sum_{S \subseteq [n]} |\hat{g}(S)|$. Next, we observe that the Fourier expansion of the normalized function $g/\|g\|_1$ is a convex combination of functions $\pm \chi_S$, and thus can be viewed, following a technique of Bruck and Smolensky [9], as an expectation over a distribution on the S 's. Using the Chernoff-Hoeffding bound and the bound on the spectral norm, we obtain a succinct representation for every such function g . The ability to represent these functions by a binary string of bounded length yields the upper bound on their number. We note that the proof approach somewhat resembles that of the upper bound on the list-decoding size of Reed-Muller codes due to Kaufman, Lovett, and Porat [17].

1.2.2 Learning Fourier-Sparse Boolean Functions

As a warmup, let us mention an easy upper bound of $O(n \cdot k^2)$. This follows by recalling that there are at most $2^{O(nk)}$ k -Fourier-sparse Boolean functions, and that each one differs from any fixed function on at least $1/k$ fraction of the inputs. Hence by the union bound, after $O(n \cdot k^2)$ samples all other functions will be eliminated.

The improved bound in Corollary 1.2 follows similarly using the list-decoding result of Theorem 1.1. Namely, we apply the union bound separately on functions of different distances from the input function. Functions that are nearby are harder to “hit” using random samples, but by the theorem, there are few of them; functions that are further away are in abundance, but they are easier to “hit” using random samples.

1.2.3 Testing Booleanity of Fourier-Sparse Functions

The testing Booleanity problem is somewhat different from typical property testing problems. Indeed, in property testing one usually has to distinguish objects that satisfy a certain property from those that are ε -far from the property for some distance parameter $\varepsilon > 0$. However, here the tester is required to decide if the function satisfies the Booleanity property or not, with no distance parameter involved. This unusual setting makes sense in this case because Fourier-sparse non-Boolean functions are always quite far from every Boolean function. More precisely, the authors of [14] used the uncertainty principle (see Proposition 2.1) to prove that every k -Fourier-sparse non-Boolean function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$ is non-Boolean on at least $\Omega(2^n/k^2)$ inputs (see Claim 2.3). This immediately implies a (non-adaptive one-sided error) tester that uses $O(k^2)$ queries: just check that f is Boolean on $O(k^2)$ uniform inputs in \mathbb{F}_2^n .

The analysis of [14] turns out to be tight, as there are k -Fourier-sparse non-Boolean functions that are not Boolean at only $\Theta(2^n/k^2)$ points. Indeed, for an even integer n , consider the function $f : \mathbb{F}_2^n \rightarrow \{0, 1, 2\}$ defined by

$$f(x_1, \dots, x_n) = \text{AND}(x_1, \dots, x_{n/2}) + \text{AND}(x_{n/2+1}, \dots, x_n), \quad (1)$$

which is not Boolean at only one point and has Fourier-sparsity $2 \cdot 2^{n/2}$ (see Claim 2.4).

Upper bound

We prove Theorem 1.3 using our learning result, Corollary 1.2. To do so, we first observe that a restriction of a k -Fourier-sparse non-Boolean function to a random subspace of dimension $O(\log k)$ is non-Boolean with high probability (see Lemma 4.1). Since a restriction to a subspace does not increase the Fourier-sparsity, this reduces our problem to testing Booleanity of k -Fourier-sparse functions on $n = O(\log k)$ variables. Then, after $O(k \cdot \log^2 k)$ samples from the subspace, if a non-Boolean value was found then we are clearly done. Otherwise, by Corollary 1.2, the samples uniquely specify a Boolean candidate for the restricted function. Such a function must be quite far from every other k -Fourier-sparse function (Boolean or not; see Claim 2.2). This enables us to decide if the restricted function equals the Boolean candidate function or not.

Lower bound

The upper bound in Theorem 1.3 gets close to the $\Omega(k)$ lower bound proven by Gur and Tamuz in [14]. For their lower bound, they considered the following two distributions: (a) the uniform distribution over all Boolean n -variable functions that depend only on their first $\log_2 k$ variables; (b) the uniform distribution over all n -variable functions that depend only on their first $\log_2 k$ variables and return a Boolean value on $k - 1$ of the assignments to the relevant variables and the value 2 otherwise. It can be easily seen that any (possibly adaptive) tester that distinguishes with some constant probability between distributions (a) and (b) has query complexity $\Omega(k)$. Since the first distribution is supported on k -Fourier-sparse Boolean functions and the second on k -Fourier-sparse non-Boolean functions, this implies that the same lower bound holds for the query complexity of testing Booleanity of k -Fourier-sparse functions.

Note that the distributions considered above are supported on $\log_2 k$ -Fourier-dimensional functions. It can be seen (say, using the uncertainty principle) that such functions are not Boolean on at least $1/k$ fraction of their inputs, so $O(k)$ random samples suffice for finding a non-Boolean value if exists. Hence, in order to get beyond the $\Omega(k)$ lower bound, we need

to consider k -Fourier-sparse functions that are not Boolean at only $o(1/k)$ fraction of the inputs – our functions will actually have $O(1/k^2)$ fraction of such inputs.

Specifically, we consider the distribution of functions obtained by composing the function f given in (1) with a random invertible affine transformation. This is the class of functions that can be represented as a sum $\mathbf{1}_{V_1} + \mathbf{1}_{V_2}$ of two indicators of affine subspaces $V_1, V_2 \subseteq \mathbb{F}_2^n$ of dimension $n/2$, which intersect at exactly one point. Intuitively, it seems that distinguishing the functions in this class from those where V_1 and V_2 have empty intersection requires the tester to learn the affine subspaces V_1 and V_2 , a task that requires $\Omega(n \cdot 2^{n/2})$ queries. We prove such a lower bound for non-adaptive one-sided error testers. Since the above functions are k -Fourier-sparse for $k = O(2^{n/2})$, the obtained lower bound is $\Omega(k \cdot \log k)$.

2 Preliminaries

Let $[n]$ denote the set $\{1, \dots, n\}$. A function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$ is *Boolean* if its image is contained in $\{0, 1\}$ and is *non-Boolean* otherwise. The *distance* between two functions $f, g : \mathbb{F}_2^n \rightarrow \mathbb{R}$, denoted $\text{dist}(f, g)$, is the number of vectors $x \in \mathbb{F}_2^n$ for which $f(x) \neq g(x)$.

Fourier Expansion

For every $S \subseteq [n]$, let $\chi_S : \mathbb{F}_2^n \rightarrow \{-1, 1\}$ denote the function defined by $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$. It is well known that the 2^n functions $\{\chi_S\}_{S \subseteq [n]}$ form an orthonormal basis of the space of functions $\mathbb{F}_2^n \rightarrow \mathbb{R}$ with respect to the inner product $\langle f, g \rangle = \mathbb{E}_x[f(x) \cdot g(x)]$, where x is distributed uniformly over \mathbb{F}_2^n . Thus, every function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$ can be uniquely represented as a linear combination $f = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \chi_S$ of this basis. This representation is called the *Fourier expansion* of f , and the numbers $\hat{f}(S)$ are referred to as its *Fourier coefficients*. The support of f is defined by $\text{supp}(f) = \{x \in \mathbb{F}_2^n \mid f(x) \neq 0\}$ and the support of \hat{f} , known as the *Fourier spectrum* of f , by $\text{supp}(\hat{f}) = \{S \subseteq [n] \mid \hat{f}(S) \neq 0\}$. We say that f is k -Fourier-sparse¹ if $|\text{supp}(\hat{f})| \leq k$. For every $p \geq 1$ we denote $\|\hat{f}\|_p = (\sum_{S \subseteq [n]} |\hat{f}(S)|^p)^{1/p}$. For $p = 1$, $\|\hat{f}\|_1$ is known as the *spectral norm* of f . *Parseval's theorem* states that $\mathbb{E}_x[f(x)^2] = \|\hat{f}\|_2^2$.

The uncertainty principle says that there is no nonzero function f for which the supports of both f and \hat{f} are small (see, e.g., [22, Exercise 3.15]). We state it below with two simple consequences.

► **Proposition 2.1** (The Uncertainty Principle). *For every nonzero function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$,*

$$|\text{supp}(f)| \cdot |\text{supp}(\hat{f})| \geq 2^n.$$

► **Claim 2.2.** *For every two distinct k -Fourier-sparse functions $f, g : \mathbb{F}_2^n \rightarrow \mathbb{R}$, $\text{dist}(f, g) \geq 2^n/(2k)$.*

Proof. Apply Proposition 2.1 to the function $f - g$, whose Fourier-sparsity is at most $2k$. ◀

► **Claim 2.3** ([14]). *For every k -Fourier-sparse function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$, if f is non-Boolean then*

$$|\{x \in \mathbb{F}_2^n \mid f(x) \notin \{0, 1\}\}| \geq \frac{2}{k^2 + k + 2} \cdot 2^n.$$

¹ Boolean functions are sometimes defined in the literature with range $\{-1, +1\}$ rather than $\{0, 1\}$. Notice that this affects the Fourier-sparsity by at most 1.

Proof. Apply Proposition 2.1 to the function $f \cdot (f - 1)$, whose Fourier-sparsity is at most

$$|\{S \Delta T \mid S, T \in \text{supp}(\hat{f})\}| + |\text{supp}(\hat{f})| \leq \binom{k}{2} + k + 1,$$

where Δ stands for symmetric difference of sets. ◀

We also need the following simple claim.

► **Claim 2.4.** *For every affine subspace $V \subseteq \mathbb{F}_2^n$ of co-dimension k , the indicator function $\mathbf{1}_V : \mathbb{F}_2^n \rightarrow \{0, 1\}$ is 2^k -Fourier-sparse.*

Proof. Since V has co-dimension k , there exist $a_1, \dots, a_k \in \mathbb{F}_2^n$ and $b_1, \dots, b_k \in \mathbb{F}_2$ such that $V = \{x \in \mathbb{F}_2^n \mid \langle x, a_i \rangle = b_i, i = 1, \dots, k\}$. For every i , let $S_i \subseteq [n]$ denote the set whose characteristic vector is a_i , and observe that for every $x \in \mathbb{F}_2^n$,

$$\mathbf{1}_V(x) = \prod_{i=1}^k \left(\frac{1 + (-1)^{b_i} \cdot \chi_{S_i}(x)}{2} \right).$$

This representation implies that $\mathbf{1}_V$ is 2^k -Fourier-sparse. ◀

Chernoff-Hoeffding Bound

► **Theorem 2.5.** *Let X_1, \dots, X_N be N identically distributed independent random variables in $[-a, +a]$ satisfying $\mathbb{E}[X_i] = \mu$ for all i . Then for every $\delta \leq 1/2$ and $N \geq C \cdot a^2 \cdot \log(1/\delta)/\varepsilon^2$, for a universal constant C , it holds that*

$$\Pr \left[\left| \mu - \frac{1}{N} \cdot \sum_{i=1}^N X_i \right| < \varepsilon \right] \geq 1 - \delta.$$

3 The List-Decoding Size of Fourier-Sparse Boolean Functions

We turn to prove Theorem 1.1, which provides an upper bound on the list-decoding size of the code of block length 2^n of all k -Fourier-sparse Boolean functions on n variables. Equivalently, for a general distance d and a function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$ we bound the number of k -Fourier-sparse Boolean functions on n variables of distance at most d from f .

We start by proving that a function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$ with small spectral norm can be well approximated by a linear combination of few functions from $\{\chi_S\}_{S \subseteq [n]}$ with coefficients of equal magnitude. This was essentially proved in [9] and we include here the proof for completeness.

► **Lemma 3.1.** *For every function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$, $\varepsilon > 0$, and $\delta \in (0, 1/2]$, there exists a collection² \mathcal{F} of $O(\|\hat{f}\|_1^2 \cdot \log(1/\delta)/\varepsilon^2)$ subsets of $[n]$ with signs $(a_S \in \{\pm 1\})_{S \in \mathcal{F}}$ such that for all but at most δ fraction of $x \in \mathbb{F}_2^n$ it holds that*

$$\left| f(x) - \frac{\|\hat{f}\|_1}{|\mathcal{F}|} \cdot \sum_{S \in \mathcal{F}} a_S \cdot \chi_S(x) \right| < \varepsilon.$$

² Repetitions of subsets in the collection \mathcal{F} are allowed.

Proof. Observe that the function f can be represented as follows.

$$f = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \chi_S = \sum_{S \subseteq [n]} \frac{|\hat{f}(S)|}{\|\hat{f}\|_1} \cdot \|\hat{f}\|_1 \cdot \text{sign}(\hat{f}(S)) \cdot \chi_S = \mathbb{E}_{S \sim D} [\|\hat{f}\|_1 \cdot \text{sign}(\hat{f}(S)) \cdot \chi_S],$$

where D is the distribution defined by $D(S) = |\hat{f}(S)|/\|\hat{f}\|_1$. Let \mathcal{F} be a collection of $|\mathcal{F}| = O(\|\hat{f}\|_1^2 \cdot \log(1/\delta)/\varepsilon^2)$ independent random samples from the distribution D . For every $x \in \mathbb{F}_2^n$, the Chernoff-Hoeffding bound (Theorem 2.5) implies that with probability at least $1 - \delta$ it holds that

$$\left| f(x) - \frac{1}{|\mathcal{F}|} \cdot \sum_{S \in \mathcal{F}} \|\hat{f}\|_1 \cdot a_S \cdot \chi_S(x) \right| < \varepsilon, \tag{2}$$

where $a_S = \text{sign}(\hat{f}(S))$. By linearity of expectation, it follows that there exist \mathcal{F} and signs $(a_S)_{S \in \mathcal{F}}$ for which (2) holds for all but at most δ fraction of $x \in \mathbb{F}_2^n$, as required. ◀

We now apply Lemma 3.1 to Fourier-sparse functions in $\mathbb{F}_2^n \rightarrow \{-1, 0, +1\}$ with bounded support size, and then, in Corollary 3.3, derive an upper bound on the number of these functions.

► **Corollary 3.2.** *Let $f : \mathbb{F}_2^n \rightarrow \{-1, 0, +1\}$ be a k -Fourier-sparse function satisfying $|\text{supp}(f)| \leq d$. Then for every $\delta \in (0, 1/2]$ there exists a collection \mathcal{F} of $O(dk \log(1/\delta)/2^n)$ subsets of $[n]$ with signs $(a_S \in \{\pm 1\})_{S \in \mathcal{F}}$ such that for all but at most δ fraction of $x \in \mathbb{F}_2^n$ it holds that*

$$\left| f(x) - \frac{\|\hat{f}\|_1}{|\mathcal{F}|} \cdot \sum_{S \in \mathcal{F}} a_S \cdot \chi_S(x) \right| < \frac{1}{2}.$$

Proof. By the Cauchy-Schwarz inequality and Parseval's theorem, we obtain that

$$\frac{\|\hat{f}\|_1^2}{k} \leq \sum_{S \subseteq [n]} \hat{f}(S)^2 = 2^{-n} \cdot \sum_{x \in \mathbb{F}_2^n} f(x)^2 \leq \frac{d}{2^n}.$$

The corollary follows from Lemma 3.1, applied with $\varepsilon = 1/2$, for $|\mathcal{F}| = O(\|\hat{f}\|_1^2 \log(1/\delta)/\varepsilon^2) = O(dk \log(1/\delta)/2^n)$. ◀

► **Corollary 3.3.** *The number of k -Fourier-sparse functions $f : \mathbb{F}_2^n \rightarrow \{-1, 0, +1\}$ satisfying $|\text{supp}(f)| \leq d$ is $2^{O(ndk \log k/2^n)}$.*

Proof. For every k -Fourier-sparse function $f : \mathbb{F}_2^n \rightarrow \{-1, 0, +1\}$ satisfying $|\text{supp}(f)| \leq d$, let \mathcal{F} and $(a_S)_{S \in \mathcal{F}}$ be as given by Corollary 3.2 for, say, $\delta = 1/(5k)$. Since the range of f is $\{-1, 0, +1\}$, it follows that the collection \mathcal{F} , the signs $(a_S)_{S \in \mathcal{F}}$, and the value of $\|\hat{f}\|_1$ define a function of distance at most $\delta \cdot 2^n$ from f . Notice that by Claim 2.2 and our choice of δ , the distance between every two distinct k -Fourier-sparse functions is larger than $2\delta \cdot 2^n$. Thus, a function of distance at most $\delta \cdot 2^n$ from f fully defines f . This implies that f can be represented by a binary string of length $O(n \cdot dk \log k/2^n)$, so the total number of such functions is $2^{O(ndk \log k/2^n)}$. ◀

The bound in Corollary 3.3 implies a bound on the number of Fourier-sparse Boolean functions of bounded distance from a given Boolean function.

► **Corollary 3.4.** *For every k -Fourier-sparse Boolean function $f : \mathbb{F}_2^n \rightarrow \{0, 1\}$, the number of k -Fourier-sparse Boolean functions of distance at most d from f is $2^{O(ndk \log k/2^n)}$.*

Proof. Let $f : \mathbb{F}_2^n \rightarrow \{0, 1\}$ be a k -Fourier-sparse Boolean function. Consider the mapping that maps every k -Fourier-sparse Boolean function $g : \mathbb{F}_2^n \rightarrow \{0, 1\}$, whose distance from f is at most d , to the function $h = f - g$. Observe that h is a $2k$ -Fourier-sparse function from \mathbb{F}_2^n to $\{-1, 0, +1\}$ satisfying $|\text{supp}(h)| \leq d$. By Corollary 3.3, the number of such functions h is bounded by $2^{O(ndk \log k/2^n)}$. Since the above mapping is bijective, this bound holds for the number of functions g as well. \blacktriangleleft

Equipped with Corollary 3.3, we restate and prove Theorem 1.1.

► **Theorem 1.1.** *For every function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$, the number of k -Fourier-sparse Boolean functions of distance at most d from f is $2^{O(ndk \log k/2^n)}$.*

Proof. If there is no k -Fourier-sparse Boolean function of distance at most d from f , then the bound trivially holds. So assume that such a function $g : \mathbb{F}_2^n \rightarrow \{0, 1\}$ exists. Observe that every k -Fourier-sparse Boolean function of distance at most d from f has distance at most $2d$ from g . Thus, by Corollary 3.4 applied to g , the number of such functions is at most $2^{O(ndk \log k/2^n)}$. \blacktriangleleft

3.1 The Sample Complexity of Learning Fourier-Sparse Boolean Functions

The *sample complexity* of learning a class of functions is the minimum number of uniform and independent random samples needed from a function in the class for specifying it with high success probability. Here we consider the class of k -Fourier-sparse Boolean functions on n variables, and show how Theorem 1.1 implies an upper bound on the sample complexity of learning it (Corollary 3.6).

► **Theorem 3.5.** *For every n , $1 < k \leq 2^n$, and a k -Fourier-sparse function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$, the following holds. The probability that when sampling $O(n \cdot k \log k)$ uniform and independent random samples from f , there exists a k -Fourier-sparse Boolean function $g \neq f$ that agrees with f on all the samples is $2^{-\Omega(n \log k)}$.*

Proof. Consider $q = O(nk \log k)$ samples $(x, f(x))$ from a k -Fourier-sparse function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$, where x is distributed uniformly and independently in \mathbb{F}_2^n . By Claim 2.2, the distance between f and every other k -Fourier-sparse function is at least $2^n/(2k)$. For an integer $\ell \in [1, \lceil \log_2 2k \rceil]$, consider all the k -Fourier-sparse Boolean functions whose distance from f is in $[2^{n-\ell}, 2^{n-\ell+1}]$. By Theorem 1.1, the number of such functions is $2^{O(nk \log k/2^\ell)}$. The probability that such a function agrees with q random independent samples of f is at most $(1 - 2^{-\ell})^q$. By the union bound, the probability that at least one of these functions agrees with the q samples is at most

$$2^{O(nk \log k/2^\ell)} \cdot (1 - 2^{-\ell})^q \leq 2^{O(nk \log k/2^\ell)} \cdot e^{-q/2^\ell} \leq 2^{-\Omega(n \log k)},$$

where the last inequality holds for an appropriate choice of $q = O(nk \log k)$. By applying the union bound over all the values of ℓ , it follows that with probability $1 - 2^{-\Omega(n \log k)}$ all the k -Fourier-sparse Boolean functions (besides f) are eliminated, completing the proof. \blacktriangleleft

The following corollary follows immediately from Theorem 3.5 and confirms Corollary 1.2.

► **Corollary 3.6.** *For every n and $1 \leq k \leq 2^n$, the number of uniform and independent random samples required for learning the class of k -Fourier-sparse Boolean functions on n variables with success probability $1 - 2^{-\Omega(n \log k)}$ is $O(n \cdot k \log k)$.*

We end with the following simple lower bound.

► **Theorem 3.7.** *For every n and $1 \leq k \leq 2^n$, the number of uniform and independent random samples required for learning the class of k -Fourier-sparse Boolean functions on n variables with constant success probability is $\Omega(k \cdot (n - \log_2 k))$.*

Proof. Assume without loss of generality that k is a power of 2. Let A be an algorithm for learning the class above with constant success probability $p > 0$ using q uniform and independent random samples. Consider the class \mathcal{G} of indicators of affine subspaces of \mathbb{F}_2^n of co-dimension $\log_2 k$ (i.e., affine subspaces of \mathbb{F}_2^n of size $2^n/k$). By Claim 2.4, the functions in \mathcal{G} are k -Fourier-sparse. Observe that their number satisfies

$$|\mathcal{G}| = 2^{\Theta(n \cdot \min(\log_2 k, n - \log_2 k))}.$$

By Yao's minimax principle, there exists a deterministic algorithm A' (obtained by fixing the random coins of A) that given evaluations of a function, chosen uniformly at random from \mathcal{G} , on a *fixed* collection of q points in \mathbb{F}_2^n , learns it with success probability p .

Now, observe that the expected number of 1-evaluations that A' receives is q/k . By Markov's inequality, the probability that A' receives at least $2q/(pk)$ 1-evaluations is at most $p/2$. It follows that for at least $p/2$ fraction of the functions in \mathcal{G} the algorithm A' receives at most $2q/(pk)$ 1-evaluations and learns them correctly. Assuming that $pk \geq 2$, the number of possible evaluation sequences on these inputs is at most

$$\sum_{i=0}^{2q/(pk)} \binom{q}{i} \leq (k \cdot pe/2)^{2q/(pk)} \leq 2^{O(q \cdot \log_2 k/k)},$$

where for the first inequality we used the standard inequality $\sum_{i=0}^t \binom{q}{i} \leq (qe/t)^t$ which holds for $t \leq q$ (see, e.g., [16, Proposition 1.4]). The above is bounded from below by $|\mathcal{G}| \cdot p/2$, implying that

$$q \geq \Omega(n \cdot \min(\log_2 k, n - \log_2 k) \cdot k / \log_2 k) \geq \Omega(k \cdot (n - \log_2 k)),$$

where the last inequality follows by considering separately the cases of $k \geq 2^{n/2}$ and $k < 2^{n/2}$. In case that $pk < 2$, the number of possible evaluation sequences is at most 2^q , and the bound follows similarly using the assumption that p is a fixed constant. ◀

4 Testing Booleanity of Fourier-Sparse Functions

In this section we prove upper and lower bounds on the query complexity of testing Booleanity of Fourier-sparse functions. For a parameter k , consider the problem in which given access to a k -Fourier-sparse function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$ one has to decide if f is Boolean, i.e., $f(x) \in \{0, 1\}$ for every $x \in \mathbb{F}_2^n$, or not, with some constant success probability.

4.1 Upper Bound

As mentioned before, Gur and Tamuz proved in [14] that every k -Fourier-sparse non-Boolean function f on n variables satisfies $f(x) \notin \{0, 1\}$ for at least $\Omega(2^n/k^2)$ inputs $x \in \mathbb{F}_2^n$ (see Claim 2.3). Thus, querying the input function f on $O(k^2)$ independent and random inputs suffices in order to catch a non-Boolean value of f if such a value exists. In the following lemma it is shown that it is not really needed to choose the $O(k^2)$ random vectors *independently*. It turns out that a restriction of a k -Fourier-sparse non-Boolean function to a random linear

subspace of size $O(k^2)$, that is, of dimension $\approx 2 \log_2 k$, is with high probability non-Boolean. Thus, the tester could randomly pick such a subspace and query f on all of its vectors. This decreases the amount of randomness used in the tester of [14] from $O(nk^2)$ to $O(n \log k)$. More importantly for us, this reduces the problem of testing Booleanity of k -Fourier-sparse functions on n variables to the case of $k = \Theta(2^{n/2})$.

► **Lemma 4.1.** *Let $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$ be a k -Fourier-sparse non-Boolean function, and denote $L = (k^2 + k + 2)/2$. Then, for every $\delta > 0$, the restriction of f to a uniformly chosen random linear subspace of dimension $r \geq \log_2(L/\delta)$ is also non-Boolean with probability at least $1 - \delta$.*

Proof. Let $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$ be a k -Fourier-sparse non-Boolean function. By Claim 2.3, there are at least $2^n/L$ vectors $x \in \mathbb{F}_2^n$ for which $f(x) \notin \{0, 1\}$. This implies that there exists a set S of at least $\log_2(2^n/L)$ linearly independent vectors in \mathbb{F}_2^n on which f is not Boolean. Consider a linear subspace $V \subseteq \mathbb{F}_2^n$ of dimension $n - 1$ chosen uniformly at random. Since the vectors in S are linearly independent, the probability that no vector in S is in V is $2^{-|S|} \leq \frac{L}{2^n}$. It follows that the restriction $f|_V$ of f to V is a k -Fourier-sparse function defined on a linear subspace of dimension $n - 1$, and its probability to be Boolean is at most $\frac{L}{2^n}$. Note that one can think of the domain of $f|_V$ as \mathbb{F}_2^{n-1} , because V and \mathbb{F}_2^{n-1} are isomorphic and a composition with an invertible linear transformation does not affect the Fourier-sparsity. Now, let us repeat the above process $n - r - 1$ additional times, until we get a linear subspace of dimension r . The probability that the function becomes Boolean in one of the steps is at most

$$\frac{L}{2^n} + \frac{L}{2^{n-1}} + \cdots + \frac{L}{2^{r+1}} \leq \frac{L}{2^r} \leq \delta,$$

and we are done. ◀

We now restate and prove Theorem 1.3, which gives an upper bound of $O(k \cdot \log^2 k)$ on the query complexity of testing Booleanity of k -Fourier-sparse functions. In the proof, we first apply Lemma 4.1 to restrict the input function to a subspace of dimension $O(\log k)$. Then, we apply Theorem 3.5 in an attempt to learn the restricted function and check if it is consistent with some k -Fourier-sparse Boolean function.

► **Theorem 1.3.** *For every k there exists a non-adaptive one-sided error tester that using $O(k \cdot \log^2 k)$ queries to an input k -Fourier-sparse function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$ decides if f is Boolean or not with constant success probability.*

Proof. Consider the tester that given access to an input k -Fourier-sparse function $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$ acts as follows:

1. Pick uniformly at random a linear subspace V of \mathbb{F}_2^n of dimension $r = \min(n, \lceil \log_2(100L) \rceil)$, where $L = (k^2 + k + 2)/2$, and let T be an invertible linear transformation mapping \mathbb{F}_2^r to V .
2. Query f on $O(r \cdot k \log k)$ random vectors chosen uniformly and independently from the subspace V . Note that these queries can be seen as uniform and independent random samples from the function $g : \mathbb{F}_2^r \rightarrow \mathbb{R}$ defined as $g = f \circ T$.
3. If there exists a k -Fourier-sparse Boolean function on r variables that agrees with the above samples of g then accept, and otherwise reject.

We turn to prove the correctness of the above tester. If f is a k -Fourier-sparse Boolean function then so is g , because a restriction to a subspace and a composition with a linear transformation leave the function k -Fourier-sparse and Boolean. Hence, in this case the tester accepts with probability 1.

On the other hand, if f is a k -Fourier-sparse non-Boolean function, then by Lemma 4.1 the restriction of f to the random subspace V of dimension r picked in Item 1, as well as the function g defined in Item 2, are also non-Boolean with probability at least 0.99. In this case, by Theorem 3.5, the probability that there is a k -Fourier-sparse Boolean function on r variables that agrees with $O(r \cdot k \log k)$ uniform and independent random samples from g is $2^{-\Omega(r \log k)}$, thus the tester correctly rejects with probability at least, say, 0.9, as required. Finally, observe that the number of queries made by the tester is $O(r \cdot k \log k) = O(k \cdot \log^2 k)$. ◀

4.2 Lower Bound

We turn to restate and prove our lower bound on the query complexity of testing Booleanity of k -Fourier-sparse functions.

► **Theorem 1.4.** *Every non-adaptive one-sided error tester for Booleanity of k -Fourier-sparse functions has query complexity $\Omega(k \cdot \log k)$.*

Proof. For a given integer k , let n be the largest even integer satisfying $k \geq 3 \cdot 2^{n/2}$. Define a distribution D_{no} over functions in $\mathbb{F}_2^n \rightarrow \{0, 1, 2\}$ as follows. Pick uniformly at random a pair (V_1, V_2) of affine subspaces satisfying $\dim(V_1) = \dim(V_2) = n/2$ and $|V_1 \cap V_2| = 1$, and output the sum of indicators $\mathbf{1}_{V_1} + \mathbf{1}_{V_2}$. Notice that, by Claim 2.4, such a function has Fourier-sparsity at most $2 \cdot 2^{n/2} \leq k$. Thus, a function chosen from D_{no} is k -Fourier-sparse and non-Boolean with probability 1.

Let T be a non-adaptive one-sided error randomized tester for Booleanity of k -Fourier-sparse functions with query complexity q and success probability at least $2/3$. By Yao's minimax principle, there exists a deterministic tester T' (obtained by fixing the random coins of T) that rejects a random function chosen from D_{no} with probability at least $2/3$. Since T is non-adaptive and has one-sided error, it follows that T' queries an input function on q fixed vectors $a_1, \dots, a_q \in \mathbb{F}_2^n$, accepts every k -Fourier-sparse Boolean function, and rejects a function chosen from D_{no} with probability at least $2/3$. We turn to prove that $q > (n \cdot 2^{n/2})/1000 = \Omega(k \cdot \log k)$.

Assume in contradiction that $q \leq (n \cdot 2^{n/2})/1000$. Let f be a random function chosen from D_{no} , that is, $f = \mathbf{1}_{V_1} + \mathbf{1}_{V_2}$ for random affine subspaces V_1 and V_2 of dimension $n/2$ satisfying $|V_1 \cap V_2| = 1$. For $i = 1, 2$, let W_i be the affine span of $\{a_1, \dots, a_q\} \cap V_i$. Let E be the event that the intersection of W_1 and W_2 is empty. We turn to prove that if the event E happens then the tester T' accepts the function f and that the probability of this event is at least 0.9. This contradicts the success probability of T' on functions chosen from D_{no} and completes the proof.

► **Lemma 4.2.** *If the event E happens then the tester T' accepts the function f .*

Proof. Assume that the event E happens, i.e., $W_1 \cap W_2 = \emptyset$. Then, there exists an affine subspace V_2' of dimension $n/2 - 1$ satisfying $W_2 \subseteq V_2' \subsetneq V_2$ and $V_1 \cap V_2' = \emptyset$. Consider the function $g = \mathbf{1}_{V_1} + \mathbf{1}_{V_2'}$. By Claim 2.4, g is a Boolean function whose Fourier-sparsity is at most $3 \cdot 2^{n/2} \leq k$, thus it is accepted by T' . However, g satisfies $g(a_i) = f(a_i)$ for every $1 \leq i \leq q$. This implies that T' cannot distinguish between g and f , so it must accept f as well. ◀

► **Lemma 4.3.** *The probability of the event E is at least 0.9.*

Proof. Denote by X the number of vectors in $\{a_1, \dots, a_q\} \cap V_1$. Since V_1 is distributed uniformly over all affine subspaces of dimension $n/2$, the probability that a_i belongs to V_1 is $2^{-n/2}$ for every $1 \leq i \leq q$. Thus, by linearity of expectation,

$$\mathbb{E}[X] = \frac{q}{2^{n/2}} \leq \frac{(n \cdot 2^{n/2})/1000}{2^{n/2}} = \frac{n}{1000}.$$

By Markov's inequality, we obtain that

$$\Pr \left[\dim(W_1) \geq \frac{n}{10} \right] \leq \Pr \left[X \geq \frac{n}{10} \right] \leq \frac{1}{100}.$$

Now, fix a choice of V_1 for which $\dim(W_1) < n/10$, and consider the randomness over the choice of V_2 . Notice that, conditioned on V_1 , V_2 is distributed uniformly over all the affine subspaces of dimension $n/2$ which contain exactly one vector from V_1 . By symmetry, every vector of V_1 has probability $|V_1|^{-1} = 2^{-n/2}$ to belong to V_2 . Thus, the probability that the vector that belongs to both V_1 and V_2 is in W_1 is $|W_1| \cdot 2^{-n/2} < 2^{n/10} \cdot 2^{-n/2} = 2^{-2n/5}$.

Finally, the probability that $W_1 \cap W_2 = \emptyset$ is at least the probability that $W_1 \cap V_2 = \emptyset$, and the latter is at least $1 - (0.01 + 2^{-2n/5}) \geq 0.9$ for every sufficiently large n . \blacktriangleleft

\blacktriangleleft

Acknowledgements. We thank Adi Akavia, Shachar Lovett, and Eric Price for useful discussions and comments.

References

- 1 Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing Reed-Muller codes. *IEEE Trans. on Information Theory*, 51(11):4032–4039, 2005. Preliminary version in RANDOM'03.
- 2 Alexandr Andoni, Rina Panigrahy, Gregory Valiant, and Li Zhang. Learning sparse polynomial functions. In *SODA*, pages 500–510, 2014.
- 3 Khanh Do Ba, Piotr Indyk, Eric Price, and David P. Woodruff. Lower bounds for sparse recovery. In *SODA*, pages 1190–1197, 2010.
- 4 Anna Bernasconi and Bruno Codenotti. Spectral analysis of Boolean functions as a graph eigenvalue problem. *IEEE Trans. on Computers*, 48(3):345–351, 1999.
- 5 Arnab Bhattacharyya. Guest column: On testing affine-invariant properties over finite fields. *SIGACT News*, 44(4):53–72, 2013.
- 6 Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. Optimal testing of Reed-Muller codes. In *FOCS*, pages 488–497, 2010.
- 7 Avrim Blum. Learning a function of r relevant variables. In *COLT*, pages 731–733, 2003.
- 8 Jean Bourgain. An improved estimate in the restricted isometry problem. In *Geometric Aspects of Functional Analysis*, volume 2116 of *Lecture Notes in Mathematics*, pages 65–70. Springer, 2014.
- 9 Jehoshua Bruck and Roman Smolensky. Polynomial threshold functions, AC^0 functions, and spectral norms. *SIAM J. Comput.*, 21(1):33–42, 1992. Preliminary version in FOCS'90.
- 10 Emmanuel J. Candès and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. on Information Theory*, 52(12):5406–5425, 2006.
- 11 Mahdi Cheraghchi, Venkatesan Guruswami, and Ameya Velingker. Restricted isometry of Fourier matrices and list decodability of random linear codes. *SIAM J. Comput.*, 42(5):1888–1914, 2013. Preliminary version in SODA'13.

- 12 Oded Goldreich and Leonid A. Levin. A hard-core predicate for all one-way functions. In *STOC*, pages 25–32, 1989.
- 13 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.
- 14 Tom Gur and Omer Tamuz. Testing Booleanity and the uncertainty principle. *Chicago J. Theor. Comput. Sci.*, 2013, 2013.
- 15 Ishay Haviv and Oded Regev. The restricted isometry property of subsampled Fourier matrices, 2015. Manuscript.
- 16 Stasys Jukna. *Extremal Combinatorics: With Applications in Computer Science*. Texts in theoretical computer science. Springer-Verlag, second edition, 2011.
- 17 Tali Kaufman, Shachar Lovett, and Ely Porat. Weight distribution and list-decoding size of Reed-Muller codes. *IEEE Trans. on Information Theory*, 58(5):2689–2696, 2012. Preliminary version in ICS’10.
- 18 Murat Kocaoglu, Karthikeyan Shanmugam, Alexandros G. Dimakis, and Adam R. Klivans. Sparse polynomial learning and graph sketching. In *NIPS*, pages 3122–3130, 2014.
- 19 Eyal Kushilevitz and Yishay Mansour. Learning decision trees using the Fourier spectrum. *SIAM J. Comput.*, 22(6):1331–1348, 1993. Preliminary version in STOC’91.
- 20 Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform, and learnability. *J. ACM*, 40(3):607–620, 1993. Preliminary version in FOCS’89.
- 21 Elchanan Mossel, Ryan O’Donnell, and Rocco A. Servedio. Learning functions of k relevant variables. *J. Comput. Syst. Sci.*, 69(3):421–434, 2004. Preliminary version in STOC’03.
- 22 Ryan O’Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.
- 23 Mark Rudelson and Roman Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61(8):1025–1045, 2008.
- 24 Madhu Sudan. Invariance in property testing. In *Property Testing – Current Research and Surveys*, volume 6390, pages 211–227. Springer, 2010.
- 25 Gregory Valiant. Finding correlations in subquadratic time, with applications to learning parities and juntas. In *FOCS*, pages 11–20, 2012.