Isoperimetric Inequalities Made Simpler

Ronen Eldan * Guy Kindler † Noam Lifshitz ‡ Dor Minzer §

Abstract

We give an alternative, simple method to prove isoperimetric inequalities over the hypercube. In particular, we show:

1. An elementary proof of classical isoperimetric inequalities of Talagrand, as well as a stronger isoperimetric result conjectured by Talagrand and recently proved by Eldan and Gross.
2. A strengthening of the Friedgut junta theorem, asserting that if the p-moment of the sensitivity of a function is constant for some $p \geqslant 1/2+\varepsilon$, then the function is close to a junta. In this language, Friedgut's theorem is the special case that $p = 1$.

1 Introduction

1.1 Isoperimetric inequalities over the hypercube

Isoperimetric inequalities are fundamental results in mathematics with a myriad of applications. The main topic of this paper is discrete isoperimetric inequalities, and more specifically isoperimetric inequalities over the hypercube $\{0,1\}^n$. Classical results along these lines are due to Margulis [Mar74], Talagrand [Tal93, Tal97, Tal94] and the KKL Theorem [KKL88].

All of these theorems discuss various measures of boundaries and relations between them for Boolean functions. For a function $f\colon \{0,1\}^n \to \{0,1\}$ and a point $x \in \{0,1\}^n$, we define the sensitivity of f at x, denoted by $s_f(x)$, to be the number of coordinates $i \in [n]$ such that $f(x \oplus e_i) \neq f(x)$. Thus, the vertex boundary V_f of a function f is defined to be the set of all x's such that $s_f(x) > 0$, and the edge boundary E_f of f is defined to be the set of edges $(x, x \oplus e_i)$ hypercube whose endpoints are assigned different values by f. The most basic isoperimetric result related to Boolean functions, known as Poincare's inequality, asserts that $\frac{|E_f|}{2^n} \geqslant \mathsf{var}(f)$, i.e. the edge boundary of a function f cannot be too small. Margulis [Mar74] improved this result, showing that for all functions f, one has that $\frac{|V_f|}{2^n} \cdot \frac{|E_f|}{2^n} \geqslant \Omega(\mathsf{var}(f)^2)$.

Talagrand considered a different notion of boundary for a function f, namely the quantity $\mathbb{E}_x\left[\sqrt{s_f(x)}\right]$, which we will refer to as the Talagrand boundary of f. Talagrand proved a number of results about this quantity [Tal93, Tal97], the most basic of which asserts that $\mathbb{E}_x\left[\sqrt{s_f(x)}\right] \geqslant \Omega(\mathsf{var}(f))$, which strengthens Margulis' theorem (via a simple application of the Cauchy-Schwarz inequality). Talagrand then went on and strengthened the result for function with small variance, proving that

$$\mathop{\mathbb{E}}_x\left[\sqrt{s_f(x)}\right] \geqslant \Omega(\mathsf{var}(f)\sqrt{\log(1/\mathsf{var}(f))})$$

*Department of Computer Science and Applied Mathematics, Weizmann Institute.
†School of Computer Science and Engineering, Hebrew University.
‡School of Mathematics, Hebrew University.
§Department of Mathematics, Massachusetts Institute of Technology. Supported by a Sloan Research Fellowship.

for all Boolean functions. While this inequality can be seen to be tight (as can be seen by considering sub-cubes), it leaves something to be desires since it is incomparable to yet another prominent isoperimetric result – the KKL theorem [KKL88].

The KKL Theorem is concerned with yet another isoperimetric quantity related to the function f, which are known as the influences. The influence of a coordinate $i \in [n]$, denoted by $I_i[f]$, is equal to number of edges in E_f along direction i divided by 2^n, i.e.

$$I_i[f] = \Pr_{x \in \{0,1\}^n} [f(x) \neq f(x \oplus e_i)].$$

The KKL theorem [KKL88] asserts that any function must have a variable that is at least somewhat influential; more quantitatively, they proved that $\max_i I_i[f] \geqslant \Omega\left(\frac{\log n}{n} \mathsf{var}(f)\right)$.

Since the isoperimetric result of Talagrand and of KKL look very different (as well as the techniques that go into their proofs), Talagrand asked whether one can establish a single isoperimetric inequality that is strong enough to capture both of these results simultaneously. In [Tal97], Talagrand conjectures such statement, and makes some progress towards it. Namely, he shows that there exists $\alpha \in (0, 1/2]$ such that for any function $f\colon \{0,1\}^n \to \{0,1\}$ it holds that

$$\mathop{\mathbb{E}}_x \left[\sqrt{s_f(x)}\right] \geqslant c\mathsf{var}(f) \left(\log\left(\frac{1}{\mathsf{var}(f)}\right)\right)^{1/2-\alpha} \left(\log\left(1 + \frac{1}{\sum_{i=1}^{n} I_i[f]^2}\right)\right)^{\alpha}, \tag{1}$$

for some absolute constant $c > 0$. This inequality can be seen to imply a weaker result along the lines of the KKL theorem [1], but not enough to fully recover it. Talagrand conjectured that the above inequality holds for $\alpha = 1/2$, a statement that is strong enough to directly imply the KKL theorem.

Recently, Eldan and Gross [EG20] introduced techniques from stochastic processes to the field, and used to prove that Talagrand's conjecture is true, i.e. that one may take $\alpha = 1/2$ in (1).

1.2 Our results

We give an elementary proof method that can be used to establish the results mentioned above, as well as strengthen Friedgut's theorem. Our results are:

1. **Classical isoperimetric inequalities.** We give a elementary proof of the inequalities above by Margulis, Talagrand and Eldan-Gross. Our proof technique also unravels a connection between these isoperimetric inequalities and other classical results in the area, such as the Friedgut junta theorem [Fri98] and Bourgain's tail theorem [Bou02]

2. **Junta theorems.** We establish a strengthening of Friedgut's theorem in which the assumption that the average sensitivity of f is small, i.e. that $\mathbb{E}_x[s_f(x)]$ is small, is replaced by the weaker condition that $\mathbb{E}_x[s_f(x)^p]$ is small, for p slightly larger than $1/2$.

In the rest of this introductory section, we describe our results in more detail.

[1] Namely that there is $\beta > 0$ such that if f is balanced, $\max_i I_i[f] \geqslant \Omega((\log n)^{\beta}/n)$.

1.3 Classical isoperimetric inequalities

We give simpler proofs for the results below, which are due to Talagrand [Tal93] and Eldan-Gross [EG20].

Theorem 1.1. *There exists an absolute constant $c > 0$, such that for all functions $f\colon \{0,1\}^n \to \{0,1\}$ we have that*

$$\mathop{\mathbb{E}}_x\left[\sqrt{s_f(x)}\right] \geqslant c\cdot \mathsf{var}(f)\sqrt{\log\left(\frac{1}{\mathsf{var}(f)}\right)}.$$

Theorem 1.2. *There exists an absolute constant $c > 0$, such that for all functions $f\colon \{0,1\}^n \to \{0,1\}$ we have that*

$$\mathop{\mathbb{E}}_x\left[\sqrt{s_f(x)}\right] \geqslant c\cdot \mathsf{var}(f)\sqrt{\log\left(1+\frac{1}{\sum\limits_{i=1}^{n} I_i[f]^2}\right)}.$$

Our technique is more general, and for example implies that the above theorems hold for all $p \in [1/2, 1]$, the above theorems hold if all square-roots are replaced with p-th power.

1.4 Junta theorems

We prove a strengthening of Friedgut's junta theorem [Fri98], which is Theorem 1.4 below.

Definition 1.3. *A function $f\colon \{0,1\}^n \to \{0,1\}$ is called a j-junta if there exists $J \subseteq [n]$ of size at most j, and $g\colon \{0,1\}^J \to \{0,1\}$ such that $f(x) = g(x_J)$ for all $x \in \{0,1\}^n$.*

We say a function $f\colon \{0,1\}^n \to \{0,1\}$ is ε-close to a j-junta if there is a j-junta $g\colon \{0,1\}^n \to \{0,1\}$ such that $\Pr_x\left[f(x) \neq g(x)\right] \leqslant \varepsilon$.

In this language, Friedgut's result asserts that if $\mathbb{E}_x\left[s_f(x)\right] \leqslant A$ (where A is typically thought of $O(1)$), then for all $\varepsilon > 0$, f is ε-close to a $2^{O(A/\varepsilon)}$-junta. On the other hand, we have seen that $\mathbb{E}_x\left[\sqrt{s_f(x)}\right] = \Omega(1)$ for all functions with constant variance. What can be said about intermediate moments of the sensitivity, say $\mathbb{E}_x\left[s_f(x)^p\right]$ for some $1/2 < p < 1$? We establish the following result:

Theorem 1.4. *For all $\varepsilon, \delta > 0$ and $A > 0$, there is $J = 2^{O_\delta((A/\varepsilon)^{1/p})}$ such that if $1/2+\delta \leqslant p \leqslant 1$, and $f\colon \{0,1\}^n \to \{0,1\}$ is a function such that $\mathbb{E}_x\left[s_f(x)^p\right] \leqslant A$, then f is ε-close to a J-junta.*

In particular, we conclude a version of the KKL theorem for pth moments of sensitivity:

Corollary 1.5. *For all $\delta > 0$, there is $c > 0$ such that the following holds. Suppose $f\colon \{0,1\}^n \to \{0,1\}$ satisfies that $\mathbb{E}_x\left[s_f(x)^p\right] \leqslant A$ for $1/2+\delta \leqslant p \leqslant 1$; then there exists $i \in [n]$ such that*

$$I_i[f] \geqslant 2^{-c(A/\mathsf{var}(f))^{1/p}}.$$

We note that the KKL theorem is this statement for $p = 1$, and the version for $p < 1$ is only stronger than the KKL theorem as always

$$\mathop{\mathbb{E}}_x\left[s_f(x)^p\right] \leqslant \mathop{\mathbb{E}}_x\left[s_f(x)\right]^p.$$

We remark that a naive application of our method establishes sub-optimal dependency between the parameters A and J, but gives some meaningful relation to the noise sensitivity theorem of Bourgain [Bou02].

More precisely, we show that if $f\colon \{0,1\}^n \to \{0,1\}$ satisfying $\mathbb{E}_x[s_f(x)^p] \leqslant A$, then f must be *noise stable*. Namely, for $\varepsilon > 0$, the $(1-\varepsilon)$-correlated distribution (x,y) is defined by taking $x \in_R \{0,1\}^n$, and for each $i \in [n]$ independently, setting $y_i = x_i$ with probability $(1-\varepsilon)$ and otherwise we sampling $y_i \in_R \{0,1\}$. In this language, we show that if $\mathbb{E}_x[s_f(x)^p] \leqslant A$, then

$$\Pr_{x,y\ (1-\varepsilon)\text{-correlated}}[f(x) = f(y)] \geqslant 1 - O(A\cdot\varepsilon^p),$$

from which one may conclude the result of Theorem 1.4 by appealing to Bourgain's result [Bou02]. This derivation however leads to sub-optimal parameters in Theorem 1.4, and to improve these parameters we have to work a bit harder; see Section 4 for more details.

2 Preliminaries

Notations. We will use big O notations throughout the paper. Sometimes, it will be easier for us to use $\lesssim$ and $\gtrsim$ notations; when we say that $X \lesssim Y$ we mean that there is an absolute constant $C > 0$ such that $X \leqslant CY$. Analogously, when we say that $X \gtrsim Y$ we mean that there is an absolute constant $C > 0$ such that $X \geqslant CY$.

2.1 Fourier analysis over the hypercube

We consider the space of real-valued functions $f : \{0,1\}^n \to \mathbb{R}$, equipped with the inner product $\langle f, g\rangle = \mathbb{E}_{x\in_R\{0,1\}^n}[f(x)g(x)]$. It is well-known that the functions $\{\chi_S\}_{S\subseteq[n]}$ defined by $\chi_S(x) = \prod\limits_{i\in S}(-1)^{x_i}$ form an orthonormal basis, so one may expand any function f uniquely as $f(x) = \sum\limits_{S\subseteq[n]}\widehat{f}(S)\chi_S(x)$ where $\widehat{f}(S) = \langle f, \chi_S\rangle$. We also consider L_p norms for $p \geqslant 1$, that are defined as $\|f\|_p = (\mathbb{E}_x[|f(x)|^p])^{1/p}$.

For $z \in \{0,1\}^n$, we denote by $z_{-i} \in \{0,1\}^{n-1}$ the vector resulting from dropping the ith coordinate of z. We denote by $(x_i = 1, x_{-i} = z_{-i})$ the n-bit vector which is 1 on the ith coordinate and is equal to z on any other coordinate.

Definition 2.1. *The derivative of f along direction $i \in [n]$ is $\partial_i f\colon \{0,1\}^n \to \mathbb{R}$ defined by $\partial_i f(z) = f(x_i = 0, x_{-i} = z_{-i}) - f(x_i = 1, x_{-i} = z_{-i})$. The gradient of f, $\nabla f\colon \{0,1\}^n \to \mathbb{R}^n$, is defined as $\nabla f(z) = (\partial_1 f(z), \ldots, \partial_n f(z))$.*

Note that for each $x \in \{0,1\}^n$, $\partial_i f(x) \neq 0$ if and only if the coordinate i is sensitive on x, and in this case its value is either 1 or -1. Thus, $\|\partial f(x)\|_2 = \sqrt{s_f(x)}$, and it will be often convenient for us to use the ℓ_2-norm of the gradient of f in place of $\sqrt{s_f(x)}$.

Fact 2.2. *Let $f\colon \{0,1\}^n \to \mathbb{R}$, and let $i \in [n]$. Then*

$$\partial_i f(x) = 2\sum_{S\ni i}\widehat{f}(S)\chi_{S\setminus\{i\}}(x).$$

In particular, we have that $\mathbb{E}_x[\partial_i f(x)] = 2\widehat{f}(\{i\})$.

Next, we define the level d Fourier weight of a function f as well as the approximate level d of a function f.

Definition 2.3. *The level d Fourier weight of f is defined to be $W_{=d}[f] = \sum\limits_{|S|=d} \widehat{f}(S)$. The approximate level d weight of a function f is defined to be $W_{\approx d}[f] = \sum\limits_{d\leqslant j<2d} W_{=j}[f]$.*

2.2 Random restrictions

For a function $f\colon \{0,1\}^n \to \mathbb{R}$, a set of coordinates $I \subseteq [n]$ and $z \in \{0,1\}^I$, we define the restriction of f according to (I,z) as $f_{I\to z}\colon \{0,1\}^{[n]\setminus I} \to \mathbb{R}$ defined by

$$f_{I\to z}(y) = f(x_I = z, x_{\bar{I}} = y).$$

We refer to the coordinates outside I as "alive", and the coordinates of I as fixed. We will often be interested in random restrictions of a function f. By that, we mean that the set I is chosen randomly by including in it each $i \in [n]$ with some probability, and picking $z \in \{0,1\}^I$ uniformly. We have the following well known fact, which establishes a relation between the approximate level d weight of f and the level 1 weight of a random restriction of f in which each $i \in [n]$ is included in I with probability $1 - 1/d$.

Fact 2.4. *Let $f\colon \{0,1\}^n \to \mathbb{R}$ and $d \in \mathbb{N}$, and let (J,z) be a random restriction where each $j \in [n]$ is included in J with probability $\frac{1}{d}$, and $z \in_R \{0,1\}^{\bar{J}}$. Then $\mathbb{E}_{J,z}\left[W_{=1}[f_{\bar{J}\to z}]\right] \gtrsim W_{\approx d}[f]$.*

2.3 Hypercontractivity

We will need the following well-known consequence of the hypercontractive inequality, stated below.

Theorem 2.5. *Let $f\colon \{0,1\}^n \to \mathbb{R}$ be a function of degree at most d. Then $\|f\|_4 \leqslant \sqrt{3}^d \|f\|_2$.*

3 Isoperimetric inequalities on the hypercube

In this section, we prove Theorems 1.1 and 1.2. We begin with the following very crude bound on the Talagrand boundary, and then improve it using random restrictions.

Lemma 3.1. *Let $f\colon \{0,1\}^n \to \mathbb{R}$. Then $\mathbb{E}\left[\|\nabla f(x)\|_2\right] \geqslant 2\sqrt{W_{=1}[f]}$.*

Proof. By the triangle inequality and Fact 2.2 we have

$$\mathop{\mathbb{E}}_{x}\left[\|\nabla f(x)\|_2\right] \geqslant \|\mathop{\mathbb{E}}_{x}\left[\nabla f(x)\right]\| = \|2(\widehat{f}(\{1\}), \ldots, \widehat{f}(\{n\}))\|_2 = 2\sqrt{\sum_{i=1}^{n} \widehat{f}(\{i\})^2} = 2\sqrt{W_{=1}[f]}.\square$$

Lemma 3.2. *Let $d \in \mathbb{N}$, and $f\colon \{0,1\}^n \to \{0,1\}$. Then $\mathbb{E}\left[\|\nabla f(x)\|_2\right] \gtrsim \sqrt{d}W_{\approx d}[f]$.*

Proof. Let (J,z) be a random restriction for f, where each $j \in [n]$ is included in J with probability $1/d$, and $z \in \{0,1\}^{\bar{J}}$ is sampled uniformly. Fix $x \in \{0,1\}^n$, and note that

$$\mathop{\mathbb{E}}_{J}\left[\|\nabla f_{\bar{J}\to x_{\bar{J}}}(x_J)\|_2\right] \leqslant \sqrt{\mathop{\mathbb{E}}_{J}\left[\|\nabla f_{\bar{J}\to x_{\bar{J}}}(x_J)\|_2^2\right]} = \sqrt{\mathop{\mathbb{E}}_{J}\left[\sum_{j} |\partial_j f(x)|^2 1_{j\in J}\right]} = \frac{\|\nabla f(x)\|_2}{\sqrt{d}}.$$

Therefore,

$$\frac{1}{\sqrt{d}}\mathbb{E}\left[\|\nabla f(x)\|_2\right] \geqslant \mathop{\mathbb{E}}_{J,z}\left[\mathop{\mathbb{E}}_{x\in\{0,1\}^J}\left[\|\nabla f_{\bar{J}\to z}(x)\|_2\right]\right]. \tag{2}$$

On the other hand, for each J and $z \in \{0,1\}^{\bar{J}}$ we may apply Lemma 3.1 on $f_{\bar{J}\to z}$ to get that

$$\mathop{\mathbb{E}}_{x\in\{0,1\}^J}\left[\|\nabla f_{\bar{J}\to z}(x)\|_2\right] \gtrsim \sqrt{W_{=1}[f_{\bar{J}\to z}]} \geqslant W_{=1}[f_{\bar{J}\to z}].$$

Taking expectation over J and z we get that

$$\mathop{\mathbb{E}}_{J,z}\left[\mathop{\mathbb{E}}_{x\in\{0,1\}^J}\left[\|\nabla f_{\bar{J}\to z}(x)\|_2\right]\right] \gtrsim \mathop{\mathbb{E}}_{J,z}\left[W_{=1}[f_{\bar{J}\to z}]\right] \gtrsim W_{\approx d}[f],$$

where the last inequality is by Fact 2.4. Combining this with equation (2) finishes the proof. □

Lemma 3.3. *Let $d \in \mathbb{N}$, and $f\colon \{0,1\}^n \to \{0,1\}$. Then $\mathbb{E}\left[\|\nabla f(x)\|_2\right] \gtrsim \sqrt{d}W_{\geqslant d}[f]$.*

Proof. Applying Lemma 3.2 with $d2^j$ for $j = 0, 1, 2, \ldots$ we get that

$$2^{-j/2}\mathop{\mathbb{E}}_{J}\left[\|\nabla f_{\bar{J}\to x_{\bar{J}}}(x_J)\|_2\right] \gtrsim \sqrt{d}W_{\approx 2^j d}[f],$$

and summing over j yields the result. □

3.1 Proof of Theorem 1.1

If $\mathsf{var}(f) \geqslant 2^{-16}$, then it is enough to prove a lower bound of $\gtrsim \mathsf{var}(f)$. Applying Lemma 3.2 on d's that are powers of 2 between 1 and n, we get that

$$\sum_{d=2^j}\frac{1}{\sqrt{d}}\mathbb{E}\left[\|\nabla f(x)\|_2\right] \gtrsim \sum_{d=2^j} W_{\approx d}[f] = \mathsf{var}(f),$$

and by geometric sum, the left hand side is $\lesssim \mathbb{E}\left[\|\nabla f(x)\|_2\right]$.

We thus assume that $\mathsf{var}(f) \leqslant 2^{-16}$, and without loss of generality the majority value of $f(x)$ is 1. Thus, $\mu(f) \leqslant 2\mathsf{var}(f)$. Setting $d = \frac{1}{8}\log(1/\mathsf{var}(f))$, we have that

$$\sum_{0<|S|\leqslant d}\widehat{f}(S)^2 = \|f^{\leqslant d}\|_2^2 = \langle f, f^{\leqslant d}\rangle \leqslant \|f\|_{4/3}\|f^{\leqslant d}\|_4 \leqslant \sqrt{3}^d\mu(f)^{3/4}\|f\|_2 = 2^d\mu(f)^{5/4} \leqslant 0.9\mathsf{var}(f),$$

where we used the bound $\|f^{\leqslant d}\|_4 \leqslant \sqrt{3}^d\|f^{\leqslant d}\|_2$ from Theorem 2.5. Therefore, $W_{\geqslant d}[f] \geqslant 0.1\mathsf{var}(f)$, and by Lemma 3.3 we conclude the result. □

3.2 Proof of Theorem 1.2

Denote $M[f] = \sum_{i=1}^{n} I_i[f]^2$; we need the following result due to [Tal96, BKS99, KK13] (the precise version we use is due to Keller and Kindler [KK13]).

Theorem 3.4. *There are $c_1, c_2 > 0$, such that $W_{\leqslant c_1\log(1/M[f])}[f] \leqslant M[f]^{c_2}$.*

We may now give the proof of Theorem 1.2.

Proof of Theorem 1.2. If $M[f] \geqslant \mathsf{var}(f)^{2/c_2}$, then the result follows from Theorem 1.1. Otherwise, fix c_1, c_2 from Theorem 3.4 and set $d = c_1 \log(1/M[f])$. By Theorem 3.4 we have that $W_{\leqslant d}[f] \leqslant M[f]^{c_2} \leqslant \mathsf{var}(f)^2 \leqslant \frac{1}{2}\mathsf{var}(f)$, and so $W_{>d}[f] \geqslant \frac{1}{2}\mathsf{var}(f)$, and the result follows from Lemma 3.3. □

We remark that Theorem 1.2 implies a stability result for the KKL theorem. Namely, if all influences of a Boolean function are at most $O(\log n/n)$, then f must have constant-size vertex boundary. More precisely:

Corollary 3.5. *For any $K > 0$ there are $c, \delta > 0$ such that the following holds. Suppose that for $f\colon \{0,1\}^n \to \{0,1\}$ we have that $\max_i I_i[f] \leqslant K\mathsf{var}(f)\frac{\log n}{n}$. Then*

$$\Pr_x\left[\|\nabla f(x)\|_2 \geqslant c\mathsf{var}(f)\sqrt{\log n}\right] \geqslant \delta\mathsf{var}(f).$$

Proof. Note that by assumption, $M[f] \leqslant K^2\frac{\log^2 n}{n} \leqslant \frac{K^2}{\sqrt{n}}$, so Theorem 1.2 implies that $Z \stackrel{def}{=} \mathbb{E}_x\left[\|\nabla f(x)\|_2\right] \geqslant K'\mathsf{var}(f)\sqrt{\log n}$ for some K' depending only on K. Also, we have that

$$\mathop{\mathbb{E}}_x\left[\|\nabla f(x)\|_2^2\right] = I_1[f] + \ldots + I_n[f] \leqslant K\mathsf{var}(f)\log n.$$

It follows from the Paley-Zygmund inequality that

$$\Pr\left[\|\nabla f(x)\|_2 \geqslant \frac{1}{2}Z\right] \geqslant \frac{3}{4}\frac{Z^2}{\mathbb{E}_x\left[\|\nabla f(x)\|_2^2\right]} \geqslant \frac{3}{4}\frac{K'^2}{K}\mathsf{var}(f),$$

and the claim is proved for $c = \frac{1}{2}K'$ and $\delta = \frac{3}{4}\frac{K'^2}{K}$. □

3.3 Extensions

3.3.1 Other L_p-norms

Lemma 3.1 applies not only in the case of L_2-norms, but rather for any L_p norm:

Lemma 3.6. *Let $1 \leqslant p \leqslant 2$ and $f\colon \{0,1\}^n \to \{-1,1\}$. Then $\mathbb{E}\left[\|\nabla f(x)\|_p\right] \geqslant W_{=1}[f]^{1/p}$.*

Proof. By Jensen and Lemma 3.1 we have

$$\mathop{\mathbb{E}}_x\left[\|\nabla f(x)\|_p\right] = \mathop{\mathbb{E}}_x\left[\|\nabla f(x)\|_2^{2/p}\right] \geqslant \mathop{\mathbb{E}}_x\left[\|\nabla f(x)\|_2\right]^{2/p} \geqslant W_{=1}[f]^{1/p}. \square$$

As in Lemma 3.2, the previous lemma quickly leads to the following conclusion.

Lemma 3.7. *Let $d \in \mathbb{N}$, $\alpha \in [1/2, 1]$ and $f\colon \{0,1\}^n \to \{0,1\}$. Then $\mathbb{E}\left[s_f(x)^\alpha\right] \gtrsim d^\alpha W_{\geqslant d}[f]$.*

Using the same argument as before and replacing invocations of Lemma 3.3 with Lemma 3.7, one may conclude variants of Theorems 1.1 and 1.2 in which the square root is replaced with any power $p \in [1/2, 1]$.

3.3.2 Talagrand boundary and noise sensitivity

We next demonstrate the connection between having small Talagrand-type boundary to being noise stable, a notion that we define next.

For a parameter $\rho \in [0,1]$ and a point $x \in \{0,1\}^n$, the distribution of ρ-correlated inputs with x, denoted as $y \sim_\rho x$, is defined by the following randomized process. For each $i \in [n]$, set $y_i = x_i$ with probability ρ, and otherwise sample $y_i \in \{0,1\}$ uniformly. The operator $T_\rho \colon L_2(\{0,1\}^n) \to L_2(\{0,1\}^n)$ is defined as

$$T_\rho f(x) = \mathop{\mathbb{E}}_{y \sim_\rho x}[f(y)].$$

Definition 3.8. *For $\varepsilon > 0$, the noise stability of f with noise rate ε is defined as $\mathsf{Stab}_{1-\varepsilon}(f) = \langle f, T_{1-\varepsilon} f\rangle$.*

The relation established in Lemma 3.7 between the Fourier weight of f and $\mathbb{E}\left[s_f(x)^\alpha\right]$ for $1/2 \leqslant \alpha \leqslant 1$ implies a relation between the latter quantity and noise stability, as follows:

Corollary 3.9. *There exists an absolute constant $C > 0$, such that for any $\delta \in [0,1/2]$ and $f \colon \{0,1\}^n \to \{0,1\}$ satisfying $\mathbb{E}\left[s_f(x)^{\frac{1}{2}+\delta}\right] \leqslant A$, we have that $\mathsf{Stab}_{1-\varepsilon}(f) \geqslant 1 - C \cdot A \cdot \varepsilon^{\frac{1}{2}+\delta}$.*

Proof.

$$1 - \mathsf{Stab}_{1-\varepsilon}(f) = \mathop{\Pr}_{x,y\ (1-\varepsilon)\text{-correlated}}[f(x) \neq f(y)] = 2\langle 1-f, T_{1-\varepsilon} f\rangle = 2\sum_{k=0}^{n}(1-(1-\varepsilon)^k)W_{=k}[f].$$

Set $d = 1/\varepsilon$; we split the sum into $k < d$ and $k \geqslant d$, and upper bound the contribution of each part separately. For $k \geqslant d$, using Lemma 3.7 we have

$$2\sum_{k=d}^{n}(1-(1-\varepsilon)^k)W_{=k}[f] \leqslant 2W_{\geqslant d}[f] \lesssim A \cdot d^{-(1/2+\delta)} = A\varepsilon^{1/2+\delta}.$$

For $k \leqslant d$, we write

$$\sum_{k=0}^{d}(1-(1-\varepsilon)^k)W_{=k}[f] \leqslant \sum_{k=1}^{d} k\varepsilon W_{=k}[f] \leqslant \varepsilon\sum_{k=1}^{d} W_{\geqslant k}[f].$$

Using Lemma 3.7 we get $W_{\geqslant k}[f] \lesssim A \cdot k^{-(1/2+\delta)}$, and so

$$\sum_{k=0}^{d}(1-(1-\varepsilon)^k)W_{=k}[f] \lesssim A \cdot \varepsilon \sum_{k=0}^{d}\frac{1}{k^{1/2+\delta}} \lesssim A \cdot \varepsilon d^{1/2-\delta} \lesssim A\varepsilon^{1/2+\delta}. \qquad \square$$

4 Improved junta theorems

4.1 Proof of Theorem 1.4

In this section, we prove Theorem 1.4. As explained in the introduction, if one does not care about getting tight dependency between the junta size J and the parameter A, then one may use Corollary 3.9 and Bourgain's noise sensitivity theorem [Bou02] and get worse parameters. The focus here it thus to get optimal dependency of J on A. For the proof, we will need the notion of noisy influences of a function f, defined next.

Definition 4.1. *The ρ-noisy influence of a coordinate i for $f\colon \{0,1\}^n \to \mathbb{R}$ is defined to be $I_i^{(\rho)}[f] = I_i[T_\rho f]$.*

Fact 4.2. *[[O'D14]] For all $f\colon \{0,1\}^n \to \{0,1\}$ we have that $\sum_{i=1}^{n} I_i^{(\rho)}[f] = O(\frac{1}{1-\rho})$.* □

Let C_1, C_2 be two absolute constants, and we think of C_2 as much larger than C_1. Given $f\colon \{0,1\}^n \to \{0,1\}$ with $\mathbb{E}_x\left[s_f(x)^p\right] \leqslant A$, we denote $d = C_1\left(\frac{A}{\varepsilon}\right)^{1/p}$, $\delta = 2^{-\frac{C_2}{2p-1}d}$ and the candidate set of junta coordinates as

$$\mathcal{J} = \left\{ i \in [n] \mid I_i[T_{1-1/2d}[f]] \geqslant \delta \right\},$$

i.e. $\mathcal{J}$ is the set of coordinates whose noisy influence at f is somewhat large. Our task is to show that f is ε-close to a $\mathcal{J}$ junta, i.e. that

$$\sum_{S: S \not\subseteq \mathcal{J}} \widehat{f}(S)^2 \leqslant \varepsilon,$$

and the rest of the proof is devoted towards this goal. Towards this end, we partition the contribution of different scales of $|S \cap \mathcal{J}|$ to the left hand side, and for each $k = 1, \ldots, \log d$ define

$$W_{k,d}[f] = \sum_{\substack{S:|S|\leqslant d, \\ 2^{k-1}\leqslant \left|S\cap\bar{\mathcal{J}}\right|\leqslant 2^k}} \widehat{f}(S)^2.$$

Thus, we may write

$$\sum_{S: S \not\subseteq \mathcal{J}} \widehat{f}(S)^2 \leqslant W_{>d}[f] + \sum_{k=1}^{\log d} W_{k,d}[f]. \tag{3}$$

The rest of the proof is devoted to bounding $W_{>d}[f]$ and $W_{k,d}[f]$ for each k. By Lemma 3.7 that

$$W_{>d}[f] \leqslant O\left(\frac{A}{d^p}\right) \leqslant \frac{\varepsilon}{16} \tag{4}$$

for sufficiently large C_1. Next, we bound $W_{k,d}[f]$. Fix k, and consider the operator $S_k = T_{1-1/d}^{\mathcal{J}} T_{1-1/2^k}^{\bar{\mathcal{J}}}$, i.e. the noise operator that applies $1/2^k$ noise on coordinates outside $\mathcal{J}$, and $1/d$ noise on coordinates of $\mathcal{J}$. Note that

$$W_{k,d}[f] \lesssim W_{k,d}[S_k f],$$

so it suffices to bound $W_{k,d}[S_k f]$. We partition $W_{k,d}[S_k f]$ according to contribution of each $j \in \bar{\mathcal{J}}$, getting:

$$W_{k,d}[S_k f] = \sum_{\substack{S:|S|\leqslant d, \\ 2^{k-1}\leqslant \left|S\cap\bar{\mathcal{J}}\right|\leqslant 2^k}} \widehat{S_k f}(S)^2 \leqslant \frac{1}{2^k} \sum_{j \notin \mathcal{J}} \sum_{\substack{S\ni j:|S|\leqslant d, \\ 2^{k-1}\leqslant \left|S\cap\bar{\mathcal{J}}\right|\leqslant 2^k}} \widehat{S_k f}(S)^2 \leqslant \frac{1}{2^k} \sum_{j \notin \mathcal{J}} \|S_k \partial_j f\|_2^2 \overset{def}{=} \tilde{W}_{k,d}[f].$$

Thus, we have that $W_{k,d}[f] \lesssim \tilde{W}_{k,d}[f]$, and the rest of the proof is devoted to bounding the right hand side. Define the operator S'_k as

$$S'_k = T_{1-1/2d}^{\mathcal{J}} T_{1-1/2^k+1/2d}^{\bar{\mathcal{J}}}.$$

We establish the following claim using random restrictions and hypercontractivity:

Claim 4.3. *For all $\tau \in (0,1)$, we have that*

$$2^k \tilde{W}_{k,d}[f] \leqslant C \cdot d^{-p} \tau^{2p-1} \mathop{\mathbb{E}}_{x}\left[s_f(x)^p\right] + C \cdot \tau^{-\frac{1}{100d}} \sum_{j \in \bar{\mathcal{J}}} \|S'_k \partial_j f\|_2^{2+\frac{1}{100d}},$$

where C is an absolute constant.

Proof. Deferred to Section 4.2. □

Given Claim 4.3, one may pull out $\|S'_k \partial_j f\|_2^{1/100d} \leqslant \delta^{1/100d}$ outside the sum, and bound the sum on the rest of the noisy influences by $O(d)$ to get a bound of the form $2^k \tilde{W}_{k,d}[f] \lesssim d^{-p} \tau^{2p-1} \mathbb{E}_x\left[s_f(x)^p\right] + \tau^{-\frac{1}{100d}} \delta^{\frac{1}{200d}}$. This idea is good enough to establish a junta theorem, yet it gives worse parameters.[2] Instead, to get the tight result with respect to the junta size, we will need a more careful bound on the sum of noisy influences outside $\mathcal{J}$. Indeed, using an improved bound of the sum of noisy influences that relates them to $\tilde{W}_{\ell,d}[f]$. This is captured in the following claim, for which we need to define

$$\varepsilon_{\ell,d} = \sum_{\substack{|S|>d \\ 2^{\ell-1} \leqslant \left|S \cap \bar{J}\right| \leqslant 2^{\ell}}} \widehat{f}(S)^2.$$

Claim 4.4. $2^{-k} \sum_{j \in \bar{\mathcal{J}}} \|S'_k \partial_j f\|_2^2 \lesssim \sum_{\ell=1}^{\log n} 2^{-|\ell-k|} (W_{\ell,d}[f] + \varepsilon_{\ell,d})$.

Proof. By definition,

$$2^{-k} \sum_{j \in \bar{\mathcal{J}}} \|S'_k \partial_j f\|_2^2 = \sum_{S} \frac{\left|S \cap \bar{J}\right|}{2^k} \widehat{S'_k f}(S)^2 = \sum_{S} \frac{\left|S \cap \bar{J}\right|}{2^k} \left(1 - \frac{1}{2^k} + \frac{1}{2d}\right)^{\left|S \cap \bar{J}\right|} \left(1 - \frac{1}{2d}\right)^{|S \cap J|} \widehat{f}(S)^2.$$

Partitioning the last sum according to $\left|S \cap \bar{J}\right|$ and dyadically partitioning so that $2^{\ell-1} \leqslant \left|S \cap \bar{J}\right| < 2^{\ell}$, we have that the last expression is at most

$$\sum_{\ell=1}^{\log n} 2^{\ell-k} \left(1 - \frac{1}{2^k} + \frac{1}{2d}\right)^{2^{\ell-1}} (W_{\ell,d}[f] + \varepsilon_{\ell,d}) \lesssim \sum_{\ell=1}^{\log n} 2^{-|\ell-k|} (W_{\ell,d}[f] + \varepsilon_{\ell,d}).$$ □

Combining Claims 4.3 and 4.4 we immediately get the following corollary:

Corollary 4.5. *There exists an absolute constant $C > 0$ such that for all $\tau > 0$,*

$$\tilde{W}_{k,d}[f] \leqslant C \cdot 2^{-k} d^{-p} \tau^{2p-1} \mathop{\mathbb{E}}_{x}\left[s_f(x)^p\right] + C \cdot \tau^{-\frac{1}{100d}} \delta^{\frac{1}{200d}} \left(\sum_{\ell=1}^{\log d} 2^{-|\ell-k|} W_{\ell,d}[f] + \sum_{\ell=1}^{\log n} 2^{-|\ell-k|} \varepsilon_{\ell,d}\right).$$

We may now give an upper bound on $\sum_{k=1}^{\log d} W_{k,d}[f]$.

[2]One has to take $\delta = 2^{-O(d \log d)}$, as opposed to $\delta = 2^{-O(d)}$ as we have taken. We remark though if one only cares about getting a larger junta of size $2^{O(d \log d)}$, the proof greatly simplifies.

Claim 4.6. *For sufficiently large absolute constant C_1 and sufficiently large C_2 in comparison to C_1, we have that* $\sum_{k=1}^{\log d} W_{k,d}[f] \leqslant \varepsilon/16$.

Proof. Let $\tau > 0$ to be determined. As $W_{k,d}[f] \lesssim \tilde{W}_{k,d}[f]$, by Claim 4.5 we have

$$\sum_{k=1}^{\log d} W_{k,d}[f] \lesssim \sum_{k=1}^{\log d} 2^{-k} d^{-p} \tau^{2p-1} A + \tau^{-\frac{1}{100d}} \delta^{\frac{1}{200d}} \left(\sum_{\ell=1}^{\log d} 2^{-|\ell-k|} W_{\ell,d}[f] + \sum_{\ell=1}^{\log n} 2^{-|\ell-k|} \varepsilon_{\ell,d} \right).$$

The first sum is clearly $\lesssim d^{-p}\tau^{2p-1}A$. The second sum is at most

$$\tau^{-\frac{1}{100d}} \delta^{\frac{1}{200d}} \sum_{\ell=1}^{\log d} W_{\ell,d}[f] \sum_{k} 2^{-|\ell-k|} \lesssim \tau^{-\frac{1}{100d}} \delta^{\frac{1}{200d}} \sum_{\ell=1}^{\log d} W_{\ell,d}[f].$$

Finally, the third sum is at most

$$\tau^{-\frac{1}{100d}} \delta^{\frac{1}{200d}} \sum_{\ell=1}^{\log n} \varepsilon_{\ell,d} \sum_{k=1}^{\log d} 2^{-|\ell-k|} \lesssim \tau^{-\frac{1}{100d}} \delta^{\frac{1}{200d}} \sum_{\ell=1}^{\log n} \varepsilon_{\ell,d} \lesssim \tau^{-\frac{1}{100d}} \delta^{\frac{1}{200d}} W_{>d}[f] \lesssim \tau^{-\frac{1}{100d}} \delta^{\frac{1}{200d}} \varepsilon,$$

where we used (4). We thus get that

$$\sum_{k=1}^{\log d} W_{k,d}[f] \leqslant C \cdot d^{-p} \tau^{2p-1} A + C \cdot \tau^{-\frac{1}{100d}} \delta^{\frac{1}{200d}} \sum_{k=1}^{\log d} W_{k,d}[f] + C \cdot \tau^{-\frac{1}{100d}} \delta^{\frac{1}{200d}} \varepsilon$$

for some absolute constant $C > 0$. We choose $\tau = \delta^{1/4}$ and then C_2 large enough so that $C\tau^{-\frac{1}{100d}}\delta^{\frac{1}{200d}} \leqslant 1/2$, getting that $\sum_{k=1}^{\log d} W_{k,d}[f] \leqslant 2C \cdot d^{-p} \delta^{(2p-1)/4} A + 2C\delta^{\frac{1}{400d}} \varepsilon$, which is at most $\varepsilon/16$ for large enough C_2. □

We can now prove Theorem 1.4.

Proof of Theorem 1.4. Combining Claim (3), inequality (4) and Claim 4.6 we get that

$$\sum_{S: S \not\subseteq \mathcal{J}} \widehat{f}(S)^2 \leqslant \varepsilon/8.$$

Thus, defining $h(x) = 1$ if $\sum_{S: S \not\subseteq \mathcal{J}} \widehat{f}(S)\chi_S(x) \geqslant 1/2$ and 0 otherwise, we get that h is a $\mathcal{J}$-junta and

$$\|f - h\|_1 = \|f - h\|_2^2 \leqslant 4 \left\| f - \sum_{S: S \not\subseteq \mathcal{J}} \widehat{f}(S)\chi_S(x) \right\|_2^2 = 4 \sum_{S: S \not\subseteq \mathcal{J}} \widehat{f}(S)^2 \leqslant \varepsilon.$$

Finally, we note that as the sum of $(1 - 1/2d)$ noisy influences of f is at most $O(d)$ by Fact 4.2, we get that

$$|\mathcal{J}| \leqslant O\left(\frac{d}{\delta}\right).$$

□

4.2 Proof of Claim 4.3

For the proof of Claim 4.3 we need a version of the hypercontractive inequality with the noise operator, as follows:

Lemma 4.7. *[[O'D14]] Let* $q > 2$*, and* $\rho \leqslant \frac{1}{\sqrt{q-1}}$*. Then for all* $f\colon \{0,1\}^n \to \mathbb{R}$ *we have* $\|T_\rho f\|_q \leqslant \|f\|_2$*.*

We also need the following claim, asserting that noise only decreases Talagrand's boundary.

Claim 4.8. *For* $f\colon \{0,1\}^n \to \mathbb{R}$ *and any noise operator* $S = T_{\rho_1} \otimes \ldots T_{\rho_n}$ *we have* $\mathbb{E}_x\left[\|\nabla Sf(x)\|_p\right] \leqslant \mathbb{E}_x\left[\|\nabla f(x)\|_p\right]$*.*

Proof.

$$\mathop{\mathbb{E}}_{x}\left[\|\nabla(Sf)(x)\|_p\right] \leqslant \mathop{\mathbb{E}}_{x}\left[\|S\nabla f(x)\|_p\right] = \mathop{\mathbb{E}}_{x}\left[\|\mathop{\mathbb{E}}_{y\sim Sx}\left[\nabla f(y)\right]\|_p\right],$$

and the result follows from the triangle inequality. □

Take (I, z) to be a random restriction so that each $i \in [n]$ is included in I with probability $1 - 1/2d$, and consider $g = S_k f$. We calculate the contribution of $\bar{\mathcal{J}}$ to the level 1 weight of a restriction of g:

$$d\mathop{\mathbb{E}}_{I,z}\left[\sum_{i\in\bar{I}\cap\bar{\mathcal{J}}} \widehat{g_{I\to z}}(\{i\})^2\right] = \sum_S d\mathop{\mathbb{E}}_I\left[\left|S\cap\bar{I}\cap\bar{\mathcal{J}}\right|\widehat{g}(S)^2\right] = \sum_S \left|S\cap\bar{\mathcal{J}}\right|\widehat{g}(S)^2.$$

A direct computation also shows that this is equal to $\sum_{j\notin\mathcal{J}} \|S_k\partial_j f\|_2^2$, hence our goal translates to bounding the contribution of $\mathcal{J}$ to the level 1 of random restrictions of g

For I, z define

$$L_{I,z} = \left\{i\in\bar{I}\cap\bar{\mathcal{J}}\;\middle|\;\left|\widehat{g_{I\to z}}(\{i\})\right|\leqslant\tau\right\},\qquad H_{I,z} = \left\{i\in\bar{I}\cap\bar{\mathcal{J}}\;\middle|\;\left|\widehat{g_{I\to z}}(\{i\})\right|>\tau\right\}.$$

Then

$$\mathop{\mathbb{E}}_{I,z}\left[\sum_{i\in\bar{I}\cap\bar{\mathcal{J}}} \widehat{g_{I\to z}}(\{i\})^2\right] = \mathop{\mathbb{E}}_{I,z}\left[\sum_{i\in L_{I,z}} \widehat{g_{I\to z}}(\{i\})^2\right] + \mathop{\mathbb{E}}_{I,z}\left[\sum_{i\in H_{I,z}} \widehat{g_{I\to z}}(\{i\})^2\right].$$

Bounding the contribution from $L_{I,z}$. For the first sum, we have

$$\begin{aligned}\sum_{i\in L_{I,z}} \widehat{g_{I\to z}}(\{i\})^2 \leqslant \left(\sum_{i\in L_{I,z}} \widehat{g_{I\to z}}(\{i\})^2\right)^p &\leqslant \left(\sum_{i\in L_{I,z}} \tau^{(2p-1)/p}\widehat{g_{I\to z}}(\{i\})^{1/p}\right)^p \\ &\leqslant \tau^{2p-1}\left(\sum_i \widehat{g_{I\to z}}(\{i\})^{1/p}\right)^p.\end{aligned}$$

In the last expression we have τ^{2p-1} times the $1/p$-norm of the vector $(\widehat{g_{I\to z}}(\{i\}))_{i\notin I}$. Thus, using the triangle inequality similarly to Lemma 3.1 we may bound the last expression by

$$\tau^{2p-1}\mathop{\mathbb{E}}_x\left[\|\nabla g_{I\to z}(x)\|_{1/p}\right].$$

Taking expectation, we get that

$$\mathop{\mathbb{E}}_{I,z}\left[\sum_{i\in L_{I,z}}\widehat{g_{I\to z}}(\{i\})^2\right]\leqslant \tau^{2p-1}\mathop{\mathbb{E}}_{I,z}\left[\mathop{\mathbb{E}}_{x}\left[\|\nabla g_{I\to z}(x)\|_{1/p}\right]\right]\leqslant \tau^{2p-1}\mathop{\mathbb{E}}_{x}\left[\mathop{\mathbb{E}}_{I,z}\left[\|\nabla g_{I\to z}(x)\|_{1/p}\right]\right].$$

Note that fixing x, z and looking at $\nabla g(x,z)$, we have by Jensen

$$\mathop{\mathbb{E}}_{I}\left[\|\nabla g_{I\to z}(x)\|_{1/p}\right]\leqslant d^{-p}\|\nabla g(x,z)\|_{1/p},$$

hence the last expression may be upper bounded by $\tau^{2p-1}d^{-p}\mathbb{E}_{x,z}\left[\|\nabla g(x,z)\|_{1/p}\right]$. Using Claim 4.8, we have that this may be upper bounded by $\tau^{2p-1}d^{-p}\mathbb{E}_{x,z}\left[\|\nabla f(x,z)\|_{1/p}\right]=\tau^{2p-1}d^{-p}\mathbb{E}_{x,z}\left[s_f(x,z)^p\right]$.

Bounding the contribution from $H_{I,z}$. We have

$$\sum_{i\in H_{I,z}}\widehat{g_{I\to z}}(\{i\})^2\leqslant \tau^{-1/100d}\sum_{i\in H_{I,z}}\left|\widehat{g_{I\to z}}(\{i\})\right|^{2+\frac{1}{100d}}\leqslant \tau^{-1/100d}\sum_{i\in \bar{I}\cap\bar{\mathcal{J}}}\left|\widehat{g_{I\to z}}(\{i\})\right|^{2+\frac{1}{100d}},$$

and we next take expectation over z:

$$\mathop{\mathbb{E}}_{z}\left[\sum_{i\in H_{I,z}}\widehat{g_{I\to z}}(\{i\})^2\right]\leqslant \tau^{-\frac{1}{100d}}\sum_{i\in \bar{I}\cap\bar{\mathcal{J}}}\mathop{\mathbb{E}}_{z}\left[\left|\widehat{g_{I\to z}}(\{i\})\right|^{2+\frac{1}{100d}}\right].$$

We may write $g=T_{1-1/2d}S'_k f$, and note that $g_{I\to z}(\{i\})$ is equal to

$$(\partial_i g)_{I\to z}=\sum_{S\subseteq I, S\ni I}\widehat{g}(S)\chi_S(z)=\sum_{S\subseteq I, S\ni i}\left(1-\frac{1}{2d}\right)^{|S|}\widehat{S'_k f}(S)\chi_S(z).$$

Thus, using hypercontractivity, i.e. Lemma 4.7 we have that

$$\begin{aligned}\mathop{\mathbb{E}}_{z}\left[\left|\widehat{g_{I\to z}}(\{i\})\right|^{2+\frac{1}{100d}}\right]&=\left\|\sum_{S\ni i}\left(1-\frac{1}{2d}\right)^{|S|}\widehat{S'_k f}(S)\chi_S(z)\right\|_{2+1/100d}^{2+1/100d}\\&\leqslant\left\|\sum_{S\ni i}\widehat{S'_k f}(S)\chi_S(z)\right\|_{2}^{2+1/100d},\end{aligned}$$

which is $\|S'_k\partial_i f\|_2^{2+\frac{1}{100d}}$. Plugging this above yields that

$$\mathop{\mathbb{E}}_{z}\left[\sum_{i\in H_{I,z}}\widehat{g_{I\to z}}(\{i\})^2\right]\leqslant \tau^{-\frac{1}{100d}}\sum_{i\in \bar{I}\cap\bar{\mathcal{J}}}\|S'_k\partial_i f\|_2^{2+\frac{1}{100d}}.$$

Taking expectation over I now gives that

$$\mathop{\mathbb{E}}_{z,I}\left[\sum_{i\in H_{I,z}}\widehat{g_{I\to z}}(\{i\})^2\right]\leqslant \tau^{-\frac{1}{100d}}\sum_{i\in\bar{\mathcal{J}}}\mathop{\mathbb{E}}_{I}\left[1_{i\in\bar{I}}\right]\|S'_k\partial_i f\|_2^{2+\frac{1}{100d}}\leqslant\frac{1}{d}\tau^{-\frac{1}{100d}}\sum_{i\in\bar{\mathcal{J}}}\|S'_k\partial_i f\|_2^{2+\frac{1}{100d}}.$$

□

References

[BKS99] Itai Benjamini, Gil Kalai, and Oded Schramm. Noise sensitivity of boolean functions and applications to percolation. *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques*, 90(1):5–43, 1999.

[Bou02] Jean Bourgain. On the distribution of the fourier spectrum of boolean functions. *Israel Journal of Mathematics*, 131(1):269–276, 2002.

[EG20] Ronen Eldan and Renan Gross. Concentration on the boolean hypercube via pathwise stochastic analysis. In *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, pages 208–221, 2020.

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

[KK13] Nathan Keller and Guy Kindler. Quantitative relation between noise sensitivity and influences. *Combinatorica*, 33(1):45–71, 2013.

[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on boolean functions (extended abstract). In *29th Annual Symposium on Foundations of Computer Science, White Plains, New York, USA, 24-26 October 1988*, pages 68–80, 1988.

[Mar74] Grigorii Aleksandrovich Margulis. Probabilistic characteristics of graphs with large connectivity. *Problemy peredachi informatsii*, 10(2):101–108, 1974.

[O'D14] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[Tal93] Michel Talagrand. Isoperimetry, logarithmic sobolev inequalities on the discrete cube, and margulis' graph connectivity theorem. *Geometric & Functional Analysis GAFA*, 3(3):295–314, 1993.

[Tal94] Michel Talagrand. On russo's approximate zero-one law. *The Annals of Probability*, 22(3):1576–1587, 1994.

[Tal96] Michel Talagrand. How much are increasing sets positively correlated? *Combinatorica*, 16(2):243–258, 1996.

[Tal97] Michel Talagrand. On boundaries and influences. *Combinatorica*, 17(2):275–285, 1997.