ISRAEL JOURNAL OF MATHEMATICS **160** (2007), 389–412
DOI: 10.1007/s11856-007-0068-9

ON THE FOURIER TAILS OF BOUNDED FUNCTIONS OVER THE DISCRETE CUBE

BY

IRIT DINUR

School of Computer Science and Engineering, Hebrew University of Jerusalem
Givat Ram, Jerusalem 91904, Israel
e-mail: dinuri@cs.huji.ac.il

AND

EHUD FRIEDGUT*

Institute of Mathematics, Hebrew University of Jerusalem
Givat Ram, Jerusalem 91904, Israel
e-mail: ehudf@math.huji.ac.il

AND

GUY KINDLER

Dept. of Computer Science and Applied Mathematics, Weizmann Institute of Science
PO Box 26, Rehovot 76100, Israel
e-mail: guy.kindler@weizmann.ac.il

AND

RYAN O'DONNELL**

School of Computer Science, Carnegie Mellon University
5000 Forbes Avenue Pittsburgh, PA 15213, USA
e-mail: odonnell@cs.cmu.edu

* Research supported in part by the Israel Science Foundation, grant no. 0329745.

** This work was completed while the author was at the Institute for Advanced Study, Princeton, NJ. The material is based upon work supported in part by the National Science Foundation under agreement No. CCR-0324906. Recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.
Received July 21, 2005

ABSTRACT

In this paper we consider bounded real-valued functions over the discrete cube, $f\colon \{-1,1\}^n \to [-1,1]$. Such functions arise naturally in theoretical computer science, combinatorics, and the theory of social choice. It is often interesting to understand when these functions essentially depend on few coordinates. Our main result is a dichotomy that includes a lower bound on how fast the Fourier coefficients of such functions can decay: we show that

$$\sum_{|S|>k} \widehat{f}(S)^2 \geq \exp(-O(k^2 \log k)),$$

unless f depends essentially only on $2^{O(k)}$ coordinates. We also show, perhaps surprisingly, that this result is sharp up to the $\log k$ factor.

The same type of result has already been proven (and shown useful) for Boolean functions [Bou02, KS]. The proof of these results relies on the Booleanity of the functions, and does not generalize to all bounded functions. In this work we handle all **bounded** functions, at the price of a much faster tail decay. As already mentioned, this rate of decay is shown to be both roughly necessary and sufficient.

Our proof incorporates the use of the noise operator with a random noise rate and some extremal properties of the Chebyshev polynomials.

1. Introduction

Boolean functions $f\colon \{-1,1\}^n \to \{-1,1\}$ — equivalently, set systems on $[n]$ — are ubiquitous in theoretical computer science and combinatorics, and also arise in the theory of social choice, in statistical physics and in classical harmonic analysis. Thus, it is of great importance to understand their basic combinatorial structure. For problems involving the uniform probability distribution over $\{-1,1\}^n$ and its Hamming graph structure, a watershed in the analysis of boolean functions came with the paper of Kahn, Kalai and Linial [KKL88]; this paper demonstrated the power of generalizing to the case of functions $f\colon \{-1,1\}^n \to \mathbb{R}$ and applying tools from harmonic analysis. Since this result, Fourier analysis has had crucial application in the many areas of research in which boolean functions arise. For notation and Fourier-related definitions in what follows please see Section 2.

Some of the most important theorems on the Fourier analysis of boolean functions show that the Fourier coefficients of boolean functions cannot decay too quickly. In other words, boolean functions must have at least some small portion of their L_2 Fourier mass on characters of high degree. There is a slight catch to such theorems, in that boolean functions that actually only depend on a constant number of coordinates have **no** Fourier weight at high levels. Such functions — called "juntas" — along with slight perturbations of them must be exempted from the statements of theorems on Fourier decay. Accordingly, the following definition is made.

Definition 1.1: A function $f\colon \{-1,1\}^n \to \mathbb{R}$ is called an (ε, j)-junta if there exists a function $g\colon \{-1,1\}^n \to \mathbb{R}$ depending on at most j coordinates such that $\|f-g\|_2^2 \le \varepsilon$.

With this definition in place, let us recall three basic junta-related theorems related to Fourier analysis of boolean functions. In each of these theorems, $f\colon \{-1,1\}^n \to \{-1,1\}$ is an arbitrary boolean function and k is any positive integer.

Theorem ([Fri98]): *Suppose* $\sum_S |S|\widehat{f}(S)^2 \le k$. *Then* f *is an* $(\varepsilon, 2^{O(k/\varepsilon)})$*-junta for every* $\varepsilon > 0$.

Theorem ([FKN02], see also [ADFS03]): *Suppose* $\sum_{|S|>1} \widehat{f}(S)^2 < \varepsilon$. *Then* f *is an* $(O(\varepsilon), 1)$*-junta.*

Theorem ([Bou02], see also [KS, KN05]): *Suppose*

$$\sum_{|S|>k} \widehat{f}(S)^2 > (\varepsilon/k)^{1/2+o(1)}.$$

Then f *is an* $(\varepsilon, 2^{O(k)}/\varepsilon^{O(1)})$*-junta.*

In this paper we prove a theorem of the same flavor as Bourgain's Theorem, (the third of the theorems quoted above) but for **real-valued, bounded** functions, $f\colon \{-1,1\}^n \to [-1,1]$. Such bounded functions often arise naturally, particularly as weighted averages of boolean functions; e.g., as Fourier transforms of boolean functions, as noise-convolutions of boolean functions, in the context of random walks on the discrete cube and in hardness-of-approximation theory in computational complexity [KKMO04].

1.1 Our results. Informally, Bourgain's Theorem states that boolean functions $f\colon \{-1,1\}^n \to \{-1,1\}$ must satisfy $\sum_{|S|>k} \widehat{f}(S)^2 > k^{-1/2-o(1)}$, unless they are close to being juntas. One may ask to what extent this lower bound depends on the fact that f's range is $\{-1,1\}$. Certainly nothing can be said for general (unbounded) functions $f\colon \{-1,1\}^n \to \mathbb{R}$, as the function

$f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i$ demonstrates. (This function has constant ℓ_2 norm, all influences are small, yet it is far from every junta.) However, we will present a nontrivial lower bound for the Fourier weight beyond level k so long as f's range is **bounded** by, say, the interval $[-1, 1]$.

For bounded functions $f\colon \{-1,1\}^n \to [-1,1]$ one cannot expect a polynomially large amount of weight beyond level k for non-juntas, certainly not the $k^{-1/2}$ of Bourgain's Theorem. A simple way to see this is to consider the function $T_{1/2}(\mathsf{Majority}_n)$. The precise definition of this function will come later in Section 2, but for now it suffices to state that this is a symmetric function which is easily seen to be $\Omega(1)$-far from every $o(n)$-junta. Furthermore, it is a weighted average of Boolean functions, hence bounded, and its Fourier weight beyond level k is of the form $2^{-\Theta(k)}$. Since the majority function is very often an extremal case (as it essentially is in Bourgain's Theorem) one might expect that a tail decay rate of $2^{-\Theta(k)}$ is maximal for bounded non-juntas. However, the relative tightness of our theorems show that this is not the case.

Our main result shows that bounded boolean functions $f\colon \{-1,1\}^n \to [-1,1]$ must satisfy $\sum_{|S|>k} \widehat{f}(S)^2 > \exp(-O(k^2 \log k))$, unless they are close to being juntas. Formally:

THEOREM 1: *Let $f\colon \{-1,1\}^n \to [-1,1]$, $k \geq 1$ and $\varepsilon > 0$. Suppose*

$$\sum_{|S|>k} \widehat{f}(S)^2 \leq \exp(-O(k^2 \log k)/\varepsilon).$$

Then f is an $(\varepsilon, 2^{O(k)}/\varepsilon^2)$-junta.

On the other hand, this theorem is tight, except, possibly, for the $\log k$ in the exponent.

THEOREM 2: *For every $j > 0$ there exists a bounded function*

$$f\colon \{-1,1\}^n \to [-1,1]$$

which is not a $(0.01, j)$-junta, such that

$$\sum_{|S|>k} \widehat{f}(S)^2 \geq \exp(-O(k^2)).$$

The main lemma used in the proof of Theorem 1 is, in our opinion, interesting in its own right. It is basically a converse of standard tail bounds, where we are interested in a lower bound for the tail probability, rather than the more popular upper bound. Our lemma is a generalization of the following well-known supergaussian tail property, which can be found as equation (4.2) in [LT91].

Lemma 1.2: *There is a universal constant K such that the following holds: Let $\ell(x) = \sum_{i=1}^{n} a_i x_i$, where the a_i's satisfy $\sum a_i^2 = 1$ and the x_i's are independent Rademacher random variables. Let $t \geq 1$ and suppose that $|a_i| < \frac{1}{Kt}$ for all i. Then*

$$\Pr[|\ell(x)| > t] \geq \exp(-Kt^2).$$

Lemma 1.2 shows a supergaussian estimate for linear functions whose coefficients are "smeared" (i.e., where each of the coefficients is small). Our main lemma shows that the supergaussian behavior is maintained, albeit with different parameters, even if terms of degree at most k are added to the function. As we will show, the bound in Lemma 1.3 is tight up to the constant K'.

Lemma 1.3 (our main lemma): *There is a universal constant K' such that the following holds: Suppose $f\colon \{-1,1\}^n \to \mathbb{R}$ has degree at most k and that $\|f^{=1}\|_2 \geq 1$. (See subsection 2.1 for definitions.) Let $t \geq 1$ and suppose that $|\widehat{f}(\{i\})| < \frac{1}{K'tk}$ for all i. Then*

$$\Pr[|f| \geq t] \geq \exp(-K't^2k^2).$$

Finally, using a random restriction technique, Lemma 1.3 yields a lower bound on the probability of large deviations for low-degree functions with small "influences", rather than for functions with smeared weight on the first level.

Theorem 3: *There is a universal constant C such that the following holds: Suppose $f\colon \{-1,1\}^n \to \mathbb{R}$ has degree at most k and that $\sum_{S\neq\emptyset} \widehat{f}(S)^2 = 1$. Let $t \geq 1$ and suppose that $\sum_{S\ni i} \widehat{f}(S)^2 \leq t^{-2}C^{-k}$ for all i. Then*

$$\Pr[|f| \geq t] \geq \exp(-Ct^2k^2 \log k).$$

1.2 An outline of the paper and the proof. The proof of our main theorem is conceptually simple. Given a function with rapidly decaying Fourier coefficients we may approximate it by truncating its Fourier expansion, thus getting a low degree polynomial. Now we face two possibilities: either this polynomial essentially depends on few coordinates, i.e., is a junta, or it has a significant component contributed by all coordinates with small influence. In the latter case we show that such a component would necessarily mean that the original function has large deviations, i.e., is not bounded by a constant.

The structure of the rest of the paper is as follows. In Section 2, we present some notation and background related to Fourier analysis and Chebyshev polynomials. In Section 3 we prove Lemma 1.3 regarding the deviation of functions with "smeared" linear part. In Section 4, we prove Theorem 3 regarding the deviation of functions where all influences are small. Our main theorem, Theorem 1, that relates the rate of decay of the Fourier coefficients of a function to its dependence on few variables, is proven in Section 5. Next, in Section 6, we state and prove our tightness results for Lemma 1.3, Theorem 3 and Theorem 1. Finally, in Appendix A, we discuss the parameters in the biased B.G.B. (hypercontractive) inequality (Theorem 6), and in Appendix B we prove some large deviation results.

2. Preliminaries

In this section we will recall the relevant notions from the Fourier analysis of boolean functions and define some basic notation. We will then recall some simple facts about the Chebyshev polynomials, which will play a role in our proofs. Finally, we will give some deviation bounds for boolean functions which follow from hypercontractivity.

2.1 Fourier notation.

MEASURES: We write $[n]$ for $\{1,\ldots,n\}$. Denote by μ_p the probability measure on the two point space $\{-1,1\}$ for which $\mu_p(-1)=p$. We write $\{-1,1\}^n_{(p)}$ for the set $\{-1,1\}^n$ equipped with the measure $\times^n\mu_p$. We will concentrate mainly on the uniform measure, and thus whenever we write $\{-1,1\}^n$ alone, we refer to $\{-1,1\}^n_{(1/2)}$. The uniform measure induces the L_2 norm on the space of real functions on $\{-1,1\}^n$, which in turn naturally defines an inner product.

FOURIER REPRESENTATION: For every $S\subseteq[n]$ define the character

$$\chi_S\colon \{-1,1\}^n\to\{-1,1\} \quad \text{by } \chi_S(x)=\prod_{i\in S}x_i.$$

The set of characters forms an orthonormal basis with respect to the inner product above. Every function $f\colon\{-1,1\}^n\to\mathbb{R}$ can thus be uniquely expanded with respect to this basis as $f=\sum\widehat{f}(S)\chi_S$. This is called the Fourier expansion of f.

The Fourier representation allows us to consider certain interesting projections. For a function $f\colon\{-1,1\}^n\to\mathbb{R}$ and a nonnegative integer k, we define

$$f^{=k}=\sum_{|S|=k}\widehat{f}(S)\chi_S,$$

and similarly,

$$f^{>k}=\sum_{|S|>k}\widehat{f}(S)\chi_S.$$

We refer to the quantity $\|f^{=k}\|_2^2 = \sum_{|S|=k} \widehat{f}(S)^2$ as the "weight of f on level k;" similarly, $\|f^{>k}\|_2^2$ we call the "weight of f above level k."

Influences: Let $f\colon \{-1,1\}^n \to \mathbb{R}$ and let $i \in [n]$. We define

$$\partial_i f\colon \{-1,1\}^n \to \mathbb{R} \quad \text{by } (\partial_i f)(x) = (f(x) - f(x \oplus i))/2,$$

where $x \oplus i$ is the vector x with the coordinate x_i replaced by $-x_i$. The "influence of i on f" is defined to be $\mathrm{Inf}_i(f) = \|\partial_i f\|_2^2$. It is an easy exercise to show that

$$\partial_i f = \sum_{S\colon i \in S} \widehat{f}(S)\chi_S$$

and hence, by Parseval's identity, we have the following formula:

$$\mathrm{Inf}_i(f) = \sum_{S \ni i} \widehat{f}(S)^2. \tag{1}$$

2.2 Noise and large deviations. We will need some large deviation bounds for low degree functions. Such results are known to follow from the hypercontractive estimate discovered independently by Bonami, Gross and Beckner [Bon70, Gro75, Bec75]; see, e.g., [Jan97, Chap. 6]. However, we use these results in the biased measure case, for which the sharpest hypercontractivity result was determined only recently, by Oleszkiewicz [Ole02]. For this reason, and because we were unable to find the precise formulations we use in the literature, we have given explicit proofs of all the theorems below in Appendix B.

Let us begin by defining the noise operator:

Definition 1: Let $0 \le \rho \le 1$. The **noise operator** T_ρ, operating on the space of functions $f\colon \{-1,1\}^n \to \mathbb{R}$, is defined by

$$T_\rho(f) = \sum_{S \subseteq [n]} \rho^{|S|} \widehat{f}(S)\chi_S.$$

The following gives an equivalent definition:

Fact 2.1: Let $0 \le \rho \le 1$. Then for $f\colon \{-1,1\}^n \to \mathbb{R}$ and $x \in \{-1,1\}^n$,

$$(T_\rho f)(x) = \mathop{\mathbb{E}}_{z}[f(x \cdot z)],$$

where $z \sim \{-1,1\}^n_{((1-\rho)/2)}$, and $\cdot$ denotes coordinate-wise multiplication.

Note that the noise operator is a weighted average. The value of $(T_\rho f)(x)$ is an average of f's over the whole space, where points receive weights proportional to (a power of) their distance from x. A simple consequence of this, which in fact is part of the motivation for the study of bounded functions, is that if f is boolean then $T_\rho f$ is bounded.

The next important attribute of the noise operator is the fact that the averaging of f has a "smoothing" effect; this results in the higher norms of $T_\rho f$ being comparable to the 2-norm of f. This notion is captured by the important hypercontractive inequality of Bonami, Gross and Beckner (henceforth B.G.B.):

THEOREM 4 (B.G.B.): *Let $r \geq 2$, and let $f\colon \{-1,1\}^n \to \mathbb{R}$. Then for $\rho = (r-1)^{-1/2}$,*

$$\|T_\rho(f)\|_r \leq \|f\|_2.$$

The following inequality is a direct corollary of Theorem 4.

THEOREM 5: *Suppose $f\colon \{-1,1\}^n \to \mathbb{R}$ has degree at most k. Then for all $r \geq 2$,*

$$\|f\|_r \leq (r-1)^{k/2}\|f\|_2.$$

The above inequality extends to the case of biased (non-uniform) product measures on $\{-1,1\}^n$, with different constants.

THEOREM 6: *For every fixed $p \in (0,1)$ there exists a constant $B(p) \geq 1$ for which the following holds: For every function $f\colon \{-1,1\}^n_{(p)} \to \mathbb{R}$ of degree at most k, and for every $r \geq 2$,*

$$\|f\|_r \leq (B(p)\cdot(r-1))^{k/2}\|f\|_2.$$

Moreover, for every p in the segment $[1/4, 3/4]$, $B(p) < 1.22$.

Theorem 6 follows from a "biased version" of the B.G.B. hypercontractive estimate. The tightest constants $B(p)$ for which Theorem 6 holds were recently found by Oleszkiewicz [Ole02]. These parameters are cited and discussed in Appendix A.

Theorem 6 leads in a straightforward way to large-deviation bounds for low degree functions on the biased discrete cube. The results stated in the remainder of this section are all proved in Appendix B.

LEMMA 2.2: *Suppose $f\colon \{-1,1\}^n_{(p)} \to \mathbb{R}$ has degree at most k and assume $\|f\|_2 = 1$. Let $B = B(p)$ be defined as in Theorem 6. Then for any $t \geq (2Be)^{k/2}$,*

$$\Pr[|f| \geq t] \leq \exp\Big(-\frac{k}{2Be}t^{2/k}\Big).$$

A useful corollary of this lemma bounds how much of a low-degree function's second moment can come from very large values:

LEMMA 2.3: *Under the hypotheses of Lemma 2.2,*

$$\mathbb{E}[f^2 \cdot \mathbf{1}_{\{f^2>t^2\}}] \le t^2 \exp\Big(-\frac{k}{2Be}t^{2/k}\Big).$$

For future use, we record a special case of Lemma 2.3.

COROLLARY 2.4: *Suppose $f\colon \{-1,1\}^n_{(p)} \to \mathbb{R}$ has degree at most k and assume $\|f\|_2 = 1$. Let $B = B(p)$ be defined as in Theorem 6, and let $t_0 = (2Be)^k$. Then*

$$\mathbb{E}[f^2 \cdot \mathbf{1}_{\{f^2>t_0^2\}}] \le 0.13.$$

Finally, we show that low-degree functions must exceed their expectation with nonnegligible probability.

LEMMA 2.5: *Suppose that $f\colon \{-1,1\}^n_{(p)} \to \mathbb{R}$ has degree at most k. Let $B = B(p)$ be as in Theorem 6. Then*

$$\Pr[f \ge \mathbb{E}[f]] \ge 0.4 \cdot (30B^2)^{-k}.$$

2.3 CHEBYSHEV POLYNOMIALS. The Chebyshev polynomials (of the first kind) are essential in the proofs of Lemma 1.3 (which leads to Theorem 1) and of Theorem 2. Recall that for each integer $k \ge 0$ there is one Chebyshev polynomial of degree k, $C_k(x)$, defined uniquely by

$$C_k(x) = \cos(k \cdot \arccos(x)).$$

(Note: The standard notation is $T_k(x)$ as opposed to $C_k(x)$; however we have made the switch because T is the standard notation for the noise operator.)

Our proof of Lemma 1.3 will require one of the many extremal properties of the Chebyshev polynomials. For a thorough treatment of the properties of Chebyshev polynomials, the reader may consult the book of Rivlin [Riv90]. Following Rivlin (but with the $T \to C$ notational switch), let us write $c_j^{(k)}$ for the degree-j coefficient of C_k, and write $\eta_0^{(k)}, \dots, \eta_n^{(k)}$ for the $k+1$ extrema of C_k in the segment $[-1,1]$, on which $|C_k| = 1$. We shall need the following lemma, which is a special case of "Remark 2" on page 112 of Rivlin.

LEMMA 2.6: *Let k be odd and let $p(x) = a_0 + a_1 x + \cdots + a_k x^k$ be a polynomial satisfying $|p(\eta_j^{(k)})| \le 1$ for all $j = 0, \dots, k$. Then $|a_1| \le |c_1^{(k)}|$, with equality only if $p = \pm C_k$.*

It may be checked that $|c_1^{(k)}| = k$ when k is odd. Thus as a simple corollary we have:

COROLLARY 2.7: *Let k be odd and let $p(x) = a_0 + a_1x + \cdots + a_kx^k$. Then there exists $0 \le j \le k$ such that $|p(\eta_j^{(k)})| \ge |a_1|/k$.*

For a technical reason we will need a further corollary:

COROLLARY 2.8: *Suppose p is a polynomial of degree at most k, where k is odd, with linear coefficient a_1. Then there exists $0 \le j \le k+1$ such that $|p(\eta_j^{(k+1)}/2)| \ge |a_1|/(2k+2)$.*

Proof: This follows by applying the previous corollary to the polynomial $\tilde{p}(x) = p(x/2)$. ∎

For our proof of Theorem 2 and other tightness results, we will need some technical estimates on the Chebyshev polynomials and their derivatives.

FACT 2.9: For k odd, $|C_k(x)| \le \frac{1}{2}|2x|^k$ for all $|x| \ge 1$.

FACT 2.10: For k odd,

$$|C_k'(x)| \text{ is } \begin{cases} \ge k/2 & \text{for } |x| \le 1/k, \\ \le (4/3)k & \text{for } |x| \le 1/2, \\ \le k|2x|^{k-1} & \text{for } |x| \ge 1/2. \end{cases}$$

3. Proof of the main lemma; Random noise with a random rate

In this section we prove Lemma 1.3, which shows that a function with a "smeared" first level obtains large values with positive (though exponentially small) probability. Lemma 1.3 is an extension of Lemma 1.2 to the case of degree k functions. In Section 6, we show that the bound in Lemma 1.3 is tight, except for the constant factor in the exponent. For the sake of clarity, let us restate the lemma before proving it.

LEMMA 1.3: *There is a universal constant K' such that the following holds: Suppose $f\colon \{-1,1\}^n \to \mathbb{R}$ has degree at most k and that $\|f^{=1}\|_2 \ge 1$. Let $t \ge 1$ and suppose that $|\widehat{f}(\{i\})| < 1/(K'tk)$ for all i. Then*

$$\Pr[|f| \ge t] \ge \exp(-K't^2k^2).$$

IDEA OF THE PROOF: The proof first considers the linear part of f. According to Lemma 1.2, when evaluated on a random point x_0, the linear part has a non-negligible probability of obtaining a large value. It may be that even in that event, the total value of f at x_0 is still small, due to cancellations contributed by the non-linear part. To evade this cancellation we introduce some random noise with rate ρ and consider $T_\rho f(x_0)$. Unfortunately, since we have no further information concerning f and x_0, it could be — as if a malicious adversary had planned it — that the average effect of the noise on the linear and non-linear parts of f would once again cancel out. The twist of our proof is that this cannot happen simultaneously for all noise rates, hence we choose ρ from a finite set without specifying in advance which rate will be chosen. To paraphrase a prominent American president ([L]): "You can fool all the noise rates some of the time, you might even be able to fool some noise rates all of the time, but you can't fool all the noise rates all of the time." The existence of a successful noise rate is shown to be equivalent to the extremal properties of the Chebyshev polynomials (the expected contribution of the noise may be expressed as a polynomial in ρ). It is interesting to note that in showing this, we do not assume anything regarding the weight of $\widehat{f}^2$ on levels higher than the first, and we only use the fact that the weights beyond level k are all zero.

Proof of Lemma 1.3: By scaling f, we may assume that $\|f^{=1}\|_2 = 1$. Let ℓ denote the linear part of f, $\ell(x) = \sum_{i=1}^n \widehat{f}(\{i\})x_i$, and let $t' = (2k+2)t \geq 1$. Note that for a proper choice of the constant K' (specifically if $K' \geq 4K$, where K is the constant in Lemma 1.2), we have that ℓ and t' satisfy the conditions of Lemma 1.2, and therefore we have that

$$\Pr[|\ell(x)| \geq t'] \geq \exp(-Kt'^2). \tag{2}$$

Let $x_0 \in \{-1,1\}^n$ be any point on which $|\ell(x_0)| \geq t'$. For every real number ρ, denote

$$p_{x_0}(\rho) = (T_\rho f)(x_0),$$

and note that since f has degree at most k, it also holds that p_{x_0} is a degree k polynomial in ρ. Moreover, one easily notes that the linear coefficient of p_{x_0} is precisely $\ell(x_0)$.

We would like to apply Corollary 2.8 to p_{x_0} now. For this purpose we need k to be odd, but we may assume this without loss of generality, increasing k by one if needed. Write $\rho_j = (1/2)\eta_j^{(k+1)}$, for $j = 0, \ldots, (k+1)$, where the $\eta_j^{(k+1)}$'s are the extremal points of the Chebyshev polynomial, as mentioned in Subsection 2.3. Corollary 2.8 implies that there is some $j(x_0)$, $0 \leq j(x_0) \leq (k+1)$, such that

$$|(T_{\rho_{j(x_0)}} f)(x_0)| = |p(\rho_{j(x_0)})| \geq |\ell(x_0)|/(2k+2) \geq t'/(2k+2) = t. \tag{3}$$

Let z be a random string from $\{-1,1\}^n_{((1-\rho_{j(x_0)})/2)}$. By Fact 2.1,

$$T_{\rho_{j(x_0)}}(x_0) = \mathop{\mathbb{E}}_z[f(x_0 \cdot z)].$$

Hence (3) implies $|\mathbb{E}_z[f(x_0 \cdot z)]| \geq t$. But $f(x_0 \cdot z)$ is a polynomial in z of degree at most k and so we can apply Lemma 2.5 to it (by replacing f by $-f$ in the proof of the lemma, we get the same bound on the probability that $-f$ goes below $-\mathbb{E}[f]$). By the choice of the ρ_j's we have $\rho_{j(x_0)} \in [-1/2, 1/2]$, and hence the bias $(1/2 - \rho_{j(x_0)}/2)$ by which z is chosen is in the segment $(1/4, 3/4)$. Therefore by Theorem 6, the constant $B = B(1/2 - \rho_{j(x_0)}/2)$ for z is at most $4/3 \ln(3) \leq 1.22$, and so Lemma 2.5 yields

$$\Pr_z[|f(x_0 \cdot z)| \geq t] \geq 0.4 \cdot 37^{-k}. \tag{4}$$

Now consider the following process of choosing a random point $y \in \{-1,1\}^n$. First, a parameter ρ is chosen uniformly from the set $\{\rho_j\}_{j=0,\ldots,k+1}$. Then a point $x \in \{-1,1\}^n$ is chosen uniformly at random, and finally y is set to be a ρ-correlated copy of x. Note that y is uniformly distributed over $\{-1,1\}^n$. From the previous discussion we see that with probability at least $\exp(-Kt'^2)/(k+2) = \exp(-O(t^2k^2))$, both $|\ell(x)| \geq t'$ and $\rho = \rho_{j(x)}$; in this case, (4) implies that $\Pr_y[|f(y)| \geq t] \geq 0.4 \cdot 37^{-k}$. Thus

$$\Pr_y[|f(y)| \geq t] \geq \exp(-O(t^2k^2)) \cdot (0.4) \cdot 37^{-k} = \exp(-O(t^2k^2)).$$

This completes the proof. ∎

4. A lower bound on large deviations

In this section we derive Theorem 3 from Lemma 1.3. In fact we use the following slightly modified version of Lemma 1.3.

LEMMA 4.1: *There is a universal constant K such that the following holds: Suppose $f\colon \{-1,1\}^n \to \mathbb{R}$ has degree at most k. Let $T \subseteq [n]$, and let $t \geq 1$. Suppose that $\sum_{i\in T} \widehat{f}(\{i\})^2 \geq 1$ and that $|\widehat{f}(\{i\})| < 1/(Ktk)$ for all $i \in T$. Then*

$$\Pr[|f| \geq t] \geq \exp(-Kt^2k^2).$$

Proof: By scaling f, we may assume that $\sum_{i\in T} \widehat{f}(\{i\})^2 = 1$. We continue with almost the same proof as of Lemma 1.3, except for the following difference: instead of Equation (2), which stated that $\Pr[|\ell(x)| \geq t'] \geq \exp(-Kt'^2)$, we claim that

$$\Pr[|\ell(x)| \geq t'] \geq 1/2\exp(-Kt'^2)$$

(and thus we lose a factor of $\frac{1}{2}$ in the final bound, compared to Lemma 1.3). This is justified since Lemma 1.2 ensures that

$$\Pr\left[\left|\sum_{i\in T} \widehat{f}(\{i\})x_i\right| \geq t'\right] \geq \exp(-Kt'^2),$$

whereas the part of f depending on the coordinates outside of T is symmetric, and hence will increase or leave unchanged the magnitude of $\ell(x)$ with probability at least 1/2.

The rest of the proof follows as in Lemma 1.3 ∎

We now continue to the proof of Theorem 3. For convenience, we first cite it again.

THEOREM 3: *There is a universal constant C such that the following holds: Suppose $f\colon \{-1,1\}^n \to \mathbb{R}$ has degree at most k and assume $\sum_{S\neq\emptyset} \widehat{f}(S)^2 = 1$. Let $t \geq 1$ and suppose that $\mathrm{Inf}_i(f) \leq t^{-2}C^{-k}$ for all i. Then*

$$\Pr[|f| \geq t] \geq \exp(-Ct^2k^2 \log k).$$

IDEA OF THE PROOF: The proof begins by first identifying a significant "slice" in the Fourier transform of f. That is, we find some s, $1 \leq s \leq \log_2 k$, for which the weight of the Fourier transform of f on levels between 2^{s-1} and 2^s is at least $1/\log k$. By performing an appropriate random restriction which fixes many of the coordinates of f, we obtain a function where much of this weight is brought down to the first level (with non-negligible probability). Since f has very small influences, we expect the restricted function to have small first level Fourier coefficients (since f has degree k we can control the amount by which the coefficients deviate from their expectation using B.G.B.); when this happens, we can apply Lemma 4.1 to the restricted function. We thus prove that random restrictions of f obtain large values with non-negligible probability, which implies Theorem 3.

Proof: Given $s \geq 1$ we write $|S| \sim 2^s$ if $S \subseteq [n]$ satisfies $2^{s-1} \leq |S| < 2^s$. Since $\sum_{1\leq|S|\leq k} \widehat{f}(S)^2 \geq 1$ there must exist s, $1 \leq s \leq (\lceil \log_2 k \rceil + 1)$, such that $\sum_{|S|\sim 2^s} \widehat{f}(S)^2 \geq 1/(2\log k)$.

Let us choose a random subset $U \subseteq [n]$ by including each coordinate independently with probability 2^{-s}. For every $i \in [n]$, let

$$\gamma_i = \sum_{S\cap U=\{i\}} \widehat{f}(S)^2.$$

Note that γ_i is 0 if i is not in U, and it is also never more than $\sum_{S\ni i} \widehat{f}(S)^2 = \mathrm{Inf}_i(f)$, hence

$$\forall i,\ \gamma_i \leq t^{-2}C^{-k}. \tag{5}$$

A simple calculation shows that

$$\begin{aligned}\mathop{\mathbb{E}}_U[\gamma_i] &= \sum_S \Pr[S \cap U = \{i\}] \cdot \widehat{f}(S)^2 \geq \sum_{|S|\sim 2^s,\ S\ni i} 2^{-s} \cdot (1-2^{-s})^{2^s} \cdot \widehat{f}(S)^2 \\ &\geq \sum_{|S|\sim 2^s,\ S\ni i} 2^{-s} \cdot (1/4) \cdot \widehat{f}(S)^2.\end{aligned}$$

Summing over i, each S with $|S| \sim 2^s$ is counted at least 2^{s-1} times, and therefore

$$\mathop{\mathbb{E}}_U\left[\sum_{i\in[n]} \gamma_i\right] \geq \frac{1}{8} \sum_{|S|\sim 2^s} \widehat{f}(S)^2 \geq \frac{1}{16 \log k}.$$

Since $0 \leq \sum_{i\in[n]} \gamma_i \leq \sum_{S\neq\emptyset} \widehat{f}(S)^2 = 1$ for every choice of U, we conclude (by a Markov-inequality argument) that

$$\Pr_U\left[\sum_{i\in U} \gamma_i \geq \frac{1}{32\log k}\right] \geq \frac{1}{32 \log k}. \tag{6}$$

Let us fix for now a set U for which

$$\sum_{i\in U} \gamma_i \geq \frac{1}{32 \log k},$$

and let y be a uniformly random assignment to the coordinates in $[n] \setminus U$. Let $f_y\colon \{-1,1\}^U \to \mathbb{R}$ denote the restriction of f obtained by fixing the coordinates in $[n] \setminus U$ to y. Considering $\widehat{f_y}(\{i\})$ as a function of y, it is simple to observe that this is a function of degree smaller than k, that it has $\{\widehat{f}(S) : S \cap U = \{i\}\}$ as Fourier coefficients, and that it has no other nonzero Fourier coefficients. Therefore by definition of γ_i, we have for all $i \in U$, that

$$\mathop{\mathbb{E}}_y[\widehat{f_y}(\{i\})^2] = \gamma_i.$$

Since $\widehat{f_y}(\{i\})$ is of degree at most k as a function of y, letting $\mathbf{1}_i(y)$ denote the indicator of the event $\widehat{f_y}(\{i\})^2 \leq (2e)^{2k}\gamma_i$ and applying Corollary 2.4, we obtain that

$$\begin{aligned} & \mathop{\mathbb{E}}_y[\widehat{f_y}(\{i\})^2 \cdot (1 - \mathbf{1}_i(y))] \leq 0.13\gamma_i \\ \Rightarrow \quad & \mathop{\mathbb{E}}_y[\widehat{f_y}(\{i\})^2 \cdot \mathbf{1}_i(y)] \geq 0.87\gamma_i \\ \Rightarrow \quad & \mathop{\mathbb{E}}_y\left[\sum_{i\in U} \widehat{f_y}(\{i\})^2 \cdot \mathbf{1}_i(y)\right] \geq \frac{0.87}{32\log k} \quad \text{(using (6))}.\end{aligned}$$

Since by definition of the $\mathbf{1}_i$'s it holds for every y that

$$\sum_{i\in U}\widehat{f_y}(\{i\})^2\cdot\mathbf{1}_i(y)\le(2e)^{2k}\left(\sum_{i\in U}\gamma_i\right)\le(2e)^{2k},$$

we conclude that

$$\Pr_y\left[\sum_{i\in U}\widehat{f_y}(\{i\})^2\cdot\mathbf{1}_i(y)\ge\frac{.87}{64\log k}\right]\ge\frac{.87}{64(2e)^{2k}\log k}. \tag{7}$$

If U is made to be a random subset again, rather than a fixed one, combining (6) with (7) (and noting that $0.87/64>0.01$) yields

$$\Pr_{U,y}\left[\sum_{i\in U}\widehat{f_y}(\{i\})^2\cdot\mathbf{1}_i(y)\ge 0.01/\log k\right]\ge\exp(-O(k)). \tag{8}$$

Let us now condition on the event that U and y satisfy the condition in (8), namely

$$\sum_{i\in U}\widehat{f_y}(\{i\})^2\cdot\mathbf{1}_i(y)\ge 0.01/\log k. \tag{9}$$

Denote $g=f_y$, $T=\{i\in U:\mathbf{1}_i(y)=1\}$, and $\sigma^2=\sum_{i\in T}\widehat{g}(\{i\})^2$. Then $\deg(g)\le k$, and, by definition of T and of the $\mathbf{1}_i$'s,

$$\max_{i\in T}|\widehat{g}(\{i\})|\le(2e)^k\sqrt{\gamma_i}\le t^{-1}(\sqrt{C}/2e)^{-k}, \tag{10}$$

where the second inequality follows from (5).

We would now like to apply Lemma 4.1 to the function g/σ, the set T and the parameter t', where $t'=\max\{1,t/\sigma\}$. To see that the conditions of the lemma hold we use (10),(9) (if $t'>1$) and the fact that in case $t'=1$ we have $\sigma\ge 1$. These imply, with much room to spare (if C is a sufficiently large constant), that for every i

$$|\widehat{(g/\sigma)}(\{i\})|\le\sigma^{-1}t^{-1}(\sqrt{C}/2e)^{-k}\le\frac{1}{Kt'k}.$$

Therefore Lemma 4.1 indeed applies, and implies that

$$\begin{aligned}\Pr_z[|f_y(z)|\ge t]&=\Pr[|g(z)|\ge t]\ge\Pr[|g(z)|\ge\sigma t']=\Pr[|g(z)/\sigma|\ge t']\\ &\ge\exp(-O((t')^2k^2))\\ &\ge\exp(-O(t^2k^2\log k)).\end{aligned} \tag{11}$$

We now combine the lower bound (8) on the probability that (9) holds, with the lower bound (11) on the probability that $|f_y(z)| > t$ in case (9) holds, to get

$$\Pr_{U,y,z}[|f_y(z)| \geq t] \geq \exp(-O(k)) \cdot \exp(-O(t^2k^2 \log k)) = \exp(-O(t^2k^2 \log k)).$$

But when U, y and z are randomly chosen, $f_y(z)$ is just the value of f on a uniformly random $x \in \{1,-1\}^n$, and therefore $\Pr_{U,y,z}[|f_y(z)| \geq t]$ is nothing more than $\Pr_x[|f(x)| \geq t]$ where x is uniformly random. This completes the proof. ∎

5. Bounded non-juntas have tails

In this section we prove Theorem 1, showing that a bounded real-valued function whose Fourier tail is very small must be close to a junta (or in the contra-positive, that a bounded non-junta must have a nonnegligible Fourier tail). The idea of the proof is as follows. Given a bounded function (i.e. $|f| < 1$), with sufficiently rapidly decaying Fourier coefficients, it can be approximated by a low degree polynomial. This polynomial, in turn, can be approximated by a function depending only on those variables with large influence. Assuming, by way of contradiction, that this approximation is not a good one we get nonnegligible weight on the coordinates with small influence; this implies, by our previous results, that the original function was not bounded.

The proof is based on a slightly tweaked version of Theorem 3.

THEOREM 7: *There is a universal constant C such that the following holds: Suppose $f\colon \{-1,1\}^n \to \mathbb{R}$ has degree at most k, $J \subseteq [n]$ and assume*

$$\sum_{S \setminus J \neq \emptyset} \widehat{f}(S)^2 \geq \varepsilon.$$

Let $t \geq \sqrt{\varepsilon}$ and suppose that $\mathrm{Inf}_i(f) \leq \varepsilon^2 t^{-2} C^{-k}$ for all $i \notin J$. Then

$$\Pr[|f| \geq t] \geq \exp(-(Ct^2k^2 \log k)/\varepsilon).$$

Proof: Rescale by a factor of $(1/\sqrt{\varepsilon})$, letting $f' = (f/\sqrt{\varepsilon})$ and $t' = (t/\sqrt{\varepsilon})$. Now we may repeat the proof of Theorem 3 for f' and t', with the following alterations: s is chosen so that $\sum_{|S \setminus J| \sim 2^s} \widehat{f'}(S)^2 \geq 1/(2 \log k)$; and U is chosen at random from $[n] \setminus J$ rather than from $[n]$. ∎

Before we prove Theorem 1, let us restate it for convenience.

THEOREM 1: *Let* $f\colon \{-1,1\}^n \to [-1,1]$, $k \geq 1$, *and* $\varepsilon > 0$. *Suppose*

$$\sum_{|S|>k} \widehat{f}(S)^2 \leq \exp(-O(k^2 \log k)/\varepsilon).$$

Then f *is an* $(\varepsilon, 2^{O(k)}/\varepsilon^2)$*-junta.*

Proof: We may assume $\varepsilon < 1$, else the statement is trivial.

Let $g = f^{\leq k}$ be the k-degree part of f, and define

$$J = \{i \in [n] : \mathrm{Inf}_i(g) \geq \varepsilon^2 C^{-k}/16\},$$

where C is the constant from Theorem 7. Also, let

$$h = \sum_{|S| \leq k,\ S \subseteq J} \widehat{f}(S)\chi_S.$$

Note that by (1) we have

$$\sum_{i\in[n]} \mathrm{Inf}_i(g) = \sum_S |S|\widehat{g}(S)^2 = \sum_{|S|\leq k} |S| \cdot \widehat{f}(S)^2 \leq k \cdot \|f\|_2^2 \leq k.$$

Thus $|J|$ can be no bigger than $k/(\varepsilon^2 C^{-k}/16) = 2^{O(k)}/\varepsilon^2$. But h depends on the coordinates of J; thus if we can show that $\|f-h\|_2^2 \leq \varepsilon$, then f is an $(\varepsilon, 2^{O(k)}/\varepsilon^2)$-junta as claimed. To do this, we will show that $\|f-g\|_2^2 \leq \varepsilon/2$ and that $\|g-h\|_2^2 \leq \varepsilon/2$.

The first of these is easy: By the assumption on f,

$$\|f-g\|_2^2 = \sum_{|S|>k} \widehat{f}(S)^2 \leq \exp(-O(k^2 \log k)/\varepsilon) \leq \exp(-O(1)/\varepsilon) \leq \varepsilon/2.$$

The fact that $\|g-h\|_2^2 \leq \varepsilon/2$ follows from Theorem 7 applied to g. To see this, assume, for the sake of contradiction, that $\|g-h\|_2^2 > \varepsilon/2$. Then applying Theorem 7 with parameters J, $\varepsilon/2$, and $t = 2$ (the definition of J insures that these parameters can be used), we get

$$\Pr[|g| \geq 2] \geq \exp(-(8Ck^2 \log k)/\varepsilon).$$

But note that since $|f| \leq 1$ always, then whenever $|g| \geq 2$ there is a contribution of at least 1 to $\|f-g\|_2^2$. Thus we get $\|f-g\|_2^2 \geq \exp(-(8Ck^2 \log k)/\varepsilon)$ which contradicts the premise of the theorem, taking a large enough constant in the $O(\cdot)$. ∎

6. Nearly matching bounds

In this section we sketch proofs that the main results Lemma 1.3, Theorem 3, and Theorem 1 are nearly tight. Specifically, Lemma 1.3 is tight up to the constant in the exponent, and Theorems 3 and 1 are tight up to the $\log k$ factor in the exponent.

To show these tightness results we construct a family of functions as follows.

Lemma 6.1: *There exists a family of functions*

$$\{\Phi_{(n,k,t)}\colon \{-1,1\}^n \to \mathbb{R}\}_{n,k,t}$$

for $n \in \mathbb{N}$, $k \in \mathbb{N}$ odd, and $t \in [1,\infty)$, satisfying the following:

1. *$\Phi_{(n,k,t)}$ is symmetric, i.e., invariant under permutations of its variables;*
2. *$\Phi_{(n,k,t)}$ is of degree at most k;*
3. *There is a universal constant $C < \infty$ such that $\|\Phi_{(n,k,t)}\|_2 \le C$;*
4. *There is a universal constant $c > 0$ such that $\lim_{n\to\infty} \|(\Phi_{(n,k,t)})^{=1}\|_2 \ge c$;*
5. *For every $\alpha \ge 1$,*

$$\Pr[|\Phi_{(n,k,t)}| \ge \alpha t] \le \exp(-\Omega(t^2 k^2 \alpha^{2/k})).$$

The family $\{\Phi_{(n,k,t)}\}$ is constructed in the next subsection. In the mean time, let us explain why once constructed, it provides the stated tightness results.

Tightness for Lemma 1.3. The lemma is tight for any k and any $t \ge (2/c)$. To see this, take $t' = (c/2)t$ and let $f = (2/c)\Phi_{(n,k,t')}$, where n is chosen large enough so that $\|(\Phi_{(n,k,t')})\|_2 \ge c/2$, using property 4. Thus f has weight at least 1 on the first level. By property 3 this weight is also bounded above by M^2; hence by the symmetry of $\Phi_{(n,k,t')}$ and f, we get

$$\forall i, \quad |\widehat{f}(i)| = (2/c)|\widehat{\Phi_{(n,k,t')}}(i)| \le (2/c)\frac{M}{\sqrt{n}} < \frac{1}{K'tk},$$

increasing n if necessary. Thus f satisfies the hypotheses of the Lemma 1.3, but using property 5 (with $\alpha = 1$) it is easily shown that

$$\Pr[|f| \ge t] \le \exp(-\Omega(t^2k^2));$$

hence Lemma 1.3 is tight up to the constant in the exponent.

Tightness for Theorem 3. It is easy to check that if we scale the function f just constructed so that it satisfies $\sum_{S\neq\emptyset} \widehat{f}(S)^2 = 1$, then its influences will be smaller than $O(k)/n \ll t^{-2}C^{-k}$ and yet $\Pr[|f| \ge t] \le \exp(-\Omega(t^2k^2))$.

Tightness for Theorem 1. The following family of functions provides our tightness result for Theorem 1: for every odd k and every n, let

$$\Psi_{(n,k)}(x) = \begin{cases} \Phi_{(n,k,1)}(x) & \text{if } |\Phi_{(n,k,1)}| \le 1, \\ 0 & \text{if } |\Phi_{(n,k,1)}| > 1. \end{cases}$$

It is not very hard to check that $\|\Psi_{(n,k)} - \Phi_{(n,k,1)}\|_2^2 \le \exp(-\Omega(k^2))$. The idea is that this quantity is simply the contribution to the squared 2-norm of $\Phi_{(n,k,1)}$ from values exceeding 1. By property 5 of Lemma 6.1, the probability that $|\Phi_{(n,k,1)}|$ exceeds 1 is at most $\exp(-\Omega(k^2))$. Furthermore, the very rapid tail decay provided in property 5 ensures that almost all the contribution to the squared 2-norm comes from constant values of $|\Phi_{(n,k,1)}|$, and thus the estimate $\exp(-\Omega(k^2))$ is of the correct order.

Since $\|(\Phi_{(n,k,1)})^{>k}\|_2^2 = 0$, it follows that $\|(\Psi_{(n,k)})^{>k}\|_2^2 \le \exp(-\Omega(k^2))$, as desired. It only remains to check that $\Psi_{(n,k)}$ is not at all close to a junta. Let k be given, sufficiently large so that $\exp(-\Omega(k^2)) < c^2/4$. Then since $\Phi_{(n,k,1)}$ has weight at least $c^2/2$ on level 1 for all n large enough, it follows that $\Psi_{(n,k)}$ has weight at least $c^2/4$ on level 1 for all n large enough. But $\Psi_{(n,k)}$ is symmetric; thus any function that depends only on, say, $n/2$ of $\Psi_{(n,k)}$'s coordinates will have L_2^2-distance at least $c^2/8$ from $\Psi_{(n,k)}$, just from the level-1 contribution alone. Thus $\Psi_{(n,k)}$ is not even a $(c^2/8, n/2)$-junta.

6.1 The family $\{\Phi_{(n,k,t)}\}$. We will now sketch the proof of Lemma 6.1 by constructing the family $\{\Phi_{(n,k,t)}\}$ and indicating why the claimed properties hold.

Definition 2: For $x \in \{-1,1\}^n$, denote $s(x) = \sum_i x_i$. Now define

$$\Phi_{(n,k,t)} = t \cdot C_k\Big(\frac{s(x)/\sqrt{n}}{10tk}\Big),$$

where C_k is the Chebyshev polynomial of degree k.

Properties 1 and 2 are immediate. We next prove property 5, which is straightforward:

$$\begin{aligned} \Pr[|\Phi_{(n,k,t)}| \ge \alpha t] &= \Pr\Big[\Big|C_k\Big(\frac{s(x)/\sqrt{n}}{10tk}\Big)\Big| \ge \alpha\Big] \\ &\le \Pr\Big[\frac{1}{2}\Big(\frac{s(x)/\sqrt{n}}{5tk}\Big)^k \ge \alpha\Big] \qquad \text{(using Fact 2.9 and } \alpha \ge 1) \\ &= \Pr[s(x)/\sqrt{n} \ge 5tk(2\alpha)^{1/k}] \\ &\le \exp(-\Omega(t^2k^2\alpha^{2/k})) \qquad \text{(Chernoff bound).} \end{aligned}$$

Property 3 follows from property 5, essentially as already indicated in the discussion of the tightness for Theorem 1. The idea is simply that most of the time, $s(x) = \Theta(\sqrt{n})$, in which case a constant is contributed to $\|\Phi_{(n,k,t)}\|_2^2$; the tail decay provided by property 5 shows that this is the bulk of the contribution to the 2-norm.

It remains to show property 4 of $\Phi_{(n,k,t)}$.

Asymptotic first-level weight of $\{\Phi_{(n,k,t)}\}$. Fix k odd and $t \geq 1$. By symmetry,

$$\|(\Phi_{(n,k,t)})^{=1}\|_2^2 = n \cdot (\widehat{\Phi_{(n,k,t)}}(\{n\}))^2,$$

and therefore

$$\begin{aligned}\|(\Phi_{(n,k,t)})^{=1}\|_2 &= |\sqrt{n} \cdot \mathop{\mathbb{E}}_{x \in \{-1,1\}^n} [\Phi_{(n,k,t)} \cdot x_n]| \\ &= \Big|\sqrt{n} \cdot \mathop{\mathbb{E}}_{x \in \{-1,1\}^n} \Big[t \cdot C_k\Big(\frac{s(x)/\sqrt{n}}{10tk}\Big) \cdot x_n\Big]\Big| \\ &= \Big|\frac{1}{2}\sqrt{n} \cdot \mathop{\mathbb{E}}_{x \in \{-1,1\}^{n-1}} \Big[t \cdot \Big(C_k\Big(\frac{s(x)+1}{10tk \cdot \sqrt{n}}\Big) - C_k\Big(\frac{s(x)-1}{10tk \cdot \sqrt{n}}\Big)\Big)\Big]\Big|\end{aligned}$$

(note that the expectation is now over $\{-1,1\}^{n-1}$, not $\{-1,1\}^n$)

$$= \Big|\frac{1}{10k} \cdot \mathop{\mathbb{E}}_{x \in \{-1,1\}^{n-1}} \Big[5tk \cdot \sqrt{n} \cdot \Big(C_k\Big(\frac{s(x)+1}{10tk \cdot \sqrt{n}}\Big) - C_k\Big(\frac{s(x)-1}{10tk \cdot \sqrt{n}}\Big)\Big)\Big]\Big|.$$

To make sense of the expression above, note that for very large n the distribution of $s(x)/\sqrt{n}$ converges to that of a standard Gaussian. Also observe that as n increases while the value of $s(x)/\sqrt{n}$ is somehow "fixed", the expression inside the expectation brackets tends to the derivative of C_k at the point $\frac{s(x)/\sqrt{n}}{10tk}$. It is elementary, albeit tedious, to check that indeed

$$\lim_{n \to \infty} \|(\Phi_{(n,k,t)})^{=1}\|_2 = \Big|\frac{1}{10k} \cdot \mathop{\mathbb{E}}_{U \sim \mathbb{N}(0,1)} \Big[C_k'\Big(\frac{U}{10tk}\Big)\Big]\Big|;$$

the proof only uses the fact that the exponential tail behavior of Gaussians overcomes the polynomial growth of the Chebyshev polynomials.

The proof of property 4 is now implied by the following

Claim 6.2: *If $k \in \mathbb{N}$ is odd and $t \geq 1$, then*

$$\Big|\mathop{\mathbb{E}}_{U \sim \mathbb{N}(0,1)} \Big[C_k'\Big(\frac{U}{10tk}\Big)\Big]\Big| \geq k/3.$$

Proof: The idea is that with high probability, $U/(10tk)$ is in the range $[-1/k, 1/k]$, where $|C_k'|$ is $\Omega(k)$. With very slight probability, $|U/(10tk)|$ is as large as $1/2$, where $|C'_k|$ is still $O(k)$. And finally, although $|C'_k|$ increases exponentially as its argument becomes large enough, the probability of $U/(10tk)$ becoming large decays at an exponentially higher rate. Explicitly:

$$\left|\mathop{\mathbb{E}}_{U\sim\mathbb{N}(0,1)}\left[C'_k\left(\frac{U}{10tk}\right)\right]\right| \geq \Pr[|U/10tk| \leq 1/k]\cdot \inf_{|x|\leq 1/k} |C'_k(x)| \tag{12}$$

$$- \Pr[|U/10tk| \geq 1/k]\cdot \sup_{|x|\leq 1/2} |C'_k(x)| \tag{13}$$

$$- 2\int_{5tk}^{\infty} |C'_k(u/10tk)|\cdot\phi(u)du, \tag{14}$$

where ϕ is the density function of a standard normal.

Using Fact 2.10 and $t \geq 1$, we get the following: (12) is at least $\Pr[|U| \leq 10](k/2) \geq 0.49k$; (13) is at most $\Pr[|U| \geq 10](4/3)k \leq 0.01k$; and (15) is at most

$$2\int_{5tk}^{\infty} k(u/10tk)^{k-1}\phi(u)du \leq 2k\int_{5k}^{\infty} k(u/10k)^{k-1}\phi(u)du.$$

Certainly this integral becomes smaller as k increases and even at $k=1$ it is no more than 0.01. Putting the three estimates together completes the proof of the claim. ∎

A. The parameters in Theorem 6

Recall that we are considering the probability space $\{-1,1\}^n_{(p)}$, the p-biased measure on the discrete cube. Define $\theta = \theta(p) = q/p$.

Definition 3: For $\theta \in (0,\infty)$, $r \geq 1$, and r' the conjugate exponent of r, we define the following constant:

$$B^{(\theta)}(r) = \frac{\theta^{1/r'} - \theta^{-1/r'}}{\theta^{1/r} - \theta^{-1/r,}},$$

where in the case $\theta = 1$ the quantity is understood by taking the limit: $B^{(1)}(r) = r - 1$.

Note that $B^{(1/\theta)}(r) = B^{(\theta)}(r)$ and $B^{(\theta)}(r') = 1/B^{(\theta)}(r)$. For each θ, the quantity $B^{(\theta)}(r)$ increases from 0 at $r=1$ to 1 at $r=2$ and to ∞ as $r\to\infty$.

Oleszkiewicz [Ole02] proves the following generalization of the B.G.B. inequality:

THEOREM 8: *Suppose $f\colon \{-1,1\}^n_{(p)} \to \mathbb{R}$ has degree at most k. Then for all $r \geq 2$,*

$$\|f\|_r \leq [B^{(\theta)}(r)]^{k/2}\|f\|_2.$$

Since the quantity $B^{(\theta)}(r)$ is sometimes inconvenient, we use an estimate:

Definition 4: For $\theta \in (0, \infty)$ and $r \geq 1$, we define the following constant:

$$B(\theta) = B(1/\theta) = \frac{\theta - 1/\theta}{2 \ln \theta} = \frac{1}{2} \frac{q - p}{\ln q - \ln p},$$

where, once again, in the case $\theta = 1$ we take the limit and define $B(1) = 1$.

FACT A.1: For all $r \geq 1$, $B^{(\theta)}(r) \leq B(\theta) \cdot (r - 1)$.

B. Large-deviation proofs

This section contains the proofs omitted from Section 2.2.

Proof (Lemma 2.2): Let $r = (t^{2/k}/Be)$. By the assumption on t we have $r \geq 2$. Then

$$\begin{aligned} \Pr[|f| \geq t] &= \Pr[|f|^r \geq t^r] \\ &\leq \mathbb{E}[|f|^r]/t^r && \text{(By Markov's inequality)} \\ &= \|f\|_r^r/t^r \\ &\leq [(B \cdot (r-1))^{k/2} \, \|f\|_2]^r/t^r && \text{(By Theorem 6)} \\ &\leq [(Br)^{k/2}/t]^r = \exp(-(k/2Be) \, t^{2/k}) && \text{(by definition of } r\text{).} \end{aligned}$$

Proof (Lemma 2.3): Let $r = t^{2/k}/Be \geq 2$ as before. Then

$$\begin{aligned} \mathbb{E}[f^2 \cdot \mathbf{1}_{\{f^2 > t^2\}}] &\leq \|f^2\|_{r/2} \cdot \|\mathbf{1}_{\{f^2 > t^2\}}\|_{(r/2)'} && \text{(by Hölder's inequality)} \\ &= \|f\|_r^2 \cdot \Pr[|f| > t]^{1-2/r} \\ &\leq (B \cdot r)^k \cdot \exp(-(k/2Be) \, t^{2/k})^{1-2/r} \\ & && \text{(by Theorem 6 and Lemma 2.2)} \\ &= t^2 \exp(-(k/2Be) \, t^{2/k}) && \text{(by definition of } r\text{).} \end{aligned}$$ ∎

Proof (Corollary 2.4): For $k = 0$, the claim is easy to verify, and we therefore assume $k \geq 1$. Applying Lemma 2.3 we get an upper bound of $(2Be)^{2k} \exp(-2Bek)$. Since $B \geq 1$, this term increases if we substitute 1 for both B and k. Hence the upper bound is at most $(2e)^2 \exp(-2e) < 0.13$. ∎

Proof (Lemma 2.5): We may assume, without loss of generality, that $\mathbb{E}[f] = 0$ and that $\|f\|_2 = 1$ (otherwise apply an appropriate linear transformation to f), this fixes the first and second moments of f. Using Corollary 2.4 we also have for every positive t a bound on the contribution to the second moment of f, coming from values larger than t in absolute value. To use this bound, we employ the following elementary fact: For any real $t > 0$, the following inequality holds for every $x \in \mathbb{R}$:

$$\mathbf{1}_{\{x<0\}} \leq 1 - \frac{x}{2t} - \frac{x^2}{2t^2} + \frac{9}{16t^2}(x^2 \cdot \mathbf{1}_{\{x^2>t^2\}}). \tag{15}$$

To see this, we check the inequality for the two cases $|x| \leq t$ and $|x| > t$. In the first case the right-hand side is $1 - x/2t - x^2/(2t^2)$, a parabola which is 1 at $x = -t$ and $x = 0$, and 0 at $x = t$; thus it is clear the right-hand side exceeds $\mathbf{1}_{\{x<0\}}$ on the range $[-t, t]$. In the second case the right-hand side is $1-x/2t+x^2/(16t^2)$, a parabola which is 1 at $x = 0$ and has its vertex at $x = 4t$ at which point it is 0; thus it is clear the right-hand exceeds $\mathbf{1}_{\{x<0\}}$ everywhere.

Now let $t_0 = (2Be)^k$, and consider (15) with parameter t_0. Substituting f for x in the inequality and taking expectations of both sides, we get

$$\begin{aligned}\Pr[f < 0] &\leq 1 - \mathbb{E}[f]/2t_0 - \mathbb{E}[f^2]/(2t_0^2) + (9/16t_0^2)\mathbb{E}[f^2 \cdot \mathbf{1}_{\{x^2>t_0^2\}}] \\ &= 1 - 1/(2t_0^2) + (9/16t_0^2)\mathbb{E}[f^2 \cdot \mathbf{1}_{\{x^2>t_0^2\}}] \\ &\leq 1 - 1/(2t_0^2) + (9/16t_0^2)(.13) \quad \text{(by Corollary 2.4)} \\ &< 1 - .4/t_0^2.\end{aligned}$$

Noting that $(2e)^2 < 30$ completes the proof. ∎

References

[ADFS03] N. Alon, I. Dinur, E. Friedgut and B. Sudakov, *Graph products, fourier analysis and spectral techniques,* Geometric and Functional Anlysis **14** (2004), 913–940.

[Bec75] W. Beckner, *Inequalities in Fourier analysis,* Annals of Mathematics **102** (1975), 159–182.

[Bon70] A. Bonami, *Etude des coefficients fourier des fonctiones de $l^p(g)$,* Annales de l'Institut Fourier **20(2)** (1970), 335–402.

[Bou02] J. Bourgain, *On the distribution of the Fourier spectrum of boolean functions,* Israel Journal of Mathematics **131** (2002), 269–276.

[Fri98] E. Friedgut, *Boolean functions with low average sensitivity depend on few coordinates,* Combinatorica **18** (1998), 27–36.

[FKN02] E. Friedgut, G. Kalai and A. Naor, *Boolean functions whose fourier transform is concentrated on the first two levels and neutral social choice,* Advances in Applied Mathematics **29** (2002), 427–437.

[Gro75] L. Gross, *Logarithmic Sobolev inequalities*, American Journal of Mathematics **97** (1975), 1061–1083.

[Jan97] S. Janson, *Gaussian Hilbert Spaces*, Cambridge Tracts in Mathematics, **129**, Cambridge University Press, Cambridge, 1997.

[KKL88] J. Kahn, G. Kalai and N. Linial, *The influence of variables on Boolean functions*, in *Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science*, IEEE Computer Society, Washington, DC, 1988, pp. 68–80.

[KN05] S. Khot and A. Naor, *Nonembeddability theorems via Fourier analysis*, Manuscript.

[KKMO04] S. Khot, G. Kindler, E. Mossel and R. O'Donnell, *Optimal inapproximability results for MAX-CUT and other 2-variable CSPs?* in *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, IEEE Computer Society, Washington, DC, 2004, pp. 146–154.

[KS] G. Kindler and S. Safra, *Noise-resistant boolean functions are juntas*, Manuscript.

[LT91] M. Ledoux and M. Talagrand, *Probability in Banach Spaces*, Ergelonisse der Mathematik und ihrer Gernzgebiete (3) [Results in Mathematics and Related Areas (3)], 23, Springer-Verlag, Berlin, 1991.

[L] A. Lincoln, *Public address*, circa, 1860.

[Ole02] K. Oleszkiewicz, *On the non-symmetric khinchine-kahane inequality*, Private communication, 2002.

[Riv90] T. J. Rivlin, *Chebyshev Polynomials*, John Wiley & Sons, New York, second edition, 1990.