A robust Khintchine inequality, and algorithms for computing optimal constants in Fourier analysis and high-dimensional geometry *

Anindya De[1**], Ilias Diakonikolas[2***], and Rocco Servedio[3†]

[1] UC Berkeley anindya@cs.berkeley.edu
[2] University of Edinburgh ilias.d@ed.ac.uk
[3] Columbia University rocco@cs.columbia.edu

Abstract. This paper makes two contributions towards determining some well-studied optimal constants in Fourier analysis of Boolean functions and high-dimensional geometry.

1. It has been known since 1994 [GL94] that every linear threshold function has squared Fourier mass at least $1/2$ on its degree-0 and degree-1 coefficients. Denote the minimum such Fourier mass by $\mathbf{W}^{\leq 1}[\mathbf{LTF}]$, where the minimum is taken over all n-variable linear threshold functions and all $n \geq 0$. Benjamini, Kalai and Schramm [BKS99] have conjectured that the true value of $\mathbf{W}^{\leq 1}[\mathbf{LTF}]$ is $2/\pi$. We make progress on this conjecture by proving that $\mathbf{W}^{\leq 1}[\mathbf{LTF}] \geq 1/2+c$ for some absolute constant $c > 0$. The key ingredient in our proof is a "robust" version of the well-known Khintchine inequality in functional analysis, which we believe may be of independent interest.
2. We give an algorithm with the following property: given any $\eta > 0$, the algorithm runs in time $2^{\mathrm{poly}(1/\eta)}$ and determines the value of $\mathbf{W}^{\leq 1}[\mathbf{LTF}]$ up to an additive error of $\pm\eta$. We give a similar $2^{\mathrm{poly}(1/\eta)}$-time algorithm to determine *Tomaszewski's constant* to within an additive error of $\pm\eta$; this is the minimum (over all origin-centered hyperplanes H) fraction of points in $\{-1,1\}^n$ that lie within Euclidean distance 1 of H. Tomaszewski's constant is conjectured to be $1/2$; lower bounds on it have been given by Holzman and Kleitman [HK92] and independently by Ben-Tal, Nemirovski and Roos [BTNR02]. Our algorithms combine tools from anti-concentration of sums of independent random variables, Fourier analysis, and Hermite analysis of linear threshold functions.

1 Introduction

This paper is inspired by a belief that simple mathematical objects should be well understood. We study two closely related kinds of simple objects: n-dimensional linear threshold functions $f(x) = \mathrm{sign}(w \cdot x - \theta)$, and n-dimensional origin-centered hyperplanes $H = \{x \in \mathbb{R}^n : w \cdot x = 0\}$. Benjamini, Kalai and Schramm [BKS99] and

* A full version of this paper may be found at http://arxiv.org/abs/1207.2229.

** Research supported by NSF award CCF-1118083.

*** Research supported by a Simons Foundation Postdoctoral Fellowship.

† Research supported in part by NSF awards CCF-0915929 and CCF-1115703.

Tomaszewski [Guy86] have posed the question of determining two universal constants related to halfspaces and origin-centered hyperplanes respectively; we refer to these quantities as "the BKS constant" and "Tomaszewski's constant." While these constants arise in various contexts including uniform-distribution learning and optimization theory, little progress has been made on determining their actual values over the past twenty years. In both cases there is an easy upper bound which is conjectured to be the correct value; Gotsman and Linial [GL94] gave the best previously known lower bound on the BKS constant in 1994, and Holzmann and Kleitman [HK92] gave the best known lower bound on Tomaszewski's constant in 1992.

We give two main results. The first of these is an improved lower bound on the BKS constant; a key ingredient in the proof is a "robust" version of the well-known Khintchine inequality, which we believe may be of independent interest. Our second main result is a pair of algorithms for computing the BKS constant and Tomaszewski's constant up to any prescribed accuracy. The first algorithm, given any $\eta > 0$, runs in time $2^{\mathrm{poly}(1/\eta)}$ and computes the BKS constant up to an additive η, and the second algorithm runs in time $2^{\mathrm{poly}(1/\eta)}$ and has the same guarantee for Tomaszewski's constant.

1.1 Background and problem statements.

First problem: low-degree Fourier weight of linear threshold functions. A *linear threshold function*, henceforth denoted simply LTF, is a function $f : \{-1,1\}^n \to \{-1,1\}$ of the form $f(x) = \mathrm{sign}(w \cdot x - \theta)$ where $w \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$ (the univariate function $\mathrm{sign} : \mathbb{R} \to \mathbb{R}$ is $\mathrm{sign}(z) = 1$ for $z \geq 0$ and $\mathrm{sign}(z) = -1$ for $z < 0$). The values $w_1, \dots, w_n$ are the *weights* and θ is the *threshold.* Linear threshold functions play a central role in many areas of computer science such as concrete complexity theory and machine learning, see e.g. [DGJ$^+$10] and the references therein.

It is well known [BKS99,Per04] that LTFs are highly noise-stable, and hence they must have a large amount of Fourier weight at low degrees. For $f : \{-1,1\}^n \to \mathbb{R}$ and $k \in [0,n]$ let us define $\mathbf{W}^k[f] = \sum_{S \subseteq [n], |S|=k} \widehat{f}^2(S)$ and $\mathbf{W}^{\leq k}[f] = \sum_{j=0}^{k} \mathbf{W}^j[f]$; we will be particularly interested in the Fourier weight of LTFs at levels 0 and 1. More precisely, for $n \in \mathbf{N}$ let $\mathbf{LTF}_n$ denote the set of all n-dimensional LTFs, and let $\mathbf{LTF} = \cup_{n=1}^{\infty} \mathbf{LTF}_n$. We define the following universal constant:

Definition 1. *Let* $\mathbf{W}^{\leq 1}[\mathbf{LTF}] \stackrel{def}{=} \inf_{h \in \mathbf{LTF}} \mathbf{W}^{\leq 1}(h) = \inf_{n \in \mathbf{N}} \mathbf{W}^{\leq 1}[\mathbf{LTF}_n]$, *where* $\mathbf{W}^{\leq 1}[\mathbf{LTF}_n] \stackrel{def}{=} \inf_{h \in \mathbf{LTF}_n} \mathbf{W}^{\leq 1}(h)$.

Benjamini, Kalai and Schramm (see Remark 3.7 of [BKS99]) and subsequently O'Donnell (see the Conjecture following Theorem 2 of Section 5.1 of [O'D12]) have conjectured that $\mathbf{W}^{\leq 1}[\mathbf{LTF}] = 2/\pi$, and hence we will sometimes refer to $\mathbf{W}^{\leq 1}[\mathbf{LTF}]$ as "the BKS constant." As $n \to \infty$, a standard analysis of the n-variable Majority function shows that $\mathbf{W}^{\leq 1}[\mathbf{LTF}] \leq 2/\pi$. Gotsman and Linial [GL94] observed that $\mathbf{W}^{\leq 1}[\mathbf{LTF}] \geq 1/2$ but until now no better lower bound was known. We note that since the universal constant $\mathbf{W}^{\leq 1}[\mathbf{LTF}]$ is obtained by taking the infimum over an infinite set, it is not *a priori* clear whether the computational problem of computing or even approximating $\mathbf{W}^{\leq 1}[\mathbf{LTF}]$ is decidable.

Jackson [Jac06] has shown that improved lower bounds on $\mathbf{W}^{\leq 1}[\mathbf{LTF}]$ translate directly into improved noise-tolerance bounds for agnostic weak learning of LTFs in the "Restricted Focus of Attention" model of Ben-David and Dichterman [BDD98]. Further motivation for studying $\mathbf{W}^{\leq 1}[f]$ comes from the fact that $\mathbf{W}^1[f]$ is closely related to the noise stability of f (see [O'D12]). In particular, if $\mathbf{NS}_\rho[f]$ represents the noise stability of f when the noise rate is $(1-\rho)/2$, then it is known that $\frac{d\mathbf{NS}_\rho[f]}{d\rho}\Big|_{\rho=0} = \mathbf{W}^1[f]$. This means that for a function f with $\mathbf{E}[f]=0$, we have $\mathbf{NS}_\rho[f] \to \rho \cdot \mathbf{W}^{\leq 1}[f]$ as $\rho \to 0$. Thus, at very large noise rates, $\mathbf{W}^1[f]$ quantifies the size of the "noisy boundary" of mean-zero functions f.

Second problem: how many hypercube points have distance at most 1 from an origin-centered hyperplane? For $n \in \mathbb{N}$ and $n > 1$, let $\mathbb{S}^{n-1}$ denote the n-dimensional sphere $\mathbb{S}^{n-1} = \{w \in \mathbb{R}^n : \|w\|_2 = 1\}$, and let $\mathbb{S} = \cup_{n>1}\mathbb{S}^{n-1}$. Each unit vector $w \in \mathbb{S}^{n-1}$ defines an origin-centered hyperplane $H_w = \{x \in \mathbb{R}^n : w \cdot x = 0\}$. Given a unit vector $w \in \mathbb{S}^{n-1}$, we define $\mathbf{T}(w) \in [0,1]$ to be $\mathbf{T}(w) = \mathbf{Pr}_{x\in\{-1,1\}^n}[|w \cdot x| \leq 1]$, the fraction of hypercube points in $\{-1,1\}^n$ that lie within Euclidean distance 1 of the hyperplane H_w. We define the following universal constant, which we call "Tomaszewski's constant:"

Definition 2. *Define* $\mathbf{T}(\mathbb{S}) \stackrel{def}{=} \inf_{w\in\mathbb{S}} \mathbf{T}(w) = \inf_{n\in\mathbb{N}} \mathbf{T}(\mathbb{S}^{n-1})$, *where* $\mathbf{T}(\mathbb{S}^{n-1}) \stackrel{def}{=} \inf_{w\in\mathbb{S}^{n-1}} \mathbf{T}(w)$.

Tomaszewski [Guy86] has conjectured that $\mathbf{T}(\mathbb{S}) = 1/2$. The main result of Holzman and Kleitman [HK92] is a proof that $3/8 \leq \mathbf{T}(\mathbb{S})$; the upper bound $\mathbf{T}(\mathbb{S}) \leq 1/2$ is witnessed by the vector $w = (1/\sqrt{2}, 1/\sqrt{2})$. As noted in [HK92], the quantity $\mathbf{T}(\mathbb{S})$ has a number of appealing geometric and probabilistic reformulations. Similar to the BKS constant, since $\mathbf{T}(\mathbb{S})$ is obtained by taking the infimum over an infinite set, it is not immediately evident that any algorithm can compute or approximate $\mathbf{T}(\mathbb{S})$. [4]

An interesting quantity in its own right, Tomaszewski's constant also arises in a range of contexts in optimization theory, see e.g. [So09,BTNR02]. In fact, the latter paper proves a lower bound of $1/3$ on the value of Tomaszewski's constant independently of [HK92], and independently conjectures that the optimal lower bound is $1/2$.

1.2 Our results.

A better lower bound for the BKS constant $\mathbf{W}^{\leq 1}[\mathbf{LTF}]$. Our first main result is the following theorem:

Theorem 1 (Lower Bound for the BKS constant). *There exists a universal constant* $c' > 0$ *such that* $\mathbf{W}^{\leq 1}[\mathbf{LTF}] \geq \frac{1}{2} + c'$.

This is the first improvement on the [GL94] lower bound of $1/2$ since 1994. We actually give two quite different proofs of this theorem, which are sketched in the "Techniques" subsection below.

[4] Whenever we speak of "an algorithm to compute or approximate" one of these constants, of course what we really mean is an algorithm that outputs the desired value *together with a proof of correctness of its output value*.

An algorithm for approximating the BKS constant $\mathbf{W}^{\leq 1}[\mathbf{LTF}]$. Our next main result shows that in fact there *is* a finite-time algorithm that approximates the BKS constant up to any desired accuracy:

Theorem 2 (Approximating the BKS constant). *There is an algorithm that, on input an accuracy parameter $\epsilon > 0$, runs in time $2^{\mathrm{poly}(1/\epsilon)}$ and outputs a value Γ_ϵ such that*

$$\mathbf{W}^{\leq 1}[\mathbf{LTF}] \leq \Gamma_\epsilon \leq \mathbf{W}^{\leq 1}[\mathbf{LTF}] + \epsilon. \tag{1}$$

An algorithm for approximating Tomaszewski's constant $\mathbf{T}(\mathbb{S})$. Our final main result is an algorithm that approximates $\mathbf{T}(\mathbb{S})$ up to any desired accuracy:

Theorem 3 (Approximating Tomaszewski's constant). *There is an algorithm that, on input $\epsilon > 0$, runs in time $2^{\mathrm{poly}(1/\epsilon)}$ and outputs a value Γ_ϵ such that*

$$\mathbf{W}^{\leq 1}[\mathbf{LTF}] \leq \Gamma_\epsilon \leq \mathbf{W}^{\leq 1}[\mathbf{LTF}] + \epsilon. \tag{2}$$

1.3 Our techniques for Theorem 1: lower-bounding the BKS constant

It is easy to show that it suffices to consider the level-1 Fourier weight $\mathbf{W}^1$ of LTFs that have threshold $\theta = 0$ and have $w \cdot x \neq 0$ for all $x \in \{-1,1\}^n$, so we confine our discussion to such zero-threshold LTFs. To explain our approaches to lower bounding $\mathbf{W}^{\leq 1}[\mathbf{LTF}]$, we recall the essentials of the simple argument of [GL94] that gives a lower bound of $1/2$. The key ingredient of their argument is the well-known Khintchine inequality from functional analysis:

Definition 3. *For a unit vector $w \in \mathbb{S}^{n-1}$ we define $\mathbf{K}(w) \overset{def}{=} \mathbf{E}_{x\in\{-1,1\}^n}[|w \cdot x|]$ to be the "Khintchine constant for w."*

The following is a classical theorem in functional analysis (we write e_i to denote the unit vector in $\mathbb{R}^n$ with a 1 in coordinate i):

Theorem 4 (Khintchine inequality, [Sza76]). *For $w \in \mathbb{S}^n$ any unit vector, we have $\mathbf{K}(w) \geq 1/\sqrt{2}$, with equality holding if and only if $w = \frac{1}{\sqrt{2}}(\pm e_i \pm e_j)$ for some $i \neq j \in [n]$.*

Szarek [Sza76] was the first to obtain the optimal constant $1/\sqrt{2}$, and subsequently several simplifications of his proof were given [Haa82,Tom87,LO94]; we shall give a simple self-contained proof in Section 2.1 below, which is quite similar to Filmus's [Fil12] translation of the [LO94] proof into "Fourier language." With Theorem 4 in hand, the Gotsman-Linial lower bound is almost immediate:

Proposition 1 ([GL94]). *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a zero-threshold LTF $f(x) = \mathrm{sign}(w \cdot x)$ where $w \in \mathbb{R}^n$ has $\|w\|_2 = 1$. Then $\mathbf{W}^1[f] \geq (\mathbf{K}(w))^2$.*

Proof. We have that $\mathbf{K}(w) = \mathbf{E}_x[f(x)(w\cdot x)] = \sum_{i=1}^n \widehat{f}(i) w_i \leq \sqrt{\sum_{i=1}^n \widehat{f}^2(i)} \cdot \sqrt{\sum_{i=1}^n w_i^2} = \sqrt{\mathbf{W}^1[f]}$ where the first equality uses the definition of f, the second is Plancherel's identity, the inequality is Cauchy-Schwarz, and the last equality uses the assumption that w is a unit vector. □

First proof of Theorem 1: A "robust" Khintchine inequality. Given the strict condition required for equality in the Khintchine inequality, it is natural to expect that if a unit vector $w \in \mathbb{R}^n$ is "far" from $\frac{1}{\sqrt{2}}(\pm e_i \pm e_j)$, then $\mathbf{K}(w)$ should be significantly larger than $1/\sqrt{2}$. We prove a robust version of the Khintchine inequality which makes this intuition precise. Given a unit vector $w \in \mathbb{S}^{n-1}$, define $d(w)$ to be $d(w) = \min \|w - w^*\|_2$, where w^* ranges over all $4\binom{n}{2}$ vectors of the form $\frac{1}{\sqrt{2}}(\pm e_i \pm e_j)$. Our "robust Khintchine" inequality is the following:

Theorem 5 (Robust Khintchine inequality). *There exists a universal constant $c > 0$ such that for any $w \in \mathbb{S}^{n-1}$, we have $\mathbf{K}(w) \geq \frac{1}{\sqrt{2}} + c \cdot d(w)$.*

Armed with our robust Khintchine inequality, the simple proof of Proposition 1 suggests a natural approach to lower-bounding $\mathbf{W}^{\leq 1}[\mathbf{LTF}]$. If w is such that $d(w)$ is "large" (at least some absolute constant), then the statement of Proposition 1 immediately gives a lower bound better than $1/2$. So the only remaining vectors w to handle are highly constrained vectors which are almost exactly of the form $\frac{1}{\sqrt{2}}(\pm e_i \pm e_j)$. A natural hope is that the Cauchy-Schwarz inequality in the proof of Proposition 1 is not tight for such highly constrained vectors, and indeed this is essentially how we proceed (modulo some simple cases in which it is easy to bound $\mathbf{W}^{\leq 1}$ above $1/2$ directly).

Second proof of Theorem 1: anticoncentration, Fourier analysis of LTFs, and LTF approximation. Our second proof of Theorem 1 employs several sophisticated ingredients from recent work on structural properties of LTFs [OS11,MORS10]. The first of these ingredients is a result (Theorem 6.1 of [OS11]) which essentially says that any LTF $f(x) = \mathrm{sign}(w \cdot x)$ can be perturbed very slightly to another LTF $f'(x) = \mathrm{sign}(w' \cdot x)$ (where both w and w' are unit vectors). The key properties of this perturbation are that (i) f and f' are extremely close, differing only on a tiny fraction of inputs in $\{-1,1\}^n$; but (ii) the linear form $w' \cdot x$ has some nontrivial "anti-concentration" when x is distributed uniformly over $\{-1,1\}^n$, meaning that very few inputs have $w' \cdot x$ very close to 0.

Why is this useful? It turns out that the anti-concentration of $w' \cdot x$, together with results on the degree-1 Fourier spectrum of "regular" halfspaces from [MORS10], lets us establish a lower bound on $\mathbf{W}^{\leq 1}[f']$ that is strictly greater than $1/2$. Then the fact that f and f' agree on almost every input in $\{-1,1\}^n$ lets us argue that the original LTF f must similarly have $\mathbf{W}^{\leq 1}[f]$ strictly greater than $1/2$. Interestingly, the lower bound on $\mathbf{W}^{\leq 1}[f']$ is proved using the Gotsman-Linial inequality $\mathbf{W}^{\leq 1}[f'] \geq (\mathbf{K}(w))^2$; in fact, the anti-concentration of $w' \cdot x$ is combined with ingredients in the simple Fourier proof of the (original, non-robust) Khintchine inequality (specifically, an upper bound on the total influence of the function $\ell(x) = |w' \cdot x|$) to obtain the result. Because of space constraints we give this second proof in the full version of the paper, following the references.

1.4 Our techniques for Theorem 2: approximating the BKS constant

As in the previous subsection, it suffices to consider only zero-threshold LTFs $\mathrm{sign}(w \cdot x)$. Our algorithm turns out to be very simple (though its analysis is not):

Let $K = \Theta(\epsilon^{-24})$. Enumerate all K-variable zero-threshold LTFs, and output the value $\Gamma_\epsilon \stackrel{\text{def}}{=} \min\{\mathbf{W}^1[f] : f$ is a zero-threshold K-variable LTF.$\}$.

It is well known (see e.g. [MT94]) that there exist $2^{\Theta(K^2)}$ distinct K-variable LTFs, and it is straightforward to confirm that they can be enumerated in output-polynomial time. Thus the above simple algorithm runs in time $2^{\mathrm{poly}(1/\epsilon)}$; the challenge is to show that the value Γ_ϵ thus obtained indeed satisfies Equation (2).

A key ingredient in our analysis is the notion of the "critical index" of an LTF f. The critical index was implicitly introduced and used in [Ser07] and was explicitly used in [DS09,DGJ$^+$10,OS11,DDFS12] and other works. To define the critical index we need to first define "regularity":

Definition 4 (regularity). *Fix $\tau > 0$. We say that a vector $w = (w_1, \dots, w_n) \in \mathbb{R}^n$ is* τ-regular *if* $\max_{i\in[n]} |w_i| \le \tau\|w\| = \tau\sqrt{w_1^2 + \cdots + w_n^2}$. *A linear form $w \cdot x$ is said to be τ-regular if w is τ-regular, and similarly an LTF is said to be τ-regular if it is of the form* $\mathrm{sign}(w \cdot x - \theta)$ *where w is τ-regular.*

Regularity is a helpful notion because if w is τ-regular then the Berry-Esséen theorem tells us that for uniform $x \in \{-1,1\}^n$, the linear form $w \cdot x$ is "distributed like a Gaussian up to error τ." This can be useful for many reasons (as we will see below).

Intuitively, the critical index of w is the first index i such that from that point on, the vector $(w_i, w_{i+1}, \dots, w_n)$ is regular. A precise definition follows:

Definition 5 (critical index). *Given a vector $w \in \mathbb{R}^n$ such that $|w_1| \ge \cdots \ge |w_n| > 0$, for $k \in [n]$ we denote by σ_k the quantity $\sqrt{\sum_{i=k}^n w_i^2}$. We define the* τ-critical index $c(w,\tau)$ of w *as the smallest index $i \in [n]$ for which $|w_i| \le \tau \cdot \sigma_i$. If this inequality does not hold for any $i \in [n]$, we define $c(w,\tau) = \infty$.*

Returning to Theorem 2, since our algorithm minimizes over a proper subset of all LTFs, it suffices to show that for any zero-threshold LTF $f = \mathrm{sign}(w \cdot x)$, there is a K-variable zero-threshold LTF g such that $\mathbf{W}^1[g] - \mathbf{W}^1[f] < \epsilon$. At a high level our proof is a case analysis based on the size of the δ-critical index $c(w,\delta)$ of the weight vector w, where we choose the parameter δ to be $\delta = \mathrm{poly}(\epsilon)$. The first case is relatively easy: if the δ-critical index is large, then it is known that the function f is very close to some K-variable LTF g. Since the two functions agree almost everywhere, it is easy to show that $|\mathbf{W}^1[f] - \mathbf{W}^1[g]| \le \epsilon$ as desired.

The case that the critical index is small is much more challenging. In this case it is by no means true that f can be well approximated by an LTF on few variables – consider, for example, the majority function. We deal with this challenge by developing a novel *variable reduction technique* which lets us construct a $\mathrm{poly}(1/\epsilon)$-variable LTF g whose level-1 Fourier weight closely matches that of f.

How is this done? The answer again comes from the critical index. Since the critical index $c(w,\delta)$ is small, we know that except for the "head" portion $\sum_{i=1}^{c(w,\delta)-1} w_i x_i$ of the linear form, the "tail" portion $\sum_{i=c(w,\delta)}^n w_i x_i$ of the linear form "behaves like a Gaussian." Guided by this intuition, our variable reduction technique proceeds in three steps. In the first step, we replace the tail coordinates $x_T = (x_{c(w,\delta)}, \dots, x_n)$ by independent Gaussian random variables and show that the degree-1 Fourier weight of the corresponding "mixed" function (which has some ± 1-valued inputs and some Gaussian inputs) is approximately equal to $\mathbf{W}^1[f]$. In the second step, we replace the tail random variable $w_T \cdot G_T$, where G_T is the vector of Gaussians from the first step, by a *single* Gaussian random variable G, where $G \sim \mathcal{N}(0, \|w_T\|^2)$. We show that this transformation exactly preserves the degree-1 weight. At this point we have reduced the number of variables from n down to $c(w, \delta)$ (which is small in this case!), but the last variable is Gaussian rather than Boolean. As suggested by the Central Limit Theorem, though, one may try to replace this Gaussian random variable by a normalized sum of independent ± 1 random variables $\sum_{i=1}^{M} z_i/\sqrt{M}$. This is exactly the third step of our variable reduction technique. Via a careful analysis, we show that by taking $M = \mathrm{poly}(1/\epsilon)$, this operation preserves the degree-1 weight up to an additive ϵ. Combining all these steps, we obtain the desired result.

1.5 Our techniques for Theorem 3: approximating Tomaszewski's constant

The first step of our proof of Theorem 3 is similar in spirit to the main structural ingredient of our proof of Theorem 2: we show that given any $\epsilon > 0$, there is a value $K_\epsilon = \mathrm{poly}(1/\epsilon)$ such that it suffices to consider linear forms $w \cdot x$ over K_ϵ-dimensional space, i.e. for any $n \in \mathbf{N}$ we have $\mathbf{T}(\mathbb{S}^{n-1}) \le \mathbf{T}(\mathbb{S}^{K_\epsilon - 1}) \le \mathbf{T}(\mathbb{S}^{n-1}) + \epsilon$. Similar to the high-level outline of Theorem 2, our proof of again proceeds by fixing any $w \in \mathbb{S}^{n-1}$ and doing a case analysis based on whether the critical index of w is "large" or "small." However, the technical details of each of these cases is quite different from the earlier proof. In the "small critical index" case we employ Gaussian anti-concentration (which is inherited by the "tail" random variable $w_T x_T$ since the tail vector w_T is regular), and in the "large critical index" case we use an anti-concentration result from [OS11].

Unlike the previous situation for the BKS constant, at this point more work remains to be done for approximating Tomaszewski's constant. While there are only $2^{\mathrm{poly}(1/\epsilon)}$ many halfspaces over $\mathrm{poly}(1/\epsilon)$ many variables and hence a brute-force enumeration could cover all of them in $2^{\mathrm{poly}(1/\epsilon)}$ time for the BKS constant, here we must contend with the fact that $\mathbb{S}^{K_\epsilon - 1}$ is an uncountably infinite set, so we cannot naively minimize over all its elements. Instead we take a dual approach and exploit the fact that while there are uncountably infinitely many vectors in $\mathbb{S}^{K_\epsilon - 1}$, there are only 2^{K_ϵ} many hypercube points in $\{-1, 1\}^{K_\epsilon}$, and (with some care) the desired infimum over all unit vectors can be formulated in the language of existential theory of the reals. We then use an algorithm for deciding existential theory of the reals (see [Ren88]) to compute the infimum. Because of space constraints we prove Theorem 3 in the full version of the paper, following the references.

Discussion. It is interesting to note that determining Tomaszewski's constant is an instance of the well-studied generic problem of understanding tails of Rademacher sums. For the sake of discussion, let us define $\mathbf{T}_{\mathsf{in}}(w, a) = \mathbf{Pr}_{x \in \{-1,1\}^n}[|w \cdot x| \le a]$ and $\mathbf{T}_{\mathsf{out}}(w, a) = \mathbf{Pr}_{x \in \{-1,1\}^n}[|w \cdot x| \ge a]$ where $w \in \mathbb{S}^{n-1}$. Further, let $\mathbf{T}_{\mathsf{in}}(a) = \inf_{w \in \mathbb{S}} \mathbf{T}_{\mathsf{in}}(w, a)$ and $\mathbf{T}_{\mathsf{out}}(a) = \inf_{w \in \mathbb{S}} \mathbf{T}_{\mathsf{out}}(w, a)$. Note that Tomaszewski's constant $\mathbf{T}(\mathbb{S})$ is simply $\mathbf{T}_{\mathsf{in}}(1)$. Much effort has been expended on getting sharp estimates for $\mathbf{T}_{\mathsf{in}}(a)$ and $\mathbf{T}_{\mathsf{out}}(a)$ for various values of a (see e.g. [Pin12,Ben04]). As a representative example, Bentkus and Dzindzalieta [BD12] proved that $\mathbf{T}_{\mathsf{in}}(a) \ge \frac{1}{4} + \frac{1}{4} \cdot \sqrt{2 - \frac{2}{a^2}}$ for

$a \in (1, \sqrt{2}]$. Similarly, Pinelis [Pin94] showed that there is an absolute constant $c > 0$ such that $\mathbf{T}_{\mathsf{out}}(a) \geq 1 - c \cdot \frac{\phi(a)}{a}$ where $\phi(x)$ is the density function of the standard normal $\mathcal{N}(0,1)$ (note this beats the standard Hoeffding bound by a factor of $1/a$).

On the complementary side, Montgomery-Smith [MS90] proved that there is an absolute constant $c' > 0$ such that $\mathbf{T}_{\mathsf{out}}(a) \geq e^{-c' \cdot a^2}$ for all $a \leq 1$. Similarly, Oleszkiewicz [Ole96] proved that $\mathbf{T}_{\mathsf{out}}(1) \geq 1/10$. The conjectured lower bound on $\mathbf{T}_{\mathsf{out}}(1)$ is $7/32$ (see [HK94]). While we have not investigated this in detail, we suspect that our techniques may be applicable to some of the above problems. Finally, we note that apart from being of intrinsic interest to functional analysts and probability theorists, the above quantities arise frequently in the optimization literature (see [HLNZ08,BTNR02]). Related tail bounds have also found applications in extremal combinatorics (see [AHS12]).

Mathematical Preliminaries. These are given in Section 2 of the full version.

2 Proof of Theorem 5: A "robust" Khintchine inequality

It will be convenient for us to reformulate Theorems 4 and 5 as follows: Let us say that a unit vector $w = (w_1, \ldots, w_n) \in \mathbb{S}^{n-1}$ is *proper* if $w_i \geq w_{i+1} \geq 0$ for all $i \in [n-1]$. Then we may state the "basic" Khintchine inequality with optimal constant, Theorem 4, in the following equivalent way:

Theorem 6 (Khintchine inequality, [Sza76]). *Let $w \in \mathbb{R}^n$ be a proper unit vector. Then $\mathbf{K}(w) \geq 1/\sqrt{2}$, with equality if and only if $w = w^* \stackrel{def}{=} (1/\sqrt{2}, 1/\sqrt{2}, 0, \ldots, 0)$.*

And we may restate our "robust" Khintchine inequality, Theorem 5, as follows:

Theorem 7 (Robust Khintchine inequality). *There exists a universal constant $c > 0$ such that the following holds: Let $w \in \mathbb{R}^n$ be a proper unit vector. Then $\mathbf{K}(w) \geq 1/\sqrt{2} + c \cdot \|w - w^*\|_2$, where $w^* \stackrel{def}{=} (1/\sqrt{2}, 1/\sqrt{2}, 0, \ldots, 0)$.*

Before we proceed with the proof of Theorem 7, we give a simple Fourier analytic proof of the "basic" Khintchine inequality with optimal constant, $\mathbf{K}(w) \geq 1/\sqrt{2}$. (We note that this is a well-known argument by now; it is given in somewhat more general form in [Ole99] and in [KLO96].) We then build on this to prove Theorem 7.

2.1 Warm-up: simple proof that $\mathbf{K}(w) \geq 1/\sqrt{2}$

We consider the function $\ell(x) = |\sum_{i=1}^n w_i x_i|$ where $\sum_i w_i^2 = 1$ and will show that $\mathbf{K}(w) = \mathbf{E}_x[\ell(x)] \geq 1/\sqrt{2}$. Noting that $\mathbf{E}_x[(\ell(x))^2] = 1$, we have $(\mathbf{E}[\ell(x)])^2 = 1 - \mathrm{Var}[\ell]$, so it suffices to show that $\mathrm{Var}[\ell] \leq 1/2$. This follows directly by combining the following claims. The first bound is an improved Poincaré inequality for even functions:

Fact 8 *(Poincaré inequality) Let $f : \{-1,1\}^n \to \mathbb{R}$ be even. Then $\mathrm{Var}[f] \leq (1/2) \cdot \mathrm{Inf}(f)$.*

Proof. Since f is even, we have that $\widehat{f}(S) = 0$ for all S with odd $|S|$. We can thus write

$$\mathrm{Inf}(f) = \sum_{S \subseteq [n], |S| \text{ even}} |S| \cdot \widehat{f}^2(S) \geq 2 \cdot \sum_{\emptyset \neq S, |S| \text{ even}} \widehat{f}^2(S) = 2 \cdot \sum_{\emptyset \neq S \subseteq [n]} \widehat{f}^2(S) = 2 \cdot \mathrm{Var}[f].$$

The second is an upper bound on the influences in ℓ as a function of the weights:

Lemma 1. *Let $\ell(x) = |\sum_{i=1}^n w_i x_i|$. For any $i \in [n]$, we have $\mathrm{Inf}_i(\ell) \le w_i^2$.*

Proof. Recall that $\mathrm{Inf}_i(\ell) = \mathbf{E}_x\left[\mathrm{Var}_{x_i}\left[\ell(x)\right]\right] = \mathbf{E}_x\left[\mathbf{E}_{x_i}[\ell^2(x)] - (\mathbf{E}_{x_i}[\ell(x)])^2\right]$. We claim that for any $x \in \{-1,1\}^n$, it holds that $\mathrm{Var}_{x_i}[\ell(x)] \le w_i^2$, which yields the lemma. To show this claim we write $\ell(x) = |w_i x_i + c_i|$, where $c_i = \sum_{j \neq i} w_j \cdot x_j$ does not depend on x_i.

Since $\ell^2(x) = c_i^2 + w_i^2 + 2c_i w_i x_i$, it follows that $\mathbf{E}_{x_i}[\ell^2(x)] = c_i^2 + w_i^2$, and clearly $\mathbf{E}_{x_i}[\ell(x)] = (1/2)\cdot(|w_i - c_i| + |w_i + c_i|)$. We consider two cases based on the relative magnitudes of c_i and w_i.

If $|c_i| \le |w_i|$, we have $\mathbf{E}_{x_i}[\ell(x)] = (1/2)\cdot(\mathrm{sign}(w_i)(w_i - c_i) + \mathrm{sign}(w_i)(w_i + c_i)) = |w_i|$. Hence, in this case $\mathrm{Var}_{x_i}[\ell(x)] = c_i^2 \le w_i^2$. If on the other hand $|c_i| > |w_i|$, then we have $\mathbf{E}_{x_i}[\ell(x)] = (1/2)\cdot(\mathrm{sign}(c_i)(c_i - w_i) + \mathrm{sign}(c_i)(c_i + w_i)) = |c_i|$, so again $\mathrm{Var}_{x_i}[\ell(x)] = w_i^2$ as desired. $\square$

The bound $\mathbf{K}(w) \ge 1/\sqrt{2}$ follows from the above two claims using the fact that ℓ is even and that $\sum_i w_i^2 = 1$.

2.2 Proof of Theorem 7

Let $w \in \mathbb{R}^n$ be a proper unit vector and denote $\tau = \|w - w^*\|_2$. To prove Theorem 7, one would intuitively want to obtain a robust version of the simple Fourier-analytic proof of Theorem 6 from the previous subsection. Recall that the latter proof boils down to the following:

$$\mathrm{Var}[\ell] \le (1/2)\cdot \mathrm{Inf}(\ell) = (1/2)\cdot \sum_{i=1}^n \mathrm{Inf}_i(\ell) \le (1/2)\cdot \sum_{i=1}^n w_i^2 = 1/2$$

where the first inequality is Fact 8 and the second is Lemma 1. While it is clear that both inequalities can be individually tight, one could hope to show that both inequalities cannot be tight simultaneously. It turns out that this intuition is not quite true, however it holds if one imposes some additional conditions on the weight vector w. The remaining cases for w that do not satisfy these conditions can be handled by elementary arguments.

We first note that without loss of generality we may assume that $w_1 = \max_i w_i > 0.3$, for otherwise Theorem 7 follows directly from the following result of König *et al*:

Theorem 9 ([KSTJ99]). *For a proper unit vector $w \in \mathbb{R}^n$, we have $\mathbf{K}(w) \ge \sqrt{2/\pi} - (1 - \sqrt{2/\pi})w_1$.*

Indeed, if $w_1 \le 0.3$, the above theorem gives that $\mathbf{K}(w) \ge 1.3\sqrt{2/\pi} - 0.3 > 0.737 > 1/\sqrt{2} + 3/100 \ge 1/\sqrt{2} + (1/50)\tau$, where the last inequality follows from the fact that $\tau \le \sqrt{2}$ (as both w and w^* are unit vectors). Hence, we will henceforth assume that $w_1 > 0.3$.

The preceding discussion leads us to the following definition:

Definition 6 (canonical vector). *We say that a proper unit vector $w \in \mathbb{R}^n$ is* canonical *if it satisfies the following conditions: (a) $w_1 \in [0.3, 1/\sqrt{2} + 1/100]$; and (b) $\tau = \|w - w^*\|_2 \ge 1/5$.*

The following lemma establishes Theorem 7 for non-canonical vectors:

Lemma 2. *Let w be a proper non-canonical vector. Then $\mathbf{K}(w) \geq 1/\sqrt{2} + (1/1000)\tau$, where $\tau = \|w - w^*\|_2$.*

The proof of Lemma 2 is elementary, using only basic facts about symmetric random variables, but sufficiently long that we give it in the full version. For canonical vectors we show:

Theorem 10. *There exist universal constants $c_1, c_2 > 0$ such that: Let $w \in \mathbb{R}^n$ be canonical. Consider the mapping $\ell(x) = |w \cdot x|$. Then at least one of the following statements is true : (1) $\mathrm{Inf}_1(\ell) \leq w_1^2 - c_1$; (2) $\mathbf{W}^{>2}[\ell] \geq c_2$.*

This proof is more involved, using Fourier analysis and critical index arguments (see the full version). We proceed now to show that for canonical vectors, Theorem 7 follows from Theorem 10. To see this we argue as follows: Let $w \in \mathbb{R}^n$ be canonical. We will show that there exists a universal constant $c > 0$ such that $\mathbf{K}(w) \geq 1/\sqrt{2} + c$; as mentioned above, since $\tau < \sqrt{2}$, this is sufficient for our purposes. Now recall that

$$\mathbf{K}(w) = \mathbf{E}_x[\ell(x)] = \widehat{\ell}(0) = \sqrt{1 - \mathrm{Var}[\ell]}. \tag{3}$$

In both cases, we will show that there exists a constant $c' > 0$ such that

$$\mathrm{Var}[\ell] \leq 1/2 - c'. \tag{4}$$

From this (3) gives $\mathbf{K}(w) \geq \sqrt{1/2 + c'} = 1/\sqrt{2} + c''$ where $c'' > 0$ is a universal constant, so to establish Theorem 7 it suffices to establish (4).

Suppose first that statement (1) of Theorem 10 holds. In this case we exploit the fact that Lemma 1 is not tight. We can write

$$\mathrm{Var}[\ell] \leq (1/2) \cdot \mathrm{Inf}(f) \leq (1/2) \cdot \left(w_1^2 - c_1 + \sum_{i=2}^{n} w_i^2 \right) \leq (1/2) - c_1/2,$$

giving (4). Now suppose that statement (2) of Theorem 10 holds, i.e. at least a c_2 fraction of the total Fourier mass of ℓ lies above level 2. Since ℓ is even, this is equivalent to the statement $\mathbf{W}^{\geq 4}[\ell] \geq c_2$. In this case, we prove a better upper bound on the variance because Fact 8 is not tight. In particular, we have

$$\mathrm{Inf}(\ell) \geq 2\mathbf{W}^2[\ell] + 4\mathbf{W}^{\geq 4}[\ell] = 2\left(\mathrm{Var}[\ell] - \mathbf{W}^{\geq 4}[\ell]\right) + 4\mathbf{W}^{\geq 4}[\ell] = 2\,\mathrm{Var}[\ell] + 2\mathbf{W}^{\geq 4}[\ell]$$

which yields $\mathrm{Var}[\ell] \leq (1/2)\mathrm{Inf}(\ell) - \mathbf{W}^{\geq 4}[\ell] \leq (1/2) - c_2$, again giving (4) as desired.

3 Proof of Theorem 1 using Theorem 5

We first observe that it suffices to prove the theorem for balanced LTFs, i.e. LTFs $f : \{-1,1\}^n \to \{-1,1\}$ with $\widehat{f}(\emptyset) = \mathbf{E}[f] = 0$. (Note that any balanced LTF can be represented with a threshold of 0, i.e. $f(x) = \mathrm{sign}(w \cdot x)$ for some $w \in \mathbb{R}^n$.) To see this, let $f : \{-1,1\}^n \to \{-1,1\}$ be an arbitrary n-variable threshold function, i.e.

$f(x) = \mathrm{sign}(w_0 + \sum_{i=1}^n w_i x_i)$, and note that we may assume that $w_0 \neq w \cdot x$ for all $x \in \{-1,1\}^n$. Consider the $(n+1)$-variable balanced LTF $g : (x, y) \to \{-1, 1\}$, where $y \in \{-1,1\}$, defined by $g(x,y) = \mathrm{sign}(w_0 y + \sum_{i=1}^n w_i x_i)$. Then it is easy to see that $\widehat{g}(y) = \mathbf{E}[f]$ and $\widehat{g}(i) = \widehat{f}(i)$ for all $i \in [n]$. Therefore, $\mathbf{W}^{\leq 1}[f] = \mathbf{W}^1[g] = \mathbf{W}^{\leq 1}[g]$.

Let $f = \mathrm{sign}(w \cdot x)$ be an LTF. We may assume that w is a proper unit vector, i.e. that $\|w\|_2 = 1$ and $w_i \geq w_{i+1} > 0$ for $i \in [n-1]$. We can also assume that $w \cdot x \neq 0$ for all $x \in \{-1,1\}^n$. We distinguish two cases: If w is "far" from w^* (i.e. the worst-case vector for the Khintchine inequality), the desired statement follows immediately from our robust inequality (Theorem 5). For the complementary case, we use a separate argument that exploits the structure of w. More formally, we have the following two cases:

Let $\tau > 0$ be a sufficiently small universal constant, to be specified.

[Case I: $\|w - w^*\|_2 \geq \tau$**].** In this case, Proposition 1 and Theorem 5 give us

$$\mathbf{W}^1[f] \geq (\mathbf{K}(w))^2 \geq (1/\sqrt{2} + c\tau)^2 \geq 1/2 + \sqrt{2}c\tau$$

which completes the proof of Theorem 1 for Case I.

[Case II: $\|w - w^*\|_2 \leq \tau$**].** In this case the idea is to consider the restrictions of f obtained by fixing the variables x_1, x_2 and argue based on their bias. See the full version for details.

References

[AHS12] N. Alon, H. Huang, and B. Sudakov. Nonnegative k-sums, fractional covers, and probability of small deviations. *J. Combin. Theory, Series B*, 102(3):784–796, 2012.

[BD12] V. Bentkus and D. Dzindzalieta. A tight Gaussian bound for weighted sums of Rademacher random variables. Technical report, preprint, 2012.

[BDD98] S. Ben-David and E. Dichterman. Learning with restricted focus of attention. *Journal of Computer and System Sciences*, 56(3):277–298, 1998.

[Ben04] V. Bentkus. On Hoeffding's Inequality. *Annals of Probability*, 32:1650–1673, 2004.

[BKS99] I. Benjamini, G. Kalai, and O. Schramm. Noise sensitivity of Boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math.*, 90:5–43, 1999.

[BTNR02] A. Ben-Tal, A. Nemirovski, and C. Roos. Robust Solutions of Uncertain Quadratic and Conic-Quadratic Problems. *SIAM J. Optimization*, 13(2):535–560, 2002.

[DDFS12] A. De, I. Diakonikolas, V. Feldman, and R. Servedio. Near-optimal solutions for the Chow Parameters Problem and low-weight approximation of halfspaces. In *Proc. 44th ACM Symposium on Theory of Computing (STOC)*, pages 729–746, 2012.

[DGJ+10] I. Diakonikolas, P. Gopalan, R. Jaiswal, R. Servedio, and E. Viola. Bounded independence fools halfspaces. *SIAM J. on Comput.*, 39(8):3441–3462, 2010.

[DS09] I. Diakonikolas and R. Servedio. Improved approximation of linear threshold functions. In *Proc. 24th CCC*, pages 161–172, 2009.

[Fil12] Y. Filmus. Khintchine-Kahane using Fourier Analysis. Posted at http://www.cs.toronto.edu/˜yuvalf/KK.pdf, 2012.

[GL94] C. Gotsman and N. Linial. Spectral properties of threshold functions. *Combinatorica*, 14(1):35–50, 1994.

[Guy86] R. K. Guy. Any answers anent these analytical enigmas? *American Math Monthly*, 93:279–281, 1986.

[Haa82] U. Haagerup. The best constants in the Khintchine inequality. *Studia Math.*, 70:231–283, 1982.

[HK92] R. Holzman and D. J. Kleitman. On the product of sign vectors and unit vectors. *Combinatorica*, 12(3):303–316, 1992.

[HK94] P. Hitczenko and S. Kwapień. On the Rademacher series. *Probability in Banach spaces*, pages 31–36, 1994. Birkhäuser, Boston, MA.

[HLNZ08] S. He, Z. Luo, J. Nie, and S. Zhang. Semidefinite Relaxation Bounds for Indefinite Homogenous Quadratic Optimization. *SIAM J. Optimization*, 19:503–523, 2008.

[Jac06] J. Jackson. Uniform-distribution learnability of noisy linear threshold functions with restricted focus of attention. In *Proc. 19th COLT*, pages 304–318, 2006.

[KLO96] S. Kwapień, R. Latala, and K. Oleszkiewicz. Comparison of moments of sums of independent random variables and differential inequalities. *J. Funct. Anal.*, 136:258–268, 1996.

[KSTJ99] H. König, C. Schütt, and N. Tomczak-Jaegermann. Projection constants of symmetric spaces and variants of Khintchine's inequality. *J. Reine Agnew. Math.*, 511:1–42, 1999.

[LO94] R. Latala and K. Oleszkiewicz. On the best constant in the Khinchin-Kahane inequality. *Studia Math.*, 109:101–104, 1994.

[MORS10] K. Matulef, R. O'Donnell, R. Rubinfeld, and R. Servedio. Testing halfspaces. *SIAM J. on Comput.*, 39(5):2004–2047, 2010.

[MS90] S. Montgomery-Smith. The distribution of Rademacher sums. In *Proceedings of the American Mathematical Society*, pages 517–522, 1990.

[MT94] W. Maass and G. Turan. How fast can a threshold gate learn? In *Computational Learning Theory and Natural Learning Systems: Volume I: Constraints and Prospects*, pages 381–414. MIT Press, 1994.

[O'D12] R. O'Donnell. Analysis of boolean functions. Technical report, 2012. Book serialization available at http://analysisofbooleanfunctions.org/.

[Ole96] K. Oleszkiewicz. On the Stein property of Rademacher sequences. *Probability and Mathematical Statistics*, 16:127–130, 1996.

[Ole99] K. Oleszkiewicz. Comparison of moments via Poincaré-type inequality. *Contemporary Mathematics*, 234:135–148, 1999.

[OS11] R. O'Donnell and R. Servedio. The Chow Parameters Problem. *SIAM J. on Comput.*, 40(1):165–199, 2011.

[Per04] Y. Peres. Noise stability of weighted majority, 2004. Available at http://arxiv.org/abs/math/0412377.

[Pin94] I. Pinelis. Extremal probabilistic problems and Hotelling's t^2 test under a symmetry condition. *Ann. Statist.*, 22:357–368, 1994.

[Pin12] I. Pinelis. An asymptotically Gaussian bound on the Rademacher tails. *Electronic Journal of Probability.*, 17:1–22, 2012.

[Ren88] J. Renegar. A faster PSPACE algorithm for deciding the existential theory of the reals. *IEEE Annual Symposium on Foundations of Computer Science*, pages 291–295, 1988.

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

[So09] A. M.-C. So. Improved approximation bound for quadratic optimization problems with orthogonality constraints. In *SODA*, pages 1201–1209, 2009.

[Sza76] S.J. Szarek. On the best constants in the Khinchine inequality. *Studia Math.*, 58:197–208, 1976.

[Tom87] B. Tomaszewski. A simple and elementary proof of the Khintchine inequality with the best constant. *Bull. Sci. Math.*, 111(2):103–109, 1987.