Noise Sensitivity on the p-Biased Hypercube

Noam Lifshitz
Einstein Institute of Mathematics
Hebrew University of Jerusalem
Israel
noamlifshitz@gmail.com

Dor Minzer
Institute for Advanced Study
Princeton, USA
minzer.dor@gmail.com

***Abstract*—The noise sensitivity of a Boolean function measures how susceptible the value of f on a typical input x to a slight perturbation of the bits of x: it is the probability $f(x)$ and $f(y)$ are different when x is a uniformly chosen n-bit Boolean string, and y is formed by flipping each bit of x with small probability ε. The noise sensitivity of a function is a key concept with applications to combinatorics, complexity theory, learning theory, percolation theory and more.**

In this paper, we investigate noise sensitivity on the p-biased hypercube, extending the theory for polynomially small p. Specifically, we give sufficient conditions for monotone functions with large groups of symmetries to be noise sensitive (which in some cases are also necessary). As an application, we show that the 2-SAT function is noise sensitive around its critical probability.

En route, we study biased versions of the invariance principle for monotone functions and give p-biased versions of Bourgain's tail theorem and the Majority is Stablest theorem, showing that in this case the correct analog of "small low degree influences" is lack of correlation with constant width DNF formulas.

***Keywords*-Analysis of Boolean Functions, Noise Sensitivity, Graph Properties.**

I. Introduction

Analysis of Boolean Functions is an integral part of theoretical computer science, with many applications in PCP's, pseudo-random generators, learning theory and more. Traditionally, one applies existing analytical results, such as the case, for example, in many PCP constructions (e.g. Friedgut's junta theorem [1] in [2]) and metric nonembeddability results (e.g. Bourgain's noise sensitivity theorem [3] in [4], [5]). But at other times, the analytical results are not known and have to be developed for the applications in TCS: such as the case in the Majority is Stablest theorem, first appearing in [6] as a conjecture and subsequently proved in [7] (introducing the invariance principle for this purpose, which has become an important tool since then). Another example is the technique of polarizing random walks from [8] introduced for constructing pseudo-random generators, that was subsequently used in the context of quantum computing to prove oracle separation results [9]. In fact, many developments in the study of Boolean functions are directly motivated from TCS perspective.

The resolution of the 2-to-2 Games Conjecture of Khot [10] is a recent such example [11], [12], [13], [14]. Therein, the correctness of the PCP construction is reduced to a purely combinatorial/analytical question about the structure of small sets with small edge boundary in the Grassmann graph, a question that is intimately related to hypercontractive inequalities on that graph. The vertices of the Grassmann graph are all ℓ-dimensional subspaces of a given vector space V of dimension k, and for the PCP application one has $\ell \ll k$, thus the graph is closely related to the p-biased hypercube for small p (identifying subspaces over V with their indicator vector in $\{0,1\}^V$). The characterization of small sets with small edge boundary [14] can be indeed viewed as a hypercontractive-type inequality, and in fact the phenomenon appearing there also appears in various of different domains [15], [16]. We remark that in general, these hypercontractive-type results are only meaningful for a subclass of functions, making them harder to apply than in usual domains.

The p-biased hypercube for small p is also prevalent in the study of critical probabilities and sharp thresholds, dating back to the work of Erdős and Rényi [17]. Therein, one has a monotone Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ – for example a graph property, where the input x is the adjacency matrix of a graph, and $f(x)$ is 1 if the graph has the property (such as connectedness and containing hamiltonian cycle) and one is interested in the behavior of the function $\mu_p(f) \stackrel{def}{=} \Pr_{x\sim\mu_p^n}[f(x)=1]$ as p varies. Here, μ_p^n denotes the probability measure on $\{0,1\}^n$ where each input bit is independently chosen to be 1 with probability p, and otherwise 0 (e.g. when x specifies a graph, this is the Erdős-Rényi random graph model). Since f is monotone, $\mu_p(f)$ is an increasing function of p, and it is natural to ask: what is its rate of increase? what is its behavior around the point it is balanced at, e.g. p_c such that $\mu_{p_c}(f) = 1/2$? It turns out that many functions of interest have a sharp increase: there is $\varepsilon(n) = o(1)$ such that in the interval $[p_c-\varepsilon(n), p_c+\varepsilon(n)]$, $\mu_p(f)$ increases being almost 0 to being almost 1 (e.g. [18], [19]). In case $\varepsilon(n) = o(p_c(1-p_c))$, we say the function has a sharp threshold.

In case p_c is bounded away from 0 and 1, one has a wide

2575-8454/19/$31.00 ©2019 IEEE
DOI 10.1109/FOCS.2019.00075

range of efficient tools to prove sharp-threshold results, most notably the combination of the Russo-Margulis lemma [20], [21] with structural results such as the KKL theorem [22] and Friedgut's Junta theorem [1]. However, in case p_c is sub-constant,[1] and especially when it is polynomially small in n, these results become ineffective and proving sharp threshold result is harder; even proving the function has a sharp threshold (without specifying the value of p_c) takes considerable effort; a notable such example is the k-SAT function, proved to have sharp threshold by Friedgut [23]. In this regime of p_c, there are Friedgut's characterization of graph properties with coarse threshold [23], and weaker (but more general) partial results of Bourgain and Hatami [23], [24].

The situation becomes more difficult when one wishes to study properties stronger than having a sharp threshold, such as noise sensitive, which is the focus of this paper. A function f is said to be noise sensitive if picking x, y according to the p-biased measure in a highly correlated manner, the random variables $f(x), f(y)$ are nearly independent. Here, by "highly correlated manner", we mean picking $x = y$ according to μ_p^n and resampling each coordinate of y with probability ε (according to μ_p).

The current paper extends the theory of noise sensitivity and structural results for small tail Boolean functions [3], [25], [7] to the p-biased setting for polynomially small p. Specifically, we prove a noise sensitivity criteria for monotone functions with large group of symmetries (including, for example, random CSPs and hypergraph properties) and p-biased versions of Bourgain's tail theorem [3] as well as Majority is Stablest theorem [7] for monotone functions. For that purpose, we use the new hypercontractive-type inequalities for the p-biased hypercube proved in [16].

A. Hypercontractivity on the Uniform Hypercube

Consider the uniform hypercube, i.e. the set $\{0,1\}^n$ with the uniform measure. The L_q norm of $f\colon \{0,1\}^n \to \mathbb{R}$, for $q \geqslant 1$, is given by

$$\|f\|_q = \left(\mathop{\mathbb{E}}_{\mathbf{x} \in_R \{0,1\}^n} [|f(\mathbf{x})|^q] \right)^{1/q}.$$

For $\rho \in (0,1)$, the noise operator T_ρ acts on functions $f\colon \{0,1\}^n \to \mathbb{R}$ in the following way: $(T_\rho f)(x)$ is the average of $f(\mathbf{y})$, on $\mathbf{y} \in \{0,1\}^n$ that are ρ-correlated with x, i.e. for every $i \in [n]$, $\mathbf{y}_i = x_i$ with probability ρ, and otherwise $\mathbf{y}_i$ is sampled independently from $\{0,1\}$.

In its most basic form, the hypercontractive inequality [26], [27], [28] on the hypercube states that there exists $\rho \in (0,1)$ ($\rho \leqslant 1/\sqrt{3}$ will do), such that T_ρ is a contraction from L_4 to L_2. That is, for every real-valued function f,

$$\|T_\rho f\|_4 \leqslant \|f\|_2. \tag{1}$$

[1]The case p_c is very close to 1 is analogous.

Note that any averaging operator, and in particular T_ρ, is a contraction from L_p to itself (e.g. $\|T_\rho f\|_4 \leqslant \|f\|_4$), and the power of the hypercontractive inequality is that the right hand side, $\|f\|_2$ is always at most $\|f\|_4$ (and potentially, especially when it is useful, it is significantly smaller).

This innocent looking inequality has several remarkable implications that we discuss next.

1) KKL and Friedgut's Theorem: For a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$, and a variable $i \in [n]$, the influence of a variable $i \in [n]$, denoted by $I_i[f]$, is the probability the ith variable affects the value of $f(x)$. That is, it is the probability that taking $\mathbf{x}$ uniformly from $\{0,1\}^n$ we have $f(\mathbf{x}) \neq f(\mathbf{x} \oplus e_i)$ ($e_i = (0,\ldots,0,1,0,\ldots,0)$, the 1 is on the ith coordinate). The total influence of the function is $I[f] = \sum\limits_{i=1}^{n} I_i[f]$.

It can show without much effort, that if f far from being constant, $I[f] = \Omega(1)$, hence there is always a variable with influence $\Omega(1/n)$. The KKL Theorem strengthens this assertion.

Theorem I.1. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be $\Omega(1)$-far from constant. Then there is a variable $i \in [n]$ with influence at least* $\mathsf{exp}(-O(I[f]))$.

A more commonly known formulation of the KKL-theorem asserts that for any function as Theorem I.1, there is a variable i whose influence is at least $\Omega(\log n/n)$, and we quickly show the how to deduce it. If $I[f] \ll \log n$ a variable guaranteed to exist from Theorem I.1 has large influence, and otherwise $I[f] \geqslant \Omega(\log n)$ and there is a variable whose influence is at least $I[f]/n = \Omega(\log n/n)$).

Friedgut's Junta theorem is a strengthening of the KKL Theorem, stating that not only does a function with small total influence has a variable with large individual influence, but actually the function nearly only depends on such variables.

Theorem I.2. *For every $f\colon \{0,1\}^n \to \{0,1\}$ and $\varepsilon > 0$ there exists $k(I[f],\varepsilon) = \mathsf{exp}(O(I[f]/\varepsilon))$, such that f is ε-close to a k-junta g.*

The KKL Theorem and the Friedgut Junta Theorem are the first applications of discrete Fourier Analysis, and in particular the hypercontractive ineuqality, in TCS.

2) Talagrand's Correlation Theorem: The FKG inequality [29] on the hypercube states that any two monotone Boolean functions $f, g\colon \{0,1\}^n \to \{0,1\}$ are positively correlated, i.e $\mathsf{cov}(f,g) \geqslant 0$. Talagrand [30] proved a stronger statement, that qualitatively says the following: if there are variables that are highly influential in both f and g, then the correlation between f and g is bounded away from 0. More precisely:

Theorem I.3. *Let f, g be monotone functions. Then*

$$\mathsf{cov}(f,g) \geqslant C \cdot \phi\left(\sum_{i=1}^{n} I_i[f] I_i[g]\right),$$

where $\phi(t) = t/\ln(t/e)$.

This proof of this theorem uses the hypercontractive inequality as well.

3) The BKS Noise Sensitivity Theorem: For any function $f\colon \{0,1\}^n \to \{0,1\}$, sampling $\mathbf{x}$ and $\mathbf{y}$ independently, the probability that $f(\mathbf{x}) \neq f(\mathbf{y})$ is equal to

$$2\Pr\left[f(\mathbf{x}) = 0\right] \cdot \Pr\left[f(\mathbf{y}) = 1\right] = 2\mathsf{var}(f).$$

A function is called *noise sensitive*, if sampling a $(1-\varepsilon)$-correlated pair of points x, y, the probability that $f(x) \neq f(y)$ is roughly the same (the formal definition has to consider a sequence of functions: $(f_n)_{n\in\mathbb{N}}$ is called noise sensitive if for any $\varepsilon, \delta > 0$ there is n_0 such that for any $n \geqslant n_0$, we have that $\Pr_{x,y\ (1-\varepsilon)\text{ correlated}}\left[f_n(x) \neq f_n(y)\right] = 2\mathsf{var}(f_n) \pm \delta$). For instance, the function $f(x) = x_1 + \ldots + x_n \pmod 2$ is easily seen to be noise sensitive, while the function $f(x) = x_1$ is not. More generally, the class of linear threshold functions ($1_{\alpha_1 x_1 + \ldots + \alpha_n x_n \geqslant t}$ for real numbers $\alpha_1, \ldots, \alpha_n, t$) are all noise insensitive, as well as functions that are correlated with them.

For a monotone function f, being noise sensitive near the critical probability p_c is a strictly stronger property than having a sharp threshold. To see it is stronger, note that if f noise sensitive and has a constant variance at p, then taking x, y that are $(1-\varepsilon)$-correlated we have that $f(x) \neq f(y)$ with constant probability, and since x, y differ roughly in $\approx \varepsilon pn$ coordinate, it follows that the total influence of f at p is at least $1/(p\varepsilon)$. The Russo-Margulis lemma [20], [21] now implies that the derivative of $\mu_z(f)$ at $z = p$ is at least $1/\varepsilon$. To see it is strictly stronger, note that the majority function $f(x) = 1_{x_1 + \ldots + x_n > n/2}$ has a sharp threshold but is noise insensitive.

Benjamini Kalai and Schramm [25] (who also defined noise sensitivity) proved the following "if and only if" characterization of monotone, noise sensitive functions. For a function f define $II[f] = \sum\limits_{i=1}^{n} I_i[f]^2$.

Theorem I.4 (BKS Theorem)**.** *A monotone function $f\colon \{0,1\}^n \to \{0,1\}$ is noise sensitive if and only if $II[f] = o(1)$. Furthermore, $II[f] = o(1)$ if and only if the correlation of f with any linear threshold function is $o(1)$.*

To prove this theorem, Benjamini Kalai and Schramm extend Talagrand's key tool from [30] that shows that if a monotone function has $\leqslant \delta$ weight on its first Fourier level, then it has $O(\delta \log(1/\delta))$ weight on the second Fourier level.

Noise sensitivity has become an important concept with variety of applications in percolation theory (e.g. [25], [31]), complexity theory (e.g. [32], [6], [4], [33] and learning theory (e.g. [34], [35]).

B. The Biased Hypercube

The vertex set of the biased hypercube is $\{0,1\}^n$, but this time it is equipped with the p-biased measure, defined by $\mu_p^n(x) = p^{|x|}(1-p)^{n-|x|}$, where $p \in (0,1)$ (i.e., to sample $x \sim \mu_p^n$, sample each coordinate as a Bernouli random variable with parameter p). In this paper, we shall think of p as a small function of n, e.g. $p(n) = 1/n^{\alpha}$ for constant $\alpha \in (0,1)$.

All of the results we discussed so far, namely the results of KKL, Friedgut, Talagrand and BKS have generalizations to the p-biased cube, and in some regimes of p – typically when p is bounded away from $0, 1$, are also tight. But when p is close to 0 or 1 – say $p = 1/n^{\alpha}$, these generalizations become very weak (and often time meaningless). The key reason for that is that while the hypercontractive inequality along the lines of (1) holds, it is very weak (specifically, the correlation taken could only be $\rho \approx p^{1/4}$), and therefore to prove stronger results, one has to prove special variants of the hypercontractive inequality and show how to use them. A similar phenomenon occurs also for different (but related) graphs:

1) The Johnson graph is a slice of the hypercube, i.e. its vertices are all x's with Hamming weight k, and two vertices are connected by an edge if their Hamming distance is 2. The Johnson graph should be thought of as the "exact" analog of the p-biased cube for $p = k/n$, and in fact the two graph are morally equivalent (in the sense that theorems proven on one can often be translated, though not automatically, to the other). A stronger hypercontractivity-type result for the Johnson graph was given in [15].
2) The Grassmann graph is the subspace analog of the Johnson graph. In this graph, the "set of coordinates", $[n]$, is replaced by a linear space V of dimension k over $\mathbb{F}_2$, and the vertices are all ℓ-dimensional subspaces of V. ℓ-dimensional subspaces can be thought of as vectors $x \in \{0,1\}^V$ of Hamming weight 2^{ℓ}, providing some analogy to the p-biased cube for $p = 2^{\ell}/2^k$ (2^k should be compared to n, and ℓ should be thought of as slowly growing to infinity).
   The "ordinary" hypercontractive inequality on the Grassmann graph is very weak – and for similar reasons it is weak on the p-biased cube. Proving strong hypercontractive-type results for this graph for a subclass of functions is an important step in the proof of the 2-to-1 Games Conjecture [13], [14].

In this paper, our primary interest is to find extensions of the BKS noise sensitivity theorem to the p-biased measure. Keller and Kindler [36] proved an extension of the BKS Theorem that is often useful for p such that $p = 1/n^{o(1)}$: if f has very small correlation with linear threshold functions (say, polynomially small in n), then it is noise sensitive. Their result, however, has the same drawbacks that stem from the use of the usual hypercontractive inequality, and is useless when p is polynomially small in n.

We give a new noise sensitivity condition (that in general is only sufficient) for functions that have sufficiently nice group of symmetries. The simplest form of our result is concerned with graph properties. In this setting, we think of $n = \binom{N}{2}$, and think of the input $x \in \{0,1\}^n$ as specifying the adjacency matrix of a given graph. A graph property $f: \{0,1\}^n \to \{0,1\}$ is a function that is closed under vertex relabeling, i.e. the action of S_N. To state our result, we define the appropriately normalized BKS parameter: $II[f] = p \sum_{i=1}^{n} I_i[f]^2$.

Theorem I.5. *Let $f: \{0,1\}^n \to \{0,1\}$ be a monotone graph property, and let $p = 1/n^{(k-1)/k}$ for an even number $k \in \mathbb{N}$. Then f is noise sensitive if and only if:*

- *$II[f] = o(1)$, and*
- *any constant width DNF formula ϕ is at most $o(1)$ correlated with f (i.e. $\mathsf{cov}(f, \phi) = o(1)$).*

The condition on the specific form of p may look odd at first, but surprisingly the theorem may fail for different values of p (a counter example is given in Section VI).

We note that this result is very natural: as in the unbiased hypercube, if $II[f] = \Omega(1)$ for a monotone f, then it is noise insensitive. In the p-biased setting however, there are more examples of noise insensitive functions, such as constant width DNF formulas. Thus, the above theorem asserts that for p as in the statement, these are the only reasons for a monotone graph property to be noise insensitive.

The proof of this result follows similar lines of Talagrand's argument [30] as presented by Keller and Kindler [36], but relies on the stronger hypercontractive inequality of [16], and in particular on large deviation bounds similar to the ones proved in [37]. We remark that since this hypercontractive inequality is only meaningful for a subclass of functions, some care has to be taken when choosing which functions to apply it on, and this is where we exploit the fact that we have a group of symmetries.

The general form of our result extends beyond graph properties and this specific value of p (see Theorems IV.14 and IV.16), but in general only provides a sufficient condition for noise sensitivity. Nevertheless, we are able to recover some known noise sensitivity results, as well as prove new ones using our result described in the next section.

C. Main results

DNF correlations: The invariance principle of [7] states that a low-degree, multilinear functions $f(x)$ over the uniform hypercube whose individual influences are small, behave similarly when one plugs into them $\mathbf{x} \in_R \{0,1\}^n$ and $\mathbf{z} \in \mathbb{R}^n$ where each $\mathbf{z}_i$ is a standard, independent Gaussian random variable $N(0,1)$. [2]

Recently, an invariance principle for a class of functions (called global functions) was established in [16] using new hypercontractive-type results. We strengthen this result for the class of monotone functions. Namely, we show that an analogous invariance principle for the p-biased cube holds for functions that are not correlated with constant width DNF formulas (see Theorem III.7 for a precise statement). Using this theorem, we are able to conclude properties of the weight distribution of Boolean functions that are not correlated with DNF formulas (that we use in the proof of our BKS-style results), as well as the following p-biased variants of Bourgain's tail theorem [3].

Theorem I.6. *For every $k \in \mathbb{N}$, $\varepsilon > 0$, there exists $\delta > 0$, such that the following holds. If a monotone function $f: \{0,1\}^n \to \{0,1\}$ has $\mathsf{var}(f) \geqslant \varepsilon$, and at most δ correlated with width k DNF formulas, then*

$$W^{\geqslant k}[f] \geqslant \frac{\mathsf{var}(f)}{2000\sqrt{k}}.$$

Using the same ideas, one can also establish analogous p-biased variants of other classical applications of the invariance principle, such as the Majority is Stablest theorem [7].

Biased version of BKS for G-symmetric functions: Let $G \subseteq S_n$ be a subgroup. A function $f: \{0,1\}^n \to \mathbb{R}$ is called G-invariant if $f(x) = f(\pi(x))$ for every $x \in \{0,1\}^n, \pi \in G$ (here $\pi(x) = (x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)})$). The general form of our noise sensitivity result, Theorem IV.16, gives sufficient conditions for a monotone G-invariant Boolean function to be noise sensitive, in terms of $II[f]$, the correlation of f with constant width DNF functions, and the "bumpiness" of the group G that we explain next.

Bumpiness.: For simplicity we explain this notion in the case of graph properties. The input $x \in \{0,1\}^n$ is viewed as an N-vertex graph (whose adjacency matrix is x, $n = \binom{N}{2}$), and the group G is the action of vertex permutations on the edges, i.e. action of S_N on $[n]$.

Let H be a constant size subgraph with $V(H)$ vertices and $E(H)$ edges. By linearity of expectation, the expected number of copies of H that appear when x is sampled from μ_p^n (which is nothing but the Erdős–Réyni model $G(N,p)$) is given by

$$\binom{N}{|V(H)|} p^{|E(H)|}.$$

Depending on $V(H), E(H)$ and p, this number could be as small as $o(1)$ and as large as $\omega(1)$. We say that we have bumpiness for p if for any constant size H, the expected number of copies of H exhibits a jump: once

[2]The two distributions are clearly very different; for start, one is discrete and the other is continuous. The meaning here is that for every test function, i.e. smooth $\phi: \mathbb{R} \to \mathbb{R}$, $\mathbb{E}_{\mathbf{x}}[\phi(f(\mathbf{x}))] \approx \mathbb{E}_{\mathbf{z}}[\phi(f(\mathbf{z}))]$.

it is asymptotically larger than $b(n) = \Theta(1)$, it must be at least $B(n)$, which grows relatively rapidly with n. For example, for $p = \Theta(1/N) = \Theta(1/\sqrt{n})$ we have a bumpy from $b(n) = \Theta(1)$ to $B(n) = \Theta(\sqrt{n})$: this follows from the fact that the expected number of copies of H is $\Theta_H(1)N^{V(H)-E(H)}$, so once it is $\Omega(1)$ we have $V(H) \geqslant E(H) + 1$ hence it is $\Omega(N) = \Omega(\sqrt{n})$.

Applications: In the k-SAT function, we think of $n = 2^k\binom{N}{k}$ and think of an input $x \in \{0,1\}^n$ as encoding a k-CNF fomrula ϕ_x, where each index specifies a width k clause and negation pattern on the variables. The k-SAT function is the function $f(x)$ that is 1 if and only if ϕ_x is satisfiable.

The k-SAT function is clearly monotone decreasing, so increasing p causes $\Pr_{x\sim\mu_p^n}[f(x) = 1]$ to be smaller. It is known that if $p \gg \frac{1}{n^{(k-1)/k}}$ (at least $\frac{k2^k}{n^{(k-1)/k}}$ will do), then f is close to the constant 0 function, and if $p \ll \frac{1}{n^{(k-1)/k}}$, then the function is close to the 1 function, thus, the interesting regime for f is around $p = \Theta_k(1)/n^{(k-1)/k}$.

Understanding the behavior of the k-SAT function has been a challenging task; just proving that it exhibits a sharp threshold, i.e. that there is c_k such that on $p \geqslant (c_k + \varepsilon)/n^{(k-1)/k}$ the function is close to the 0 function and on $p \leqslant (c_k - \varepsilon)/n^{(k-1)/k}$ the function is close to the 1 function, takes significant amount of work [23]. Benjamini conjectured in fact that more is true, namely that the k-SAT function is noise sensitive on p's in which it is far from being constant. Using our noise sensitivity condition, we reduce this task to proving upper bounds on the total influence of k-SAT:

Corollary I.7. *Let f be the k-SAT function. Then f is noise sensitive on p if and only if $I_p[f] = o(\sqrt{n/p})$.*

We remark that for all monotone functions, $I_p[f] \leqslant \sqrt{n/p}$, so the theorem states that if the total influence of the k-SAT function is asymptotically smaller than the maximum, then it is noise sensitive. For $k = 2$, using results from [38], we are able to prove the required upper bound on the total influence of 2-SAT, thereby proving:

Corollary I.8. *The 2-SAT function is noise sensitive on any p in which it is far from being constant.*

We also recover some of the noise sensitivity results of Lubetzky and Steif [39] for several graph properties, including connectivity, containing a perfect matching and containing a long cycle.

D. General BKS-style result fails

In light of Theorem I.6, it is tempting to wonder if the restriction of the particular form of p is necessary. That is, whether the following speculation is correct: if a monotone graph property $f\colon \{0,1\}^n \to \{0,1\}$ has $o(1)$ correlations with constant width DNF formulas and $II[f] = o(1)$, then f is noise sensitive. If true, that would have been a very clean analog of the BKS Theorem. Unfortunately, in Section VI we show that this is too much to ask for, and that there are noise insensitive f's with $II[f] = o(1)$ and $o(1)$ correlation with constant width DNF formulas.

E. More related works

Schramm and Steif [31] provide an algorithm-based method of proving noise sensitivity. They show that if there is a randomized algorithm A that reads expectedly $o(n)$ variables of x and determines $f(x)$ when $x \sim \mu_p^n$, then f is noise sensitive. This theorem is not an "if an only if" statement (e.g. the k-clique for $k \approx 2\log n$ is known to be noise sensitive, but there is no algorithm reading $o(n)$ bits that can compute it), and sometimes it is not clear whether it applies or not; for $k \geqslant 3$, k-SAT function is also an instance where the Schramm-Steif method does not work. [3]

II. Preliminaries

A. The p-biased Fourier decomposition

For a parameter $p \in (0,1)$, the p-biased measure on $\{0,1\}$ assigns probability p to 1 and probability $1-p$ to 0. Let μ_p^n be the product distribution of μ_p over $\{0,1\}^n$, i.e. $\mu_p^n(x) = p^{|x|}(1-p)^{n-|x|}$. In the paper the parameter p may (and mostly will) depend on n, and will always be at most $1/2$. Note that the mean of each coordinate in the p-biased distribution is p, and its standard deviation is $\sigma \overset{def}{=} \sqrt{p(1-p)}$.

We consider the space of functions real valued functions on $\{0,1\}^n$, equipped with the inner product $\langle f, g\rangle = \mathbb{E}_{\mathbf{x}\sim\mu_p^n}[f(\mathbf{x})g(\mathbf{x})]$. The p-biased Fourier-Walsh basis is $\{\chi_S\}_{S\subseteq[n]}$, where $\chi_S\colon \{0,1\}^n \to \mathbb{R}$ is given by $\chi_S(x) = \prod\limits_{i\in S}\frac{x_i - p}{\sigma}$. The set $\{\chi_S\}_{S\subseteq[n]}$ forms an orthonormal basis, thus any function f can be written as

$$f(x) = \sum_{S\subseteq[n]} \widehat{f}(S)\chi_S(x),$$

where the coefficients are given by $\widehat{f}(S) = \langle f, \chi_S\rangle$.

Fact II.1 (Plancheral/Parseval equality)**.** *For any $f, g\colon \{0,1\}^n \to \mathbb{R}$, $\langle f, g\rangle = \sum\limits_{S\subseteq[n]} \widehat{f}(S)\widehat{g}(S)$.*

Definition II.2. *For $f\colon \{0,1\}^n \to \mathbb{R}$ and $1 \leqslant k \leqslant n$, the level k weight of f is $W^{=k}[f] = \sum\limits_{|S|=k}\widehat{f}(S)^2$.*

Definition II.3. *For $q \geqslant 1$, the q-norm of $f\colon \{0,1\}^n \to \mathbb{R}$ is $\|f\|_q = \left(\mathbb{E}_{\mathbf{x}\sim\mu_p^n}[|f(\mathbf{x})|^q]\right)^{1/q}$.*

[3] If, for p_c there was an algorithm reading $o(n)$ bits of an input $x \in \{0,1\}^n$ (specifying a k-SAT formula) and deciding satisfiability, then if x is unsatisfiable it would find an a unsatisfiable sub-formula with $o(p_c n) = o(N)$ clauses (recall $n = 2^k\binom{N}{k}$ and $p_c = \Theta_k(1/N^{k-1})$). This algorithm would then be able to find $o(N)$ contradictions for random k-SAT formulas any $p \geqslant p_c$ of the same order, and thus a $2^{o(N)}$ resolution proof of unsatisfiability, contradicting the result of [40].

We denote the average of a function f according to μ_p by $\mu_p(f)$, and sometimes omit the subscript p whenever it is clear from the context. For $x \in \{0,1\}^n$ and $S \subseteq [n]$, x_S is the string in $\{0,1\}^S$ corresponding to the S coordinates of x.

1) Derivatives:

Definition II.4. *The derivative of $f\colon \{0,1\}^n \to \mathbb{R}$ with respect to $S \subseteq [n]$ is the function $\partial_S f\colon \{0,1\}^{[n]\setminus S} \to \mathbb{R}$ defined by*

$$\partial_S f(z) = \sum_{a \in \{0,1\}^S} (-1)^{|S|-|a|} f(x_S = a, x_{\overline{S}} = z).$$

Here, $y = (x_S = a, x_{\overline{S}} = z)$ is the point whose in which $y_S = a$ and $y_{\overline{S}} = z$.

The square of the 2-norm of $\partial_S f$ is called the generalized influence of f on S (in case S contains a single variable, this definition coincides with the usually definition of influences). Expanding each term in the definition, one gets the following Fourier expression for the derivative.

Lemma II.5. *For any $f\colon \{0,1\}^n \to \mathbb{R}$ and $S \subseteq [n]$, $\partial_S f(z) = \frac{1}{\sigma^{|S|}} \sum\limits_{T \supseteq S} \widehat{f}(T)\chi_{T\setminus S}(z)$.*

Therefore, by Parseval we have that the S-generalized influence is

$$\frac{1}{\sigma^{2|S|}} \sum_{T \supseteq S} \widehat{f}(T)^2. \tag{2}$$

Sometimes it will be convenient for us to consider the derivative as multiplied by $\sigma^{|S|}$, we denote that by

$$\widetilde{\partial}_S f(z) = \sum_{T \supseteq S} \widehat{f}(T)\chi_{T\setminus S}(z).$$

Thus, we have $\|\widetilde{\partial}_S f\|_2^2 = \sum\limits_{T \supseteq S} \widehat{f}(T)^2$.

2) Influences:

Definition II.6. *For $f\colon \{0,1\}^n \to \mathbb{R}$, $p \in (0,1)$ and $i \in [n]$, the p-biased influence of variable i is $I_i[f] = \|\partial_i f\|_2^2$. The p-biased total influence of f is $I[f] = \sum\limits_{i=1}^{n} I_i[f]$.*

The influences of monotone functions f are closely related to the singleton Fourier coefficients, stated precisely in the following fact.

Fact II.7. *For any monotone $f\colon \{0,1\}^n \to \{0,1\}$ we have:*

1) *for all $i \in [n]$, $\widehat{f}(\{i\}) = \sigma I_i[f]$.*
2) *$I[f] \leqslant \sqrt{\frac{n}{\sigma^2}}$.*

3) The noise operator: Let $\varepsilon > 0$. For any point $x \in \{0,1\}^n$, consider the probability distribution $N_\varepsilon x$ over $y \in \{0,1\}^n$ defined as follows: for every $i \in [n]$ independently, set $\mathbf{y}_i = x_i$ with probability $1-\varepsilon$, and otherwise resample $\mathbf{y}_i \sim \mu_p$. The points (x,y) when $\mathbf{x} \sim \mu_p^n$ and $\mathbf{y} \sim N_\varepsilon x$, are called $(1-\varepsilon)$ *correlated.*

The noise operator $T_{1-\varepsilon}$ acts on a function f on the hypercube by averaging over $N_\varepsilon x$. That is, $T_{1-\varepsilon} f\colon \{0,1\}^n \to \mathbb{R}$ is defined by

$$T_{1-\varepsilon} f(x) = \mathop{\mathbb{E}}_{\mathbf{y} \sim N_\varepsilon x}[f(\mathbf{y})].$$

The operator $T_{1-\varepsilon}$ has the Fourier–Walsh basis $\{\chi_S\}_{S \subseteq [n]}$ as eigenvectors, with eigenvalues $(1-\varepsilon)^{|S|}$ respectively. Since $T_{1-\varepsilon}$ is a linear operator, we have $T_{1-\varepsilon} f(x) = \sum\limits_{S \subseteq [n]} (1-\varepsilon)^{|S|} \widehat{f}(S)\chi_S(x)$.

B. Refined hypercontractivity

In this section we recall the refined hypercontracitivity of [16].

Theorem II.8. *For any $f\colon \{0,1\}^n \to \mathbb{R}$ it holds $\|T_{\frac{1}{2\sqrt{3}}} f\|_4^4 \leqslant \sum\limits_{S \subseteq [n]} \|\partial_S f\|_2^2 \|\widetilde{\partial}_S f\|_2^2$.*

We also have the more general form for higher norms:

Theorem II.9. *For any $f\colon \{0,1\}^n \to \mathbb{R}$ and $m \in \mathbb{N}$, it holds that $\|T_{\frac{1}{4m^2}} f\|_{2m}^{2m} \leqslant \sum\limits_{S \subseteq [n]} \|\partial_S f\|_2^{2m-2} \|\widetilde{\partial}_S f\|_2^2$.*

For our proofs we require a noiseless version of the hypercontractivity above for low-degree functions, which is easily deduced.

Theorem II.10. *Let $f\colon \{0,1\}^n \to \mathbb{R}$ be a function of degree at most d, and $m \in \mathbb{N}$. We have that*

$$\|f\|_{2m}^{2m} \leqslant (2m)^{8md} \sum_{S \subseteq [n]} \|\partial_S f\|_2^{2m-2} \|\widetilde{\partial}_S f\|_2^2.$$

Proof: Let g be the function such that $f(x) = T_{1/4m^2} g(x)$, i.e. $g = \sum\limits_S (4m^2)^{|S|} \widehat{f}(S)\chi_S(x)$. Applying Theorem II.9 on g we get that $\|f\|_{2m}^{2m} = \|T_{\frac{1}{4m^2}} g\|_{2m}^{2m} \leqslant \sum\limits_{S \subseteq [n]} \|\partial_S g\|_2^{2m-2} \|\widetilde{\partial}_S g\|_2^2$. By Parseval,

$$\begin{aligned}\|\widetilde{\partial}_S g\|_2^2 &= \sum_{T \supseteq S} \widehat{g}(T)^2 = \sum_{T \supseteq S} (4m^2)^{2|T|} \widehat{f}(T)^2 \\ &\leqslant (4m^2)^{2d} \sum_{T \supseteq S} \widehat{f}(T)^2 = (2m)^{4d} \|\widetilde{\partial}_S f\|_2^2,\end{aligned}$$

and thus also $\|\partial_S g\|_2^2 \leqslant (2m)^{4d} \|\partial_S f\|_2^2$. Plugging these two bounds finishes the proof. ■

C. A concentration bound

Definition II.11. *A function f is called (r,ε) quasi-regular if for every S of size at most r, $\|\partial_S f\|_2^2 \leqslant \varepsilon$.*

A function f is simply called ε quasi-regular if for every S, $\|\partial_S f\|_2^2 \leqslant \varepsilon$.

Lemma II.12. *Let $f\colon \{0,1\}^n \to \mathbb{R}$ be a degree d, ε-quasi-regular function, and let $t \geqslant e^{20d}$. Then*

$$\Pr_{\mathbf{x} \sim \mu_p^n}\left[|f(\mathbf{x})| \geqslant t\sqrt{\varepsilon}\right] \leqslant e^{-2d \cdot t^{1/10d}}.$$

Proof: Let $2m$ be a parameter to be determined. Taking power $2m$ and using Markov's inequality, we get that $\Pr_{\mathbf{x}\sim\mu_p^n}[|f(\mathbf{x})|\geqslant t\sqrt{\varepsilon}]\leqslant\frac{\|f\|_{2m}^{2m}}{t^{2m}\varepsilon^m}$. By Theorem II.10 and the quasi-regularity of f we have

$$\begin{aligned}\|f\|_{2m}^{2m} &\leqslant (2m)^{8md}\sum_{S\subseteq[n]}\|\partial_S f\|_2^{2m-2}\|\widetilde{\partial}_S f\|_2^2\\ &\leqslant (2m)^{8md}\varepsilon^{m-1}\sum_{S\subseteq[n]}\|\widetilde{\partial}_S f\|_2^2\\ &= (2m)^{8md}\varepsilon^{m-1}\sum_{T}2^{|T|}\widehat{f}(T)^2\\ &\leqslant 2^d(2m)^{8md}\varepsilon^{m-1}\|f\|_2^2\\ &\leqslant m^{10md}\varepsilon^m.\end{aligned}$$

Therefore, $\Pr_{x\sim\mu_p^n}[|f(x)|\geqslant t\sqrt{\varepsilon}]\leqslant m^{10md}/t^{2m}$. Choosing $m=\lceil t^{1/10d}\rceil$ gives the desired bound. ∎

D. Functions with groups of symmetries

Let $G\subseteq S_n$ be a group, and $f\colon\{0,1\}^n\to\mathbb{R}$. Recall that f is G-invariant if for every $\pi\in G$ and $x\in\{0,1\}^n$, we have that $f(\pi x)=f(x)$. Important examples of G's and functions that are invariant under G's action are graph properties: we think of n as $\binom{N}{2}$, where each coordinate corresponds to a potential edge in a graph, and the input $x\in\{0,1\}^n$ specifies which edges exist (those edges with corresponding coordinate 1). In this case we take $G\subseteq S_n$ to be the permutations corresponding to graph isomorphism (i.e. action of S_N on $[n]$), and functions that are invariant under G are called graph properties.

Examples.: typical examples of graph properties are connectedness – $f(x)=1$ if and only if the graph described by x is connected, containing a large cycle and containing a prescribed, fixed subgraph H. In some applications we will also consider hypergraph properties.

Lemma II.13. *Let $f\colon\{0,1\}^n\to\mathbb{R}$ be G invariant. Then for any $S\subseteq[n]$, $\pi\in G$ we have*

1) $\widehat{f}(S)=\widehat{f}(\pi(S))$.
2) $\|\partial_S f\|_2=\|\partial_{\pi(S)}f\|_2$.

Proof: For the first item, $\widehat{f}(\pi(S)) = \mathbb{E}_{\mathbf{x}\sim\mu_p^n}\left[f(\mathbf{x})\chi_{\pi(S)}(\mathbf{x})\right] = \mathbb{E}_{\mathbf{x}\sim\mu_p^n}\left[f(\mathbf{x})\chi_S(\pi^{-1}(\mathbf{x}))\right]$. Since f is G invariant, $f(x)=f(\pi^{-1}(x))$ so letting $y=\pi^{-1}(x)$ and noting that it is distributed according to μ_p^n, we have that

$$\begin{aligned}\widehat{f}(\pi(S)) &= \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_p^n}\left[f(\pi^{-1}(\mathbf{x}))\chi_S(\pi^{-1}(\mathbf{x}))\right]\\ &= \mathop{\mathbb{E}}_{\mathbf{y}\sim\mu_p^n}\left[f(\mathbf{y})\chi_S(\mathbf{y})\right]=\widehat{f}(S).\end{aligned}$$

The second item follows immediately by (2). ∎

Definition II.14. *For a set $S\subseteq[n]$ and a group $G\subseteq S_n$, the orbit of G on S is denoted by $\mathsf{orb}_G(S)$ is defined to be $\{\pi(S)\mid\pi\in G\}$.*

III. DNF CORRELATION RESULTS

In this section we show that the error term arising in the invariance principle we have on the p-biased cube is related to the correlation the function has with constant width DNF formulas. Throughout this section we assume $p\leqslant 1/2$ (analogous results may be shown for $p>1/2$ by considering the dual function $g(x)=1-f(\bar{x})$). We first recall the invariance principle of [16].

A. Invariance Principle

The Fourier expansion of $f\colon\{0,1\}^n\to\mathbb{R}$ over the p-biased measure is given by $f(x)=\sum_{S\subseteq[n]}\widehat{f}(S)\chi_S(x)$, and using it we extend the definition of f to $\mathbb{R}^n$ (since the left hand side makes sense for all $x\in\mathbb{R}^n$).

Let $\mathbf{z}=(\mathbf{z}_1,\dots,\mathbf{z}_n)$ be Gaussian random variables with mean p and standard deviation σ. Define

$$\mathsf{Err}(f)\stackrel{def}{=}\mathsf{var}(f)\sum_{0<|S|}\sigma^{2|S|}\|\partial_S f\|_2^4.$$

Theorem III.1. *Let $\phi\colon\mathbb{R}\to\mathbb{R}$ be smooth with third derivative bounded by B, and let $f\colon\{0,1\}^n\to\mathbb{R}$ be of degree d. Then*

$$\left|\mathop{\mathbb{E}}_{\mathbf{x}}[\phi(f(\mathbf{x}))]-\mathop{\mathbb{E}}_{\mathbf{z}}[\phi(f(\mathbf{z}))]\right|\leqslant B10^{4d+1}\sqrt{\mathsf{Err}(f)}.$$

This is the basic form of the invariance principle for smooth functions ϕ, analogous to the basic invariance principle of [7] (see also [41, Chapter 11]). Therein, using approximation arguments and anti-concentration bounds in Gaussian space, the basic invariance principle is extended for wider class of functions ϕ with only slightly worse error bounds. The same arguments apply also here, so we do not repeat the proofs and only state the generalization that we need for the purpose of this paper.

Theorem III.2. *Define $\zeta\colon\mathbb{R}\to[0,\infty)$ by $\zeta(t)=(t-1_{t\geqslant 1/2})^2$. Then for every $f\colon\{0,1\}^n\to\mathbb{R}$ of degree d we have*

$$\left|\mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_p^n}[\zeta(f(\mathbf{x}))]-\mathop{\mathbb{E}}_{\mathbf{z}\sim N(p,\sigma)}[\zeta(f(\mathbf{z}))]\right|\leqslant e^{O(d)}\cdot\mathsf{Err}(f)^{1/3}.$$

The significance of the above version of the invariance principle, is that the function ζ measures the distance of a given function f from Boolean. Thus, if f has small Fourier weight above level d, then its truncation $f^{\leqslant d}$ is close to Boolean, and so if $\mathsf{Err}(f^{\leqslant d})$ is small, we conclude that $f(\mathbf{z})$ is also close to Boolean.

The Levy distance between two random variables $\mathbf{R},\mathbf{S}$ is denoted by $d_L(\mathbf{R},\mathbf{S})$ and is defined to be the infimum over the set of λ's on which we have $\Pr[\mathbf{S}\leqslant t-\lambda]-\lambda\leqslant\Pr[\mathbf{R}\leqslant t]\leqslant\Pr[\mathbf{S}\leqslant t+\lambda]+\lambda$ for all $t\in\mathbb{R}$.

Theorem III.3. *For every $f\colon\{0,1\}^n\to\mathbb{R}$ of degree d we have $d_L(f(\mathbf{x}),f(\mathbf{z}))\leqslant e^{O(d)}\cdot\mathsf{Err}(f)^{1/8}$.*

B. Corollaries for quasi-random functions

Using the technique of Kindler and O'Donnell [42], one has the following corollary.

Theorem III.4. *For every k, there are $r \in \mathbb{N}, \tau > 0$ such that the following holds. Let $f\colon \{0,1\}^n \to \{0,1\}$ be a function with constant variance. If $\max_{0<|S|\leqslant r} \|\partial_S f^{\leqslant k}\|_2^2 \leqslant \tau$, then*

$$\sum_{|S|\geqslant k} \widehat{f}(S)^2 \geqslant c\frac{\mathsf{var}(f)}{\sqrt{k}}.$$

In the next section we state and prove a qualitatively stronger version of the above result. The proof of this theorem follows the same lines, and hence we omit it. We remark that one may prove in a similar fashion p-biased versions of many classical results, such as the Majority is Stablest Theorem [7].

C. DNF correlations

Let us assume now that the function f is monotone. In this section we show that in Theorem III.4, the quasi-randomness condition could be relaxed to "small correlation with DNFs". The general strategy is to show that assuming the converse (namely that the tail of the function is small, or that its stability is small), we show (by invoking the invariance principle) that the error term is large on the low-degree part of f, i.e. $\mathsf{Err}(f^{<k})$ must be large. This step follows the argument of [42]. In the second step, we show that if $\mathsf{Err}(f^{<k})$ is large, then we may find a DNF formula correlated with f. To simplify notations we write $\mathsf{Err}_k(f) = \mathsf{Err}(f^{<k})$.

1) Step 1: small tail implies large error term:

Proposition III.5. *For every $k \in \mathbb{N}$, $\varepsilon > 0$ there is $\delta > 0$ such that the following holds. If $f\colon \{0,1\}^n \to \{0,1\}$ has $W^{\geqslant k}[f] \leqslant \frac{\mathsf{var}(f)}{2000\sqrt{k}}$, then*

$$\mathsf{Err}_k(f) \geqslant e^{-O(k)}\mathsf{var}(f)^8.$$

Proof: In the proof we define 2 more functions: $g = f^{<k}$, and $h(z)$ which is the rounding of $g(z)$ into Boolean for any $z \in \mathbb{R}^n$. We use [42, Theorem 2.8] on h, stating that for any Boolean function $h\colon \mathbb{R}^n \to \{0,1\}$, $W^{\geqslant k}[h] \geqslant \Omega(\mathsf{var}(h)/\sqrt{k})$, and using invariance we show that unless $\mathsf{Err}_k(f)$ is significant, $W^{\geqslant k}[f]$ would be large as well.

If $\mathsf{Err}_k(f) \geqslant C^{-k}100^{-8}$ for some constant C, we are clearly done, so assume otherwise.

Let $g = f^{<k}$, $\mathbf{z} \in \mathbb{R}^n$ where for each i, $\mathbf{z}_i$ is an independent Gaussian random variables with mean p and variance σ, and $\mathbf{x} \in \{0,1\}^n$ where for each i, $\mathbf{x}_i$ is an independent p-biased bit, and let $\zeta(t) = (t - 1_{t\geqslant 1/2})^2$.

Then:

$$W^{\geqslant k}[f] = \mathop{\mathbb{E}}_{\mathbf{x}}\left[(f(\mathbf{x}) - g(\mathbf{x}))^2\right] \geqslant \mathop{\mathbb{E}}_{\mathbf{x}}[\zeta(g(\mathbf{x}))],$$

since the expectation on the right hand side measures the distance of g from the closest Boolean function, and f is a specific one. By Theorem III.2,

$$\mathop{\mathbb{E}}_{\mathbf{x}}[\zeta(g(\mathbf{x}))] \geqslant \mathop{\mathbb{E}}_{\mathbf{z}}[\zeta(g(\mathbf{z}))] - e^{O(k)} \cdot \mathsf{Err}_k(g)^{1/3}, \quad (3)$$

and note that $\mathsf{Err}_k(g) = \mathsf{Err}_k(f)$. Additionally, defining $h(z) = 1_{g(z)\geqslant 1/2}$ we have

$$\mathop{\mathbb{E}}_{\mathbf{z}}[\zeta(g(\mathbf{z}))] = \mathop{\mathbb{E}}_{\mathbf{z}}\left[(h(\mathbf{z}) - g(\mathbf{z}))^2\right] \geqslant W^{\geqslant k}[h] \geqslant \frac{\mathsf{var}(h)}{10\sqrt{k}}. \quad (4)$$

The second transition holds because g is a degree $< k$ function, and $W^{\geqslant k}[f]$ is minimum the distance of h from a degree $< k$ function, and the last transition we used [42, Theorem 2.8]. Combining (3) and (4) we get

$$W^{\geqslant k}[f] \geqslant \frac{\mathsf{var}(h)}{10\sqrt{k}} - e^{O(k)} \cdot \mathsf{Err}_k(f)^{1/3}, \quad (5)$$

so it suffices to show that the variance of h is at least a multiple of the variance of f minus an error term (i.e. depending on Err). To show that, first note that

$$\Pr_{\mathbf{x}}[g(\mathbf{x}) \geqslant 2/3], \Pr_{\mathbf{x}}[g(\mathbf{x}) \leqslant 1/3] \geqslant \mathsf{var}(f)/2,$$

otherwise if one of them failed, say the first, then we would have that $g(\mathbf{x}) \in [1/3, \infty)$ with probability at least $1 - \mathsf{var}(f)/2$. Since $f(\mathbf{x}) = 0$ with probability at least $\mathsf{var}(f)$, we would have $f(\mathbf{x}) = 0, g(\mathbf{x}) > 1/3$ with probability at least $\mathsf{var}(f)/2$ and thus

$$W^{\geqslant k}[f] = \|f - g\|_2^2 \geqslant \frac{\mathsf{var}(f)}{2}\frac{1}{9},$$

contradiction to the assumption that f has small tail.

Using the invariance principle for Levy distance, Theorem III.3, and the assumption $\mathsf{Err}_k(g) \leqslant C^{-k}100^{-8}$ (recall that $\mathsf{Err}_k(g) = \mathsf{Err}_k(f)$), we get that both $\Pr_{\mathbf{z}}[h(\mathbf{z}) \leqslant 1/3 + 1/100]$ and $\Pr_{\mathbf{z}}[h(\mathbf{z}) \geqslant 2/3 - 1/100]$ are at least $\mathsf{var}(f)/2 - e^{-O(k)}\mathsf{Err}_k(g)^{1/8}$, and therefore

$$\begin{aligned}\mathsf{var}(h) &\geqslant (1/3 - 2/100)^2 \cdot \left(\mathsf{var}(f)/2 - e^{O(k)}\mathsf{Err}_k(f)^{1/8}\right)^2 \\ &\geqslant \frac{1}{196}\mathsf{var}(f) - e^{O(k)}\mathsf{Err}_k(f)^{1/8}.\end{aligned}$$

Plugging this into (5) we get

$$W^{\geqslant k}[f] \geqslant \frac{\mathsf{var}(f)}{1960\sqrt{k}} - e^{O(k)}\mathsf{Err}_k(f)^{1/8},$$

using the assumption of the proposition that upper bounds the left hand side, we conclude that

$$\mathsf{Err}_k(f) \geqslant e^{-O(k)}\mathsf{var}(f)^8.$$

■

2) Step 2: large error term implies correlation with DNF formula: We show that if $\mathsf{Err}_k(f)$ is large, then f is correlated with a DNF formula. We shall use the following simple relation between generalized influences and *boosters* for a function f. For a subset $S \subseteq [n]$ and $a \in \{0,1\}^S$, the function $f_{S\to a}$ is the restriction of f where the bits of S are fixed to a. A restriction (S,a) of f is called a δ-booster, if $\mu(f_{S\to a}) \geqslant \mu(f) + \delta$. We have the following easy relation between generalized influences and restrictions (see Section A for a proof).

Lemma III.6. *Let f be a monotone function, and let S be of size at most k, and assume $p \leqslant 1/2$. Then*

$$\|\partial_S f\|_2^2 \leqslant 8^k(\mu(f_{S\to 1}) - \mu(f)).$$

Stated otherwise, if $\|\partial_S f\|_2^2 \geqslant \delta$, then $(S,\vec{1})$ is a $8^{-k}\delta$ booster for f.

Recall that $\mathsf{Err}_k(f) = \mathsf{var}(f) \sum_{0<|S|\leqslant k} \sigma^{2|S|} \|\partial_S f^{\leqslant k}\|_2^4 = \mathsf{var}(f) \sum_{0<|S|\leqslant k} \|\partial_S f^{\leqslant k}\|_2^2 \|\tilde{\partial}_S f^{\leqslant k}\|_2^2$.

Theorem III.7. *For every $k \in \mathbb{N}$, $\varepsilon > 0$ there exists $\delta > 0$, such that if a monotone $f\colon \{0,1\}^n \to \{0,1\}$ has $\mathsf{Err}_k(f) \geqslant \varepsilon$, then there exists a DNF formula ϕ of width k such that $\mathsf{Cov}[f,\phi] \geqslant \delta$. Furthermore, $\delta = \exp(-\exp(O(k)))\varepsilon^{2^k+4}$, and each one of the clauses of ϕ is a δ-booster for f.*

The rest of this section is devoted to the proof of this result. This step is done by several smaller claims, the first of which easily follows from the fact that $\mathsf{Err}_k(f)$ only contains derivatives of order at most k. Denote $g = f^{<k}$.

Claim III.8. *There exists $0 < j \leqslant k$, $C(k,\varepsilon) = \varepsilon/k > 0$, such that $\sum_{\substack{S\subseteq[n]\\|S|=j}} \sigma^{2|S|}\|\partial_S g\|_2^4 \geqslant C(k,\varepsilon)$.*

Proof: In the definition of $\mathsf{Err}_k(f)$, we see that only generalized derivatives of g of order $\leqslant k$ are considered, so there is one order that contributes at least $\frac{1}{k}$ fraction of $\mathsf{Err}_k(f)$. ∎

Fix $C(k,\varepsilon)$ from the above claim, and define the set $\mathcal{F} = \left\{ S \mid \|\partial_S g\|_2^2 \geqslant C_2(k,\varepsilon) \right\}$, for $C_2(k,\varepsilon) = 2^{-k-1}C(k,\varepsilon)$. We have

$$\begin{aligned} C(k,\varepsilon) &\leqslant \sum_{\substack{S\subseteq[n]\\|S|=j}} \|\tilde{\partial}_S g\|_2^2 \|\partial_S g\|_2^2 \\ &\leqslant 4^{2k}\sigma^{2j}|\mathcal{F}| + C_2(k,\varepsilon) \sum_{\substack{S\subseteq[n]\\|S|=j}} \|\tilde{\partial}_S g\|_2^2, \end{aligned}$$

for the second inequality we used $\|\partial_S g\|_2^2 \leqslant 4^k$, $\|\tilde{\partial}_S g\|_2^2 \leqslant 4^k\sigma^{2j}$ to upper bound the contribution from $S \in \mathcal{F}$. [4] The second sum is easily seen to be equal to

$$\sum_{|T|\leqslant k} \binom{|T|}{j} \widehat{f}(T)^2 \leqslant 2^k,$$

so the second term is at most $C(k,\varepsilon)/2$ and hence $|\mathcal{F}| \geqslant 4^{-2k}\sigma^{-2j} \cdot \frac{1}{2}C(k,\varepsilon) \geqslant C_3(k,\varepsilon)p^{-j}$ (for appropriate constant $C_3(k,\varepsilon) > 0$). Also, since the sum of the degree j generalized influences of g is at most 2^k, we get that $|\mathcal{F}| \leqslant \frac{2^k}{C_2(k,\varepsilon)}p^{-j}$.

Define the DNF formula $J_{\mathcal{F}}(x) = \bigvee_{S\in\mathcal{F}} \bigwedge_{i\in S} x_i$ and the function $h_{\mathcal{F}}(x) = \sum_{S\in\mathcal{F}} 1_{x_S=1}$.

Claim III.9. $\Pr_{\mathbf{x}}[J_{\mathcal{F}}(\mathbf{x}) = 1] \geqslant C_4(k,\varepsilon)$ *for* $C_4(k,\varepsilon) \geqslant e^{-O(k)}\varepsilon^5$.

Proof: We use the second moment method. By Cauchy–Schwarz $\Pr_{\mathbf{x}}[J_{\mathcal{F}}(\mathbf{x}) = 1] \geqslant \frac{\mathbb{E}_{\mathbf{x}}[h_{\mathcal{F}}(\mathbf{x})]^2}{\mathbb{E}_{\mathbf{x}}[h_{\mathcal{F}}(\mathbf{x})^2]}$. The numerator is the square of $p^j|\mathcal{F}| \geqslant C_3(k,\varepsilon)$. Let $\mathcal{F}_A$ be the family of all sets in $\mathcal{F}$ containing A. The denominator is at most

$$\sum_{|A|\leqslant j} \sum_{\substack{S_1,S_2\in\mathcal{F}\\S_1\cap S_2=A}} p^{2j-|A|} \leqslant \sum_{|A|\leqslant j} p^{2j-|A|}|\mathcal{F}_A|^2.$$

Next, we upper bound the size of $\mathcal{F}_A$ for any A of size at most j. The idea is to show an upper bound and a lower bound on the sum of all generalized influences of $S \in \mathcal{F}_A$. On the one hand, since $S \in \mathcal{F}_A$ have large influence, it is at least $C_2(k,\varepsilon)|\mathcal{F}_A|$. On the other hand, we may bound their sum using their Fourier expression as follows:

$$\begin{aligned} \sum_{\substack{S\supseteq A\\|S|=j}} \|\partial_S g\|_2^2 &= \sigma^{-2j} \sum_{T\supseteq A} \binom{|T|-|A|}{j-|A|} \widehat{g}^2(T) \\ &\leqslant 2^k\sigma^{-2j}\|\tilde{\partial}_A g\|_2^2 \leqslant 8^k\sigma^{-2j+2|A|}. \end{aligned}$$

In the last inequality we used $\|\tilde{\partial}_A g\|_2^2 \leqslant 4^k\sigma^{2|A|}$. Hence $|\mathcal{F}_A| \leqslant \frac{8^k}{C_2(k,\varepsilon)}\sigma^{-2j+2|A|}$. Upper bounding one of the $|\mathcal{F}_A|$ factors using the above bound, we see that the denominator is at most

$$\begin{aligned} &\frac{8^k}{C_2(k,\varepsilon)} \sum_{|A|\leqslant j} p^{2j-|A|}\sigma^{-2j+2|A|}|\mathcal{F}_A| \\ &= \frac{8^k}{C_2(k,\varepsilon)} p^j \sum_{|A|\leqslant j} (1-p)^{|A|-j}|\mathcal{F}_A| \\ &\leqslant \frac{16^k}{C_2(k,\varepsilon)} p^j \sum_{|A|\leqslant j} |\mathcal{F}_A|. \end{aligned}$$

In the last inequality we used $p \leqslant 1/2$. In the above sum, each element of $\mathcal{F}$ is counted $2^j \leqslant 2^k$ times, so it is at most $2^k|\mathcal{F}|$, and the denominator is at most

$$\frac{32^k}{C_2(k,\varepsilon)} p^j |\mathcal{F}| \leqslant \frac{64^k}{C_2(k,\varepsilon)^2},$$

[4] To see this, note that by Parseval $\|\partial_S g\|_2^2 \leqslant \|\partial_S f\|_2^2$ and since f is Boolean, $|S| \leqslant j$, the values of $\partial_S f$ are at most 2^j in absolute value, hence $\|\partial_S f\|_2^2 \leqslant 2^{2j} \leqslant 4^k$.

in the last inequality we used the upper bound on the size of $\mathcal{F}$. The claim follows for $C_4(k,\varepsilon) = 64^{-k}C_3(k,\varepsilon)^2C_2(k,\varepsilon)^2$. ■

The above claim asserts that the DNF J has significant probability of being 1, and it is tempting to argue that it is correlated with f. This may indeed be the case, but we do not know how to argue that. The issue is that while for each $S \in \mathcal{F}$, restricting $x_S = 1$ increases the average of f significantly, there could be delicate overlaps between the points that contribute to this increase on different S's. We circumvent this issue by identifying "potential overlaps" and replacing them if it is necessary.

Define a sequence $\varepsilon_0 = C_2(k,\varepsilon)$, and recursively $\varepsilon_{i+1} = 2^{-8k}\varepsilon_i^2$. This sequence should be thought of as: a set S of size $j-i$ would be considered influential if its generalized influence is at least ε_i (note that for $i=0$, this is the original set $\mathcal{F}$ we started with). We consider all sets $T \subseteq [n]$ of size $j-i$ starting with $i=1$ to $i=j-1$. If $\|\partial_T g\|_2^2 \geqslant \varepsilon_i$ we remove from $\mathcal{F}$ all supersets of T and add T to $\mathcal{F}$ instead. Since whenever we remove a set S we add a proper subset of it T, we only increase $\mathbb{E}_{\mathbf{x}}[J_{\mathcal{F}}(\mathbf{x})]$ during this process and in particular in the end it is at least $C_4(k,\varepsilon)$. We choose a uniform-sized subset of $\mathcal{F}$ that contributes at least $\frac{1}{j}$ of the measure of $J_{\mathcal{F}}$, and redefine $\mathcal{F}$ to be that set, call that size s, so that we have $\mathbb{E}_{\mathbf{x}}[J_{\mathcal{F}}(\mathbf{x})] \geqslant C_5(k,\varepsilon)$ (for $C_5(k,\varepsilon) = C_4/k$).

We may assume that $p^s \leqslant 2^{-5k-2}\varepsilon_{j-s}$, otherwise for any $S \in \mathcal{F}$, the function $\bigwedge_{i\in S} x_i$ is δ correlated with f for $\delta = p^s 8^{-k}\varepsilon_{j-s} \geqslant \mathsf{exp}(-\mathsf{exp}(O(k)))\varepsilon^{2^k}$ (using Lemma III.6), and we are done.

Denote $\mu = \mathbb{E}_x[h_{\mathcal{F}}(x)]$, and note that $\mu \geqslant \mathbb{E}_{\mathbf{x}}[J_{\mathcal{F}}(\mathbf{x})] \geqslant C_5$. If $\mu \leqslant 2^{-5k}\varepsilon_{j-s}$ we continue in the argument with $C_6 \stackrel{def}{=} C_5$, and otherwise we remove elements from $\mathcal{F}$ until we have $\mu \leqslant 2^{-5k-1}\varepsilon_{j-s}$, and in the end of the process have $2^{-5k-1}\varepsilon_{j-s} - p^s \leqslant \mu \leqslant 2^{-5k-1}\varepsilon_{j-s}$, implying (by $p^s \leqslant 2^{-5k-2}\varepsilon_{j-s}$) that $\mu \geqslant C_6(k,\varepsilon) \stackrel{def}{=} 2^{-5k-2}\varepsilon_{j-s}$.

Claim III.10. $\mathbb{E}_{\mathbf{x}}\left[h_{\mathcal{F}}(\mathbf{x})^2\right] \leqslant \mu^2 + \mu + 2^{-5k}\mu\varepsilon_{j-s}$

Proof: The proof is similar to the proof of Claim III.9, and we use the notation $\mathcal{F}_A$ therein. We have

$$\mathop{\mathbb{E}}_{\mathbf{x}}\left[h_{\mathcal{F}}(\mathbf{x})^2\right] \leqslant \sum_{|A|\leqslant s}\sum_{\substack{S_1,S_2\in\mathcal{F}\\ S_1\cap S_2=A}} p^{2s-|A|} \leqslant \sum_{|A|\leqslant s} p^{2s-|A|}\left|\mathcal{F}_A\right|^2.$$

From $A=\emptyset$, we get that the contribution is μ^2, and from all A's of size s we get a contribution of μ. For any $0<r<s$ and A of size r, we upper bound $|\mathcal{F}_A|$ (whenever it is non-empty) by considering the sum of generalized influences of S containing A.

$$\begin{aligned}\sum_{\substack{S\supseteq A\\|S|=s}}\|\partial_S g\|_2^2 &= \sigma^{-2s}\sum_{T\supseteq A}\binom{|T|-r}{s-r}\widehat{g}^2(T)\\ &\leqslant 2^k\sigma^{-2s+2|A|}\|\partial_A g\|_2^2 \leqslant 2^k\sigma^{-2s+2|A|}\varepsilon_{j-r},\end{aligned}$$

in the last inequality, we used the fact that as $\mathcal{F}$ contains proper supersets of A, we have $\|\partial_A g\|_2^2 \leqslant \varepsilon_{j-r}$. On the other hand, the sum of generalized influences of S that contain A is at least $\varepsilon_{j-s}|\mathcal{F}_A|$, and we conclude $|\mathcal{F}_A| \leqslant \frac{1}{\varepsilon_{j-s}}2^k\sigma^{-2s+2|A|}\varepsilon_{j-r}$. Since $r<s$ we have by the definition of ε_i that this is at most $2^{-8k}2^k\sigma^{-2s+2|A|}\varepsilon_{j-s} \leqslant 2^{-6k}p^{-s+|A|}\varepsilon_{j-s}$. Thus we get

$$\begin{aligned}\mathop{\mathbb{E}}_{x}\left[h_{\mathcal{F}}(x)^2\right] &\leqslant \mu^2+\mu+2^{-6k}\varepsilon_{j-s}\sum_{0<|A|<s}p^s|\mathcal{F}_A|\\ &\leqslant \mu^2+\mu+2^{-6k}\varepsilon_{j-s}\cdot 2^k|\mathcal{F}|p^s\\ &= \mu^2+\mu+2^{-5k}\mu\varepsilon_{j-s}.\end{aligned}$$

■

Claim III.11. *$\mathsf{cov}(f,J_{\mathcal{F}}) \geqslant C_7(k,\delta)$ for $C_7(k,\varepsilon) = \mathsf{exp}(-\mathsf{exp}(O(k)))\varepsilon^{2^k+4}$.*

Proof: Note that for every x, $f(x)J_{\mathcal{F}}(x) \geqslant f(x)h_{\mathcal{F}}(x) - (h_{\mathcal{F}}(x)^2 - h_{\mathcal{F}}(x))$. Indeed, if $h_{\mathcal{F}}(x)=0,1$ then we have equality, and if $h_{\mathcal{F}}(x)\geqslant 2$ we have that the right hand side is at most $2h_{\mathcal{F}}(x)-h_{\mathcal{F}}(x)^2 \leqslant 0$. Therefore,

$$\mathop{\mathbb{E}}_{\mathbf{x}}[f(\mathbf{x})J_{\mathcal{F}}(\mathbf{x})] \geqslant \mathop{\mathbb{E}}_{\mathbf{x}}[f(\mathbf{x})h_{\mathcal{F}}(\mathbf{x})] + \mathop{\mathbb{E}}_{\mathbf{x}}[h_{\mathcal{F}}(\mathbf{x})] - \mathop{\mathbb{E}}_{\mathbf{x}}\left[h_{\mathcal{F}}(\mathbf{x})^2\right].$$

The first expectation is at least $\mu\cdot(\mu(f)+8^{-k}\varepsilon_{j-s})$ by linearity of expectation (and Lemma III.6, that implies that for every $S\in\mathcal{F}$, $\mu(f_{S\to 1}) \geqslant \mu(f)+8^{-k}\varepsilon_{j-s}$) . The second expectation is, by definition, equal to μ. The third expectation is upper bounded using the last claim. Thus, we get that

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathbf{x}}[f(\mathbf{x})J_{\mathcal{F}}(\mathbf{x})] &\geqslant \mu\cdot(\mu(f)+8^{-k}\varepsilon_{j-s}) - \mu^2 - 2^{-5k}\mu\varepsilon_{j-s}\\ &\geqslant \mu\cdot(\mu(f)+8^{-k}\varepsilon_{j-s}) - 2^{-5k+1}\mu\varepsilon_{j-s}\\ &\geqslant \mu\cdot(\mu(f)+16^{-k}\varepsilon_{j-s}).\end{aligned}$$

The second inequality holds since $\mu \leqslant 2^{-5k}\varepsilon_{j-s}$. Since $\mathbb{E}_{\mathbf{x}}[J_{\mathcal{F}}(\mathbf{x})] \leqslant \mathbb{E}_{\mathbf{x}}[h_{\mathcal{F}}(\mathbf{x})] = \mu$, we get that

$$\begin{aligned}\mathsf{cov}(f,J_{\mathcal{F}}) &\geqslant \mu\cdot(\mu(f)+2^{-3k}\varepsilon_{j-s}) - \mu(f)\mu\\ &= 2^{-3k}\cdot\mu\varepsilon_{j-s} = C_7(k,\varepsilon).\end{aligned}$$

Finally, to get the dependency of $C_7(k,\varepsilon)$ on k,ε, one sees that $\varepsilon_{j-s} \geqslant \mathsf{exp}(-\mathsf{exp}(O(k)))\varepsilon^{2^{k-1}}$, and $\mu \geqslant C_6(k,\varepsilon) \geqslant \mathsf{exp}(-O(k))\min(\varepsilon^{2^{k-1}},\varepsilon^4)$. ■

D. Corollaries

Combining the two steps, one has the following result:

Theorem III.12. *For every $k\in\mathbb{N}$, $\varepsilon>0$ there is $\delta>0$ such that the following holds. If a monotone $f\colon\{0,1\}^n\to\{0,1\}$ with $\mathsf{var}(f)\geqslant\varepsilon$ has $W^{\geqslant k}[f] \leqslant \frac{\mathsf{var}(f)}{2000\sqrt{k}}$, then f is δ correlated with a DNF formula ϕ of width k. Additionally, each term S of ϕ is a δ-booster for f, i.e. $\mu(f_{S\to 1}) \geqslant \mu(f)+\delta$.*

As remarked earlier, one may also prove analogous p-biased results for other classical applications of the invariance principle, such as the Majority is Stablest theorem of [7].

IV. Biased analogs of the BKS Theorem

In this section we prove our results for noise sensitivity.

Definition IV.1. *Let $f\colon \{0,1\}^n \to \{0,1\}$, and let $\varepsilon > 0$. The noise sensitivity of f is*

$$\mathsf{NS}_\varepsilon(f) = \Pr_{\mathbf{x},\mathbf{y}\ \varepsilon\ correlated}[f(\mathbf{x}) \neq f(\mathbf{y})].$$

Recall that a function f is called noise sensitive if for every $\varepsilon > 0$, $\mathsf{NS}_\varepsilon(f) = 2\mathsf{var}(f) + o_n(1)$.

In this section, we present sufficient conditions for functions f invariant under well behaved groups G to be noise sensitive at μ_p^n. First, recall the well-known relation between the noise sensitivity of f and its Fourier coefficients:

Fact IV.2. *For any $f\colon \{0,1\}^n \to \{0,1\}^n$ and $\varepsilon \in (0,1)$, $\mathsf{NS}_\varepsilon(f) = 2\mathsf{var}(f) - 2\sum\limits_{i=1}^{n}(1-\varepsilon)^i W^{=i}[f]$.*

A. BKS analog for some graph properties

In this section, we prove the basic form of our result, and in later sections we generalize it.

Theorem IV.3. *Let $p = \Theta(1/n^{1/2})$. For every $\varepsilon, \delta > 0$. there are $\eta > 0$, $d \in \mathbb{N}$ such that if a monotone graph property $f\colon \{0,1\}^n \to \{0,1\}$ satisfies:*

1) *The correlation of f with any width d DNF formula is at most η, and*
2) *$I[f] \leqslant \eta\sqrt{\frac{n}{\sigma^2}}$,*

then $\mathsf{NS}_\varepsilon(f) \geqslant 2\mathsf{var}(f) - \delta$.

Remark IV.4. *It is easy to check that for any f symmetric under a transitive group, and in particular for graph properties, the conditions $I[f] = o(\sqrt{n/\sigma^2})$ and $II[f] = o(1)$ are equivalent.*

To prove Theorem IV.3, by Fact IV.2 it suffices to prove an upper bound on the weight of f on the low levels, say on $i \leqslant r$, since the contribution from $i > r$ is at most $(1-\varepsilon)^r \leqslant \delta/2$ for large enough $r(\varepsilon, \delta)$.

Lemma IV.5. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a monotone graph property, $p = \Theta(1)/\sqrt{n}$, $k \in \mathbb{N}$ and $\delta > 0$. There is $\eta(k,\delta) > 0$ such that if:*

1) *The correlation of f with any width k DNF formula is at most η,*
2) *$I_p[f] \leqslant \eta\sqrt{\frac{n}{\sigma^2}}$,*

then $W^{=k}[f] \leqslant \delta$.

Using Fact IV.2, one sees that Theorem IV.3 is easily implied by the above lemma. The rest of this section is dedicated for the proof of this lemma.

Define $g(x) = f^{=k}(x)$, i.e. the homogenous degree k part of the Fourier decomposition of f, and let $G \subseteq S_n$ be the action of vertex permutations on $[n]$. Since f is invariant under the action of G, so is g. Since the number of unlabeled graphs with d edges is some constant $C(d)$, we have that the number of different orbits of $S \subseteq [n]$ of size k is bounded by $C(k)$; we partition these orbits into several different classes, and bound the contribution of each one of them.

We shall need the following observation about G: the orbit of S under G may only be of size $\Theta_d(n^{v/2})$ (if S describes a v-vertex graph). Thus, there exist constants $A_d, a_d > 0$, such that if the orbit size exceeds $A_d n^{v/2}$, then it must be at least of size $a_d n^{(v+1)/2}$.

Consider $\mathcal{P} = \left\{S \mid |S| = k, \sigma^{2k}\left|\mathsf{orb}_G(S)\right| \leqslant A_k\right\}$, and denote $W = \sum\limits_{S\in\mathcal{P}} \|\widetilde{\partial}_S g\|_2^2$. Since there are at most $C(k)$ different orbits for characters of size d, there is $S \in \mathcal{P}$ such that $\sum\limits_{T\in\mathsf{orb}_G(S)} \|\widetilde{\partial}_S g\|_2^2 \geqslant \frac{W}{C(k)}$. Since all terms on the right hand side are equal we get that $\left|\mathsf{orb}_G(S)\right| \|\tilde{\partial}_S g\|_2^2 \geqslant \frac{W}{C(k)}$, and therefore

$$\sum_{T\in\mathsf{orb}_G(S)} \|\widetilde{\partial}_S g\|_2^2 \cdot \|\partial_S g\|_2^2 = \left|\mathsf{orb}_G(S)\right| \cdot \|\widetilde{\partial}_S g\|_2^2 \cdot \|\partial_S g\|_2^2$$
$$= \frac{1}{\sigma^{2k}\mathsf{orb}_G(S)}\left(\left|\mathsf{orb}_G(S)\right| \cdot \|\widetilde{\partial}_S g\|_2^2\right)^2 \geqslant \frac{W^2}{C(k)^2 A_k}.$$

By Theorem III.7, if $W > \delta/2$, the above inequality implies that f is η_1-correlated with a width k DNF formula for some $\eta_1(k,\delta) > 0$. Therefore, restricting $\eta \leqslant \eta_1$, we have $W \leqslant \delta/2$. Note that $W \geqslant \sum\limits_{S:\exists T\in\mathcal{P}, T\subseteq S} \widehat{f}(S)^2$, so we may drop these character from g and upper bound the contribution from the rest, i.e. redefine

$$g(x) = \sum_{\substack{|S|=k\\ T\subseteq S\to T\notin\mathcal{P}}} \widehat{f}(S)\chi_S(x),$$

and have that $W^{=k}[f] \leqslant \frac{\delta}{2} + \|g\|_2^2$.

Claim IV.6. *For any $S \neq \emptyset$, we have $\|\partial_S g\|_2^2 \leqslant c_k \frac{1}{\sqrt{n}}$ for some $c_k > 0$.*

Proof: Note that if $S \in \mathcal{P}$, then the generalized influence of g on S is 0. Thus, we may assume that $\sigma^{2k}\left|\mathsf{orb}_G(S)\right| > A_k$, and in particular $\left|\mathsf{orb}_G(S)\right| > A_k p^{-k} \geqslant A_k n^{k/2}$, and by the definition of A_k this implies that the orbit of S is of size at least $a_d n^{(k+1)/2}$. Therefore, $\sigma^{2k}\left|\mathsf{orb}_G(S)\right| \geqslant a_k\sqrt{n}$, and since all generalized influences of g in the orbit of S are equal we get

$$\sum_{T\in\mathsf{orb}_G(S)} \|\widetilde{\partial}_T g\|_2^2 = \sigma^{2k} \sum_{T\in\mathsf{orb}_G(S)} \|\partial_T g\|_2^2$$
$$= \sigma^{2k}\left|\mathsf{orb}_G(S)\right| \|\partial_S g\|_2^2 \geqslant a_k\sqrt{n}\|\partial_S g\|_2^2.$$

On the other hand

$$\sum_{T\in \mathsf{orb}_G(S)} \|\widetilde{\partial}_T g\|_2^2 \leqslant \sum_{T}\sum_{Q\supseteq T} \widehat{g}(Q)^2 = \sum_{Q} 2^{|Q|}\widehat{g}(Q)^2 \leqslant 2^k\|g\|_2^2 \leqslant 2^k.$$

We used the fact that all non-zero Fourier coefficients of g are of size $\leqslant k$ and Parseval. By combining the upper and lower bounds we conclude that there is $c_k > 0$ such that $\|\partial_S g\|_2^2 \leqslant c_k \frac{1}{\sqrt{n}}$. ∎

The next and final part of the proof bounds $\|g\|_2^2$, using a variant of an argument of [36].

Claim IV.7. *$\|g\|_2^2 \leqslant C_k n^{-1/4}\sigma^2 I_p[f]\ln^{10d}\left(\frac{\sqrt{n/\sigma^2}}{I_p[f]}\right)$ for some $C_k > 0$.*

Before proving this claim, we show that it quickly finishes the proof. Indeed, taking into account that $\sigma = \Theta(1)/n^{1/4}$, the bound that we have on $\|g\|_2^2$ is

$$C_k \frac{I_p[f]}{\sqrt{n/\sigma^2}}\ln^{10d}\left(\frac{\sqrt{n/\sigma^2}}{I_p[f]}\right)$$

Thus, we may take small enough $\eta_2(k) > 0$ so that given $I_p[f] \leqslant \eta_2(k)\sqrt{n/\sigma^2}$, the last expression is at most δ. Taking $\eta = \min(\eta_1, \eta_2)$ finishes the proof.

Proof of Claim IV.7: Choose a partition $[n] = \mathbf{I}\cup\mathbf{J}$ randomly, where each $i\in[n]$ it put in $\mathbf{I}$ with probability $1/k$ and otherwise in $\mathbf{J}$. Since each S of size k intersects $\mathbf{I}$ in a single element with probability $k\cdot\frac{1}{k}(1-1/k)^{k-1} \geqslant \frac{1}{e}$, we have that

$$\mathop{\mathbb{E}}_{\mathbf{I},\mathbf{J}}\left[\sum_{S:|S\cap\mathbf{I}|=1}\widehat{g}(S)^2\right] \geqslant \frac{1}{e}\|g\|_2^2. \tag{6}$$

Thus, there are I, J that give value at least $\frac{1}{e}\|g\|_2^2$ for the left hand side. Fix such I and J, and bound the left hand side. Write $x = (y,z)$ where $y\in\{0,1\}^I$, $z\in\{0,1\}^J$ and for each $i\in I$ define

$$g_i(z) = \frac{1}{\sigma}\sum_{S\subseteq J, |S|=k-1}\widehat{g}(S\cup\{i\})\chi_S(z).$$

Then note that

$$\sum_{S:|S\cap I|=1}\widehat{g}(S)^2 = \sigma\sum_{i\in I}\langle f, g_i\chi_i\rangle = \sigma\sum_{i\in I}\langle f\chi_i, g_i\rangle. \tag{7}$$

The following claim bounds each summand in (7) separately.

Claim IV.8. *Let ξ be the bound on the $S\neq\emptyset$ derivatives from Claim IV.6. Then for every $i\in I$, $\langle f\chi_i, g_i\rangle$ is at most*

$$T\sqrt{\xi}\sigma I_i[f] + C_k\sqrt{\xi}e^{-T^{1/10k}}\sqrt{\mathop{\mathbb{E}}_{z}\left[\left|\mathop{\mathbb{E}}_{y}[f(y,z)\chi_i(y)]\right|^2\right]}.$$

Proof:

$$\begin{aligned}\langle f\chi_i, g_i\rangle &= \mathop{\mathbb{E}}_{\substack{\mathbf{y}\sim\mu_p^I\\ \mathbf{z}\sim\mu_p^J}}[f(\mathbf{y},\mathbf{z})\chi_i(\mathbf{y})g_i(\mathbf{z})]\\ &= \mathop{\mathbb{E}}_{\mathbf{z}\sim\mu_p^J}\left[g_i(\mathbf{z})\mathop{\mathbb{E}}_{\mathbf{y}}[f(\mathbf{y},\mathbf{z})\chi_i(\mathbf{y})]\right]\\ &\leqslant \mathop{\mathbb{E}}_{\mathbf{z}\sim\mu_p^J}\left[|g_i(\mathbf{z})|\left|\mathop{\mathbb{E}}_{\mathbf{y}}[f(\mathbf{y},\mathbf{z})\chi_i(\mathbf{y})]\right|\right].\end{aligned}$$

Writing $|g_i(z)| = \int_0^\infty 1_{s\leqslant|g_i(z)|}ds$, we get that

$$\langle f\chi_i, g_i\rangle \leqslant \int_0^\infty \mathop{\mathbb{E}}_{\mathbf{z}\sim\mu_p^J}\left[1_{s\leqslant|g_i(z)|}\cdot\left|\mathop{\mathbb{E}}_{\mathbf{y}}[f(\mathbf{y},\mathbf{z})\chi_i(\mathbf{y})]\right|\right]ds.$$

Recall that by Claim IV.6, all $S\neq\emptyset$ derivatives are ξ pseudo-random for $\xi = c_d\frac{1}{\sqrt{n}}$. We partition the range of integration $[0,\infty)$ into $[0, T\sqrt{\xi}]$, call this integral α and $[T\sqrt{\xi},\infty)$, call this integral β, for $T > e^{20d}$ to be chosen later. For α, since the value of an indicator is at most 1 we get that

$$\begin{aligned}\alpha &\leqslant \int_0^{T\sqrt{\xi}}\mathop{\mathbb{E}}_{\mathbf{z}\sim\mu_p^J}\left[\left|\mathop{\mathbb{E}}_{\mathbf{y}}[f(\mathbf{y},\mathbf{z})\chi_i(\mathbf{y})]\right|\right]ds\\ &= T\sqrt{\xi}\mathop{\mathbb{E}}_{\mathbf{z}\sim\mu_p^J}\left[\left|\mathop{\mathbb{E}}_{\mathbf{y}}[f(\mathbf{y},\mathbf{z})\chi_i(\mathbf{y})]\right|\right]\\ &\leqslant T\sqrt{\xi}\mathop{\mathbb{E}}_{\mathbf{z},\mathbf{y}_{-i}}\left[\left|\mathop{\mathbb{E}}_{\mathbf{y}_i}[f(\mathbf{y},\mathbf{z})\chi_i(\mathbf{y})]\right|\right],\end{aligned}$$

here we denote y_{-i} by all variables of y except coordinate i, and used the triangle inequality. Note that if variable i is not influential on (y_{-i}, z) for f, the contribution to the expectation is 0, and when it is influential the contribution is σ, therefore we get that $\alpha \leqslant T\sqrt{\xi}\sigma I_i[f]$.

For β, we have by Cauchy-Schwarz (on z) that

$$\beta \leqslant \int_{T\sqrt{\xi}}^\infty \sqrt{\mathop{\mathbb{E}}_{\mathbf{z}}\left[1_{s\leqslant|g_i(\mathbf{z})|}\right]}\sqrt{\mathop{\mathbb{E}}_{\mathbf{z}}\left[\left|\mathop{\mathbb{E}}_{\mathbf{y}}[f(\mathbf{y},\mathbf{z})\chi_i(\mathbf{y})]\right|^2\right]}\,ds.$$

The second term does not depend on s, thus it can be pulled outside the integral. The expectation of the first term is the probability $|g_i(\mathbf{z})| \geqslant s$. Making the change of variables $s = x\sqrt{\xi}$, we get that β is at most

$$\sqrt{\xi\mathop{\mathbb{E}}_{z}\left[\left|\mathop{\mathbb{E}}_{\mathbf{y}}[f(\mathbf{y},\mathbf{z})\chi_i(\mathbf{y})]\right|^2\right]}\int_T^\infty\sqrt{\Pr_{\mathbf{z}}\left[|g_i(\mathbf{z})|\geqslant x\sqrt{\xi}\right]}\,dx.$$

For all S, by Parseval $\|\partial_S g_i\|_2^2 \leqslant \|\partial_{S\cup\{i\}}g\|_2^2 \leqslant \sqrt{\xi}$ where the last inequality is by Claim IV.6. Therefore, since $x \geqslant T \geqslant e^{20k}$ we may apply Lemma II.12 to bound the probability inside the integral, i.e.

$$\beta \leqslant \sqrt{\xi}\sqrt{\mathop{\mathbb{E}}_{\mathbf{z}}\left[\left|\mathop{\mathbb{E}}_{\mathbf{y}}[f(\mathbf{y},\mathbf{z})\chi_i(\mathbf{y})]\right|^2\right]}\int_T^\infty e^{-x^{1/10k}}dx.$$

To evaluate the last integral, we change variables $s = x^{1/10k}$ to get that it is equal to

$$10k \int_{T^{1/10k}}^{\infty} s^{10k-1} e^{-s} \, ds \leqslant C_k e^{-T^{1/10k}},$$

since the integrand is exponentially decaying starting from large enough $s(k)$. Hence,

$$\beta \leqslant C_k \sqrt{\xi} e^{-T^{1/10k}} \sqrt{\mathop{\mathbb{E}}_{\mathbf{z}}\left[\left|\mathop{\mathbb{E}}_{\mathbf{y}}[f(\mathbf{y},\mathbf{z})\chi_i(\mathbf{y})]\right|^2\right]}.$$

■

Plugging Claim IV.8 to (7), we get that it is at most

$$T\sqrt{\xi}\sigma^2 \sum_{i\in I} I_i[f] + C_k \sigma \sqrt{\xi} e^{-T^{1/10k}} \sum_{i\in I} \sqrt{\mathop{\mathbb{E}}_{\mathbf{z}}\left[\left|\mathop{\mathbb{E}}_{\mathbf{y}}[f(\mathbf{y},\mathbf{z})\chi_i(\mathbf{y})]\right|^2\right]}.$$

Let us observe that the second sum is upper bounded by $\sqrt{n}$. Indeed, by Cauchy-Schwarz it is at most

$$\sqrt{n}\sqrt{\sum_{i\in I} \mathop{\mathbb{E}}_{\mathbf{z}}\left[\left|\mathop{\mathbb{E}}_{\mathbf{y}}[f(\mathbf{y},\mathbf{z})\chi_i(\mathbf{y})]\right|^2\right]} = \sqrt{n}\sqrt{\mathop{\mathbb{E}}_{\mathbf{z}}\left[\sum_{i\in I}\left|\mathop{\mathbb{E}}_{\mathbf{y}}[f(\mathbf{y},\mathbf{z})\chi_i(\mathbf{y})]\right|^2\right]}.$$

The inner-most expectation is the Fourier coefficient $\widehat{f_{\mathbf{z}}}(\{i\})$, where $f_{\mathbf{z}}$ the function f with z fixed, so the sum of their squares is at most 1. Therefore, (7) $\leqslant T\sqrt{\xi}\sigma^2 I_p[f] + C_k\sqrt{\xi}\sigma\sqrt{n}e^{-T^{1/10k}}$. Choosing $T = \ln^{10k}\left(\frac{\sqrt{n/\sigma^2}}{I[f]}\right)$ (which is large enough by assumption on $I[f]$), we get (7) $\leqslant C_k\sqrt{\xi}\sigma^2 I_p[f]\ln^{10k}\left(\frac{\sqrt{n/\sigma^2}}{I_p[f]}\right)$. Therefore, from (6) (or more precisely the choice of I, J that we have made), and the value of ξ, it follows that $\|g\|_2^2 \leqslant C_k n^{-1/4}\sigma^2 I_p[f]\ln^{10k}\left(\frac{\sqrt{n/\sigma^2}}{I_p[f]}\right)$. ■

B. General graph properties

Theorem IV.9. *Let $k \in \mathbb{N}$ be even, and $p = \Theta(1/n^{(k-1)/k})$. Then for every $\varepsilon, \delta > 0$. there are $\eta > 0$ and $d \in \mathbb{N}$ such that if a monotone graph property $f\colon \{0,1\}^n \to \{0,1\}$ satisfies:*

1) *The correlation of f with any width d DNF formula is at most η.*
2) $I_p[f] \leqslant \eta\sqrt{\frac{n}{\sigma^2}}$.

then $\mathsf{NS}_\varepsilon(f) \geqslant 2\mathsf{var}(f) - \delta$.

The proof of this Theorem goes along the lines of the proof of Theorem IV.3, with small modifications, and to explain them we first re-examine the proof of Theorem IV.3. In the proof we considered generalized derivatives on S's of two types: the collection $\mathcal{P}$ of S's that have $\sigma^{2|S|}\,|\mathsf{orb}_G(S)|$ smaller than some constant A_d, and the rest. The first type is bounded using the assumption that the function is not correlated with DNF formulas. The second type is bounded using refined hypercontractivity, and the crucial property used about the constant A_d, is that once $\sigma^{2|S|}\,|\mathsf{orb}_G(S)| > A_d$ for S of size d, $\sigma^{2|S|}\,|\mathsf{orb}_G(S)|$ has to be in fact much larger than that ($c_d\sqrt{n}$ there). That is, this expression "bumps" from constant value A_d to polynomial value in n. This bump is then used in Claim IV.6 to show that the function g consisting of coefficients are still to be accounted for, is quasi-random, and all of its generalized influences are at most 1 over the "bump".

Finally, to express the bound on $\|g\|_2$ in terms of $I[f]/\sqrt{n/\sigma^2}$, we use the precise value of p (and more specifically, its relation to the bump we had).

To prove Theorem IV.9 one may use the same outline with the following modifications:

1) For p such as in the theorem, for any d there is A_d such that whenever $\sigma^{2|S|}\,|\mathsf{orb}_G(S)| > A_d$, one has $\sigma^{2|S|}\,|\mathsf{orb}_G(S)| > c_d n^{1/k}$. To see this, note that the left hand side is $c_d n^\alpha$ for $\alpha > 0$ of the form $m_1\frac{k-1}{k} + m_2\frac{1}{2}$ for $m_1, m_2 \in \mathbb{Z}$, and using the fact k is even this must be at least $1/k$. Consequently, taking this A_d, the bound on $\mathcal{P}$ stays as is, and the bound in Claim IV.6 changes to $c_d n^{1/k}$.
2) In the end, the bound on $W^{=k}[f]$ we would get is $C_d n^{-1/2k}\sigma^2 I[f]\ln^{10d}\left(\frac{\sqrt{n/\sigma^2}}{I[f]}\right)$, and for our specific p we have $n^{-1/2k}\sigma^2 = \frac{C_d}{\sqrt{n/\sigma^2}}$ and the bound may be rewritten as $C_d\frac{I[f]}{\sqrt{n/\sigma^2}}\ln^{10d}\left(\frac{\sqrt{n/\sigma^2}}{I[f]}\right)$. In particular, provided that$I[f] = o(\sqrt{n/\sigma^2})$ we get that $W^{=k}[f] = o(1)$.

It is now clear that the only property that we used for graph properties, is that in this case the symmetry group $G \subset S_n$ would be bumpy: namely, that if the expected number of occurrences of copies of a subgraph S (which is equal to $p^{|S|}\,|\mathsf{orb}_G(S)| \approx \sigma^{2|S|}\,|\mathsf{orb}_G(S)|$) exceeds some constant A_d, then in fact it should at least some polynomial in n.

Continuing in this fashion, one may state similar statement about hypergraph properties and further about sufficiently nice general group of symmetries. In the following section we state such result. We remark however, that while the previous result, Theorem IV.9, gives a sufficient condition for noise sensitivity that is also necessary, the more general result is does not give a condition that is necessary (but is sufficient).

C. Results for general groups of symmetry

Let $G \subseteq S_n$ be a group, $p(n)$ be a probability parameter, and let $S \subseteq [n]$. We associate with S the parameter $e_p(S) \overset{def}{=} \sigma^{2|S|}\,|\mathsf{orb}_G(S)|$ that roughly counts the expected number of $T \in \mathsf{orb}_G(S)$ to have $x_T = 1$ when $x \sim \mu_p^n$

(since $\sigma^{2|S|} \approx p^{|S|}$ when $p = o(1)$ and $|S| = O(1)$). As discussed in the previous section, the key property we used in Theorems IV.3 and IV.9, is that this expectation is bumpy – namely, for any $d \in \mathbb{N}$ there is a constant A_d, such that if S of size d has $e_p(S) > A_d$, then in fact $e_p(S)$ is significantly larger (and in fact some power of n). We formalize this property below.

Definition IV.10. *Let $p(n) \in (0,1)$, $B\colon \mathbb{N} \to \mathbb{N}$ and $G \subseteq S_n$. The group G is called (p, B)-bumpy if for any $d \in \mathbb{N}$, there are constants $A_d, c_d > 0$ such that if $S \subseteq [n]$ of size d has $e_p(S) > A_d$, then $e_p(S) \geqslant c_d B(n)$.*

Another property we used, is that for any $d \in \mathbb{N}$, the number of different orbits of S of size d only depends on d.

Definition IV.11. *A group $G \subseteq S_n$ is called bounded, if for any $d \in \mathbb{N}$ there is $C_d > 0$ such that the number of different orbits of sets of size d from $[n]$ is at most C_d.*

We can now state our general result.

Theorem IV.12. *Let $G \subseteq S_n$ be a bounded, (p, n^τ)-bumpy group for $\tau > 0$. For every $\varepsilon, \delta > 0$, there are $\eta > 0, d \in \mathbb{N}$ such that if a monotone, G-invariant function $f\colon \{0,1\}^n \to \{0,1\}$ has:*

1) *The correlation of f with any width d DNF formula is at most η.*
2) $I_p[f] \leqslant \eta\sqrt{\frac{n}{\sigma^2}}$.

then $W^{=d}[f] \leqslant C_d \frac{\sigma^2}{n^{\tau/2}} I[f] \ln^{10d}\left(\frac{\sqrt{n/\sigma^2}}{I[f]}\right)$.

For this statement to be meaningful, we need, at the very least to have $\frac{\sigma^2}{n^{\tau/2}} I[f] = o(1)$, but in general it is not sufficient (since the logarithmic term may be large) and also may not be necessary.

Remark IV.13. *In some cases, such as k-uniform hypergraphs properties for constant k, and $p = \Theta(1/n^{(r-1)/r})$ for r divisible by k, the condition $\frac{\sigma^2}{n^{\tau/2}} I_p[f] = o(1)$ is sufficient and necessary.*

The following result gives a sufficient condition for noise sensitivity and follows immediately by the bound of Theorem IV.12.

Theorem IV.14. *Let $G \subseteq S_n$ be a bounded, (p, n^τ)-bumpy group for $\tau > 0$. For every $\varepsilon, \delta > 0$, there are $\eta > 0, d \in \mathbb{N}$ such that if a monotone, G-invariant function $f\colon \{0,1\}^n \to \{0,1\}$ has:*

1) *The correlation of f with any width d DNF formula is at most η.*
2) $I_p[f] \leqslant \frac{n^{\tau/2}}{\sigma^2} \frac{1}{\log^{1/\eta} n}$.

then $\mathsf{NS}_\varepsilon(f) \geqslant 2\mathsf{var}(f) - \delta$.

D. General groups

Some interesting functions f have critical probability not of the form $p = \frac{1}{n^r}$ for a rational number r, but some may have more complicated critical probabilities such as $p = \frac{\log n}{n^r}$ (for example, the connectivity property); or for some function f, we do not know p precisely, but only some bounds on it. In the first case, the "bump" of $e_p(S)$ occur at a different point, $\mathsf{poly}_d(\log n)$ value (instead of constant value A_d), and in the second case, we do not really know where or if the bump occurs.

Our noise sensitivity method extends to some of these cases as well, at least when the correlation with DNF formulas is very small.

Definition IV.15. *We say a group G is $(p, b(n), B(n))$ bumpy, if for every d there is $a_d, A_d > 0$ such that if for some S of size d we have $e_p(S) \geqslant b(n)^{A_d}$, then in fact $e_p(S) \geqslant a_d B(n)$.*

Clearly, every group is $(p, b(n), b(n))$ bumpy (and this is sometimes useful when we have little information about p).

Theorem IV.16. *Let $b, B\colon \mathbb{N} \to \mathbb{N}$ be two increasing functions, and let $G \subseteq S_n$ be a bounded, $(p, b(n), B(n))$ bumpy group. For every $\varepsilon, \delta > 0$, there are $D, d \in \mathbb{N}$ such that if a monotone, G-invariant function $f\colon \{0,1\}^n \to \{0,1\}$ has:*

1) *The correlation of f with any width d DNF formula is at most $\frac{1}{b(n)^D}$.*
2) $I[f] \leqslant \frac{\sqrt{B(n)}}{\sigma^2} \frac{1}{\log^D(n)}$.

then $\mathsf{NS}_\varepsilon(f) \geqslant 2\mathsf{var}(f) - \delta$.

The proof of this theorem follows the proof of Theorem IV.3 closely as well. The difference is that generalized influences of the first type are now $\mathcal{P} = \{\, S \mid e_p(S) \leqslant b(n) \}$. Repeating the argument therein, if $W = \sum_{S \in \mathcal{P}} \|\partial_S f\|_2^2$ is at least $\delta/2$, then $\sum_{S \in \mathcal{P}} \|\partial_S f\|_2^2 \|\widetilde{\partial}_S f\|_2^2 \geqslant c_d W^2 \frac{1}{b(n)} \stackrel{def}{=} \varepsilon$ for some $c_d > 0$. Therefore, using Theorem III.7, we conclude that f is ρ-correlated with DNF formula of width d, where

$$\rho = c'_d \varepsilon^{2^d + 4} = C(d, \delta) \frac{1}{b(n)^{2^d + 4}} > \frac{1}{b(n)^D},$$

for some $D(d, \delta)$, and contradiction to the first assumption of the theorem.

The rest of the proof proceeds exactly as there, and we omit it.

V. Applications

In this section we use Theorems IV.3, IV.9, and IV.14 to prove noise sensitivity for specific interesting functions.

A. k-SAT

Let $n = \binom{N}{k}$. A k-CNF formula ϕ can be described by a bit string $x \in \{0,1\}^n$, where each coordinates corresponds to a potential clause, and its value is 1 if and only if it appears in the formula. Let k-SAT: $\{0,1\}^n \to \{0,1\}$ be the function that gives x the value 1 if and only if the formula described by x is unsatisfiable, meaning every Boolean assignments to its N variables fails to satisfy at least one clause.

Friedgut [23] showed that k-SAT problem has a sharp threshold at $p = \frac{c(k,n)}{n^{(k-1)/k}}$ for some bounded $c(k,n)$ (the value of $c(k,n)$ is not known and it is conjectured to only depend on k). Theorem IV.12 implies a sufficient condition (which is also necessary) for the k-SAT function to be noise sensitive.

Corollary V.1. *Let f be the k-SAT function, and let p be a point where $\mu_p(f)$ is a constant bounded away from 0 and 1. Then f is noise sensitive at p if and only if $I_p[f] = o(\sqrt{n/\sigma^2})$.*

Proof: An argument by Friedgut [23] shows that $f = k$-SAT is not correlated with constant width DNF formulas (for completeness we include a proof in Section A). Thus, if $I_p[f] = o(\sqrt{n/\sigma^2})$ the function would be noise sensitive by Theorem IV.12. If, on the other hand, we had $I_p[f] = \Omega(\sqrt{n/\sigma^2})$, then by Claim A.1 the function has constant weight on level 1 and in particular it is not noise sensitive. ■

Though we suspect that the total influence of k-SAT should be $o(\sqrt{n/\sigma^2})$ for any constant k, we do not know how to show that, except for the case $k = 2$.

2-SAT.: In [38], the 2-SAT function is studied in great details, and in particular using their techniques one may bound the total influence of 2-SAT near its critical probability. Denoting $n = 4\binom{N}{2}$ where N is the number of variables in the 2-SAT formula, the critical probability of 2SAT is known to be $p = 1/(2N)$.

Corollary V.2. *Let f be the 2SAT function. For any $p = \frac{1}{2N}(1 + \frac{\lambda}{N^{1/3}})$ for $\lambda \in [-1,1]$, we have $I_p[f] = o(\sqrt{n/\sigma^2})$. Hence, f is noise sensitive on any such p.*

Proof: We shall use a few results from [38], so we begin with some terminology. Given a 2SAT formula ϕ_x encoded by x, we may construct a directed graph $H_x = (V,E)$ whose vertex set has size $2N$ and corresponds to variables and their negations, and edges correspond to clauses, i.e. if the clause $(u \vee v)$ appears in ϕ_x then we add the implication edges $(\bar{u}, v)$, $(\bar{v}, u)$.

For the graph H_x, one has the following simple facts (for a proof, see for example [38]). We denote the fact "in H_x there is a path from u to v" by $u \rightsquigarrow_x v$.

Proposition V.3 ([38])**.** *For every x, we have:*

1) *ϕ_x is satisfiable if and only if H_x does not contain a cycle that contains a variable and its negation, which we call a contradictory cycle.*
2) *Reversal: for all $u, v \in V$, if $u \rightsquigarrow_x v$, then it has a path from $\bar{v} \rightsquigarrow_x \bar{u}$.*

Since f is symmetric under transitive group, all variables have the same influence and it suffices to bound the influence of the first variable, say (u,v). Since f is monotone,

$$I_1[f] = \mu_p(f_{x_1 \to 1}) - \mu_p(f_{x_1 \to 0}) = \mathop{\mathbb{E}}_{\mathbf{x}}[f(1,\mathbf{x}) - f(0,\mathbf{x})].$$

Therefore, x's that contribute to $I_1[f]$ are x's in which $H_{(0,x)}$ does not contain a contradictory cycle, but $H_{(1,x)}$ does, so we know that in $(1,x)$ we have a negative cycle passing through (u,v), say

$$z \to \ldots \to \bar{z} \to \ldots \to u \to v \to \ldots \to z.$$

Since $\bar{z} \rightsquigarrow_x u$, by reversal we have a path from $\bar{u} \rightsquigarrow_x z$, and then using $z \rightsquigarrow_x u$ we have a path from $\bar{u}$ to u. Similarly, since $v \rightsquigarrow_x \bar{z}$, by reversal $z \rightsquigarrow_x \bar{v}$, so using the path from v to z we have $v \rightsquigarrow_x \bar{v}$. We note that both paths we constructed do not use the edge (u,v).

Thus, whenever (u,v) appears in a negative cycle, there is a path from w to $\bar{w}$ and s to $\bar{s}$ for some $w \in \{u, \bar{u}\}$ and $s \in \{v, \bar{v}\}$. Thus, $I_1[f] \leqslant 4\Pr_{\mathbf{x}}[\bar{u} \rightsquigarrow_{\mathbf{x}} u \wedge v \rightsquigarrow_{\mathbf{x}} \bar{v}]$. Let us take $\lambda = 1$, then applying [38, Theorem 3.2 (ii)], the probability that $H_{\mathbf{x}}$ contains a path from u to $\bar{u}$ and from v to $\bar{v}$ is at most

$$\Pr_{\mathbf{x}}[\bar{u} \rightsquigarrow_{\mathbf{x}} u] \Pr_{\mathbf{x}}[v \rightsquigarrow_{\mathbf{x}} \bar{v}] + O(N^{-2/3}),$$

and applying [38, Theorem 3.1 (ii)] each probability is $O(N^{-1/3})$, so we get that $\Pr_{\mathbf{x}}[\bar{u} \rightsquigarrow_{\mathbf{x}} u \wedge v \rightsquigarrow_{\mathbf{x}} \bar{v}] = O(N^{-2/3})$. Now observe that this probability only decreases when $\lambda < 1$, thus we have this inequality for every $\lambda \in [-1,1]$. In particular, $I_1[f] = O(N^{-2/3})$ for every p in our range and consequently $I_p[f] = O(nN^{-2/3}) = O(n^{2/3}) = o(\sqrt{n/\sigma^2})$ (in the last transition we used $\sigma^2 = \Theta(1/\sqrt{n})$). ■

B. Lubetzky-Steif

Based on Theorem IV.16, we give alternative proofs for some noise sensitivity results obtained in [39].

1) Large minimum degree:

Corollary V.4. *For any constant $k \in \mathbb{N}$, let $f_k \colon \{0,1\}^n \to \{0,1\}$ be the graph property of having minimum degree $\geqslant k$ in the graph. Let $p(n)$ be such that $\mu_p(f_k)$ is constant bounded away from 0 and 1.*

Then f_k is noise sensitive, that is $\mathsf{NS}(f_k) = 2\mathsf{var}(f_k) + o(1)$

Proof: Fix k and consider f_k. Write $n = \binom{N}{2}$, so that N is the number of vertices in the graph f_k operates on. It is well known (see [43]) that $\mu_p(f_k)$ is constant bounded away from $0,1$ only when

$$p = \frac{\ln N + (k-1)\ln\ln N + O(1)}{N},$$

so assume p is of this form. To show noise sensitivity, we use Theorem IV.16 with $b(n) = \log n, B(n) = n^{1/2}$, and to do that we argue that (1) f_k is not δ-correlated with $d =$

$O(1)$ width DNF formulas for any $\delta = 1/\mathsf{polylog}(n)$, and (2) $I[f_k] = O_k(\sqrt{n})$ (3) the symmetry group G of f_k is $(p, b(n), B(n))$.

Plugging this three properties into Theorem IV.16 establishes the result, and next we prove each one of these properties.

- Fix d, and assume for contradiction f_k is δ-correlated with DNF formula of constant width. Since f is monotone, we may assume the DNF formula to be monotone (otherwise we may remove negated variables and not decrease the correlation). Using Proposition A.5, f_k has a δ-booster of size d, say S. By the bound on the width of the formula, S is a set of at most d edges, and hence it involves at most $2d$ vertices of the graph. Call this set of vertices V, and let S_2 denote a clique on V. By monotonicity
$$\mu((f_k)_{S_2 \to \vec{1}}) \geqslant \mu((f_k)_{S \to 1}) \geqslant \mu(f_k) + \delta.$$
However as we argue next, this implies $\delta = O_k(1/N)$. Indeed, note that
$$\mu((f_k)_{S_2 \to \vec{1}}) - \mu(f_k) = \mathop{\mathbb{E}}_{\mathbf{a} \in \{0,1\}^n}\left[f_k(x_{S_2} = \vec{1}, \mathbf{a}_{\overline{S}_2}) - f_k(\mathbf{a})\right], \quad 1)$$
so the difference is at most the probability the graph a describes has minimum degree $< k$, but when the clique on V is inserted it has minimum degree $\geqslant k$. But the only vertices whose degree is different between the two graphs is different are vertices of V, so this probability is at most
$$\Pr_{\mathbf{a}}\left[\exists v \in V \mathsf{deg}(v) < k\right] \leqslant \sum_{v \in V} \Pr_{\mathbf{a}}\left[\mathsf{deg}(v) < k\right].$$
Note that all of these probabilities are equal and are at most the probability a vertex v in $G(N,p)$ has degree $< k$. The latter probability may be upper bounded as
$$\begin{aligned}&\sum_{i=0}^{k-1}\binom{N}{i}p^i(1-p)^{N-i}\\ &\leqslant C_k\sum_{i=0}^{k-1}N^ip^i(1-p)^N \leqslant C_k\sum_{i=0}^{k-1}\ln^i(n)e^{-pN}\\ &\leqslant C_k\sum_{i=0}^{k-1}\ln^i(n)e^{-\ln N-(k-1)\ln\ln n}\\ &\leqslant \frac{C_k}{N}.\end{aligned}$$
- We upper bound the influence variables $i \in [n]$, that by symmetry are all equal. The influence of i is equal to $\mu((f_k)_{\{i\} \to 1}) - \mu(f_k)$, which by the previous argument is $O_k(1/N)$. Therefore, $I_p[f_k] = O_k(n/N) = O_k(N)$.
- For any constant d, if $e_p(S) \geqslant A_d \log^d(n)$, then $e_p(S) \geqslant N = \Theta(n^{1/2})$. This is true, since by expanding the left hand side is $\Theta_d(\log^d(n)N^{-d}N^v)$ where v is the number of vertices in the graph described by S, so since it is significantly larger than $\log^d(n)$, we must have $v \geqslant d+1$.

■

Remark V.5. *As shown in [39], Theorem V.6 implies noise sensitivity for several natural graph properties including connectivity, containing Hamiltonian cycle and containing a perfect matching.*

2) Containing a large cycle: Let $n = \binom{N}{2}$.

Corollary V.6. *Let $\ell \colon \mathbb{N} \to \mathbb{N}$ such that $\ell(n) = \omega(1)$, $\ell(n) = O(N^{1/3})$, and let $f_\ell \colon \{0,1\}^n \to \{0,1\}$ be the graph property of containing a cycle of length $\geqslant \ell(n)$. Let $p(n)$ be such that $\mu_p(f_k)$ is constant bounded away from* 0 *and* 1. *Then f is noise sensitive.*

Proof: It is well know that for $\ell(n) = O(N^{1/3})$, we have $p(n) \leqslant (1 + O(N^{-1/3}))/N$. We apply Theorem IV.16 with $b(n) = 2^{\sqrt{\log n}}$, $B(n) = \sqrt{n}$, and check the conditions (the $(p, b(n), B(n))$ bumpiness is clear):

Assume f is δ-correlated with width d DNF formula (which as before may assume to be monotone since f is monotone). Then by Proposition A.5 f has a δ-booster of size d. I.e., there is a set of edges S of size d, such that fixing them to 1, the probability to contains a cycle of length ℓ increase by at least δ.
On the other hand, observe that since $d \ll \ell$, $\mu(f_{S \to \vec{1}}) - \mu(f)$ is at most the probability a graph $\mathbf{H} \sim G(N,p)$ has a path between a pair of vertices u, v that appear in S (perhaps in different edges). By the union bound the latter probability is at most $d^2 \cdot \Pr_{\mathbf{H} \sim G(N,p)}\left[u \rightsquigarrow_{\mathbf{H}} v\right]$.
Denote by $C(u)$ the connected components of u, and note that $\Pr_{\mathbf{H} \sim G(N,p)}\left[u \rightsquigarrow_{\mathbf{H}} v\right]$ is at most
$$\begin{aligned}&\Pr_{\mathbf{H}}\left[|C(u)| \geqslant N^{2/3}\log N\right]\\ &+ \Pr_{\mathbf{H}}\left[N^{2/3} \leqslant |C(u)| \leqslant N^{2/3}\log N\right] \cdot \frac{N^{2/3}\log N}{N}\\ &+ \sum_{t=0}^{\log N}\Pr_{\mathbf{H}}\left[\frac{N^{2/3}}{2^t} \leqslant |C(u)| \leqslant \frac{N^{2/3}}{2^{t+1}}\right] \cdot \frac{N^{2/3}}{2^{t+1}N}.\end{aligned}$$
We recall some basic facts regarding connected components of $G(N,p)$: first they are all of size $\leqslant N^{2/3}\log N$ except with probability $O(1/N)$ (see [44]), hence we get a bound on the first term. For the second and third, we use the fact that for $t \geqslant 0$, the probability the connected component of u is of size

$\geqslant N^{2/3}/2^t$ is at most $O(2^{t/2}N^{-1/3})$ [5]. Therefore, we get $\Pr_{\mathbf{H}\sim G(N,p)}[u \rightsquigarrow_{\mathbf{H}} v]$ is at most

$$O(1/N) + O(N^{-2/3}\log N) + \sum_{t=0}^{\log N} \frac{N^{-2/3}}{2^{t/2}} = O(N^{-2/3}\log N).$$

so $\delta = O(N^{-2/3}\log N)$.

2) The argument from before in particular shows that $I_i[f] = O(N^{-2/3}\log N)$ for every $i \in [n]$, and hence $I[f] = O(N^{4/3}\log N) = o(n^{3/4-0.01})$.

■

VI. General BKS fails

In this section, we show a counter example to the following speculation. We shall think of an input as a graph over N vertices, and $n = \binom{N}{2}$.

Speculation VI.1. *For every $p = o(1)$, if a graph property $f\colon \{0,1\}^n \to \{0,1\}$ is $o(1)$ correlated with constant width DNF formulas, has $II[f] = o(1)$, then f is noise sensitive.*

The example we consider is as follows. Let $p = \frac{1}{N^{5/8}}$, and consider the random variable $\mathbf{Z}$ defined to be the number of triangles in a $G(N,p)$ graph. It is a straightforward computation to show that $\mathbf{Z}$ has expectation $\mu = \binom{N}{3}p^3 = \Theta(N^{9/8})$ and variance $\sigma^2 = \mu + O(N^4p^5) = \mu + O(N^{7/8})$. Let t be the median of $\mathbf{Z}$, i.e. the least integer t for which $\Pr[\mathbf{Z} \geqslant t] \geqslant \frac{1}{2}$, and set $f(G) = f_\triangle(G) = 1_{\#\triangle(G)\geqslant t}$.

Theorem VI.2. *The function $f = f_\triangle$ is a counter example to the above speculation. More precisely, we have the following properties:*

1) $II[f_\triangle] = o(1)$.
2) *$f_\triangle$ is $o(1)$ correlated with any constant width DNF.*
3) *$f_\triangle$ is noise insensitive, and more precisely $W^{=3}[f_\triangle] \geqslant \Omega(1)$.*

Proof: We shall use the following Gaussian-type behavior of $\mathbf{Z}$, proved in [46]. First, by Chebyshev's inequality we have $\Pr[|\mathbf{Z}-\mu| \geqslant 10\sigma] \leqslant 1/100$, thus $t \in [\mu - 10\sigma, \mu + 10\sigma]$.

Lemma VI.3 ([46, Theorem 4.11])**.** *There are constants c, C such that for all integers $k \in [\mu - 10\sigma, \mu + 10\sigma]$ we have that $\Pr[\mathbf{Z} = k]$ is between c/σ and C/σ.*

From this lemma it is already clear that $\mathbb{E}_{\mathbf{G}}[f(\mathbf{G})] = \frac{1}{2} + o(1)$. Next, we show that $II[f] = o(1)$. Recall that since f is symmetric under transitive group, we have that $II[f] = npI_i[f]^2$. Below we upper bound the probability that i is influential. An edge $i = (u,v)$ is influential if sampling the input $\mathbf{G}$ on all coordinates except i, $\mathbf{G}$ has $< t$ triangles but $\mathbf{G} \cup \{i\}$ has $\geqslant t$ triangles. The expected number of triangles i is adjacent to is $Np^2 = N^{-1/4}$, and by an easy computation the probability it is adjacent to at least 8 is at most $O(N^8p^{16}) = O(1/n)$. Thus, letting E be the event that i is adjacent to at most 8 triangles, we have that

$$I_i[f] \leqslant \Pr_{\mathbf{G}}[\bar{E}] + \Pr_{\mathbf{G}}[i \text{ influential} \mid E].$$

Clearly, conditioned on E, for i to be influential we need it to be adjacent to at least one triangle (that happens with probability $O(Np^2)$) and that $\mathbf{G}$ would contain between $t-8$ and $t-1$ triangles. Therefore,

$$\begin{aligned} I_i[f] &= O(1/n) + O(Np^2)\Pr_{\mathbf{G}}[t-8 \leqslant \mathbf{Z}[\mathbf{G}] \leqslant t-1] \\ &\leqslant O(1/n) + O(Np^2)\Pr_{\mathbf{G}'}[t-8 \leqslant \mathbf{Z}[\mathbf{G}'] \leqslant t-1], \end{aligned}$$

where $\mathbf{G}' \sim G(N,p)$ without the restriction of not containing the edge i (here we use the fact that the probability i is in $\mathbf{G}'$ is $1-p \geqslant 1/2$, so removing this conditioning increases the probability by at most constant factor). Since the probability $\mathbf{G}'$ contains between $t-8$ and $t-1$ triangles is $O(\sigma^{-1})$, we conclude that $I_i[f] \leqslant O(1/n) + O(Np^2\sigma^{-1}) = O(N^{-13/16})$, and therefore $II[f] = npO(N^{-26/16}) = O(N^{-1/4})$, and the first bullet is proved.

To show the second bullet, suppose f has ε correlation with width k DNF where $\varepsilon = \Omega(1), k = O(1)$. Then by Proposition A.5 there exists a set of k edges J, such that $\mu_p(g) \geqslant \mu_p(f) + \varepsilon$ for g which is the function f when the variables of J are set to 1. We may assume without loss of generality J is a clique of size k (by possibly restricting more edges). Note that since the size of J is constant, the probability edges in J participate in any triangles other than the ones appearing in J is at most $JNp^2 = o(1)$, thus we get that the difference $\mu_p(g) - \mu_p(f)$ is (up to $o(1)$) at most the probability that a randomly chosen graph G has between t and $t - J^3$ triangles. This probability is $o(1)$ by Lemma VI.3, and contradiction to the assumption that $\varepsilon = \Omega(1)$. Thus, the second bullet is proved.

Next we prove the third bullet. Let C_3 be potential triangles on an N-vertex graph, then $W^{=3}[f]$ is at least

$$\sum_{S\in C_3} \widehat{f}(S)^2 \geqslant \binom{N}{3}p^3 \left| \sum_{a\in\{0,1\}^S} (-1)^{|S|-|a|}\mu_p(f_{S\to a}) \right|^2 \geqslant \binom{N}{3}p^3\Big(A - 3B\Big)^2,$$

where we choose a specific triangle S (arbitrarily) with edges $i_1 = (u,v), i_2 = (v,w), i_3 = (u,w)$, and

[5] The idea is as in [45, Theorem 7]. Using their notation (except our N is n in their notation), setting $H = N^{1/3}2^{-t/2}$ therein, if $|C(u)| \geqslant N^{2/3}/2^t$ then either $S_\gamma \geqslant H$, or else in $< H$ steps, the connected components of u has gotten to size $N^{2/3}/2^t$. For the probability of the first event, the proof therein gives $\Pr[S_\gamma \geqslant H] \leqslant O(2^{t/2}N^{-1/3})$. For the second event, one may calculate the expected number of new members discovered at each step is at most $O((1+O(N^{-1/3}))^H) = O(1)$, so the expected number of members in $< H$ steps is $O(H)$, and hence by Markov the probability it exceeds $N^{2/3}/2^t$ is at most $O(2^{t/2}N^{-1/3})$.

define $A = \left|\mu_p(f_{S\to(1,1,1)}) - \mu_p(f_{S\to(0,0,0)})\right|$, $B = \left|\mu_p(f_{S\to(1,1,0)}) - \mu_p(f_{S\to(1,0,0)})\right|$. Note that A is at least the probability that sampling a random graph $\mathbf{G} \sim G(N,p)$, $\mathbf{G} \setminus \{i_1,i_2,i_3\}$ has $t-1$ triangles, which is at least

$$\Pr_{\mathbf{G}\sim G(N,p)}[\mathbf{Z}[\mathbf{G}] = t-1] - 3p = \Omega(1/\sigma),$$

where we used Lemma VI.3 and $p = o(\sigma^{-1})$. Therefore $A \geqslant \Omega(1/\sigma)$.

Next, we show that $B = o(1/\sigma)$. Consider the distribution $G_{u,v.w}$ over graphs $G(N,p)$ conditioned on $(i_1,i_2,i_3) \to (1,0,0)$. Using a similar argument to the computation of the influence, the difference $\mu_p(f_{S\to(1,1,0)}) - \mu_p(f_{S\to(1,0,0)})$ is at most the probability i_2 is adjacent $\geqslant 8$ triangles (which is $O(1/n)$), plus the probability i_2 is adjacent to some triangle (which is $O(Np^2) = o(1)$), times the probability that sampling $\mathbf{G} \sim G_{u,v,w}$, $\mathbf{G}\setminus\{i_1\}$ has between $t-8$ and $t-1$ triangles, i.e.

$$B \leqslant O(1/n)+o(1)\Pr_{\mathbf{G}\sim G_{u,v,w}}[t-8 \leqslant \mathbf{Z}[\mathbf{G}\setminus\{i_1\}] \leqslant t-1].$$

We evaluate the last probability. The distribution of $\mathbf{G}\setminus\{i_1\}$ is $\mathbf{G} \sim G(N,p)$ conditioned on i_1,i_2,i_3 being 0. Since the probability of this conditioning is at least $(1-p)^3 = \Omega(1)$, we may drop it and only increase the probability by a constant factor. Thus,

$$\begin{aligned}
&\Pr_{\mathbf{G}\sim G_{u,v,w}}[t-8 \leqslant \mathbf{Z}[\mathbf{G}\setminus\{i_1\}] \leqslant t-1] \\
&\leqslant O(1)\Pr_{\mathbf{G}\sim G(n,p)}[t-8 \leqslant \mathbf{Z}[\mathbf{G}] \leqslant t-1] \\
&= O(\sigma^{-1}),
\end{aligned}$$

where we used Lemma VI.3. Therefore we get

$$B \leqslant O(1/n) + o(1)\cdot O(\sigma^{-1}) = o(\sigma^{-1}).$$

Plugging this in, we get that

$$\begin{aligned}
W^{=3}[f] &\geqslant \binom{N}{3} p^3(\Omega(1/\sigma) - o(1/\sigma))^2 \\
&= \Omega(N^3p^3\sigma^{-2}) = \Omega(1),
\end{aligned}$$

and the third bullet is proved. ■

VII. Open Problems

We list a few open problems that we believe are interesting.

Question 1. *Show that the total influence of the k-SAT problem is $o(\sqrt{n/\sigma^2})$ for any constant k.*

For a predicate $P\colon \{0,1\}^k \to \{0,1\}$, an instance in the constraint satisfaction problem (CSP(P) in short) with variables $X = \{X_1,\dots,X_N\}$ is a set of k-tuples $E \subseteq X^k$, and the instance is called satisfiable if there is a Boolean assignment to the variables, $A\colon X \to \{0,1\}$ such that for every $e \in E$, say $e = (X_{i_1},\dots,X_{i_k})$, we have $P(A(X_{i_1}),\dots,A(X_{i_k})) = 1$.

The P-SAT function receives an input $x \in \{0,1\}^{N^k}$ describing an instance of CSP(P) (each coordinate specifying whether the corresponding k-tuple appears in E, and P-SAT$(x) = 1$ if and only if x is satisfiable.

Question 2. *For what predicates $P\colon \{0,1\}^k \to \{0,1\}$ is the P-SAT function noise sensitive?*

Acknowledgments

We would like to thank Itai Benjamini for raising the question of the noise sensitivity of the 2-SAT function that motivated much of this work. We would also like to thank Gadi Kozma and Avi Wigderson for helpful conversations.

Author D.M was supported by NSF grant CCF-1412958 and Rothschild Fellowship.

References

[1] E. Friedgut, "Boolean functions with low average sensitivity depend on few coordinates," *Combinatorica*, vol. 18, no. 1, pp. 27–35, 1998. [Online]. Available: http://dx.doi.org/10.1007/PL00009809

[2] I. Dinur and S. Safra, "On the hardness of approximating minimum vertex cover," *Ann. of Math. (2)*, vol. 162, no. 1, pp. 439–485, 2005. [Online]. Available: https://doi.org/10.4007/annals.2005.162.439

[3] J. Bourgain, "On the distribution of the Fourier spectrum of Boolean functions," *Israel J. of Math.*, vol. (131), pp. 269–276, 2002.

[4] S. Khot and A. Naor, "Nonembeddability theorems via Fourier analysis," in *FOCS 2005*, 2005, pp. 101–112. [Online]. Available: https://doi.org/10.1109/SFCS.2005.54

[5] ——, "The Andoni-Krauthgamer-Razenshteyn characterization of sketchable norms fails for sketchable metrics," in *SODA 2019*, 2019, pp. 1814–1824. [Online]. Available: https://doi.org/10.1137/1.9781611975482.109

[6] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell, "Optimal inapproximability results for MAX-CUT and other 2-variable CSPs?" *SIAM J. Comput.*, vol. 37, no. 1, pp. 319–357, 2007. [Online]. Available: https://doi.org/10.1137/S0097539705447372

[7] E. Mossel, R. O'Donnell, and K. Oleszkiewicz, "Noise stability of functions with low influences: invariance and optimality," in *FOCS 2005*, 2005, pp. 21–30.

[8] E. Chattopadhyay, P. Hatami, K. Hosseini, and S. Lovett, "Pseudorandom generators from polarizing random walks," in *CCC 2018*, 2018, pp. 1:1–1:21. [Online]. Available: https://doi.org/10.4230/LIPIcs.CCC.2018.1

[9] R. Raz and A. Tal, "Oracle separation of BQP and PH," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 25, p. 107, 2018. [Online]. Available: https://eccc.weizmann.ac.il/report/2018/107

[10] S. Khot, "On the power of unique 2-prover 1-round games," in *CCC 2002*, 2002, p. 25. [Online]. Available: http://dx.doi.org/10.1109/CCC.2002.1004334

[11] S. Khot, D. Minzer, and M. Safra, "On independent sets, 2-to-2 games, and Grassmann graphs," in *STOC 2017*, 2017, pp. 576–589. [Online]. Available: http://doi.acm.org/10.1145/3055399.3055432

[12] I. Dinur, S. Khot, G. Kindler, D. Minzer, and M. Safra, "Towards a proof of the 2-to-1 games conjecture?" in *STOC 2018*, 2018, pp. 376–389. [Online]. Available: http://doi.acm.org/10.1145/3188745.3188804

[13] ——, "On non-optimally expanding sets in grassmann graphs," in *STOC 2018*, 2018, pp. 940–951. [Online]. Available: http://doi.acm.org/10.1145/3188745.3188806

[14] S. Khot, D. Minzer, and M. Safra, "Pseudorandom sets in grassmann graph have near-perfect expansion," in *FOCS 2018*, 2018, pp. 592–601. [Online]. Available: https://doi.org/10.1109/FOCS.2018.00062

[15] S. Khot, D. Minzer, D. Moshkovitz, and M. Safra, "Small set expansion in the Johnson Graph," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 25, p. 78, 2018. [Online]. Available: https://eccc.weizmann.ac.il/report/2018/078

[16] P. Keevash, N. Lifshitz, E. Long, and D. Minzer, "Hypercontractivity for global functions and sharp thresholds," *arXiv preprint arXiv:1906.05568*, 2019.

[17] P. Erdős and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci*, vol. 5, no. 1, pp. 17–60, 1960.

[18] B. Bollobás and A. G. Thomason, "Threshold functions," *Combinatorica*, vol. 7, no. 1, pp. 35–38, 1987.

[19] E. Friedgut and G. Kalai, "Every monotone graph property has a sharp threshold," *Proceedings of the American mathematical Society*, vol. 124, no. 10, pp. 2993–3002, 1996.

[20] L. Russo, "An approximate zero-one law," *Z. Wahrsch. Verw. Gebiete*, vol. 61(1), pp. 129–139, 1982.

[21] G. A. Margulis, "Probabilistic characteristics of graphs with large connectivity," *Problemy peredachi informatsii*, vol. 10, no. 2, pp. 101–108, 1974.

[22] J. Kahn, G. Kalai, and N. Linial, "The influence of variables on Boolean functions," in *FOCS 1988*, 1988, pp. 68–80.

[23] E. Friedgut and J. Bourgain, "Sharp thresholds of graph properties, and the k-SAT problem," *Journal of the American mathematical Society*, vol. 12, no. 4, pp. 1017–1054, 1999.

[24] H. Hatami, "A structure theorem for Boolean functions with small total influences," *Annals of Mathematics*, vol. 176, no. 1, pp. 509–533, 2012.

[25] I. Benjamini, G. Kalai, and O. Schramm, "Noise sensitivity of boolean functions and applications to percolation," *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques*, vol. 90, no. 1, pp. 5–43, 1999.

[26] A. Bonami, "Étude des coefficients de fourier des fonctions de $l^p(g)$," *Annales de l'Institut Fourier*, vol. 20, no. 2, pp. 335–402, 1970.

[27] W. Beckner, "Inequalities in Fourier Analysis," *Annals of Mathematics*, vol. 102, no. 1, pp. 159–182, 1975. [Online]. Available: http://www.jstor.org/stable/1970980

[28] L. Gross, "Logarithmic Sobolev Inequalities," *American Journal of Mathematics*, vol. 97, no. 4, pp. 1061–1083, 1975. [Online]. Available: http://www.jstor.org/stable/2373688

[29] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre, "Correlation inequalities on some partially ordered sets," *Comm. Math. Phys.*, vol. 22, pp. 89–103, 1971.

[30] M. Talagrand, "How much are increasing sets positively correlated?" *Combinatorica*, vol. 16, no. 2, pp. 243–258, Jun 1996. [Online]. Available: https://doi.org/10.1007/BF01844850

[31] O. Schramm and J. E. Steif, "Quantitative noise sensitivity and exceptional times for percolation," *Ann. of Math. (2)*, vol. 171, no. 2, pp. 619–672, 2010. [Online]. Available: https://doi.org/10.4007/annals.2010.171.619

[32] J. Håstad, "Some optimal inapproximability results," *J. ACM*, vol. 48, no. 4, pp. 798–859, July 2001.

[33] R. O'Donnell, "Hardness amplification within NP," in *STOC 2002*, 2002, pp. 751–760. [Online]. Available: https://doi.org/10.1145/509907.510015

[34] N. H. Bshouty, J. C. Jackson, and C. Tamon, "More efficient PAC-learning of DNF with membership queries under the uniform distribution," in *COLT 1999*, 1999, pp. 286–295. [Online]. Available: https://doi.org/10.1145/307400.307472

[35] A. R. Klivans, R. O'Donnell, and R. A. Servedio, "Learning intersections and thresholds of halfspaces," *J. Comput. Syst. Sci.*, vol. 68, no. 4, pp. 808–840, 2004. [Online]. Available: https://doi.org/10.1016/j.jcss.2003.11.002

[36] N. Keller and G. Kindler, "Quantitative relation between noise sensitivity and influences," *Combinatorica*, vol. 33, no. 1, pp. 45–71, 2013. [Online]. Available: https://doi.org/10.1007/s00493-013-2719-2

[37] I. Dinur, E. Friedgut, G. Kindler, and R. O'Donnell, "On the Fourier tails of bounded functions over the discrete cube," in *STOC 2006*, 2006, pp. 437–446. [Online]. Available: https://doi.org/10.1145/1132516.1132580

[38] B. Bollobás, C. Borgs, J. T. Chayes, J. H. Kim, and D. B. Wilson, "The scaling window of the 2-SAT transition," *Random Struct. Algorithms*, vol. 18, no. 3, pp. 201–256, May 2001. [Online]. Available: https://doi.org/10.1002/rsa.1006

[39] E. Lubetzky and J. E. Steif, "Strong noise sensitivity and random graphs," *The Annals of Probability*, vol. 43, no. 6, pp. 3239–3278, 2015.

[40] V. Chvátal and E. Szemerédi, "Many hard examples for resolution," *Journal of the ACM (JACM)*, vol. 35, no. 4, pp. 759–768, 1988.

[41] R. O'Donnell, *Analysis of boolean functions*. Cambridge University Press, 2014.

[42] G. Kindler and R. O'Donnell, "Gaussian Noise Sensitivity and Fourier Tails," in *CCC 2012*, 2012, pp. 137–147. [Online]. Available: https://doi.org/10.1109/CCC.2012.35

[43] B. Bollobás, *Random graphs*, 2nd ed., ser. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2001, vol. 73. [Online]. Available: https://doi.org/10.1017/CBO9780511814068

[44] N. Alon and J. H. Spencer, *The Probabilistic Method*, 4th ed. Wiley Publishing, 2016.

[45] A. Nachmias and Y. Peres, "The critical random graph, with martingales," *Israel Journal of Mathematics*, vol. 176, no. 1, pp. 29–41, 2010.

[46] A. Röllin and N. Ross, "Local limit theorems via Landau-Kolmogorov inequalities," *Bernoulli*, vol. 21, no. 2, pp. 851–880, 2015. [Online]. Available: https://doi.org/10.3150/13-BEJ590

[47] R. Aharoni and N. Linial, "Minimal non-two-colorable hypergraphs and minimal unsatisfiable formulas," *Journal of Combinatorial Theory, Series A*, vol. 43, no. 2, pp. 196–204, 1986.

APPENDIX

A. *Proof of Lemma III.6*

Assume without loss of generality $|S| = k$. By definition,

$$\|\partial_S f\|_2^2 = \mathop{\mathbb{E}}_{\mathbf{x}}\left[\left(\sum_{a\in\{0,1\}^S}(-1)^{k-|a|}f_{S\to a}(\mathbf{x}_{\overline{S}})\right)^2\right]$$
$$= \mathop{\mathbb{E}}_{\mathbf{x}}\left[\left(\sum_{a\in\{0,1\}^S}(-1)^{k-|a|}(f_{S\to a}(\mathbf{x}_{\overline{S}}) - f(\mathbf{x}))\right)^2\right].$$

Using Cauchy-Schwarz, this is at most

$$2^k \mathop{\mathbb{E}}_{\mathbf{x}}\left[\sum_{a\in\{0,1\}^S}(f_{S\to(a,1)}(\mathbf{x}_{\overline{S}}) - f(\mathbf{x}))^2\right].$$

For all a and x, $(f_{S\to a}(x) - f(x))^2 \leqslant f_{S\to 1}(x_{\overline{S}}) - f_{S\to 0}(x_{\overline{S}})$, so we get that $\|\partial_S f\|_2^2 \leqslant 4^k\left(\mu(f_{S\to 1}) - \mu(f_{S\to 0})\right)$ and it suffices to show that $\mu(f_{S\to 1}) - \mu(f_{S\to 0}) \leqslant 2^k(\mu(f_{S\to 1}) - \mu(f))$. Indeed, by conditioning on x_S, we have

$$\mu(f) = \sum_{a\in\{0,1\}^S} p^{|a|}(1-p)^{k-|a|}\mu(f_{S\to a}).$$

Using monotonicity for all $a \neq \vec{0}$, we have $\mu(f_{S\to a}) \leqslant \mu(f_{S\to\vec{1}})$ so we get

$$\mu(f) \leqslant (1-p)^k\mu(f_{S\to 0}) + (1-(1-p)^k)\mu(f_{S\to 1}),$$

and rearranging gives $\mu(f_{S\to 1}) - \mu(f_{S\to 0}) \leqslant (1-p)^{-k}(\mu(f_{S\to 1}) - \mu(f))$. Since $p \leqslant \frac{1}{2}$, we are done.

Claim A.1. *If a monotone function f has $I[f] \geqslant c\sqrt{n/\sigma^2}$, then $W^{=1}[f] \geqslant c^2$.*

Proof: By Fact II.7, $I_i[f] = \sigma\widehat{f}(\{i\})$ and hence by Cauchy-Schwarz $\sigma I[f] = \sum_{i=1}^{n}\widehat{f}(\{i\}) \leqslant \sqrt{n}\sqrt{W^{=1}[f]}$. Using the lower bound on $I[f]$ and rearranging finishes the proof. ∎

Claim A.2. *Let $p \leqslant \frac{1}{2}$. If a monotone function f is ε-correlated with a DNF formula ϕ of width w, then $W^{=D}[f] \geqslant \delta$ for D, δ that only depend on w, ε.*

Proof: By Cauchy-Schwarz,

$$\delta = \mathsf{cov}(f,\phi) = \sum_{S\neq\emptyset}\widehat{f}(S)\widehat{\phi}(S) = \sum_{d=1}^{n}\sum_{|S|=d}\widehat{f}(S)\widehat{\phi}(S)$$
$$\leqslant \sum_{d=1}^{n}\sqrt{W^{=d}[f]W^{=d}[\phi]}.$$

Since ϕ has width w, $pI[\phi] \leqslant 4w$ (see [41, Chapter 8, Exercise 8.26]), and therefore for $k = 16w/\varepsilon^2$ we have $W^{\geqslant k}[\phi] \leqslant \frac{1}{4}\varepsilon^2$. Therefore, by Cauchy-Schwarz

$$\sum_{d=k}^{n}\sqrt{W^{=d}[f]W^{=d}[\phi]} \leqslant \sqrt{W^{\geqslant k}[f]\cdot W^{\geqslant k}[\phi]}$$
$$\leqslant \sqrt{W^{\geqslant k}[\phi]} \leqslant \frac{1}{2}\varepsilon.$$

Plugging this into the first inequality, we conclude that $\sum_{d=1}^{k-1}\sqrt{W^{=d}[f]W^{=d}[\phi]} \geqslant \frac{1}{2}\varepsilon$. Therefore, there is $1 \leqslant D(k) = D(w,\varepsilon) \leqslant k-1$ for which the Dth summand is at least $\varepsilon/(2k)$, implying $W^{=d}[f] \geqslant \varepsilon^2/(2k)^2$. ∎

Claim A.3. *Let f be the k-SAT function, and let p be such that $\mu_p(f)$ is constant bounded away from* 0 *and* 1. *Then the correlation of f with any constant width DNF formula is $o(1)$.*

Proof: We think of the k-SAT problem as a hypergraph, where vertices are variables and their negations and hyperedges correspond to clauses in the formula the input x describes. Write $n = \binom{N}{k}$, and recall that the relevant range of p is $p = \Theta_k(1)/N^{k-1} = \Theta_k(1)/n^{1-1/k}$.

Suppose f is δ-correlated with width d DNF formula ϕ, for constants $\delta > 0, d \in \mathbb{N}$. Since f is monotone, we may assume without loss of generality ϕ is monotone, otherwise we could increase the correlation by flipping negations on variables.

The formula ϕ may contain a clause $\mathsf{AND}_{i\in S}x_i$ that is a certificate for the function f (i.e. small set of clauses that cannot be satisfied together, e.g. for $k = 2$, this could be $(x \vee y) \wedge (\bar{y} \vee x) \wedge (\bar{x} \vee z) \wedge (\bar{z} \vee \bar{x})$). Suppose there is such S, and let $S' \subseteq S$ be a set of clauses that is still a certificate that minimizes the ratio between the number of hyperedges it has and the number of variables it has.

Proposition A.4 ([47]). *The number of hyperedges in S' is greater than the number of vertices in it.*

Proof: Assume towards contradiction this is not the case. Consider the biparatite graph in which the left side consists of the variables of S', the right side consists of the hyperedges of S', and (v, e) is an edge if v or its negation appear in e. Then assumption and minimality of S' implies that Hall's condition holds, and hence the graph has a perfect matching. Setting variables to values $0, 1$ according to whether they appear with their negation in the hyperedge that is matched with them, we see that S' is satisfiable, and in particular it is not a certificate for f. ■

Hence, any certificate S for f contains S' that has more hyperedges than variables, so the total μ_p mass of x's covered by certificates of f is at most

$$c_d \sum_{s=1}^{d-1} N^s p^{s+1} \leqslant 2c_d N p^2 = o(1),$$

and since the expectation of ϕ is constant, we may delete all clauses that are certificates and only decrease correlation by $o(1)$. Then, by proposition A.5 we may choose a clause S of ϕ such that

$$\mu_p(f_{S\to 1}) \geqslant \mu_p(f) + \delta - o(1) \geqslant \mu_p(f) + \frac{\delta}{2}.$$

We may choose a possibly larger S that has only a single satisfying assignment (still of constant size $\leqslant 2^{dk^2}$) , and by monotonicity still have $\mu_p(f_{S\to 1}) \geqslant \mu_p(f) + \delta/2$. Fixing such S, we show that $\mu_p(f_{S\to 1}) = \mu_p(f) + o(1)$, and hence $\delta = o(1)$ and contradiction.

Let V be the set of $r = O(1)$ variables participating in hyperedges of S. Fix this set, and consider the following process that samples a k-SAT instance $\mathbf{x}$:

1) Sample $\mathbf{y}$: hyperedges of size k only among $[N] \setminus V$, each with probability p.
2) Sample $\mathbf{z}$: hyperedges of size $\leqslant k-1$ from $[N]$, each with probability $r^k p$ (these hyperedges correspond to original hyperedges that had some variables from V).

Note that after the first step, by monotonicity the probability that $f(\mathbf{y}, 0) = 1$ is at most $\mu_p(f)$. We claim that after the second step, the probability that $f(\mathbf{y}, \mathbf{z}) = 1$ is at least $\mu_p(f_{S\to 1}) - o(1) \geqslant \mu_p(f) + \delta/4$. To see that, note that choosing $\mathbf{x}_{\bar{S}}$ from μ_p, (a) the probability an hyperedge clause with only vertices in V would appear is $O(p) = o(1)$, (b) hyperedges that only contain vertices from $[N] \setminus V$ have equal probability in this process (c) hyperedges that contain some vertices from $[N] \setminus V$ and some from V have higher probability of appearing in $\mathbf{z}$ (taking into account the forced value of variables in V).

Denoting $q_y = \Pr_{\mathbf{z}}[f(\mathbf{y}, z) = 1]$, the above argument implies that $\mathbb{E}_{\mathbf{y}}[q_{\mathbf{y}}] \geqslant \mu_p(f) + \delta/4$. Note that on the second stage, the expected number of hyperedges added is at most $kN^{k-1} r^k p = O_{k,d}(1)$, so except with probability $\delta/8$, it adds only $C = O_{k,d,\delta}(1)$ hyperedges to $\mathbf{x}$. Thus, adding $\mathbf{z}'$, which is C random hyperedges of size $\leqslant k-1$, instead of the second stage would and letting $q'_y = \Pr_{\mathbf{z}'}[f(y, \mathbf{z}')]$, we have $\mathbb{E}_{\mathbf{y}}[q'_{\mathbf{y}}] \geqslant \mu_p(f) + \delta/8$. Since hyperedges of size 1 are the most restrictive (i.e. increase q'_y the most), we may assume that these hyperedges in $\mathbf{z}'$ are of size 1.

For each y, let $\mathsf{SAT}(y)$ denote the set of satisfying assignments of the formula given by y. The above implies that choosing $\mathbf{y}$ randomly and C dictators at random $\mathbf{D}_1, \ldots, \mathbf{D}_C$ (noting that choosing a hyperedge of size 1 among to forcing the value of a variable, i.e. intersecting with dictatorship), we have that

$$\mathsf{SAT}(y) \cap (\mathbf{D}_1 \cap \ldots \cap \mathbf{D}_C) = \emptyset,$$

with probability at least $\mu_p(f) + \delta/8$.

Fix y. We show that the choice of each $\mathbf{D}_i$ could be replaced by choosing $w(n)$ random k-clauses (unions of k-dictatorships), where we only require $w(n) \to \infty$. We start from replacing the last one, $\mathbf{D}_C$, with $w(n)$ random k-clauses. For $D = (D_1, \ldots, D_{C-1})$, let $\alpha(D)$ be the probability $\mathbf{D}_C$ makes the intersection empty, and consider only D with $\alpha(D) \geqslant \delta/32C$. Then this means that $\mathsf{SAT}(y) \cap (D_1 \cap \ldots \cap D_{C-1})$ is contained in at least $2\alpha(D)N$ dictatorships, call their intersection W. Choosing a k-clause randomly, the probability it intersects W in $\emptyset$ is at least

$$\left(\frac{2\alpha(D)N}{2N}\right)^k = \alpha(D)^k \geqslant (\delta/32C)^k,$$

so choosing $w(n)$ k-clauses independently, the probability their intersection intersects W in $\emptyset$ is $1 - o(1)$. Therefore, we see that choosing $(\mathbf{D}_1, \ldots, \mathbf{D}_{C-1}, \mathbf{T}_C)$ where $\mathbf{T}_C$ is the intersection of $w(n)$ random k-clauses,

$$\mathsf{SAT}(y) \cap (\mathbf{D}_1 \cap \ldots \cap \mathbf{D}_{C-1} \cap \mathbf{T}_C) = \emptyset$$

with probability $\geqslant q'_y - \delta/32C - o(1)$ (the $-\delta/32C$ is accounts for contribution from D's with $\alpha(D) \leqslant \delta/32C$ that we ignored). We may now interchange the order in which we sample things, i.e. first sample $\mathbf{T}_C$ and then $\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_{C-1}$, and then apply the argument on $\mathbf{D}_{C-1}$ as before.

Hence, we conclude that choosing $\mathbf{T}_1, \ldots, \mathbf{T}_C$, each one being an intersection of $w(n)$ independently chosen k-clauses, we have

$$\mathsf{SAT}(y) \cap (\mathbf{T}_1 \cap \ldots \cap \mathbf{T}_C) = \emptyset$$

with probability at least $q'_y - C \cdot \delta/32C - C \cdot o(1) \geqslant q'_y - \delta/16$.

From this argument it follows, that for every y, choosing $\mathbf{T}$, a collection of $Cw(n)$ random k-clauses, we have $\Pr_{\mathbf{T}}[f(y \vee e_{\mathbf{T}}) = 1] \geqslant q'_y - \delta/16$, and hence

$$\Pr_{\mathbf{y},\mathbf{T}}[f(\mathbf{y} \vee e_{\mathbf{T}}) = 1] \geqslant \mathop{\mathbb{E}}_{\mathbf{y}}[q'_{\mathbf{y}}] - \delta/16 \geqslant \mu_p(f) + \delta/16.$$

Finally, we reach a contradiction by showing that this is false. Let $q = 2Cw(n)/n$, and note that we may replace by choice of "exactly $Cw(n)$ k-clauses" by "sampling each k-clause with probability q"since the probability we would get $\leqslant Cw(n)$ clauses this way is $o(1)$, and have

$$\Pr_{\mathbf{y}\sim\mu_p^n,\mathbf{z}\sim\mu_q^n}[f(\mathbf{y}\vee\mathbf{z})=1] \geqslant \mu_p(f)+\delta/16-o(1) \geqslant \mu_p(f)+\delta/32.$$

Note that in $\mathbf{y}\vee\mathbf{z}$, the bits are independent and each one is 1 with probability at most $p+q$, and hence by monotonicity the last inequality implies $\mu_{p+q}(f) \geqslant \mu_p(f) + \delta/32$. However by the mean-value theorem and the Russo-Margulis Lemma [20], [21] (see also [41]) we have $\mu_{p+q}(f)-\mu_p(f) = qI_{p'}[f]$ for some $p' \in (p, p+q)$. Since f is monotone, by Fact II.7 we have $I_{p'} \leqslant \sqrt{n/p(1-p)}$, and recalling that $p = O_k(1)/n^{1-1/k}$ we get

$$\mu_{p+q}(f)-\mu_p(f) \leqslant O_k(1)\frac{w(n)}{n}\cdot\sqrt{n^{2-1/k}} = O_k(1)\frac{w(n)}{n^{1/2k}},$$

which tends to 0 by choosing, say, $w(n) = \log n$, and contradiction. ■

Proposition A.5. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a monotone function, and let $\phi\colon \{0,1\}^n \to \{0,1\}$ be a monotone DNF formula. If f, ϕ are δ correlated, then at least one of the clauses of ϕ is a δ-booster of f.*

Proof: Assume this is not the case, let $S_1, \dots, S_r$ be the clauses of ϕ, and let ϕ_i be the disjunction of clauses $S_1, \dots, S_i$. By telescoping we may write $\mathbb{E}_{\mathbf{x}}[f(\mathbf{x})\phi(\mathbf{x})] = \sum_{i=0}^{r-1}\mathbb{E}_{\mathbf{x}}[f(\mathbf{x})(\phi_{i+1}(\mathbf{x})-\phi_i(\mathbf{x}))]$. The inner expectation is

$$\Pr_{\mathbf{x}}[\phi_{i+1}(\mathbf{x})=1,\phi_i(\mathbf{x})=0]\cdot\mathbb{E}_{\mathbf{x}}\left[f(\mathbf{x})\mid \mathbf{x}_{S_{i+1}}=1,\phi_i(\mathbf{x})=0\right].$$

Let C be the subcube $x_{S_{i+1}} = \vec{1}$, and consider the μ_p measure on it. The conditional expectation above is $\mathbb{E}_{\mathbf{x}\in C}[f(\mathbf{x})\mid\phi_i(\mathbf{x})=0]$. Since f, ϕ_i are increasing in C, it follows by the FKG inequality [29] that

$$\begin{aligned}\mathbb{E}_{\mathbf{x}\in C}[f(\mathbf{x})\mid\phi_i(\mathbf{x})=0] &\leqslant \mathbb{E}_{\mathbf{x}\in C}[f(\mathbf{x})] = \mu(f_{S_{i+1}\to\vec{1}})\\ &< \mu_p(f)+\delta.\end{aligned}$$

Therefore, $\mathbb{E}_{\mathbf{x}}[f(\mathbf{x})\phi(\mathbf{x})] < \sum_{i=0}^{r-1}\Pr_{\mathbf{x}}[\phi_{i+1}(\mathbf{x})=1,\phi_i(\mathbf{x})=0]\cdot(\mu_p(f)+\delta) = \mu_p(\phi)(\mu_p(f)+\delta)$, and in particular $\mathsf{cov}(f,\phi) < \delta$ and contradiction. ■