Electronic Colloquium on Computational Complexity, Report No. 9 (2020)

Theorems of KKL, Friedgut, and Talagrand via Random Restrictions and Log-Sobolev Inequality

Esty Kelman* Subhash Khot† Guy Kindler‡ Dor Minzer§ Muli Safra¶

Abstract

We give alternate proofs for three related results in analysis of Boolean functions, namely the KKL Theorem, Friedgut's Junta Theorem, and Talagrand's strengthening of the KKL Theorem. We follow a new approach: looking at the first Fourier level of the function after a suitable random restriction and applying the Log-Sobolev inequality appropriately. In particular, we avoid using the hypercontractive inequality that is common to the original proofs. Our proofs might serve as an alternate, uniform exposition to these theorems and the techniques might benefit further research.

1 Introduction

Let us consider the Boolean cube $\{0,1\}^n$ equipped with the uniform measure and let $f\colon \{0,1\}^n \to \{0,1\}$ be a function. The influence of a coordinate $i \in [n]$, denoted by $I_i[f]$, is defined to be $\Pr_x\left[f(x) \neq f(x \oplus e_i)\right]$, where $x \in \{0,1\}^n$ is sampled uniformly and $x \oplus e_i$ denotes the input x with the i^{th} bit flipped. The total influence of f is $I[f] = \sum_{i=1}^{n} I_i[f]$. One of the most basic inequalities, known as Poincare's inequality, states that $I[f] \geqslant \mathsf{var}(f)$, where $\mathsf{var}(f)$ is the variance of the random variable $f(x)$ when $x \in \{0,1\}^n$ is sampled uniformly. In general, Poincare's inequality may be tight, which raises the following question: can it be the case that not only $I[f] \approx \mathsf{var}(f)$, but actually $I_i[f] \approx \frac{\mathsf{var}(f)}{n}$ for all $i \in [n]$? In other words, can all influences of f be as small as possible simultaneously? The landmark result of Kahn, Kalai, and Linial [KKL88] gives a negative answer to this question:

Theorem 1.1. *There exists an absolute constant $c > 0$, such that for any $f\colon \{0,1\}^n \to \{0,1\}$, there is a coordinate $i \in [n]$ with $I_i[f] \geqslant c \cdot \frac{\log n}{n}\,\mathsf{var}(f)$.*

The KKL Theorem and its strengthenings by Friedgut [Fri98] and Talagrand [Tal94] are foundational results in analysis of Boolean functions. These have found several applications, e.g. to the threshold phenomena, computational learning theory, extremal combinatorics, communication complexity, hardness of approximation, non-embeddability results in metric geometry, and coding theory [FK96, OS07, DF09, GKK+08, CKK+06, DS05, KR08, KR09, DKSV06, KKM+16]. Before we discuss the theorems of Friedgut and Talagrand, let us state a dimension-free variant of the KKL Theorem (that is morally equivalent to Theorem 1.1 and is easily implied by the techniques in [KKL88]).

*School of Computer Science, Tel Aviv University.

†Courant institute of Mathematical Sciences, New York University. Supported by the NSF Award CCF-1813438, the Simons Collaboration on Algorithms and Geometry, and the Simons Investigator Award.

‡Einstein Institute of Mathematics, Hebrew University of Jerusalem.

§Institute for Advanced Study. Supported partially by NSF grant DMS-1638352 and Rothschild Fellowship.

¶School of Computer Science, Tel Aviv University. Supported by the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (Grant agreement No. 835152).

ISSN 1433-8092

Theorem 1.2. *There exists an absolute constant $K > 0$, such that for any $f\colon \{0,1\}^n \to \{0,1\}$, there is a coordinate $i \in [n]$ with $I_i[f] \geqslant 2^{-K\frac{I[f]}{\mathsf{var}(f)}}$.*

We note that Theorem 1.2 implies Theorem 1.1: if $I[f] \geqslant \frac{\log n}{2K}\mathsf{var}(f)$, then clearly there is a corodinate $i \in [n]$ such that $I_i[f] \geqslant \frac{I[f]}{n} \geqslant \frac{1}{2K}\frac{\log n}{n}\mathsf{var}(f)$. Otherwise, by Theorem 1.2, there is a coordinate $i \in [n]$ such that $I_i[f] \geqslant 2^{-K\frac{I[f]}{\mathsf{var}(f)}} \geqslant \frac{1}{\sqrt{n}}$ and we are done either way. Friedgut's Junta Theorem can now be stated as below.

Theorem 1.3. *There exists an absolute constant $K > 0$, such that for any $f\colon \{0,1\}^n \to \{0,1\}$ and $\varepsilon > 0$, the function f is ε-close to a function $g\colon \{0,1\}^n \to \{0,1\}$ (in Hamming distance) that depends on at most $2^{K\frac{I[f]}{\varepsilon}}$ coordinates.*

Morally speaking, Theorem 1.3 states that not only that there is a coordinate with significant influence as in Theorem 1.2, but actually all coordinates that have smaller influence, combined, barely affect the output of the function f (and this is how its proof proceeds). Talagrand's strengthening of the KKL Theorem is stated below.

Theorem 1.4. *There exists an absolute constant $c > 0$, such that for any $f\colon \{0,1\}^n \to \{0,1\}$,*

$$\sum_{i=1}^{n} \frac{I_i[f]}{\log(1/I_i[f])} \geqslant c \cdot \mathsf{var}(f).$$

We note that Theorem 1.4 implies Theorem 1.2 as follows: suppose on the contrary that all influences $I_i[f]$ are at most $2^{-K\frac{I[f]}{\mathsf{var}(f)}}$. Then the "Talagrand sum" as above is at most $\frac{\mathsf{var}(f)}{KI[f]}\sum_{i=1}^{n} I_i[f] = \frac{\mathsf{var}(f)}{K}$, a contradiction for a large enough constant K.

A key technique used in the original proofs of all the theorems above is the hypercontractive inequality (stated in Section 2.3). The use of this inequality is, by now, nearly ubiquitous in analysis of Boolean functions. Still, using this inequality might impose limitations of its own, limiting the discovery of new results, both qualitatively and quantitatively. As far as we know, researchers in this area have wondered whether there is "life" beyond the hypercontractive inequality, and certainly there have been efforts to prove the KKL Theorem (and its strengthenings) without using it. In particular, proofs using "only" the Log-Sobolev inequality (stated in Section 2.3) for the KKL Theorem and Friedgut's Junta Theorem are known [FS07] (their argument though does not seem to extend to Talagrand's Theorem). There is also a recent proof of the KKL Theorem (as well as Talagrand's result and some strengthenings) using stochastic calculus [EG19].

In this paper, we prove Theorems 1.1, 1.2, 1.3, 1.4 using "only" the Log-Sobolev inequality. Since the hypercontractive inequality and the Log-Sobolev inequality are equivalent to each other and both have separate not-so-difficult proofs as well, whether one uses one or the other is, admittedly, splitting hairs. Still, another interesting aspect of this paper is that our proof approach is very different from all earlier proofs. We look at the first Fourier level of the function after a suitable random restriction and apply the Log-Sobolev inequality appropriately. The approach is, in our subjective opinion, more direct, natural, and less mysterious, though the overall proofs are not necessarily "easier". The additional structural information implicit in these proofs might benefit further research. In Section 3, we describe the basic skeleton that is common to all our proofs and the main technical lemma, Lemma 3.5. The paper might have some benefit from expository perspective as all our proofs are uniformly built around the same skeleton.

Before proceeding to formal proofs, we illustrate here the underlying intuition and how it morally explains the KKL Theorem (translating the intuition into a formal proof takes some effort). We assume here that the reader is somewhat familiar with the area and the standard terminology. Let $f : \{0,1\}^n \to \{0,1\}$ be a balanced function and suppose all its influences are at most (say) $\frac{1}{\sqrt{n}}$. We hope to conclude that the total influence is then $\Omega(\log n)$. Suppose f has degree d (we are referring to the so-called average degree, but never mind). Consider a $\frac{1}{d}$-random restriction $f_{\bar{J}\to z}$ of the function where each coordinate stays alive with probability $\frac{1}{d}$ independently and denoting the set of alive coordinates as J, the coordinates in $\bar{J} = [n] \setminus J$ are set to a uniformly random setting z. Since f has degree d, we expect that the restricted function $f_{\bar{J}\to z}$ has constant Fourier weight at the first level and ideally, is even a dictatorship function (indeed, if the Fourier weight at the first level exceeds a certain threshold, a Boolean function is necessarily a dictatorship). Suppose, for the sake of illustration, that the restricted function $f_{\bar{J}\to z}$ is always a dictatorship function. However, it could be the dictatorship of a different coordinate for different settings of z. Let $A_j \subseteq \{0,1\}^{\bar{J}}$ consist of those settings of z for which $f_{\bar{J}\to z}$ is the dictatorship of coordinate $j \in J$. We note that the fractional size of A_j, denoted $\mu(A_j)$, is at most the influence of the coordinate j (why?) and hence $\mu(A_j) \leqslant \frac{1}{\sqrt{n}}$ for all $j \in J$. Now we simply note that since the sets $A_1, \dots, A_{|J|}$ are all polynomially small in size and form a partition of $\{0,1\}^{\bar{J}}$, at least $\frac{\log n}{n}$ fraction of the edges in the hypercube $\{0,1\}^{\bar{J}}$ are across some A_j and $A_{j'}$ with $j \neq j'$. These edges, along with the fact that A_j and $A_{j'}$ are restrictions leading to dictatorships of j and j' respectively, contribute $\Omega(\log n)$ to the total influence of the function f as desired (why?)! We use here the standard isoperimetric result on the hypercube that for a small set $A \subseteq \{0,1\}^n$, at least $\frac{\log(1/\mu(A)))}{n}$ fraction of hypercube edges incident on it, go outside of A (this is also a special case of the Log-Sobolev inequality, see Lemma 2.7).

2 Preliminaries

We denote $[n] = \{1, 2, \dots, n\}$. We write $X \gtrsim Y$ to say that there exists an absolute constant $c > 0$ such that $X \geqslant c \cdot Y$.

2.1 Standard Fourier Analysis

We consider the space of real-valued functions $f : \{0,1\}^n \to \mathbb{R}$, equipped with the inner product $\langle f, g\rangle = \mathbb{E}_{x \in_R \{0,1\}^n}[f(x)g(x)]$. Here and throughout the paper, we consider the uniform distribution over $\{0,1\}^n$. It is well-known that the collection of functions $\chi_S : \{0,1\}^n \to \{-1,1\}$, one for each subset $S \subseteq [n]$, defined as $\chi_S(x) = (-1)^{\oplus_{i\in S} x_i}$, is an orthonormal basis w.r.t. the said inner product. Thus each function $f\colon \{0,1\}^n \to \mathbb{R}$ can be written uniquely as

$$f(x) = \sum_{S\subseteq[n]} \hat{f}(S)\chi_S(x), \qquad\qquad \text{where } \hat{f}(S) = \langle f, \chi_S\rangle.$$

Since the basis $\{\chi_S\}_{S\subseteq[n]}$ is orthonormal, one has the Plancherel/Parseval equality:

Fact 2.1. *For any $f, g\colon \{0,1\}^n \to \mathbb{R}$, we have $\langle f, g\rangle = \sum_{S\subseteq[n]} \hat{f}(S)\hat{g}(S)$. Also*

$$\langle f, f\rangle = \mathop{\mathbb{E}}_x\left[f(x)^2\right] = \|f\|_2^2 = \sum_{S\subseteq[n]} \hat{f}(S)^2.$$

We will also consider other L_p norms of functions for $p \geqslant 1$ (mostly L_1-norm), similarly defined as $\|f\|_p = \left(\mathbb{E}_x\left[|f(x)|^p\right]\right)^{1/p}$. It will be useful to consider the "Fourier weight" of a function on a given "level".

Definition 2.2. *For integer $d \geqslant 1$, the level d Fourier weight of a function $f\colon \{0,1\}^n \to \mathbb{R}$ is $W_{=d}[f] = \sum_{|S|=d} \hat{f}(S)^2$. Also, its Fourier weight on the "chunk" d is $W_{\approx d}[f] = \sum_{d \leqslant j < 2d} W_{=j}[f]$.*

For a noise parameter $\varepsilon \in (0,1)$, the noise operator $T_{1-\varepsilon}$ is defined as follows. For a function $f : \{0,1\}^n \to \mathbb{R}$, the function $T_{1-\varepsilon} f$ is

$$T_{1-\varepsilon} f(x) = \mathop{\mathbb{E}}_{y \sim_\varepsilon x}[f(y)],$$

where the input y is obtained from input x by resembling each coordinate of x with probability ε independently. It is well-known that the Fourier representation of $T_{1-\varepsilon} f$ is

$$T_{1-\varepsilon} f = \sum_{S \subseteq [n]} (1-\varepsilon)^{|S|} \hat{f}(S) \chi_S.$$

2.2 Discrete Derivatives and Influences

For a coordinate $i \in [n]$, the discrete derivatives of f along the i^{th} direction is a function $\partial_i f\colon \{0,1\}^{n-1} \to \mathbb{R}$ defined as

$$\partial_i f(y) = f(x_{-i} = y, x_i = 1) - f(x_{-i} = y, x_i = 0).$$

Definition 2.3. *The L_p-influence of a coordinate $i \in [n]$ is defined as $I_i^p[f] = \|\partial_i f\|_p^p$. The L_p total-influence is $I^p[f] = \sum_{i=1}^{n} I_i^p[f]$. We stress here that in the notation $I_i^p[f]$ and $I^p[f]$ herein, the "p" is a super-script and not an exponent.*

We will be concerned with only L_2 and L_1 influences. In the literature, the notion usually refers to L_2-influences, so in this case the superscript p is omitted, writing $I_i[f] = I_i^2[f]$ and $I[f] = I^2[f]$ for the individual and total influence respectively. We note that for Boolean functions, all the L_p-influences are equal. We will be concerned with the more general case of bounded functions, i.e. functions taking values in the interval $[-1,1]$, and state our variants of Theorems 1.1, 1.2, 1.3, and 1.4 using L_1-influences instead. We remark that for bounded functions, one has $I_i^p[f] \leqslant I_i^q[f]$ for $p \geqslant q \geqslant 1$. In particular and via Cauchy-Schwartz, $I_i[f] \leqslant I_i^1[f] \leqslant \sqrt{I_i[f]}$. Using the Fourier expansion of the discrete derivatives and Parseval equality gives the following standard formula for the total L_2-influence.

Fact 2.4. *For any $f\colon \{0,1\}^n \to \mathbb{R}$, we have $I[f] = 4 \sum_{S \subseteq [n]} |S|\, \hat{f}(S)^2$. In particular, by an averaging argument, for any $\varepsilon > 0$, $\sum_{|S| \geqslant I[f]/\varepsilon} \hat{f}(S)^2 \leqslant \varepsilon$.*

2.3 Hypercontractive Inequality and Log-Sobolev Inequality

The hypercontractive inequality states that for each $\varepsilon > 0$, there is $p > 2$ such that $T_{1-\varepsilon}$ is a contraction from L_2 to L_p, i.e. that $\|T_{1-\varepsilon} f\|_p \leqslant \|f\|_2$ for any $f\colon \{0,1\}^n \to \mathbb{R}$. The inequality has an equivalent form (which is often times used) that does not involve the noise operator $T_{1-\varepsilon}$, and is instead concerned with bounded degree functions.

The degree of a function f, denoted $\mathsf{deg}(f)$, is the maximum of $|S|$ over all S such that $\hat{f}(S) \neq 0$. The Bonami-Beckner hypercontractive inequality [Bon70, Bec75] asserts that the L_p-norm and the L_2-norm of a low-degree function are comparable. More precisely, for any $f : \{0,1\}^n \to \mathbb{R}$ and any $p > 2$,

Theorem 2.5. $\|f\|_p \leqslant (p-1)^{\mathsf{deg}(f)/2} \|f\|_2$.

To motivate the Log-Sobolev inequality and its relationship to the hypercontractive inequality, let us rewrite the above as

$$\mathsf{deg}(f) \geqslant \frac{2}{\log(p-1)} \log\left(\frac{\|f\|_p}{\|f\|_2}\right). \tag{1}$$

Instead of looking at the maximal degree of a non-zero monomial that appears in f, one may consider the average degree of f, defined as $\sum_S |S|\, \hat{f}(S)^2$, where the weight given to a characters S equals the squared Fourier coefficient $\hat{f}(S)^2$. When f is $\{-1,1\}$-valued, the squared Fourier coefficients sum up to 1, giving a probability distribution over them, explaining the term "average degree". As noted, the average degree is same as the total influence $I[f]$ (up to the factor 4). The Log-Sobolev inequality, established by Gross [Gro75], can be seen as the limiting case of the above inequality as $p \to 2$ and replacing the degree by average degree (see [Gro75], [O'D14, Chapter 10.1] and [O'D14, Pages 319-320] for the equivalence between the two inequalities and also separate inductive proofs). Towards stating this inequality, one needs the notion of entropy of a non-negative function $h\colon \{0,1\}^n \to [0,\infty)$:

$$\mathsf{Ent}(h) := \mathop{\mathbb{E}}_x [h(x)] \log\left(\frac{1}{\mathbb{E}[h(x)]}\right) - \mathop{\mathbb{E}}_x \left[h(x) \log\left(\frac{1}{h(x)}\right)\right],$$

with the convention that $0\log(1/0) = 0$. The Log-Sobolev inequality is (note that the entropy is of the non-negative function f^2):

Theorem 2.6. *For any $f\colon \{0,1\}^n \to \mathbb{R}$, we have $I[f] \geqslant \frac{1}{2}\mathsf{Ent}(f^2)$.*

A simple corollary of this inequality, when $f : \{0,1\}^n \to \{0,1\}$ is Boolean, is below. This is also known as the standard isoperimetric inequality for the Boolean hypercube.

Lemma 2.7. *For any $f : \{0,1\}^n \to \{0,1\}, \beta = \mathbb{E}[f] \leqslant \frac{1}{2}$, we have $I[f] \geqslant \frac{1}{2}\beta\log(1/\beta)$.*

It will be more convenient for us to use the following easy consequence of the Log-Sobolev inequality.

Lemma 2.8. *There exists an absolute constant $K > 0$, such that for any $f\colon \{0,1\}^n \to [-1,1]$, we have*

$$I[f] \ \gtrsim\ \|f\|_2^2 \log\left(\frac{1}{\|f\|_2^2}\right) - K \cdot \|f\|_1^{\frac{1}{2}} \|f\|_2.$$

Proof. By Theorem 2.6, $I[f] \gtrsim \mathsf{Ent}(f^2)$, so it is enough to show that the entropy of f^2 is at least the right hand side. Indeed, the first term in the definition of the entropy is precisely $\|f\|_2^2 \log(1/\|f\|_2^2)$. The second term is (using Cauchy-Schwarz and that $t^2\log^2(1/t^2) \lesssim |t|$ for $t \in [-1,1]$)

$$\mathop{\mathbb{E}}_x \left[f(x)^2 \log\left(\frac{1}{f(x)^2}\right)\right] \leqslant \sqrt{\mathop{\mathbb{E}}_x \left[f(x)^2 \log^2\left(\frac{1}{f(x)^2}\right)\right] \mathop{\mathbb{E}}_x [f(x)^2]} \lesssim \sqrt{\mathop{\mathbb{E}}_x [|f(x)|] \mathop{\mathbb{E}}_x [f(x)^2]} = \|f\|_1^{\frac{1}{2}} \|f\|_2.$$

□

2.4 Random Restrictions

Let $J \subseteq [n]$ be a subset of coordinates thought of as "alive" and coordinates in $\bar{J} = [n] \setminus J$ thought of as "restricted". Given a function $f\colon \{0,1\}^n \to \mathbb{R}$ and a setting $z \in \{0,1\}^{\bar{J}}$, we denote by $f_{\bar{J}\to z}$, the restriction of f to the domain $z \times \{0,1\}^J$. More precisely, $f_{\bar{J}\to z}\colon \{0,1\}^J \to \mathbb{R}$ is defined as $f_{\bar{J}\to z}(y) = f(x_{\bar{J}} = z, x_J = y)$. The following standard fact gives the Fourier coefficients of the restricted function:

Fact 2.9. *For any $T \subseteq J$, we have $\hat{f}_{\bar{J}\to z}(T) = \sum_{S\subseteq \bar{J}} \hat{f}(S\cup T)\chi_S(z)$.*

For a parameter $\delta > 0$, a δ-*random restriction* is the function $f_{\bar{J}\to z}$ after choosing J to be a random subset of $[n]$ in which each $j \in [n]$ is included with probability δ independently and choosing $z \in \{0,1\}^{\bar{J}}$ uniformly. Using Fact 2.9 and Parseval, one can easily compute the expectated squared Fourier coefficient of a random restriction and then the expected level d Fourier weight.

Fact 2.10. *Let $f\colon \{0,1\}^n \to \mathbb{R}$ and $T \subseteq J$. Then $\mathbb{E}_z\left[|\hat{f}_{\bar{J}\to z}(T)|^2\right] = \sum_{S\subseteq \bar{J}} \hat{f}(S\cup T)^2$.*

Fact 2.11. *Let $f\colon \{0,1\}^n \to \mathbb{R}$, $d \geqslant 1$ be an integer, and $\delta \in [0,1]$. Let $f_{\bar{J}\to z}$ denote the δ-random restriction. Then*

$$\mathop{\mathbb{E}}_{J,z}\left[W_{=d}[f_{\bar{J}\to z}]\right] = \sum_S \hat{f}(S)^2 \cdot \Pr_J\left[|J\cap S| = d\right].$$

3 A Basic Argument towards the KKL Theorem

In this section, we prove the lemma below. It proves the KKL Theorem in the special case when the function $f : \{0,1\}^n \to [0,1]$ has a constant fraction of its Fourier weight on some "chunk". Alternately, it proves the KKL Theorem at a loss of $\log\log n$ factor. More importantly, the proof illustrates the basic approach underlying all the subsequent proofs.

Lemma 3.1. *Let $f\colon \{0,1\}^n \to [0,1]$ be a function and $d \geqslant 1$ be an integer. Then there exists a coordinate $i \in [n]$ such that $I_i^1[f] \gtrsim \frac{\log n}{n} W_{\approx d}[f]$.*

We make some remarks before proceeding to the proof. Firstly, we note that the lemma holds for bounded functions and with respect to the L_1-influences. Secondly, we note that if a constant fraction of the Fourier weight is on some chunk, i.e. if for some d, $W_{\approx d}[f] \gtrsim \mathsf{var}(f)$, then there is a coordinate $i \in [n]$ with $I_i^1[f] \gtrsim \frac{\log n}{n}\mathsf{var}(f)$, proving the KKL Theorem. Thirdly, we note that it proves the KKL Theorem at a loss of factor $\log\log n$ as follows. We may assume that $I[f] \leqslant \frac{1}{2}\log n\, \mathsf{var}(f)$. Since $\mathsf{var}(f) = \sum_{1\leqslant |S|} \hat{f}(S)^2$ and $I[f] = 4\sum_S |S|\hat{f}(S)^2$, by Markov's inequality, we have

$$\sum_{1\leqslant |S|\leqslant \log n} \hat{f}(S)^2 \geqslant \frac{1}{2}\,\mathsf{var}(f).$$

Thus, by partitioning the interval $[1, \log n]$ into $\log\log n$ dyadic intervals $\bigcup_{k=0}^{\log\log n}[2^k, 2^{k+1})$, it follows that there is some $1 \leqslant d \leqslant \log n$ such that $W_{\approx d} \gtrsim \frac{\mathsf{var}(f)}{\log\log n}$. Hence by the lemma, there is a coordinate $i \in [n]$ such that $I_i^1[f] \gtrsim \frac{\log n}{n} W_{\approx d}[f] \gtrsim \frac{\log n}{n}\frac{\mathsf{var}(f)}{\log\log n}$.

3.1 Proof of Lemma 3.1

We now prove Lemma 3.1. The proof formalizes the intuition described at the end of the introductory section. One begins by considering a $\frac{1}{d}$-random restriction of the function, notes that the expected Fourier weight at the first level of the restricted function is at least $W_{\approx d}[f]$, and then one examines the coefficients at the first level and applies Log-Sobolev appropriately.

Weight on the First Level after Random Restriction.

Let $f_{\bar{J}\to z}$ be a $\frac{1}{d}$-random restriction, J being the set of coordinates left alive. We have by Fact 2.11 that

$$\mathop{\mathbb{E}}_{J,z}[W_{=1}[f_{\bar{J}\to z}]] \geqslant \sum_{d\leqslant|S|<2d} \hat{f}(S)^2 \cdot \Pr_J[|S\cap J|=1] \gtrsim W_{\approx d}[f], \tag{2}$$

where we used the simple fact that for any set S with $d \leqslant |S| < 2d$, the probability it intersects J in a single element is constant. For the rest of the argument, we fix some $J \subseteq [n]$ such that (it exists due to Equation (2))

$$\mathop{\mathbb{E}}_{z}[W_{=1}[f_{\bar{J}\to z}]] \gtrsim W_{\approx d}[f]. \tag{3}$$

Relating First Level Coefficients after Restriction and Influences of f

We now consider the first level coefficients of the restricted function $f_{\bar{J}\to z}$ and somehow relate them to the influences of the original function f. We note that J is the set of alive coordinates. For each $j \in J$, define a function $g_j\colon \{0,1\}^{\bar{J}} \to \mathbb{R}$ by $g_j(z) = \hat{f}_{\bar{J}\to z}(\{j\})$. That is, $g_j(z)$ is the j^{th} coefficient of the first level (= linear part) of the restricted function. By definition, $W_{=1}[f_{\bar{J}\to z}] = \sum_{j\in J} g_j(z)^2$. Let $p_j = \|g_j\|_2^2 = \mathbb{E}_z\left[g_j(z)^2\right]$. For the sake of future reference, let $q_j = \|g_j\|_1$. Thus (3) can be re-stated as

$$\mathop{\mathbb{E}}_{z}[W_{=1}[f_{\bar{J}\to z}]] = \sum_{j\in J} p_j \ \gtrsim W_{\approx d}[f]. \tag{4}$$

Since f is bounded, so is its restriction, and hence $|g_j(z)| \leqslant 1$ for every z, j.

Lemma 3.2. *$p_j = \|g_j\|_2^2$ and $q_j = \|g_j\|_1$ satisfy*

- $q_j = \|g_j\|_1 \leqslant \frac{1}{2} \cdot I_j^1[f]$.
- $p_j = \|g_j\|_2^2 \leqslant \frac{1}{4} \cdot I_j[f]$.
- $p_j \leqslant q_j \leqslant \sqrt{p_j}$.

Proof. The third item is because of the boundedness $|g_j(z)| \leqslant 1$ and Cauchy-Schwartz. Towards the first two items, we note that

$$g_j(z) = \hat{f}_{\bar{J}\to z}(\{j\}) = \mathop{\mathbb{E}}_{y}\left[f(z,y)\chi_{\{j\}}(y_j)\right] = \mathop{\mathbb{E}}_{y_{-j}}\left[\frac{f(z,y_{-j},y_j=0)-f(z,y_{-j},y_j=1)}{2}\right].$$

Taking expectation over z gives (and using Cauchy-Schwartz in the second case)

$$\|g_j\|_1 = \mathop{\mathbb{E}}_{z}[|g_j(z)|] \leqslant \mathop{\mathbb{E}}_{z,y_{-j}}\left[\left|\frac{f(z,y_{-j},y_j=0)-f(z,y_{-j},y_j=1)}{2}\right|\right] = \frac{1}{2}\cdot I_j^1[f].$$

$$\|g_j\|_2^2 = \mathop{\mathbb{E}}_{z}\left[|g_j(z)|^2\right] \leqslant \mathop{\mathbb{E}}_{z,y_{-j}}\left[\left|\frac{f(z,y_{-j},y_j=0)-f(z,y_{-j},y_j=1)}{2}\right|^2\right] = \frac{1}{4}\cdot I_j[f]. \qquad \square$$

Summing the previous inequality over all $j \in J$, we conclude that:

Lemma 3.3. $\sum_{j \in J} q_j = \sum_{j \in J} \|g_j\|_1 \leqslant \frac{1}{2} \cdot I^1[f]$.

The following lower bound on $I[f]$ is a key observation.

Lemma 3.4. $\sum_{j \in J} I[g_j] \leqslant I[f]$.

Proof. We lower bound $I[f]$ by $\sum_{i \in \bar{J}} I_i[f]$. Fix some $i \in \bar{J}$ for now. As before, z and y denote the inputs on the parts $\bar{J}$ and J respectively.

$$I_i[f] = \mathop{\mathbb{E}}_{z,y}\left[|f(z,y) - f(z \oplus e_i, y)|^2\right] = \mathop{\mathbb{E}}_{z}\left[\|f_{\bar{J} \to z} - f_{\bar{J} \to z \oplus e_i}\|_2^2\right].$$

By Parseval, we express the squared norm in terms of Fourier coefficients and then lower bound by considering only coefficients of size one.

$$I_i[f] = \mathop{\mathbb{E}}_{z}\left[\sum_{T \subseteq J} |\hat{f}_{\bar{J} \to z}(T) - \hat{f}_{\bar{J} \to z \oplus e_i}(T)|^2\right] \geqslant \mathop{\mathbb{E}}_{z}\left[\sum_{j \in J} |\hat{f}_{\bar{J} \to z}(\{j\}) - \hat{f}_{\bar{J} \to z \oplus e_i}(\{j\})|^2\right].$$

The latter are simply $g_j(z)$ and $g_j(z \oplus e_i)$ by definition and hence

$$I_i[f] \geqslant \sum_{j \in J} \mathop{\mathbb{E}}_{z}\left[|g_j(z) - g_j(z \oplus e_i)|^2\right] = \sum_{j \in J} I_i[g_j].$$

Summing over $i \in \bar{J}$ gives

$$I[f] \geqslant \sum_{i \in \bar{J}} I_i[f] \geqslant \sum_{i \in \bar{J}} \sum_{j \in J} I_i[g_j] = \sum_{j \in J} \sum_{i \in \bar{J}} I_i[g_j] = \sum_{j \in J} I[g_j].$$

$\square$

The Main Argument

Our main argument tries to obtain a lower bound on $I[f]$ as follows. Using Lemma 3.4 and the Log-Sobolev Lemma 2.8,

$$I[f] \geqslant \sum_{j \in J} I[g_j] \geqslant \sum_{j \in J} \left(p_j \log(1/p_j) - K\sqrt{q_j} \cdot \sqrt{p_j}\right).$$

Using Cauchy-Schwartz, we get

$$I[f] \geqslant \sum_{j \in J} p_j \log(1/p_j) - K\sqrt{\sum_{j \in J} q_j} \cdot \sqrt{\sum_{j \in J} p_j}.$$

By Lemma 3.3, $\sum_{j \in J} q_j \leqslant I^1[f]$, so we get our main technical inequality

$$I[f] \geqslant \sum_{j \in J} p_j \log(1/p_j) - K\sqrt{I^1[f]} \cdot \sqrt{\sum_{j \in J} p_j}. \tag{5}$$

We recall that $p_j \leqslant \frac{1}{2} I_j^1[f]$ by Lemma 3.2. Letting $W := \sum_{j \in J} p_j = \mathbb{E}_z\left[W_{=1}[f_{\bar{J} \to z}]\right]$, we rewrite this inequality, for future reference, as below. We note that in application, J is the subset of alive coordinates after a random restriction. In our proof of KKL and Friedgut Theorems, the set J is fixed so as to maximizes the expected first level Fourier weight. In the proof of Talagrand Theorem, we average over the choice of J as well.

Lemma 3.5. *Let $f\colon \{0,1\}^n \to [0,1]$ be a function and $J \subseteq [n]$. Then*

$$I[f] \geqslant \log\left(\frac{1}{\max_{j\in J} I^1_j[f]}\right)\cdot W - K\sqrt{I^1[f]}\cdot\sqrt{W}, \qquad W = \mathop{\mathbb{E}}_{z}\left[W_{=1}[f_{\bar{J}\to z}]\right] = \sum_{S\subseteq[n], |S\cap J|=1} \hat{f}(S)^2.$$

The proof of Lemma 3.1 is now completed immediately. We may assume that for all coordinates $i \in [n]$, $I^1_i[f] \leqslant \frac{\log n}{n}\, W_{\approx d}[f] \leqslant \frac{1}{\sqrt{n}}$ as otherwise we are done already. This implies that the total L_1-influence $I^1[f] \leqslant \log n\; W_{\approx d}[f]$. Lemma 3.5 (= Equation (5)) then gives (the log-factor therein is at least $\frac{1}{2}\log n$ since all L_1-influences are at most $\frac{1}{\sqrt{n}}$)

$$I[f] \geqslant \frac{1}{2}\log n\cdot W - K\sqrt{\log n\cdot W_{\approx d}[f]}\cdot\sqrt{W}, \qquad\qquad W \gtrsim W_{\approx d}[f].$$

Clearly, the first term above dominates the second, giving $I[f] \geqslant \frac{1}{4}\log n\cdot W \gtrsim \log n\cdot W_{\approx d}[f]$, implying now that there is a coordinate with in fact L_2-influence $\gtrsim \frac{\log n}{n}\, W_{\approx d}[f]$.

4 The KKL Theorem

We now prove the KKL Theorem, stated below for a bounded function, with respect to L_1-influences, and in a slightly different form.

Theorem 4.1. *There exists an absolute constant $c > 0$ such that the following holds. Let $f\colon \{0,1\}^n \to [0,1]$ be a function. Then either $I^1[f] \geqslant c\cdot\log n\,\mathsf{var}(f)$, or there is a coordinate $i \in [n]$ such that $I^1_i[f] \geqslant \frac{1}{\sqrt{n}}$.*

It will be more convenient for us to prove a dimension-independent version of the KKL Theorem below. It is easily seen to imply the statement above.

Theorem 4.2. *There exists an absolute constant $C > 0$ such that the following holds. Let $f\colon \{0,1\}^n \to [0,1]$ be a function. Then there is a coordinate $i \in [n]$ such that $I^1_i[f] \geqslant 2^{-C\cdot\frac{I^1[f]}{\mathsf{var}(f)}}$.*

In the proof of Lemma 3.1, we only "utilized" Fourier weight from a single chunk of Fourier coefficients, i.e. those of size in the range $[d, 2d)$, and this led to a loss of factor $\log\log n$ if used towards the KKL Theorem. In this section, we show how to utilize and combine the Fourier weight from multiple chunks, avoiding this loss. The idea is to "partition" f into chunks as $f = \hat{f}(\emptyset) + \sum_{d=2^k, k\geqslant 0} h^*_d$, apply the main technical inequality (5) to each chunk h^*_d, and then "sum up". A natural way to partition is to let $h^*_d = \sum_{d\leqslant|S|<2d}\hat{f}(S)\chi_S$. The problem with this approach however is that the chunk functions h^*_d as here are not necessarily bounded functions and the earlier arguments cannot be applied directly. To get around this, we instead consider a soft notion of chunks, $f \approx \hat{f}(\emptyset) + \sum_{d=2^k, k\geqslant 0} h_d$, that behaves similarly, that is

$$\sum_d \mathsf{var}(h_d) = \Theta(\mathsf{var}(f)), \qquad \sum_d I_i[h_d] = \Theta(I_i[f]), \qquad \sum_d I[h_d] = \Theta(I[f]),$$

and in addition, preserves boundedness and the L_1-influences of each soft chunk h_d are bounded by those of the original function!

4.1 Soft Chunks

Definition 4.3. *Let* $f\colon \{0,1\}^n \to [0,1]$ *be a function and let* $d \geqslant 1$ *be integer (thought of as a power of* 2*). The soft chunk of* f *of degree* d *is given by the function* $h_d\colon \{0,1\}^n \to [0,1]$ *defined by* $h_d = (T_{1-\frac{1}{2d}} - T_{1-\frac{1}{d}})f$.

The following lemma summarizes the useful properties of soft chunks (the proof appears in Appendix A.1). We point out, in particular, that the L_1-influences of the soft chunk are upper bounded by those of the original function (up to a factor 2).

Lemma 4.4. *Let* $f\colon \{0,1\}^n \to [0,1]$ *and for integer* $d = 2^k, k \geqslant 0$*, let* $h_d = (T_{1-\frac{1}{2d}} - T_{1-\frac{1}{d}})f$ *denote the soft chunk of* f *of degree* d*. Then (the sums are over* $d = 2^k, k \geqslant 0$ *and* $i \in [n]$ *is arbitrary)*

- h_d *is bounded in* $[-1,1]$*,* $\hat{h}(\emptyset) = 0$.
- $I_i^1[h_d] \leqslant 2I_i^1[f]$.
- *For any* $S \subseteq [n]$*,* $d \leqslant |S| < 2d$*, we have* $|\hat{f}(S)| \lesssim |\hat{h}_d(S)| \leqslant |\hat{f}(S)|$*. In particular, we have lower bounds*
$$\|h_d\|_2^2 \geqslant W_{\approx d}[h_d] \gtrsim W_{\approx d}[f], \quad \sum_d I_i[h_d] \gtrsim I_i[f], \quad \sum_d I[h_d] \gtrsim I[f].$$
- *And the upper bounds,*
$$\sum_d \|h_d\|_2^2 \leqslant \mathsf{var}(f), \quad \sum_d I[h_d] \leqslant I[f].$$

For technical reasons, we will be able to "utilize" only those chunks that have a significant amount of Fourier weight, referred to as the good chunks. It will turn out that the good chunks still capture a constant fraction of the variance of f, so this will not be a problem. Towards this end, we have (proof appears in Appendix A.2)

Lemma 4.5. *Let*
$$D_{\mathsf{good}} := \left\{ d = 2^k, k \geqslant 0 \mid \ W_{\approx d}[f] \geqslant \frac{\mathsf{var}(f)^2}{16 \cdot I^1[f]} \right\}.$$
Then
$$\sum_{d \in D_{\mathsf{good}}} W_{\approx d}[f] \gtrsim \mathsf{var}(f).$$

4.2 Proof of Theorem 4.2

Assume, for the sake of contradiction, that for all coordinates $i \in [n]$, $I_i^1[f] \leqslant 2^{-C \cdot \frac{I^1[f]}{\mathsf{var}(f)}}$ where C is a large enough constant chosen later. Let $h_d, d \in D_{\mathsf{good}}$ be any good soft chunk. We recall that

- h_d is a bounded function.
- Its L_1-influences are upper bounded by those of f up to a factor 2 (and hence also the total L_1-influence).
- $W_{\approx d}[h_d] \gtrsim W_{\approx d}[f] \geqslant \frac{\mathsf{var}(f)^2}{16 \cdot I^1[f]}$.

We apply Lemma 3.5 to the function h_d, considering $\frac{1}{d}$-random restriction, and letting J to be the subset of alive coordinates (fixed so as to maximize expected weight at first Fourier level). This yields the inequality

$$I[h_d] \geqslant \log\left(\frac{1}{\max_{j\in J} I^1_j[h_d]}\right)\cdot W - K\sqrt{I^1[h_d]}\cdot\sqrt{W}, \qquad W \gtrsim W_{\approx d}[h_d] \gtrsim W_{\approx d}[f] \geqslant \frac{\mathsf{var}(f)^2}{16\cdot I^1[f]}.$$

Since the L_1-influences of h_d are bounded by those of f, in particular all of them at most $2^{-C\cdot\frac{I^1[f]}{\mathsf{var}(f)}}$, we get

$$I[h_d] \ \geqslant\ C\cdot\frac{I^1[f]}{\mathsf{var}(f)}\cdot W - K\sqrt{I^1[f]}\cdot\sqrt{W}, \qquad W \gtrsim W_{\approx d}[f] \geqslant \frac{\mathsf{var}(f)^2}{16\cdot I^1[f]}.$$

It is easily seen that for a large enough constant C, the first term dominates the second term (this is why we considered only the good chunks) and thus

$$I[h_d] \ \gtrsim\ C\cdot\frac{I^1[f]}{\mathsf{var}(f)}\cdot W_{\approx d}[f].$$

Now summing over all good d gives a contradiction:

$$I^1[f] \geqslant I[f] \geqslant \sum_{d\in D_{\mathsf{good}}} I[h_d] \ \gtrsim\ C\cdot\frac{I^1[f]}{\mathsf{var}(f)}\sum_{d\in D_{\mathsf{good}}} W_{\approx d}[f] \ \gtrsim\ C\cdot\frac{I^1[f]}{\mathsf{var}(f)}\cdot\mathsf{var}(f) = C\cdot I^1[f].$$

We used Lemma 4.4 in the second step and Lemma 4.5 in the second-last step. Taking the constant C large enough gives a contradiction.

5 The Friedgut's Junta Theorem

Friedgut's Junta Theorem (restated below) is proved by a careful adjustment to the argument in the previous section.

Theorem 5.1. *There is an absolute constant $C > 0$ such that the following holds. For every function $f\colon \{0,1\}^n \to [0,1]$ and for every $\varepsilon > 0$, there exists a function $g\colon \{0,1\}^n \to [0,1]$ depending on at most $2^{C\cdot I^1[f]/\varepsilon}$ variables such that $\|f-g\|_2^2 \lesssim \varepsilon$.*

We provide a proof sketch. While in the proof of the KKL Theorem, we may assume that all influences are small, this is not the case with Friedgut's Theroem. Here we "separate out" the set L of coordinates with "non-negligible" influence and apply the previous argument to the remaining set $\bar{L} = [n]\setminus L$. Towards this end, let

$$L = \left\{ i \mid I^1_i[f] \geqslant \tau := 2^{-C\cdot I^1[f]/\varepsilon} \right\}.$$

Clearly, $|L| \leqslant \frac{I^1[f]}{\tau} \leqslant 2^{2C\cdot I^1[f]/\varepsilon}$. Let $g = \sum_{S\subseteq L} \hat{f}(S)\chi_S$. It is easily observed that

- g depends only on the coordinates of L.
- g is also bounded in $[0,1]$ since g is simply the average of f over coordinates in $\bar{L}$ and
- for the same reason, L_1-influences of g are bounded by those of f.

Let $\varphi = f - g$. We will show that $\|\varphi\|_2^2 \lesssim \varepsilon$. Clearly, φ is bounded in $[-1, 1]$ and its L_1 influences are also bounded by those of f up to a factor 2. We intend to apply the same argument used to prove the KKL Theorem to φ, except that all "action" happens only on the set of coordinates $\bar{L}$. More specifically:

- The "size" of any Fourier term is counted as $|S \cap \bar{L}|$ instead of as $|S|$.
- For an integer $d \geqslant 1$ (thought of as power of 2), the Fourier weight on the corresponding chunk is defined as
$$W^{\bar{L}}_{\approx d}[\varphi] := \sum_{d \leqslant |S \cap \bar{L}| < 2d} \hat{f}(S)^2.$$
- Towards defining the soft chunk h_d of φ, the noise operator is applied only to coordinates in $\bar{L}$. We denote this as
$$h_d = \left(T^{\bar{L}}_{1-\frac{1}{2d}} - T^{\bar{L}}_{1-\frac{1}{d}} \right) \varphi.$$
- In a random restriction, only coordinates in $\bar{L}$ may stay alive. That is, a $\frac{1}{d}$-random restriction amounts to letting J to be a random subset of $\bar{L}$ where every coordinate in $\bar{L}$ is included with probability $\frac{1}{d}$ and then the coordinates outside J (including those in L) are set uniformly at random.
- Since $J \subseteq \bar{L}$, we have $I^1_j[\varphi] \leqslant 2^{-C \cdot I^1[f]/\varepsilon}$ for all $j \in J$.

Modulo these considerations, we repeat the proof in Section 4.2. We apply Lemma 3.5 to the function h_d, considering $\frac{1}{d}$-random restriction, and letting J be the subset of alive coordinates (fixed so as to maximize expected weight at first Fourier level). This yields the inequality

$$I[h_d] \geqslant \log \left(\frac{1}{\max_{j \in J} I^1_j[h_d]} \right) \cdot W - K\sqrt{I^1[h_d]} \cdot \sqrt{W}, \qquad W \gtrsim W^{\bar{L}}_{\approx d}[h_d] \gtrsim W^{\bar{L}}_{\approx d}[\varphi].$$

Since the L_1-influences of h_d are bounded by those of φ which are in turn bounded by those of f and those for coordinates in $J \subseteq \bar{L}$ are at most $2^{-C \cdot \frac{I^1[f]}{\varepsilon}}$, we get

$$I[h_d] \geqslant C \cdot \frac{I^1[f]}{\varepsilon} \cdot W - K\sqrt{I^1[f]} \cdot \sqrt{W}, \qquad W \gtrsim W^{\bar{L}}_{\approx d}[\varphi].$$

Let D_{good} be the subset of $d = 2^k$ such that $W^{\bar{L}}_{\approx d}[\varphi] \geqslant \frac{\varepsilon^2}{16 I^1[f]}$ so that for such good d and for large enough constant C, the first term above dominates the second and we get

$$I[h_d] \gtrsim C \cdot \frac{I^1[f]}{\varepsilon} W^{\bar{L}}_{\approx d}[\varphi].$$

Now summing over all good $d \in D_{\mathsf{good}}$ gives:

$$I^1[f] \geqslant I^1[\varphi] \geqslant I[\varphi] \geqslant \sum_{d \in D_{\mathsf{good}}} I[h_d] \gtrsim C \cdot \frac{I^1[f]}{\varepsilon} \cdot \sum_{d \in D_{\mathsf{good}}} W^{\bar{L}}_{\approx d}[\phi].$$

By Lemma 4.5 (applied to φ), the last sum is at least $\gtrsim \mathsf{var}(\varphi)$ and we get $\mathsf{var}(\varphi) \lesssim \varepsilon$ as desired. An astute reader might object that the definition of the good soft chunks here seems different than that in Lemma 4.5, i.e. the threshold is set at $\frac{\varepsilon^2}{16 I^1[f]}$ instead of $\frac{\mathsf{var}(\varphi)^2}{16 I^1[\varphi]}$ therein. However since $I^1[\varphi] \leqslant I^1[f]$ and we could assume a priori that $\mathsf{var}(\varphi) \geqslant \varepsilon$ (otherwise we would already be done), this slight difference only works in our favor.

6 The Talagrand's Theorem

In this section, we prove Talagrand's Theorem, restated as Theorem 6.2 later. For now we prove the following weaker theorem to illustrate the main idea.

Theorem 6.1. *Let any $f\colon \{0,1\}^n \to [0,1]$ be a function and $d \geqslant 1$ an integer (thought of as power of 2). Then one of these two conclusions holds:*

- *(Case 1):* $\sum\limits_{j\in[n]} \frac{I_j[f]}{\log(1/I_j^1[f])} \gtrsim \frac{d(W_{\approx d}[f])^2}{I[f]}$.
- *(Case 2):* $\sum\limits_{j\in[n]} \frac{I_j^1[f]}{\log(1/I_j^1[f])} \gtrsim dW_{\approx d}[f]$.

We make a few remarks. On the left hand side of the inequalities, what appear in the numerators are the L_2-influences in Case 1 and L_1-influences in Case 2. This distinction will be important later. In both cases, in the denominator, it does not matter whether we write L_1 or L_2 influences since their logarithms are the same up to a factor 2 (since $I_j[f] \leqslant I_j^1[f] \leqslant \sqrt{I_j[f]}$). If one pretends that all non-zero Fourier coefficients of f have size between d and $2d$, we have $W_{\approx d}[f] = \mathsf{var}(f)$ and $I[f] = \Theta(d \cdot \mathsf{var}(f))$ and we get $\mathsf{var}(f)$ on the right hand side in Case 1 and (even better) $d \cdot \mathsf{var}(f)$ in Case 2, giving Talagrand's Theorem.

We now prove Theorem 6.1. Consider a $\frac{1}{d}$-random restriction as in Section 3 letting J to be the set of coordinates alive. As therein, let $g_j(z) = \hat{f}_{\bar{J}\to z}(\{j\})$, $p_j = \|g_j\|_2^2$, $q_j = \|g_j\|_1$. Unlike therein however, we will not fix the set J and instead take expectation over its choice. Exactly as in Equation (5), we get

$$I[f] \geqslant \sum_{j\in J} p_j \log(1/p_j) - K \sum_{j\in J} \sqrt{q_j}\sqrt{p_j}.$$

We now divide into two cases depending on whether or not, on the right hand side, the first term dominates the second. It will be more convenient to do this after considering expectation over choice of J.

Case 1: $\mathbb{E}_J\left[\sum_{j\in J} p_j \log(1/p_j)\right] \geqslant 2 \cdot \mathbb{E}_J\left[K \sum_{j\in J} \sqrt{q_j}\sqrt{p_j}\right]$.

In this case, we get

$$I[f] \;\gtrsim\; \mathop{\mathbb{E}}_J\left[\sum_{j\in J} p_j \log(1/p_j)\right].$$

Cauchy-Schwartz gives,

$$\mathop{\mathbb{E}}_J\left[\sum_{j\in J} \frac{p_j}{\log(1/p_j)}\right] \cdot \mathop{\mathbb{E}}_J\left[\sum_{j\in J} p_j \log(1/p_j)\right] \geqslant \left(\mathop{\mathbb{E}}_J\left[\sum_{j\in J} p_j\right]\right)^2.$$

The second term is bounded by $I[f]$ (as above) and on the right hand side we have, $\mathbb{E}_J\left[\sum_{j\in J} p_j\right] \gtrsim W_{\approx d}[f]$. This gives

$$\mathop{\mathbb{E}}_J\left[\sum_{j\in J} \frac{p_j}{\log(1/p_j)}\right] \;\gtrsim\; \frac{(W_{\approx d}[f])^2}{I[f]}.$$

Replacing p_j by its upper bound $I_j[f]$ in the numerator and its upper bound $I_j^1[f]$ in the denominator, and noting that each coordinate appears in J with probability $\frac{1}{d}$, gives the desired inequality

$$\sum_{j\in[n]} \frac{I_j[f]}{\log(1/I_j^1[f])} \gtrsim \frac{d(W_{\approx d}[f])^2}{I[f]}.$$

Case 2: $\mathbb{E}_J\left[\sum_{j\in J}\sqrt{q_j}\sqrt{p_j}\right] \gtrsim \mathbb{E}_J\left[\sum_{j\in J} p_j \log(1/p_j)\right]$.

In this case, Cauchy-Schwartz gives

$$\mathop{\mathbb{E}}_J\left[\sum_{j\in J}\frac{q_j}{\log(1/p_j)}\right]\cdot\mathop{\mathbb{E}}_J\left[\sum_{j\in J}p_j\log(1/p_j)\right] \geqslant \left(\mathop{\mathbb{E}}_J\left[\sum_{j\in J}\sqrt{q_j}\sqrt{p_j}\right]\right)^2 \gtrsim \left(\mathop{\mathbb{E}}_J\left[\sum_{j\in J}p_j\log(1/p_j)\right]\right)^2.$$

Canceling $\mathbb{E}_J\left[\sum_{j\in J} p_j\log(1/p_j)\right]$ from both sides gives

$$\mathop{\mathbb{E}}_J\left[\sum_{j\in J}\frac{q_j}{\log(1/p_j)}\right] \gtrsim \mathop{\mathbb{E}}_J\left[\sum_{j\in J}p_j\log(1/p_j)\right] \geqslant \mathop{\mathbb{E}}_J\left[\sum_{j\in J}p_j\right].$$

As before, the right hand side is $\gtrsim W_{\approx d}[f]$, and q_j, p_j are upper bounded by $I_j^1[f]$, and each coordinate appears in J with probability $\frac{1}{d}$. This gives the desired inequality

$$\sum_{j\in[n]} \frac{I_j^1[f]}{\log(1/I_j^1[f])} \gtrsim d\, W_{\approx d}[f].$$

6.1 Talagrand's Theorem by Combining Chunks: First Attempt

We (re-)state Talagrand's Theorem below.

Theorem 6.2. *For any $f\colon\{0,1\}^n\to[0,1]$, we have $\sum\limits_{j\in[n]}\frac{I_j^1[f]}{\log(1/I_j^1[f])} \gtrsim \mathsf{var}(f)$.*

We attempt to prove this result by splitting

$$f = \hat{f}(\emptyset) + \sum_{d=2^k,k\geqslant 0} h_d,$$

where $h_d = \sum_{d\leqslant|S|<2d}\hat{f}(S)\chi_S$ are the chunks of f. The strategy is to apply Theorem 6.1 to each chunk h_d separately and "sum up" or "combine" the outcomes. A crucial observation is that the L_2-influences indeed sum up, that is

$$I_i[f] = \sum_d I_i[h_d].$$

This strategy (almost) works with a careful consideration of whether the Case 1 or the Case 2 applies for different chunks. The catch, as before, is that the chunks h_d are not necessarily bounded functions and their L_1-influences might not be under control. To get around this issue, we instead work with the soft chunks as before and the full proof is completed in the next sub-section. For now, we pretend that the chunks h_d are bounded functions and see how the proof proceeds. We also pretend that the L_1-influences of h_d are upper bounded by those of f (both these conditions do hold when soft chunks are considered!).

We apply Theorem 6.1 to h_d. Noting that $W_{\approx d}[h_d] \geqslant W_{\approx d}[f]$, $I[h_d] = \Theta(d \cdot W_{\approx d}[f])$, and that L_1-influences of h_d are upper bounded by those of f, we conclude that for every $d = 2^k, k \geqslant 0$, one of these conclusions holds (perhaps both conclusions hold and if so, we pick one arbitrarily):

- (Case 1): $\sum_{j \in [n]} \frac{I_j[h_d]}{\log(1/I_j^1[f])} \gtrsim W_{\approx d}[f]$. Let D' be the set of such d.
- (Case 2): $\sum_{j \in [n]} \frac{I_j^1[f]}{\log(1/I_j^1[f])} \gtrsim d\, W_{\approx d}[f]$. Let D'' be the set of such d.

Now we complete the proof as follows. Since $\mathsf{var}(f) = \sum_{d \in D'} W_{\approx d}[f] + \sum_{d \in D''} W_{\approx d}[f]$, either of the two sums is at least $\frac{1}{2}\mathsf{var}(f)$. If the first sum is, then (crucially using the fact that L_2-influences sum up)

$$\sum_{j \in [n]} \frac{I_j[f]}{\log(1/I_j^1[f])} \geqslant \sum_{j \in [n]} \sum_{d \in D'} \frac{I_j[h_d]}{\log(1/I_j^1[f])} = \sum_{d \in D'} \sum_{j \in [n]} \frac{I_j[h_d]}{\log(1/I_j^1[f])} \gtrsim \sum_{d \in D'} W_{\approx d} \gtrsim \mathsf{var}(f),$$

as desired. Otherwise, we may assume $\sum_{d \in D''} W_{\approx d}[f] \geqslant \frac{1}{2}\mathsf{var}(f)$. Since d ranges only over powers of 2, it follows that there is some $d \in D''$ such that $d\, W_{\approx d}[f] \gtrsim \mathsf{var}(f)$ (why!). Using this particular choice of d in the Case 2 above, we get as desired

$$\sum_{j \in [n]} \frac{I_j^1[f]}{\log(1/I_j^1[f])} \gtrsim \mathsf{var}(f).$$

6.2 Talagrand's Theorem by Combining Soft Chunks

We now complete the proof of Talagrand's Theorem 6.2. We carry out the same proof as in the previous sub-section, except that we use the soft chunks $h_d = (T_{1-\frac{1}{2d}} - T_{1-\frac{1}{d}})f$. It holds that $W_{\approx d}[h_d] = \Theta(W_{\approx d}[f])$. However one place we need to be careful about is that we required that $I[h_d] = \Theta(d\, W_{\approx d}[f])$. This need not be true in general. Hence we restrict ourselves to only those $d \in D_{\mathsf{good}}$ for which this condition holds. The lemma below shows that there is still a constant fraction of variance on these good chunks and this is enough to complete the proof (the sets D' and D'' above are subsets of D_{good} now).

Lemma 6.3. *Let*

$$D_{\mathsf{good}} = \left\{ d = 2^k, k \geqslant 0 \mid\ d\, W_{\approx d}[f] \geqslant \frac{1}{40} I[h_d] \right\}.$$

Then

$$\sum_{d \in D_{\mathsf{good}}} W_{\approx d}[f] \geqslant \frac{1}{2}\mathsf{var}(f).$$

Proof. As we will see, it suffices to show that $\sum_d \frac{I[h_d]}{d} \leqslant 20\,\mathsf{var}(f)$. To see that, as $\mathsf{var}(f) = \sum_{S \neq \emptyset} \hat{f}(S)^2$, it is enough to show that for each $S \neq \emptyset$, the term $\hat{f}(S)^2$ appears in the sum $\sum_d \frac{I[h_d]}{d}$ with a multiplicative factor of at most 20. Note that this factor is

$$|S| \sum_d \frac{1}{d} \left(\left(1 - \frac{1}{2d}\right)^{|S|} - \left(1 - \frac{1}{d}\right)^{|S|} \right)^2.$$

We analyze the contribution from $d \leqslant |S|$ and $d > |S|$ separately, showing that each one of them contributes at most $\frac{10}{|S|}$. Let k be such that $2^k \leqslant |S| < 2^{k+1}$. The first part is bounded as

$$\sum_{d \leqslant |S|} \frac{1}{d}(1 - \frac{1}{2d})^{2|S|} \leqslant \sum_{d \leqslant |S|} \frac{1}{d} e^{-|S|/d} \leqslant \sum_{j=0}^{k} \frac{1}{2^j} e^{-2^{k-j}} \leqslant 2^{-k} \sum_{\ell=0}^{\infty} 2^{\ell} e^{-2^{\ell}} \leqslant 5 \cdot 2^{-k} \leqslant \frac{10}{|S|}.$$

The second part is bounded as (we approximate $1 - r\alpha \leqslant (1-\alpha)^r$ in this range).

$$\sum_{d > |S|} \frac{1}{d} \left(1 - (1 - \frac{1}{d})^{|S|}\right)^2 \leqslant \sum_{d > |S|} \frac{1}{d} \leqslant \sum_{r=k+1} \frac{1}{2^r} = \frac{2}{2^{k+1}} \leqslant \frac{2}{|S|}.$$

This shows that $\sum_d \frac{I[h_d]}{d} \leqslant 20\,\mathsf{var}(f)$. Now we complete the proof of the lemma as:

$$\begin{aligned}
\sum_{d \in D_{\mathsf{good}}} W_{\approx d}[f] &= \sum_d W_{\approx d}[f] - \sum_{d \notin D_{\mathsf{good}}} W_{\approx d}[f] \\
&\geqslant \mathsf{var}(f) - \frac{1}{40} \sum_{d \notin D_{\mathsf{good}}} \frac{I[h_d]}{d} \\
&\geqslant \mathsf{var}(f) - \frac{1}{40} \cdot 20 \cdot \mathsf{var}(f) \geqslant \frac{\mathsf{var(f)}}{2}.
\end{aligned}$$

□

References

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

[Bon70] Aline Bonami. étude des coefficients de Fourier des fonctions de $l^p(g)$. *Annales de l'Institut Fourier*, 20(2):335–402, 1970.

[CKK+06] Shuchi Chawla, Robert Krauthgamer, Ravi Kumar, Yuval Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. *Computational Complexity*, 15(2):94–114, 2006.

[DF09] Irit Dinur and Ehud Friedgut. Intersecting families are essentially contained in juntas. *Combinatorics, Probability & Computing*, 18(1-2):107–122, 2009.

[DKSV06] Nikhil R. Devanur, Subhash Khot, Rishi Saket, and Nisheeth K. Vishnoi. Integrality gaps for sparsest cut and minimum linear arrangement problems. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, Seattle, WA, USA, May 21-23, 2006*, pages 537–546, 2006.

[DS05] Irit Dinur and Samuel Safra. On the hardness of approximating minimum vertex cover. *Annals of mathematics*, pages 439–485, 2005.

[EG19] Ronen Eldan and Renan Gross. Stability of talagrand's influence inequality. *CoRR*, abs/1909.12067, 2019.

[FK96] Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American mathematical Society*, 124(10):2993–3002, 1996.

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

[FS07] Dvir Falik and Alex Samorodnitsky. Edge-isoperimetric inequalities and influences. *Combinatorics, Probability & Computing*, 16(5):693–712, 2007.

[GKK+08] Dmitry Gavinsky, Julia Kempe, Iordanis Kerenidis, Ran Raz, and Ronald de Wolf. Exponential separation for one-way quantum communication complexity, with applications to cryptography. *SIAM J. Comput.*, 38(5):1695–1708, 2008.

[Gro75] Leonard Gross. Logarithmic sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.

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

[KKM+16] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, and R. Urbankez. Comparing the bit-map and block-map decoding thresholds of reed-muller codes on bms channels. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 1755–1759, July 2016.

[KR08] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within 2-epsilon. *J. Comput. Syst. Sci.*, 74(3):335–349, 2008.

[KR09] Robert Krauthgamer and Yuval Rabani. Improved lower bounds for embeddings intol$_1$\$. *SIAM J. Comput.*, 38(6):2487–2498, 2009.

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

[OS07] Ryan O'Donnell and Rocco A. Servedio. Learning monotone decision trees in polynomial time. *SIAM J. Comput.*, 37(3):827–844, 2007.

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

A Missing Proofs

A.1 Proof of Lemma 4.4

Towards the first property, we note that both functions $T_{1-1/2d}f$ and $T_{1-1/d}f$, being averages of f, are bounded in the interval $[0,1]$. Towards the second property, we note that $I_i^1[h_d] \leqslant I_i^1[T_{1-1/2d}f]+I_i^1[T_{1-1/d}f]$ and that the latter are at most $I_i^1[f]$, again because $T_{1-1/2d}f$ and $T_{1-1/d}f$ are averages of f. Towards the third property, we note that by definition

$$\hat{h}_d(S) = \left(\left(1-\frac{1}{2d}\right)^{|S|} - \left(1-\frac{1}{d}\right)^{|S|}\right)\hat{f}(S).$$

The multiplicative factor in front of $\hat{f}(S)$, when $d \leqslant |S| \leqslant 2d$, is easily seen to be a constant. Towards the last property, we note that for each set $S \neq \emptyset$, its contribution to $\mathsf{var}(f)$ is $\hat{f}(S)^2$ and to $I[f]$ is $|S|\hat{f}(S)^2$, whereas the corresponding contributions to $\sum_d \|h_d\|_2^2$ and $\sum_d I[h_d]$ are similar up to the multiplicative factor

$$\sum_d \left(\left(1-\frac{1}{2d}\right)^{|S|} - \left(1-\frac{1}{d}\right)^{|S|}\right)^2.$$

It is enough to show that this sum is at most 1. Indeed, since each summand is square of a number in the range $[0,1]$, we can ignore the squares and then it is just a telescoping sum upper bounded by 1.

A.2 Proof of Lemma 4.5

In the following, sums run over all d that are powers of 2 unless the sum is restricted explicitly to a subset. Clearly, $\sum_d W_{\approx d}[f] = \mathsf{var}(f)$, so it is enough to show that this sum over only those $d \notin D_{\mathsf{good}}$ is at most $\frac{1}{2}\mathsf{var}(f)$. We consider two cases: those d that are "large", that is $d \geqslant T = \frac{4I^1[f]}{\mathsf{var}(f)}$, and those d that are "not large" but not in D_{good}. In the first case, we use Markov and in the second case, we note that there are only a few summands. Indeed, in the first case (using $I[f] \leqslant I^1[f]$),

$$I[f] \;\geqslant\; T \cdot \sum_{d \geqslant T} W_{\approx d}[f], \quad \text{implying that} \quad \sum_{d \geqslant T} W_{\approx d}[f] \;\leqslant\; \frac{I[f]}{T} = \frac{I[f]\,\mathsf{var}(f)}{4I^1[f]} \;\leqslant\; \frac{1}{4}\mathsf{var}(f).$$

In the second case, $d \leqslant T$, so there are at most $\log T$ chunks and when $d \notin D_{\mathsf{good}}$, we have $W_{\approx d}[f] \leqslant \frac{\mathsf{var}(f)^2}{16 \cdot I^1[f]} = \frac{\mathsf{var}(f)}{4T}$. Hence

$$\sum_{d \leqslant T, d \notin D_{\mathsf{good}}} W_{\approx d}[f] \leqslant \log T \cdot \frac{1}{4T} \cdot \mathsf{var}(f) \leqslant \frac{1}{4}\mathsf{var}(f).$$

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il