Testing Halfspaces*

Kevin Matulef[†] Ryan O'Donnell[‡] Ronitt Rubinfeld[§] Rocco A. Servedio[¶]

Abstract

This paper addresses the problem of testing whether a Boolean-valued function f is a halfspace, i.e. a function of the form $f(x) = \mathrm{sgn}(w \cdot x - \theta)$. We consider halfspaces over the continuous domain $\mathbf{R}^n$ (endowed with the standard multivariate Gaussian distribution) as well as halfspaces over the Boolean cube $\{-1, 1\}^n$ (endowed with the uniform distribution). In both cases we give an algorithm that distinguishes halfspaces from functions that are ϵ-far from any halfspace using only $\mathrm{poly}(\frac{1}{\epsilon})$ queries, independent of the dimension n.

Two simple structural results about halfspaces are at the heart of our approach for the Gaussian distribution: the first gives an exact relationship between the expected value of a halfspace f and the sum of the squares of f's degree-1 Hermite coefficients, and the second shows that any function that approximately satisfies this relationship is close to a halfspace. We prove analogous results for the Boolean cube $\{-1, 1\}^n$ (with Fourier coefficients in place of Hermite coefficients) for balanced halfspaces in which all degree-1 Fourier coefficients are small. Dealing with general halfspaces over $\{-1, 1\}^n$ poses significant additional complications and requires other ingredients. These include "cross-consistency" versions of the results mentioned above for pairs of halfspaces with the same weights but different thresholds; new structural results relating the largest degree-1 Fourier coefficient and the largest weight in unbalanced halfspaces; and algorithmic techniques from recent work on testing juntas [FKR+02].

1 Introduction

A *halfspace* is a function of the form $f(x) = \mathrm{sgn}(w_1 x_1 + \cdots + w_n x_n - \theta)$. Halfspaces, also known as *linear threshold functions* (abbreviated *LTFs* throughout this paper), are a simple yet powerful class of functions. For decades they have played an important role in complexity theory, optimization, and voting theory, as well as a central role in machine learning (see e.g. [HMP+93, Yao90, Blo62, Nov62, MP68, STC00] and related references).

The relationship between learning and property testing has been the subject of much recent work (see e.g. the references cited in the survey [Ron07]). In a typical learning setup we are given access (via queries or random examples) to an unknown function f from a class $\mathcal{C}$, and we are asked to produce a hypothesis that approximates f. In contrast, in property testing we are given query-access to an arbitrary f and we would like to distinguish whether f is a member of $\mathcal{C}$ or ϵ-far from every member of $\mathcal{C}$. Though it is well known that any proper learning algorithm can be used as a testing algorithm[1] [GGR98], testing potentially requires fewer queries than learning (and indeed when this is the case, a query-efficient testing algorithm can be used to check whether f is close to $\mathcal{C}$ before bothering to run a more query-intensive learning algorithm).

Several classes of Boolean functions that are of interest in learning theory have recently been studied from a testing perspective. For example, [PRS02] show how to test dictator functions, monomials, and $O(1)$-term monotone DNFs with query complexity $O(\frac{1}{\epsilon})$, and [FKR+02] show how to test k-juntas with query complexity $\mathrm{poly}(k, \frac{1}{\epsilon})$. Most recently, [DLM+07] gave a general method (which incorporates ideas and techniques from learning theory) for testing several different function classes corresponding to the property of "having a concise representation." These classes include size-s decision trees, s-term DNF formulas, and s-sparse polynomials; in all cases the tester of [DLM+07] makes only $\mathrm{poly}(s, \frac{1}{\epsilon})$ queries, independent of n.

In this work, we consider the problem of testing LTFs. We feel that this is a natural question to consider, given the immense amount of research that has been dedicated to various versions of the LTF learning problem.[2] Our main result is that LTFs can be tested with a number of queries that is independent of n. In proving our main result we establish a range of new structural results about LTFs, which essentially characterize LTFs in terms of their degree-0 and degree-

*Supported by an NSF Graduate Fellowship, NSF grant 012702-001, NSF grant 0514771, NSF grant 0732334, NSF CAREER award CCF-0347282, NSF award CCF-0523664, and by a Sloan Foundation fellowship.

†MIT.

‡CMU.

§MIT.

¶Columbia.

[1] A learning algorithm is *proper* if the hypothesis that it outputs is itself a member of $\mathcal{C}$.

[2] Since LTFs have been studied so intensively from the learning perspective, it is natural to wonder whether the learning-based approach of [DLM+07] can be used for testing LTFs. However, the techniques in [DLM+07] only apply to classes of functions that are "well-approximated by juntas," in the sense that every function in the class is close to depending on few variables. This is not the case with LTFs, since some LTFs (such as the majority function) are not at all close to depending on few variables. Thus it seems that fundamentally different techniques are required.

1 Fourier coefficients. These results have already proved useful in other work; indeed, structural results from this paper play a crucial role in the recent algorithm of [OS08] which efficiently *learns* any LTF given only its degree-0 and degree-1 Fourier coefficients. As we describe in the conclusion, we are hopeful that our techniques will help resolve other open questions in areas such as derandomization as well.

We start by describing our testing results in more detail.

Our Results. We adopt the standard property testing model, in which the testing algorithm is allowed black-box query access to an unknown function f and must minimize the number of times it queries f. The algorithm must with high probability pass all functions that are LTFs and with high probability fail all functions that have distance at least ϵ from any LTF. Our main algorithmic results are the following:

1. We first consider functions that map $\mathbf{R}^n \to \{-1,1\}$, where we measure the distance between functions with respect to the standard n-dimensional Gaussian distribution. In this setting we give a poly$(\frac{1}{\epsilon})$ query algorithm for testing LTFs with two-sided error.

2. [Main Result.] We next consider functions that map $\{-1,1\}^n \to \{-1,1\}$, where (as is standard in property testing) we measure the distance between functions with respect to the uniform distribution over $\{-1,1\}^n$. In this setting we also give a poly$(\frac{1}{\epsilon})$ query algorithm for testing LTFs with two-sided error.

Discussion. We remark that the dependence on $\frac{1}{\epsilon}$ in our bounds is only polynomial, rather than exponential or tower-type as in some other property testing algorithms. We note that in contrast to testing, any learning algorithm — even one with black-box query access to f — must make at least $\Omega(\frac{n}{\epsilon})$ queries to learn an unknown LTF to accuracy ϵ (this follows easily from, e.g., the results of [KMT93]). Thus the query complexity of learning is linear in n, while the query complexity of our testing algorithm is independent of n.

Note that the assumption that our testing algorithm has query access to f (as opposed to, say, access only to random labeled examples) is necessary to achieve a complexity independent of n. Any LTF testing algorithm with access only to uniform random examples $(x, f(x))$ for $f : \{-1,1\}^n \to \{-1,1\}$ must use at least $\Omega(\log n)$ examples (an easy argument shows that with fewer examples, the distribution on examples labeled according to a truly random function is statistically indistinguishable from the distribution on examples labeled according to a randomly chosen variable from $\{x_1, \ldots, x_n\}$).

We also note that while it is slightly unusual to consider property testing under the standard multivariate Gaussian distribution, our results are much simpler to establish in this setting because the rotational invariance essentially means that we can deal with a 1-dimensional problem. We moreover observe that it seems essentially *necessary* to solve the LTF testing problem in the Gaussian domain in order to solve the problem in the standard $\{-1,1\}^n$ uniform distribution framework; to see this, observe that an unknown function $f : \{-1,1\}^n \to \{-1,1\}$ to be tested could in fact have the structure $f(x_1, \ldots, x_{dm}) = \tilde{f}\left(\frac{x_1+\cdots+x_m}{\sqrt{m}}, \ldots, \frac{x_{(d-1)m+1}+\cdots+x_{dm}}{\sqrt{m}}\right)$ in which case the arguments to $\tilde{f}$ behave very much like d independent standard Gaussian random variables.

Characterizations and Techniques. We establish new structural results about LTFs which essentially characterize LTFs in terms of their degree-0 and degree-1 Fourier coefficients. For functions mapping $\{-1,1\}^n$ to $\{-1,1\}$ it has long been known [Cho61] that any linear threshold function f is *completely specified* by the $n+1$ parameters consisting of its degree-0 and degree-1 Fourier coefficients (also referred to as its *Chow parameters*). While this specification has been used to *learn* LTFs in various contexts [BDJ+98, Gol06, Ser07, OS08], it is not clear how it can be used to construct efficient *testers* (for one thing this specification involves $n+1$ parameters, and in testing we want a query complexity independent of n). Intuitively, we get around this difficulty by giving new characterizations of LTFs as those functions that satisfy a particular relationship between just *two* parameters, namely the degree-0 Fourier coefficient and the sum of the squared degree-1 Fourier coefficients. Moreover, our characterizations are robust in that if a function approximately satisfies the relationship, then it must be close to an LTF. This is what makes the characterizations useful for testing.

We first consider functions mapping $\mathbf{R}^n$ to $\{-1,1\}$ where we view $\mathbf{R}^n$ as endowed with the standard n-dimensional Gaussian distribution. Our characterization is particularly clean in this setting and illustrates the essential approach that also underlies the much more involved Boolean case. On one hand, it is not hard to show that for every LTF f, the sum of the squares of the degree-1 Hermite coefficients[3] of f is equal to a particular function of the mean of f — regardless of *which* LTF f is. We call this function W; it is essentially the

[3]These are analogues of the Fourier coefficients for L^2 functions over $\mathbf{R}^n$ with respect to the Gaussian measure.

square of the "Gaussian isoperimetric" function.

Conversely, Theorem 2.1 shows that if $f : \mathbf{R}^n \to \{-1,1\}$ is *any* function for which the sum of the squares of the degree-1 Hermite coefficients is within $\pm\epsilon^3$ of $W(\mathbf{E}[f])$, then f must be $O(\epsilon)$-close to an LTF — in fact to an LTF whose n weights are the n degree-1 Hermite coefficients of f. The value $\mathbf{E}[f]$ can clearly be estimated by sampling, and moreover it can be shown that a simple approach of sampling f on pairs of correlated inputs can be used to obtain an accurate estimate of the sum of the squares of the degree-1 Hermite coefficients. We thus obtain a simple and efficient test for LTFs under the Gaussian distribution and thereby establish Result 1. This is done in Section 2.

In Section 3 we take a step toward handling general LTFs over $\{-1,1\}^n$ by developing an analogous characterization and testing algorithm for the class of *balanced regular* LTFs over $\{-1,1\}^n$; these are LTFs with $\mathbf{E}[f] = 0$ for which all degree-1 Fourier coefficients are small. The heart of this characterization is a pair of results, Theorems 3.2 and 3.3, which give Boolean-cube analogues of our characterization of Gaussian LTFs. Theorem 3.2 states that the sum of the squares of the degree-1 Fourier coefficients of any balanced regular LTF is approximately $W(0) = \frac{2}{\pi}$. Theorem 3.3 states that any function f whose degree-1 Fourier coefficients are all small and whose squares sum to roughly $\frac{2}{\pi}$ is in fact close to an LTF — in fact, to one whose weights are the degree-1 Fourier coefficients of f. Similar to the Gaussian setting, we can estimate $\mathbf{E}[f]$ by uniform sampling and can estimate the sum of squares of degree-1 Fourier coefficients by sampling f on pairs of correlated inputs. An additional algorithmic step is also required here, namely checking that all the degree-1 Fourier coefficients of f are indeed small; it turns out that this can be done by estimating the sum of *fourth* powers of the degree-1 Fourier coefficients, which can again be obtained by sampling f on (4-tuples of) correlated inputs.

The general case of testing arbitrary LTFs over $\{-1,1\}^n$ is substantially more complex and is dealt with in Section 6 of the full version of this paper [MORS07]. We give a detailed outline of the algorithm in Section 4. Very roughly speaking, the algorithm has three main conceptual steps:

- First the algorithm implicitly identifies a set of $O(1)$ many variables that have "large" degree-1 Fourier coefficients. Even a single such variable cannot be explicitly identified using $o(\log n)$ queries; we perform the implicit identification using $O(1)$ queries by adapting an algorithmic technique from [FKR+02].
- Second, the algorithm analyzes the regular subfunctions that are obtained by restricting these implicitly identified variables; in particular, it checks that there is a single set of weights for the unrestricted variables such that the different restrictions can all be expressed as LTFs with these weights (but different thresholds) over the unrestricted variables. Roughly speaking, this is done using a generalized version of the regular LTF test that tests whether a *pair* of functions are close to LTFs over the same linear form but with different thresholds. The key technical ingredients enabling this are Theorem 37 and 38 from the full paper, which generalize Theorems 3.2 and 3.3 in two ways (to pairs of functions, and to functions which may have nonzero expectation).
- Finally, the algorithm checks that there exists a single set of weights for the restricted variables that is compatible with the different biases of the different restricted functions. If this is the case then the overall function is close to the LTF obtained by combining these two sets of weights for the unrestricted and restricted variables. (Intuitively, since there are only $O(1)$ restricted variables there are only $O(1)$ possible sets of weights to check here.)

Notation and Preliminaries. Except in Section 2, throughout this paper f will denote a function from $\{-1,1\}^n$ to $\{-1,1\}$ (in Section 2 f will denote a function from $\mathbf{R}^n$ to $\{-1,1\}$). We say that a Boolean-valued function g is ϵ-*far* from f if $\Pr[f(x) \neq g(x)] \geq \epsilon$; for f defined over the domain $\{-1,1\}^n$ this probability is with respect to the uniform distribution, and for f defined over $\mathbf{R}^n$ the probability is with respect to the standard n-dimensional Gaussian distribution.

A *linear threshold function*, or LTF, is a Boolean-valued function of the form $f(x) = \mathrm{sgn}(w_1x_1 + ... + w_nx_n - \theta)$ where $w_1, ..., w_n, \theta \in \mathbf{R}$. The w_i's are called *weights*, and θ is called the *threshold*. The sgn function is 1 on arguments ≥ 0, and -1 otherwise.

We make extensive use of Fourier analysis of functions $f : \{-1,1\}^n \to \{-1,1\}$ and Hermite analysis of functions $f : \mathbf{R}^n \to \{-1,1\}$. We present the necessary background on Fourier and Hermite analysis in Appendix A of the full paper [MORS07]

2 A Tester for LTFs over $\mathbf{R}^n$

In this section we consider functions f that map $\mathbf{R}^n$ to $\{-1,1\}$, where we view $\mathbf{R}^n$ as endowed with the standard n-dimensional Gaussian distribution. Recall that a draw of x from this distribution over $\mathbf{R}^n$ is obtained by drawing each coordinate x_i independently from the standard one-dimensional Gaussian distribution with mean zero and variance 1. In this section we will use Hermite analysis on functions.

Gaussian LTF facts. Let $f : \mathbf{R}^n \to \{-1,1\}$ be an LTF, $f(x) = \mathrm{sgn}(w \cdot x - \theta)$, and assume by normalization that $\|w\| = 1$. Now the n-dimensional Gaussian distribution is spherically symmetric, as is the class of LTFs. Thus there is a sense in which all LTFs with a given threshold θ are "the same" in the Gaussian setting. (This is very much untrue in the discrete setting of $\{-1,1\}^n$.) We can thus derive Hermite-analytic facts about all LTFs by studying one particular LTF; say, $f(x) = \mathrm{sgn}(e_1 \cdot x - \theta)$. In this case, the picture is essentially 1-dimensional; i.e., we can think of simply $h : \mathbf{R} \to \{-1,1\}$ defined by $h(x) = \mathrm{sgn}(x - \theta)$, where x is a single standard Gaussian. The only parameter now is $\theta \in \mathbf{R}$. Let us give some simple definitions and facts concerning this function:

DEFINITION 2.1. *Let $h_\theta : \mathbf{R} \to \{-1,1\}$ be the function of one Gaussian random variable x given by $h_\theta(x) = \mathrm{sgn}(x - \theta)$. We write ϕ for the p.d.f. of a standard Gaussian; i.e., $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$.*

1. *We define the function $\mu : \mathbf{R} \cup \{\pm\infty\} \to [-1,1]$ by $\mu(\theta) = \widehat{h_\theta}(0) = \mathbf{E}[h_\theta]$. Explicitly, $\mu(\theta) = -1 + 2\int_\theta^\infty \phi$. Note that μ is a monotone strictly decreasing function, and it follows that μ is invertible.*

2. *We have that $\widehat{h_\theta}(1) = \mathbf{E}[h_\theta(x)x] = 2\phi(\theta)$ (by an easy explicit calculation). We define the function $W : [-1,1] \to [0, 2/\pi]$ by $W(\nu) = (2\phi(\mu^{-1}(\nu)))^2$. Equivalently, W is defined so that $W(\mathbf{E}[h_\theta]) = \widehat{h_\theta}(1)^2$; i.e., W tells us what the squared degree-1 Hermite coefficient should be, given the mean. We remark that W is a function symmetric about 0, with a peak at $W(0) = \frac{2}{\pi}$.*

PROPOSITION 2.1.

1. *If x denotes a standard Gaussian random variable, then $\mathbf{E}[|x - \theta|] = 2\phi(\theta) - \theta\mu(\theta)$.*

2. *$|\mu'| \le \sqrt{2/\pi}$ everywhere, and $|W'| < 1$ everywhere.*

3. *If $|\nu| = 1 - \eta$ then $W(\nu) = \Theta(\eta^2 \log(1/\eta))$.*

Proof. The first statement is because both equal $\mathbf{E}[h_\theta(x)(x - \theta)]$. The bound on μ's derivative holds because $\mu' = -2\phi$. The bound on W's derivative is because $W'(\nu) = 4\phi(\theta)\theta$, where $\theta = \mu^{-1}(\nu)$, and this expression is maximized at $\theta = \pm 1$, where it is $.96788\ldots < 1$. Finally, the last statement follows ultimately from the fact that $1 - \mu(\theta) \sim 2\phi(\theta)/|\theta|$ for $|\theta| \ge 1$.

Having understood the degree-0 and degree-1 Hermite coefficients for the "1-dimensional" LTF $f : \mathbf{R}^n \to \{-1,1\}$ given by $f(x) = \mathrm{sgn}(x_1 - \theta)$, we can immediately derive analogues for general LTFs:

PROPOSITION 2.2. *Let $f : \mathbf{R}^n \to \{-1,1\}$ be the LTF $f(x) = \mathrm{sgn}(w \cdot x - \theta)$, where $w \in \mathbf{R}^n$. By scaling, assume that $\|w\| = 1$. Then:*

1. *$\hat{f}(0) = \mathbf{E}[f] = \mu(\theta)$*

2. *$\hat{f}(e_i) = \sqrt{W(\mathbf{E}[f])}\, w_i$.*

3. *$\sum_{i=1}^n \hat{f}(e_i)^2 = W(\mathbf{E}[f])$.*

Proof. The third statement follows from the second, which we will prove. The first statement is left to the reader. We have $\hat{f}(e_i) = \mathbf{E}_x[\mathrm{sgn}(w \cdot x - \theta)x_i]$. Now $w \cdot x$ is distributed as a standard 1-dimensional Gaussian. Further, $w \cdot x$ and x_i are jointly Gaussian with covariance $\mathbf{E}[(w \cdot x)x_i] = w_i$. Hence $(w \cdot x, x_i)$ has the same distribution as $(y, w_i y + \sqrt{1 - w_i^2} \cdot z)$ where y and z are independent standard 1-dimensional Gaussians. Thus

$$
\begin{aligned}
&\mathop{\mathbf{E}}_x[\mathrm{sgn}(w \cdot x - \theta)x_1] \\
&\quad = \mathbf{E}[\mathrm{sgn}(y - \theta)(w_i y + \sqrt{1 - w_i^2} \cdot z)] \\
&\quad = w_i \widehat{h_\theta}(1) + \mathbf{E}[\mathrm{sgn}(y - \theta)\sqrt{1 - w_i^2} \cdot z] \\
&\quad = w_i \sqrt{W(\mathbf{E}[h_\theta])} + 0 \\
&\quad = \sqrt{W(\mathbf{E}[f])}\, w_i,
\end{aligned}
$$

as desired.

The second item in the above proposition leads us to an interesting observation: if $f(x) = \mathrm{sgn}(w_1 x_1 + \cdots + w_n x_n - \theta)$ is any LTF, then its vector of degree-1 Hermite coefficients, $(\hat{f}(e_1), \ldots, \hat{f}(e_n))$, is parallel to its vector of weights, $(w_1, \ldots, w_n)$.

The tester. We now give a simple algorithm and prove that it accepts any LTF with probability at least 2/3 and rejects any function that is $O(\epsilon)$-far from all LTFs with probability at least 2/3. The algorithm is nonadaptive and has two-sided error; the analysis of the two-sided confidence error is standard and will be omitted.

Given an input parameter $\epsilon > 0$, the algorithm works as follows:

1. Let $\tilde{\mu}$ denote an estimate of $\mathbf{E}[f]$ that is accurate to within additive accuracy $\pm\epsilon^3$.

2. Let $\tilde{\sigma}^2$ denote an estimate of $\sum_{i=1}^n \hat{f}(e_i)^2$ that is accurate to within additive accuracy $\pm\epsilon^3$.

3. If $|\tilde{\sigma}^2 - W(\tilde{\mu})| \leq 2\epsilon^3$ then output "yes," otherwise output "no."

The first step can be performed simply by making $O(1/\epsilon^6)$ independent draws from the Gaussian distribution, querying f on each draw, and letting $\tilde{\mu}$ be the corresponding empirical estimate of $\mathbf{E}[f]$; the result will be $\pm\epsilon^3$-accurate with high probability.

We describe how to perform the second step of estimating $\sum_{i=1}^n \hat{f}(e_i)^2$ in Section 3 of the full version of the paper, see Lemma 16 in particular [MORS07]. As described there, the number of queries required is $O(1/\epsilon^{12})$ for a $\pm\epsilon^3$-accurate additive estimate. (We briefly note that the estimate of $\sum_{i=1}^n \hat{f}(e_i)^2$ is obtained in a fairly straightforward way from an estimate of $\mathbf{E}[f(x)f(y)]$ where x and y are η-correlated n-dimensional Gaussians for a suitably small η; more precisely, x and z are both drawn independently from the usual n-dimensional Gaussian distribution and y is set to be $\eta x + \sqrt{1-\eta^2}\cdot z$. Thus in some sense the 2-query test "does $f(x)$ equal $f(y)$ for x, y generated as described above?" is at the heart of our algorithm for the Gaussian distribution.)

We now analyze the correctness of the test. The "yes" case is quite easy: Since $\tilde{\mu}$ is within $\pm\epsilon^3$ of $\mathbf{E}[f]$, and since $|W'| \leq 1$ for all x (by Proposition 2.1 item 2), we conclude that $W(\tilde{\mu})$ is within $\pm\epsilon^3$ of the true value $W(\mathbf{E}[f])$. But since f is an LTF, this value is precisely $\sum_{i=1}^n \hat{f}(e_i)^2$, by Proposition 2.2 item 3. Now $\tilde{\sigma}^2$ is within $\pm\epsilon^3$ of $\sum_{i=1}^n \hat{f}(e_i)^2$, and so the test indeed outputs "yes".

As for the "no" case, the following theorem implies that any function f which passes the test with high probability is $O(\epsilon)$-close to an LTF (either a constant function ± 1 or a specific LTF defined by $\mathbf{E}[f]$ and f's degree-1 Hermite coefficients):

Theorem 2.1. *Assume that $|\mathbf{E}[f]| \leq 1-\epsilon$. If $|\sum_{i=1}^n \hat{f}(e_i)^2 - W(\mathbf{E}[f])| \leq 4\epsilon^3$, then f is $O(\epsilon)$-close to an LTF (in fact to an LTF whose coefficients are the Hermite coefficients $\hat{f}(e_i)$).*

Proof. Let $\sigma = \sqrt{\sum_i \hat{f}(e_i)^2}$, let $t = \mu^{-1}(\mathbf{E}[f])$, and let $h(x) = \frac{1}{\sigma}\sum \hat{f}(e_i)x_i - t$. We will show that f and the LTF $\mathrm{sgn}(h)$ are $O(\epsilon)$-close, by showing that both functions are correlated similarly with h. We have

$$\mathbf{E}[fh] = \frac{1}{\sigma}\sum_i \hat{f}(e_i)^2 - t\mathbf{E}[f] = \sigma - t\mathbf{E}[f],$$

where the first equality uses Plancherel. On the other hand, by Proposition 2.1 (item 1), we have

$$\begin{aligned}\mathbf{E}[|h|] &= 2\phi(t) - t\mu(t) \\ &= 2\phi(\mu^{-1}(\mathbf{E}[f])) - t\mathbf{E}[f] \\ &= \sqrt{W(\mathbf{E}[f])} - t\mathbf{E}[f]\end{aligned}$$

and thus

$$\begin{aligned}\mathbf{E}[h(\mathrm{sgn}(h)-f)] &= \mathbf{E}[|h| - fh] \\ &= \sqrt{W(\mathbf{E}[f])} - \sigma \\ &\leq \frac{4\epsilon^3}{\sqrt{W(\mathbf{E}[f])}} \\ &\leq C\epsilon^2,\end{aligned}$$

where $C > 0$ is some universal constant. Here the first inequality follows easily from $W(\mathbf{E}[f])$ being $4\epsilon^3$-close to σ^2 (see Fact 5 from the full paper) and the second follows from the assumption that $|\mathbf{E}[f]| \leq 1-\epsilon$, which by Proposition 2.1 (item 3) implies that $\sqrt{W(\mathbf{E}[f])} \geq \Omega(\epsilon)$.

Now given that $\mathbf{E}[h(\mathrm{sgn}(h)-f)] \leq C\epsilon^2$, the value of $\Pr[f(x) \neq \mathrm{sgn}(h(x))]$ is greatest if the points of disagreement are those on which h is smallest. Let p denote $\Pr[f \neq \mathrm{sgn}(h)]$. Since h is a normal random variable with variance 1, it is easy to see that $\Pr[|h| \leq p/2] \leq \frac{1}{\sqrt{2\pi}}p \leq p/2$. It follows that f and $\mathrm{sgn}(h)$ disagree on a set of measure at least $p/2$, over which $|h|$ is at least $p/2$. Thus, $\mathbf{E}[h(\mathrm{sgn}(h)-f)] \geq 2\cdot(p/2)\cdot(p/2) = p^2/2$. Combining this with the above, it follows that $p \leq \sqrt{2C}\cdot\epsilon$, and we are done.

3 A Tester for Balanced Regular LTFs over $\{-1,1\}^n$

It is natural to hope that an algorithm similar to the one we employed in the Gaussian case — estimating the sum of squares of the degree-1 Fourier coefficients of the function, and checking that it matches up with W of the function's mean — can be used for LTFs over $\{-1,1\}^n$ as well. It turns out that LTFs which are what we call "regular" — i.e., they have all their degree-1 Fourier coefficients small in magnitude — are amenable to the basic approach from Section 2, but LTFs which have large degree-1 Fourier coefficients pose significant additional complications. For intuition, consider $\mathrm{Maj}(x) = \mathrm{sgn}(x_1 + \cdots + x_n)$ as an example of a highly regular LTF and $\mathrm{sgn}(x_1)$ as an example of an LTF which is highly non-regular. In the first case, the argument $x_1 + \cdots + x_n$ behaves very much like a Gaussian random variable so it is not too surprising that the Gaussian approach can be made to work; but in the second case, the ± 1-valued random variable x_1 is very unlike a Gaussian.

We defer the general case to Section 4, and here present a tester for *balanced, regular* LTFs.

Definition 3.1. *We say that any function $f : \{-1,1\}^n \to \{-1,1\}$ is "τ-regular" if $|\hat{f}(i)| \leq \tau$ for all $i \in [n]$.*

Definition 3.2. *We say that an LTF $f : \{-1,1\}^n \to \{-1,1\}$ is "balanced" if it has threshold zero and mean* *zero. We define* $\mathrm{LTF}_{n,\tau}$ *to be the class of all balanced,* τ*-regular LTFs.*

The balanced regular LTF subcase gives an important conceptual ingredient in the testing algorithm for general LTFs and admits a relatively self-contained presentation. As we discuss in Section 4, though, significant additional work is required to get rid of either the "balanced" or "regular" restriction.

The following theorem shows that we can test the class $\mathrm{LTF}_{n,\tau}$ with a constant number of queries:

THEOREM 3.1. *Fix any* $\tau > 0$. *There is an* $O(1/\tau^8)$*-query algorithm* A *that satisfies the following property: Let* ϵ *be any value* $\epsilon \geq C\tau^{1/6}$*, where* C *is an absolute constant. Then if* A *is run with input* ϵ *and black-box access to any* $f : \{-1,1\}^n \to \{-1,1\}$,

- *if* $f \in \mathrm{LTF}_{n,\tau}$ *then* A *outputs "yes" with probability at least* $2/3$*;*
- *if* f *is* ϵ*-far from every function in* $\mathrm{LTF}_{n,\tau}$ *then* A *outputs "no" with probability at least* $2/3$.

The algorithm A in Theorem 3.1 has two steps. The purpose of Step 1 is to check that f is roughly τ-regular; if it is not, then the test rejects since f is certainly not a τ-regular LTF. In Step 2, A checks that $\sum_{i=1}^n \hat{f}(i)^2 \approx W(0) = \frac{2}{\pi}$. This check is based on the idea (see Section 3.1) that for *any* regular function f, the degree-1 Fourier weight is close to $\frac{2}{\pi}$ if and only if f is close to being an LTF. (Note the correspondence between this statement and the results of Section 2 in the case $\mathbf{E}[f] = 0$.)

We now describe algorithm A, which takes as input a parameter $\epsilon \geq C\tau^{1/6}$:

1. First A estimates $\sum_{i=1}^n \hat{f}(i)^4$ to within an additive $\pm\tau^2$. If the estimate is greater than $2\tau^2$ then A halts and outputs "no," otherwise it continues.
2. Next A estimates $\sum_{i=1}^n \hat{f}(i)^2$ to within an additive $\pm C_1\tau^{1/3}$ (where $C_1 > 0$ is an absolute constant specified below). If this estimate is within an additive $\pm 2C_1\tau^{1/3}$ of $\frac{2}{\pi}$ then A outputs "yes", otherwise it outputs "no."

A description of how the sums of powers of degree-1 Fourier coefficients can be estimated is given in Section 3 of the full paper, see Lemma 16 in particular [MORS07]. The basic idea is that for randomly chosen strings $x^1, x^2, \cdots, x^{p-1}$, and a randomly chosen "noise vector" $\mu \in \{-1,1\}^n$ whose bits are independently 1 with probability $\frac{1}{2} + \frac{1}{2}\eta$, the estimatable quantity $\mathbf{E}[f(x^1)f(x^2)\cdots f(x^{p-1})f(x^1 \odot x^2 \odot \cdots \odot x^{p-1} \odot \mu)]$ (where $\odot$ denotes coordinatewise multiplication) is equal to $\sum_{S\subseteq[n]} \eta^{|S|}\hat{f}(S)^p$, which in turn is very close to $\mathbf{E}[f] + \eta\sum_{i=1}^n \hat{f}(S)^p$ for small values of η.

In Section 3.1, we prove two theorems showing that balanced regular LTFs are essentially characterized by the property $\sum_{i=1}^n \hat{f}(i)^2 \approx \frac{2}{\pi}$. In Section 3.2 we prove correctness of the test.

3.1 Useful theorems about $\mathrm{LTF}_{n,\tau}$**.** The first theorem of this section tells us that any $f \in \mathrm{LTF}_{n,\tau}$ has sum of squares of degree-1 Fourier coefficients very close to $\frac{2}{\pi}$. The next theorem is a sort of dual; it states that any Boolean function f whose degree-1 Fourier coefficients are all small and have sum of squares $\approx \frac{2}{\pi}$ is close to being a balanced regular LTF (in fact, to the LTF whose weights equal f's degree-1 Fourier coefficients). Note the similarity in spirit between these results and the characterization of LTFs with respect to the Gaussian distribution that was provided by Proposition 2.2 item 3 and Theorem 2.1.

THEOREM 3.2. *Let* $f \in \mathrm{LTF}_{n,\tau}$*. Then* $\left|\sum_{i=1}^n \hat{f}(i)^2 - \frac{2}{\pi}\right| \leq O(\tau^{2/3})$.

Proof. Let $\rho > 0$ be small (chosen later). Using Proposition 7.1 and Theorem 5 of [KKMO07], we have

$$\sum_S \rho^{|S|}\hat{f}(S)^2 = \frac{2}{\pi}\arcsin\rho \pm O(\tau).$$

On the LHS side we have that $\hat{f}(S) = 0$ for all even $|S|$ since f is an odd function, and therefore, $|\sum_S \rho^{|S|}\hat{f}(S)^2 - \rho\sum_{|S|=1}\hat{f}(S)^2| \leq \rho^3\sum_{|S|\geq 3}\hat{f}(S)^2 \leq \rho^3$. On the RHS, by a Taylor expansion we have $\frac{2}{\pi}\arcsin\rho = \frac{2}{\pi}\rho + O(\rho^3)$. We thus conclude

$$\rho\sum_{i=1}^n \hat{f}(i)^2 = \frac{2}{\pi}\rho \pm O(\rho^3 + \tau).$$

Dividing by ρ and optimizing with $\rho = \Theta(\tau^{1/3})$ completes the proof.

THEOREM 3.3. *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *be any function such that* $|\hat{f}(i)| \leq \tau$ *for all* i *and* $|\sum_{i=1}^n \hat{f}(i)^2 - \frac{2}{\pi}| \leq \gamma$. *Write* $\ell(x) := \sum_{i=1}^n \hat{f}(i)x_i$. *Then* f *and* $\mathrm{sgn}(\ell(x))$ *are* $O(\sqrt{\gamma+\tau})$*-close.*

Proof. First note that if $\gamma > 1/3$ then the claimed bound is vacuous, so we may assume that $\gamma \leq 1/3$. Let $L := \sqrt{\sum_{i=1}^n \hat{f}(i)^2}$; note that by our assumption on γ we have $L \geq \frac{1}{2}$. We have:

$$(2/\pi) - \gamma \quad \leq \quad \sum_{i=1}^n \hat{f}(i)^2$$

$$
\begin{aligned}
(3.1) \qquad &= \mathbf{E}[f\ell] \\
(3.2) \qquad &\leq \mathbf{E}[|\ell|] \\
(3.3) \qquad &\leq \sqrt{2/\pi}\cdot L + O(\tau) \\
&\leq \sqrt{2/\pi}\sqrt{2/\pi+\gamma} + O(\tau) \\
&\leq (2/\pi) + O(\gamma) + O(\tau).
\end{aligned}
$$

The equality in (3.1) is Plancherel's identity, and the inequality in (3.2) is because f is a ± 1-valued function. The inequality (3.3) holds for the following reason: $\ell(x)$ is a linear form over random ± 1's in which all the coefficients are at most τ in absolute value. Hence we expect it to act like a Gaussian (up to $O(\tau)$ error) with standard deviation L, which would have expected absolute value $\sqrt{2/\pi}\cdot L$. See Propositions 58 and 59 in the full paper for the precise justification. Comparing the overall left- and right-hand sides, we conclude that $\mathbf{E}[|\ell|] - \mathbf{E}[f\ell] \leq O(\gamma) + O(\tau)$.

Let ϵ denote the fraction of points in $\{-1,1\}^n$ on which f and $\mathrm{sgn}(\ell)$ disagree. Given that there is a ϵ fraction of disagreement, the value $\mathbf{E}[|\ell|] - \mathbf{E}[f\ell]$ is smallest if the disagreement points are precisely those points on which $|\ell(x)|$ takes the smallest value. Now again we use the fact that ℓ should act like a Gaussian with standard deviation L, up to some error $O(\tau/L) \leq O(2\tau)$; we can assume this error is at most $\epsilon/4$, since if $\epsilon \leq O(\tau)$ then the theorem already holds. Hence we have (see Theorem 55 in the full paper for precise justification)

$$
\begin{aligned}
&\Pr[|\ell| \leq \epsilon/8] \\
&\quad = \Pr[|\ell/L| \leq \epsilon/8L] \\
&\quad \leq \Pr[|N(0,1)| \leq \epsilon/8L] + \epsilon/4 \leq \epsilon/8L + \epsilon/4 \\
&\quad \leq \epsilon/2
\end{aligned}
$$

since $L \geq 1/2$. It follows that at least an $\epsilon/2$ fraction of inputs x have both $f(x) \neq \mathrm{sgn}(\ell(x))$ and $|\ell(x)| > \epsilon/8$. This implies that $\mathbf{E}[|\ell|] - \mathbf{E}[f\ell] \geq 2\cdot(\epsilon/2)\cdot(\epsilon/8) = \epsilon^2/8$. Combining this with the previous bound $\mathbf{E}[|\ell|] - \mathbf{E}[f\ell] \leq O(\gamma) + O(\tau)$, we get $\epsilon^2/8 \leq O(\gamma) + O(\tau)$ which gives the desired result.

3.2 Proof of Theorem 3.1. First observe that for any Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, if $|\hat{f}(i)| \leq \tau$ for all i then $\sum_{i\in T} \hat{f}(i)^4 \leq \tau^2 \sum_{i \in T} \hat{f}(i)^2 \leq \tau^2$, using Parseval. On the other hand, if $|\hat{f}(i)| \geq 2\tau^{1/2}$ for some i, then $\sum_{i=1}^n \hat{f}(i)^4$ is certainly at least $16\tau^2$.

Suppose first that the function f being tested belongs to $\mathrm{LTF}_{n,\tau}$. As explained above, in this case f will with high probability pass Step 1 and continue to Step 2. By Theorem 3.2 the true value of $\sum_{i=1}^n \hat{f}(i)^2$ is within an additive $O(\tau^{2/3})$ of $\frac{2}{\pi}$; since $O(\tau^{2/3}) \leq C_1\tau^{1/3}$ the algorithm outputs "yes" with high probability. So the algorithm behaves correctly on functions in $\mathrm{LTF}_{n,\tau}$.

Now suppose $f : \{-1,1\}^n \to \{-1,1\}$ is such that the algorithm outputs "yes" with high probability; we show that f must be ϵ-close to some function in $\mathrm{LTF}_{n,\tau}$. Since there is a low probability that A outputs "no" in Step 1 on f, it must be the case that each $|\hat{f}(i)|$ is at most $2\tau^{1/2}$. Since f outputs "yes" with high probability in Step 2, it must be the case that $\sum_{i=1}^n \hat{f}(i)^2$ is within an additive $O(\tau^{1/3})$ of $\frac{2}{\pi}$. Plugging in $2\tau^{1/2}$ for "τ" and $O(\tau^{1/3})$ for "γ" in Theorem 3.3, we have that f is $C\tau^{1/6}$-close to $\mathrm{sgn}(\ell(x))$ where C is some absolute constant. This proves the correctness of A.

To analyze the query complexity, note that Corollary 13 in the full paper tells us that Step 1 requires $O(1/\tau^8)$ many queries, and Step 2 only $O(1/\tau^{4/3})$, so the total query complexity is $O(1/\tau^8)$.

4 A constant-query algorithm for testing arbitrary LTFs over $\{-1,1\}^n$

In this section we outline the ideas that underly the constant-query test for general LTFs. We stress that many technical details, in some cases subtle ones, have been suppressed.

As we saw in Section 3, it is possible to test a function f for being close to a balanced τ-regular LTF. The key observation was that such functions have $\sum_{i=1}^n \hat{f}(i)^2$ approximately equal to $\frac{2}{\pi}$ if and only if they are close to LTFs. Furthermore, in this case, the functions are actually close to being the sign of their degree-1 Fourier part. It remains to extend the test described there to handle general LTFs which may be unbalanced and/or non-regular.

A clear approach suggests itself for handling unbalanced regular LTFs using the $W(\cdot)$ function as in Section 2. This is to try to show that for f an arbitrary τ-regular function, the following holds: $\sum_{i=1}^n \hat{f}(i)^2$ is approximately equal to $W(\mathbf{E}[f])$ if and only if f is close to an LTF — in particular, close to an LTF whose linear form is the degree-1 Fourier part of f. The "only if" direction here is not too much more difficult than Theorem 3.3 (see Theorem 38 in Section 6.2 of the full paper), although the result degrades as the function's mean gets close to 1 or -1. However the "if" direction turns out to present a significant probabilistic difficulty.

In the proof of Theorem 3.2, the special case of mean-zero, we appeal to two results from [KKMO07]. The first shows that a balanced τ-regular LTF can be represented with "small weights" (small compared to their sum-of-squares); the second shows that $\sum_S \rho^{|S|}\hat{f}(S)^2$ is close to $\frac{2}{\pi}\arcsin\rho$ for balanced LTFs with small weights. It is not too hard to appropriately generalize the second of these to unbalanced LTFs with small weights (see Theorem 37 in Section 6.2 of the full paper) . However generalizing the first result to unbalanced LTFs is quite complicated, and requires the following theorem, which we prove in Section 6.1 of the full paper [MORS07] :[4]

THEOREM 4.1. *Let* $f(x) = \mathrm{sgn}(w_1x_1 + \cdots + w_nx_n - \theta)$ *be an LTF such that* $\sum_i w_i^2 = 1$ *and* $\delta := |w_1| \geq |w_i|$ *for all* $i \in [n]$. *Let* $0 \leq \epsilon \leq 1$ *be such that* $|\mathbf{E}[f]| \leq 1 - \epsilon$. *Then* $|\hat{f}(1)| \geq \Omega(\delta\epsilon^6 \log(1/\epsilon))$.

We now discuss removing the regularity condition; this requires additional analytic work and moreover requires that several new algorithmic ingredients be added to the test. Given any Boolean function f, Parseval's inequality implies that $J := \{i : |\hat{f}(i)| \geq \tau^2\}$ has cardinality at most $1/\tau^4$. Let us pretend for now that the testing algorithm could somehow know the set J. (If we allowed the algorithm $\Theta(\log n)$ many queries, it *could* in fact exactly identify some set like J. However with constantly many queries this is not possible. We ignore this problem for the time being, and will discuss how to get around it at the end of this section.)

Our algorithm first checks whether it is the case that for all but an ϵ fraction of restrictions ρ to J, the restricted function f_ρ is ϵ-close to a constant function. If this is the case, then f is an LTF if and only if f is close to an LTF which depends only on the variables in J. So in this case the tester simply enumerates over "all" LTFs over J and checks whether f seems close to any of them. (Note that since J is of constant size there are at most constantly many LTFs to check here.)

It remains to deal with the case that for at least an ϵ fraction of restrictions to J, the restricted function is ϵ-far from a constant function. In this case, it can be shown using Theorem 4.1 that if f is an LTF then in fact *every* restriction of the variables in J yields a regular subfunction. So it can use the testing procedure for (general mean) regular LTFs already described to check that for most restrictions π, the restricted function f_π is close to an LTF — indeed, close to an LTF whose linear form is its own degree-1 Fourier part.

This is a good start, but it is not enough. At this point the tester is confident that most restricted functions f_π are close to LTFs whose linear forms are their own degree-1 Fourier parts — but in a true LTF, all of these restricted functions are expressible using a *common* linear form. Thus the tester needs to test *pairwise consistency* among the linear parts of the different f_π's.

[4]Readers familiar with the notion of influence will recall that for any LTF f we have $\mathrm{Inf}_i(f) = |\hat{f}(i)|$ for each i. Thus Theorem 4.1 may roughly be viewed as saying that "every not-too-biased LTF with a large weight has an influential variable."

To do this, recall that when the algorithm tests that a restricted function f_π is close to an LTF, the actual test is that there is near-equality in the inequality $\sum_{|S|=1} \widehat{f_\pi}(S)^2 \leq W(\mathbf{E}[f_\pi])$. If this holds for both f_π and $f_{\pi'}$, the algorithm can further check that the degree-1 parts of f_π and $f_{\pi'}$ are essentially parallel (i.e., equivalent) by testing that near-equality holds in the Cauchy-Schwarz inequality $\sum_{|S|=1} \widehat{f_\pi}(S)\widehat{f_{\pi'}}(S) \leq \sqrt{W(\mathbf{E}[f_\pi])}\sqrt{W(\mathbf{E}[f_{\pi'}])}$. Thus to become convinced that most restricted f_π's are close to LTFs over the *same* linear form, the tester can pick a particular f_{π^*} and check that $\sum_{|S|=1} \widehat{f_{\pi^*}}(S)\widehat{f_\pi}(S) \approx \sqrt{W(\mathbf{E}[f_{\pi^*}])} \cdot \sqrt{W(\mathbf{E}[f_\pi])}$ for most π's. (At this point there is one caveat. As mentioned earlier, the general-mean LTF tests degrade when the function being tested has mean close to 1 or -1. For the above-described test to work, f_{π^*} needs to have mean somewhat bounded away from 1 and -1, so it is important that the algorithm uses a restriction π^* that has $|\mathbf{E}[f]|$ bounded away from 1. Fortunately, finding such a restriction is not a problem since we are in the case in which at least an ϵ fraction of restrictions have this property.)

Now the algorithm has tested that there is a single linear form ℓ (with small weights) such that for most restrictions π to J, f_π is close to being expressible as an LTF with linear form ℓ. It only remains for the tester to check that the thresholds — or essentially equivalently, for small-weight linear forms, the means — of these restricted functions are consistent with some arbitrary weight linear form on the variables in J. It can be shown that there are at most $2^{\mathrm{poly}(|J|)}$ essentially different such linear forms $w \cdot \pi - \theta$, and thus the tester can just enumerate all of them and check whether for most π's it holds that $\mathbf{E}[f_\pi]$ is close to the mean of the threshold function $\mathrm{sgn}(\ell - (\theta - w \cdot \pi))$. This will happen for one such linear form if and only if f is close to being expressible as the LTF $h(\pi, x) = \mathrm{sgn}(w \cdot \pi + \ell - \theta)$.

This completes the sketch of the testing algorithm, modulo the explanation of how the tester can get around "knowing" what the set J is. Looking carefully at what the tester needs to do with J, it turns out that it suffices for it to be able to query f on random strings and correlated tuples of strings, subject to given restrictions π to J. This can be done essentially by borrowing a technique from the paper [FKR+02] (see the discussion after Theorem 42 in Section 6.4.2. of the full paper).

In Section 6 of the full version of the paper, we make all these ideas precise and prove the following, which is our main result [MORS07] :

THEOREM 4.2. *There is an algorithm* **Test-LTF** *for testing whether an arbitrary black-box* $f : \{-1,1\}^n \to \{-1,1\}$ *is an LTF versus* ϵ*-far from any LTF. The algo-*

rithm has two-sided error and makes at most $\mathrm{poly}(1/\epsilon)$ *queries to* f.

We remark that the algorithm described above is adaptive, but using the same techniques found in [FKR+02], the algorithm can be made nonadaptive with a polynomial factor increase in the query complexity.

5 Conclusion

As mentioned in the introduction, the techniques of this paper were instrumental in the recent algorithm of [OS08] for the "Chow Parameters Problem" (i.e. the problem of efficiently learning an unknown LTF given only its degree-0 and degree-1 Fourier coefficients). A goal for future work is to see whether our techniques can be applied to other interesting and related problems. One such problem, suggested to us by R. Santhanam [San08], is the following: give an efficient *deterministic* algorithm which, given an explicit weights-based representation of an LTF $f(x) = \mathrm{sgn}(w_1x_1 + \cdots + w_nx_n - \theta)$, outputs an estimate of $\mathbf{E}[f]$ (i.e. the fraction of inputs that satisfy f) to within an additive $\pm\epsilon$. This is arguably the simplest explicit open derandomization problem of which we are aware. The work of [Ser06] gives a deterministic algorithm which runs in time polynomial in n but exponential in $1/\epsilon$. Can the insights into LTFs from the current paper be used to obtain a truly polynomial algorithm?

References

[BDJ+98] A. Birkendorf, E. Dichterman, J. Jackson, N. Klasner, and H.U. Simon. On restricted-focus-of-attention learnability of Boolean functions. *Machine Learning*, 30:89–123, 1998.

[Blo62] H. Block. The Perceptron: a model for brain functioning. *Reviews of Modern Physics*, 34:123–135, 1962.

[Cho61] C.K. Chow. On the characterization of threshold functions. In *Proceedings of the Symposium on Switching Circuit Theory and Logical Design (FOCS)*, pages 34–38, 1961.

[DLM+07] I. Diakonikolas, H. Lee, K. Matulef, K. Onak, R. Rubinfeld, R. Servedio, and A. Wan. Testing for concise representations. In *Proc. 48th Ann. Symposium on Computer Science (FOCS)*, pages 549–558, 2007.

[FKR+02] E. Fischer, G. Kindler, D. Ron, S. Safra, and A. Samorodnitsky. Testing juntas. In *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science*, pages 103–112, 2002.

[GGR98] O. Goldreich, S. Goldwaser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45:653–750, 1998.

[Gol06] P. Goldberg. A Bound on the Precision Required to Estimate a Boolean Perceptron from its Average Satisfying Assignment. *SIAM Journal on Discrete Mathematics*, 20:328–343, 2006.

[HMP+93] A. Hajnal, W. Maass, P. Pudlak, M. Szegedy, and G. Turan. Threshold circuits of bounded depth. *Journal of Computer and System Sciences*, 46:129–154, 1993.

[KKMO07] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for Max-Cut and other 2-variable CSPs? *SIAM Journal on Computing*, 37(1):319–357, 2007.

[KMT93] S. Kulkarni, S. Mitter, and J. Tsitsiklis. Active learning using arbitrary binary valued queries. *Machine Learning*, 11:23–35, 1993.

[MORS07] K. Matulef, R. O'Donnell, R. Rubinfeld, and R. Servedio. Testing Halfspaces. Technical Report 128, Electronic Colloquium in Computational Complexity, 2007.

[MP68] M. Minsky and S. Papert. *Perceptrons: an introduction to computational geometry*. MIT Press, Cambridge, MA, 1968.

[Nov62] A. Novikoff. On convergence proofs on perceptrons. In *Proceedings of the Symposium on Mathematical Theory of Automata*, volume XII, pages 615–622, 1962.

[OS08] R. O'Donnell and R. Servedio. The Chow Parameters Problem. To appear in Proceedings of the 40th Annual Symposium on Theory of Computing (STOC), 2008.

[PRS02] M. Parnas, D. Ron, and A. Samorodnitsky. Testing basic boolean formulae. *SIAM J. Disc. Math.*, 16:20–46, 2002.

[Ron07] D. Ron. Property testing: A learning theory perspective. COLT 2007 Invited Talk, slides available at http://www.eng.tau.ac.il/ danar/Public-ppt/colt07.ppt, 2007.

[San08] R. Santhanam. . Personal communication, 2008.

[Ser06] R. Servedio. Every linear threshold function has a low-weight approximator. In *Proceedings of the 21st Conference on Computational Complexity (CCC)*, pages 18–30, 2006.

[Ser07] R. Servedio. Every linear threshold function has a low-weight approximator. *Computational Complexity*, 16(2):180–209, 2007.

[STC00] J. Shawe-Taylor and N. Cristianini. *An introduction to support vector machines*. Cambridge University Press, 2000.

[Yao90] A. Yao. On ACC and threshold circuits. In *Proceedings of the Thirty-First Annual Symposium on Foundations of Computer Science*, pages 619–627, 1990.