Noise-Resistant Boolean-Functions are Juntas

Guy Kindler[*] Shmuel Safra[*]

April 17, 2003

Abstract

We consider Boolean functions over n binary variables, and a general p-biased, product measure over the inputs. We show that if f is of low-degree, that is, so that the weight of f on the Fourier-Walsh products of size larger than k is small, then f is close to a junta, namely, a function which depends only on very small, related to k however unrelated to n, number of variables.

We conclude that juntas are the only highly noise-resistant Boolean functions.

Furthermore, we manage to utilize such a statement to prove an alternative switching lemma, one which may prove useful in the study of computational-complexity lower-bounds, in particular to a completely analytical proof that any AC^0 function is close to low-degree.

1 Introduction

An n-voters-binary voting system is one in which each voter casts a binary value –either "for" or "against" some proposition– and the system decides the outcome of the vote according to the input values. In other words, such a system is a Boolean function over n Boolean variables. Now suppose that the votes are transmitted over a noisy channel and therefore might each, independently flip with some probability λ. Say we would like to design a voting system that is robust against such noise. That is, one in which the probability of the outcome to change due to such noise is small. One way to ensure the probability the outcome changes when noise is introduced small, is to consider those voting systems in which only one of the voters count —such systems are referred to as *dictatorships*— in which case the probability the outcome of the system changes is simply the probability that the dictator's vote has change, which is λ. Can one come up with a system with smaller noise-sensitivity? Do other systems, which take into account a nonnegligible fraction of the votes but are nevertheless noise-resistant exist? A simple corollary of the main theorem of this paper answers this question with a definitive "no": Corollary 4.1 states that any noise resistant voting system is almost completely determined by a small number of the voters, a number depending only on the noise-probability λ.

To be a bit more formal, let us denote a Boolean function f over the n-dimensional binary cube as

$$\mathsf{f}\colon \{-1,1\}^n \rightarrow \{-1,1\}$$

[*] School of Mathematical Sciences, Tel Aviv University, Israel.

(We arbitrarily decide to represent the binary range, both for the input variables and the output, by the values -1 and 1.)

By interpolation, one can represent any real function over the binary cube as a polynomial. Since on the discrete cube variables take their values in $\{-1, 1\}$, all exponents are without loss of generality either 1 or 0. Hence the relevant monomials are all products of subsets of variables S

$$\chi_S(x) \doteq \prod_{i \in S} x_i$$

These functions (one for each subset S of the variables) are multiplicative, and are thus referred to as a *characters*. The representation of a function $\mathsf{f}\colon \{-1,1\}^n \to \mathbb{R}$ as a polynomial is referred to as its Fourier-Walsh expansion. The coefficient of the character $S \subseteq [n]$ of a function f in this expansion is denoted $\widehat{\mathsf{f}}(S)$, and thus

$$\mathsf{f} = \sum_S \widehat{\mathsf{f}}(S) \cdot \chi_S$$

Now assume we have a Boolean function that turns out to be close to a degree-k polynomial, that is where the sum of the squares of coefficients for all characters of size larger than k is smaller than $k^{-(\frac{1}{2}+\eta)}$. Bourgain [Bou01] proven that any such function is almost completely determined by a small number of variables – a number exponential in k nevertheless completely independent of the number of variables n.

A function that depends on at most J variables is called, in accordance with the dictatorship terminology, a *J-junta*. A sufficient condition for a function to be close to a junta, such as Theorem 1, has many implications in various fields.

For example, a Boolean function over the binary-hypercube can be considered as a binary-string of length 2^n, having one entry for for each binary representation of an index between 1 to 2^n. A code is accordingly a subset of the Boolean functions, in which every pair of functions disagree on at least the *distance* d of the functions' plausible inputs. The Hadamard code then consists of all multiplicative functions, that is, all characters.

The long-code, which has been used in numerous hardness of approximation results [BGS98, Hås97, Hås99, DS02, Kho02], would comprise all increasing dictatorship (all functions $\mathsf{f}(x) = x_i$). When using the long-code to prove hardness of approximation, one is typically required to show that a Boolean function that satisfies a certain condition, must have a short list-decoding as a code-word. The sufficient condition above, for a function to be close to a junta, implies exactly that. The code-word associated with any variable $i \in [n]$ cannot have a nonnegligible correlation with f unless it is one of the variables in the junta determining f.

The biased distribution. We note that for some results (e.g. [DS02]), the correlation is weighted according to the p-biased measure on $\{-1,1\}^n$, for which Theorem 1 does not extend (as shown below). In the p-biased measure over the binary cube, each variable is independently set to -1 with probability p and to 1 with probability $1-p$, namely

$$\mu_p(x) \doteq p^{|x|}(1-p)^{n-|x|}$$

where $|x|$ denotes the number of variables assigned -1 in x.

Another, seemingly completely different field, in which analysis with respect to the biased measure is essential, is that of graph properties: a graph-property is a set of graphs closed under all permutations of the vertices. An example for a graph-property is all graphs containing a triangle, or all the connected graphs. A graph G over a given set of vertices can be represented by a sequence of bits, one for each potential edge, that determine which edges belong to G. Hence a graph-property can be represented as a Boolean function $\mathsf{f}\colon V^2 \to \{-1,1\}$ which is invariant under permutation of the vertices.

One may be specifically interested in monotone graph properties, in which the function is monotone. For a monotone property f, one can consider the probability that a graph satisfies the property, when it is randomly chosen according to the Erdos-Renyi random-graphs distribution, namely where the graph contains each potential edge with probability p, independently of other edges. This corresponds to choosing the input for f according to the p-biased measure. Since f in monotone, the probability of a graph to have the property is increasing in p. Some monotone properties have a *sharp-threshold*: there exists a p so that the probability for f to be satisfied is very small, and such that the probability of f to hold for $p+\gamma$, for a very small γ, is very high. One way of proving that f has a sharp threshold at p, is by showing that if it has a coarse threshold it must be a junta. Note that a symmetric property cannot be a junta unless it is constant. Therefore by a contrapositive argument, f must have a sharp threshold.

Another application of such a sufficient condition for a function to be a junta is with regards to circuit complexity, and in particular as to lower-bounds on the computational complexity of some functions. One of the most basic tools in these regards are switching lemmas [Ajt83, Has86], which were utilized, for example, to show the *parity* function ($=\chi_{[n]}$) is not computable in AC^0 (the class of circuits of non bounded fan-in and of constant depth).

An alternative view and an extension of that lower-bound, as well as an alternative to the original proof of Hastad is the result of Linial, Mansour and Nisan [LMN89], where it is proven that all functions computed by an AC^0 circuit are close to being low-degree, namely have almost all of their weight on characters of small size (the weight of f on characters of size at most k is defined by $\sum_{|S|\le k}\widehat{\mathsf{f}}(S)^2$). This immediately excludes the Parity function, and moreover, shows that many other functions are not in AC^0. The main technical tool in the proof of [LMN89] is still the switching lemma of [Has86].

A typical switching lemma shows that a random restriction of a Boolean function in a given class, is with high probability a very simple function (e.g. depends on a constant number of variables). Completely analytical proofs for switching lemmas has been sought after for some time. In this paper (see Section 4), we show how our results imply certain switching lemmas, which can perhaps be of help in the study of AC^0.

Our Results

In this paper we extend the result of [Bou01] to the case of the p-biased measure μ_p. For a bias p, we deal with the expansion of the Boolean function f as a linear combination of p-biased Walsh-products (the biased-measure analogue of the Fourier-expansion of f, as defined in [Tal94]. See also Section 2). It states that if f has very small weight on 'large' Walsh-products, then it must be close to a junta. Let us first define formally the notion of being close to the junta, and then state our main results.

(ϵ, J)-juntas. A Boolean function f is an (ϵ, J)-junta with respect to the biased measure μ_p, if it is 'almost-completely determined' by the coordinates in a set $\mathcal{J}\subseteq[n]$ of size at most J. That is, if there exists a function f' which only depends on coordinates in $\mathcal{J}$, such that if a random input x is chosen according to μ_p, then $\Pr_x[\mathsf{f}(x) = \mathsf{f}'(x)] \geq 1 - \epsilon$.

Theorem 1. *Let $\mathsf{f}\colon \mathcal{P}(-1,1)^n \to \{-1,1\}$ be a Boolean function satisfying $\sum_{|S|>k} |\widehat{\mathsf{f}}(S)|^2 \leq (\epsilon/k)^2$, where the coefficients $\widehat{\mathsf{f}}(S)$ are taken with respect to the p-biased Walsh-products. Then f is an $[O(\epsilon \log(1/p)/p^2), J]$-junta with respect to μ_p, for $J = O(\epsilon^{-2}k^3p^{-k})$.*

Better parameters than the ones above are achieved if the bias p is essentially constant.

Theorem 2. *Fix a positive integer ℓ. Then there exists a constant $\delta_p > 0$, such that for every $\epsilon > 0$ and every Boolean function $\mathsf{f}\colon \mathcal{P}([n]) \to \{-1,1\}$ satisfying*

$$\sum_{|S|>k} |\widehat{\mathsf{f}}(S)|^2 \leq \left(\frac{\epsilon}{k}\right)^{(\ell+1)/\ell}$$

is an $(O(\epsilon), J)$-junta, where*

$$J = O\left(\delta_p^{-4k} \epsilon^{(\ell+1)/\ell} k^{(2\ell+1)/\ell}\right)$$

The parameters obtained by the above theorems are different than those in [Bou01].

Theorem [Bou01]. *Let $\mathsf{f}\colon \{-1,1\}^n \to \{-1,1\}$ be a Boolean function, and let k be a positive integer and $\epsilon, \eta > 0$ any fixed constants. Then there is a constant $c_{\eta,\epsilon}$, such that if $\sum_{|S|>k} |\widehat{\mathsf{f}}(S)|^2 < c_\eta k^{-\frac{1}{2}-\eta}$, then there exists a Boolean function $\mathsf{h}\colon \mathcal{P}([n]) \to \{-1,1\}$ which depends on $k10^k$ variables, and for which $\mathsf{f}(x) = \mathsf{h}(x)$ for all but at most an ϵ-fraction of the inputs $x \in \{-1,1\}^n$.*

The main difference between our results and that of Bourgain, is the dependence on k of the threshold on $\sum_{|S|>k} \left|\widehat{\mathsf{f}}(S)\right|^2$, beyond which f is ensured to be close to a junta of constant size. Here, it is required that this weight be bounded by $O(k^{-(\ell+1)/\ell})$, while the result in [Bou01] requires only a bound of order $K^{-1/2-\eta}$ for every constant $\eta > 0$. Our proof of Theorem 5 is, however, different than that of Bourgain, and its technique may of of independent interest. Also note that the ensured distance of f from a junta can be greatly improved, if one assumes that $\sum_{|S|>k} \left|\widehat{\mathsf{f}}(S)\right|^2$ is *very* small (see Section 6 and Section 7).

As a direct corollary of Theorem 1, we have that juntas are the only noise-resistant Boolean functions. The λ-noise-sensitivity of a Boolean function f with respect to μ_p, is the probability that f yields the same value when evaluated on a random input x, and then re-evaluated on an input x', where a λ-fraction of the coordinates are randomly re-assigned (an exact definition appears in Section 4). We show that a Boolean function f whose noise-sensitivity is small, must be ϵ-close to a junta of size independent of n.

Corollary 1.1. *For any parameter $\lambda > 0$, fix $k = \log_{(1-\lambda)}(1/2)$. Then every Boolean function $\mathsf{f}\colon \mathcal{P}([n]) \to \{-1,1\}$ whose λ-noise-sensitivity with respect to μ_p^n is bounded by $(\epsilon/k)^2$, is an $[O(\epsilon \log(1/p)/p^2), J]$-junta, where*

$$J = O\left(\frac{k^3}{\epsilon^2 p^k}\right)$$

A similar corollary, with improved parameters for constant bias p, follows from Theorem 2.

*The O notation here hides constants which are independent of ϵ, k, and n. However they may depend on the bias, p, and on ℓ.

Structure of the paper

In Section 2, we present the biased Walsh-products which replace the usual Fourier-basis. In this Section we also define other basic notions such as restrictions and variations of a function f, and show their connections with the expansion of f as a combination of biased Walsh-products. In addition, we state in the same section a biased version for the Beckner-Bonami hyper-contractive estimate. The proof of this estimate is deferred to Section 9. In Section 3 we give a simple conceptual proof of Theorem 1, which contains the main arguments in this paper (in fact a slightly stronger result is shown).

In Section 4, we prove Corollary 1.1, showing that a function with small noise-sensitivity is close to a junta. We also show why Theorem 1 does not hold "out of the box" for the biased case. In Section 5 we show a version of the switching lemma that is obtained from our results.

In Section 6 we show an alternative proof for a theorem of [FKN01], showing that a Boolean function which is almost linear is close to a dictatorship. This proof not only extends to the case of biased-measure, but is also extendable to the case of higher frequencies. This is done in Section 7, yielding a theorem that is similar to Theorem 2, but holds only for functions whose weight beyond the k'th level is extremely small. Finally, in Section 8 we prove Theorem 2 simply by plugging the result from Section 7 into the proof of Theorem 5. We then obtain, in the same appendix, a corollary similar to Corollary 1.1.

2 Preliminaries

It will be more convenient in the sequel to deal with Boolean functions of the form $\mathsf{f}\colon \mathcal{P}([n]) \to \{-1, 1\}$, where $\mathcal{P}([n])$ denotes the power-set of $[n] \doteq \{1, 2, \ldots, n\}$. The elements of $\mathcal{P}([n])$ can, of course, be easily identified with those of $\{-1, 1\}^n$ (the value of the i'th variable determines whether the argument of f contains i or not).

The biased measure is defined, in this notation, as follows. For a finite set I and $0 < p < 1$, define a probability measure μ_p^I on $\mathcal{P}(I)$ by

$$\forall\ A \subseteq I\,, \qquad \mu_p^I(A) \doteq p^{|A|}(1-p)^{|I \setminus A|}$$

Throughout this paper we assume (without loss of generality) that $0 < p \leq 1/2$. Also, we abbreviate μ_p^n for $\mu_p^{[n]}$.

2.1 Discrete Fourier Expansion

We next define the basic notions we need concerning the space of real-valued functions over $\mathcal{P}([n])$.

Inner-products and norms. The biased inner-product of two real-valued functions f, g over $\mathcal{P}([n])$ is defined by $\langle \mathsf{f}, \mathsf{g} \rangle \doteq \mathbb{E}_{x \sim \mu_p^n}\left[\mathsf{f}(x)\mathsf{g}(x)\right]$. The q-norm of a function $\mathsf{f}\colon \mathcal{P}([n]) \to \mathbb{R}$, is defined by $\|\mathsf{f}\|_q \doteq \left(\mathbb{E}_{x \sim \mu_p^n}\left[|\mathsf{f}(x)|^q\right]\right)^{1/q}$.

Fourier basis. The usual Fourier basis for the space of functions $\mathsf{f}\colon \mathcal{P}([n]) \to \mathbb{R}$ is not orthonormal (or even orthogonal) with respect to the biased inner-product. Following [Tal94] we define an analogue basis, which is orthonormal with respect to the biased inner-product (for $p = 1/2$, it is the usual Walsh/Fourier basis). Like the Fourier basis it is a "tensorised" basis, containing products of functions each of which depending on only one coordinate, and having expectation zero and variance one.

Definition 1 (biased Walsh-Products). *Let $0 < p < 1$. For every $i \in [n]$, we define the i'th p-biased Rademacher function $\chi_{\{i\}} : \mathcal{P}([n]) \to \mathbb{R}$ by*

$$\chi_{\{i\}}(x) \doteq \begin{cases} \sqrt{p/(1-p)} & i \notin x \\ -\sqrt{(1-p)/p} & i \in x \end{cases}$$

For every set $S \subseteq [n]$, the p-biased Walsh-product that corresponds to it is then defined by $\chi_S \doteq \prod_{i \in S} \chi_{\{i\}}$. It is said that χ_S has frequency $|S|$, or that it has size $|S|$.

Since the set of biased Walsh-products forms an orthonormal basis, we have that every function $\mathsf{f} : \mathcal{P}([n]) \to \mathbb{R}$ can be written as a linear combination $\mathsf{f} = \sum_{S \subseteq [n]} \widehat{\mathsf{f}}(S)\chi_S$, called the Fourier expansion of f, where $\widehat{\mathsf{f}}(S) = \langle f, \chi_S \rangle$.

2.2 Projections

An important aspect of the Fourier representation is that it enables the definition and analysis of simple but important orthonormal projections of f.

Frequency separation. The first two projections, which are crucial in this work, separate the Walsh-products into low-frequencies and high-frequencies. For a given k and a function f denote $\mathsf{f}^{\leq k} = \sum_{|S| \leq k} \widehat{\mathsf{f}}(S)\chi_S$ and $\mathsf{f}^{>k} = \sum_{|S| > k} \widehat{\mathsf{f}}(S)\chi_S$.

The averaging projection. Let $I \subseteq [n]$ be a set of coordinates. For a function $\mathsf{f} : \mathcal{P}([n]) \to \mathbb{R}$, consider the function obtained from it by averaging , for each element $x \in \mathcal{P}([n])$, over all distinct input settings y such that $x \cap I = y \cap I$ (the average is weighted according to the biased weights). This is a real-valued function, $\mathsf{Avg}_I[\mathsf{f}] : \mathcal{P}([n]) \to \mathbb{R}$, that depends only on $\bar{I} = [n] \setminus I$, and is formally defined as $\mathsf{Avg}_I[\mathsf{f}](x) \doteq \mathbb{E}_{z \sim \mu_p^I}[\mathsf{f}((x \setminus I) \cup z)]$. One can easily verify that Avg_I is the projection onto the set of p-biased Walsh-products whose support is disjoint from I, namely $\mathsf{Avg}_I[\mathsf{f}] = \sum_{S \cap I = \emptyset} \widehat{\mathsf{f}}(S)\chi_S$.

2.3 Variations

The *variation* of a Boolean function $\mathsf{f} \colon \mathcal{P}([n]) \to \{-1, 1\}$ on a subset $I \subseteq [n]$ of the coordinates measures the dependency of f on I. The variation of f on a singleton $\{i\}$ coincides with the classical definition ([BL89, KKL88]) of the influence of the i'th coordinate on f. We define the variation of f on I with respect to μ_p to be twice the probability that f yields different values, given two random inputs that agree on all the coordinates outside I, that is

$$\mathsf{Vr}_{\mathsf{f}}(I) = 2 \Pr_{\substack{y \sim \mu_p^{[n] \setminus I} \\ z_1, z_2 \sim \mu_p^I}} [\mathsf{f}(y \cup z_1) \neq \mathsf{f}(y \cup z_2)]$$

(the 2 factor is for compatibility with the definition of the influence of a coordinate).

Next is an alternative definition of the variation (as can be easily verified), which extends to the case of non-Boolean functions. The bias-parameter p here is implicit.

Definition 2 (variation). *The* variation *of a function* $\mathsf{f}\colon \mathcal{P}([n]) \to \mathbb{R}$ *on a set* $I \subseteq [n]$ *of coordinates is defined by* $\mathsf{Vr}_{\mathsf{f}}(I) \doteq \|\mathsf{f} - \mathsf{Avg}_I[\mathsf{f}]\|_2^2 = \sum_{S\cap I\neq\emptyset} \widehat{\mathsf{f}}^2(S)$.

It easily follows from the above definition that the variation is sub-additive, namely $\mathsf{Vr}_{\mathsf{f}}(I_1 \cup I_2) \leq \mathsf{Vr}_{\mathsf{f}}(I_1) + \mathsf{Vr}_{\mathsf{f}}(I_2)$.

The following proposition justifies our view of the variation as a measure of dependency, showing that a Boolean function f whose variation on a given set of coordinates is small, is indeed almost independent of the coordinates in that set. Putting it differently, if the variation of f on the complement of a set I is small, than f is close to a Boolean function g which depends only on the coordinates in I.

Proposition 2.1. *Let* $\mathsf{f}\colon \mathcal{P}([n]) \to \{-1,1\}$ *be a Boolean function, and let* $I \subseteq [n]$ *be a set of coordinates. Then there exists a Boolean function* $\mathsf{g}\colon \mathcal{P}([n]) \to \{-1,1\}$ *which depends only on coordinates from* I*, and satisfies* $\Pr_{x\sim\mu_p^n}[\mathsf{f}(x) \neq \mathsf{g}(x)] < \mathsf{Vr}_{\mathsf{f}}([n]\setminus I)/2$.

Proof. Denote $\bar{I} \doteq [n]\setminus I$, and let $\mathsf{g} \doteq \mathrm{sign}\left(\mathsf{Avg}_{\bar{I}}[\mathsf{f}]\right)$ (we arbitrarily set $\mathrm{sign}(0) \doteq 1$). Then g depends only on coordinates from I. Let us show that $\mathsf{f}(x) = \mathsf{g}(x)$ for most x's.

For $y \in \mathcal{P}(I)$, denote

$$\alpha(y) \doteq \Pr_{z\sim\mu_p^{\bar{I}}}[\mathsf{f}(y\cup z) \neq \mathsf{g}(y\cup z)]$$

and note that $\alpha(y) \leq 1/2$ for all y. Therefore, we have

$$\begin{aligned}\Pr_{x\sim\mu_p^n}[\mathsf{f}(x) \neq \mathsf{g}(x)] &= \mathop{\mathbb{E}}_{y\sim\mu_p^I}[\alpha(y)] \leq \mathop{\mathbb{E}}_{y\sim\mu_p^I}[2\alpha(y)(1-\alpha(y))] = \\ &= \Pr_{\substack{y\sim\mu_p^I\\ z_1,z_2\sim\mu_p^{\bar{I}}}}[\mathsf{f}(y\cup z_1) \neq \mathsf{f}(y\cup z_2)] = \frac{1}{2}\mathsf{Vr}_{\mathsf{f}}(\bar{I})\end{aligned}$$

■

Low-frequency variation. According to Definition 2, the variation of f on a set I equals its weight on the set of characters χ_S such that S intersects I. We separate the contribution of the low-degree characters to the variation, defining for any function $\mathsf{f} : \mathcal{P}([n]) \to \mathbb{R}$ and a subset $I \subseteq [n]$, $\mathsf{Vr}_{\mathsf{f}}^{\leq k}(I) \doteq \sum_{\substack{S\cap I\neq\emptyset\\ |S|\leq k}} \widehat{\mathsf{f}}^2(S)$. For shortness, denote $\mathsf{Vr}_{\mathsf{f}}^{\leq k}(i) \doteq \mathsf{Vr}_{\mathsf{f}}^{\leq k}(\{i\})$.

2.4 Restrictions

For a given set of coordinates $I \subseteq [n]$ and every $x \in \mathcal{P}([n]\setminus I)$, let us denote by $\mathsf{f}_I[x]\colon \mathcal{P}(I) \to \{-1,1\}$ the Boolean function defined by $\forall\, y \in \mathcal{P}(I)\,,\ \mathsf{f}_I[x](y) \doteq \mathsf{f}(x\cup y)$. The Fourier expansion of $\mathsf{f}_I[x]$ can be deduced from the Fourier expansion of f. For every $S \subseteq I$, it is easily seen that

$$\widehat{\mathsf{f}_I[x]}(S) = \sum_{\substack{T\subseteq[n]\\ T\cap I=S}} \widehat{\mathsf{f}}(T)\chi_{T\setminus S}(x)$$

where $\chi_{T\setminus S}$ can in fact be replaced by χ_T.

Note that the variation on a subset I of a function f can be expressed in terms of restrictions – it is the expected variance of $\mathsf{f}_I[x]$, over all input settings $x \sim \mu_p^{\bar{I}}$ outside I. This leads immediately to the following claim.

Claim 2.2. *Let* f *be a Boolean function, and let* $I \subseteq [n]$. *Then* $\mathsf{Vr}_{\mathsf{f}}(I) = \mathbb{E}_{x \sim \mu_p^{\bar{I}}}\left[\mathsf{Vr}_{\mathsf{f}_I[x]}(I)\right]$.

2.5 Bonami-Beckner Inequality

We define for every $0 \le \delta \le 1$ an operator T_δ over real-valued functions $\mathsf{f}\colon \mathcal{P}([n]) \to \mathbb{R}$. At each point x, $T_\delta[\mathsf{f}](x)$ is the expected value of f when a $(1-\delta)$-fraction of the coordinates in x are randomly re-assigned (we say that a $(1-\delta)$-noise is applied to x). Thus

$$T_\delta[\mathsf{f}](x) = \mathop{\mathbb{E}}_{I \sim \mu^n_{(1-\delta)},\, z \sim \mu_p^I}\left[\mathsf{f}((x \setminus I) \cup z)\right]$$

Since T_δ is obviously a linear operator, and by evaluating it on biased Walsh-products, one easily verifies that $T_\delta[\mathsf{f}] = \sum_S \delta^{|S|}\widehat{\mathsf{f}}(x)\chi_S$

Bonami and Beckner independently proved that, in the case of uniform measure, T_δ is hyper-contractive for appropriate values of δ:

Theorem 3. *Let* $q \ge r \ge 1$, *and let* $\mathsf{f}\colon \mathcal{P}([n]) \to \mathbb{R}$. *Then in the uniform case, namely when the norms are taken with respect to* $\mu^n_{1/2}$,

$$\|\, T_\delta[\mathsf{f}]\,\|_q \le \|\mathsf{f}\|_r \quad \textit{for any} \quad \delta \le \sqrt{(r-1)/(q-1)}\,.$$

In [Fri98], a special case of Theorem 3 was shown to hold for the biased case as well. We prove another special case of this theorem, which is sufficient for our purposes.

Theorem 4. *For every* $p > 0$ *there exists a parameter* $\delta_p > 0$, *such that for every* $\delta \le \delta_p$ *and every function* $\mathsf{f}\colon \mathcal{P}([n]) \to \mathbb{R}$, $\|\, T_\delta[\mathsf{f}]\,\|_4 \le \|\mathsf{f}\|_2$, *where the norms and the operator* T_δ *are taken with respect to* μ_p^n.

This theorem is proven in Section 9. Note that in the sequel all the parameters denoted δ_p refer, unless noted otherwise, to the best parameter δ_p for which Theorem 4 holds. This best parameter was, in fact, found recently by K. Oleszkiewicz.

Theorem [Ole02]. *Let* δ_p *denote the largest parameter for which Theorem 4 holds. Then*

$$\delta_p = (1 + p^{-1/2}(1-p)^{-1/2})^{-1/2} = O(p^{1/4})$$

3 The Main Arguments

This section introduces the main ideas used throughout this paper. The main result of this section, Theorem 1, shows that a function f whose weight is concentrated on low-frequencies is close to a junta. In fact we show a slightly more concrete result, showing exactly which are the coordinates that determine most values of f. It is the set $\mathcal{J}$ of coordinates i whose low-degree influence on f is large (namely the coordinates i for which $\mathsf{Vr}_{\mathsf{f}}\!\le k(i)$ is large enough).

The parameters which are achieved here are improved in Section 8 for constant biases p, by repeating the proof and plugging in the parameters obtained from Theorem 10, which is proven in Section 7.

Definition 3. *Let* $\mathsf{f}\colon \mathcal{P}([n]) \to \{-1,1\}$ *be a Boolean function. For every parameter* $\tau > 0$ *and integer* $k > 0$*, let*

$$\mathcal{J}_{k,\tau}(f) \doteq \left\{ i \in [n] \mid \mathsf{Vr}_{\mathsf{f}}^{\leq k}(i) > \tau \right\}$$

Since for every Boolean function f, $\|\mathsf{f}\|_2^2 = 1$, one easily observes that $|\mathcal{J}_{k,\tau}(f)| \leq k/\tau$ for every such function.

The main result of this section is the following theorem, which is a more specified version of Theorem 1.

Theorem 5. *Every Boolean function* $\mathsf{f}\colon \mathcal{P}([n]) \to \{-1,1\}$ *satisfying* $\|\mathsf{f}^{>k}\|_2^2 \leq (\epsilon/k)^2$*, is* $O(\epsilon\log(1/p)/p^2)$*-close to a Boolean function dominated by the coordinates in* $\mathcal{J}_{k,\tau}(f)$*, where*

$$\tau \doteq {\delta_p}^{4k}\epsilon^2/k^2 = \Theta(\frac{p^k\epsilon^2}{k^2})$$

Hence such a function f *is an* $[O(\epsilon\log(1/p)/p^2), J]$*-junta for* $J = O(\epsilon^{-2}k^3p^{-k})$*.*

Note that the O-notation here, and throughout this paper, only hides constants which are independent of ϵ, k, n, and p.

Proof. Denote $\mathcal{J} \doteq \mathcal{J}_{k,\tau}(f)$, and $\bar{\mathcal{J}} \doteq [n] \setminus \mathcal{J}$. To prove Theorem 5, it is enough to show that the variation of f on $\bar{\mathcal{J}}$ is dominated by $\epsilon\log(1/p)/p^2$. To show this, we take a random partition of $\bar{\mathcal{J}}$ into r subsets $I_1, \ldots, I_r$, where $r \doteq k^2/\epsilon$. We show that the expectation of the variation of f on each of these subsets is very small, and then use a probabilistic argument to deduce that the variation of f on their union is small as well.

The following lemma, which contains the main arguments in the proof of Theorem 5, is proven in the next subsection.

Lemma 3.1. *There exists a global constant* C*, such that*

$$\mathop{\mathbb{E}}_{I \sim \mu_{1/r}^{\bar{\mathcal{J}}}} [\mathsf{Vr}_{\mathsf{f}}(I)] \leq \frac{C\log(1/p)}{p^2}\left({\delta_p}^{-4k}\tau + k^2/r^2 + \|\mathsf{f}^{>k}\|_2^2\right)$$

Note that the distribution of each subset in the partition is $\mu_p^{\bar{\mathcal{J}}}$. Hence Lemma 3.1 implies, using the linearity of expectation and the sub-additivity of the variation, that the variation of f on $\bar{\mathcal{J}}$ is small, namely

$$\mathsf{Vr}_{\mathsf{f}}(\bar{\mathcal{J}}) \leq \mathbb{E}\left[\sum_{h=1}^{r} \mathsf{Vr}_{\mathsf{f}}(I_h)\right] \leq \frac{Cr\log(1/p)}{p^2}\left({\delta_p}^{-4k}\tau + k^2/r^2 + \|\mathsf{f}^{>k}\|_2^2\right)$$

From Proposition 2.1 we thus obtain that f is $\frac{Cr\log(1/p)}{2p^2}\left({\delta_p}^{-4k}\tau + k^2/r^2 + \|\mathsf{f}^{>k}\|_2^2\right)$-close to a Boolean function that depends only on the coordinates of $\mathcal{J}$. This completes the proof of Theorem 5, since

$$\frac{Cr\log(1/p)}{2p^2}\left({\delta_p}^{-4k}\tau + k^2/r^2 + \|\mathsf{f}^{>k}\|_2^2\right) \leq \frac{Cr\log(1/p)}{2p^2}\left({\delta_p}^{-4k}\tau + k^2/r^2 + \frac{\epsilon^2}{k^2}\right) = O\left(\frac{\epsilon\log(1/P)}{p^2}\right)$$

∎

3.1 The Variation of f on Random Subsets of $\bar{\mathcal{J}}$ is Small

We now give the proof of Lemma 3.1, which proceeds by showing that on a random I chosen according to $\mu_{1/r}^{\bar{\mathcal{J}}}$, the following holds with high probability: for almost all input settings x outside I the weight of the restriction $\mathsf{f}_I[x]$ on Walsh-products of size higher than 1 is very small. Once this is shown, we apply a corollary of Theorem 8, which is proven later, stating that in this case $\mathsf{f}_I[x]$ must be close to either a dictatorship or a constant. We then show that $\mathsf{f}_I[x]$ cannot be close to a non-constant dictatorship too often, hence it is usually almost constant, and the lemma follows.

3.2 When $k = 1$

We start by considering the case where a Boolean function f has low weight on Walsh-products of size higher than one. In this case f must be close to a dictatorship, as was proven in [FKN01] for the uniform case, and as we prove for the biased case in Section 6. The result for the case $k = 1$ does not extend directly to the case of higher frequencies, since it uses the fact that the values of Walsh-products of size one on a random assignment are independent.

The following proposition follows from Corollary 6.1, which is proven in Section 6. For the convenience of the reader, we cite Corollary 6.1 here.

Corollary 6.1. *Let $\mathsf{f}\colon \mathcal{P}([n]) \to \{-1,1\}$ be a Boolean function, and let $\epsilon \doteq \|\mathsf{f}^{>1}\|_2^2$. Assume that $\epsilon \le \frac{p^2}{20(\log(\frac{1}{p^2})+5)}$. Then f is $\left(1 + \frac{30}{p^2}\exp(-\frac{p^2}{20\epsilon})\right)\epsilon$-close to some Boolean dictatorship.*

Proposition 3.2. *There exists a global constant M so that given any Boolean function $\mathsf{f}\colon \mathcal{P}(m) \to \{-1,1\}$, either there exists a coordinate i such that $|\widehat{\mathsf{f}}(\{i\})| > \sqrt{p}$, or*

$$\mathcal{V}(\mathsf{f}) = \mathsf{Vr}_{\mathsf{f}}([m]) = \|\mathsf{f}^{>0}\|_2^2 \le \frac{M\log(1/p)}{p^2} \cdot \|\mathsf{f}^{>1}\|_2^2$$

Proof. Setting $\epsilon = \|\mathsf{f}^{>1}\|_2^2$, one notes that there is nothing to prove in the case $\epsilon > \frac{p^2}{20(\log(\frac{1}{p^2})+5)}$, since taking M to be a large-enough constant, the right-hand side of the inequality is larger than 1.

If $\epsilon \le \frac{p^2}{20(\log(\frac{1}{p^2})+5)}$, Corollary 6.1 is applicable, and we have that f is 2ϵ-close either to a constant dictatorship, or to a non-constant Boolean dictatorship g. An easy calculation shows that if g is a non-constant Boolean dictatorship dominated by the i'th coordinate then $|\widehat{\mathsf{g}}(\{i\})| = 2\sqrt{p(1-p)} \ge \sqrt{2p}$. Hence in the second case there must be a coordinate i for which $|\widehat{\mathsf{f}}(\{i\})| > \sqrt{p}$. In the first case, it is easy to verify that the second inequality holds in Proposition 3.2. ∎

Proposition 3.2 implies that there are two kinds of Boolean functions f. A Boolean function is either of the 'dictatorship-type', namely for some i it has a very large coefficient of the form $\widehat{\mathsf{f}}(\{i\})$; or it is of the 'bounded-variance' type, namely its variance is bounded by its weight on frequencies higher frequencies.

3.3 Few Non-Constant Dictatorships

The proof of Lemma 3.1 continues as follows. Let I be a random set of coordinates, as specified in Lemma 3.1, and consider the restriction $\mathsf{f}_I[x]$ for a random $x \in \mathcal{P}(\bar{I})$, where $\bar{I} \doteq [n] \setminus I$. We first show that there may only be a few values of x for which $\mathsf{f}_I[x]$ is of the 'dictatorship-type'. Then we show by a simple combinatorial argument that for a random x, the expected weight of $\mathsf{f}_I[x]$ on frequencies above 1 is small. It therefore follows from Proposition 3.2 that the expected variance of $\mathsf{f}_I[x]$ (which equals $\mathsf{Vr}_\mathsf{f}(I)$) is small as well.

Dictatorship-type restrictions. For a given $I \subseteq [n]$ denote the 'dictatorship set' by

$$\mathcal{D}_I \doteq \left\{ x \in \mathcal{P}(\bar{I}) \,\middle|\, \exists\, i \in I \text{ for which } \left|\widehat{\mathsf{f}_I[x]}(\{i\})\right| > \sqrt{p} \right\}$$

To bound the measure of $\mathcal{D}_I$, we use the fact that the coefficient of χ_i in $\mathsf{f}_I[x]$ is a function of x that is concentrated on low-frequencies, and has small norm (since every $i \in I$ has small variation). We thus use the following lemma, which utilizes Theorem 4 to show that such a function cannot often attain large values. This implies that the coefficient of χ_i is almost never high enough for it to become a dictator.

Lemma 3.3. *Let $0 < \alpha < \beta$ be any parameters. Then for any function $\mathsf{g}\colon \mathcal{P}([m]) \to \mathbb{R}$*

$$\Pr_{x\sim\mu_p^m}\left[|\mathsf{g}(x)| > \beta\right] \le \alpha^{-4}{\delta_p}^{-4k}\|\mathsf{g}^{\le k}\|_2^4 + (\beta-\alpha)^{-2}\|\mathsf{g}^{>k}\|_2^2$$

Proof. We start with a simpler claim, dealing only with the low frequencies portion of the function considered.

Claim 3.4. *Let $\mathsf{g}\colon \mathcal{P}([m]) \to \mathbb{R}$ be a real-valued function such that $\mathsf{g}^{>k} = 0$, and let $\alpha < 1$ be a positive parameter; then*

$$\Pr_{x\sim\mu_p^m}\left[|\mathsf{g}(x)| > \alpha\right] \le \alpha^{-4}{\delta_p}^{-4k}\|\mathsf{g}\|_2^4$$

Proof. By applying Markov's inequality for $|\mathsf{g}|^4$ and then applying Theorem 4,
we have $\alpha^4 \cdot \Pr_{x\sim\mu_p^m}\left[|\mathsf{g}(x)| > \alpha\right] \le \|\mathsf{g}\|_4^4 \le {\delta_p}^{-4k}\|\mathsf{g}\|_2^4$. ∎

Now to obtain Lemma 3.3, we break g into its low-frequency and its high-frequency parts:

$$\Pr_{x\sim\mu_p^m}\left[|\mathsf{g}(x)| > \beta\right] \le \Pr_{x\sim\mu_p^m}\left[\left|\mathsf{g}^{\le k}(x)\right| > \alpha\right] + \Pr_{x\sim\mu_p^m}\left[\left|\mathsf{g}^{>k}(x)\right| > \beta-\alpha\right] \le$$

$$\le \alpha^{-4}{\delta_p}^{-4k}\|\mathsf{g}^{\le k}\|_2^4 + \Pr_{x\sim\mu_p^m}\left[(\mathsf{g}^{>k}(x))^2 > (\beta-\alpha)^2\right] \le \alpha^{-4}{\delta_p}^{-4k}\|\mathsf{g}^{\le k}\|_2^4 + (\beta-\alpha)^{-2}\|\mathsf{g}^{>k}\|_2^2$$

∎

Now fix $i \in I$ and consider the function $g_i\colon \mathcal{P}(\widehat{I}) \to \mathbb{R}$, which assigns to every x the coefficient of χ_i in f_I. That is,

$$g_i(x) = \widehat{\mathsf{f}_I[x]}(\{i\})$$

For $\mathsf{f}_I[x]$ to be a dictatorship, one of the g_i's must evaluate to at least $\sqrt{p}$ in absolute value. Applying lemma 3.3, with parameters $\alpha = \sqrt{p}/2$ and $\beta = \sqrt{p}$, we get a bound on the probability, for a random x, that $\mathsf{f}_I[x]$ is a dictatorship.

$$
\begin{aligned}
\Pr_{x\sim\mu_p^{\bar{I}}}[x\in\mathcal{D}_I] &\leq \sum_{i\in I}\Pr_{x\sim\mu_p^{\bar{I}}}[|g_i(x)|>\sqrt{p}]\leq \\
&= 16p^{-2}\delta_p{}^{-4k}\sum_{i\in I}\|\mathsf{g}_i^{\leq k}\|_2^4+\frac{4}{p}\sum_{i\in I}\|\mathsf{g}^{>k}\|_2^2= \\
&= 16p^{-2}\delta_p{}^{-4k}\sum_{i\in I}\Big\|\sum_{\substack{|S|\leq k\\ S\cap I=\{i\}}}\widehat{\mathsf{f}}(S)\chi_S\Big\|_2^4 + \frac{4}{p}\sum_{i\in I}\Big\|\sum_{\substack{|S|>k\\ S\cap I=\{i\}}}\widehat{\mathsf{f}}(S)\chi_S\Big\|_2^2\leq \\
&\leq 16p^{-2}\delta_p{}^{-4k}\sum_{i\in I}\Big(\sum_{\substack{|S|\leq k\\ S\cap I=\{i\}}}\widehat{\mathsf{f}}^2(S)\Big)^2+\frac{4}{p}\|\mathsf{f}^{>k}\|_2^2
\end{aligned}
$$

Since $\sum_{|S\cap I|=1}\widehat{\mathsf{f}}^2(S)\leq 1$, it follows that

$$
\sum_{i\in I}\Big(\sum_{\substack{|S|\leq k\\ S\cap I=\{i\}}}\widehat{\mathsf{f}}^2(S)\Big)^2\leq\max_{i\in I}\sum_{\substack{|S|\leq k\\ S\cap I=\{i\}}}\widehat{\mathsf{f}}^2(S)=\max_{i\in I}\mathsf{Vr}_{\mathsf{f}}^{\leq k}(i)<\tau
$$

Altogether this implies that for some constant M_1,

$$
\Pr_{x\sim\mu_p^{\bar{I}}}[x\in\mathcal{D}_I]\leq M_1p^{-2}\delta_p{}^{-4k}\tau+M_1p^{-1}\|\mathsf{f}^{>k}\|_2^2
$$

3.4 Restrictions are Expectedly of Small Variation

We are now ready to prove that the variation of f on I is, with high probability, quite small. First, note that for an x such that $x\notin\mathcal{D}_I$, Proposition 3.2 asserts that

$$
\mathsf{Vr}_{\mathsf{f}_I[x]}(I)\leq\frac{M\log(1/p)}{p^2}\sum_{|R|>1}\widehat{\mathsf{f}_I[x]}^2(R)
$$

and by Claim 2.2 we have that

$$
\begin{aligned}
\mathop{\mathbb{E}}_{I\sim\mu_{1/r}^{\mathcal{J}}}[\mathsf{Vr}_{\mathsf{f}}(I)]=\mathop{\mathbb{E}}_{\substack{I\sim\mu_{1/r}^{\mathcal{J}}\\ x\sim\mu_p^{\bar{I}}}}\left[\mathsf{Vr}_{\mathsf{f}_I[x]}(I)\right] &\leq \Pr_{\substack{I\sim\mu_{1/r}^{\mathcal{J}}\\ x\sim\mu_p^{\bar{I}}}}[x\in\mathcal{D}_I]+\mathop{\mathbb{E}}_{\substack{I\sim\mu_{1/r}^{\mathcal{J}}\\ x\sim\mu_p^{\bar{I}}}}\left[\frac{M\log(1/p)}{p^2}\sum_{|R|>1}\widehat{\mathsf{f}_I[x]}^2(R)\right]\leq \\
&\leq M_1p^{-2}\delta_p{}^{-4k}\tau+M_1p^{-1}\|\mathsf{f}^{>k}\|_2^2+\frac{M\log(1/p)}{p^2}\mathop{\mathbb{E}}_{I\sim\mu_{1/r}^{\mathcal{J}}}\left[\sum_{|S\cap I|>1}\widehat{\mathsf{f}}^2(S)\right]\leq \\
&\leq M_1p^{-2}\delta_p{}^{-4k}\tau+M_1p^{-1}\|\mathsf{f}^{>k}\|_2^2+\frac{M\log(1/p)}{p^2}\left(\|\mathsf{f}^{>k}\|_2^2+\mathop{\mathbb{E}}_{I\sim\mu_{1/r}^{\mathcal{J}}}\left[\sum_{\substack{|S|\leq k\\ |S\cap I|>1}}\widehat{\mathsf{f}}^2(S)\right]\right)
\end{aligned}
$$

Now note that $\displaystyle\mathop{\mathbb{E}}_{I\sim\mu_{1/r}^{\mathcal{J}}}\Big[\sum_{\substack{|S|\leq k\\ |S\cap I|>1}}\widehat{\mathsf{f}}^2(S)\Big]\leq\frac{k^2/r^2}{1-k/r}\leq\frac{2k^2}{r^2}$:

This follows since the total weight of all Walsh-products is bounded by 1, and since for a single Walsh-product supported by S,

$$\Pr_I[|S \cap I| > 1] \le \sum_{i=2}^{k} \binom{k}{i} r^{-i}(1-1/r)^{k-i} \le \sum_{i=2}^{k} k^i r^{-i} \le \frac{k^2/r^2}{1-k/r}$$

Therefore, we get that overall, the expectation of the variation is bounded by

$$\mathop{\mathbb{E}}_{I\sim\mu_{1/r}^{\bar{\mathcal{J}}}}[\mathsf{Vr}_{\mathsf{f}}(I)] \le M_1 p^{-2}\delta_p{}^{-4k}\tau + M_1 p^{-1}\|\mathsf{f}^{>k}\|_2^2 + \frac{M\log(1/p)}{p^2}\left(\frac{2k^2}{r^2} + \|\mathsf{f}^{>k}\|_2^2\right)$$

This completes the proof of Lemma 3.1.

4 Discussion of Theorem 5

4.1 A Corollary Concerning Noise-Sensitivity

Let us next translate Theorem 5 from the language of Fourier coefficient to that of noise-sensitivity.

Definition 4. *The λ-noise-sensitivity of a Boolean function $\mathsf{f}\colon [n] \to \{-1,1\}$ (with respect to μ_p) is defined by*

$$\mathrm{NS}_{\lambda,p}(\mathsf{f}) \doteq \Pr_{x\sim\mu_p^n,\ I\sim\mu_\lambda^n,\ z\sim\mu_p^I}\left[\mathsf{f}(x) \ne \mathsf{f}\Big((x\setminus I)\cup z\Big)\right]$$

The noise-sensitivity can be also formulated in terms of the Fourier-expansion of f as follows.

Proposition 4.1. *Let $\mathsf{f}\colon \mathcal{P}([n]) \to \{-1,1\}$ be a Boolean function. Then for every parameter λ,*

$$\mathrm{NS}_{\lambda,p}(\mathsf{f}) = \frac{1}{2} - \frac{1}{2}\sum_S (1-\lambda)^{|S|}\widehat{\mathsf{f}}(S)^2$$

Proof. If X and Y are two random variables obtaining values in $\{-1,1\}$, then

$$1 - 2\Pr[X \ne Y] = \mathbb{E}[XY]$$

So

$$\begin{aligned}
1 - 2\mathrm{NS}_{\lambda,p}(\mathsf{f}) &= \mathop{\mathbb{E}}_{x\sim\mu_p^n,\ I\sim\mu_\lambda^n,\ z\sim\mu_p^I}\left[\mathsf{f}(x)\mathsf{f}\Big((x\setminus I)\cup z\Big)\right] \\
&= \sum_{S,T\subseteq[n]} \widehat{\mathsf{f}}(S)\widehat{f}(T)\mathbb{E}_{x,\ z,\ I}[\chi_S(x)\chi_T((x\setminus I)\cup z)] \\
&= \sum_{S,T\subseteq[n]} \widehat{\mathsf{f}}(S)\widehat{f}(T)\mathbb{E}_I\Big[\mathbb{E}_{x,\ z}\Big[\chi_S(x)\chi_{T\setminus I}(x)\chi_{T\cap I}(z)\Big]\Big] \qquad (1)
\end{aligned}$$

In each term of (1) the inner expectation is zero unless $T \cap I = \emptyset$, since the expectation of $\chi_{T\cap I}(z)$ is zero. In case $T \cap I = \emptyset$ the expectation is still zero unless $S = T$, since it is the inner-product of two biased Walsh-products, and if indeed $S = T$, the expectation equals 1. Therefore we have

$$1 - 2\mathrm{NS}_{\lambda,p}(\mathsf{f}) = \sum_{S\subseteq[n]} \widehat{\mathsf{f}}^2(S)\Pr_I[S\cap I = \emptyset] = \sum_{S\subseteq[n]}(1-\lambda)^{|S|}\widehat{\mathsf{f}}^2(S)$$

which implies the desired identity. ∎

Using Proposition 4.1 and Theorem 5, we obtain the Corollary 1.1. Let us cite it here for convenience, and then prove it.

Corollary 1.1. *For any parameter $\lambda > 0$, fix $k = \log_{(1-\lambda)}(1/2)$. Then every Boolean function $\mathsf{f}\colon \mathcal{P}([n]) \to \{-1,1\}$ whose λ-noise-sensitivity with respect to μ_p^n is bounded by $(\epsilon/k)^2$, is an $[O(\epsilon \log(1/p)/p^2), J]$-junta, where*

$$J = O\left(\frac{k^3}{\epsilon^2 p^k}\right)$$

Proof. Let f be a Boolean function as stated in Corollary 1.1. Then $\sum_S \widehat{\mathsf{f}}(S)^2 = \|\mathsf{f}\|_2^2 = 1$, and hence using by Proposition 4.1 we have

$$\begin{aligned}\mathrm{NS}_{\lambda,p}(\mathsf{f}) = (\epsilon/k)^2 &\geq \frac{1}{2} - \frac{1}{2}\sum_S (1-\lambda)^{|S|}\widehat{\mathsf{f}}(S)^2 \geq \frac{1}{2} - \frac{1}{2}\left(\sum_{|S|\leq k}\widehat{\mathsf{f}}(S)^2 + \frac{1}{2}\sum_{|S|>k}\widehat{\mathsf{f}}(S)^2\right)\\ &= \frac{1}{2} - \frac{1}{2}\left(1 - \frac{1}{2}\sum_{|S|>k}\widehat{\mathsf{f}}(S)^2\right) = \frac{1}{4}\sum_{|S|>k}\widehat{\mathsf{f}}(S)^2\end{aligned}$$

we thus obtain

$$\sum_{|S|>k}\widehat{\mathsf{f}}(S)^2 \leq 4(\epsilon/k)^2$$

Corollary 1.1 now follows from Theorem 5. ■

4.2 Tightness

It is worth noting that Bourgain's result cannot hold "out of the box" for the case of general p. To demonstrate this we consider the graph-property of having an induced triangle. Let f be a Boolean function with one variable x_e for each edge e of the undirected complete graph on m vertices (f has $n \doteq \binom{m}{2}$ variables). One can view an assignment to the variables of f as a graph on m vertices. Define the value of f for such an assignment to be 1 if the graph contains an induced triangle and -1 otherwise.

To describe the properties of f, we need the notion of average sensitivity.

Definition 5 (average sensitivity). *Let $\mathsf{f} : \mathcal{P}([n]) \to \{-1,1\}$ be a Boolean function. For every input x for f let*

$$\gamma(x) \doteq \sharp\,\{i \mid \mathsf{f}(x \setminus \{i\}) \neq \mathsf{f}(x \cup \{i\})\}$$

The average sensitivity of f with respect to μ_p is defined by

$$\mathrm{as}_p(\mathsf{f}) \doteq \mathop{\mathbb{E}}_{x\sim\mu_p}[\gamma(x)]$$

Here we will be interested in the following property of the average sensitivity.

Proposition 4.2. *Let $\mathsf{f} : \mathcal{P}([n]) \to \{-1,1\}$ be a Boolean function. Then for every $p < \frac{1}{2}$ and k,*

$$\|\mathsf{f}^{>k}\|_2^2 < \frac{p \cdot \mathrm{as}_p(\mathsf{f})}{k}$$

where both $\mathsf{f}^{>k}$ and the norm are taken with respect to μ_p.

Proof. We omit the simple proof. ■

Let us return to the function f defined above. It is known that there exists a parameter $p = p_m$, $p \approx \frac{1}{m}$, for which

$$\frac{1}{4} \leq \Pr_{x \sim \mu_p}[\mathsf{f}(x) = 1] \leq \frac{3}{4} \tag{*}$$

and

$$p \cdot \mathrm{as}_p(\mathsf{f}) < C$$

where C is a constant which is independent of n. It hence follows from Proposition 4.2 that

$$\|\mathsf{f}^{>k}\|_2^2 < \frac{C}{k} \tag{**}$$

Now suppose that Theorem 1 holds with respect to μ_p. By applying it to f with parameters $\epsilon = \eta = 1/10$, one easily obtains from (**) that f is ϵ-close to some constant-sized junta. This means that one can predict whether a graph has an induced triangle, up to probability 1/10, by only looking at a constant number of edges. However since $p \approx \frac{1}{m}$, it follows that with probability $1 - o(1)$ the inspected edges are all absent from the graph, and therefore provide no prediction as to whether it contains a triangle.

5 A Switching Lemma

Switching lemmas are a tool of great importance in the study of Boolean functions. A typical switching lemma shows that a random restriction of a Boolean function in a given class, is with high probability a very simple function (e.g. depends on a constant number of variables). As shown below Theorem 1 yields a switching lemma, and our results extend it to the biased case.

Let us begin with the switching lemma obtained from Theorem 1.

Theorem 6. *Let $\epsilon > 0$ and $K > 0$ be some parameters, and let $\mathsf{f} : \mathcal{P}([n]) \to \{-1, 1\}$ be a Boolean function that can be described by a decision-tree with l nodes. Let $I \subseteq [n]$ be a random subset, containing each coordinate with probability $K/\log(l/\epsilon)$, and let $x \in \mathcal{P}([n] \setminus I)$ be chosen according to $\mu_{1/2}^{[n]\setminus I}$.*

Then with probability $(1-\epsilon)$, $\mathsf{f}_I[x]$ is an (ϵ, J)-junta, where

$$J \leq 200^{K \cdot (c/\epsilon)^{1/eK}}$$

and c is some global constant.

Before we prove Theorem 6, let us examine how it compares with the following (well known) switching lemma of Hastad [Has86].

Theorem [Has86]. *Let $\mathsf{f} : \mathcal{P}([n]) \to \{-1, 1\}$ be Boolean function that can be written as a Boolean OR over AND's of at most m literals (that is, f is an m-DNF). Let $I \subseteq [n]$ be a random set, $I \sim \mu_\rho^n$, and let $x \in \mathcal{P}([n] \setminus I)$ be chosen according to $\mu_{1/2}^{[n]\setminus I}$. Then with probability at least $(1 - (7\rho m)^s)$, $\mathsf{f}_I[x]$ can be described by a decision tree of height at most s.*

Note that in Theorem 5, if the number of AND's is polynomial in the number of inputs, then f has a decision tree where the number of nodes is polynomial in n. Therefore Theorem 6 holds for such functions, and even for many other functions. Its main drawback, however, is that it only yields an approximation for the random restriction, and not a full description of it.

Proof of Theorem 6. We only give the highlights of the proof (more details are given in the similar proof, of Theorem 7 below). The first step is to approximate f by a Boolean function of small degree. This is accomplished by the following well-known claim.

Claim 5.1. *Let* f *be a Boolean function that is computed by a decision-tree with* l *nodes. Then there exists a Boolean function* g *of degree* $m \doteq \log(l/\epsilon)$*, such that* $\|\mathsf{f} - \mathsf{g}\|_2^2 \leq \epsilon$*.*

Let g be as in the above claim. Let $c = c_{3/4,\epsilon}$ be the constant that corresponds to taking $\eta = 3/4$ in Theorem 1, and set $t \doteq (c/\epsilon)^{1/eK}$ and $k \doteq (t+1)K$. Denoting $\mathsf{h} \doteq \mathsf{g}_I[x]$, one calculates that $\mathbb{E}_{x,I}\left[\|\mathsf{h}^{>k}\|_2^2\right] < c\epsilon k^{-3/4}$, and hence $\Pr_{x,I}\left[\|\mathsf{h}^{>k}\|_2^2 > ck^{-3/4}\right] < \epsilon$. Applying Theorem 1, we thus have that with probability at least $(1-\epsilon)$, h is an $(\epsilon, k10^k)$-junta. This gives Theorem 6 immediately. ■

We now state the switching lemma that is obtained for the biased case from our results.

Theorem 7. *Let* $\epsilon > 0$ *and* $k > 0$ *be some parameters, and let* $\mathsf{f} : \mathcal{P}([n]) \to \{-1, 1\}$ *be a Boolean function that can be described by a decision-tree with* l *nodes. Let* $I \subseteq [n]$ *be a random subset, containing each coordinate with probability* $\epsilon^{3/k}k/(50e\log_p(l/\epsilon))$*. Let* $x \in \mathcal{P}([n] \setminus I)$ *be chosen according to* $\mu_p^{[n]\setminus I}$*.*

Then with probability $(1-\epsilon)$*,* $\mathsf{f}_I[x]$ *is an* $(O(\epsilon\log(1/p)/p^2), J)$*-junta, where* $J \leq O(\epsilon^{-2}k^3p^{-k})$*.*

Proof. We first use a simple analogue of Claim 5.1 for the biased case, as follows.

Claim 5.2. *Let* f *be a Boolean function that is computed by a decision-tree with* l *nodes. Then there exists a Boolean function* g *of degree* $m \doteq \log_p(l/\epsilon)$ *with respect to* μ_p*, such that* $\|\mathsf{f} - \mathsf{g}\|_2^2 \leq 4\epsilon$*.*

Proof. Define g to be the function obtained by clipping the decision tree for f beyond depth m, putting arbitrary values in the newly formed leaves. When computing f on a random input using its decision tree, one easily notes that the probability of reaching the clipped section is at most ϵ, and thus $\|\mathsf{f} - \mathsf{g}\|_2^2 \leq 4\epsilon$. ■

Let g be a Boolean function of degree m as in the above claim, and let $\mathsf{h} \doteq \mathsf{g}_I[x]$. We need to prove that with probability at least $(1-\epsilon)$, over the choice of I and x, h is an $(O(\epsilon\log(1/p)/p^2), J)$-junta.

$$\mathop{\mathbb{E}}_{x,I}\left[\|\mathsf{h}^{>k}\|_2^2\right] = \mathop{\mathbb{E}}_{I}\left[\sum_{|S\cap I|>k} \widehat{\mathsf{g}}(S)^2\right] = \sum_{S}\left(\widehat{\mathsf{g}}(S)^2 \cdot \Pr_{I}\left[|S \cap I| > k\right]\right)$$

Now let $\alpha \doteq \epsilon^{3/k}/(50e)$. Since g is of degree m, it follows that for every S

$$\begin{aligned}\Pr_i\left[|S\cap I|>k\right] &\leq \sum_{i=k+1}^{m}\binom{m}{i}\left(\frac{\alpha k}{m}\right)^i\left(1-\frac{\alpha k}{m}\right)^i \leq \\ &\leq \sum_{i=k}^{m}\frac{m^i}{i!}\left(\frac{\alpha k}{m}\right)^i \leq \sum_{i=k}^{m}\sqrt{2\pi i}(e\alpha)^i\left(\frac{k}{i}\right)^i = \sum_{i=k}^{m}\sqrt{2\pi i}(e\alpha)^i\left(1-\frac{i-k}{i}\right)^i \leq \\ &\leq \sum_{i=k}^{m}\sqrt{2\pi i}(e\alpha)^i e^{k-i} \leq \sqrt{2\pi k}(e\alpha)^k\sum_{i=k}^{m}(i-k+1)\alpha^i \leq 2\sqrt{2\pi k}(e\alpha)^k = \\ &= \frac{2\sqrt{2\pi k}\cdot\epsilon^3}{50^k} \leq \frac{\epsilon^3}{k^2}\end{aligned}$$

We thus have

$$\mathop{\mathbb{E}}_{x,I}\left[\|\mathsf{h}^{>k}\|_2^2\right] \leq \frac{\epsilon^3}{k^2}$$

and thus with probability at least $(1-\epsilon)$, $\|\mathsf{h}^{>k}\|_2^2 \leq \frac{\epsilon^3}{k^2}$. When this occurs, then according to Theorem 5 h is an $(O(\epsilon\log(1/p)/p^2), J)$-junta. ■

6 Biased FKN

In this section, we prove that a Boolean function f which is close to being linear, namely a function for which $\|\mathsf{f}^{>1}\|_2^2$ is small, is in fact close to being a dictatorship. This is an easy corollary of the following theorem, which contains a slightly different statement. It considers a *real-valued* function f for which $\mathsf{f}^{>1}\equiv 0$, and shows that if f is close to being Boolean, it must also be close to a real-valued dictatorship, namely to a real-valued function which depends on at most one coordinate.

Theorem 8. *Let $\mathsf{f}\colon \mathcal{P}([n]) \to \mathbb{R}$ be a linear real valued function, namely $\mathsf{f}^{>1}=0$. Let $\epsilon \doteq \|\,|\mathsf{f}|-1\,\|_2^2$ measure the squared distance of f from the nearest Boolean function, and suppose that $\epsilon \leq \frac{p^2}{20(\log(\frac{1}{p^2})+5)}$.*

Then, denoting by i_o the index such that $\left|\widehat{\mathsf{f}}(\{i_o\})\right|$ is maximal, we have

$$\|\mathsf{f}-\left(\widehat{\mathsf{f}}(\emptyset)+\widehat{\mathsf{f}}(\{i_o\})\chi_{\{i_o\}}\right)\|_2^2 < \left(1+\frac{30}{p^2}\exp(-\frac{p^2}{20\epsilon})\right)\epsilon$$

Before we prove Theorem 8, we state and prove the following corollary.

Corollary 6.1. *Let $\mathsf{f}\colon \mathcal{P}([n]) \to \{-1,1\}$ be a Boolean function, and let $\epsilon \doteq \|\mathsf{f}^{>1}\|_2^2$. Assume that $\epsilon \leq \frac{p^2}{20(\log(\frac{1}{p^2})+5)}$. Then f is $\left(1+\frac{30}{p^2}\exp(-\frac{p^2}{20\epsilon})\right)\epsilon$-close to some Boolean dictatorship.*

Proof. Note that $\|\,\left|\mathsf{f}^{\leq 1}\right|-1\,\|_2^2 \leq \|\mathsf{f}^{>1}\|_2^2 = \epsilon$. Hence according to Theorem 8, there is some coordinate $i_o \in [n]$ such that

$$\begin{aligned}\mathsf{Vr}_{\mathsf{f}}([n]\setminus\{i_o\}) &= \|\mathsf{f}-\mathsf{Avg}_{[n]\setminus\{i_o\}}[\mathsf{f}]\|_2^2 = \|\mathsf{f}-\left(\widehat{\mathsf{f}}(\emptyset)+\widehat{\mathsf{f}}(\{i_o\})\chi_{\{i_o\}}\right)\|_2^2 \leq \\ &\leq \epsilon + \|\mathsf{f}^{\leq 1}-\left(\widehat{\mathsf{f}}(\emptyset)+\widehat{\mathsf{f}}(\{i_o\})\chi_{\{i_o\}}\right)\|_2^2 < 2\left(1+\frac{30}{p^2}\exp(-\frac{p^2}{20\epsilon})\right)\epsilon\end{aligned}$$

It follows from Proposition 2.1 that there exists a Boolean function g that depends only on the coordinate i_o (and is thus a dictatorship), such that

$$\Pr_{x\sim\mu_p^n}[\mathsf{f}(x)\neq\mathsf{g}(x)] < \left(1+\frac{30}{p^2}\exp(-\frac{p^2}{20\epsilon})\right)\epsilon$$

Therefore f is a $\left(\left(1+\frac{30}{p^2}\exp(-\frac{p^2}{20\epsilon})\right)\epsilon,\, 1\right)$-junta. ∎

Proof of Theorem 8: For simplicity, we write $\mathsf{f} = a_0 + \sum_{i=1}^n a_i\chi_{\{i\}}$, and assume without loss of generality that $|a_1| \geq |a_2| \geq \ldots \geq |a_n|$. Also, define $q = 1-p$.

Our goal is to prove that $\sum_{i=2}^n |a_i|^2 < (1+O(\frac{1}{p^2})e^{O(-p^2/\epsilon)})\epsilon$. As a first step, we show that none of the terms $a_2,\ldots,a_n$ can be large.

Claim 6.2. $|a_2| \leq \frac{2\sqrt{q\epsilon}}{\sqrt{p}}$.

Proof. Recall first that each character $\chi_{\{i\}}$ attains two values, whose difference equals $\sqrt{p/q}+\sqrt{q/p}$.

We prove the claim by contradiction. For each given setting of the values of $x_3,\ldots,x_n$, consider the values of f attained by assigning x_1, x_2. Suppose that $|a_2|\left(\sqrt{p/q}+\sqrt{q/p}\right) \geq \frac{3}{2}$. In that case, since $|a_1| \geq |a_2|$, we have that the difference between the maximal value and the minimal value obtained by assigning x_1 and x_2 is at least 3. At least one of these values is therefore within distance at least $\frac{1}{2}$ from the nearest Boolean value. It follows that with probability at least p^2 over the choices of $x\sim\mu_p^n$, we have $|\,|\mathsf{f}(x)|-1| \geq \frac{1}{2}$. Therefore $\||\mathsf{f}(x)|-1\|_2^2 \geq \frac{p^2}{4} > \epsilon$, a contradiction.

Now suppose that

$$\frac{3}{2} \geq |a_2|\left(\sqrt{p/q}+\sqrt{q/p}\right) > \frac{2\sqrt{\epsilon}}{p}$$

In that case, fix any assignment for the variables x_1 and $x_3,\ldots,x_n$, and consider the values of f attained for the two possible assignment of x_2. If one of these values is within distance at most $\frac{\sqrt{\epsilon}}{p}$ from, say, 1, then the other is within distance at least $\frac{\sqrt{\epsilon}}{p}$ from 1, and within distance at least $\frac{3}{2}-\frac{\sqrt{\epsilon}}{p} > \frac{\sqrt{\epsilon}}{p}$ from (-1). It follows that for a random input x, there is at least probability p to have $|\,|\mathsf{f}(x)|-1| > \frac{\sqrt{\epsilon}}{p}$. Therefore in this case $\||\mathsf{f}(x)|-1\|_2^2 > \epsilon$, again a contradiction.

The only option not leading to contradiction is therefore that $|a_2|\left(\sqrt{p/q}+\sqrt{q/p}\right) \leq \frac{2\sqrt{\epsilon}}{p}$, in which case one easily obtains that $|a_2| \leq \frac{2\sqrt{\epsilon q}}{\sqrt{p}}$. ∎

According to Claim 6.2, for every $2\leq i\leq n,\ |a_i|^2 \leq 4q\epsilon/p$. We thus choose $m\in\{2,\ldots,n\}$ to be the minimal index satisfying

$$\sum_{i=m}^n |a_i|^2 \leq (\frac{4q}{p}+2)\epsilon \tag{2}$$

Denote $I\doteq\{m,\ldots,n\}$. Then

$$\epsilon \geq \|\,|\mathsf{f}|-1\,\|_2^2 = \mathop{\mathbb{E}}_{x\sim\mu_p^n}\left[(|\mathsf{f}(x)|-1)^2\right] = \mathop{\mathbb{E}}_{y\sim\mu_p^{\bar{I}}}\left[\mathop{\mathbb{E}}_{z\sim\mu_p^I}\left[(|\mathsf{f}(y\cup z)|-1)^2\right]\right] =$$
$$= \mathop{\mathbb{E}}_{y\sim\mu_p^{\bar{I}}}\left[\|\,|\mathsf{f}_I[y]|-1\,\|_2^2\right]$$

hence for some $y \in \mathcal{P}(\bar{I})$, $\| \, |\mathsf{f}_I[y]| - 1 \, \|_2^2 \leq \epsilon$. Now $\mathsf{f}_I[y]$ has the form

$$\mathsf{f}_I[y] = b + \sum_{i=m}^{n} a_i \chi_i$$

for some b, and therefore it satisfies the conditions of Theorem 8, with the additional property that $\|\mathsf{f}_I^{>0}[y]\|_2^2 \leq (\frac{4q}{p} + 2)\epsilon$. We use the following lemma, which deals with such a situation.

Lemma 6.3. *Let* $\mathsf{f} \colon \mathcal{P}([n]) \to \mathbb{R}$ *be a function satisfying* $\mathsf{f}^{>1} \equiv 0$. *Let* $\epsilon \doteq \| |\mathsf{f}| - 1 \|_2^2$, *and suppose that* $\|\mathsf{f}^{>0}\|_2^2 < (\frac{4q}{p} + 2)\epsilon$ *and that* $\epsilon \leq p^2/30$. *Then it also holds that*

$$\|\mathsf{f}^{>0}\|_2^2 < \left(1 + \frac{30}{p^2} \exp\left(-\frac{p^2}{20\epsilon}\right)\right)\epsilon$$

Before proving Lemma 6.3, let us show how it concludes the proof of Theorem 8. Lemma 6.3 implies that

$$\sum_{i=m}^{n} |a_i|^2 = \|\mathsf{f}_I^{>0}[y]\|_2^2 < \left(1 + \frac{30}{p^2} \exp(-\frac{p^2}{20\epsilon})\right)\epsilon$$

If $m = 2$, this is what we wanted to show. If $m > 2$, noting that the bound above is smaller than 2ϵ, we obtain from Claim 6.2 that m is not the minimal index satisfying (2), a contradiction. ■

Proof of Lemma 6.3

We now return to the proof of Lemma 6.3. For convenience, we write $\mathsf{f} = b + \sum_{i=1}^{n} a_i \chi_{\{i\}}$.

Proof Overview. The norm $\|\mathsf{f}^{>0}\|_2^2$ is in fact the variance (not variation) of f. Now, since the variance of $|\mathsf{f}|$ is bounded by $\| \, |\mathsf{f}| - 1 \, \|_2^2$ (this expression is minimized by replacing 1 with the expectation of $|\mathsf{f}|$), we have $\mathcal{V}(|\mathsf{f}|) < \epsilon$. Lemma 6.3 follows by showing that $\mathcal{V}(\mathsf{f})$ is essentially bounded by $\mathcal{V}(|\mathsf{f}|)$.

First, we show that the expectation of f, b, is well separated from zero. This holds since $|\mathsf{f}|$ is ϵ-close to 1 on the one hand, and $(\frac{4q}{p} + 2)\epsilon$-close to $|b|$ on the other hand. ¿From the above it follows that $\mathrm{sign}(\mathsf{f}(x)) = \mathrm{sign}(b)$ for almost all inputs x, since the weight of the non-constant part of f is rather small, and cannot move the value of f over to the other side of zero very often. This implies that $|\mathbb{E}\mathsf{f}| \approx \mathbb{E}|\mathsf{f}|$ and hence that $\mathcal{V}(\mathsf{f}) \approx \mathcal{V}(|\mathsf{f}|)$.

We have

$$\| \, |b| - 1 \, \|_2 \leq \| \, |\mathsf{f}| - |b| \, \|_2 + \| \, |\mathsf{f}| - 1 \, \|_2 \leq \|\mathsf{f} - b\|_2 + \sqrt{\epsilon} \leq \left(1 + \sqrt{\frac{4q}{p} + 2}\,\right)\sqrt{\epsilon}$$

and hence

$$|b| \geq 1 - \left(1 + \sqrt{\frac{4q}{p} + 2}\,\right)\sqrt{\epsilon} \geq 1 - \left(1 + \sqrt{\frac{4q + 2p}{p}}\,\right)\sqrt{\epsilon} \geq 1 - \left(1 + \frac{2}{\sqrt{p}}\right)\sqrt{\epsilon} \geq \frac{1}{2}$$

We assume without loss of generality that b is positive. Writing $|\mathsf{f}| - \mathsf{f} = 2|\mathsf{f}|\mathbf{1}_{\{\mathsf{f}<0\}}$, we have

$$\mathbb{E}|\mathsf{f}| - \mathbb{E}\mathsf{f} \leq 2\mathbb{E}|\mathsf{f}|\mathbf{1}_{\{\mathsf{f}<0\}} \tag{3}$$

To show that the expectations on the left-hand side are approximately equal, we bound the term on the right-hand side using the following special case of Azuma's inequality (see [Sch99] for a proof).

Theorem 9 (Azuma's inequality). *Let $X = \sum_{i=1}^{n} X_i$ be a sum of independent random variables with zero expectation, such that the absolute value of each x_i is bounded by d_i. Then*

$$\Pr\left[|X| > t\right] \leq 2 \exp\left(\frac{-t^2}{\sum_{i=1}^{n} d_i^2}\right)$$

The absolute value of a Rademacher function $\chi_{\{i\}}$ is bounded by some constant $\sqrt{q/p} \leq 1/\sqrt{p}$. Denoting $\lambda \doteq \sum_i |a_i|^2$, we have, by applying Azuma's inequality to $\sum_{i=1}^{n} a_i \chi_i$, that

$$\begin{aligned}
\mathbb{E}|\mathsf{f}|\mathbf{1}_{\{\mathsf{f}<0\}} &= \int_{t=0}^{\infty} \Pr\left[\mathsf{f} < -t\right] dt = \int_{t=0}^{\infty} \Pr\left[b + \sum_i a_i \chi_i < -t\right] dt = \\
&= \int_{t=b}^{\infty} \Pr\left[\sum_i a_i \chi_i < -t\right] dt \leq 2 \int_{t=b}^{\infty} \exp\left(\frac{-pt^2}{\lambda}\right) dt \leq \\
&\leq \frac{\lambda}{pb} \int_{t=b}^{\infty} \frac{2pt}{\lambda} \exp\left(\frac{-pt^2}{\lambda}\right) dt \leq \frac{\lambda}{pb} \exp\left(\frac{-pb^2}{\lambda}\right)
\end{aligned}$$

Now since $\lambda < (\frac{4q}{p} + 2)\epsilon$ and $b > \frac{1}{2}$, we have

$$\mathbb{E}|\mathsf{f}|\mathbf{1}_{\{\mathsf{f}<0\}} = \frac{(\frac{4q}{p} + 2)\epsilon}{pb} \exp\left(\frac{-pb^2}{(\frac{4q}{p} + 2)\epsilon}\right) \leq \frac{10\epsilon}{p^2} \exp\left(-\frac{p^2}{20\epsilon}\right)$$

It now follows from (3) that

$$\begin{aligned}
\epsilon > \left\| |\mathsf{f}| - 1 \right\|_2^2 \geq \mathcal{V}(|\mathsf{f}|) = \|\mathsf{f}\|_2^2 - \mathbb{E}|\mathsf{f}|^2 = \mathcal{V}(\mathsf{f}) + \mathbb{E}\mathsf{f}^2 - (\mathbb{E}|\mathsf{f}|)^2 = \\
= \mathcal{V}(\mathsf{f}) + (\mathbb{E}\mathsf{f} + \mathbb{E}|\mathsf{f}|)(\mathbb{E}\mathsf{f} - \mathbb{E}|\mathsf{f}|)
\end{aligned} \tag{4}$$

Now note that

$$\mathbb{E}\mathsf{f} + \mathbb{E}|\mathsf{f}| \leq 2\|\mathsf{f}\|_1 \leq 2\|\mathsf{f}\|_2 \leq 2(\left\||\mathsf{f}| - 1\right\|_2 + 1) \leq 3,$$

hence it follows from (4) that

$$\|\mathsf{f}^{>0}\|_2^2 = \mathcal{V}(\mathsf{f}) \leq \epsilon + \frac{30\epsilon}{p^2} \exp\left(-\frac{p^2}{20\epsilon}\right)$$

which completes the proof.

7 Extending FKN to Higher Frequencies

Following, is an extension of Theorem 8 to the case where f is concentrated on Walsh-products of size at most k rather than 1. It examines the distance between f and a junta, in the case where the weight of f on frequencies higher than k becomes very small – exponentially small in k. For a high-frequency weight in this range, the bound on the distance of f from a junta behaves much better, as a function of the weight, than the bound given in Theorem 5: The squared 2-norm distance from a (real-valued) junta is shown to be at most $1 + o(1)$ times the weight on high frequencies. We do not know whether the small range for which we prove this estimate is a weakness of our proof, or whether this really is the range where the squared 2-norm distance from a junta behaves according to this estimate.

In Section 8 it is shown that the following theorem may be used to also somewhat improve the parameters in Theorem 5 for larger high-frequency weights.

Theorem 10 (high-frequency FKN). *There exists a global constant M for which the following holds. Let $\mathsf{f}\colon \mathcal{P}([n]) \to \mathbb{R}$ be a real valued function of degree k, namely $\mathsf{f}^{>k} \equiv 0$, and take $\tau \doteq \delta_p^{16k}/M$. Let $\epsilon \doteq \| |\mathsf{f}| - 1 \|_2^2$ measure the squared distance of f from the nearest Boolean function, and suppose that $\epsilon < \tau$. Then*

$$\mathsf{Vr}_{\mathsf{f}}([n] \setminus \mathcal{J}_{k,\tau}(\mathsf{f})) \leq \epsilon\big(1 + 1064(\delta_p)^{-4k}(2\epsilon)^{1/4}\big)$$

7.1 Proof of Theorem 10

Set $\mathcal{J} \doteq \mathcal{J}_{\tau,k}$, and $\bar{\mathcal{J}} \doteq [n] \setminus \mathcal{J}$. We therefore need to show that $\mathsf{Vr}_{\mathsf{f}}(\bar{\mathcal{J}}) \leq \epsilon\big(1 + 1064(\delta_p)^{-4k}(2\epsilon)^{1/4}\big)$ Suppose, w.l.o.g., that $\bar{\mathcal{J}}$ is not empty. We consider sets $I \subseteq \bar{\mathcal{J}}$ that satisfy $\mathsf{Vr}_{\mathsf{f}}(I) \leq 3\tau$, and take $I \subseteq \bar{\mathcal{J}}$ to be a maximal set with this property.

Program of Proof. In the proof of Theorem 8 we used the fact that the variation on a set I of coordinates is also the variation on I of any restriction $\mathsf{f}_I[x]$. We could thus fix x and focus only on $\mathsf{f}_I[x]$, as in Lemma 6.3. Here this is not the case, however according to Claim 2.2 the variation on I of f, which is bounded by τ, is the average of the variations on I of restrictions of the form $\mathsf{f}_I[x]$. The proof thus begins by bounding the deviation of the variations on I of restrictions $\mathsf{f}_I[x]$, showing that the contribution of restriction with high variation to this average is very small. For restrictions $\mathsf{f}_I[x]$ where the variation on I is not very high, it is shown that the squared 2-norm distance of $\mathsf{f}_I[x]$ from the nearest Boolean function is essentially bounded below by $\mathsf{Vr}_{\mathsf{f}_I[x]}(I)$.

By averaging over all restrictions, this implies that the distance of f from the nearest Boolean function is essentially bounded below by $\mathsf{Vr}_{\mathsf{f}}(I)$, and therefore $\mathsf{Vr}_{\mathsf{f}}(I) < \epsilon\big(1 + 1064(\delta_p)^{-4k}(2\epsilon)^{1/4}\big)$. This completes the proof, since if $I = \bar{\mathcal{J}}$ we are obviously done, but on the other hand, if $I \neq \bar{\mathcal{J}}$, one can add a coordinate to I, keeping its variation below 3τ, in contradiction to the maximality of I.

Bounding high variations of restrictions

To show that there cannot be too many restrictions $\mathsf{f}_I[x]$ with large variation, we need the following lemma, proven in the next section.

Lemma 7.1. *Let $\mathsf{g}_1, \ldots, \mathsf{g}_m \colon \mathcal{P}([l]) \to \mathbb{R}$ be real-valued functions such that $\mathsf{g}_i^{>k} \equiv 0$ for every i. Then for every $\alpha \geq 0$,*

$$\Pr_{x \sim \mu_p^m}\Big[\sum |\mathsf{g}_i(x)|^2 > \alpha\Big] \leq 256\alpha^{-2}(\delta_p)^{-4k}\left(\sum_{i=1}^{m} \|\mathsf{g}_i\|_2^2\right)^2$$

For shortness, denote $\eta \doteq \mathsf{Vr}_{\mathsf{f}}(I)$ (then $\eta < 3\tau$), and let

$$\mathcal{D} \doteq \big\{x \in \mathcal{P}(\bar{I}) \mid \mathsf{Vr}_{\mathsf{f}_I[x]}(I) > \eta^{3/4}\big\}$$

be the set of restrictions whose variation is much higher than expected.

Proposition 7.2.

$$\mathop{\mathbb{E}}_{x\sim\mu_p^{\bar{I}}}\left[\mathsf{Vr}_{\mathsf{f}_I[x]}(I)\mathbf{1}_{\{x\in\mathcal{D}\}}\right] < 512(\delta_p)^{-4k}\eta^{5/4}$$

Proof. For a non-empty set $S \subseteq I$, define for every $x \in \bar{I}$,

$$g_S(x) \doteq \widehat{\mathsf{f}_I[x]}(S)$$

Then each g_S is a function of degree at most $k-1$, and for every x,

$$\mathsf{Vr}_{\mathsf{f}_I[x]}(I) = \sum_{\substack{S\subseteq I\\ S\neq\emptyset}} g_S^2(x)$$

It follows that

$$\sum_{\substack{S\subseteq I\\ S\neq\emptyset}} \|g_S\|_2^2 = \mathop{\mathbb{E}}_x\Big[\sum_{\substack{S\subseteq I\\ S\neq\emptyset}} g_S^2(x)\Big] = \mathop{\mathbb{E}}_x\left[\mathsf{Vr}_{\mathsf{f}_I[x]}(I)\right] = \mathsf{Vr}_{\mathsf{f}}(I) = \eta$$

Hence

$$\begin{aligned}
\mathbb{E}\mathsf{Vr}_{\mathsf{f}_I[x]}(I)\mathbf{1}_{\{x\in\mathcal{D}\}} =& \\
&= \int_{t=0}^{\infty} \Pr\left[\sum g_S(x)^2 \geq \max(t,\eta^{3/4})\right]dt = \\
&= \int_{t=0}^{\eta^{3/4}} \Pr\left[\sum g_S(x)^2 \geq \eta^{3/4}\right]dt + \\
&\quad + \int_{t=\eta^{3/4}}^{\infty} \Pr\left[\sum g_S(x)^2 \geq t\right]dt \leq
\end{aligned}$$

(using Lemma 7.1)

$$\begin{aligned}
&\leq \eta^{3/4}\cdot 256(\delta_p)^{-4k}\eta^{-3/2}\eta^2 + 256(\delta_p)^{-4k}\eta^2\int_{t=\eta^{3/4}}^{\infty} t^{-2}dt = \\
&= 512(\delta_p)^{-4k}\eta^{5/4}
\end{aligned}$$

■

Bounding $\mathsf{Vr}_{\mathsf{f}_I[x]}(I)$ for $x \notin \mathcal{D}$

Proposition 7.3. *For every $x \notin \mathcal{D}$,*

$$\|\,|\mathsf{f}_I[x]| - 1\,\|_2^2 \geq \mathsf{Vr}_{\mathsf{f}_I[x]}(I) - 20(\delta_p)^{-4k}\eta^{3/2}$$

Proof. Define

$$\mathcal{C} \doteq \left\{x\in\bar{I} \;\middle|\; \left|\,|\widehat{\mathsf{f}_I[x]}(\emptyset)| - 1\right| > \frac{1}{2}\right\}$$

We proof the statement separately for $x \in \mathcal{C}\setminus\mathcal{D}$ and for $x \notin \mathcal{C}\cup\mathcal{D}$.

The case $x \in \mathcal{C} \setminus \mathcal{D}$. It suffices to show that in this case for most $y \in \mathcal{P}(I)$, $\Big| |\mathsf{f}_I[x](y)| - 1 \Big| \geq 1/4$. Note that $\mathsf{f}_I[x] - \widehat{\mathsf{f}_I[x]}(\emptyset)$ is a function of degree at most k, and that since $x \notin \mathcal{D}$,

$$\|\mathsf{f}_I[x] - \widehat{\mathsf{f}_I[x]}(\emptyset)\|_2^2 = \mathsf{Vr}_{\mathsf{f}_I[x]}(I) \leq \eta^{3/4}$$

Hence by Claim 3.4

$$\Pr_{y \sim \mu_p^I}\left[\left|\mathsf{f}_I[x](y) - \widehat{\mathsf{f}_I[x]}(\emptyset)\right| > 1/4\right] < 4^4(\delta_p)^{-4k}\eta^{3/2}$$

It follows that with probability at least $1 - 4^4(\delta_p)^{-4k}\eta^{3/2}$, $\Big| |\mathsf{f}_I[x](y)| - 1 \Big| > 1/4$. Therefore in this case

$$\| \, |\mathsf{f}_I[x]| - 1 \, \|_2^2 \geq \frac{1}{16}(1 - 4^4(\delta_p)^{-4k}\eta^{3/2}) \gg \tau^{3/4} > \eta^{3/4} \geq \mathsf{Vr}_{\mathsf{f}_I[x]}(I)$$

The case $x \notin \mathcal{C} \cup \mathcal{D}$. Recall that $\mathsf{Vr}_{\mathsf{f}_I[x]}(I) = \mathcal{V}(\mathsf{f}_I[x])$ and note that $\| \, |\mathsf{f}_I[x]| - 1 \, \|_2^2$ is bounded from below by $\mathcal{V}(|\mathsf{f}_I[x]|)$. We thus show that $\mathcal{V}(|\mathsf{f}_I[x]|) \gtrsim \mathcal{V}(\mathsf{f}_I[x])$. For this purpose, we assume without loss of generality that $\widehat{\mathsf{f}_I[x]}(\emptyset)$ is positive (it is therefore at least $1/2$ and at most $3/2$). One sees that

$$\begin{aligned}
\mathcal{V}(\mathsf{f}_I[x]) - \mathcal{V}(|\mathsf{f}_I[x]|) &= \| \, \mathsf{f}_I[x]\|_1^2 - \left|\widehat{\mathsf{f}_I[x]}(\emptyset)\right|^2 = \\
&= \left(\| \, \mathsf{f}_I[x]\|_1 + \widehat{\mathsf{f}_I[x]}(\emptyset)\right)\left(\| \, \mathsf{f}_I[x]\|_1 - \widehat{\mathsf{f}_I[x]}(\emptyset)\right) \leq \\
&\leq \left(\| \, \mathsf{f}_I[x]\|_2 + \widehat{\mathsf{f}_I[x]}(\emptyset)\right)\left(\| \, \mathsf{f}_I[x]\|_1 - \widehat{\mathsf{f}_I[x]}(\emptyset)\right) \leq \\
&\leq \left(\left(\mathcal{V}(\mathsf{f}_I[x]) + \widehat{\mathsf{f}_I[x]}(\emptyset)^2\right)^{1/2} + \widehat{\mathsf{f}_I[x]}(\emptyset)\right)\left(\| \, \mathsf{f}_I[x]\|_1 - \widehat{\mathsf{f}_I[x]}(\emptyset)\right) \leq \\
&\leq 6\left(\| \, \mathsf{f}_I[x]\|_1 - \widehat{\mathsf{f}_I[x]}(\emptyset)\right) = \\
&= 6 \mathop{\mathbb{E}}_{y \sim \mu_p^I}\left[| \, \mathsf{f}_I[x](y) \, | - \mathsf{f}_I[x](y)\right] = \\
&= 6 \mathop{\mathbb{E}}_{y \sim \mu_p^I}\left[| \, \mathsf{f}_I[x](y) \, | \cdot \mathbf{1}_{\{\mathsf{f}_I[x](y) < 0\}}\right] \leq 6 \int_{t=0}^{\infty} \Pr\left[\mathsf{f}_I[x] < -t\right] dt = \\
&= 6 \int_{t=0}^{\infty} \Pr\left[\mathsf{f}_I[x] - \widehat{\mathsf{f}_I[x]}(\emptyset) + \widehat{\mathsf{f}_I[x]}(\emptyset) > -t\right] dt \leq
\end{aligned}$$

(using claim 3.4)

$$\leq 6(\delta_p)^{-4k}\eta^{3/2} \int_{t=\widehat{\mathsf{f}_I[x]}(\emptyset)}^{\infty} t^{-4} dt \leq 20(\delta_p)^{-4k}\eta^{3/2}$$

Hence we are done. ∎

Completion of the argument

¿From Proposition 7.3 and Proposition 7.2, we have

$$\begin{aligned}\eta = \mathsf{Vr}_{\mathsf{f}}(I) &= \mathop{\mathbb{E}}_{x\sim\mu_p^{\bar{I}}}\big[\mathsf{Vr}_{\mathsf{f}_I[x]}(I)\big] = \\ &= \mathop{\mathbb{E}}_{x\sim\mu_p^{\bar{I}}}\big[\mathsf{Vr}_{\mathsf{f}_I[x]}(I)\cdot \mathbf{1}_{\{x\in\mathcal{D}\}}\big] + \mathop{\mathbb{E}}_{x\sim\mu_p^{\bar{I}}}\big[\mathsf{Vr}_{\mathsf{f}_I[x]}(I)\cdot \mathbf{1}_{\{x\notin\mathcal{D}\}}\big] < \\ &< 512(\delta_p)^{-4k}\eta^{5/4} + \mathop{\mathbb{E}}_{x\sim\mu_p^{\bar{I}}}\big[\|\,|\mathsf{f}_I[x]| - 1\,\|_2^2\cdot \mathbf{1}_{\{x\notin\mathcal{D}\}}\big] + 20(\delta_p)^{-4k}\eta^{3/2} \leq \\ &\leq 532(\delta_p)^{-4k}\eta^{5/4} + \mathop{\mathbb{E}}_{x\sim\mu_p^{\bar{I}}}\big[\|\,|\mathsf{f}_I[x]| - 1\,\|_2^2\big] = \\ &= 532(\delta_p)^{-4k}(\mathsf{Vr}_{\mathsf{f}}(I))^{5/4} + \|\,|\mathsf{f}| - 1\,\|_2^2 = \\ &= 532(\delta_p)^{-4k}\eta^{5/4} + \epsilon\end{aligned}$$

¿From which it follows that

$$\eta\Big(1 - 532(\delta_p)^{-4k}\eta^{1/4}\Big) < \epsilon \tag{5}$$

We now select τ to be

$$\frac{(\delta_p)^{16k}}{3(1064)^4}$$

Since $\eta < 3\tau$, we have $532(\delta_p)^{-4k}\eta^{1/4} < 1/2$, and thus Equation (5) yields $\eta < 2\epsilon$. Recall that $\frac{1}{1-x} \leq 1 + 2x$ for $0 < x < 1/2$, hence putting this into Equation (5) again, we get

$$\begin{aligned}\mathsf{Vr}_{\mathsf{f}}(I) = \eta &< \frac{\epsilon}{1 - 532(\delta_p)^{-4k}\eta^{1/4}} < \epsilon\big(1 + 1064(\delta_p)^{-4k}\eta^{1/4}\big) < \\ &< \epsilon\big(1 + 1064(\delta_p)^{-4k}(2\epsilon)^{1/4}\big)\end{aligned}$$

thus completing the proof.

7.2 Proof of Lemma 7.1

Before we prove Lemma 7.1, we need the following technical observation.

Lemma 7.4. *Let $\lambda_1, \ldots, \lambda_m$ be (not all zero) real numbers, and let $y_1, \ldots, y_n$ be independent random variables, distributed uniformly on $\{-1, 1\}$. Then*

$$\Pr\left[\Big(\sum_{i=1}^n \lambda_i y_i\Big)^2 > \frac{1}{4}\sum_{i=1}^n \lambda_i^2\right] > \frac{1}{16}$$

Proof. Set $\lambda^2 \doteq \sum_i \lambda_i^2$, and let

$$p(t) \doteq \Pr\Big[\Big(\sum_{i=1}^n \lambda_i y_i\Big)^2 > t\Big]$$

Then

$$\lambda^2 = \mathbb{E}\Big(\sum_{i=1}^{n}\lambda_i y_i\Big)^2 = \int_{t=0}^{\infty} p(t)dt = \int_{t=0}^{\lambda^2/4} p(t)dt + \int_{t=\lambda^2/4}^{8\lambda^2} p(t)dt + \int_{t=8\lambda^2}^{\infty} p(t)dt \le$$
$$\le \lambda^2/4 + 8\lambda^2 p(\lambda^2/4) + \int_{t=8\lambda^2}^{\infty} p(t)dt \tag{6}$$

Let us bound the last term on the right-hand side of (6). We use Azuma's inequality (Theorem 9).

$$\int_{t=8\lambda^2}^{\infty} \Pr\big[\Big(\sum_{i=1}^{n}\lambda_i y_i\Big)^2 > t\big]dt < 2\int_{t=8\lambda^2}^{\infty} \exp\left(-\frac{t}{2\lambda^2}\right)dt = 4\lambda^2\exp\left(-\frac{8\lambda^2}{2\lambda^2}\right) < \lambda^2/4$$

Putting this back into (6) we have

$$p(\lambda^2/4) > \frac{\lambda^2/2}{8\lambda^2} = 1/16$$

which is what we wanted. ∎

Proof of Lemma 7.1: For $x \in \mathcal{P}([l])$ and $y \in \mathcal{P}([l+m]\setminus[l])$, let

$$\mathsf{g}(x\cup y) \doteq \sum_{i\notin y}\mathsf{g}_i(x) - \sum_{i\in y}\mathsf{g}_i(x) = \sum_{i=1}^{m} v_i(y)\mathsf{g}_i(x)$$

where v_i is the i'th Rademacher function for bias 1/2. Then g contains mixed walsh-products (with some biased Rademacher functions and some uniform Rademachers) of size at most $k+1$, and

$$\|\mathsf{g}\|^2_{2,\mu_p^l\times\mu_{1/2}^m} = \sum_{i=1}^{m}\|\mathsf{g}_i\|_2^2$$

According to Lemma 7.4,

$$\Pr_{x\sim\mu_p^m}\Big[\sum|\mathsf{g}_i(x)|^2 > \alpha\Big] \le 16 \Pr_{\substack{x\sim\mu_p^l\\ y\sim\mu_{1/2}^m}}\big[\mathsf{g}(x\cup y)^2 > \alpha/4\big] \tag{7}$$

To bound the right-hand side of (7), we use Claim 3.4 with respect to the measure† $\mu_p^l \times \mu_{1/2}^m$. We obtain, for some global constant δ_p (here δ_p is the minimum between the δ_p that is valid in Theorem 4 for the uniform measure and for the biased measure)

$$\Pr_{x\sim\mu_p^m}\Big[\sum|\mathsf{g}_i(x)|^2 > \alpha\Big] \le 16\cdot 16\alpha^{-2}(\delta_p)^{-4k}\|\mathsf{g}\|_2^4 \le 256\alpha^{-2}(\delta_p)^{-4k}\left(\sum_{i=1}^{m}\|\mathsf{g}_i\|_2^2\right)^2$$

∎

† Claim 3.4 requires Theorem 4. As is shown in [Bec75], this theorem can be applied to a product of two-point spaces, even if each is equipped with a different measure. In our case the coordinates of x lie in two-point spaces with a biased measure, and the coordinates of y are uniformly distributed

8 Improving the Junta Threshold

Building on the strengthening of Theorem 8 by Theorem 10, we turn to prove Theorem 2, improving the tradeoff between $\|\mathsf{f}^{>k}\|_2^2$ and the distance of f from a junta. This is only an improvement in terms of the dependency on ϵ and k – the dependency on p is much worse. Essentially, Theorem 2 is only applicable for a constant bias parameter p, and the dependency on p appears in its statement only for completeness. For convenience, let us first restate Theorem 2.

Theorem 2. *Fix a positive integer ℓ, and set $t \doteq (\delta_p)^{16/\ell}/M$, where M is the constant from Theorem 10, and $d \doteq \frac{1}{2}\left(\frac{p}{2}\right)^{\ell/2t}$.*

Then for every positive integer k, every Boolean function $\mathsf{f}\colon \mathcal{P}([n]) \to \{-1,1\}$ satisfying

$$\|\mathsf{f}^{>k}\|_2^2 \leq d^4\left(\frac{\epsilon}{k}\right)^{\frac{\ell+1}{\ell}},$$

is $O(\epsilon)$-close to a Boolean function dominated by the coordinates in $\mathcal{J}_{k,\tau}(\mathsf{f})$, where $\tau \doteq d^4 {\delta_p}^{4k}\left(\frac{\epsilon}{k}\right)^{(\ell+1)/\ell}$

Proof. The main argument in the proof of Theorem 2 is contained in the following lemma, which we prove in Subsection 8.1 below.

Lemma 8.1. *Fix a positive integer ℓ and set $t \doteq (\delta_p)^{16/\ell}/M$, where M is the constant from Theorem 10, and $d \doteq \frac{1}{2}\left(\frac{p}{2}\right)^{\ell/2t}$. Then for every Boolean function $\mathsf{f}\colon \mathcal{P}([n]) \to \{-1,1\}$, and any parameter $\tau > 0$ and positive integers k and r,*

$$\mathop{\mathbb{E}}_{I\sim\mu_{1/r}^{\bar{\mathcal{J}}}}[\mathsf{Vr}_{\mathsf{f}}(I)] \leq \frac{16}{d^4}{\delta_p}^{-4k}\tau + \frac{4}{d^2}\|\mathsf{f}^{>k}\|_2^2 + \frac{16}{d}\left(\frac{(k/r)^{\ell+1}}{1-k/r} + \|\mathsf{f}^{>k}\|_2^2\right)$$

where $\mathcal{J} \doteq \mathcal{J}_{k,\tau}(\mathsf{f})$.

Following the proof of Theorem 5, but using the parameters of Lemma 8.1, we obtain the following Lemma.

Lemma 8.2. *Fix a positive integer ℓ, and set $t \doteq (\delta_p)^{16/\ell}/M$, where M is the constant from Theorem 10. Then for every Boolean function $\mathsf{f}\colon \mathcal{P}([n]) \to \{-1,1\}$, and any parameter $\tau > 0$ and positive integers k and r,*

$$\mathsf{Vr}_{\mathsf{f}}([n] \setminus \mathcal{J}_{k,\tau}(\mathsf{f})) \leq O\left(\frac{r}{d^4}\left({\delta_p}^{-4k}\tau + \|\mathsf{f}^{>k}\|_2^2 + \frac{(k/r)^{\ell+1}}{1-k/r}\right)\right)$$

where $d \doteq \frac{1}{2}\left(\frac{p}{2}\right)^{\ell/2t}$

Taking

$$r \doteq k\left(\frac{k}{\epsilon}\right)^{1/\ell} \quad \text{and} \quad \tau \doteq d^4 {\delta_p}^{4k}\left(\frac{\epsilon}{k}\right)^{(\ell+1)/\ell}$$

in Lemma 8.2, we obtain Theorem 2. ∎

Noise-Sensitivity

Similarly to Corollary 1.1, which is obtained from Theorem 5, Theorem 2 yields a result concerning noise-insensitive Boolean functions with improved parameters.

Corollary 8.3. *For every positive integer ℓ, there exists a function $\phi : (0, \frac{1}{2}) \to \mathbb{R}^+$, such that the following holds. For any parameter $\lambda > 0$, fix $k = \log_{(1-\lambda)}(1/2)$. Then every Boolean function $\mathsf{f}\colon \mathcal{P}([n]) \to \{-1, 1\}$ whose λ-noise-sensitivity with respect to μ_p^n is bounded by $\phi(p)(\epsilon/k)^{\frac{\ell+1}{\ell}}$ is an $[O(\epsilon), J]$-junta, where*

$$J = O\left(k\phi(p)p^k \left(\frac{\epsilon}{k}\right)^{\frac{\ell+1}{\ell}}\right)$$

Proof. Set $t \doteq (\delta_p)^{16/\ell}/M$, where M is the constant from Theorem 10, $d \doteq \frac{1}{2}\left(\frac{p}{2}\right)^{\ell/2t}$, and $\phi(p) \doteq d^4$.

Now let f be a Boolean function as stated in Corollary 8.3. Then according to proposition 4.1,

$$\phi(p)\left(\frac{\epsilon}{k}\right)^{\frac{\ell+1}{\ell}} \geq \mathrm{NS}_{\lambda,p}(\mathsf{f}) \geq \frac{1}{2} - \frac{1}{2}\sum_S (1-\lambda)^{|S|}\widehat{\mathsf{f}}(S)^2 \geq \frac{1}{2} - \frac{1}{2}\left(\sum_{|S| \leq k} \widehat{\mathsf{f}}(S)^2 + \frac{1}{2}\sum_{|S|>k} \widehat{\mathsf{f}}(S)^2\right)$$

Since f is Boolean, we have $\sum_S \widehat{\mathsf{f}}(S)^2 = \|\mathsf{f}\|_2^2 = 1$, hence we obtain from the above inequality that

$$\sum_{|S|>k} \widehat{\mathsf{f}}(S)^2 \leq \phi(p)\left(\frac{\epsilon}{k}\right)^{\frac{\ell+1}{\ell}}$$

Corollary 1.1 now follows from Theorem 5, recalling that $\delta_p = \Theta(p^{1/4})$. ∎

8.1 Proof of Lemma 8.1

The proof of Lemma 8.1 is similar to that of Lemma 3.1. First, it is shown that on a random I chosen according to $\mu_{1/r}^{\mathcal{J}}$, it is very likely that for almost all input settings x outside I the weight of $\mathsf{f}_I[x]$ is concentrated on Walsh-products of size at most ℓ. Theorem 10 is then applied, showing that in this case $\mathsf{f}_I[x]$ must either have a large Walsh-coefficient, or be close to constant. It is then shown that in fact the first alternative almost never occurs, hence the lemma follows.

8.2 When $k = \ell$

We start by a corollary of Theorem 10, showing that a Boolean function that is concentrated on Walsh-products of size at most ℓ, either has a large Walsh-coefficient, or is very close to constant.

Corollary 8.4. *Fix a positive integer ℓ, and set $t \doteq (\delta_p)^{16/\ell}/M$, where M is the constant from Theorem 10. Then for every Boolean function $\mathsf{g}\colon \mathcal{P}(I) \to \{-1, 1\}$ the following holds. Denoting $\epsilon \doteq \|\mathsf{g}^{>\ell}\|_2^2$, either*

$$\mathsf{Vr}_{\mathsf{g}}(I) < 32\left(\frac{2}{p}\right)^{\ell/t}\epsilon$$

or there exists a non-empty subset $T \subseteq I$ such that $|\widehat{\mathsf{g}}(T)| > \frac{1}{2}\left(\frac{p}{2}\right)^{\ell/2t}$.

Proof. If $\epsilon \geq \frac{1}{32}\left(\frac{p}{2}\right)^{\ell/t}$, then there is nothing to prove, since the right-hand side of the second inequality in the statement of Corollary 8.4 is bigger than 1.

If $\epsilon < \frac{1}{32}\left(\frac{p}{2}\right)^{\ell/t}$ then in particular $\epsilon < t$, and therefore according to Theorem 10, there exists a Boolean function h which only depends on the coordinates of $\mathcal{J}_{\ell,t}(\mathsf{g})$, such that the distance between f and g is bounded by $\epsilon\left(1 + 1064(\delta_p)^{-4\ell}(2\epsilon)^{1/4}\right) < 2\epsilon$. If h is constant, then

$$\mathsf{Vr}_{\mathsf{g}}(I) \leq 2\left(\|\mathsf{g} - \mathsf{h}\|_2^2 + \mathsf{Vr}_{\mathsf{h}}(I)\right) \leq 8\epsilon$$

Since h is Boolean, if it is not constant, then $\|\mathsf{h}^{>0}\|_2^2 \geq p^{\left|\mathcal{J}_{\ell,t}(\mathsf{g})\right|} \geq p^{\ell/t}$. Therefore, since there there are less than $2^{\ell/t}$ non-empty subsets of $\mathcal{J}_{\ell,t}(\mathsf{g})$, there exists a non-empty subset $T \subseteq \mathcal{J}_{\ell,t}(\mathsf{g})$ for which $\widehat{\mathsf{h}}(T)^2 \geq \left(\frac{p}{2}\right)^{\ell/t}$. It follows that $|\widehat{\mathsf{g}}(T)| \geq \frac{1}{2}\left(\frac{p}{2}\right)^{\ell/2t}$, thus completing the proof. ■

8.3 Few Large Coefficients

We now return to the proof of Lemma 8.1 which states, in essence, that for most I's and x's $\mathsf{f}_I[x]$ is almost constant. We begin by giving an upper-estimate on the (weighted) number of restrictions $\mathsf{f}_I[x]$ that can be far from constant. The next subsection will show that indeed most restriction are almost constant.

For a given $I \subseteq [n]$ denote

$$\mathcal{D}_I \doteq \left\{ x \in \mathcal{P}(\bar{I}) \;\middle|\; \exists T \in I \quad |T| \leq \ell \;,\; \left|\widehat{\mathsf{f}_I[x]}(T)\right| > d \right\}$$

(Recall that $d = \frac{1}{2}\left(\frac{p}{2}\right)^{\ell/2t}$, where t is as in Corollary 8.4). To bound the measure of $\mathcal{D}_I$, we note that the coefficient of χ_T in $\mathsf{f}_I[x]$ is a function of x that is concentrated on low-frequencies, and has small norm (since every $i \in I$ has small variation). Hence according to Theorem 4, it cannot often attain large values, and therefore the coefficient of χ_T almost never reaches d.

Fix $T \subseteq I$ to be a non-empty set of size at most ℓ, and consider the function $g_T : \mathcal{P}(\widehat{I}) \to \mathbb{R}$, which assigns to every x the coefficient of χ_T in f_I. That is,

$$g_T(x) = \widehat{\mathsf{f}_I[x]}(T)$$

For x to be in $\mathcal{D}$, one of the g_T's must evaluate to at least d in absolute value. Applying lemma 3.3, with parameters $\alpha = d/2$ and $\beta = d$, we get a bound on the probability, for a random x, that $\mathsf{f}_I[x]$ is a dictatorship.

$$\begin{aligned}
\Pr_{x \sim \mu_p^{\bar{I}}}[x \in \mathcal{D}_I] &\leq \sum_{\substack{T \subseteq I \\ T \neq \emptyset}} \Pr_{x \sim \mu_p^{\bar{I}}}[|g_T(x)| > d] \leq \\
&= 16 d^{-4} \delta_p{}^{-4k} \sum_{\substack{T \subseteq I \\ T \neq \emptyset}} \|\mathsf{g}_T^{\leq k}\|_2^4 + \frac{4}{d^2} \sum_{T \subseteq I} \|\mathsf{g}^{>k}\|_2^2 = \\
&= 16 d^{-4} \delta_p{}^{-4k} \sum_{\substack{T \subseteq I \\ T \neq \emptyset}} \Bigg\| \sum_{\substack{S \subseteq [n], |S| \leq k \\ S \cap I = T}} \widehat{\mathsf{f}}(S) \chi_S \Bigg\|_2^4 + \frac{4}{d^2} \sum_{\substack{T \subseteq I \\ T \neq \emptyset}} \Bigg\| \sum_{\substack{S \subseteq [n], |S| > k \\ S \cap I = T}} \widehat{\mathsf{f}}(S) \chi_S \Bigg\|_2^2 \leq \\
&\leq 16 d^{-4} \delta_p{}^{-4k} \sum_{\substack{T \subseteq I \\ T \neq \emptyset}} \Bigg(\sum_{\substack{S \subseteq [n], |S| \leq k \\ S \cap I = T}} \widehat{\mathsf{f}}^2(S) \Bigg)^2 + \frac{4}{d^2} \|\mathsf{f}^{>k}\|_2^2
\end{aligned}$$

Since the sum of $\widehat{\mathsf{f}}^2(S)$ over *all* S's equals 1, we have

$$\sum_{\substack{T\subseteq I\\ T\neq\emptyset}}\left(\sum_{\substack{S\subseteq[n],|S|\leq k\\ S\cap I=T}}\widehat{\mathsf{f}}^2(S)\right)^2 \leq \max_{\substack{T\subseteq I\\ T\neq\emptyset}}\sum_{\substack{S\subseteq[n],|S|\leq k\\ S\cap I=T}}\widehat{\mathsf{f}}^2(S) \leq \max_{i\in I}\mathsf{Vr}^{\leq k}_{\bar{\mathsf{f}}}(i) < \tau$$

Altogether this implies that,

$$\Pr_{x\sim\mu_p^{\bar{I}}}[x\in\mathcal{D}_I] \leq 16d^{-4}{\delta_p}^{-4k}\tau + \frac{4}{d^2}\|\mathsf{f}^{>k}\|_2^2$$

8.4 Restrictions are Mostly Constant

We are now ready to prove that the restrictions $\mathsf{f}_I[x]$ are mostly constant. First, note that for an x such that $x\notin\mathcal{D}_I$, Corollary 8.4 asserts that

$$\mathsf{Vr}_{\mathsf{f}_I[x]}(I) \leq \frac{16}{d}\sum_{|R|>\ell}\widehat{\mathsf{f}_I[x]}^2(R)$$

and by Claim 2.2 we have that

$$\begin{aligned}\mathop{\mathbb{E}}_{I\sim\mu_{1/r}^{\bar{\mathcal{J}}}}[\mathsf{Vr}_{\mathsf{f}}(I)] = \mathop{\mathbb{E}}_{\substack{I\sim\mu_{1/r}^{\bar{\mathcal{J}}}\\ x\sim\mu_p^{\bar{I}}}}\left[\mathsf{Vr}_{\mathsf{f}_I[x]}(I)\right] \quad &\leq \Pr_{\substack{I\sim\mu_{1/r}^{\bar{\mathcal{J}}}\\ x\sim\mu_p^{\bar{I}}}}[x\in\mathcal{D}_I] + \mathop{\mathbb{E}}_{\substack{I\sim\mu_{1/r}^{\bar{\mathcal{J}}}\\ x\sim\mu_p^{\bar{I}}}}\left[\frac{16}{d}\sum_{|R|>\ell}\widehat{\mathsf{f}_I[x]}^2(R)\right] \quad \leq \\ &\leq 16d^{-4}{\delta_p}^{-4k}\tau + \frac{4}{d^2}\|\mathsf{f}^{>k}\|_2^2 + \frac{16}{d}\mathop{\mathbb{E}}_{I\sim\mu_{1/r}^{\bar{\mathcal{J}}}}\left[\sum_{|S\cap I|>\ell}\widehat{\mathsf{f}}^2(S)\right]\end{aligned}$$

Now note that

$$\mathop{\mathbb{E}}_{I\sim\mu_{1/r}^{\bar{\mathcal{J}}}}\Big[\sum_{\substack{|S|\leq k\\ |S\cap I|>\ell}}\widehat{\mathsf{f}}^2(S)\Big] \leq \frac{(k/r)^{\ell+1}}{1-k/r}\ :$$

This holds since for $S\subseteq[n]$ with $|S|\leq k$,

$$\Pr_I[|S\cap I|>\ell] \leq \sum_{i=\ell+1}^{k}\binom{k}{i}r^{-i}(1-1/r)^{k-i} \leq \sum_{i=\ell+1}^{k}k^i r^{-i} \leq \frac{(k/r)^{\ell+1}}{1-k/r}$$

and since the total weight of all Walsh-products is bounded by 1.

Therefore, we get that the overall probability of disagreement with the majority is bounded by

$$\mathop{\mathbb{E}}_{I\sim\mu_{1/r}^{\bar{\mathcal{J}}}}[\mathsf{Vr}_{\mathsf{f}}(I)] \leq 16d^{-4}{\delta_p}^{-4k}\tau + \frac{4}{d^2}\|\mathsf{f}^{>k}\|_2^2 + \frac{16}{d}\left(\frac{(k/r)^{\ell+1}}{1-k/r} + \|\mathsf{f}^{>k}\|_2^2\right)$$

This completes the proof of Lemma 8.1.

9 The Biased Bonami-Beckner Inequality

In this chapter we prove Theorem 4. As is shown in [Bec75], it is enough to prove the statement for the two-point space, namely for the case where $\mathsf{f}\colon \{\emptyset, \{1\}\} \to \mathbb{R}$. By homogeneity, we may assume that $\mathsf{f}(\emptyset) = 1$, and denote $t \doteq \mathsf{f}(\{1\})$. Also, for convenience we denote $q \doteq 1 - p$. It is easy to verify that

$$T_\delta[\mathsf{f}](\emptyset) = q + \delta p + (1-\delta)pt \ , \quad \text{and} \quad T_\delta[\mathsf{f}](\{1\}) = (1-\delta)q + (p + \delta q)t$$

Hence we should show that

Lemma 9.1. *for every $p \in (0,1)$ there exists a $\delta = \delta_p > 0$, such that for every t,*

$$\left[q(q + \delta p + (1-\delta)pt)^4 + p((1-\delta)q + (p+\delta q)t)^4\right]^{1/4} \le \sqrt{q + pt^2}$$

Proof. The following inequality is equivalent to the inequality stated in lemma 9.1.

$$q(q + \delta p + (1-\delta)pt)^4 + p((1-\delta)q + (p+\delta q)t)^4 - (q + pt^2)^2 \le 0 \qquad (\circ)$$

To prove it for every t (for an appropriate δ_p), it is obviously enough to consider only non-negative t's in $(\circ)$. Moreover, notice that replacing t by $1/t$ in $(\circ)$ and multiplying by t^4 yields the same inequality, only where the roles of p and q are reversed. We can therefore find a parameter δ that is suitable only for t's in the segment $[0,1]$, and then take δ_p to be the minimum between the δ parameters obtained for p and that which is obtained for q (it is known that $\|T_\delta(\mathsf{f})\|_4$ is a decreasing function in δ).

We are thus left with the task of proving that for every $p \in (0,1)$, there exists a positive δ, such that $(\circ)$ holds for every $t \in [0,1]$. Denoting

$$A_\delta(t) \doteq q(q + \delta p + (1-\delta)pt)^4 + p((1-\delta)q + (p+\delta q)t)^4$$

and

$$B(t) \doteq (q + pt^2)^2$$

this is equivalent to finding a δ such that

$$\forall\ t \in [0,1] \qquad \ln A_\delta(t) - \ln B(t) \le 0 \qquad (*)$$

For every δ, $A_\delta(1) = B(1) = 1$, and hence $(*)$ holds there as an equality. We would like to show that the derivative, with respect to t, of the left-hand side of $(*)$ is non-negative in the segment $[0,1]$ for an appropriate δ. This would imply that for this δ, $(*)$ holds throughout the segment.

Since it can also be verified that $A'_\delta(1) = B'(1) = 1$ (where F' denotes the derivative of F with respect to t), we have that the derivative of the left-hand side of $(*)$ zeros at 1 as well. Hence by similar arguments as above, if we prove that the second derivative of the left-hand side of $(*)$ is non-positive for $t \in [0,1]$, for a suitable positive δ, we are done.

What we show, in fact, is that

$$\max_{t \in [0,1]} (\ln A_0 - \ln B)''(t) < 0$$

That is, we show that for $\delta = 0$, the maximum of the second derivative in $(*)$ is strictly negative. Since the maximum is a continuous function of δ, we have that there also exists a positive δ for which the maximum is still negative, which completes the proof.

Taking two derivatives of the left-hand side of $(*)$, we obtain

$$(\ln A_0 - \ln B)''(t) = -4p\left(\frac{p}{(q+pt)^2} + \frac{q-pt^2}{(q+pt^2)^2}\right)$$

Taking a common denominator, we get

$$(\ln A_0 - \ln B)''(t) = -4pq(q+pt)^{-2}(q+pt^2)^{-2}\left(q + 2pqt + (3p^2 - pq)t^2 - 2p^2t^3\right)$$

The third clause in the above expression equals

$$2p^2(t^2 - t^3) + pq(t - t^2) + p^2t^2 + pqt + q$$

which is at least q for every $t \in [0, 1]$. It follows that the second derivative is strictly negative for every such t. The proof of Lemma 9.1, and hence the proof of Theorem 4, is thus complete. ■

References

[Ajt83] M. Ajtai. Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24:1–48, 1983.

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

[BGS98] Mihir Bellare, Oded Goldreich, and Madhu Sudan. Free bits, PCPs, and nonapproximability—towards tight results. *SIAM Journal on Computing*, 27(3):804–915, June 1998.

[BL89] M. Ben-Or and N. Linial. Collective coin flipping. *ADVCR: Advances in Computing Research*, 5, 1989.

[Bou01] J. Bourgain. On the distribution of the fourier spectrum of boolean functions. to appear in Israel J. of Math., 2001.

[DS02] I. Dinur and S. Safra. On the importance of being biased. In *Proc. 34th ACM Symp. on Theory of Computing*, 2002.

[FKN01] E. Friedgut, G. Kalai, and A. Naor. Boolean functions whose fourier transform is concentrated on the first two levels and neutral social choice. To appear in Advances in Applied Mathematics, 2001.

[Fri98] Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[Has86] J. Hastad. *Computational limitations for small depth circuits.* PhD thesis, MIT Press, 1986.

[Hås97] Johan Håstad. Some optimal inapproximability results. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 1–10, El Paso, Texas, 4–6 May 1997.

[Hås99] Johan Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Math.*, 182(1):105–142, 1999.

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th ACM Symp. on Theory of Computing*, 2002.

[KKL88] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions. In IEEE, editor, *29th annual Symposium on Foundations of Computer Science, October 24–26, 1988, White Plains, New York*, pages 68–80. IEEE Computer Society Press, 1988.

[LMN89] Linial, Mansour, and Nisan. Constant depth circuits, fourier transform and learnabiltiy. In *FOCS: IEEE Symposium on Foundations of Computer Science (FOCS)*, 1989.

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

[Sch99] Gideon Schechtman. Concentration, results and applications. *Preprint submitted to Elsevier Preprint*, 1999.

[Tal94] M. Talagrand. On Russo's approximate 0-1 law. *The Annals of Probability*, 22(3):1576–1587, 1994.