Noise-Resistant Boolean-Functions are Juntas

Guy Kindler
Einstein School of Mathematics
The Hebrew University, Israel

Shmuel Safra
School of Mathematical Sciences
Tel Aviv University, Israel

March 28, 2004

Abstract

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

We conclude that any highly noise-resistant Boolean function must be a junta.

Furthermore, we utilize our results to prove a switching lemma, one which may prove useful in the study of computational-complexity lower-bounds for AC^0 or related classes.

1 Introduction

A Boolean function of the form $\mathsf{f} : \{1, -1\}^n \to \{1, -1\}$ (we use 1 and -1 as the two Boolean values) is said to be *symmetric* if the symmetry group over its coordinates is transitive, namely if the affect on f of each coordinate is the same as that of every other coordinate. It is said that f is a J-junta, if there are J or less coordinates whose values completely determine the value of f. Symmetric functions and J-juntas (for small J's) can be viewed as the opposite ends of a wide spectrum, which on one sides contains functions which depend equally on all of their variables, and on the other contains functions that depend greatly on a few of their coordinates but ignore all others.

Social choice. Symmetric functions and juntas play major roles in several areas of science. In the theory of social choice, for example, Boolean functions f represent voting systems: a person votes by assigning one of the coordinates of f to be either 1 or -1, and once the voting is over, f is evaluated on the generated assignment to obtain the result of the election. For the voting system to be fair, f must be symmetric, and if on the other hand f is a J-junta, then just J people control the result of the election.

Noise-sensitivity. Consider a model where each voter decides on her vote randomly, and independently of other voters. Also, suppose that after the election is over, each voter regrets her vote with some probability λ, in which case she casts her vote again according to the same distribution as before. When the election is re-evaluated according to the new votes, the probability that the result of the election be changed is called the λ-noise sensitivity of the election, and is a parameter of the voting system f (the λ-noise sensitivity also depends on the initial distribution according to which each vote is selected).

Symmetry breaking. A natural question arises: how low can the λ-noise sensitivity be for a symmetric voting system. Bourgain considered the setting where each initial vote is chosen uniformly, and in this case showed* [Bou01] that if the noise sensitivity is smaller than a certain threshold, the voting system cannot be symmetric. In fact, he showed that the only functions whose noise sensitivity is smaller than the aforementioned threshold are juntas that depends on a relatively small number of variables, or functions which are very close to being juntas.

Bourgain's result is one of several other "symmetry breaking" results. The first such result by [KKL88], considered the *average sensitivity* of a Boolean function. This is the average, taken over all assignments for a Boolean function f, of the number of individual coordinates that when flipped cause the value of f to change. They showed that a balanced Boolean function $\mathsf{f}\colon \{1,-1\}^n \to \{1,-1\}$ cannot be symmetric if its average sensitivity is below $\log n$ (f is balanced if it obtains the value 1 just as often as it obtains -1).

Moreover, it was shown in [Fri98] that when the average sensitivity becomes significantly lower than $\log n$, symmetry is shattered, and the only Boolean functions left are juntas. Specifically, he showed that if the average sensitivity of a Boolean function f is smaller than k, and $\epsilon > 0$ is any parameter, then f must be ϵ-close to a $9^{k/\epsilon}$-junta (f is ϵ-close to g if $\Pr_x[\mathsf{f}(x) \neq \mathsf{g}(x)] \leq \epsilon$).

Symmetry breaking in complexity. Symmetry breaking results play a crucial role also in complexity, and in numerous hardness of approximation results ([BGS98, Hås97, Hås99, DS98, Kho02] and many others). In these contexts a Boolean function $\mathsf{f} : \{1,-1\}^n \to \{1,-1\}$ is seen as a binary string, the entries of which are indexed by the set $\{1,-1\}^n$, and is used to encode elements in $[n] \doteq \{1,2,\ldots,n\}$ via the *long-code*. The long-code encoding of an element $i \in [n]$ is the 1-junta, also called *dictatorship*, defined by $\mathsf{f}(x) \doteq x_i$.

Typically when using the long-code in a hardness-of-approximation result, a construction or a test of some sort are used to verify that a Boolean function (manifested as a string) satisfies a certain condition. It is then proved that a legal long-code word satisfies the condition, and moreover, that any Boolean function which satisfies the condition has a short list-decoding, namely it must be associated with a small set of coordinates in $[n]$.

The construction in [Kho02], for example, verifies that certain Boolean functions have small noise-sensitivity, and by Bourgain's result, mentioned above, concludes that each of them is close to a junta. The list-decoding of a function is thus the coordinates which dominate the junta associated with it. Other proofs (e.g. [Hås97, Hås99]) use symmetry-breaking results which do not imply that a given function is close to a junta, but rather that there must be a short list of coordinates which affect it much more than other coordinates.

The biased distribution. So far, we only discussed the uniform distribution over $\{1,-1\}^n$. For many uses, however, it is interesting to consider the *p-biased distribution*, μ_p, where each coordinate of the input is independently chosen to be -1 with probability p and 1 with probability $(1-p)$. Note that the distance between two Boolean functions f and g, namely the probability that $\mathsf{f}(x) \neq \mathsf{g}(x)$ for random x, may change drastically when the distribution of x becomes biased. Hence a function which is close to a junta with respect to one bias may be far from all juntas with respect to another.

Several combinatorial properties of Boolean function change when viewed with respect to biased distributions. This was first used in a hardness-of-approximation result by Dinur and the second

*Bourgain's result is actually stated as a theorem concerning the tail of the Fourier transform of Boolean functions. However, it translates easily to the language of noise-sensitivity.

author [DS98] to show hardness for the vertex-cover problem. The proof in [DS98] used the biased version of the result in [Fri98]. It seems that there is a lot of potential for using biased distributions in hardness results, and for that purpose there might be need for symmetry-breaking results which apply for such distributions.

Graph properties. Another 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 permutations of the vertices. A graph G over n vertices can be represented by an element $x \in \{1,-1\}^m$ for $m = \binom{n}{2}$, where each coordinate in x specifies whether a certain edge belongs to G. A graph-property can thus be represented as a Boolean function $\mathsf{f}\colon \{1,-1\}^m \to \{1,-1\}$ which is invariant under a certain set of permutations.

Note that the set of vertex-permutations is transitive over the edges, so graph properties make symmetric Boolean functions. Results concerning symmetric Boolean functions therefore apply to graph properties. When considering graph properties over random graphs, however, most graph properties show interesting behavior only for relatively sparse graphs, namely graphs which contain each edge with some probability p for $p \ll 1/2$. To study such properties, one is therefore interested in results for symmetric functions, that apply for biased distributions.

Our Results

The main contribution of this paper is analogous to the result in [Bou01], but applies in the case of the biased measure. Our proof is also conceptually simpler than that of Bourgain, and the techniques used may be of independent interest. The parameters achieved by our results are, however, somewhat worse than those of [Bou01].

Biased Walsh Products. As mentioned in a footnote above, Bourgain's theorem is formulated in terms of the Fourier expansion of a Boolean function f, which is obtained by writing f as a linear combination of *Walsh products*. The following theorem is the statement of our result in terms of the expansion of a Boolean function as a combination of *p-biased Walsh products*, which is the p-biased version of the Fourier expansion (the definition of p-biased Walsh products follows [Tal94], and appears in the next section).It states that if the weight of the expansion of a Boolean function f on high-frequency Walsh-products is small, then f must be close to a junta. We now give a formal definition of 'being close to a junta', and then state our results.

(ϵ, J)-juntas. A Boolean function f is said to be an (ϵ, J)-junta with respect to the biased measure μ_p, if there exists a J-junta f' such that

$$\Pr_{x \sim \mu_p}[\mathsf{f}(x) \neq \mathsf{f}'(x)] \leq \epsilon$$

Theorem 1. *Let $\mathsf{f}\colon \{1,-1\}^n \to \{1,-1\}$ be a Boolean function satisfying*

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

where the coefficients $\widehat{\mathsf{f}}(S)$ are taken with respect to the p-biased expansion of f. Then f is an

$[O(\sqrt{\epsilon} \cdot k \cdot \log(1/p)/p^2), J]$*-junta with respect to* μ_p*, where*

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

We achieve better parameters if the bias p is essentially constant.

Theorem 2. *For every positive integer* ℓ*, there exists a positive function* $\phi_\ell(p)$ *satisfying the following.*

For every positive integer k*, every Boolean function* $\mathsf{f}\colon \{1,-1\}^n \to \{1,-1\}$ *satisfying*

$$\|\mathsf{f}^{>k}\|_2^2 \leq \epsilon \ ,$$

is an (η, J)*-junta*[†]*, where*

$$J = O\left(\frac{k}{\epsilon p^k}\right) \qquad \text{and} \qquad \eta = \phi_\ell(p) \cdot k\epsilon^{\ell/(\ell+1)} \ .$$

For comparison, we also cite here the result from [Bou01].

Theorem [Bou01]. *Let* $\epsilon, \eta > 0$ *any fixed constants, and let* k *be a positive integer. Then there is a constant* $c_{\eta,\epsilon}$*, such that any Boolean function* $\mathsf{f} : \{1,-1\}^n \to \{1,-1\}$ *satisfying*

$$\sum_{|S|>k} |\widehat{\mathsf{f}}(S)|^2 < c_\eta k^{-\frac{1}{2}-\eta}$$

is an $(\epsilon, k10^k)$*-junta.*

The main difference between our results and that of Bourgain, is how the the threshold on the high-frequency weight, beyond which f is ensured to be close to a junta, behaves as a function of k. While our results require that this weight be bounded by $O(k^{-(\ell+1)/\ell})$, Theorem 1 requires only a bound of order $K^{-1/2-\eta}$ (for η arbitrarily small).

Let us now translate Theorem 1 to the language of noise-sensitivity. Recall that 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. 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* $\mathsf{f}\colon \{1,-1\}^n \to \{1,-1\}$ *whose* λ*-noise-sensitivity with respect to* μ_p^n *is bounded by* ϵ*, is an* $[O(\sqrt{\epsilon} \cdot k \cdot \log(1/p)/p^2), J]$*-junta, where*

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

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

[†]The O notation here, as well as everywhere in this paper, means "up to a factor which is a global constant".

Corollary 1.2. *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 $\mathsf{f}\colon \{1,-1\}^n \to \{1,-1\}$ whose λ-noise-sensitivity with respect to μ_p^n is bounded by $\phi(p)(\epsilon/k)^{\frac{\ell+1}{\ell}}$ is an (η, J)-junta, where*

$$J = O\left(\frac{k}{\epsilon p^k}\right) \qquad \text{and} \qquad \eta = \phi_\ell(p) \cdot k\epsilon^{\ell/(\ell+1)} .$$

The asymptotic distance from a junta. So far we were mostly interested in the threshold on high-frequency weight, beyond which a Boolean function must be a junta. In the case where the high-frequency weight tends to zero, however, it is natural to ask how the distance from a junta behaves as a a function of that weight. We give an asymptotically optimal bound, up to a constant factor, on the distance from a junta in that case. Our bound comes into effect when the weight of f on Walsh-products larger than k becomes smaller than some negative exponent in k.

Theorem 3. *There exists a constant M such that for every fixed positive integer k, the following holds. Let $\mathsf{f}\colon \mathcal{P}([n]) \to \{1,-1\}$ be a Boolean function, let $\epsilon \doteq \sum_{|S|>k} \left|\widehat{\mathsf{f}}(S)\right|^2$ and denote $\tau \doteq \delta_p^{16k}/M$. If $\epsilon < \tau$ then f is an $\Big(\big(1 + 1064(\delta_p)^{-4k}(2\epsilon)^{1/4}\big)\epsilon\, ,\, k/\tau\Big)$-junta.*

Theorem 3 was proven in [FKN01] for the case $k = 1$, and with respect to the uniform measure $\mu_{1/2}$. We prove Theorem 3 by first giving an alternative proof for the case $k = 1$, which is valid with respect to every μ_p, and then extending the proof to the case $k > 1$.

A switching lemma. Another context in which our results may be of interest is that of switching lemmas. Switching lemmas are used (e.g. in [Ajt83, Has86]) to study complexity classes such as AC^0 (the class of circuits of non bounded fan-in and of constant depth), which lie low in the complexity class hierarchy. For example, a switching lemma is used by Hastad in [Has86] to show that the *parity* function ($= \chi_{[n]}$) is not computable in AC^0 .

Hastad's result is improved and extended in [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 frequency. 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 have been sought after for some time$^{\rightarrow}$. 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. do they leave it as an open question!?

Structure of the paper

$^{\rightarrow}$In Section 2 we define the biased Walsh products, and the biased Fourier expansion of functions. verify! We also define other notions related to Boolean functions, and show their connections with the biased Fourier expansion. In Section 3 we prove Theorem 1 in a slightly stronger form, which not only states that the given function is close to a junta, but also points to the influential coordinates in it.

In Section 4 we prove Corollary 1.1 by translating Theorem 1 from the language of Fourier expansion to the language of noise-sensitivity. We also give an example to show that 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 the theorem of [FKN01], showing that a Boolean function which is almost linear is close to a dictatorship. This proof not only holds in 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 whose parameters are much better, 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 6. By translating it into the language of noise-sensitivity, we obtain Corollary 1.2.

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 formally defined, in this notation, as follows. For every 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 $\langle \mathsf{f}, \mathsf{g} \rangle \doteq \mathbb{E}_{x\sim\mu_p^n}[\mathsf{f}(x)\mathsf{g}(x)]$. For every real q, $q \geq 1$, the q-norm of a function $\mathsf{f}\colon \mathcal{P}([n]) \to \mathbb{R}$, is defined by

$$\|\mathsf{f}\|_q \doteq \left(\mathop{\mathbb{E}}_{x\sim\mu_p^n}[|\mathsf{f}(x)|^q]\right)^{1/q}$$

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 biased 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

$$\mathsf{Avg}_I[\mathsf{f}] = \sum_{S \cap I = \emptyset} \widehat{\mathsf{f}}(S)\chi_S \tag{1}$$

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

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

Note that for any two identically distributed and independent random variable X, Y taking values in $\{1, -1\}$, $\Pr[X \neq Y] = V(X)/2$, where $V[X]$ denotes the variance (not variation) of X.

The following is therefore an equivalent definition for the variation, but it extends for non Boolean functions as well.

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*

$$\mathop{\mathbb{E}}_{y\sim\mu_p^{[n]\setminus I}}\left[\mathop{V}_{z\sim\mu_p^I}\left[\mathsf{f}(y\cup z)\right]\right]$$

The following claim, which follows directly from Definition 2 and from (1), expresses the variation in terms of the averages, and in terms of the Fourier expansion of a given function f.

Claim 2.1. *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* $\mathsf{Vr_f}(I) \doteq \|\mathsf{f} - \mathsf{Avg}_I\,[\mathsf{f}]\|_2^2 = \sum_{S\cap I\neq\emptyset} \widehat{\mathsf{f}}^2(S)$.

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

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 2.2. *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*

$$\Pr_{x\sim\mu_p^n}\left[\mathsf{f}(x) \neq \mathsf{g}(x)\right] < \mathsf{Vr_f}([n]\setminus I)/2$$

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

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

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

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

■

2.4 Restrictions

Let $\mathsf{f}\colon \mathcal{P}([n]) \to \mathbb{R}$ be any real-valued function. For a given set of coordinates $I \subseteq [n]$ and any $x \in \mathcal{P}([n]\setminus I)$, let us denote by $\mathsf{f}_I[x]\colon \mathcal{P}(I) \to \{1,-1\}$ the restriction of f 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 as follows. For every $S \subseteq I$, it is easily seen that

$$\widehat{\mathsf{f}_I[x]}(S) = \sum_{\substack{T\subseteq[n]\\ T\cap I=S}} \widehat{\mathsf{f}}(T)\chi_{T\setminus S}(x) \tag{2}$$

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

The variation of a function f on a subset I can be expressed in terms of the restrictions of f to I.

Claim 2.3. *Let $\mathsf{f}\colon \mathcal{P}([n]) \to \mathbb{R}$, let $I \subseteq [n]$, and denote $\bar{I} \doteq [n] \setminus I$. Then*

$$\mathsf{Vr}_{\mathsf{f}}(I) = \mathop{\mathbb{E}}_{x\sim\mu_p^{\bar{I}}}\left[\mathop{V}_{z\sim\mu_p^{I}}\left[\mathsf{f}_I[x](z)\right]\right] = \mathop{\mathbb{E}}_{x\sim\mu_p^{\bar{I}}}\left[\mathsf{Vr}_{\mathsf{f}_I[x]}(I)\right]$$

where $V_{z\sim\mu_p^I}\left[\mathsf{f}_I[x](z)\right]$ denotes the variance of $\mathsf{f}_I[x](z)$, where z is distributed according to μ_p^I.

Proof. Both the first and the second inequalities follow immediately from Definition 2. ∎

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 [Bon70], and later Beckner [Bec75], independently proved that in the case of uniform measure, T_δ is hyper-contractive for appropriate values of δ.

Theorem 4. *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 4 was shown to hold for the biased case as well. We need another special case of this theorem, which is sufficient for our purposes.

Theorem 5. *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.*

Note that in the sequel all the parameters denoted δ_p refer, unless noted otherwise, to the best parameter δ_p for which Theorem 5 holds. This best parameter was, in fact, found recently by K. Oleszkiewicz.

Theorem [Ole02]. *Let δ_p denote the largest parameter for which Theorem 5 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 influence on $\mathsf{f}^{\leq k}$ is large (namely the coordinates i for which $\mathsf{Vr}_{\mathsf{f}^{\leq k}}(\{i\})$ 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 11, 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*

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

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

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

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

Note that the O-notation here, and throughout this paper, only hides global constants.

Proof. Denote $\mathcal{J} \doteq \mathcal{J}_{k,\tau}(f)$, and $\bar{\mathcal{J}} \doteq [n] \setminus \mathcal{J}$. To prove Theorem 6, 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 $r \doteq k/\sqrt{\epsilon}$. 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 contains the main arguments in the proof of Theorem 6.

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

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

Let us show how Lemma 3.1 implies Theorem 6. Note that when a random partition $I_1, \ldots, I_r$ of $[n]$ is chosen, the distribution of each subset I_j in the partition is $\mu_{1/r}^{\bar{\mathcal{J}}}$. 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

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

From Proposition 2.2 we thus obtain that f is $\frac{Cr\log(1/p)}{2p^2}\Big((\delta_p)^{-4k}\cdot\tau + k^2/r^2 + \|\mathsf{f}^{>k}\|_2^2\Big)$-close to a Boolean function that depends only on the coordinates of $\mathcal{J}$. This completes the proof of Theorem 6, since

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

■

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

We now turn to the proof of Lemma 3.1. The idea of the proof is to choose I, to consider random restrictions of the form $\mathsf{f}_I[x]$, and to show that most of these restrictions are almost constant. Using Claim 2.3, this implies that $\mathsf{Vr}_{\mathsf{f}}[I]$ is small.

In the interesting case we expect that $r \gg k$, in which case the Fourier expansion of most of the restrictions $\mathsf{f}_I[x]$ will have very low weight on frequencies higher than 1. The result of [FKN01], which we extend below to the biased case, ensures that such functions are close to being a dictatorship (namely a 1-junta). However since we know that the coordinates in I have little influence, we can show that most of these restrictions are in fact almost constant.

Almost linear functions

Let us cite here a corollary that is proven in Section 6, which is a biased version of the result in [FKN01].

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 $\epsilon \leq \frac{p^2}{40(\log(\frac{1}{p^2})+6)}$. Then f is $\Big(1+\frac{60}{p^2}\exp(-\frac{p^2}{40\epsilon})\Big)\epsilon$-close to some Boolean dictatorship.*

For the proof of Lemma 3.1 we need Proposition 3.2 below, which follows from Corollary 6.1. Proposition 3.2 essentially states 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 higher frequencies.

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*

$$\mathcal{V}(\mathsf{f}) = \mathsf{Vr}_{\mathsf{f}}([m]) = \|\mathsf{f}^{>0}\|_2^2 \leq \frac{M\log(1/p)}{p^2}\cdot\|\mathsf{f}^{>1}\|_2^2 \tag{3}$$

Proof. Setting $\epsilon = \|\mathsf{f}^{>1}\|_2^2$, one notes that there is nothing to prove in the case $\epsilon > \frac{p^2}{40(\log(\frac{1}{p^2})+6)}$, since taking M to be a large-enough constant, the right-hand side of (3) is larger than 1.

If $\epsilon \leq \frac{p^2}{40(\log(\frac{1}{p^2})+6)}$, 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 $|\widehat{\mathsf{g}}(\{i\})| = 2\sqrt{p(1-p)} \geq \sqrt{2p}$. Hence if f is 2ϵ-close to such a function there must be a coordinate i for which $|\widehat{\mathsf{f}}(\{i\})| > \sqrt{p}$. If f is 2ϵ-close to a constant dictatorship then (3) obviously holds. ∎

Few Non-Constant Dictatorships

Let I be a random set of coordinates, as specified in Lemma 3.1. Denote $\bar{I} \doteq [n] \setminus I$, and consider the restriction $\mathsf{f}_I[x]$ for a random $x \in \mathcal{P}(\bar{I})$.

The proof of Lemma 3.1 continues as follows. We first show that there may only be a few values of x for which $\mathsf{f}_I[x]$ is of a '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 $\mathsf{f}_I[x]$ 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 consider the coefficient of χ_i in $\mathsf{f}_I[x]$ as a function of x. As explained in more detail below, it is a function of small norm (since every $i \in I$ has small influence), and has a low weight on higher frequencies. To bound the probability, over a random restriction $\mathsf{f}_I[x]$, that the coefficient of χ_i is large, we use the following bounds on large deviations of low-degree functions.

Claim 3.3. *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*

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

Proof. By applying Markov's inequality for $|\mathsf{g}|^4$ and then applying Theorem 5, we have $\alpha^4 \cdot \Pr_{x\sim\mu_p^m}\left[|\mathsf{g}(x)| > \alpha\right] \leq \|\mathsf{g}\|_4^4 \leq (\delta_p)^{-4k}\|\mathsf{g}\|_2^4$. ∎

Lemma 3.4. *Let $0 < \alpha < \beta$ be any parameters. Then for any function $\mathsf{g}\colon \mathcal{P}([m]) \to \mathbb{R}$*

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

Proof. We break g into its low-frequency and its high-frequency parts, and note that

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

∎

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. To bound the probability of a random restriction $\mathsf{f}_I[x]$ to be a dictatorship, we use (2) to get the Fourier expansion of g_i, and then apply Lemma 3.4 with parameters $\alpha = \sqrt{p}/2$ and $\beta = \sqrt{p}$ to it.

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

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

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

Altogether this implies that for some constant M_1,

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

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 thus by Claim 2.3 we have

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

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

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

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

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

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

This completes the proof of Lemma 3.1.

4 Discussion of Theorem 6

4.1 A Corollary Concerning Noise-Sensitivity

Let us next translate Theorem 6 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*

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

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

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

So

$$\begin{aligned}
1-2\mathrm{NS}_{\lambda,p}(\mathsf{f}) &= \mathop{\mathbb{E}}_{x\sim\mu_p^n\,,\ I\sim\mu_\lambda^n\,,\ z\sim\mu_p^I}\Big[\mathsf{f}(x)\mathsf{f}\Big((x\setminus I)\cup z\Big)\Big] \\
&= \sum_{S,T\subseteq[n]} \widehat{\mathsf{f}}(S)\widehat{f}(T)\mathbb{E}_{x,\ z,\ I}[\chi_S(x)\chi_T((x\setminus I)\cup z)] \\
&= \sum_{S,T\subseteq[n]} \widehat{\mathsf{f}}(S)\widehat{f}(T)\mathbb{E}_I\Big[\mathbb{E}_{x,\ z}\Big[\chi_S(x)\chi_{T\setminus I}(x)\chi_{T\cap I}(z)\Big]\Big] \qquad (4)
\end{aligned}$$

In each term of (4) 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 6, 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 $\mathsf{f}\colon \{1,-1\}^n \to \{1,-1\}$ whose λ-noise-sensitivity with respect to μ_p^n is bounded by ϵ, is an $[O(\sqrt{\epsilon}\cdot k \cdot \log(1/p)/p^2), J]$-junta, where*

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

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

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

$$= \frac{1}{2} - \frac{1}{2}\left(1 - \frac{1}{2}\sum_{|S|>k}\widehat{\mathsf{f}}(S)^2\right) = \frac{1}{4}\sum_{|S|>k}\widehat{\mathsf{f}}(S)^2$$

we thus obtain

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

Corollary 1.1 now follows from Theorem 6. ∎

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*

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

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

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

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*,*

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

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

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

and

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

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

$$\|\mathsf{f}^{>k}\|_2^2 < \frac{C}{k} \tag{5}$$

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 (5) 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 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* $K/\log(l/\epsilon)$*, 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*

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

and c *is some global constant.*

Before we prove Theorem 7, 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 7 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 7. We only give the highlights of the proof (more details are given in the similar proof, of Theorem 8 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* $\|\mathsf{f} - \mathsf{g}\|_2^2 \le \epsilon$*.*

Let g be as in the above claim. Let $c = c_{1/4,\epsilon}$ be the constant that corresponds to taking $\eta = 1/4$ in Bourgain's Theorem, 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 Bourgain's theorem, we thus have that with probability at least $(1 - \epsilon)$, h is an $(\epsilon, k10^k)$-junta. This gives Theorem 7 immediately. ∎

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

Theorem 8. *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* $J \le O(\epsilon^{-2} k^3 p^{-k})$*.*

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* $\|\mathsf{f} - \mathsf{g}\|_2^2 \le 4\epsilon$*.*

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 $\|\mathsf{f} - \mathsf{g}\|_2^2 \le 4\epsilon$. ∎

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

$$\Pr_i [|S \cap I| > k] \leq \sum_{i=k+1}^{m} \binom{m}{i} \left(\frac{\alpha k}{m}\right)^i \left(1 - \frac{\alpha k}{m}\right)^i \leq$$

$$\leq \sum_{i=k}^{m} \frac{m^i}{i!} \left(\frac{\alpha k}{m}\right)^i \leq \sum_{i=k}^{m} \sqrt{2\pi i}(e\alpha)^i \left(\frac{k}{i}\right)^i = \sum_{i=k}^{m} \sqrt{2\pi i}(e\alpha)^i \left(1 - \frac{i-k}{i}\right)^i \leq$$

$$\leq \sum_{i=k}^{m} \sqrt{2\pi i}(e\alpha)^i e^{k-i} \leq \sqrt{2\pi k}(e\alpha)^k \sum_{i=k}^{m} (i-k+1)\alpha^i \leq 2\sqrt{2\pi k}(e\alpha)^k =$$

$$= \frac{2\sqrt{2\pi k} \cdot \epsilon^3}{50^k} \leq \frac{\epsilon^3}{k^2}$$

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 6, h is an $(O(\epsilon \log(1/p)/p^2), J)$-junta. ∎

6 Biased FKN

In this section, we prove that if a Boolean function f is close to being linear, namely $\|\mathsf{f}^{>1}\|_2^2$ is small, then it 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 9. *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*

$$\epsilon \leq \frac{p^2}{40(\log(\frac{1}{p^2}) + 6)} \tag{6}$$

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

$$\|\mathsf{f} - \left(\widehat{\mathsf{f}}(\emptyset) + \widehat{\mathsf{f}}(\{i_o\})\chi_{\{i_o\}}\right)\|_2^2 < \left(1 + \frac{60}{p^2}\exp(-\frac{p^2}{40\epsilon})\right)\epsilon$$

Before we prove Theorem 9, 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* $\epsilon \leq \frac{p^2}{40(\log(\frac{1}{p^2})+6)}$. *Then* f *is* $\left(1 + \frac{60}{p^2}\exp(-\frac{p^2}{40\epsilon})\right)\epsilon$*-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 9, there is some coordinate $i_o \in [n]$ such that

$$\mathsf{Vr}_{\mathsf{f}}([n] \setminus \{i_o\}) = \|\mathsf{f} - \mathsf{Avg}_{[n]\setminus\{i_o\}}[\mathsf{f}]\|_2^2 = \|\mathsf{f} - \left(\widehat{\mathsf{f}}(\emptyset) + \widehat{\mathsf{f}}(\{i_o\})\chi_{\{i_o\}}\right)\|_2^2 \leq$$

$$\leq \epsilon + \|\mathsf{f}^{\leq 1} - \left(\widehat{\mathsf{f}}(\emptyset) + \widehat{\mathsf{f}}(\{i_o\})\chi_{\{i_o\}}\right)\|_2^2 < 2\left(1 + \frac{60}{p^2}\exp(-\frac{p^2}{40\epsilon})\right)\epsilon$$

It follows from Proposition 2.2 that there exists a Boolean function g that depends only on the coordinate i_o (and is thus a dictatorship), such that

$$\Pr_{x\sim\mu_p^n}[\mathsf{f}(x)\neq\mathsf{g}(x)] < \left(1+\frac{60}{p^2}\exp(-\frac{p^2}{40\epsilon})\right)\epsilon$$

Therefore f is a $\left(\left(1+\frac{60}{p^2}\exp(-\frac{p^2}{40\epsilon})\right)\epsilon\,,\,1\right)$-junta. ∎

Proof of Theorem 9: For simplicity, we write $\mathsf{f} = a_0 + \sum_{i=1}^{n} a_i\chi_{\{i\}}$, and assume without loss of generality that $|a_1| \geq |a_2| \geq \ldots \geq |a_n|$. We also denote $q = 1-p$. Our goal is then to prove that $\sum_{i=2}^{n}|a_i|^2 < \left(1+\frac{60}{p^2}\exp(-\frac{p^2}{40\epsilon})\right)\epsilon$. 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, the difference of which is $\sqrt{p/q}+\sqrt{q/p}$.

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 $|\,|\mathsf{f}(x)|-1| \geq \frac{1}{2}$. Together with (6) we thus obtain $\||\mathsf{f}(x)|-1\|_2^2 \geq \frac{p^2}{4} > \epsilon$, 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 can thus choose $m \in \{2,\ldots,n\}$ to be the minimal index satisfying

$$\sum_{i=m}^{n}|a_i|^2 \leq (\frac{4q}{p}+2)\epsilon \tag{7}$$

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 9, 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* $\epsilon \leq p^2/60$. *Then it also holds that*

$$\|\mathsf{f}^{>0}\|_2^2 < \left(1 + \frac{60}{p^2} \exp\left(-\frac{p^2}{40\epsilon} \right) \right) \epsilon$$

Before proving Lemma 6.3, let us show how it concludes the proof of Theorem 9. We apply Lemma 6.3 to $\mathsf{f}_I[y]$, noting that $\mathsf{f}_I[y]$ satisfies its conditions, and obtain that

$$\sum_{i=m}^{n} |a_i|^2 = \|\mathsf{f}_I^{>0}[y]\|_2^2 < \left(1 + \frac{60}{p^2} \exp(-\frac{p^2}{40\epsilon}) \right) \epsilon$$

If $m = 2$, this is what we wanted to show. If $m > 2$, noting that by (6) the bound above is smaller than 2ϵ, we obtain from Claim 6.2 that m is not the minimal index satisfying (7), 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 $\text{sign}(\mathsf{f}(x)) = \text{sign}(b)$ 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 $\mathcal{V}(\mathsf{f}) \approx \mathcal{V}(|\mathsf{f}|)$.

Let us now commence with the actual proof. According to the definition of ϵ,

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

and hence, using (6),

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

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

$$\mathbb{E}|\mathsf{f}| - \mathbb{E}\mathsf{f} \leq 2\mathbb{E}|\mathsf{f}|\mathbf{1}_{\{\mathsf{f}<0\}} \tag{8}$$

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 10 (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*

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

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

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

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

$$\mathbb{E}|\mathsf{f}|\mathbf{1}_{\{\mathsf{f}<0\}} = \frac{2(\frac{4q}{p} + 2)\epsilon}{pb}\exp\left(\frac{-pb^2}{2(\frac{4q}{p} + 2)\epsilon}\right) \leq \frac{20\epsilon}{p^2}\exp\left(-\frac{p^2}{40\epsilon}\right) \tag{9}$$

It now follows from (8) and (9) that

$$\begin{aligned}
\epsilon > \||\mathsf{f}| - 1\|_2^2 \geq \mathcal{V}(|\mathsf{f}|) = \|\mathsf{f}\|_2^2 - \mathbb{E}|\mathsf{f}|^2 = \mathcal{V}(\mathsf{f}) + \mathbb{E}\mathsf{f}^2 - (\mathbb{E}|\mathsf{f}|)^2 = \\
= \mathcal{V}(\mathsf{f}) + (\mathbb{E}\mathsf{f} + \mathbb{E}|\mathsf{f}|)(\mathbb{E}\mathsf{f} - \mathbb{E}|\mathsf{f}|) \geq \qquad (10) \\
\geq \mathcal{V}(\mathsf{f}) - (\mathbb{E}\mathsf{f} + \mathbb{E}|\mathsf{f}|) \cdot \frac{20\epsilon}{p^2}\exp\left(-\frac{p^2}{40\epsilon}\right) \qquad (11)
\end{aligned}$$

Noting that

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

we have that (10) implies

$$\|\mathsf{f}^{>0}\|_2^2 = \mathcal{V}(\mathsf{f}) \leq \epsilon + \frac{60\epsilon}{p^2}\exp\left(-\frac{p^2}{40\epsilon}\right)$$

which completes the proof.

7 Extending FKN to Higher Frequencies

Following, is an extension of the proof method of Theorem 9, 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 6: 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. $\rightarrow$ 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 the following theorem is used to improve the parameters in Theorem 6.

Theorem 11 (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*

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

Note that theorem 3 follows from Theorem 11, using the same proof as in Corollary 6.1.

7.1 Proof of Theorem 11

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 9 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 2.3 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* $\alpha \geq 0$,

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

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

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

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

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

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 $y \in \mathcal{P}(I)$, $\left| |\mathsf{f}_I[x](y)| - 1 \right| \geq 1/4$. Note that $\mathsf{f}_I[x] - \widehat{\mathsf{f}_I[x]}(\emptyset)$ is a function of degree at most k, and that since $x \notin \mathcal{D}$,

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

Hence by Claim 3.3

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

It follows that with probability at least $1 - 4^4(\delta_p)^{-4k}\eta^{3/2}$, $\left| |\mathsf{f}_I[x](y)| - 1 \right| > 1/4$. 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 3.3)

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

Hence we are done. ∎

Completion of the argument

From Proposition 7.3 and Proposition 7.2, we have

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

From which it follows that

$$\eta\Big(1 - 532(\delta_p)^{-4k}\eta^{1/4}\Big) < \epsilon \tag{12}$$

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 (12) yields $\eta < 2\epsilon$. Recall that $\frac{1}{1-x} \leq 1 + 2x$ for $0 < x < 1/2$, hence putting this into Equation (12) 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

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

Then

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

$$\leq \lambda^2/4 + 8\lambda^2 p(\lambda^2/4) + \int_{t=8\lambda^2}^{\infty} p(t)dt \tag{13}$$

Let us bound the last term on the right-hand side of (13). We use Azuma's inequality (Theorem 10).

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

Putting this back into (13) 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,

$$\Pr_{x \sim \mu_p^m}\left[\sum |\mathsf{g}_i(x)|^2 > \alpha\right] \leq 16 \Pr_{\substack{x \sim \mu_p^l \\ y \sim \mu_{1/2}^m}}\left[\mathsf{g}(x \cup y)^2 > \alpha/4\right] \tag{14}$$

To bound the right-hand side of (14), we use Claim 3.3 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 5 for the uniform measure and for the biased measure)

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

■

‡ Claim 3.3 requires Theorem 5. 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 9 by Theorem 11, 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 considerably 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. *For every positive integer ℓ, there exists a positive function $\phi_\ell(p)$ satisfying the following.*

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

$$\|\mathsf{f}^{>k}\|_2^2 \leq \epsilon ,$$

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

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 $t \doteq \left(\delta_p\right)^{16/\ell}/M$, where M is the constant from Theorem 11, 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,*

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

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

Following the proof of Theorem 6, but using the parameters of Lemma 8.1, we obtain the following Lemma.

Lemma 8.2. *Fix a positive integer ℓ, and set $t \doteq \left(\delta_p\right)^{16/\ell}/M$, where M is the constant from Theorem 11. 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,*

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

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

Taking

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

in Lemma 8.2, we obtain Theorem 2 with $\phi_\ell(p) = d^{-4}$. ∎

Noise-Sensitivity. Note that Corollary 1.2 easily follows from Theorem 2. It can be derived by plugging the parameters of Theorem 2 into the proof that was used to obtain Corollary 1.1 from Theorem 6. We do not repeat the proof here.

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 $\mu_{1/r}^{\bar{\mathcal{J}}}$, 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 11 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 11, 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 8.3. *Fix a positive integer ℓ, and set $t \doteq (\delta_p)^{16/\ell}/M$, where M is the constant from Theorem 11. 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 8.3 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 11, 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 $\epsilon\big(1 + 1064(\delta_p)^{-4\ell}(2\epsilon)^{1/4}\big) < 2\epsilon$. If h is constant, then

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

Since h is Boolean, if it is not constant, then $\|\mathsf{h}^{>0}\|_2^2 \geq p^{|\mathcal{J}_{\ell,t}(\mathsf{g})|} \geq p^{\ell/t}$. 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 8.3). 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 5, 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 3.4, 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.

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

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

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

Altogether this implies that,

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

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 8.3 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 2.3 we have that

$$\mathop{\mathbb{E}}_{I\sim\mu_{1/r}^{\bar{\mathcal{J}}}}[\mathsf{Vr}_{\mathsf{f}}(I)] = \mathop{\mathbb{E}}_{\substack{I\sim\mu_{1/r}^{\bar{\mathcal{J}}}\\ x\sim\mu_p^{\bar{I}}}}\left[\mathsf{Vr}_{\mathsf{f}_{I[x]}}(I)\right] \le \Pr_{\substack{I\sim\mu_{1/r}^{\bar{\mathcal{J}}}\\ x\sim\mu_p^{\bar{I}}}}[x\in\mathcal{D}_I] + \mathop{\mathbb{E}}_{\substack{I\sim\mu_{1/r}^{\bar{\mathcal{J}}}\\ x\sim\mu_p^{\bar{I}}}}\left[\frac{16}{d}\sum_{|R|>\ell}\widehat{\mathsf{f}_I[x]}^2(R)\right] \le$$

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

Now note that

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

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

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

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

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

This completes the proof of Lemma 8.1.

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.

[Bon70] Aline Bonami. Etude des coefficients fourier des fonctiones de $l^p(g)$. *Ann. Inst. Fourier*, 20(2):335–402, 1970.

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

[DS98] I. Dinur and S. Safra. Monotone-minimum-satisfying assignment is NP-hard for almost polynomial factors. Manuscript, 1998.

[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.