Electronic Colloquium on Computational Complexity, Report No. 128 (2007)

Testing Halfspaces

Kevin Matulef*
MIT
matulef@mit.edu

Ryan O'Donnell
Carnegie Mellon University
odonnell@cs.cmu.edu

Ronitt Rubinfeld†
MIT
ronitt@theory.csail.mit.edu

Rocco A. Servedio‡
Columbia University
rocco@cs.columbia.edu

November 9, 2007

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].

*Supported in part by an NSF graduate fellowship.
†Supported in part by NSF grant 012702-001.
‡Supported in part by NSF CAREER award CCF-0347282, by NSF award CCF-0523664, and by a Sloan Foundation fellowship.

ISSN 1433-8092

Contents

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 are also known as *threshold functions* or *linear threshold functions*; for brevity we shall often refer to them in this paper as LTFs. LTFs are a simple yet powerful class of functions, which for decades have played an important role in fields such as complexity theory, optimization, and machine learning (see e.g. [HMP+93, Yao90, Blo62, Nov62, MP68, STC00]).

In this work, we focus on the *halfspace testing* problem: given query access to a function, we would like to distinguish whether it is an LTF or whether it is ϵ-far from any LTF. This is in contrast to the proper *halfspace learning* problem: given examples labeled according to an unknown LTF (either random examples or queries to the function), find an LTF that it is ϵ-close to. Though any proper learning algorithm can be used as a testing algorithm (see, e.g., the observations of [GGR98]), testing potentially requires fewer queries. Indeed, in situations where query access is available, a query-efficient testing algorithm can be used to check whether a function is close to a halfspace, before bothering to run a more intensive algorithm to learn which halfspace it is close to.

Our main result is to show that the halfspace testing problem can be solved with a number of queries that is *independent* of n. In doing so, we establish new structural results about LTFs which essentially characterize LTFs in terms of their degree-0 and degree-1 Fourier coefficients.

We note that 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 under the uniform distribution on $\{-1,1\}^n$ (this follows easily from, e.g., the results of [KMT93]). Thus the complexity of learning is linear in n, as opposed to our testing bounds which are independent of n.

We start by describing our testing results in more detail.

Our Results. We consider 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 have the property and with high probability fail all functions that have distance at least ϵ from any function with the property. 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 $\mathrm{poly}(\frac{1}{\epsilon})$ 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 $\mathrm{poly}(\frac{1}{\epsilon})$ query algorithm for testing LTFs with two-sided error.

Results 1 and 2 show that in two natural settings we can test a highly geometric property — whether or not the -1 and $+1$ values defined by f are linearly separable — with a number of queries that is independent of the dimension of the space. Moreover, the dependence on $\frac{1}{\epsilon}$ is only polynomial, rather than exponential or tower-type as in some other property testing algorithms.

While it is slightly unusual to consider property testing under the standard multivariate Gaussian distribution, we remark that 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.

We 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 $\{x_1, \ldots, x_n\}$).

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], 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[1] 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 square of the "Gaussian isoperimetric" function.

Conversely, Theorem 20 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 4.

In Section 5 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 24 and 25, which give Boolean-cube analogues of our characterization of Gaussian LTFs. Theorem 24 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 25 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

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

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. 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 Theorems 37 and 38, which generalize Theorems 24 and 25 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.)

Related Work. Various classes of Boolean functions have recently been studied from a testing perspective. [PRS02] shows how to test dictator functions, monomials, and $O(1)$-term monotone DNFs with query complexity $O(\frac{1}{\epsilon})$. [FKR+02] gave algorithms for testing k-juntas with query complexities that are low-order polynomials in k and $1/\epsilon$. On the other hand, [FLN+02] showed that any algorithm for testing monotonicity must have a query complexity which increases with n. See also [AKK+03, BLR93, GGL+00] and references therein for other work on testing various classes of Boolean functions.

In [DLM+07] a general method is given for testing functions that have concise representations in various formats; among other things this work shows that the class of *decision lists* (a subclass of LTFs) is testable using poly$(\frac{1}{\epsilon})$ queries. The method of [DLM+07] does not apply to LTFs in general since it requires that the functions in question be "well approximated" by juntas, which clearly does not hold for general LTFs.

Outline of the Paper. In Section 2 we give some notation and preliminary facts used throughout the paper. In Section 3 we describe a subroutine for estimating sums of powers of Fourier and Hermite coefficients, based on the notion of Noise Stability. Section 4 contains our algorithm for testing general LTFs over Gaussian Space. Section 5 contains an algorithm for testing *balanced, regular* LTFs over $\{-1,1\}^n$, a "warm-up" to our main result. Finally, Section 6 contains our main result, a general algorithm for testing LTFs over $\{-1,1\}^n$

2 Notation and Preliminaries.

Except in Section 4, throughout this paper f will denote a function from $\{-1,1\}^n$ to $\{-1,1\}$ (in Section 4 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.

We make extensive use of Fourier analysis of functions $f : \{-1,1\}^n \to \{-1,1\}$ and Hermite analysis of functions $f : \mathbf{R}^n \to \{-1,1\}$. In this section we summarize some facts we will need regarding Fourier analysis of functions $f : \{-1,1\}^n \to \{-1,1\}$ and Hermite analysis of functions $f : \mathbf{R}^n \to \{-1,1\}$. For more information on Fourier analysis see, e.g., [Šte00]; for more information on Hermite analysis see, e.g., [LT91].

Fourier analysis. Here we consider functions $f : \{-1,1\}^n \to \mathbf{R}$, and we think of the inputs x to f as being distributed according to the uniform probability distribution. The set of such functions forms a 2^n-dimensional inner product space with inner product given by $\langle f, g\rangle = \mathbf{E}_x[f(x)g(x)]$. The set of functions $(\chi_S)_{S\subseteq[n]}$ defined by $\chi_S(x) = \prod_{i\in S} x_i$ forms a complete orthonormal basis for this space. We will also often write simply x_S for $\prod_{i\in S} x_i$. Given a function $f : \{-1,1\}^n \to \mathbf{R}$ we define its *Fourier coefficients* by $\hat{f}(S) = \mathbf{E}_x[f(x)x_S]$, and we have that $f(x) = \sum_S \hat{f}(S)x_S$. We will be particularly interested in f's *degree*-1 coefficients, i.e., $\hat{f}(S)$ for $|S| = 1$; we will write these as $\hat{f}(i)$ rather than $\hat{f}(\{i\})$. Finally, we have *Plancherel's identity* $\langle f, g\rangle = \sum_S \hat{f}(S)\hat{g}(S)$, which has as a special case *Parseval's identity*, $\mathbf{E}_x[f(x)^2] = \sum_S \hat{f}(S)^2$. From this it follows that for every $f : \{-1,1\}^n \to \{-1,1\}$ we have $\sum_S \hat{f}(S)^2 = 1$.

Hermite analysis. Here we consider functions $f : \mathbf{R}^n \to \mathbf{R}$, and we think of the inputs x to f as being distributed according to the standard n-dimensional Gaussian probability distribution. We treat the set of square-integrable functions as an inner product space with inner product $\langle f, g\rangle = \mathbf{E}_x[f(x)g(x)]$ as before. In the case $n = 1$, there is a sequence of *Hermite polynomials* $h_0 \equiv 1, h_1(x) = x, h_2(x) = (x^2 - 1)/\sqrt{2}, \ldots$ that form a complete orthonormal basis for the space; they can be defined via $\exp(\lambda x - \lambda^2/2) = \sum_{d=0}^{\infty}(\lambda^d/\sqrt{d!})h_d(x)$. In the case of general n, given $S \in \mathbf{N}^n$, we have that the collection of n-variate polynomials $H_S(x) := \prod_{i=1}^n h_{S_i}(x_i)$ forms a complete orthonormal basis for the space. Given a square-integrable function $f : \mathbf{R}^n \to \mathbf{R}$ we define its *Hermite coefficients* by $\hat{f}(S) = \langle f, H_S\rangle$ for $S \in \mathbf{N}^n$ and we have that $f(x) = \sum_S \hat{f}(S)H_S(x)$ (the equality holding in L^2). Again, we will be particularly interested in f's "degree-1" coefficients, i.e., $\hat{f}(e_i)$, where e_i is the vector which is 1 in the ith coordinate and 0 elsewhere. Recall that this is simply $\mathbf{E}_x[f(x)x_i]$. Plancherel and Parseval's identities also hold in this setting.

We will also use the following definitions:

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

Definition 2. *We say that* $f : \{-1,1\}^n \to \{-1,1\}$ *is "*τ*-regular" if* $|\hat{f}(i)| \leq \tau$ *for all* $i \in [n]$.

Definition 3. *A function* $f : \{-1,1\}^n \to \{-1,1\}$ *is said to be a "junta on* $J \subset [n]$*" if* f *only depends on the coordinates in* J. *Typically we think of* J *as a "small" set in this case.*

Definition 4. *For* $a, b \in \mathbf{R}$ *we write* $a \stackrel{\eta}{\approx} b$ *to indicate that* $|a - b| \leq O(\eta)$.

and the following simple facts:

Fact 5. *Suppose* A *and* B *are nonnegative and* $|A - B| \leq \eta$. *Then* $|\sqrt{A} - \sqrt{B}| \leq \eta/\sqrt{B}$.

Proof. $|\sqrt{A} - \sqrt{B}| = \frac{|A-B|}{\sqrt{A}+\sqrt{B}} \leq \frac{\eta}{\sqrt{B}}$. □

Fact 6. *If* X *is a random variable taking values in the range* $[-1,1]$, *its expectation can be estimated to within an additive* $\pm\epsilon$, *with confidence* $1 - \delta$, *using* $O(\log(1/\delta)/\epsilon^2)$ *queries.*

Proof. This follows from a standard additive Chernoff bound. □

3 Tools for Estimating Sums of Powers of Fourier and Hermite Coefficients

In this section we show how to estimate the sum $\sum_{i=1}^{n} \hat{f}(i)^2$ for functions over a boolean domain, and the sum $\sum_{i=1}^{n} \hat{f}(e_i)^2$ for functions over gaussian space. This subroutine lies at the heart of our testing algorithms. We actually prove a more general theorem, showing how to estimate $\sum_{i=1}^{n} \hat{f}(i)^p$ for any integer $p \geq 2$. Estimating the special case of $\sum_{i=1}^{n} \hat{f}(i)^4$ allows us to distinguish whether a function has a single large $|\hat{f}(i)|$, or whether all $|\hat{f}(i)|$ are small. The main results in this section are Corollary 13 (along with its analogue for Gaussian space, Lemma 16), and Lemma 15.

3.1 Noise Stability.

Definition 7. *(Noise stability for Boolean functions.) Let $f, g : \{-1,1\}^n \to \{-1,1\}$, let $\eta \in [0,1]$, and let (x,y) be a pair of η-correlated random inputs — i.e., x is a uniformly random string and y is formed by setting $y_i = x_i$ with probability η and letting y_i be uniform otherwise, independently for each i. We define*

$$\mathbb{S}_\eta(f,g) = \mathbf{E}[f(x)g(y)].$$

Fact 8. *In the above setting, $\mathbb{S}_\eta(f,g) = \sum_{S\subseteq[n]} \hat{f}(S)\hat{g}(S)\eta^{|S|}$.*

Definition 9. *(Noise stability for Gaussian functions.) Let $f, g : \mathbf{R}^n \to \mathbf{R}$ be in $L^2(\mathbf{R}^n)$ with respect to the Gaussian measure, let $\eta \in [0,1]$, and let (x,y) be a pair of η-correlated n-dimensional Gaussians. I.e., each pair of coordinates (x_i, y_i) is chosen independently as follows: x_i is a standard 1-dimensional Gaussian, and $y_i = \eta x_i + \sqrt{1-\eta^2}\cdot z_i$, where z_i is an independent standard Gaussian. We define*

$$\mathbb{S}_\eta(f,g) = \mathbf{E}[f(x)g(y)].$$

Fact 10. *In the above setting, $\mathbb{S}_\eta(f,g) = \sum_{S\in\mathbf{N}^n} \hat{f}(S)\hat{g}(S)\eta^{|S|}$, where $|S|$ denotes $\sum_{i=1}^{n} S_i$.*

3.2 Estimating sums of powers of Fourier coefficients.

For $x = (x_1, \ldots, x_n)$ and $S \subseteq [n]$ we write x_S for the monomial $\prod_{i\in S} x_i$. The following lemma generalizes Fact 8:

Lemma 11. *Fix $p \geq 2$. Let $f_1, \ldots, f_p$ be p functions $f_i : \{-1,1\}^n \to \{-1,1\}$. Fix any set $T \subseteq [n]$. Let $x^1, \ldots, x^{p-1}$ be independent uniform random strings in $\{-1,1\}^n$ and let y be a random string whose bits are independently chosen with $\Pr[y_i = 1] = \frac{1}{2}$ for $i \notin T$ and $\Pr[y_i = 1] = \frac{1}{2} + \frac{1}{2}\eta$ for $i \in T$. Let $\odot$ denote coordinate-wise multiplication. Then*

$$\mathbf{E}[f_1(x^1)f_2(x^2)\cdots f_{p-1}(x^{p-1})f_p(x^1\odot x^2\odot\cdots\odot x^{p-1}\odot y)] = \sum_{S\subseteq T} \eta^{|S|}\hat{f}_1(S)\hat{f}_2(S)\cdots\hat{f}_p(S).$$

Proof. We have

$$\begin{aligned}
&\mathbf{E}[f_1(x^1)f_2(x^2)\cdots f_{p-1}(x^{p-1})f_p(x^1\odot x^2\odot\cdots\odot x^{p-1}\odot y)]\\
=\ &\mathbf{E}[\sum_{S_1,\ldots,S_p\subseteq[n]} \hat{f}_1(S_1)\cdots\hat{f}_{p-1}(S_{p-1})\hat{f}_p(S_p)\cdot(x^1)_{S_1}\cdots(x^{p-1})_{S_{p-1}}(x^1\odot x^2\odot\cdots\odot x^{p-1}\odot y)_{S_p}]\\
=\ &\sum_{S_1,\ldots,S_p\subseteq[n]} \hat{f}_1(S_1)\cdots\hat{f}_{p-1}(S_{p-1})\hat{f}_p(S_p)\cdot\mathbf{E}[(x^1)_{S_1\Delta S_p}\cdots(x^{p-1})_{S_{p-1}\Delta S_p}y_{S_p}]
\end{aligned}$$

Now recalling that $x^1, \ldots, x^{p-1}$ and y are all independent and the definition of y, we have that the only nonzero terms in the above sum occur when $S_1 = \cdots = S_{p-1} = S_p \subseteq T$; in this case the expectation is $\eta^{|S_p|}$. This proves the lemma. $\square$

Lemma 12. *Let $p \geq 2$. Suppose we have black-box access to $f_1, \dots, f_p : \{-1,1\}^n \to \{-1,1\}$. Then for any $T \subseteq [n]$, we can estimate the sum of products of degree-1 Fourier coefficients*

$$\sum_{i \in T} \hat{f}_1(i) \cdots \hat{f}_p(i)$$

to within an additive η, with confidence $1-\delta$, using $O(p \cdot \log(1/\delta)/\eta^4)$ queries.

Proof. Let $x^1, \dots, x^p$ be independent uniform random strings in $\{-1,1\}^n$ and let y be as in the previous lemma. Empirically estimate

$$\mathbf{E}[f_1(x^1)f_2(x^2)\cdots f_p(x^p)] \quad \text{and} \quad \mathbf{E}[f_1(x^1)f_2(x^2)\cdots f_{p-1}(x^{p-1})f_p(x^1 \odot x^2 \odot \cdots \odot x^{p-1} \odot y)] \tag{1}$$

to within an additive $\pm\eta^2$, using $O(1/\eta^4)$ samples. By the previous lemma these two quantities are exactly equal to

$$\hat{f}_1(\emptyset) \cdots \hat{f}_p(\emptyset) \qquad \text{and} \qquad \sum_{S \subseteq T} \eta^{|S|} \hat{f}_1(S)\hat{f}_2(S) \cdots \hat{f}_p(S)$$

respectively. Subtracting the former estimate from the latter yields

$$\sum_{|S|>0, S \subseteq T} \eta^{|S|} \hat{f}_1(S) \cdots \hat{f}_p(S)$$

to within an additive $O(\eta^2)$, and this itself is within η^2 of

$$\sum_{|S|=1, S \subseteq T} \eta \hat{f}_1(S) \cdots \hat{f}_p(S)$$

because the difference is

$$\begin{aligned}
\sum_{|S|>1, S \subseteq T} \eta^{|S|} \hat{f}_1(S) \cdots \hat{f}_p(S) &\leq \eta^2 \sum_{|S|>1, S \subseteq T} |\hat{f}_1(S) \cdots \hat{f}_p(S)| \\
&\leq \eta^2 \sqrt{\sum_{|S|>1, S \subseteq T} \hat{f}_1(S)^2} \sqrt{\sum_{|S|>1, S \subseteq T} (\hat{f}_2(S) \cdots \hat{f}_p(S))^2} \qquad (2) \\
&\leq \eta^2 \cdot 1 \cdot \sqrt{\sum_{|S|>1, S \subseteq T} \hat{f}_2(S)^2} \leq \eta^2 \qquad (3)
\end{aligned}$$

where (2) is Cauchy-Schwarz and (3) uses the fact that the sum of the squares of the Fourier coefficients of a Boolean function is at most 1. Thus we have $\eta \cdot \sum_{i \in T} \hat{f}_1(i) \cdots \hat{f}_p(i)$ to within an additive $O(\eta^2)$; dividing by η gives us the required estimate within $O(\eta)$. $\square$

Taking all f_i's to be the same function f, we have

Corollary 13. *Fix $p \geq 2$ and fix any $T \subseteq [n]$. Given black-box access to $f : \{-1,1\}^n \to \{-1,1\}$, we can estimate $\sum_{i \in T} \hat{f}(i)^p$ to an additive $\pm\eta$, with confidence $1-\delta$, using $O(p \cdot \log(1/\delta)/\eta^4)$ queries.*

Proposition 14. *If every $i \in T$ has $|\hat{f}(i)| < \alpha$, then $\sum_{i \in T} \hat{f}(i)^4 < \alpha^2 \sum_{i \in T} \hat{f}(i)^2 \leq \alpha^2$.*

Lemma 15. *Fix any $T \subseteq [n]$. There is an $O(\log(1/\delta)/\tau^{16})$-query test* **Non-Regular**(τ, δ, T) *which, given query access to $f : \{-1,1\}^n \to \{-1,1\}$, behaves as follows: with probability $1-\delta$,*

- *If $|\hat{f}(i)| \geq \tau$ for some $i \in T$ then the test accepts;*
- *If every $i \in T$ has $|\hat{f}(i)| < \tau^2/4$ then the test rejects.*

Proof. The test is to estimate $\sum_{i \in T} \hat{f}(i)^4$ to within an additive $\pm\tau^4/4$ and then accept if and only if the estimate is at most $\tau^4/2$. If $|\hat{f}(i)| \geq \tau$ for some i then clearly $\sum_{i=1}^n \hat{f}(i)^4 \geq \tau^4$ so the test will accept since the estimate will be at least $3\tau^4/4$. On the other hand, if each $i \in T$ has $|\hat{f}(i)| < \tau^2/4$, then $\sum_{i \in T} \hat{f}(i)^4 < \tau^4/16$ by Proposition 14 and so the test will reject since the estimate will be less than $5\tau^4/16$. $\square$

3.3 Estimating sums of powers of Hermite coefficients.

Here we let $\hat{f}(e_i)$ denote the i-th degree-1 Hermite coefficient of $f : \mathbf{R}^n \to \mathbf{R}$ as described in Section 4.

For the Gaussian distribution we require only the following lemma, which can be proved in a straightforward way following the arguments in Section 3.2 and using Fact 10.

Lemma 16. *Given black-box access to $f : \mathbf{R}^n \to \{-1, 1\}$, we can estimate $\sum_{i=1}^n \hat{f}(e_i)^2$ to within an additive η, with confidence $1 - \delta$, using $O(\log(1/\delta)/\eta^4)$ queries.*

4 A Tester for General LTFs over $\mathbb{R}^n$

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 17. *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 18. 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 $.96788\cdots < 1$. Finally, the last statement follows ultimately from the fact that $1 - \mu(\theta) \sim 2\phi(\theta)/|\theta|$ for $|\theta| \ge 1$. $\square$

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 19. *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:*

$$1.\ \hat{f}(0) = \mathbf{E}[f] = \mu(\theta). \qquad 2.\ \hat{f}(e_i) = \sqrt{W(\mathbf{E}[f])}w_i. \qquad 3.\ \sum_{i=1}^{n} \hat{f}(e_i)^2 = W(\mathbf{E}[f]).$$

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

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

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. The second step of estimating $\sum_{i=1}^n \hat{f}(e_i)^2$ was described in section 3.

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 18 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 19 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 20. *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 18 (item 1), we have

$$\mathbf{E}[|h|] = 2\phi(t) - t\mu(t) = 2\phi(\mu^{-1}(\mathbf{E}[f])) - t\,\mathbf{E}[f] = \sqrt{W(\mathbf{E}[f])} - t\,\mathbf{E}[f], \quad \text{and thus}$$
$$\mathbf{E}[h(\mathrm{sgn}(h) - f)] = \mathbf{E}[|h| - fh] = \sqrt{W(\mathbf{E}[f])} - \sigma \le \frac{4\epsilon^3}{\sqrt{W(\mathbf{E}[f])}} \le C\epsilon^2,$$

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) and the second follows from the assumption that $|\mathbf{E}[f]| \le 1 - \epsilon$, which by Proposition 18 (item 3) implies that $\sqrt{W(\mathbf{E}[f])} \ge \Omega(\epsilon)$.

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

5 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 4, 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 halfspace and $\mathrm{sgn}(x_1)$ as an example of a halfspace 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 6, and here present a tester for *balanced, regular* LTFs.

Definition 21. *We say that any function $f : \{-1,1\}^n \to \{-1,1\}$ is "τ-regular" if $|\hat{f}(i)| \le \tau$ for all $i \in [n]$.*

Definition 22. *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 6, 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 23. *Fix any $\tau > 0$. There is an $O(1/\tau^8)$-query algorithm A that satisfies the following property: Let ϵ be any value $\epsilon \ge C\tau^{1/6}$, 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 23 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 halfspace. 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 5.1) 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 4 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, see Corollary 13 in particular.

In Section 5.1, 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 5.2 we prove correctness of the test.

5.1 Two 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 19 item 3 and Theorem 20.

Theorem 24. *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 25. *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 $\gamma \le 1/3$. Let $L := \sqrt{\sum_{i=1}^n \hat{f}(i)^2}$; note that by our assumption on γ we have $L \ge \frac{1}{2}$. We have:

$$\begin{aligned} (2/\pi) - \gamma &\le \sum_{i=1}^n \hat{f}(i)^2 = \mathbf{E}[f\ell] \le \mathbf{E}[|\ell|] &(4)\\ &\le \sqrt{2/\pi} \cdot L + O(\tau) &(5)\\ &\le \sqrt{2/\pi}\sqrt{2/\pi + \gamma} + O(\tau) \le (2/\pi) + O(\gamma) + O(\tau). \end{aligned}$$

The equality in (4) is Plancherel's identity, and the latter inequality is because f is a ± 1-valued function. The inequality (5) 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 Appendix A for the precise justification. Comparing the overall left- and right-hand sides, we conclude that $\mathbf{E}[|\ell|] - \mathbf{E}[f\ell] \le O(\gamma) + O(\tau)$.

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 $O(\tau/L) \le O(2\tau)$; we can assume this error is at most $\epsilon/4$, since if $\epsilon \le O(\tau)$ then the theorem already holds. Hence we have (see Theorem 55 for precise justification)

$$\Pr[|\ell| \le \epsilon/8] = \Pr[|\ell/L| \le \epsilon/8L] \le \Pr[|N(0,1)| \le \epsilon/8L] + \epsilon/4 \le \epsilon/8L + \epsilon/4 \le \epsilon/2,$$

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

5.2 Proving correctness of the test.

First observe that for any Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, if $|\hat{f}(i)| \le \tau$ for all i then $\sum_{i \in T} \hat{f}(i)^4 \le \tau^2 \sum_{i \in T} \hat{f}(i)^2 \le \tau^2$, using Parseval. On the other hand, if $|\hat{f}(i)| \ge 2\tau^{1/2}$ 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 24 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 $O(\tau^{2/3}) \le C_1\tau^{1/3}$ 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 25, 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 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 $O(1/\tau^8)$. This completes the proof of Theorem 23. $\square$

6 A Tester for General LTFs over $\{-1,1\}^n$

In this section we give our main result, a constant-query tester for general halfspaces over $\{-1,1\}^n$. We start with a very high-level overview of our approach.

As we saw in Section 5, 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 4. 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 25 (see Theorem 38 in Section 6.2), 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 24, the special case of mean-zero, we appealed 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). However generalizing the first result to unbalanced LTFs is quite complicated, and requires the following theorem, which we prove in Section 6.1:[2]

Theorem 26. *Let $f(x) = \mathrm{sgn}(w_1 x_1 + \cdots + w_n x_n - \theta)$ 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 26 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.

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

[2]Readers familiar with the notion of influence (Definition 60) will recall that for any LTF f we have $\mathrm{Inf}_i(f) = |\hat{f}(i)|$ for each i. Thus Theorem 26 may roughly be viewed as saying that "every not-too-biased LTF with a large weight has an influential variable."

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).

In the remainder of this section we make all these ideas precise and prove the following, which is our main result:

Theorem 27. *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 algorithm has two-sided error and makes at most $\mathrm{poly}(1/\epsilon)$ queries to f.*

Remark 28. *The algorithm described above is adaptive. We note that similar to [FKR+02], the algorithm can be made nonadaptive with a polynomial factor increase in the query complexity (see Remark 44 in Section 6.4.2).*

Section 6.1 gives the proof of Theorem 26. Section 6.2 gives two theorems essentially characterizing LTFs; these theorems are the main tools in proving the correctness of our test. Section 6.3 gives an overview of the algorithm, which is presented in Sections 6.4 and 6.5. Section 6.6 proves correctness of the test.

6.1 On the structure of LTFs: relating weights, influences and biases

In this section we prove a structural theorem about LTFs. The theorem says that an LTF's most influential variable has influence at least polynomial in the size of the LTF's largest weight and the size of the LTF's bias.

Theorem 26. *Let $f(x) = \mathrm{sgn}(\sum_{i=1}^n a_i x_i - \theta)$ be an LTF such that $\sum_i a_i^2 = 1$ and $\delta \stackrel{def}{=} |a_1| \geq |a_i|$ for all $i \in [n]$. Let $0 \leq \epsilon \leq 1$ be such that $|\mathbf{E}[f]| = 1 - \epsilon$. Then $\mathrm{Inf}_1(f) = \Omega(\delta \epsilon^6 \log(1/\epsilon))$.*

Even the $\theta = 0$ case of the theorem, corresponding to $\epsilon = 1$, is somewhat tricky to prove. It appeared first as Proposition 10.2 of [KKMO07]. A substantially more intricate proof is required for the general statement; indeed, the arguments of [KKMO07] occur in somewhat modified form as Cases 1.a and 1.b of our proof below.

We note that it is easy to give an *upper* bound on $\mathrm{Inf}_1(f)$ in terms of either δ or ϵ: it is immediate that $\mathrm{Inf}_1(f) \leq O(\epsilon)$, and from Proposition 64 we have that $\mathrm{Inf}_1(f) \leq O(\delta)$. We suspect that $\Theta(\delta\epsilon)$ may be the optimal bound for Theorem 26.

6.1.1 Useful tools for proving Theorem 26.

We first observe that

$$\mathrm{Inf}_1(f) = \Pr\big[|a_2x_2 + \cdots + a_nx_n - \theta| \le \delta\big]. \tag{6}$$

We shall prove Theorem 26 by lower bounding the right hand side of (6).

At many points in the proof of Theorem 26 we will use the following fact, which is a simple consequence of "Poincaré's inequality" — i.e., the fact that the sum of a function's influences is at least its variance:

Fact 29. *Let $g : \{-1,1\}^\ell \to \{-1,1\}$ be a linear threshold function $g(x) = \mathrm{sgn}(\sum_{i=1}^{\ell} a_ix_i - \theta)$ with $|a_1| \ge |a_i|$ for all $i = 1,\dots,\ell$. Then $\mathrm{Inf}_1(g) \ge \mathrm{Var}[g]/\ell$.*

Proof. Poincaré's inequality says that $\sum_{i=1}^{\ell} \mathrm{Inf}_i(g) \ge \mathrm{Var}[g]$ for any Boolean function g. Since $|a_1| \ge |a_i|$ for all i (Proposition 61), we have $\mathrm{Inf}_1(g) \ge \mathrm{Inf}_i(g)$, and the fact follows. □

The following easily verified fact is also useful:

Fact 30. *Let $g : \{-1,1\}^\ell \to \{-1,1\}$ be a linear threshold function $g(x) = \mathrm{sgn}(\sum_{i=1}^{\ell} a_ix_i - \theta)$ with $|a_1| > |\theta|$. Then $\mathrm{Var}[g] = \Omega(1)$.*

Proof. Since $|a_1| > |\theta|$, one of the two restrictions obtained by fixing the first variable outputs 1 at least half the time, and the other outputs -1 at least half the time. This implies that $1/4 \le \Pr[g(x) = 1] < 3/4$, which gives $\mathrm{Var}[g] = \Omega(1)$. □

We will also often use the Berry-Esseen theorem, Theorem 55. For definiteness, we will write C for the implicit constant in the $O(\cdot)$ of the statement, and we note that for every interval A we in fact have $|\Pr[\ell(x)/\sigma \in A] - \Pr[X \in A]| \le 2C\tau/\sigma$.

Finally, we will also use the Hoeffding bound:

Theorem 31. *Fix any $0 \ne w \in \mathbf{R}^n$ and write $\|w\|$ for $\sqrt{w_1^2 + \cdots + w_n^2}$. For any $\gamma > 0$, we have*

$$\Pr_{x\in\{-1,1\}^n}[w \cdot x \ge \gamma\|w\|] \le e^{-\gamma^2/2} \quad \textit{and} \quad \Pr_{x\in\{-1,1\}^n}[w \cdot x \le -\gamma\|w\|] \le e^{-\gamma^2/2}.$$

6.1.2 The idea behind Theorem 26.

We give a high-level outline of the proof before delving into the technical details. Here and throughout the proof we suppose for convenience that $\delta = |a_1| \ge |a_2| \ge \cdots \ge |a_n| \ge 0$.

We first consider the case (Case 1) that the biggest weight δ is small relative to ϵ. We show that with probability $\Omega(\epsilon^2)$, the "tail" $a_\beta x_\beta + \cdots + a_nx_n$ of the linear form (for a suitably chosen β) takes a value in $[\theta - 1, \theta + 1]$; this means that the effective threshold for the "head" $a_2x_2 + \cdots + a_{\beta-1}x_{\beta-1}$ is in the range $[-1, 1]$. In this event, a modified version of the [KKMO07] proof shows that the probability that $a_2x_2 + \cdots + a_{\beta-1}x_{\beta-1}$ lies within $\pm\delta$ of the effective threshold is $\Omega(\delta)$; this gives us an overall probability bound of $\Omega(\delta\epsilon^2)$ for (6) in Case 1.

We next consider the case (Case 2) that the biggest weight δ is large. We define the "critical index" of the sequence $a_1,\dots,a_n$ to be the first index $k \in [n]$ at which the Berry-Esseen theorem applied to the sequence $a_k,\dots,a_n$ has a small error term; see Definition 35 below. (This quantity was implicitly defined and used in [Ser07].) We proceed to consider different cases depending on the size of the critical index.

Case 2.a handles the case in which the critical index k is "large" (larger than $\Theta(\log(1/\epsilon)/\epsilon^4)$. Intuitively, in this case the weights $a_1,\dots,a_k$ decrease exponentially and the value $\sum_{j\ge k'} a_j^2$ is very small, where $k' = \Theta(\log(1/\epsilon)/\epsilon^4)$. The rough idea in this case is that the effective number of relevant variables is at most k', so we can use Fact 29 to get a lower bound on Inf_1. (There are various subcases here for technical reasons but this is the main idea behind all of them.)

Case 2.b handles the case in which the critical index k is "small" (smaller than $\Theta(\log(1/\epsilon)/\epsilon^4)$). Intuitively, in this case the value $\sigma_k \stackrel{\text{def}}{=} \sqrt{\sum_{j \geq k} a_j^2}$ is large, so the random variable $a_k x_k + \cdots + a_n x_n$ behaves like a Gaussian random variable $N(0, \sigma_k)$ (recall that since k is the critical index, the Berry-Esseen error is "small"). Now there are several different subcases depending on the relative sizes of σ_k and θ, and on the relative sizes of δ and θ. In some of these cases we argue that "many" restrictions to the tail variables $x_k, \ldots, x_n$ yield a resulting LTF which has "large" variance; in these cases we can use Fact 29 to argue that for any such restriction the influence of x_1 is large, so the overall influence of x_1 cannot be too small. In the other cases we use the Berry-Esseen theorem to approximate the random variable $a_k x_k + \cdots + a_n x_n$ by a Gaussian $N(0, \sigma_k)$, and use properties of the Gaussian to argue that the analogue to expression (6) (with a Gaussian in place of $a_k x_k + \cdots + a_n x_n$) is not too small.

6.1.3 The detailed proof of Theorem 26.

We suppose without loss of generality that $\mathbf{E}[f] = -1 + \epsilon$, i.e. that $\theta \geq 0$. We have the following two useful facts:

Fact 32. *We have* $0 \leq \theta \leq \sqrt{2 \ln(2/\epsilon)}$.

Proof. The lower bound is by assumption, and the upper bound follows from the Hoeffding bound and the fact that $\mathbf{E}[f] = -1 + \epsilon$. □

Fact 33. *Let S be any subset of variables $x_1, \ldots, x_n$. For at least an $\epsilon/4$ fraction of restrictions ρ that fix the variables in S and leave other variables free, we have $\mathbf{E}[f_\rho] \geq -1 + \epsilon/4$.*

Proof. If this were not the case then we would have $\mathbf{E}[f] < (\epsilon/4) \cdot 1 + (1 - \epsilon/4)(-1 + \epsilon/4) < -1 + \epsilon$, which contradicts the fact that $\mathbf{E}[f] = -1 + \epsilon$. □

Now we consider the cases outlined in the previous subsection. Recall that C is the absolute constant in the Berry-Esseen theorem; we shall suppose w.l.o.g. that C is a positive integer. Let $C_1 > 0$ be a suitably large (relative to C) absolute constant to be chosen later.

Case 1: $\delta \leq \epsilon^2/C_1$. We will show that in Case 1 we actually have $\mathrm{Inf}_1(f) = \Omega(\delta \epsilon^2)$.

Let us define $T \stackrel{\text{def}}{=} \{\beta, \ldots, n\}$ where $\beta \in [n]$ is the last value such that $\sum_{i=\beta}^n a_i^2 \geq \frac{1}{2}$. Since each $|a_i|$ is at most $\epsilon^2/C_1 \leq 1/C_1$ (because we are in Case 1), we certainly have that $\sum_{i \in T} a_i^2 \in [\frac{1}{2}, \frac{3}{4}]$ by choosing C_1 suitably large.

We first show that the tail sum $\sum_{i \in T} a_i x_i$ lands in the interval $[\theta - 1, \theta + 1]$ with fairly high probability:

Lemma 34. *We have*

$$\Pr\left[\sum_{i \in T} a_i x_i \in [\theta - 1, \theta + 1]\right] \geq \epsilon^2/18.$$

Proof. Let σ_T denote $\left(\sum_{i \in T} a_i^2\right)^{1/2}$. As noted above we have $\sqrt{4/3} \leq \sigma_T^{-1} \leq \sqrt{2}$. We thus have

$$\begin{aligned}
\Pr\left[\sum_{i \in T} a_i x_i \in [\theta - 1, \theta + 1]\right] &= \Pr\left[\sigma_T^{-1} \sum_{i \in T} a_i x_i \in \sigma_T^{-1}[\theta - 1, \theta + 1]\right] \\
&\geq \Phi([\sigma_T^{-1}\theta - \sigma_T^{-1}, \sigma_T^{-1}\theta + \sigma_T^{-1}]) - 2C\delta\sigma_T^{-1} \qquad (7) \\
&> \Phi([\sigma_T^{-1}\theta - \sigma_T^{-1}, \sigma_T^{-1}\theta + \sigma_T^{-1}]) - 2\sqrt{2}C\delta \qquad (8)
\end{aligned}$$

where (7) follows from the Berry-Esseen theorem using the fact that each $|a_i| \le \delta$.

If $0 \le \theta \le 1$, then clearly the interval $[\sigma_T^{-1}\theta - \sigma_T^{-1}, \sigma_T^{-1}\theta + \sigma_T^{-1}]$ contains the interval $[0, 1]$. Since $\Phi([0, 1]) \ge \frac{1}{3}$, the bound $\delta \le \epsilon^2/C_1$ easily gives that (8) is at least $\epsilon^2/18$ as required, for a suitably large choice of C_1.

If $\theta > 1$, then using our bounds on σ_T^{-1} we have that

$$
\begin{aligned}
\Phi([\sigma_T^{-1}\theta - \sigma_T^{-1}, \sigma_T^{-1}\theta + \sigma_T^{-1}]) &\ge \Phi([\sqrt{2}\cdot\theta - \sqrt{4/3}, \sqrt{2}\cdot\theta + \sqrt{4/3}) \\
&> \Phi([\sqrt{2}\cdot\theta - \sqrt{4/3}, \sqrt{2}\cdot\theta]) \\
&> \sqrt{4/3}\cdot\phi(\sqrt{2}\cdot\theta) \\
&\ge \sqrt{4/3}\cdot\phi(2\sqrt{\ln(2/\epsilon)}) \qquad (9) \\
&= \sqrt{\frac{4}{3}}\cdot\frac{1}{\sqrt{2\pi}}\cdot\frac{\epsilon^2}{4} > \frac{\epsilon^2}{9}. \qquad (10)
\end{aligned}
$$

Here (9) follows from Fact 32 and (10) follows from definition of $\phi(\cdot)$. Since $\delta \le \epsilon^2/C_1$, again with a suitably large choice of C_1 we easily have $2\sqrt{2}C\delta \le \epsilon^2/18$, and thus (8) is at least $\epsilon^2/18$ as required and the lemma is proved. □

Now consider any fixed setting of $x_\beta, \ldots, x_n$ such that the tail $\sum_{i\in T} a_i x_i$ comes out in the interval $[\theta - 1, \theta + 1]$, say $\sum_{i\in T} a_i x_i = \theta - \tau$ where $|\tau| \le 1$. We show that the head $a_2x_2 + \cdots + a_{\beta-1}x_{\beta-1}$ lies in $[\tau - \delta, \tau + \delta]$ with probability $\Omega(\delta)$; with Lemma 34, this implies that the overall probability (6) is $\Omega(\delta\epsilon^2)$.

Let $\alpha \stackrel{\text{def}}{=} C_1^2/8$, let $S \stackrel{\text{def}}{=} \{\alpha, \ldots, \beta - 1\}$, and let $R \stackrel{\text{def}}{=} \{2, \ldots, \alpha - 1\}$. Since $\delta \le \epsilon^2/C_1$, we have that $\sum_{i=1}^{\alpha-1} a_i^2 \le 1/8$, so consequently $1/8 \le \sum_{i\in S} a_i^2 \le 1/2$. Letting σ_S denote $(\sum_{i\in S} a_i^2)^{1/2}$, we have $\sqrt{2} \le \sigma_S^{-1} \le 2\sqrt{2}$.

We now consider two cases depending on the magnitude of a_α. Let $C_2 \stackrel{\text{def}}{=} C_1/4$.

Case 1.a: $|a_\alpha| \le \delta/C_2$. In this case we use the Berry-Esseen theorem on S to obtain

$$
\begin{aligned}
\Pr\left[\sum_{i\in S} a_i x_i \in [\tau - \delta, \tau + \delta]\right] &= \Pr\left[\sigma_S^{-1}\sum_{i\in S} a_i x_i \in \sigma_S^{-1}[\tau - \delta, \tau + \delta]\right] \\
&\ge \Phi([\sigma_S^{-1}\tau - \sigma_S^{-1}\delta, \sigma_S^{-1}\tau + \sigma_S^{-1}\delta]) - 2C(\delta/C_2)\sigma_S^{-1}. \qquad (11)
\end{aligned}
$$

Using our bounds on τ and σ_S^{-1}, we have that the $\Phi(\cdot)$ term of (11) is at least $(\sqrt{2}\delta)\cdot\phi(2\sqrt{2}) > \delta/100$. Since the error term $2C(\delta/C_2)\sigma_S^{-1}$ is at most $\delta/200$ for a suitably large choice of C_1 relative to C (recall that $C_2 = C_1/4$), we have (11) $\ge \delta/200$. Now for any setting of $x_\alpha, \ldots, x_{\beta-1}$ such that $\sum_{i\in S} a_i x_i$ lies in $[\tau - \delta, \tau + \delta]$, since each of $|a_2|, \ldots, |a_{\alpha-1}|$ is at most δ there is (at least one) corresponding setting of $x_2, \ldots, x_{\alpha-1}$ such that $\sum_{i\in(R\cup S)} a_i x_i$ also lies in $[\tau - \delta, \tau + \delta]$. (Intuitively, one can think of successively setting each bit $x_{\alpha-1}, x_{\alpha-2}, \ldots, x_j, \ldots, x_2$ in such a way as to always keep $\sum_{i=j}^{\beta-1} a_i x_i$ in $[\tau-\delta, \tau+\delta]$). So the overall probability that $a_2x_2 + \cdots + a_{\beta-1}x_{\beta-1}$ lies in $[\tau - \delta, \tau + \delta]$ is at least $(\delta/200)\cdot 2^{-\alpha+2} = \Omega(\delta)$, and we are done with Case 1.a.

Case 1.b: $a_\alpha > \delta/C_2$. Similar to Case 2 of [KKMO07], we again use the Berry-Esseen theorem on S, now using the bound that $|a_i| \le \delta$ for each $i \in S$ and bounding the probability of a larger interval $[\tau - C_2\delta, \tau +$

$C_2\delta]$:

$$\begin{aligned}
&\Pr\left[\sum_{i\in S} a_i x_i \in [\tau - C_2\delta, \tau + C_2\delta]\right] \\
&\quad = \Pr\left[\sigma_S^{-1}\sum_{i\in S} a_i x_i \in \sigma_S^{-1}[\tau - C_2\delta, \tau + C_2\delta]\right] \\
&\quad \geq \Phi([\sigma_S^{-1}\tau - \sigma_S^{-1}C_2\delta, \sigma_S^{-1}\tau + \sigma_S^{-1}C_2\delta]) - 2C\delta\sigma_S^{-1} \qquad (12) \\
&\quad \geq \Phi([2\sqrt{2} - \sqrt{2}C_2\delta, 2\sqrt{2}]) - 4\sqrt{2}C\delta \qquad (13)
\end{aligned}$$

In (12) we have used the Berry-Esseen theorem and in (13) we have used our bounds on σ_S^{-1} and τ. Now recalling that $\delta \leq \epsilon^2/C_1 \leq 1/C_1$ and $C_2 = C_1/4$, we have $\sqrt{2}C_2\delta < 2\sqrt{2}$, and hence

$$(13) \quad \geq \quad \sqrt{2}C_2\delta \cdot \phi(2\sqrt{2}) - 4\sqrt{2}C\delta > C\delta \qquad (14)$$

where the second inequality follows by choosing C_1 (and hence C_2) to be a sufficiently large constant multiple of C. Now for any setting of $x_\alpha, \dots, x_{\beta-1}$ such that $\sum_{i\in S} a_i x_i = t$ lies in $[\tau - C_2\delta, \tau + C_2\delta]$, since $\delta/C_2 \leq |a_2|, \dots, |a_{\alpha-1}| \leq \delta$, there is at least one setting of the bits $x_2, \dots, x_{\alpha-1}$ for which $t + \sum_{i=2}^{\alpha-1} a_i x_i$ lies in $[\tau - \delta, \tau + \delta]$. (Since, as is easily verified from the definitions of α and C_2, we have $(\alpha - 2)\delta/C_2 \geq C_2\delta$, the magnitude of $a_2, \dots, a_{\alpha-1}$ is large enough to get from $\tau - C_2\delta$ to τ; and since each $|a_i|$ is at most δ, once the interval $[\tau - \delta, \tau + \delta]$ is reached a suitable choice of signs will keep the sum in the right interval.) So in Case 1.b. the overall probability that $a_2x_2 + \cdots + a_{\beta-1}x_{\beta-1}$ lies in $[\tau - \delta, \tau + \delta]$ is at least $C\delta \cdot 2^{-\alpha+2} = \Omega(\delta)$, and we are done with Case 1.b..

We turn to the remaining case in which δ is "large:"

Case 2: $\delta > \epsilon^2/C_1$. Let us introduce the following definition which is implicit in [Ser07]:

Definition 35. *Let $a_1, \dots, a_n$ be a sequence of values such that $|a_1| \geq \cdots \geq |a_n| \geq 0$. The* critical index *of the sequence is the smallest value of $k \in [n]$ such that*

$$\frac{C|a_k|}{\sqrt{\sum_{j=k}^n a_j^2}} \leq C_3\delta\epsilon^2. \qquad (15)$$

Here $C_3 > 0$ is a (suitably small) absolute constant specified below. (Note that the LHS value $C|a_k|/\sqrt{\sum_{j=k}^n a_j^2}$ is an upper bound on the Berry-Esseen error when the theorem is applied to $a_kx_k + \cdots + a_nx_n$.)

Throughout the rest of the proof we write k to denote the critical index of $a_1, \dots, a_n$. Observe that $k > 1$ since we have

$$\frac{C|a_1|}{\sqrt{\sum_{j=1}^n a_j^2}} = C\delta > \frac{C\epsilon^2}{C_1} \geq \frac{C\delta\epsilon^2}{C_1} > C_3\delta\epsilon^2$$

where the final bound holds for a suitably small constant choice of C_3.

We first consider the case that the critical index k is large. In the following $C_4 > 0$ denotes a suitably large absolute constant.

Case 2.a: $k > C_4 \ln(1/\epsilon)/\epsilon^4 + 1$. In this case we define $k' \stackrel{\text{def}}{=} \lceil C_4 \ln(1/\epsilon)/\epsilon^4 \rceil + 1$. Let us also define $\sigma_{k'} \stackrel{\text{def}}{=} \sqrt{\sum_{j=k'}^n a_j^2}$. The following claim shows that $\sigma_{k'}$ is small:

Claim 36. *We have $\sigma_{k'} \leq \frac{\epsilon^3}{10C_1}$.*

Proof. For $i \in [n]$ let us write A_i to denote $\sum_{j=i}^{n} a_j^2$; note that $A_1 = 1$ and $A_i = a_i^2 + A_{i+1}$. For ease of notation let us write ζ to denote $\delta\epsilon^2 C_3/C$.

Since we are in Case 2.a, for any $1 \leq i < k'$ we have $a_i^2 > \zeta A_i = \zeta a_i^2 + \zeta A_{i+1}$, or equivalently $(1-\zeta)a_i^2 > \zeta A_{i+1}$. Adding $(1-\zeta)A_{i+1}$ to both sides gives $(1-\zeta)(a_i^2 + A_{i+1}) = (1-\zeta)A_i > A_{i+1}$. So consequently we have

$$A_{k'} < (1-\zeta)^{k'-1} \leq (1-\zeta)^{C_4 \ln(1/\epsilon)/\epsilon^4} \leq (1 - \epsilon^4 C_3/(CC_1))^{C_4 \ln(1/\epsilon)/\epsilon^4} \leq \left(\frac{\epsilon^3}{10C_1}\right)^2,$$

where in the third inequality we used $\delta > \epsilon^2/C_1$ (which holds since we are in Case 2) and the fourth inequality holds for a suitable choice of the absolute constant C_4. This proves the claim. $\square$

At this point we know δ is "large" (at least ϵ^2/C_1) and $\sigma_{k'}$ is "small" (at most $\frac{\epsilon^3}{10C_1}$). We consider two cases depending on whether θ is large or small.

Case 2.a.i: $\theta < \epsilon^2/(2C_1)$. In this case we have $0 \leq \theta < \delta/2$. Since $4\sigma_{k'} < \epsilon^2/(2C_1) < \delta/2$, the Hoeffding bound gives that a random restriction that fixes variables $x_{k'}, \ldots, x_n$ gives $|a_{k'}x_{k'} + \cdots + a_n x_n| > 4\sigma_{k'}$ with probability at most $e^{-8} < 1/100$. Consequently we have that for at least $99/100$ of all restrictions ρ to $x_{k'}, \ldots, x_n$, the resulting function f_ρ (on variables $x_1, \ldots, x_{k'-1}$) is $f_\rho(x) = \mathrm{sgn}(a_1 x_1 + \cdots + a_{k'-1}x_{k'-1} - \theta_\rho)$ where $-\delta/2 \leq \theta_\rho < \delta$. Facts 29 and 30 now imply that each such f_ρ has $\mathrm{Inf}_1(f_\rho) = \Omega(1)/k' = \Omega(1) \cdot \epsilon^4/\ln(1/\epsilon)$, so consequently $\mathrm{Inf}_1(f)$ is also $\Omega(1) \cdot \epsilon^4/\log(1/\epsilon)$, which certainly suffices for Theorem 26. This concludes Case 2.a.i.

Case 2.a.ii: $\theta \geq \epsilon^2/(2C_1)$. We now apply the Hoeffding bound (Theorem 31) to $a_{k'}x_{k'} + \cdots + a_n x_n$ with $\gamma = 2\sqrt{\ln(8/\epsilon)}$. This gives that $a_{k'}x_{k'} + \cdots + a_n x_n < -2\sqrt{\ln(8/\epsilon)} \cdot \sigma_{k'}$ with probability at most $\epsilon^2/8$. Since $2\sqrt{\ln(8/\epsilon)} \cdot \sigma_{k'} < \epsilon^2/(2C_1) \leq \theta$, we have that for at least a $1 - \epsilon^2/8$ fraction of all restrictions ρ to $x_{k'}, \ldots, x_n$, the resulting function f_ρ (on variables $x_1, \ldots, x_{k'-1}$) is $f_\rho(x) = \mathrm{sgn}(a_1 x_1 + \cdots + a_{k'-1}x_{k'-1} - \theta_\rho)$ where $\theta_\rho > 0$. i.e. $\mathbf{E}[f_\rho] < 0$. Together with Fact 33, this implies that for at least an $\epsilon/4 - \epsilon^2/8 > \epsilon/8$ fraction of restrictions ρ, we have $-1 + \epsilon/4 \leq \mathbf{E}[f_\rho] < 0$. Each such f_ρ has $\mathrm{Var}[f_\rho] = \Omega(\epsilon)$, so by Fact 29 has $\mathrm{Inf}_1(f_\rho) = \Omega(\epsilon)/k' = \Omega(\epsilon^5/\log(1/\epsilon))$. Consequently we have that $\mathrm{Inf}_1(f) = \Omega(\epsilon^6/\log(1/\epsilon))$ which is certainly $\Omega(\delta\epsilon^6/\log(1/\epsilon))$. This concludes Case 2.a.ii.

Case 2.b: $k \leq C_4 \log(1/\epsilon)/\epsilon^4 + 1$. We now define $\sigma_k \stackrel{\mathrm{def}}{=} \sqrt{\sum_{j=k}^{n} a_j^2}$ and work with this quantity. First we consider a subcase in which σ_k is "small" relative to θ; this case can be handled using essentially the same arguments as Case 2.a.ii.

Case 2.b.i: $\sigma_k < \theta/(2\sqrt{\ln(8/\epsilon)})$. As above, the Hoeffding bound (now applied to $a_k x_k + \cdots + a_n x_n$) gives that $a_k x_k + \cdots + a_n x_n < -2\sqrt{\ln(8/\epsilon)} \cdot \sigma_k$ with probability at most $\epsilon^2/8$, so for at least a $1 - \epsilon^2/8$ fraction of restrictions ρ to $x_k, \ldots, x_n$ we have $\mathbf{E}[f_\rho] < 0$. Using Fact 33, the argument from Case 2.a.ii again gives that $\mathrm{Inf}_1(f) = \Omega(\epsilon^6/\log(1/\epsilon))$, and we are done with Case 2.b.i.

Case 2.b.ii: $\sigma_k \geq \theta/(2\sqrt{\ln(8/\epsilon)})$. In this case we shall show that $N(0, \sigma_k)$, the zero-mean Gaussian distribution with variance σ_k, assigns at least $2C_3\delta\epsilon^2$ probability weight to the interval $[\theta - \delta/2, \theta + \delta/2]$. In other words, writing Φ_{σ_k} to denote the c.d.f. of $N(0, \sigma_k)$, we shall show

$$\Phi_{\sigma_k}([\theta - \delta/2, \theta + \delta/2]) \geq 3C_3\delta\epsilon^2. \tag{16}$$

Given (16), by the Berry-Esseen theorem and the definition of the critical index we obtain

$$\Pr\left[\sum_{i=k}^{n} a_k \in [\theta - \delta/2, \theta + \delta/2]\right] \geq 3C_3\delta\epsilon^2 - 2C_3\delta\epsilon^2 = C_3\delta\epsilon^2. \tag{17}$$

For any restriction ρ that gives $a_k x_k + \cdots + a_n x_n \in [\theta - \delta/2, \theta + \delta/2]$, Fact 30 gives $\mathrm{Var}[f_\rho] = \Omega(1)$ and hence Fact 29 gives $\mathrm{Inf}_1(f_\rho) = \Omega(1)/k = \Omega(\epsilon^4/\log(1/\epsilon))$. By (17) we thus have $\mathrm{Inf}_1(f) = \Omega(C_3\delta\epsilon^6\log(1/\epsilon))$, which is the desired result.

We turn to proving (16). Let ϕ_{σ_k} denote the c.d.f. of $N(0,\sigma_k)$, i.e. $\phi_{\sigma_k}(x) \stackrel{\text{def}}{=} (1/\sigma_k\sqrt{2\pi})e^{-x^2/2\sigma_k^2}$. We first observe that since $\sigma_k \geq \theta/(2\sqrt{\ln 8/\epsilon})$, we have

$$\phi_{\sigma_k}(\theta) \geq \Omega(1/\sigma_k)\cdot\epsilon^2 \geq 6C_3\epsilon^2, \tag{18}$$

where the second bound holds for a suitably small choice of the absolute constant C_3 and uses $\sigma_k \leq 1$.

We consider two different cases depending on the relative sizes of δ and θ.

Case 2.b.ii.A: $\delta/2 \geq \theta$. In this case we have that $[0,\delta/2] \subseteq [\theta-\delta/2, \theta+\delta/2]$ and it suffices to show that $\Phi_{\sigma_k}([0,\delta/2]) \geq 3\delta\epsilon^2 C_3$.

If $\delta \geq \sigma_k$, then we have

$$\Phi_{\sigma_k}([0,\delta/2]) \geq \Phi_{\sigma_k}([0,\sigma_k/2]) \geq 3C_3 \geq 3C_3\delta\epsilon^2$$

by a suitable choice of the absolute constant C_3. On the other hand, if $\delta < \sigma_k$ then we have

$$\Phi_{\sigma_k}([0,\delta/2]) \geq (\delta/2)\phi_{\sigma_k}(\delta/2) \geq (\delta/2)\phi_{\sigma_k}(\sigma_k/2) \geq 3C_3\delta \geq 3C_3\delta\epsilon^2$$

for a suitable choice of the absolute constant C_3. This gives Case 2.b.ii.A.

Case 2.b.ii.B: $\delta/2 < \theta$. In this case we have

$$\Phi_{\sigma_k}([\theta-\delta/2, \theta+\delta/2]) \geq \Phi_{\sigma_k}([\theta-\delta/2,\theta]) \geq (\delta/2)\cdot\phi_{\sigma_k}(\theta) \geq 3C_3\delta\epsilon^2$$

where the final inequality is obtained using (18). This concludes Case 2.b.ii.B, and with it the proof of Theorem 26. ☐

6.2 Two theorems about LTFs

In this section we prove two theorems that essentially characterize LTFs. These theorems are the analogues of Theorems 24 and 25 in Section 5.1.

The following is the main theorem used in proving the completeness of our test. Roughly speaking, it says that if $f_1 = \mathrm{sgn}(w\cdot x - \theta_1)$, $f_2 = \mathrm{sgn}(w\cdot x - \theta_2)$ are two regular LTFs with the same weights (but possibly different thresholds), then the the inner product of their degree-1 Fourier coefficients is essentially determined by their means.

Theorem 37. *Let f_1 be a τ-regular LTF. Then*

$$\left|\sum_{i=1}^n \widehat{f_1}(i)^2 - W(\mathbf{E}[f_1])\right| \leq \tau^{1/6}. \tag{19}$$

Further, suppose $f_2 : \{-1,1\}^n \to \{-1,1\}$ is another τ-regular LTFs that can be expressed using the same linear form as f_1; i.e., $f_k(x) = \mathrm{sgn}(w\cdot x - \theta_k)$ for some w, θ_1, θ_2. Then

$$\left|\left(\sum_{i=1}^n \widehat{f_1}(i)\widehat{f_2}(i)\right)^2 - W(\mathbf{E}[f_1])W(\mathbf{E}[f_2])\right| \leq \tau^{1/6}. \tag{20}$$

(We assume in this theorem that τ is less than a sufficiently small constant.)

Proof. We first dispense with the case that $|\mathbf{E}[f_1]| \geq 1 - \tau^{1/10}$. In this case, Proposition 2.2 of Talagrand [Tal96] implies that $\sum_{i=1}^n \widehat{f_1}(i)^2 \leq O(\tau^{2/10}\log(1/\tau))$, and Proposition 18 (item 3) implies that $W(\mathbf{E}[f_1]) \leq O(\tau^{2/10}\log(1/\tau))$. Thus

$$\left|\sum_{i=1}^n \widehat{f_1}(i)^2 - W(\mathbf{E}[f_1])\right| \leq O(\tau^{1/5}\log(1/\tau)) \leq \tau^{1/6},$$

so (19) indeed holds. Further, in this case we have

$$\left(\sum_{i=1}^n \widehat{f_1}(i)\widehat{f_2}(i)\right)^2 \overset{\text{Cauchy-Schwarz}}{\leq} \left(\sum_{i=1}^n \widehat{f_1}(i)^2\right)\left(\sum_{i=1}^n \widehat{f_2}(i)^2\right) \leq O(\tau^{1/5}\log(1/\tau)) \cdot 1,$$

and also $W(\mathbf{E}[f_1])W(\mathbf{E}[f_2]) \leq O(\tau^{1/5}\log(1/\tau)) \cdot \frac{2}{\pi}$. Thus (20) holds as well.

We may now assume that $|\mathbf{E}[f_1]| < 1 - \tau^{1/10}$. Without loss of generality, assume that the linear form w defining f_1 (and f_2) has $\|w\| = 1$ and $|w_1| \geq |w_i|$ for all i. Then from Theorem 26 it follows that

$$\tau \geq \mathrm{Inf}_1(f_1) \geq \Omega(|w_1|\tau^{6/10}\log(1/\tau))$$

which implies that $|w_1| \leq O(\tau^{2/5})$. Note that by Proposition 57, this implies that

$$\mathbf{E}[f_k] \overset{\tau^{2/5}}{\approx} \mu(\theta_k), \quad k = 1, 2. \tag{21}$$

Let (x, y) denote a pair of η-correlated random binary strings, where $\eta = \tau^{1/5}$. By definition of $\mathbb{S}_\eta$, we have

$$\mathbb{S}_\eta(f_1, f_2) = 2\Pr[(w \cdot x, w \cdot y) \in A \cup B] - 1,$$

where $A = [\theta_1, \infty) \times [\theta_2, \infty)$ and $B = (-\infty, \theta_1] \times (-\infty, \theta_2]$. Using the same multidimensional Berry-Esseen-based reasoning as in the proof of Proposition 10.1 of [KKMO07], the fact that $|w_i| \leq O(\tau^{2/5})$ holds for all i implies

$$\Pr[(w \cdot x, w \cdot y) \in A \cup B] \overset{\tau^{2/5}}{\approx} \Pr[(X, Y) \in A \cup B],$$

where (X, Y) is a pair of η-correlated standard Gaussians. (Note that the error in the above approximation also depends multiplicatively on constant powers of $1 + \eta$ and of $1 - \eta$, but these are just constants, since $|\eta|$ is bounded away from 1.) It follows that

$$\mathbb{S}_\eta(f_1, f_2) \overset{\tau^{2/5}}{\approx} \mathbb{S}_\eta(h_1, h_2), \tag{22}$$

where $h_k : \mathbf{R} \to \{-1, 1\}$ is the function of one Gaussian variable $h_k(X) = \mathrm{sgn}(X - \theta_k)$.

Using the Fourier and Hermite expansions, we can write Equation (22) as follows:

$$\begin{aligned}\widehat{f_1}(\emptyset)\widehat{f_2}(\emptyset) + \eta \cdot \left(\sum_{i=1}^n \widehat{f_1}(i)\widehat{f_2}(i)\right) &+ \sum_{|S|\geq 2} \eta^{|S|}\widehat{f_1}(S)\widehat{f_2}(S) \\ &\overset{\tau^{2/5}}{\approx} \widehat{h_1}(0)\widehat{h_2}(0) + \eta \cdot \widehat{h_1}(1)\widehat{h_2}(1) + \sum_{j\geq 2} \eta^j \widehat{h_1}(j)\widehat{h_2}(j). \end{aligned} \tag{23}$$

Now by Cauchy-Schwarz (and using the fact that $\eta \geq 0$) we have

$$\left|\sum_{|S|\geq 2} \eta^{|S|}\widehat{f_1}(S)\widehat{f_2}(S)\right| \leq \sqrt{\sum_{|S|\geq 2} \eta^{|S|}\widehat{f_1}(S)^2}\sqrt{\sum_{|S|\geq 2} \eta^{|S|}\widehat{f_2}(S)^2} \leq \eta^2\sqrt{\sum_S \widehat{f_1}(S)^2}\sqrt{\sum_S \widehat{f_2}(S)^2} = \eta^2.$$

The analogous result holds for h_1 and h_2. If we substitute these into Equation (23) and also use

$$\widehat{h_k}(0) = \mathbf{E}[h_k] = \mu(\theta_k) \overset{\tau^{2/5}}{\approx} \mathbf{E}[f_k] = \widehat{f_k}(\emptyset)$$

which follows from Equation (21), we get:

$$\eta \cdot \left(\sum_{i=1}^{n} \widehat{f_1}(i)\widehat{f_2}(i) \right) \overset{\tau^{2/5}+\eta^2}{\approx} \eta \cdot \widehat{h_1}(1)\widehat{h_2}(1) = \eta \cdot 2\phi(\theta_1) \cdot 2\phi(\theta_2),$$

where the equality is by the comment in Definition 17 (item 2). Dividing by η and using $\tau^{2/5}/\eta + \eta = 2\tau^{1/5}$ in the error estimate, we get

$$\sum_{i=1}^{n} \widehat{f_1}(i)\widehat{f_2}(i) \overset{\tau^{1/5}}{\approx} 2\phi(\theta_1) \cdot 2\phi(\theta_2) = \sqrt{W(\mu(\theta_1))W(\mu(\theta_2))}. \tag{24}$$

Since we can apply this with f_1 and f_2 equal, we may also conclude

$$\sum_{i=1}^{n} \widehat{f_k}(i)^2 \overset{\tau^{1/5}}{\approx} W(\mu(\theta_k)) \tag{25}$$

for each $k = 1, 2$.

Using the Mean Value Theorem, the fact that $|W'| \leq 1$ on $[-1, 1]$, and Equation (21), we conclude

$$\sum_{i=1}^{n} \widehat{f_k}(i)^2 \overset{\tau^{1/5}}{\approx} W(\mathbf{E}[f_k])$$

for each $k = 1, 2$, establishing (19). Similar reasoning applied to the square of Equation (24) yields

$$\left(\sum_{i=1}^{n} \widehat{f_1}(i)\widehat{f_2}(i) \right)^2 \overset{\tau^{1/5}}{\approx} W(\mathbf{E}[f_1])W(\mathbf{E}[f_2]),$$

implying (20). The proof is complete. □

The next theorem is a sort of dual of the previous theorem and will be the main theorem we use in proving the soundness of our test. Very roughly speaking, it says that for any Boolean function g and any τ-regular Boolean function f that satisfies certain conditions, if the inner product of the degree-1 Fourier coefficients of f and g is close to the "right" value (see Theorem 37), then g is close to a particular linear threshold function whose weights are the degree-1 Fourier coefficients of f.

Theorem 38. *Let $f, g : \{-1, 1\}^n \to \{-1, 1\}$, and suppose that:*

1. *f is τ-regular and $|\mathbf{E}[f]| \leq 1 - \tau^{2/9}$;*
2. *$|\sum_{i=1}^{n} \hat{f}(i)^2 - W(\mathbf{E}[f])| \leq \tau$;*
3. *$|(\sum_{i=1}^{n} \hat{f}(i)\hat{g}(i))^2 - W(\mathbf{E}[f])W(\mathbf{E}[g])| \leq \tau$, and $\sum_{i=1}^{n} \hat{f}(i)\hat{g}(i) \geq -\tau$.*

Write $\ell(x)$ for the linear form $\sum_{i=1}^{n}(\hat{f}(i)/\sigma)x_i$, where $\sigma = \sqrt{\sum_{i=1}^{n} \hat{f}(i)^2}$. Then there exists $\theta \in \mathbf{R}$ such that $g(x)$ is $O(\tau^{1/9})$-close to the function $\mathrm{sgn}(\ell(x) - \theta)$. Moreover, we have that each coefficient $(\hat{f}(i)/\sigma)$ of $\ell(x)$ is at most $O(\tau^{7/9})$.

Proof. We may assume $|\mathbf{E}[g]| \le 1 - \tau^{1/9}$, since otherwise g is $\tau^{1/9}$-close to a constant function, which may of course be expressed in the desired form. Using this assumption, the fact that $|\mathbf{E}[f]| \le 1 - \tau^{2/9}$, and the final item in Proposition 18, it follows that

$$W(\mathbf{E}[g]) \ge \Omega(\tau^{2/9}) \quad \text{and} \quad W(\mathbf{E}[f]) \ge \Omega(\tau^{4/9}). \tag{26}$$

The latter above, combined with assumption 2 of the theorem, also yields

$$\sigma \ge \Omega(\tau^{2/9}). \tag{27}$$

Note that the second assertion of the theorem follows immediately from the τ-regularity of f and (27).

Let $\theta = \mu^{-1}(\mathbf{E}[g])$. We will show that g is $O(\tau^{1/9})$-close to $\mathrm{sgn}(h)$, where $h(x) = \ell(x) - \theta$, and thus prove the first assertion of the theorem.

Let us consider $\mathbf{E}[gh]$. By Plancherel and the fact that h is affine, we have

$$\mathbf{E}[gh] = \sum_{|S|\le 1} \hat{g}(S)\hat{h}(S) = \sum_{i=1}^{n} \frac{\hat{g}(i)\hat{f}(i)}{\sigma} - \theta\,\mathbf{E}[g]. \tag{28}$$

On the other hand,

$$\mathbf{E}[gh] \le \mathbf{E}[|h|] \overset{\tau}{\approx} \mathbf{E}[|X-\theta|] = 2\phi(\theta) - \theta\mu(\theta) = \sqrt{W(\mathbf{E}[g])} - \theta\,\mathbf{E}[g], \tag{29}$$

where the inequality is because g is ± 1-valued, the following approximation is by Proposition 58, the following equality is by Proposition 59, and the last equality is by definition of θ. Combining Equation (28) and Equation (29) we get

$$\mathbf{E}[|h|] - \mathbf{E}[gh] \le \left(\sqrt{W(\mathbf{E}[g])} - \sum_{i=1}^{n} \frac{\hat{g}(i)\hat{f}(i)}{\sigma}\right) + O(\tau). \tag{30}$$

We now wish to show the parenthesized expression in (30) is small. Using Fact 5 and the first part of assumption 3 of the theorem, we have

$$\left|\left|\sum_{i=1}^{n} \hat{f}(i)\hat{g}(i)\right| - \sqrt{W(\mathbf{E}[f])}\sqrt{W(\mathbf{E}[g])}\right| \le \frac{\tau}{\sqrt{W(\mathbf{E}[f])}\sqrt{W(\mathbf{E}[g])}} \le O(\tau^{6/9}), \tag{31}$$

where we used (26) for the final inequality. We can remove the inner absolute value on the left of (31) by using the second part of assumption 3 and observing that 2τ is negligible compared with $O(\tau^{6/9})$, i.e. we obtain

$$\left|\sum_{i=1}^{n} \hat{f}(i)\hat{g}(i) - \sqrt{W(\mathbf{E}[f])}\sqrt{W(\mathbf{E}[g])}\right| \le O(\tau^{6/9}), \tag{32}$$

We can also use Fact 5 and the first part of assumption 2 of the theorem to get $|\sigma - \sqrt{W(\mathbf{E}[f])}| \le \tau/\sqrt{W(\mathbf{E}[f])} \le O(\tau^{7/9})$. Since $|W(\mathbf{E}[g])| = O(1)$, we thus have

$$\left|\sigma\sqrt{W(\mathbf{E}[g])} - \sqrt{W(\mathbf{E}[f])}\sqrt{W(\mathbf{E}[g])}\right| \le O(\tau^{7/9}). \tag{33}$$

Combining (33) and (32), we have

$$\left|\sum_{i=1}^{n} \hat{f}(i)\hat{g}(i) - \sigma\sqrt{W(\mathbf{E}[g])}\right| \le O(\tau^{6/9}).$$

Dividing through by σ and using (27), this gives that

$$\left| \sum_{i=1}^{n} \frac{\hat{g}(i)\hat{f}(i)}{\sigma} - \sqrt{W(\mathbf{E}[g])} \right| \leq O(\tau^{4/9}).$$

Substituting this into (30) yields

$$\mathbf{E}[|h|] - \mathbf{E}[gh] \leq O(\tau^{4/9}). \tag{34}$$

Let ϵ denote the fraction of points in $\{-1,1\}^n$ on which g and $\mathrm{sgn}(h)$ disagree. Suppose first that that $\epsilon < 12\tau/\sigma$. Since $\sigma \geq \Omega(\tau^{2/9})$ by (27), in this case we have that $\epsilon \leq O(\tau^{7/9})$. Thus we may assume that $\epsilon \geq 12\tau/\sigma$. We may apply Theorem 56 as follows since $\epsilon\sigma/12 \geq \tau \geq \max_i |\hat{f}(i)|$:

$$\Pr[|h(x)| \leq \epsilon\sigma/12] \leq \frac{6\epsilon\sigma/12}{\sigma} = \frac{\epsilon}{2}.$$

It follows that at least an $\epsilon/2$ fraction of inputs x have both $g(x) \neq \mathrm{sgn}(h(x))$ and $|h(x)| > \epsilon\sigma/12$. This implies that $\mathbf{E}[|h|] - \mathbf{E}[gh] \geq 2 \cdot (\epsilon/2) \cdot (\epsilon\sigma/12) = \epsilon^2\sigma/12$. Combining this with the previous bound (34), and recalling that $\sigma \geq \Omega(\tau^{2/9})$, we get that $\epsilon^2 \leq O(\tau^{2/9})$ and thus $\epsilon \leq O(\tau^{1/9})$. This proves that g is $O(\tau^{1/9})$-close to $\mathrm{sgn}(h)$, as desired. □

6.3 Overview of the testing algorithm

We are given $\epsilon > 0$ and black-box access to an unknown $f : \{-1,1\}^n \to \{-1,1\}$, and our goal is to test whether f is an LTF versus ϵ-far from every LTF.

Our testing algorithm **Test-LTF** operates in three phases. The first two phases make queries to the black-box function f; the third phase is a deterministic test making no queries.

In the first phase the algorithm "isolates" a set J that consists of s "influential" coordinates. Essentially, this set J consists of those coordinates i such that $|\hat{f}(i)|$ is large. We call this phase **Isolate-Variables**; in Section 6.4.1 we present the **Isolate-Variables** algorithm and prove a theorem describing its behavior.

We note that one can show that it is possible to *identify* a set J as described above using $\Theta(\log n)$ queries using an approach based on binary search. However, since we want to use a number of queries that is independent of n, we cannot actually afford to explicitly identify the set J (note that indeed this set J is not part of the output that **Isolate-Variables** produces). The approach we use to "isolate" J without identifying it is based in part on ideas from [FKR$^+$02].

In the second phase, the algorithm generates a set $\pi^1, \ldots, \pi^M$ of i.i.d. uniform random strings in $\{-1,1\}^s$; these strings will play the role of restrictions to J. The algorithm then uses the output of **Isolate-Variables** to estimate various parameters of the restricted functions $f_{\pi^1}, \ldots, f_{\pi^M}$. More specifically, for each restriction π^i, the algorithm estimates the mean $\mathbf{E}[f_{\pi^i}]$, the sum of squares of degree-1 Fourier coefficients $\sum_k \widehat{f_{\pi^i}}(k)^2$, and the sum of fourth powers of degree-1 Fourier coefficients $\sum_k \widehat{f_{\pi^i}}(k)^4$; and for each pair of restrictions π^i, π^j, the algorithm estimates the inner product of degree-1 Fourier coefficients $\sum_{k \notin J} \widehat{f_{\pi^i}}(k)\widehat{f_{\pi^j}}(k)$. We call this phase **Estimate-Parameters-Of-Restrictions**; see Section 6.4.2 where we present this algorithm and prove a theorem describing its behavior.

After these two query phases have been performed, in the third phase the algorithm does some computation on the parameters that it has obtained for the restrictions $\pi^1, \ldots, \pi^M$, and either accepts or rejects. In Section 6.5 we give a description of the entire algorithm **Test-LTF** and prove Theorem 27.

6.4 The querying portions of the algorithm

6.4.1 Isolating variables.

Isolate-Variables(inputs are $\tau, \delta > 0$, and black-box access to $f : \{-1,1\}^n \to \{-1,1\}$)

1. Let $\ell = \lceil 1/(\tau^{16}\delta) \rceil$. Randomly partition the set $[n]$ into ℓ "bins" (subsets $B_1, \ldots, B_\ell$) by assigning each $i \in [n]$ to a uniformly selected B_j.
2. Run **Non-Regular**$(\tau^2, \delta/\ell, B_j)$ on each set B_j and let I be the set of those bins B_j such that **Non-Regular** accepts. Let $s = |I|$.
3. Output $(B_1, \ldots, B_\ell, I)$.

We require the following:

Definition 39. *Let $B_1, \ldots, B_\ell$ be a partition of $[n]$ and I be a subset of $\{B_1, \ldots, B_\ell\}$. We say that $(B_1, \ldots, B_\ell, I)$ is* isolationist *if the following conditions hold:*

1. *If $\max_{i \in B_j} |\hat{f}(i)| \geq \tau^2$ then $B_j \in I$;*
2. *If $B_j \in I$ then $\max_{i \in B_j} |\hat{f}(i)| \geq \tau^2/4$;*
3. *If $B_j \in I$ then the second-largest value of $|\hat{f}(i)|$ for $i \in B_j$ is less than $\tau^4/32$.*

Given $(B_1, \ldots, B_\ell, I)$ we define the set J to be

$$J := \bigcup_{B_j \in I} \{\operatorname{argmax}_{k \in B_j} |\hat{f}(k)|\}. \tag{35}$$

The following lemma is useful:

Lemma 40. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any function. With probability $1 - O(\delta)$, the sets $B_1, \ldots, B_\ell$ have the following property: for all j, the set B_j contains at most one element i such that $|\hat{f}(i)| \geq \tau^4/32$.*

Proof. Parseval's identity gives us that there are at most $1024/\tau^8$ many variables i such that $|\hat{f}(i)| \geq \tau^4/32$. For each such variable, the probability that any other such variable is assigned to its bin is at most $(1024/\tau^8)/\ell \leq 1024\tau^8\delta$. A union bound over all (at most $1024/\tau^8$ many) such variables gives that with probability at least $1 - O(\delta)$, each variable x_i with $|\hat{f}(i)| \geq \tau^4/32$ is the only variable that occurs in its bin. This gives the lemma. □

Theorem 41. *Let $f : \{-1,1\}^n \to \{-1,1\}$, and let $\tau, \delta > 0$ be given. Define $s_{\max} = 16/\tau^4$ and $\ell = \lceil 1/(\tau^{16}\delta) \rceil$. Then with probability $1 - O(\delta)$,*

1. *Algorithm* **Isolate-Variables** *outputs a list $(B_1, \ldots, B_\ell, I)$ that is isolationist;*
2. *The corresponding set J has $|J| = |I| \leq s_{\max}$, and J contains all coordinates $i \in [n]$ such that $|\hat{f}(i)| \geq \tau^2$.*

The algorithm makes $\widetilde{O}(1/(\delta\tau^{48}))$ queries to f.

Proof. Part (1) of the theorem follows from Lemma 40 and Lemma 15. Note that Lemma 40 contributes $O(\delta)$ to the failure probability, and since the algorithm runs **Non-Regular** ℓ times with confidence parameter set to δ/ℓ, Lemma 15 contributes another $O(\delta)$ to the failure probability.

We now show that if part (1) holds then so does part (2). Observe that since $(B_1, \ldots, B_\ell, I)$ is isolationist, for each $B_j \in I$ there is precisely one element that achieves the maximum value of $|\hat{f}(k)|$; thus $|J \cap B_j| = 1$ for all $B_j \in I$ and $|J| = |I|$. It is easy to see that $|J| \leq 16/\tau^4$; this follows immediately from Parseval's identity and part 2 of Definition 39.

For the query complexity, observe that **Isolate-Variables** makes $O(1/(\tau^{16}\delta))$ calls to **Non-Regular**$(\tau^2, \delta/\ell, B_j)$, each of which requires $\widetilde{O}(1/\tau^{32})$ queries to f, for an overall query complexity of

$$\widetilde{O}\left(\frac{1}{\delta\tau^{48}}\right)$$

queries. $\square$

6.4.2 Estimating Parameters of Restrictions.

Estimate-Parameters-Of-Restrictions (inputs are $\tau, \eta, \delta > 0$, $M \in \mathbf{Z}^+$, an isolationist list $(B_1, \ldots, B_\ell, I)$ where $|I| = s$, and black-box access to $f : \{-1,1\}^n \to \{-1,1\}$)

0. Let $\delta' := O(\frac{\delta\eta^2}{M^2} \cdot \log(\frac{M^2}{\delta\eta^2}))$.
1. For $i = 1, \ldots, M$ let π^i be an i.i.d. uniform string from $\{-1,1\}^s$.
2. For $i = 1, \ldots, M$ do the following:
   (a) Make $N_\mu := O(\log(1/\delta')/\eta^2)$ calls to **Random-String**(π^i, I, δ', f) to obtain N_μ strings w. Let $\widetilde{\mu}^i$ be the average value of $f(w)$ over the N_μ strings.
   (b) Make $N_\kappa := O(\log(1/\delta')/\eta^2)$ calls to **Correlated-4Tuple**$(\pi^i, \pi^i, I, \delta', f, \eta)$ to obtain N_κ pairs of 4-tuples $(w^1, x^1, y^1, z^1), (w^2, x^2, y^2, z^2)$. Run algorithm **Estimate-Sum-Of-Fourths** on the output of these calls and let $\widetilde{\kappa}^i$ be the value it returns. If $\widetilde{\kappa}^i < 0$ or $\widetilde{\kappa}^i > 1$ then set $\widetilde{\kappa}^i$ to 0 or 1 respectively.
3. For $i, j = 1, \ldots, M$ do the following: Make $N_\rho := O(\log(1/\delta')/\eta^2)$ calls to **Correlated-Pair**$(\pi^i, \pi^j, I, \delta', f, \eta)$ to obtain N_ρ pairs of pairs $(w^1, x^1), (w^2, x^2)$. Run algorithm **Estimate-Inner-Product** on the output of these calls and let $\widetilde{\rho}^{i,j}$ be the value it returns. If $|\widetilde{\rho}^{i,j}| > 1$ then set $\widetilde{\rho}^{i,j}$ to $\mathrm{sgn}(\widetilde{\rho}^{i,j})$.
4. For $i = 1, \ldots, M$, set $(\widetilde{\sigma}^i)^2$ to $(\widetilde{\rho}^{i,i})^2$.

Theorem 42. *Let $f : \{-1,1\}^n \to \{-1,1\}$, $\tau, \eta, \delta > 0$, $M \in \mathbf{Z}^+$, and let $(B_1, \ldots, B_\ell, I)$ be an isolationist list where $|I| = s \leq s_{\max} = 16/\tau^4$. Then with probability at least $1 - \delta$, algorithm* **Estimate-Parameters-Of-Restrictions** *outputs a list of tuples $(\pi^1, \widetilde{\mu}^1, \widetilde{\sigma}^1, \widetilde{\kappa}^1), \ldots, (\pi^M, \widetilde{\mu}^M, \widetilde{\sigma}^M, \widetilde{\kappa}^M)$ and a matrix $(\widetilde{\rho}^{i,j})_{1 \leq i,j \leq M}$ with the following properties:*

1. *Each π^i is an element of $\{-1,1\}^s$; further, the strings $(\pi^i)_{i \geq 1}$ are i.i.d. uniform elements of $\{-1,1\}^s$.*
2. *The quantities $\widetilde{\mu}^i, \widetilde{\rho}^{i,j}$ are real numbers in the range $[-1,1]$, and the quantities $\widetilde{\sigma}^i, \widetilde{\kappa}^i$, are real numbers in the range $[0,1]$.*

3. *For the set J corresponding to $(B_1, \ldots, B_\ell, I)$ as in (35), the following properties hold. (In (a)-(d) below, f_{π^i} denotes the restricted function obtained by substituting π^i's bits for the coordinates of J as follows: for each $k = 1, \ldots, s$, the restriction assigns the value π^i_k to the (unique) variable in $J \cap B_k$.)*

   (a) *For each $i = 1, \ldots, M$,*
$$|\widetilde{\mu}^i - \mathbf{E}[f_{\pi^i}]| \leq \eta.$$

   (b) *For each $i = 1, \ldots, M$,*
$$|\widetilde{\kappa}^i - \sum_{|S|=1} \widehat{f_{\pi^i}}(S)^4| \leq \eta.$$

   (c) *For all $1 \leq i, j \leq M$,*
$$|\widetilde{\rho}^{i,j} - \sum_{|S|=1} \widehat{f_{\pi^i}}(S)\widehat{f_{\pi^j}}(S)| \leq \eta.$$

   (d) *For each $i = 1, \ldots, M$,*
$$|(\widetilde{\sigma}^i)^2 - \sum_{|S|=1} \widehat{f_{\pi^i}}(S)^2| \leq \eta.$$

The algorithm makes $\widetilde{O}\left(\frac{M^2}{\eta^2 \tau^{36}}\right)$ queries to f.

The proof of Theorem 42 uses the ideas from Section 3 as well as certain ideas from [FKR$^+$02]. It appears in Section 6.4.3.

6.4.3 Proof of Theorem 42.

The proof of Theorem 42 follows as a sequence of lemmas. First a word of terminology: for $x \in \{-1,1\}^n$, and π a restriction of the variables in J, we say that x *is compatible with* π if for every $j \in J$ the value of x_j is the value assigned to variable j by π.

The goal of Step 2(a) is to obtain estimates $\widetilde{\mu}^i$ of the means $\mathbf{E}[f_{\pi^i}]$ of the restricted functions f_{π^i}. Thus to execute Step 2(a) of **Estimate-Parameters-Of-Restrictions** we would like to be able to draw uniform strings $x \in \{-1,1\}^n$ conditioned on their being compatible with particular restrictions π^i of the variables in J. Similarly, to estimate sums of squares, fourth powers, etc. of degree-1 Fourier coefficients of restricted functions, recalling Section 3 we would like to be able to draw pairs, 4-tuples, etc. of bitwise correlated strings subject to their being compatible with the restriction

The subroutine **Correlated-4Tuple**, described below, lets us achieve this. (The subroutines **Random-Pair** and **Correlated-Pair** will be obtained as special cases of **Correlated-4Tuple**.) The basic approach, which is taken from [FKR$^+$02], is to work with each block B_j separately: for each block we repeatedly draw correlated assignments until we find ones that agree with the restriction on the variable of J in that block. Once assignments have been independently obtained for all blocks they are combined to obtain the final desired 4-tuple of strings. (For technical reasons, the algorithm actually generates a pair of 4-tuples as seen below.)

Correlated-4Tuple (Inputs are $\pi^1, \pi^2 \in \{-1,1\}^s$, a set I of s bins, $\delta' > 0$, black-box access to $f : \{-1,1\}^n \to \{-1,1\}$, and $\eta \geq 0$. Outputs are two 4-tuples (w^1, x^1, y^1, z^1) and (w^2, x^2, y^2, z^2), each in $(\{-1,1\}^n)^4$.)

1. For each $B_j \in I$, do the following $O(\log(s/\delta'))$ times:

   (a) Draw six independent uniform assignments (call them w^{1j}, x^{1j}, y^{1j} and w^{2j}, x^{2j}, y^{2j}) to the variables in B_j. Let z^{1j} be an assignment to the same variables obtained by independently assigning each variable in B_j the same value it has in $w^{1j} \odot x^{1j} \odot y^{1j}$ with probability $\frac{1}{2} + \frac{1}{2}\eta$ and the opposite value with probability $\frac{1}{2} - \frac{1}{2}\eta$. Let z^{2j} be obtained independently exactly like z^{1j} (in particular we use $w^{1j} \odot x^{1j} \odot y^{1j}$, <u>not</u> $w^{2j} \odot x^{2j} \odot y^{2j}$, to obtain z^{2j}). Let
   $$P = \{i \in B_j : (w^{jk})_i = (x^{jk})_i = (y^{jk})_i = (z^{jk})_i = \pi^k_j \text{ for } k = 1, 2\}.$$
   i.e. P is the set of those $i \in B_j$ such that for $k = 1, 2$, assignments w^{jk}, x^{jk}, y^{jk} and z^{jk} all set bit i the same way that restriction π^k sets π^k_j.

   (b) Run **Non-Regular**$(\tau^2/4, \delta'/(s\log(s/\delta')), P, f)$.

2. If any call of **Non-Regular** above returned "accept," let $(w^{1j}, x^{1j}, y^{1j}, z^{1j})$, $(w^{2j}, x^{2j}, y^{2j}, z^{2j})$ denote the pair of assignments corresponding to the call that accepted. If no call returned "accept," stop everything and FAIL.

3. For $k = 1, 2$ let (w^k, x^k, y^k, z^k) be obtained as follows:

   - For each $i \notin \cup_{B_j \in I} B_j$, set $(w^k)_i, (x^k)_i, (y^k)_i$ independently to ± 1. Similar to 1(a) above, set both $(z^1)_i$ and $(z^2)_i$ independently to $w^1_i \odot x^1_i \odot y^1_i$ with probability $\frac{1}{2} + \frac{1}{2}\eta$.
   - For each bin $B_j \in I$, set the corresponding bits of w according to w^j; the corresponding bits of x according to x^j; the corresponding bits of y according to y^j; and the corresponding bits of z according to z^j.

   Return the 4-tuples (w^1, x^1, y^1, z^1) and (w^2, x^2, y^2, z^2).

Lemma 43. *Each time* **Correlated-4Tuple**$(\pi^1, \pi^2, I, \delta', f)$ *is invoked by* **Estimate-Parameters-Of-Restrictions**, *with probability* $1 - O(\delta')$ *it outputs two 4-tuples* $(w^1, x^1, y^1, z^1), (w^2, x^2, y^2, z^2)$*, each in* $(\{-1,1\}^n)^4$*, such that:*

- *For* $k = 1, 2$ *we have that* w^k, x^k, y^k *and* z^k *are all compatible with* π^k *on* J*;*
- *For* $k = 1, 2$*, for each* $i \notin J$*, the bits* $(w^k)_i, (x^k)_i, (y^k)_i$ *are each independent uniform* ± 1 *values independent of everything else;*
- *For* $k = 1, 2$*, for each* $i \notin J$*, the bit* $(z^k)_i$ *is independently randomly equal to* $(w^1)_i \odot (x^1)_i \odot (y^1)_i$ *with probability* $\frac{1}{2} + \frac{1}{2}\eta$*.*

Proof. Fix any $B_j \in I$, and consider a particular execution of Step 1(a). Let ℓ_j denote the unique element of $J \cap B_j$. By Definition 39 we have that $|\hat{f}(\ell_j)| \geq \tau^2/4$ and $|\hat{f}(k)| < \tau^4/32$ for all $k \in B_j$ such that $k \neq \ell_j$. Now consider the corresponding execution of Step 1(b). Assuming that **Non-Regular** does not make an error, if $\ell_j \in P$ then **Non-Regular** will accept by Lemma 15, and if $\ell_j \notin P$ then by Lemma 15 we have that **Non-Regular** will reject. It is not hard to see (using the fact that $\eta \geq 0$) that the element ℓ_j

belongs to P with probability $\Theta(1)$, so the probability that $O(\log(s/\delta'))$ repetitions of 1(a) and 1(b) will pass for a given B_j without any "accept" occurring is at most $c^{O(\log(s/\delta'))}$, where c is an absolute constant less than 1. Thus the total failure probability resulting from step 2 ("stop everything and fail") is at most $s2^{-O(\log(s/\delta'))} \leq \delta'$. Since each invocation of **Non-Regular** errs with probability at most $\delta'/(s\log(s/\delta'))$ and there are $O(s\log(s/\delta))$ invocations, the total failure probability from the invocations of **Non-Regular** is at most $O(\delta')$.

Once Step 3 is reached, we have that for each j,

- Each of w^{jk}, x^{jk}, y^{jk} is a uniform independent assignment to the variables in B_j conditioned on $(w^{jk})_{\ell_j}, (x^{jk})_{\ell_j}, (y^{jk})_{\ell_j}$ each being set according to the restriction π^k;
- Each bit $z^{jk}_{\ell_j}$ is compatible with π^k_j. For each variable $i \neq \ell_j$ in B_j, the bit z^{jk}_i is independently set to $w^{j1}_i \odot x^{j1}_i \odot y^{j1}_i$ with probability $\frac{1}{2} + \frac{1}{2}\eta$.

By independence of the successive iterations of Step 1 for different B_j's, it follows that the final output strings (w^1, x^1, y^1, z^1) and (w^2, x^2, y^2, z^2) are distributed as claimed in the lemma. □

Remark 44. *The overall algorithm* **Test-LTF** *is nonadaptive because the calls to* **Non-Regular** *(which involve queries to f) in* **Correlated-4Tuple** *are only performed for those B_j which belong to I, and the set I was determined by the outcomes of earlier calls to* **Non-Regular** *(and hence earlier queries to f). The algorithm could be made nonadaptive by modifying* **Correlated-4Tuple** *to always perform Step 1 on all ℓ blocks $B_1, \dots, B_\ell$. Once all these queries were completed for all calls to* **Correlated-4Tuple** *(and thus all queries to f for the entire algorithm were done), the algorithm could simply ignore the results of Step 1 for those sets B_j that do not belong to I. Thus, as claimed earlier, there is an nonadaptive version of the algorithm with somewhat – but only polynomially – higher query complexity (because of the extra calls to* **Non-Regular** *for sets $B_j \notin I$).*

The subroutine **Random-String**(π^i, I, δ', f) can be implemented simply by invoking the subroutine **Correlated-4Tuple**$(\pi^i, \pi^i, I, \delta, f, 0)$ to obtain a pair $(w^1, x^1, y^1, z^1), (w^2, x^2, y^2, z^2)$ and then discarding all components but w^1. This string w^1 is uniform conditioned on being consistent with the restriction π^i. We then easily obtain:

Lemma 45. *If $(B_1, \dots, B_\ell, I)$ is isolationist, then with probability at least $1-\delta'_1$ (where $\delta'_1 := O(MN_\mu\delta')$), each of the M values $\widetilde{\mu}^1, \dots, \widetilde{\mu}^M$ obtained in Step 2(a) of* **Estimate-Parameters-Of-Restriction** *satisfies $|\widetilde{\mu}^i - \mathbf{E}[f_{\pi^i}]| \leq \eta$.*

Proof. Step 2(a) makes a total of MN_μ many calls to **Correlated-4Tuple**, each of which incurs failure probability $O(\delta')$. Assuming the calls to **Correlated-4Tuple** all succeed, by the choice of N_μ each of the M applications of the Chernoff bound contributes another δ' to the failure probability, for an overall failure probability as claimed. □

Now we turn to part 3(b) of Theorem 42, corresponding to Step 2(b) of **Estimate-Parameters-Of-Restrictions**. We have:

Lemma 46. *There is an algorithm* **Estimate-Sum-Of-Fourths** *with the following property: Suppose the algorithm is given as input values $\eta, \delta > 0$, black-box access to f, and the output of N_κ many calls to* **Correlated-4Tuple**$(\pi, \pi, I, \delta, f, \eta)$*. Then with probability $1-\delta$ the algorithm outputs a value v such that*

$$|v - \sum_{k \in [n], k \notin J} \widehat{f_\pi}(k)^4| \leq \eta.$$

Proof. The algorithm is essentially that of Lemma 12. Consider the proof of Lemma 12 in the case where there is only one function f_π and $p = 4$. For the LHS of (1), we would like to empirically estimate $\mathbf{E}[f_\pi(\alpha^1)f_\pi(\alpha^2)f_\pi(\alpha^3)f_\pi(\alpha^4)]$ where $\alpha^1, \ldots, \alpha^4$ are independent uniform strings conditioned on being compatible with π. Such strings can be obtained by taking each $\alpha^1 = w^1, \alpha^2 = w^2, \alpha^3 = x^1$ and $\alpha^4 = x^2$ where $(w^1, x^1, y^1, z^1), (w^2, x^2, y^2, z^2)$ is the output of a call to **Correlated-4Tuple**$(\pi, \pi, I, \delta, f, \eta)$.

For the RHS of (1), we would like to empirically estimate $\mathbf{E}[f_\pi(\alpha^1)f_\pi(\alpha^2)f_\pi(\alpha^3)f_\pi(\alpha^4)]$ where each of $\alpha^1, \alpha^2, \alpha^3$ is independent and uniform conditioned on being compatible with π, and α^4 is compatible with π and has each bit $(\alpha^4)_i$ for $i \notin J$ independently set equal to $(\alpha^1 \odot \alpha^2 \odot \alpha^3)_i$ with probability $\frac{1}{2} + \frac{1}{2}\eta$. By Lemma 43, such strings can be obtained by taking $\alpha^1 = w^1$, $\alpha^2 = x^1$, $\alpha^3 = y^1$, and $\alpha^4 = z^1$. The corollary now follows from Lemma 12. $\square$

Observing that the two restrictions that are arguments to **Correlated-4Tuple** in Step 2(b) are both π^i, Lemma 48 directly gives us part 3(b) of Theorem 42:

Lemma 47. *If $(B_1, \ldots, B_\ell, I)$ is isolationist, then with probability at least $1-\delta_2'$ (where $\delta_2' := O(MN_\kappa\delta')$), each of the M values $\widetilde{\kappa}^i$ obtained in Step 2(b) of* **Estimate-Parameters-Of-Restrictions** *satisfies $|\widetilde{\kappa}^i - \sum_{|S|=1} \widehat{f_{\pi^i}}(S)^4| \leq \eta$.*

Now we turn to parts 3(c)-(d) of Theorem 42, corresponding to Steps 3 and 4 of the algorithm. The subroutine **Correlated-Pair**$(\pi^i, \pi^j, I, \delta', f, \eta)$ works simply by invoking **Correlated-4Tuple**$(\pi^i, \pi^j, I, \delta', f, \eta)$ to obtain a pair $(w^1, x^1, y^1, z^1), (w^2, x^2, y^2, z^2)$ and outputting (u^1, z^1), (u^2, z^2) where each $u^k = (w^k \odot x^k \odot y^k)$. The following corollary of Lemma 12 describes the behavior of algorithm **Estimate-Inner-Product**:

Lemma 48. *There is an algorithm* **Estimate-Inner-Product** *with the following property: Suppose the algorithm is given as input values $\eta, \delta > 0$, black-box access to f, and the output of N_ρ many successful calls to* **Correlated-Pair**$(\pi^1, \pi^2, I, \delta, f, \eta)$. *Then with probability $1 - \delta$ the algorithm outputs a value v such that*

$$|v - \sum_{k\in[n],k\notin J} \widehat{f_{\pi^1}}(k)\widehat{f_{\pi^2}}(k)| \leq \eta.$$

Proof. Again the algorithm is essentially that of Lemma 12. Consider the proof of Lemma 12 in the case where there are $p = 2$ functions f_{π^1} and f_{π^2}. For the LHS of (1), we would like to empirically estimate $\mathbf{E}[f_{\pi_1}(\alpha^1)f_{\pi^2}(\alpha^2)]$ where α^1, α^2 are independent uniform strings conditioned on being compatible with restrictions π^1 and π^2 respectively. Such strings can be obtained by taking each α^k to be u^k where $(u^1, z^1), (u^2, z^2)$ is the output of a call to **Correlated-Pair**$(\pi^1, \pi^2, I, \delta, f\eta)$.

For the RHS of (1), we would like to empirically estimate $\mathbf{E}[f_{\pi_1}(\alpha^1)f_{\pi^2}(\alpha^2)]$ where α^1 is uniform conditioned on being compatible with π^1 and α^2 is compatible with π^2 and has each bit $(\alpha^2)_i$ for $i \notin J$ independently set equal to $(\alpha^1)_i$ with probability $\frac{1}{2} + \frac{1}{2}\eta$. By Lemma 43 and the definition of **Correlated-Pair**, such strings can be obtained by taking $\alpha^1 = u^1$ and $\alpha^2 = z^2$. The corollary now follows from Lemma 12. $\square$

Lemma 48 gives us parts 3(c)-(d) of Theorem 42:

Lemma 49. *If $(B_1, \ldots, B_\ell, I)$ is isolationist, then with probability at least $1-\delta_3'$ (where $\delta_3' := O(M^2N_\rho\delta')$) both of the following events occur: each of the M^2 values $(\widetilde{\rho}^{i,j})^2$ obtained in Step 3 of* **Estimate-Parameters-Of-Restrictions** *satisfies $|\widetilde{\rho}^{i,j} - \sum_{|S|=1} \widehat{f_{\pi^i}}(S)\widehat{f_{\pi^j}}(S)| \leq \eta$, and each of the M values $(\widetilde{\sigma}^i)^2$ obtained in Step 4 satisfies $|(\widetilde{\sigma}^i)^2 - \sum_{|S|=1} \widehat{f_{\pi^i}}(S)^2| \leq \eta$.*

This essentially concludes the proof of parts 1-3 of Theorem 42. The overall failure probability is $O(\delta'_1 + \delta'_2 + \delta'_3)$; by our initial choice of δ' this is $O(\delta)$.

It remains only to analyze the query complexity. It is not hard to see that the query complexity is dominated by Step 3. This step makes $M^2 N_\rho = \widetilde{O}(M^2/\eta^2)$ invocations to **Correlated-4Tuple**$(\pi^i, \pi^j, I, \delta', f, \eta)$; at each of these invocations **Correlated-4Tuple** makes at most

$$O(s_{\max} \log(s_{\max}/\delta') = \widetilde{O}(1/\tau^4)$$

many invocations to **Non-Regular**$(\tau^2/4, \delta', P, f)$, each of which requires

$$O(\log(s_{\max} \log(s_{\max}/\delta')/\delta')/\tau^{32})) = \widetilde{O}(1/\tau^{32})$$

queries by Lemma 15. Thus the overall number of queries is at most

$$\widetilde{O}\left(\frac{M^2}{\eta^2 \tau^{36}}\right).$$

This concludes the proof of Theorem 42. □

6.5 The full algorithm

We are given black-box access to $f : \{-1,1\}^n \to \{-1,1\}$, and also a "closeness parameter" $\epsilon > 0$. Our goal is to distinguish between f being an LTF and f being ϵ-far from every LTF, using $\mathrm{poly}(1/\epsilon)$ many queries. For simplicity of exposition, we will end up distinguishing from being $O(\epsilon)$-far from every LTF. The algorithm for the test is given below, followed by a high-level conceptual explanation of the various steps it performs.

Test-LTF (inputs are $\epsilon > 0$ and black-box access to $f : \{-1,1\}^n \to \{-1,1\}$)

0. Let $\tau = \epsilon^K$, a "regularity parameter", where K is a large universal constant to be specified later.[a] Let δ be a sufficiently small absolute constant.

   We will also take $\eta = \tau$ (the error parameter for **Estimate-Parameters-Of-Restrictions**), $s_{\max} = 16/\tau^4$, and $M = \mathrm{poly}(s_{\max}) \log(1/\delta)/\epsilon^2$.

1. Run **Isolate-Variables**(τ, δ) to obtain output $(B_1, \ldots, B_\ell, I)$. This implicitly defines some set $J \subset [n]$ and explicitly defines its cardinality (the same as the cardinality of I), some s with $s \le s_{\max}$.

2. Run **Estimate-Parameters-Of-Restrictions**$(\tau, \eta, \delta, M, (B_1, \ldots, B_\ell, I), f)$. This produces a list of restrictions $\pi^i \in \{-1,1\}^s$ and real values $\widetilde{\mu}^i, (\widetilde{\sigma}^i)^2, \widetilde{\kappa}^i, \widetilde{\rho}^{i,j}$ where $1 \le i, j \le M$.

3. At this point there are two cases depending on whether or not the fraction of i's for which $|\widetilde{\mu}^i| \ge 1 - \epsilon$ is at least $1 - \epsilon$:

   (a) (The case that for at least a $1-\epsilon$ fraction of i's, $|\widetilde{\mu}^i| \ge 1 - \epsilon$.)
   In this case, enumerate all possible length-s integer vectors w with entries up to $2^{O(s \log s)}$ in absolute value, and also all possible integer thresholds θ in the same range. For each pair (w, θ), check whether $\mathrm{sgn}(w \cdot \pi^i - \theta) = \mathrm{sgn}(\widetilde{\mu}^i)$ holds for at least a $1 - 20\epsilon$ fraction of the values $1 \le i \le M$. If this ever holds, ACCEPT. If it fails for all (w, θ), REJECT.

   (b) (The case that for at least an ϵ fraction of i's, $|\widetilde{\mu}^i| < 1 - \epsilon$.)
   In this case, pick any i^* such that $|\widetilde{\mu}^{i^*}| < 1 - \epsilon$. Then:

      i. Check that $\widetilde{\kappa}^{i^*} \le 2\tau$. If this fails, REJECT.

      ii. Check that $|(\widetilde{\sigma}^{i^*})^2 - W(\widetilde{\mu}^{i^*})| \le 2\tau^{1/12}$. If this fails, REJECT.

      iii. Check that both $|(\widetilde{\rho}^{i^*,i})^2 - W(\widetilde{\mu}^{i^*})W(\widetilde{\mu}^i)| \le 2\tau^{1/12}$ and $\widetilde{\rho}^{i^*,i} \ge -\eta$ hold for all $1 \le i \le M$. If this fails, REJECT.

      iv. Enumerate all possible length-s vectors w whose entries are integer multiples of $\sqrt{\tau}/s$, up to $2^{O(s \log s)}\sqrt{\ln(1/\tau)}$ in absolute value, and also all possible thresholds θ with the same properties. For each pair (w, θ), check that $|\widetilde{\mu}^i - \mu(\theta - w \cdot \pi^i)| \le 5\sqrt{\tau}$ holds for all π^i's. If this ever happens, ACCEPT. If it fails for all (w, θ), REJECT.

[a] We will eventually take $K = 108$.

Note that all parameters described in the test are fixed polynomials in ϵ. Further, the query complexity of both **Isolate-Variables** and **Estimate-Parameters-Of-Restrictions** is polynomial in all parameters (see Theorems 41, 42). Thus the overall query complexity is $\mathrm{poly}(1/\epsilon)$. As given, the test is adaptive, since **Estimate-Parameters-Of-Restrictions** depends on the output of **Isolate-Variables** . However, in remark 44 we discuss how the test can easily be made nonadaptive with only a polynomial blowup in query complexity.

In Section 6.6 we will show that indeed this test correctly distinguishes (with probability at least $2/3$) LTFs from functions that are $O(\epsilon)$-far from being LTFs. Thus our main testing result, Theorem 27, holds as claimed.

6.5.1 Conceptual explanation of the test.

Here we provide a high-level description of the ideas underlying the various stages of the test. The following discussion should not be viewed in the light of mathematical statements but rather as narrative exposition to aid in understanding the test and its analysis. (It may also be useful to refer back to the sketch at the beginning of Section 6.)

In Step 1, the idea is that J is (roughly) the set of variables i such that $|\hat{f}(i)| \geq \tau^2$.

In Step 2, each π^i is an i.i.d. uniform random restriction of the variables in J. Each value $\widetilde{\mu}^i$ is an estimate of $\mathbf{E}[f_{\pi^i}]$, each $(\widetilde{\sigma}^i)^2$ is an estimate of $\sum_k \widehat{f_{\pi^i}}(k)^2$, each $\widetilde{\kappa}^i$ is an estimate of $\sum_k \widehat{f_{\pi^i}}(k)^4$, and each $\widetilde{\rho}^{i,j}$ is an estimate of $\sum_k \widehat{f_{\pi^i}}(k)\widehat{f_{\pi^j}}(k)$.

The idea of Step 3(a) is that in this case, almost every restriction π of the variables in J causes f_π to be very close to a constant function 1 or -1. If this is the case, then f is close to an LTF if and only if it is close to an LTF which is a junta over the variables in J. Step 3(a) enumerates over every possible LTF over the variables in J and checks each one to see if it is close to f.

If the algorithm reaches Step 3(b), then a non-negligible fraction of restrictions π have $|\mathbf{E}[f_\pi]|$ bounded away from 1. We claim that when f is an LTF, this implies that at least one of those restrictions should be τ-regular, and moreover all restrictions should be $\sqrt{\tau}$-regular (these claims are argued using Proposition 62 and Theorem 26, respectively). Step 3(b)(i) verifies that one such restriction π^{i^*} is indeed $\sqrt{\tau}$-regular.

Step 3(b)(ii) checks that the sum of squares of degree-1 Fourier coefficients $\sum_k \widehat{f}_{\pi^{i^*}}(k)^2$ is close to the "correct" value $W(\mathbf{E}[f_{\pi^{i^*}}])$ that the sum should take if $f_{\pi^{i^*}}$ were a $\sqrt{\tau}$-regular LTF (see the first inequality in the conclusion of Theorem 37). If this check passes, Step 3(b)(iii) checks that every other restriction f_{π^i} is such that the inner product of its degree-1 Fourier coefficients with those of $f_{\pi^{i^*}}$, namely $\sum_{k \notin J} \widehat{f_{\pi^i}}(k)\widehat{f_{\pi^{i^*}}}(k)$, is close to the "correct" value $W(\mathbf{E}[f_{\pi^i}])W(\mathbf{E}[f_{\pi^{i^*}}])$ that it should take if f_{π^i} and $f_{\pi^{i^*}}$ were LTFs with the same linear part (see Theorem 37 again).

At this point in Step 3(b), if all these checks have passed then every restriction f_π is close to a function of the form $\mathrm{sgn}(\ell(x) - \theta_\pi)$ with the same linear part (that is based on the degree-1 Fourier coefficients of $f_{\pi^{i^*}}$, see Theorem 38). Finally, Step 3(b)(iv) exhaustively checks "all" possible weight vectors w for the variables in J to see if there is any weight vector that is consistent with all restrictions f_{π^i}. The idea is that if f passes this final check as well, then combining w with ℓ we obtain an LTF that f must be close to.

6.6 Proving correctness of the test

In this section we prove that the algorithm **Test-LTF** is both complete and sound. At many points in these arguments we will need that our large sample $\pi^1, \dots, \pi^M$ of i.i.d. uniform restrictions is representative of the whole set of all 2^s restrictions, in the sense that empirical estimates of various probabilities obtained from the sample are close to the true probabilities over all restrictions. The following proposition collects the various statements of this sort that we will need. All proofs are straightforward Chernoff bounds.

Proposition 50. *After running Steps 0,1 and 2 of* **Test-LTF**, *with probability at least $1 - O(\delta)$ (with respect to the choice of the i.i.d. $\pi^1, \dots, \pi^M$'s in* **Estimate-Parameters-Of-Restrictions***) the following all simultaneously hold:*

1. *The true fraction of restrictions π to J for which $|\mathbf{E}[f_\pi]| \geq 1 - 2\epsilon$ is within an additive $\epsilon/2$ of the fraction of the π^i's for which this holds. Further, the same is true about occurrences of $|\mathbf{E}[f_\pi]| \geq 1 - \epsilon/2$.*

2. *For every pair (w^*, θ^*), where w^* is a length-s integer vector with entries at most $2^{O(s \log s)}$ in absolute value and θ^* is an integer in the same range, the true fraction of restrictions π to J for which*

$$|\mathbf{E}[f_\pi] - \mathrm{sgn}(w^* \cdot \pi - \theta^*)| \leq 3/5$$

is within an additive ϵ of the fraction of π^i's for which this holds. Further, the same is true about occurrences of $\mathrm{sgn}(\mathbf{E}[f_\pi]) = \mathrm{sgn}(w^* \cdot \pi - \theta^*)$.

3. *For every fixed restriction π^* to J, the true fraction of restrictions π to J for which we have*

$$|(\sum_{|S|=1} \widehat{f_{\pi^*}}(S)\widehat{f_\pi}(S))^2 - W(\mathbf{E}[f_{\pi^*}])W(\mathbf{E}[f_\pi])| \le 3\tau^{1/12}$$

is within an ϵ fraction of the true fraction of π^i's for which this holds.

4. *For every fixed pair (w^*, θ^*), where w^* is a length-s vector with entries that are integer multiples of $\sqrt{\tau}/s$ at most $2^{O(s \log s)}\sqrt{\ln(1/\tau)}$ in absolute value and θ^* is an integer multiple of $\sqrt{\tau}/s$ in the same range, the true fraction of restrictions π to J for which*

$$|\mathbf{E}[f_\pi] - \mu(\theta^* - w^* \cdot \pi)| \le 6\sqrt{\tau}$$

is within an additive ϵ of the fraction of π^i's for which this holds.

Proof. All of the claimed statements can be proved simply by using Chernoff bounds (using the fact that the π^i's are i.i.d. and M is large enough) and union bounds. For example, regarding item 4, for any particular (w^*, θ^*), a Chernoff bound implies that the true fraction and the empirical fraction differ by more than ϵ with probability at most $\exp(-\Omega(\epsilon^2 M)) \le \delta/2^{\mathrm{poly}(s)}$, using the fact that $M \ge \mathrm{poly}(s)\log(1/\delta)/\epsilon$. Thus we may union bound over all $2^{\mathrm{poly}(s)}$ possible (w^*, θ^*) to get that the statement of item 4 holds except with probability at most δ. The other statement and the other items follow by similar or easier considerations. $\square$

6.6.1 Completeness of the test.

Theorem 51. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any LTF. Then f passes* **Test-LTF** *with probability at least $2/3$.*

Proof. Steps 1 and 2 of the test, where querying to f occurs, are the places where the test has randomness. We have that Step 1 succeeds except with probability at most δ; assuming it succeeds, the set J becomes implicitly defined according to (35). Step 2 also succeeds except with probability at most δ; assuming it succeeds, we obtain restrictions π^i and estimates $\widetilde{\mu}^i, (\widetilde{\sigma}^i)^2, \widetilde{\kappa}^i, \widetilde{\rho}^{i,j}$ that satisfy the conclusion of Theorem 42, with $\eta := \tau$. Finally, in Proposition 50 (which relates the empirical properties of the restrictions to the true properties), all conclusions hold except with probability at most $O(\delta)$. Thus all of these assumptions together hold with probability at least $1 - O(\delta)$, which is at least $2/3$ when we take δ to be a sufficiently small constant. Note that we have not yet used the fact that f is an LTF.

We will now show that given that all of these assumptions hold, the fact that f is an LTF implies that the deterministic part of the test, Step 3, returns ACCEPT. We consider the two cases that can occur:

Case 3(a): for at least a $1-\epsilon$ fraction of i's, $|\widetilde{\mu}^i| \ge 1-\epsilon$. Since Theorem 42 implies that $|\widetilde{\mu}^i - \mathbf{E}[f_{\pi^i}]| \le \eta$, and since $\eta \ll \epsilon$, in this case we have that for at least a $1-\epsilon$ fraction of the i's it holds that $|\mathbf{E}[f_{\pi^i}]| \ge 1 - \epsilon - \eta \ge 1 - 2\epsilon$. Applying Proposition 50 item 1, we get that $|\mathbf{E}[f_\pi]| \ge 1 - 2\epsilon$ for at least a $1 - 2\epsilon$ fraction of all 2^s restrictions π on J. It follows that f is $2\epsilon \cdot \frac{1}{2} + (1-2\epsilon)\cdot\epsilon \le 2\epsilon$-close to being a junta on J. Thus by Proposition 63 we have that f is 2ϵ-close to being an LTF on J.

Write this LTF on J as $g(\pi) = \mathrm{sgn}(w^* \cdot \pi - \theta^*)$, where w^* is an integer vector with entries at most $2^{O(s \log s)}$ in absolute value and θ^* is also an integer in this range. (Since $|J| \le s$, any LTF on J can be expressed thus by the well-known result of Muroga *et al.* [MTT61].) Since f is 2ϵ-close to g, we know that for at least a $1 - 10\epsilon$ fraction of the restrictions π to J, $f_\pi(x)$ takes the value $g(\pi)$ on at least a $4/5$ fraction of inputs x. I.e., $|\mathbf{E}[f_\pi] - \mathrm{sgn}(w^* \cdot \pi - \theta^*)| \le 3/5$ for at least a $1 - 10\epsilon$ fraction of all π's. Using Proposition 50 item 2 we conclude that $|\mathbf{E}[f_{\pi^i}] - \mathrm{sgn}(w^* \cdot \pi^i - \theta^*)| \le 3/5$ for at least a $1 - 20\epsilon$ fraction of the π^i's. But for these π^i's we additionally have $|\widetilde{\mu}^i - \mathrm{sgn}(w^* \cdot \pi^i - \theta^*)| \le 3/5 + \eta < 1$ and hence $\mathrm{sgn}(\widetilde{\mu}^i) = \mathrm{sgn}(w^* \cdot \pi^i - \theta^*)$. Thus Step 3(a) returns ACCEPT once it tries (w^*, θ^*).

Case 3(b): for at least an ϵ fraction of i's, $|\widetilde{\mu}^i| < 1-\epsilon$. In this case we need to show that Steps i.–iv. pass.

To begin, since $|\widetilde{\mu}^i - \mathbf{E}[f_{\pi^i}]| \le \eta \ll \epsilon/2$ for all i, we have that for at least an ϵ fraction of the i's, $|\mathbf{E}[f_{\pi^i}]| \le 1-\epsilon/2$. Thus by Proposition 50 item 1, we know that among all 2^s restrictions π to J, the true fraction of restrictions for which $|\mathbf{E}[f_{\pi^i}]| \le 1-\epsilon/2$ is at least $\epsilon/2$.

On the other hand, since J contains all coordinates j with $|\hat{f}(j)| \ge \tau^2$, we know from Proposition 62 that f_π is not τ-regular for at most a τ fraction of the 2^s restrictions π to J. Since $\tau \ll \epsilon/2$, we conclude that there must exist some restriction π_0 to the coordinates in J for which both $|\mathbf{E}[f_{\pi_0}]| \le 1-\epsilon/2$ and f_{π_0} is τ-regular.

Express f as $f(\pi,x) = \mathrm{sgn}(w'\cdot\pi + \ell\cdot x - \theta')$, where π denotes the inputs in J, x denotes the inputs not in J, and ℓ is normalized so that $\|\ell\| = 1$. We've established that the LTF $f_{\pi_0}(x) = \mathrm{sgn}(\ell\cdot x - (\theta' - w'\cdot\pi_0))$ has $|\mathbf{E}[f_{\pi_0}]| \le 1-\epsilon/2$ and is τ-regular. Applying Theorem 26, we conclude that all coefficients in ℓ are, in absolute value, at most $O(\tau/(\epsilon^6\log(1/\epsilon))) \le \Omega(\sqrt{\tau})$; here use the fact that $K \gg 12$.. In particular, we've established:

Claim 52. *There is a linear form ℓ with $\|\ell\| = 1$ and all coefficients of magnitude at most $\Omega(\sqrt{\tau})$, such that the following two statements hold: 1. For every restriction π to J, the LTF f_π is expressed as $f_\pi(x) = \mathrm{sgn}(\ell\cdot x - (\theta' - w'\cdot\pi))$. 2. For every restriction π to J, f_π is $\sqrt{\tau}$-regular.*

The second statement in the claim follows immediately from the first statement and Proposition 64, taking the constant in the $\Omega(\cdot)$ to be sufficiently small.

We now show that Steps 3b(i)–(iv) all pass. Since f_π is $\sqrt{\tau}$-regular for all π, in particular $f_{\pi^{i^*}}$ is $\sqrt{\tau}$-regular. Hence $\sum_{|S|=1}\widehat{f_{\pi^{i^*}}}(S)^4 \le \tau$ (see Proposition 14) and so $\widetilde{\kappa}^{i^*} \le \tau + \eta \le 2\tau$. Thus Step 3b(i) passes.

Regarding Step 3b(ii), Claim 52 implies in particular that $f_{\pi^{i^*}}$ is $\sqrt{\tau}$-regular. Hence we may apply the first part of Theorem 37 to conclude that $\sum_{|S|=1}\widehat{f_{\pi^{i^*}}}(S)^2$ is within $\tau^{1/12}$ of $W(\mathbf{E}[f_{\pi^{i^*}}])$. The former quantity is within η of $(\widetilde{\sigma}^{i^*})^2$; the latter quantity is within η of $W(\widetilde{\mu}^{i^*})$ (using $|W'| \le 1$). Thus indeed $(\widetilde{\sigma}^{i^*})^2$ is within $\tau^{1/12} + \eta + \eta \le 2\tau^{1/12}$ of $W(\widetilde{\mu}^{i^*})$, and Step 3b(ii) passes.

The fact that the first condition in Step 3b(iii) passes follows very similarly, using the second part of Theorem 37 (a small difference being that we can only say that $W(\mathbf{E}[f_{\pi^{i^*}}])W(\mathbf{E}[f_{\pi^i}])$ is within, say, 3η of $W(\widetilde{\mu}^{i^*})W(\widetilde{\mu}^i)$). As for the second condition in Step 3b(iii), since f is an LTF, for any pair of restrictions π, π' to J, the functions f_π and $f_{\pi'}$ are LTFs expressible using the same linear form. This implies that f_π and $f_{\pi'}$ are both unate functions with the same orientation, a condition which easily yields that $\widehat{f_\pi}(j)$ and $\widehat{f_{\pi'}}(j)$ never have opposite sign for any j. We thus have that $\sum_{|S|=1}\widehat{f_{\pi^i}}(S)\widehat{f_{\pi^{i^*}}}(S) \ge 0$ and so indeed the condition $\widetilde{\rho}^{i^*,i} \ge -\eta$ holds for all i. Thus Step 3b(iii) passes.

Finally we come to Step 3b(iv). Claim 52 tells us that for every restriction π^i, we have $f_{\pi^i}(x) = \mathrm{sgn}(\ell\cdot x - (\theta' - w'\cdot\pi^i))$, where ℓ is a linear form with 2-norm 1 and all coefficients of magnitude at most $\Omega(\sqrt{\tau})$. Applying Proposition 57 we conclude that $|\mathbf{E}[f_\pi] - \mu(\theta' - w'\cdot\pi^i)| \le \sqrt{\tau}$ holds for all i (again, ensuring the constant in the $\Omega(\cdot)$ is small enough). Using the technical Lemma 53 below, we infer that there is a vector w^* whose entries are integer multiples of $\sqrt{\tau}/s$ at most $2^{O(s\log s)}\sqrt{\ln(1/\tau)}$ in absolute value, and an integer multiple θ^* of $\sqrt{\tau}/s$, also at most $2^{O(s\log s)}\sqrt{\ln(1/\tau)}$ in absolute value, such that $|\mathbf{E}[f_{\pi^i}] - \mu(\theta^* - w^*\cdot\pi^i)| \le 4\sqrt{\tau}$ holds for all π^i. By increasing the $4\sqrt{\tau}$ to $4\sqrt{\tau} + \eta \le 5\sqrt{\tau}$, we can make the same statement with $\widetilde{\mu}^i$ in place of $\mathbf{E}[f_{\pi^i}]$. Thus Step 3(b)(iv) will return ACCEPT once it tries (w^*,θ^*). $\square$

Lemma 53. *Suppose that $|\mathbf{E}[f_\pi] - \mu(\theta' - w'\cdot\pi)| \le \sqrt{\tau}$ holds for some set Π of π's. Then there is a vector w^* whose entries are integer multiples of $\sqrt{\tau}/s$ at most $2^{O(s\log s)}\sqrt{\ln(1/\tau)}$ in absolute value, and an integer* *multiple θ^* of $\sqrt{\tau}/s$, also at most $2^{O(s\log s)}\sqrt{\ln(1/\eta)}$ in absolute value, such that $|\mathbf{E}[f_\pi]-\mu(\theta^*-w^*\cdot\pi)| \le 4\eta^{1/6}$ also holds for all $\pi \in \Pi$.*

Proof. Let us express the given estimates as

$$\left\{\mathbf{E}[f_\pi]-\sqrt{\tau} \le \mu(\theta'-w'\cdot\pi) \le \mathbf{E}[f_\pi]+\sqrt{\tau}\right\}_{\pi\in\Pi} \tag{36}$$

We would prefer all of the upper bounds $\mathbf{E}[f_\pi]+\sqrt{\tau}$ and lower bounds $\mathbf{E}[f_\pi]-\sqrt{\tau}$ in these double inequalities to have absolute value either equal to 1, or at most $1-\sqrt{\tau}$. It is easy to see that one can get this after introducing some quantities $1 \le K_\pi, K'_\pi \le 2$ and writing instead

$$\left\{\mathbf{E}[f_\pi]-K_\pi\sqrt{\tau} \le \mu(\theta'-w'\cdot\pi) \le \mathbf{E}[f_\pi]+K'_\pi\sqrt{\tau}\right\}_{\pi\in\Pi}. \tag{37}$$

Using the fact that μ is a monotone function, we can apply μ^{-1} and further rewrite (37) as

$$\left\{c_\pi \le \theta'-w'\cdot\pi \le C_\pi\right\}_{\pi\in\Pi}, \tag{38}$$

where each $|c_\pi|, |C_\pi|$ is either ∞ (meaning the associated inequality actually drops out) or is at most $\mu^{-1}(-1+\sqrt{\tau}) \le O(\sqrt{\ln(1/\tau)})$. Now (38) may actually be thought of as a "linear program" in the entries of w' and in θ' — one which we know is feasible.

By standard results in linear programming [Chv83] we know that if such a linear program is feasible, it has a feasible solution in which the variables take values that are not too large. In particular, we can take as an upper bound for the variables

$$\mathcal{L} = \frac{|\max_A \det(A)|}{|\min_B \det(B)|}, \tag{39}$$

where B ranges over all nonsingular square submatrices of the constraint matrix and A ranges over all square submatrices of the constraint matrix with a portion of the "right-side vector" substituted in as a column. Note that the constraint matrix from (38) contains only ± 1's and that the right-side vector contains numbers at most $O(\sqrt{\ln(1/\tau)})$ in magnitude. Thus the minimum in the denominator of (39) is at least 1 and the maximum in the numerator of (39) is at most $O(\sqrt{\ln(1/\tau)})\cdot(s+1)!$; hence $\mathcal{L} \le 2^{O(s\log s)}\sqrt{\ln(1/\tau)}$.

Having made this conclusion, we may recast and slightly weaken (37) by saying that there exist a pair (w'', θ''), with entries all at most $\mathcal{L}$ in absolute value, such that

$$\left\{\mathbf{E}[f_\pi]-2\sqrt{\tau} \le \mu(\theta''-w''\cdot\pi) \le \mathbf{E}[f_\pi]+2\sqrt{\tau}\right\}_{\pi\in\Pi}$$

Finally, suppose we round the entries of w'' to the nearest integer multiples of $\sqrt{\tau}/s$ forming w^*, and we similarly round θ'' to θ^*. Then $|(\theta''-w''\cdot\pi)-(\theta^*-w^*\cdot\pi)| \le 2\sqrt{\tau}$ for every π. Since $|\mu'| \le \sqrt{2/\pi} \le 1$ we can thus conclude that the inequalities

$$\left\{\mathbf{E}[f_\pi]-4\sqrt{\tau} \le \mu(\theta^*-w^*\cdot\pi) \le \mathbf{E}[f_\pi]+4\sqrt{\tau}\right\}_{\pi\in\Pi}$$

also hold, completing the proof. □

6.6.2 Soundness of the test.

Theorem 54. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a function that passes* **Test-LTF** *with probability more than $1/3$. Then f is $O(\epsilon)$-close to an LTF.*

Proof. As mentioned at the beginning of the proof of Theorem 51, for any f, with probability at least $1-O(\delta)$ Step 1 of the algorithm succeeds (implicitly producing J), Step 2 of the algorithm succeeds (producing the π^i's, etc.), and all of the items in Proposition 50 hold. So if an f passes the test with probability more than $1/3 \ge O(\delta)$, it must be the case that f passes the deterministic portion of the test, Step 3, despite the above three conditions holding. We will show that in this case f must be $O(\epsilon)$-close to an LTF. We now divide into two cases according to whether f passes the test in Step 3(a) or Step 3(b).

Case 3(a). In this case we have that for at least a $1-\epsilon$ fraction of π^i's, $|\widetilde{\mu}^i| \geq 1-\epsilon$ and hence $|\mathbf{E}[f_{\pi^i}]| \geq 1-\epsilon-\eta \geq 1-2\epsilon$. By Proposition 50 item 1we conclude:

$$\text{For at least a } 1-2\epsilon \text{ fraction of all restrictions } \pi \text{ to } J,\ |\mathbf{E}[f_\pi]| \geq 1-2\epsilon. \tag{40}$$

Also, since the test passed, there is some pair (w^*, θ^*) such that $\mathrm{sgn}(w^* \cdot \pi^i - \theta^*) = \mathrm{sgn}(\widetilde{\mu}^i)$ for at least a $1-20\epsilon$ fraction of the π^i's. Now except for at most an ϵ fraction of the π^i's we have $|\mathbf{E}[f_{\pi^i}]| \geq 1-2\epsilon \geq \frac{2}{3}$ and $|\widetilde{\mu}^i - \mathbf{E}[f_{\pi^i}]| \leq \eta < \frac{1}{3}$ whence $\mathrm{sgn}(\widetilde{\mu}^i) = \mathrm{sgn}(\mathbf{E}[f_{\pi^i}])$. Hence $\mathrm{sgn}(w^* \cdot \pi^i - \theta^*) = \mathrm{sgn}(\mathbf{E}[f_{\pi^i}])$ for at least a $1-20\epsilon-\epsilon \geq 1-21\epsilon$ fraction of the π^i's. By Proposition 50 item 2 we conclude:

$$\text{For at least a } 1-22\epsilon \text{ fraction of all restrictions } \pi \text{ to } J,\ \mathrm{sgn}(\mathbf{E}[f_\pi]) = \mathrm{sgn}(w^* \cdot \pi - \theta^*). \tag{41}$$

Combining (40) and (41), we conclude that except for a $22\epsilon + 2\epsilon \leq 24\epsilon$ fraction of restrictions π to J, f_π is ϵ-close, as a function of the bits outside J, to the constant $\mathrm{sgn}(w^* \cdot \pi - \theta^*)$. Thus f is $24\epsilon + (1-24\epsilon)\epsilon \leq 25\epsilon$-close to the J-junta LTF $\pi \mapsto \mathrm{sgn}(w^* \cdot \pi - \theta^*)$. This completes the proof in Case 3(a).

Case 3(b). In this case, write π^* for π^{i^*}. Since $|\widetilde{\mu}^{i^*}| \leq 1-\epsilon$, we have that $|\mathbf{E}[f_{\pi^*}]| \leq 1-\epsilon+\eta \leq 1-\epsilon/2$. Once we pass Step 3(b)(i) we have $\widetilde{\kappa}^{i^*} \leq 2\tau$ which implies $\sum_{|S|=1} \widehat{f_{\pi^*}}(S)^4 \leq 2\tau + \eta \leq 3\tau$. This in turn implies that f_{π^*} is $(3\tau)^{1/4} \leq 2\tau^{1/4}$-regular. Once we pass Step 3(b)(ii), we additionally have $|\sum_{|S|=1} \widehat{f_{\pi^*}}(S)^2 - W(\mathbf{E}[f_{\pi^*}])| \leq 2\tau^{1/12} + \eta + \eta \leq 3\tau^{1/12}$, where we've also used that $W(\widetilde{\mu}^{i^*})$ is within η of $W(\mathbf{E}[f_{\pi^*}])$ (since $|W'| \leq 1$).

Summarizing:

$$f_{\pi^*} \text{ is } 2\tau^{1/4}\text{-regular and satisfies } |\mathbf{E}[f_{\pi^*}]| < 1-\epsilon/2,\ \left|\sum_{|S|=1} \widehat{f_{\pi^*}}(S)^2 - W(\mathbf{E}[f_{\pi^*}])\right| \leq 3\tau^{1/12}. \tag{42}$$

Since Step 3(b)(iii) passes we have that both $|(\widetilde{\rho}^{i^*,i})^2 - W(\widetilde{\mu}^{i^*})W(\widetilde{\mu}^i)| \leq 2\tau^{1/12}$ and $\widetilde{\rho}^{i^*,i} \geq -\eta$ hold for all i's. These conditions imply $|(\sum_{|S|=1} \widehat{f_{\pi^*}}(S)\widehat{f_{\pi^i}}(S))^2 - W(\mathbf{E}[f_{\pi^*}])W(\mathbf{E}[f_{\pi^i}])| \leq 2\tau^{1/12} + 4\eta \leq 3\tau^{1/12}$ and $\sum_{|S|=1} \widehat{f_{\pi^*}}(S)\widehat{f_{\pi^i}}(S) \geq -2\eta$ hold for all i. Applying Proposition 50 item 3 we conclude:

For at least a $1-\epsilon$ fraction of the restrictions π to J, both

$$\left|\left(\sum_{|S|=1} \widehat{f_{\pi^*}}(S)\widehat{f_\pi}(S)\right)^2 - W(\mathbf{E}[f_{\pi^*}])W(\mathbf{E}[f_\pi])\right| \leq 3\tau^{1/12} \text{ and } \sum_{|S|=1} \widehat{f_{\pi^*}}(S)\widehat{f_{\pi^i}}(S) \geq -2\eta. \tag{43}$$

We can use (42) and (43) in Theorem 38, with f_{π^*} playing the role of f, the good f_π's from (43) playing the roles of g and the "τ" parameter of Theorem 38 set to $3\tau^{1/12}$. (This requires us to ensure $K \gg 54$.) We conclude:

There is a fixed vector ℓ with $\|\ell\| = 1$ and $|\ell_j| \leq O(\tau^{7/108})$ for each j

such that for at least a $1-\epsilon$ fraction of restrictions π to J,

$$f_\pi(x) \text{ is } O(\tau^{1/108})\text{-close to the LTF } g_\pi(x) = \mathrm{sgn}(\ell \cdot x - \theta_\pi). \tag{44}$$

We now finally use the fact that Step 3(b)(iv) passes to get a pair (w^*, θ^*) such that $|\widetilde{\mu}^i - \mu(\theta^* - w^* \cdot \pi^i)| \leq 5\sqrt{\tau} \Rightarrow |\mathbf{E}[f_{\pi^i}] - \mu(\theta^* - w^* \cdot \pi^i)| \leq 6\sqrt{\tau}$ holds for all π^i's. By Proposition 50 item 4 we may conclude that:

$$\text{For at least a } 1-\epsilon/2 \text{ fraction of restrictions } \pi \text{ to } J,\ |\mathbf{E}[f_\pi] - \mu(\theta^* - w^* \cdot \pi)| \leq 6\sqrt{\tau}. \tag{45}$$

Define the LTF $h : \{-1,1\}^n \to \{-1,1\}$ by $h(\pi, x) = \mathrm{sgn}(w^* \cdot \pi + \ell \cdot x - \theta^*)$. We will complete the proof by showing that f is $O(\tau^{1/108})$-close to h.

We have that the conclusions of (44) and (45) hold simultaneously for at least a $1 - 2\epsilon$ fraction of restrictions π; call these the "good" restrictions. For the remaining "bad" restrictions π' we will make no claim on how close to each other $f_{\pi'}$ and $h_{\pi'}$ may be. However, these bad restrictions contribute at most 2ϵ to the distance between f and h, which is negligible compared to $O(\tau^{1/108})$. Thus it suffices for us to show that for any good restriction π, we have that f_π and h_π are oh-so-close, namely, $O(\tau^{1/108})$-close. So assume π is a good restriction. In that case we have that f_π is $O(\tau^{1/108})$-close to g_π, so it suffices to show that g_π is $O(\tau^{1/108})$-close to h_π. We have $h_\pi(x) = \mathrm{sgn}(\ell \cdot x - (\theta^* - w^* \cdot \pi))$, and since $\|\ell\| = 1$ and $|\ell_j| \le O(\alpha^{7/108})$ for each j, Proposition 57 implies that $\mathbf{E}[h_\pi] \overset{\tau^{7/108}}{\approx} \mu(\theta^* - w^* \cdot \pi)$. Since π is a good restriction, using (45) we have that $\mathbf{E}[h_\pi] \overset{6\sqrt{\tau}}{\approx} \mathbf{E}[f_\pi]$. This certainly implies $\mathbf{E}[h_\pi] \overset{\alpha^{1/108}}{\approx} \mathbf{E}[g_\pi]$ since f_π and g_π are $O(\alpha^{1/108})$-close. But now it follows that indeed g_π is $O(\alpha^{1/108})$-close to h_π because the functions are both LTFs expressible with the same linear form and thus either $g_\pi \ge h_\pi$ pointwise or $h_\pi \ge g_\pi$ pointwise, either of which implies that the distance between the two functions is proportional to the difference of their means.

Finally, we've shown that f is $O(\tau^{1/108})$-close to an LTF. Taking $K = 108$ completes the proof. $\square$

References

[AKK+03] N. Alon, T. Kaufman, M. Krivelevich, S. Litsyn, and D. Ron. Testing low-degree polynomials over GF(2). In *Proceedings of RANDOM-APPROX*, pages 188–199, 2003.

[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.

[BLR93] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comp. Sys. Sci.*, 47:549–595, 1993. Earlier version in STOC'90.

[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.

[Chv83] V. Chvátal. *Linear Programming*. W. H. Freeman, 1983.

[DLM+07] I. Diakonikolas, H. Lee, K. Matulef, K. Onak, R. Rubinfeld, R. Servedio, and A. Wan. Testing for concise representations. To appear in FOCS 2007, 2007.

[Fel68] W. Feller. *An introduction to probability theory and its applications*. John Wiley & Sons, 1968.

[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.

[FLN+02] E. Fischer, E. Lehman, I. Newman, S. Raskhodnikova, R. Rubinfeld, and A. Samrodnitsky. Monotonicity testing over general poset domains. In *Proc. 34th Annual ACM Symposium on the Theory of Computing*, pages 474–483, 2002.

[FP04] A. Fiat and D. Pechyony. Decision trees: More theoretical justification for practical algorithms. In *Algorithmic Learning Theory, 15th International Conference (ALT 2004)*, pages 156–170, 2004.

[GGL+00] O. Goldreich, S. Goldwasser, E. Lehman, D. Ron, and A. Samordinsky. Testing monotonicity. *Combinatorica*, 20(3):301–337, 2000.

[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.

[LT91] M. Ledoux and M. Talagrand. *Probability in Banach Spaces*. Springer, 1991.

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

[MTT61] S. Muroga, I. Toda, and S. Takasu. Theory of majority switching elements. *J. Franklin Institute*, 271:376–418, 1961.

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

[Pet95] V. V. Petrov. *Limit theorems of probability theory*. Oxford Science Publications, Oxford, England, 1995.

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

[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.

[Šte00] D. Štefankovič. Fourier transform in computer science. Master's thesis, University of Chicago, 2000.

[Tal96] M. Talagrand. How much are increasing sets positively correlated? *Combinatorica*, 16(2):243–258, 1996.

[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.

A Basic Theorems about Gaussians and LTFs

A.1 Gaussian basics.

We will often require the Berry-Esseen theorem, a version of the Central Limit Theorem with error bounds (see, e.g., [Fel68]):

Theorem 55. *Let $\ell(x) = c_1x_1 + \cdots + c_nx_n$ be a linear form over the random ± 1 bits x_i. Assume that $|c_i| \leq \tau$ for all i, and write $\sigma = \sqrt{\sum c_i^2}$. Write F for the c.d.f. of $\ell(x)/\sigma$; i.e., $F(t) = \Pr[\ell(x)/\sigma \leq t]$. Then for all $t \in \mathbf{R}$,*

$$|F(t) - \Phi(t)| \leq O(\tau/\sigma) \cdot \frac{1}{1 + |t|^3},$$

where Φ denotes the c.d.f. of X, a standard Gaussian random variable. In particular, if $A \subseteq \mathbf{R}$ is any interval then $\Pr[\ell(x)/\sigma \in A] \overset{\tau/\sigma}{\approx} \Pr[X \in A]$.

A special case of this theorem, with a sharper constant, is sometimes useful (the following can be found in [Pet95]):

Theorem 56. *In the setup of Theorem 55, for any $\lambda \geq \tau$ and any $\theta \in \mathbf{R}$ it holds that $\Pr[|\ell(x) - \theta| \leq \lambda] \leq 6\lambda/\sigma$.*

We will use the following proposition:

Proposition 57. *Let $f(x) = \mathrm{sgn}(c \cdot x - u)$ be an LTF such that $\sum_i c_i^2 = 1$ and $|c_i| \leq \tau$ for all i. Then we have $\mathbf{E}[f] \overset{\tau}{\approx} \mu(u)$.*

This is an almost immediate consequence of the Berry-Esseen theorem. Next we prove the following more difficult statement, which gives an approximation for the expected magnitude of the linear form $c \cdot x - u$ itself:

Proposition 58. *Let $\ell(x) = \sum c_i x_i$ be a linear form over $\{-1,1\}^n$ and assume $|c_i| \leq \tau$ for all i. let $\sigma = \sqrt{\sum c_i^2}$ and let $u \in \mathbf{R}$. Then*

$$\mathbf{E}[|\ell - u|] \overset{\tau}{\approx} \mathbf{E}[|\sigma X - u|],$$

where X is a standard Gaussian random variable.

Proof. The result is certainly true if $\sigma = 0$, so we may assume $\sigma > 0$. Using the fact that $\mathbf{E}[R] = \int_0^\infty \Pr[R > s]\, ds$ for any nonnegative random variable R for which $\mathbf{E}[R] < \infty$, we have that

$$\begin{aligned}
\mathbf{E}[|\ell - u|] &= \int_0^\infty \Pr[|\ell - u| > s]\, ds \\
&= \int_0^\infty \Pr[\ell > u + s] + \Pr[\ell < u - s]\, ds \\
&= \int_0^\infty (1 - F((u+s)/\sigma) + F((u-s)/\sigma)\, ds \qquad (46)
\end{aligned}$$

where we have written F for the c.d.f. of $\ell(x)/\sigma$. We shall apply Berry-Esseen to $\ell(x)$. Since $\sum_{i=1}^n \mathbf{E}[|cx_i|^3] = \sum_{i=1}^n |c_i|^3 \leq \tau \sum_{i=1}^n c_i^2 = \tau\sigma^2$, Berry-Esseen tells us that for all $z \in \mathbf{R}$ we have $|F(z) - \Phi(z)| \leq O(\tau/\sigma)/(1 + |z|^3)$. It follows that (46) $\leq (A) + (B)$, where

$$(A) = \int_0^\infty 1 - \Phi((u+s)/\sigma) + \Phi((u-s)/\sigma)\, ds$$

and

$$(B) \quad = \quad O(\tau/\sigma) \cdot \int_0^\infty \left(\frac{1}{1+|(u+s)/\sigma|^3} + \frac{1}{1+|(u-s)/\sigma|^3} \right) ds.$$

It is easy to see that

$$(B) = O(\tau/\sigma) \cdot \int_{-\infty}^\infty \frac{1}{1+|x/\sigma|^3}\, dx = O(\tau).$$

For (A), observe that (A) can be re-expressed as

$$\int_0^\infty \Pr[X > (u+s)/\sigma] + \Pr[X < (u-s)/\sigma] ds = \int_0^\infty \Pr[|\sigma X - u| > s]\, ds.$$

Again using the fact that $\mathbf{E}[R] = \int_0^\infty \Pr[R > s]\, ds$ for any nonnegative random variable R for which $\mathbf{E}[R] < \infty$, this equals $\mathbf{E}[|\sigma X - u|]$. This gives the desired bound. $\square$

Proposition 59. *Using the notation above,* $\mathbf{E}[|\sigma X - u|] = \sigma \cdot 2\phi(u/\sigma) - u\mu(u/\sigma)$. *(This remains sensible even for* $\sigma = 0$.*)*

Proof.

$$\mathbf{E}[|\sigma X - u| = \mathbf{E}[\mathrm{sgn}(\sigma X - u)(\sigma X - u)] = \sigma \hat{g}(1) - u\, \mathbf{E}[g],$$

where $g : \mathbf{R} \to \mathbf{R}$ is the function $g(X) = \mathrm{sgn}(X - u/\sigma)$. But $\mathbf{E}[g] = \mu(u/\sigma)$ and $\hat{g}(1) = 2\phi(u/\sigma)$ (see Definition 17). $\square$

A.2 LTF basics.

We collect here some easy propositions about LTFs. First, we need to recall the general notion of "influences" for Boolean functions:

Definition 60. *Given* $f : \{-1,1\}^n \to \{-1,1\}$ *and* $i \in [n]$, *the* influence *of variable* i *is defined as* $\mathrm{Inf}_i(f) = \Pr_x[f(x^{i-}) \neq f(x^{i+})]$, *where* x^{i-} *and* x^{i+} *denote* x *with the* i*'th bit set to* -1 *or* 1 *respectively.*

It is well-known that if f is a unate function then $\mathrm{Inf}_i(f) = |\hat{f}(i)|$. In particular, this holds for LTFs (which are unate).

The next proposition, relating the rank of the weights to the rank of the influences/degree-1 Fourier coefficients, is very elementary; an explicit proof appears in, e.g., [FP04].

Proposition 61. *Let* $f = \mathrm{sgn}(w_1x_1 + \cdots + w_nx_n - \theta)$ *be an LTF such that* $|w_1| \geq |w_i|$ *for all* $i \in [n]$. *Then* $|\mathrm{Inf}_1(f)| \geq |\mathrm{Inf}_i(f)|$ *for all* $i \in [n]$.

Next, we show that LTFs typically become regular when their most influential coordinates are restricted:

Proposition 62. *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *be an LTF and let* $J \supseteq \{j : |\hat{f}(i)| \geq \beta\}$. *Then* f_π *is not* (β/η)*-regular for at most an* η *fraction of all restrictions* π *to* J.

Proof. Since f is an LTF, $|\hat{f}(j)| = \mathrm{Inf}_j(f)$; thus every coordinate outside J has influence at most β on f. Let k be a coordinate outside of J of maximum influence. Note that since f is an LTF, k is a coordinate of maximum influence for f_π *under every restriction* π *to* J; this follows from Proposition 61. But $\mathrm{Inf}_k(f) = \mathrm{Avg}_\pi(\mathrm{Inf}_k(f_\pi)) = \mathrm{Avg}_\pi(|\widehat{f_\pi}(k)|)$ and so

$$\beta \geq \mathrm{Inf}_k(f) = \mathrm{Avg}_\pi(\text{regularity of } f_\pi).$$

The result now follows by Markov's inequality. $\square$

Next, a proposition on LTFs that are close to juntas:

Proposition 63. *Let* $f = \mathrm{sgn}(w_1x_1 + \cdots + w_nx_n - \theta)$ *be an LTF which is* ϵ*-close to being some junta on the set* J. *Then* f *is in fact* ϵ*-close to being the LTF on* J *given by* $\mathrm{sgn}(\sum_{i\in J} w_ix_i - \theta)$.

Proof. Assume without loss of generality that $J = \{1, \ldots, r\}$. Given any values for $x_1, \ldots, x_r$, let $f_{x_1,\ldots,x_r}$ denote the restricted version of f, a function of the remaining variables $x_{r+1}, \ldots, x_n$. Now without even using the fact that f is an LTF, we know that the junta over $\{-1, 1\}^r$ to which f is closest is given by mapping $x_1, \ldots, x_r$ to the more common value of $f_{x_1,\ldots,x_r}$. But this more common value is certainly $\mathrm{sgn}(w_1x_1 + \cdots + w_rx_r - \theta)$, by the symmetry of the variables $x_{r+1}, \ldots, x_n$. This completes the proof. □

Finally, we show a partial converse to our Theorem 26:

Proposition 64. *Suppose* $f(x) = \mathrm{sgn}(a_1x_1 + \cdots + a_nx_n - \theta)$ *is an LTF with* $\sum_{i=1}^n a_i^2 = 1$ *and* $|a_i| \leq \delta$ *for all* i. *Then* f *is* $O(\delta)$*-regular; i.e.,* $\mathrm{Inf}_i(f) \leq O(\delta)$ *for all* i.

Proof. Without loss of generality we may assume that $\delta = |a_1| \geq |a_i|$ for all i. By Proposition 61 we need to show that $\mathrm{Inf}_1(f) \leq O(\delta)$. Now observe that

$$\mathrm{Inf}_1(f) = \Pr\big[|a_2x_2 + \cdots + a_nx_n - \theta| \leq \delta\big].$$

If $\delta \geq 1/2$ then clearly $\mathrm{Inf}_1(f) \leq 2\delta$ so we may assume $\delta < 1/2$. By the Berry-Esseen theorem, the probability (6) above is within an additive $O(\delta/\sqrt{1-\delta^2}) = O(\delta)$ of the probability that $|X - \theta| \leq \delta$, where X is a mean-zero Gaussian with variance $1 - \delta^2$. This latter probability is at most $O(\delta/\sqrt{1-\delta^2}) = O(\delta)$, so indeed we have $\mathrm{Inf}_1(f) \leq O(\delta)$. □

ECCC ISSN 1433-8092
http://eccc.hpi-web.de/