COMP760, LECTURES 9-11: HYPERCONTRACTIVITY, FRIEDGUT'S THEOREM, KKL INEQUALITY, CHANG'S LEMMA

HAMED HATAMI

We start the study of an important tool in harmonic analysis. Namely, the *hypercontractivity* of the noise operator. The ideas and results were developed by different people (Bonami, Beckner, Ornstein-Uhlenbeck, Gross, Nelson) in different contexts.

1. The noise operator

We begin by formally introducing the noise operator. Let μ_p denote the Bernoulli distribution with success probability p (that is $\mu_p(\{1\}) = p$ and $\mu_p(\{0\}) = 1-p$). Let μ_p^n denote the corresponding product probability measure on $\{0,1\}^n$. In other words for $y \in \{0,1\}^n$, we have $\mu_p^n(y) = p^{\sum y_i}(1-p)^{n-\sum y_i}$.

Definition 1.1. *Let $0 \leq \rho \leq 1$ and set $p = \frac{1}{2}(1-\rho)$. For a function $f : \mathbb{Z}_2^n \to \mathbb{C}$, define $T_\rho f : \mathbb{Z}_2^n \to \mathbb{C}$ by*

$$T_\rho f(x) = \mathbb{E}_{y \sim \mu_p^n} f(x+y).$$

Remark 1.2. There are several equivalent ways to define the noise operator. First observe that for every $x \in \mathbb{Z}_2^n$, we have

$$\mathbb{E}_{y \sim \mu_p^n}\left[f(x+y)\right] = 2^n \mathbb{E}_y\left[f(x+y)\mu_p^n(y)\right] = 2^n f * \mu_p^n(x),$$

where in the second expected value $y \in \mathbb{Z}_2^n$ is chosen according to the uniform distribution. We can also write $T_\rho f(x) = \mathbb{E}_z\left[f(z)\right]$, where

$$z_i = \begin{cases} x_i & \text{with probability } 1-p, \\ 1-x_i & \text{with probability } p, \end{cases} \tag{1}$$

independently for each i. In other words z is a noisy copy of x (each coordinate is flipped with probability p). ■

Remark 1.3. It is easy to check that T_ρ is a linear operator:

$$T_\rho(f + \lambda g) = T_\rho f + \lambda T_\rho g.$$

■

Note that T_ρ has a smoothing property. When $\rho = 1$, we have $T_\rho f = f$, but as one decreases ρ, the function $T_\rho f$ "converges" to the constant $\mathbb{E}[f]$ and indeed, for $\rho = 0$, we have $T_\rho f = \mathbb{E}[f]$. Note $T_\rho f(x)$ takes the average of f evaluated at points sampled according to z. When $\rho = 1$, the random variable z is concentrated on point x, and thus the average is just over x so we obtain the original function f. As ρ decreases, the variable z becomes more spread out. Finally $\rho = 0$, we lose the information about x and z is distributed uniformly over all points in $\mathbb{Z}_2^n$. Therefore in this case we get the constant function $\mathbb{E}[f]$. Recall from Lecture 3, when introducing the concept of convolution, we saw that if S is the Hamming ball of radius r around 0 in $\mathbb{Z}_2^n$, then $f * 1_S(x)$ is the

These notes are scribed by Anil Ada.

average of f over the Hamming ball of radius r around x. The noise operator, which is basically a convolution itself, has a smoother definition and the Hamming ball is replaced by a distribution centered at x.

Let us now see the effect of the noise operator on the Fourier spectrum.

Lemma 1.4. *If $f : \mathbb{Z}_2^n \to \mathbb{C}$, then*

$$T_\rho f = \sum_{S \subseteq [n]} \rho^{|S|} \widehat{f}(S) \chi_S.$$

Proof. Since T_ρ is linear it suffices to show that for every $S \subseteq [n]$, we have

$$T_\rho \chi_S = \rho^{|S|} \chi_S.$$

Indeed we have

$$\begin{aligned} T_\rho \chi_S(x) &= \mathbb{E}_{y \sim \mu_p^n} \chi_S(x+y) = \chi_S(x) \mathbb{E}_{y \sim \mu_p^n} \chi_S(y) = \chi_S(x) \mathbb{E}_{y_i \sim \mu_p} \prod_{i \in S} (-1)^{y_i} \\ &= \chi_S(x) \prod_{i \in S} \mathbb{E}_{y_i \sim \mu_p} (-1)^{y_i} = \chi_S(x) \rho^{|S|}. \end{aligned}$$

□

In other words, the noise operator dampens the high frequency Fourier coefficients, and the dampening effect increases exponentially with the frequency.

In the above proof, we utilized the fact that the noise operator acts on each coordinate independently. In fact, in many results regarding the noise operator we can employ the same trick: analyze the effect of the noise in one coordinate and then use the direct product structure to obtain the desired result.

Given the smoothing property of the noise operator, it is not surprising that T_ρ is contractive:

Theorem 1.5 (Contractivity)**.** *For $1 \le p \le \infty$, the operator T_ρ is a contractive operator from L_p to L_p. That is,*

$$\|T_\rho f\|_p \le \|f\|_p.$$

Proof. A simple application of Minkowski's Inequality (Lecture 1, Theorem 3.2) gives the result.

$$\begin{aligned} \|T_\rho f\|_p &= \left(\mathbb{E}_x \left| \mathbb{E}_{y \sim \mu_p^n} f(x+y) \right|^p \right)^{1/p} \\ &\le \mathbb{E}_{y \sim \mu_p^n} \left(\mathbb{E}_x |f(x+y)|^p \right)^{1/p} \\ &= \|f\|_p. \end{aligned}$$

□

The main theorem we are going to prove in this lecture says that not only T_ρ is contractive, but it is also *hypercontractive*. Before stating this theorem and presenting its proof, we introduce some notation.

As stated before, the direct product structure of $\mathbb{Z}_2^n$ is very useful and is often exploited in proofs. For this reason we introduce some notation for product probability spaces. For a distribution μ over X and a distribution ν over Y, consider the product probability distribution $\mu \times \nu$. Consider $f : (X \times Y, \mu \times \nu) \to \mathbb{C}$. We define $\|f\|_{L_p(\nu)}$ to be the function $x \mapsto \|f_x\|_{L_p(\nu)}$, where $f_x = f(x, \cdot)$. Similarly, define $\|f\|_{L_p(\mu)}$ to be the function $y \mapsto \|f_y\|_{L_p(\mu)}$, where $f_y = f(\cdot, y)$.

Given a subset $S \subset [n]$, we can view a function $f : \mathbb{Z}_2^n \to \mathbb{C}$ as a function $f : \mathbb{Z}_2^S \times \mathbb{Z}_2^{\bar{S}} \to \mathbb{C}$. Then it is straightforward to verify,

$$\|f\|_q = \left\| \|f\|_{L_q(\mathbb{Z}_2^S)} \right\|_{L_q(\mathbb{Z}_2^{\bar{S}})}. \tag{2}$$

It can be instructive to see how $\left\| \|f\|_{L_p(\mathbb{Z}_2^S)} \right\|_{L_q(\mathbb{Z}_2^{\bar{S}})}$ expands out:

$$\left\| \|f\|_{L_p(\mathbb{Z}_2^S)} \right\|_{L_q(\mathbb{Z}_2^{\bar{S}})} = \left(\mathbb{E}_{y \in \mathbb{Z}_2^{\bar{S}}} \left| \|f_y\|_{L_p(\mathbb{Z}_2^S)} \right|^q \right)^{1/q} = \left(\mathbb{E}_{y \in \mathbb{Z}_2^{\bar{S}}} \left| \left(\mathbb{E}_{x \in \mathbb{Z}_2^S} |f_y(x)|^p \right)^{1/p} \right|^q \right)^{1/q}. \tag{3}$$

Equation (2) follows immediately as $f_y(x) = f(x, y)$.

As in the proof of Theorem 1.5, a simple application of Minkowski's Inequality gives

$$\left\| \|f\|_{L_1(\mu)} \right\|_{L_p(\nu)} \leq \left\| \|f\|_{L_p(\nu)} \right\|_{L_1(\mu)}.$$

Indeed,

$$\left\| \|f\|_{L_1(\mu)} \right\|_{L_p(\nu)} = \|\mathbb{E}_{x\sim\mu}|f(x,\cdot)|\|_{L_p(\nu)} \leq \mathbb{E}_{x\sim\mu} \| \ |f(x,\cdot)| \ \|_{L_p(\nu)} = \left\| \|f\|_{L_p(\nu)} \right\|_{L_1(\mu)}.$$

This is in fact a special case of a more general inequality:

Theorem 1.6 (Generalized Minkowski's Inequality)**.** *For $1 \leq p \leq q \leq \infty$, we have*

$$\left\| \|f\|_{L_p(\nu)} \right\|_{L_q(\mu)} \leq \left\| \|f\|_{L_q(\mu)} \right\|_{L_p(\nu)}.$$

Now we have all the tools we need to prove the Bonami-Beckner inequality. Recall that L_p norms are increasing on probability space. That is for $1 \leq p \leq q \leq \infty$ we have $\|f\|_p \leq \|f\|_q$. The Bonami-Beckner inequality says that if we sufficiently smooth f by applying the operator T_ρ, we can reverse the direction of this inequality.

Theorem 1.7 (Hypercontractivity - Bonami 1970, Beckner 1975, Nelson 1973, Gross 1975)**.** *Let $1 < p \leq q < \infty$. Then for $0 \leq \rho \leq \sqrt{\frac{p-1}{q-1}}$,*

$$\|T_\rho f\|_q \leq \|f\|_p.$$

Proof. The proof is by induction on n. First we prove the inequality for $n = 1$ and then exploit the direct product structure to prove it for all n.

Consider $f : \mathbb{Z}_2^n \to \mathbb{C}$ and set $\alpha = \frac{1}{2}(1-\rho)$. Then

$$\begin{aligned}
\|T_\rho f\|_q &= \left(\mathbb{E}_x \left|\mathbb{E}_{y\sim\mu_\alpha} f(x+y)\right|^q\right)^{1/q} \\
&= \left(\frac{1}{2}\big((1-\alpha)|f(0)| + \alpha|f(1)|\big)^q + \frac{1}{2}\big(\alpha|f(0)| + (1-\alpha)|f(1)|\big)^q \right)^{1/q} \\
&\leq \left(\frac{1}{2}|f(0)|^p + \frac{1}{2}|f(1)|^p \right)^{1/p} \\
&= \|f\|_p.
\end{aligned}$$

Above, the inequality can be derived using some standard methods from calculus.

We move on to the induction step. For $S \subseteq [n]$, let T_ρ^S denote the noise operator applied to the coordinates in S. That is, it is an operator on the function $f(\cdot, x_{\bar{S}})$, where $x_{\bar{S}}$ denotes the variables

x_i for $i \notin S$. Let $S = \{1\}$. In light of Equation (3), we have

$$\begin{aligned}
\|T_\rho f\|_q &= \|T_\rho^S T_\rho^{\bar{S}} f\|_q \\
&= \Big\| \|T_\rho^S T_\rho^{\bar{S}} f\|_{L_q(\mathbb{Z}_2^S)} \Big\|_{L_q(\mathbb{Z}_2^{\bar{S}})} && \text{(Equation (2))} \\
&\le \Big\| \|T_\rho^{\bar{S}} f\|_{L_p(\mathbb{Z}_2^S)} \Big\|_{L_q(\mathbb{Z}_2^{\bar{S}})} && \text{(Induction Hypothesis)} \\
&\le \Big\| \|T_\rho^{\bar{S}} f\|_{L_q(\mathbb{Z}_2^{\bar{S}})} \Big\|_{L_p(\mathbb{Z}_2^S)} && \text{(Generalized Minkowski)} \\
&\le \Big\| \|f\|_{L_p(\mathbb{Z}_2^{\bar{S}})} \Big\|_{L_p(\mathbb{Z}_2^S)} && \text{(Induction Hypothesis)} \\
&= \|f\|_p && \text{(Equation (2)).}
\end{aligned}$$

□

Lecture 10

We start with a very useful corollary of the Bonami-Beckner inequality.

Corollary 1.8. *Let $f : \mathbb{Z}_2^n \to \mathbb{C}$ be a function and $k > 0$ be an integer. Then for $1 < p \le 2$*

$$\|f^{\le k}\|_2 \le \left(\frac{1}{\sqrt{p-1}}\right)^k \|f\|_p,$$

and for $2 \le q < \infty$,

$$\|f^{\le k}\|_q \le \left(\sqrt{q-1}\right)^k \|f\|_2.$$

Proof. In the case of $1 < p \le 2$, we can apply the Bonami-Beckner inequality with $\rho = \sqrt{p-1}$ and get

$$\|T_\rho f\|_2 \le \|f\|_p.$$

Observe that

$$\|T_\rho f\|_2^2 = \sum_S \rho^{2|S|} |\widehat{f}(S)|^2 \ge \rho^{2k} \sum_{S:|S|\le k} |\widehat{f}(S)|^2 = \rho^{2k} \|f^{\le k}\|_2^2.$$

Therefore

$$\|f^{\le k}\|_2 \le \frac{1}{\rho^k} \|f\|_p = \left(\frac{1}{\sqrt{p-1}}\right)^k \|f\|_p.$$

Case $q \ge 2$ follows by duality. Let p satisfy $\frac{1}{p} + \frac{1}{q} = 1$. Note that $1 < p \le 2$ so we can apply the first part using the L_p norm. Since L_p and L_q are dual norms we have (see question 2 of assignment 1):

$$\|f^{\le k}\|_q = \sup_{g \ne 0} \frac{\langle f^{\le k}, g\rangle}{\|g\|_p} \le \sup_{g \ne 0} \frac{\langle f^{\le k}, g\rangle}{(\sqrt{p-1})^k \|g^{\le k}\|_2} = (\sqrt{q-1})^k \sup_{g \ne 0} \frac{\langle f^{\le k}, g^{\le k}\rangle}{\|g^{\le k}\|_2}.$$

Since the dual of the L_2 norm is the L_2 norm itself,

$$(\sqrt{q-1})^k \sup_{g\neq 0} \frac{\langle f^{\leq k}, g^{\leq k}\rangle}{\|g^{\leq k}\|_2} = (\sqrt{q-1})^k \|f^{\leq k}\|_2 \leq (\sqrt{q-1})^k \|f\|_2.$$

□

Exercise 1.9. Prove the $q \geq 2$ case of Corollary 1.8 by applying the Bonami-Beckner inequality to $g = \sum_S \rho^{-|S|}\widehat{f}(S)\chi_S$. ■

Recall that the L_p norms are increasing, that is, $\|f\|_p \leq \|f\|_q$ when $1 \leq p \leq q \leq \infty$. An immediate consequence of Corollary 1.8 is that if $\deg(f) \leq k$, then for $1 < p \leq 2$,

$$\|f\|_p \leq \|f\|_2 \leq \left(\frac{1}{\sqrt{p-1}}\right)^k \|f\|_p,$$

and for $2 \leq q < \infty$,

$$\|f\|_2 \leq \|f\|_q \leq \left(\sqrt{q-1}\right)^k \|f\|_2.$$

Remark 1.10. The above inequalities show that the function is "flat". If there are large fluctuations in f, then we cannot hope to have such strong equivalences between the different norms. In this sense, one can think of the Bonami-Beckner inequality as a concentration inequality. Indeed, viewing f as a random variable, by bounding the q-norms in terms of the 2-norm, we are essentially bounding the moments of f in terms of the standard deviation of f. ■

Remark 1.11. The case of $\deg(f) = 1$ is known as Khintchine inequality: for $a_1, a_2, \ldots, a_n \in \mathbb{C}$,

$$\sqrt{p-1}\left(\sum_i |a_i|^2\right)^{1/2} \leq \left(\mathbb{E}\left|\sum_i \epsilon_i a_i\right|^p\right)^{1/p} \leq \left(\sum_i |a_i|^2\right)^{1/2},$$

where the expectation is over $\{\epsilon_i\}$ which are ± 1 valued i.i.d. random variables with $\Pr[\epsilon_i = 1] = 1/2$. By setting $f = \sum_i a_i \chi_{\{i\}}$, we see the correspondence immediately. ■

2. Influence and Friedgut's Theorem

The Bonami-Beckner inequality is a powerful tool in the analysis of Boolean functions. We will see some important applications, but first, we introduce the notion of *influence*.

Definition 2.1 (Influence)**.** *Let $f : \mathbb{Z}_2^n \to \{0,1\}$. The influence of the ith variable on f is the probability that resampling the ith coordinate changes the the value of f. That is,*

$$I_i(f) = \Pr[f(x_1, \ldots, x_i, \ldots, x_n) \neq f(x_1, \ldots, y_i, \ldots, x_n)],$$

where $x_1, x_2, \ldots, x_n$ and y_i are independently sampled. Equivalently[1]*,*

$$I_i(f) = \frac{1}{2}\Pr[f(x) \neq f(x + e_i)].$$

The total influence of f is

$$I_f = \sum_{i=1}^n I_i(f).$$

Note that with our definition, $0 \leq I_i(f) \leq 1/2$ and $0 \leq I_f \leq n/2$.

[1]In the literature the influence is sometimes defined as $I_i(f) = \Pr[f(x) \neq f(x + e_i)]$.

Remark 2.2. Considering the support of f, $\mathrm{Supp}(f) = \{x : f(x) \neq 0\}$, as a subset of the hypercube $\mathcal{Q}_n$, I_f corresponds to the *edge boundary* of $\mathrm{Supp}(f)$. For a subset S of the hypercube, the edge boundary of S, denoted ∂S, is the set of edges of $\mathcal{Q}_n$ with one end point in S and the other endpoint outside of S. It follows by definition that

$$I_f = \frac{|\partial \mathrm{Supp}(f)|}{2^n}.$$

■

When studying influences, it is natural to consider $f_{(i)} : \mathbb{Z}_2^n \to \{-1, 0, 1\}$ (sometimes referred to as the ith derivative of f), which is defined as

$$f_{(i)}(x) = f(x) - f(x + e_i).$$

Indeed, since $|f_{(i)}(x)| = |f_{(i)}(x)|^2$, we have

$$I_i(f) = \frac{1}{2}\mathbb{E}_x|f_{(i)}(x)| = \frac{1}{2}\mathbb{E}_x|f_{(i)}(x)|^2 = \frac{1}{2}\|f_{(i)}\|_2^2.$$

The Fourier expansion of $f_{(i)}$ is

$$f_{(i)}(x) = \sum_S \widehat{f}(S)\chi_S(x) - \widehat{f}(S)\chi_S(x + e_i) = 2\sum_{S:i\in S} \widehat{f}(S)\chi_S(x),$$

and therefore

$$I_i(f) = 2\sum_{S:i\in S} |\widehat{f}(S)|^2.$$

We can also get a nice expression for the total influence of f (and hence for the edge boundary of $\mathrm{Supp}(f)$) in terms of its Fourier coefficients:

$$I_f = \sum_i 2\sum_{S:i\in S} |\widehat{f}(S)|^2 = 2\sum_S |S||\widehat{f}(S)|^2.$$

With this, we can get a simple lower bound for the total influence in terms of the variance of f. Note that $\mathrm{Var}(f) = \mathbb{E}[f^2] - (\mathbb{E}[f])^2 = \sum_{S:S\neq\emptyset} |\widehat{f}(S)|^2$ and therefore

$$I_f \geq 2\mathrm{Var}(f). \tag{4}$$

This bound in general can be quite weak. A stronger bound can be obtained by a discrete isoperimetric inequality.

Theorem 2.3 (Edge Isoperimetric Inequality)**.** *For S a subset of the vertices of the hypercube $\mathcal{Q}_n$ we have*

$$|\partial S| \geq -|S|\log_2 \frac{|S|}{2^n}.$$

Equality is achieved when S is a subcube.

Proof. The proof is quite straightforward using induction on n. The base case, $n = 1$, is easily verified so we directly move to the induction step. Partition $\mathcal{Q}_n$ into two disjoint subcubes $\mathcal{Q}_{n-1}^1$ and $\mathcal{Q}_{n-1}^2$ of dimension $n-1$ each. Similarly partition S into two sets $S_1 = S \cap V(\mathcal{Q}_{n-1}^1)$ and $S_2 = S \cap V(\mathcal{Q}_{n-1}^2)$. Without loss of generality assume $|S_1| = |S_2| + t$. Now the boundary of S will have edges from the boundary of S_1 in $\mathcal{Q}^1_{n-1}$, edges from the boundary of S_2 in $\mathcal{Q}^2_{n-1}$, and also at least t edges that must go between the two subcubes. Using the induction hypothesis, we have

$$\begin{aligned}|\partial S| &\geq |S_1|(n-1-\log|S_1|)+|S_2|(n-1-\log|S_2|)+t\\ &= |S_1|n-|S_1|-|S_1|\log|S_1|+|S_2|n-|S_2|-|S_2|\log|S_2|+t\\ &= |S_1|n-|S_1|\log|S_1|+|S_2|n-|S_2|\log|S_2|-2|S_2|.\end{aligned}$$

Note that $|S|(n-\log|S|)=|S_1|n+|S_2|n-(|S_1|+|S_2|)\log(|S_1|+|S_2|)$, so we are done provided

$$|S_1|\log|S_1|+|S_2|\log|S_2|+2|S_2|\leq(|S_1|+|S_2|)\log(|S_1|+|S_2|).$$

This inequality is easily derived using simple manipulations. □

Defining f such that $\mathrm{Supp}(f)=S$, we can rewrite the Edge Isoperimetric Inequality in terms of the total influence:

$$I_f\geq-\mathbb{E}[f]\log_2\mathbb{E}[f]. \tag{5}$$

Consider a balanced function $f:\mathbb{Z}_2^n\to\{0,1\}$, i.e. $\mathbb{E}[f]=1/2$. Both lower bounds (4) and (5) on I_f imply that $I_f\geq 1/2$, which shows

$$\max_i I_i(f)\geq\frac{1}{2n}.$$

Note that the lower bound $I_f\geq 1/2$ is tight for half-cubes, i.e. for $f(x)=x_i$ or $f(x)=-x_i$ for some i. Two questions naturally arise:

(1) (Ben-Or Linial) How small can $\max_i I_i(f)$ be for balanced functions?
(2) What are the functions with small total influence?

We first give an answer to the second question. For this we need to define a *junta*.

Definition 2.4. *A Boolean function $f:\{0,1\}^n\to\{0,1\}$ is called a k-junta if there exists $J\subseteq[n]$ of size at most k and $g:\{0,1\}^J\to\{0,1\}$ such that $f(x)=g(x_J)$. In other words, f is a k-junta if its output only depends on at most k input coordinates.*

Observe that if f is a k-junta then $I_f\leq k/2$. This is because every variable that f does not depend on has influence 0, and every other variable has influence at most 1/2. Friedgut's Theorem gives a partial converse to this observation and states that a Boolean function with small total influence is well approximated by a k-junta with a small k.

Theorem 2.5 (Friedgut)**.** *Let $f:\{0,1\}^n\to\{0,1\}$ be a Boolean function. Then there exists a $2^{O(I_f/\epsilon)}$-junta $g:\{0,1\}^n\to\{0,1\}$ such that*

$$\Pr[f(x)\neq g(x)]\leq\epsilon.$$

Proof. First note that the probabilistic approximation can be interpreted in terms of the L_2 difference:

$$\|f-g\|_2^2=\Pr[f(x)\neq g(x)].$$

Let J be the set of most influential variables of f, that is, $J=\{i\in[n]\mid I_i(f)\geq\delta\}$ for some parameter δ to be determined later. It is natural to try to find a g that depends only on the variables in J. Define h to be

$$h=\sum_{S\subseteq J}\widehat{f}(S)\chi_S.$$

Clearly h depends only on the variables in J, but it is not necessarily a Boolean function. Nevertheless we can round h to make it Boolean. Let $g(x)=1$ if $h(x)>1/2$ and let $g(x)=0$ if $h(x)\leq 1/2$.

By rounding we haven't lost much in the following sense. If $\|f-h\|_2^2 \le \epsilon$, then $\|f-g\|_2^2 \le 4\epsilon$. This is easy to see since for any x with $f(x) \neq g(x)$, $|f(x)-h(x)|^2 \ge 1/4$.

Thus our task reduces to showing that

$$\|f-h\|_2^2 \le \frac{\epsilon}{4}.$$

By Parseval we have

$$\|f-h\|_2^2 = \sum_S (\widehat{f}(S) - \widehat{h}(S))^2 = \sum_{S \not\subseteq J} \widehat{f}(S)^2.$$

So we want to upper bound the ℓ_2 mass of f on sets S with $S \not\subseteq J$. To do this we will divide the above sum into two parts, the low degree part and the high degree part, and deal with them separately.

Intuitively, a function with small total influence should not have large ℓ_2 mass on high degree characters as high degree characters, viewed as 0/1 valued functions, have large total influence. This intuition is easy to formalize. Set $k = 4I_f/\epsilon$. Then

$$I_f = 2\sum_S |S||\widehat{f}(S)|^2 \ge 2k \sum_{|S|\ge k} |\widehat{f}(S)|^2,$$

which implies

$$\sum_{S:|S|\ge k} |\widehat{f}(S)|^2 \le \frac{I_f}{2k} \le \frac{\epsilon}{8}.$$

Thus,

$$\|f-h\|_2^2 \le \frac{\epsilon}{8} + \sum_{\substack{S:|S|<k \\ S\not\subseteq J}} \widehat{f}(S)^2.$$

Now to bound the low degree part we will use Bonami-Beckner inequality (the form given in Corollary 1.8). First observe that

$$\sum_{\substack{S:|S|<k \\ S\not\subseteq J}} \widehat{f}(S)^2 = \sum_{i\notin J} \sum_{\substack{S:|S|<k \\ i\in S}} \widehat{f}(S)^2. \tag{6}$$

We want to bound the inside sum on the RHS above. Recall that

$$f_{(i)}(x) = f(x) - f(x+e_i) = 2\sum_{i\in S} \widehat{f}(S)\chi_S(x),$$

So the quantity we want to bound is $\|f_{(i)}^{<k}\|_2^2$. We apply Corollary 1.8 with $p = 4/3$ to get

$$\|f_{(i)}^{<k}\|_2 \le 3^{k/2}\|f_{(i)}\|_{4/3},$$

and so

$$\|f_{(i)}^{<k}\|_2^2 \le 3^k \left(\mathbb{E}_x|f_{(i)}(x)|^{4/3}\right)^{3/2} = 3^k \left(\mathbb{E}_x|f_{(i)}(x)|^2\right)^{3/2}.$$

Recall that $I_i(f) = \frac{1}{2}\|f_{(i)}\|_2^2$. Also, since $i \notin J$, $I_i(f) < \delta$. Thus,

$$\|f_{(i)}^{<k}\|_2^2 \le 3^k 2^{3/2} I_i(f)^{3/2} \le 3^k 2^{3/2}\delta^{1/2} I_i(f).$$

Equivalently,

$$4\sum_{\substack{S:|S|<k \\ i\in S}} \widehat{f}(S)^2 \le 3^k 2^{3/2}\delta^{1/2} I_i(f).$$

Going back to (6), we have

$$\sum_{i\notin J}\sum_{\substack{S:|S|<k\\ i\in S}}\widehat{f}(S)^2 \le \sum_{i\notin J} 3^k\delta^{1/2}I_i(f) \le 3^k\delta^{1/2}I_f.$$

Putting things together

$$\|f-h\|_2^2 \le \frac{\epsilon}{8} + 3^k\delta^{1/2}I_f \le \frac{\epsilon}{8}+\frac{\epsilon}{8} = \frac{\epsilon}{4},$$

when δ is set to be sufficiently small. Recall that we set $k = 4I_f/\epsilon$, i.e. $I_f = k\epsilon/4$. Now a simple calculation shows that we can pick any δ with

$$\delta \le \frac{1}{43^{2k}k^2}.$$

For instance we can set $\delta = 1/3^{3k+2}$. With this δ, we have

$$|J| \le \frac{I_f}{\delta} \le 2^{O(I_f/\epsilon)},$$

as required.

□

Lecture 11

3. Kahn-Kalai-Linial Theorem

In this lecture we are going to prove the Kahn-Kalai-Linial (KKL) Theorem that says that every balanced function has an influential variable, that is, there is some $i \in [n]$ such that $I_i(f) = \Omega(\frac{\log n}{n})$. The proof is essentially the same as Friedgut's Theorem[2]. We separate the Fourier spectrum of f into high degree and low degree parts. The high degree part is easy to handle and for the low degree part we apply the Bonami-Beckner inequality. The reason why Bonami-Beckner inequality is effective can be seen as follows. For $1 \le p < 2$, when g is a Boolean function, we have $\mathbb{E}[|g|] = \mathbb{E}[|g|^p] = \mathbb{E}[|g|^2]$, which implies that $\|g\|_p = \|g\|_2^{2/p}$. Now if $\|g\|_2 =: \delta$ is small, then $\|g\|_p = \delta \cdot \delta^{(2/p-1)}$ is very small. So applying Corollary 1.8 to g, we get a good bound on $\|g^{<k}\|_2$ and gain a factor of $\delta^{2/p-1}$.

Theorem 3.1 (Kahn-Kalai-Linial). *Let $f : \mathbb{Z}_2^n \to \{0,1\}$ be such that $\mathbb{E}[f] = \alpha$. If $\delta = \max_i I_i(f)$, then*

$$I_f \ge \Omega\left(\alpha(1-\alpha)\log 1/\delta\right)$$

In particular

$$\delta \ge \Omega\left(\alpha(1-\alpha)\frac{\log n}{n}\right).$$

[2]Historically the KKL Theorem came before Friedgut's Theorem.

Proof. Recall that $\mathrm{Var}(f) = \mathbb{E}[f^2] - (\mathbb{E}[f])^2 = \alpha - \alpha^2 = \alpha(1-\alpha)$. Also since $\mathbb{E}[f]^2 = \widehat{f}(\emptyset)^2$,

$$\mathrm{Var}(f) = \sum_{S:|S|\geq 1} |\widehat{f}(S)|^2.$$

In Lecture 10 we observed that $\mathrm{Var}(f) \leq \frac{1}{2} I_f$ and that this leads to the bound $\delta \geq \frac{1}{2n}$ for balanced functions. Our goal now is to obtain a better upper bound on the variance, which will lead to a better lower bound on δ. In particular we are aiming for the upper bound

$$\mathrm{Var}(f) = \sum_{S:|S|\geq 1} |\widehat{f}(S)|^2 \lesssim \frac{I_f}{\log 1/\delta}.$$

Our strategy will be as in the proof of Friedgut's Theorem. We divide the sum into the low degree and high degree parts, and upper bound each part separately.

Recall that

$$I_f = 2\sum_{S} |S||\widehat{f}(S)|^2 \geq 2k \sum_{|S|>k} |\widehat{f}(S)|^2.$$

This implies

$$\sum_{|S|>k} |\widehat{f}(S)|^2 \leq \frac{I_f}{2k}.$$

Setting $k \approx \log 1/\delta$, the upper bound above is what we want for $\mathrm{Var}(f)$. So with this choice of k, we would like to show an upper bound on the low degree part that is negligible compared to $I_f/2k$.

To handle the low degree part, we will apply Bonami-Beckner inequality to $\|f_{(i)}\|_2$ with $p = 3/2$:

$$\sum_{1\leq|S|\leq k} |\widehat{f}(S)|^2 \leq \sum_{i=1}^{n} \sum_{\substack{i\in S\\ |S|\leq k}} |\widehat{f}(S)|^2 = \frac{1}{4}\sum_{i=1}^{n} \|f_{(i)}^{\leq k}\|_2^2 \leq \frac{1}{4}\sum_{i=1}^{n} 2^k \|f_{(i)}\|_{3/2}^2.$$

Using the fact that $|f_{(i)}(x)| \in \{0,1\}$, we have

$$\frac{1}{4}\sum_{i=1}^{n} 2^k \|f_{(i)}\|_{3/2}^2 = \frac{1}{4}2^k \sum_{i=1}^{n} \|f_{(i)}\|_2^{8/3} = \frac{1}{4}2^k \sum_{i=1}^{n} (2I_i(f))^{4/3} \leq 2^k \delta^{1/3} \sum_{i=1}^{n} I_i(f) = 2^k\delta^{1/3} I_f.$$

Putting things together we get

$$\alpha(1-\alpha) = \sum_{S:|S|\geq 1} |\widehat{f}(S)|^2 \leq \frac{I_f}{2k} + 2^k \delta^{1/3} I_f.$$

Setting $k = \frac{1}{10}\log 1/\delta$ shows

$$\frac{1}{10}\alpha(1-\alpha)\log 1/\delta \leq I_f.$$

We also know that $I_f \leq \delta n$. These upper and lower bounds on I_f imply by a straightforward calculation that

$$\delta \geq \Omega\left(\alpha(1-\alpha)\frac{\log n}{n}\right).$$

□

The KKL Theorem is tight, which can be seen by considering the *tribes* function. Let

$$f(x) = \bigvee_{i=1}^{m} \bigwedge_{j=1}^{k} x_{ij},$$

where $k = \log n - \log \ln n$ and $m = n/k$. Without loss of generality consider the first variable. For x_1 to be able to change the output, all other variables in the first clause must be set to 1, and all other clauses must be evaluating to 0. Thus,

$$\begin{aligned} I_1(f) &= \frac{1}{2}\Pr[f(x) \neq f(x+e_i)] = \frac{1}{2}(1-2^{-k})^{m-1} \cdot 2^{-k+1} \\ &= 2^{-k}(1-2^{-k})^{m-1} = \frac{\ln n}{n}(1-\frac{\ln n}{n})^{m-1} = \frac{\ln n}{n}(1-o(1)). \end{aligned}$$

Now we will see some corollaries to KKL Theorem and some related conjectures.

Corollary 3.2. *If a balanced function $f : \mathbb{Z}_2^n \to \{0,1\}$ satisfies $I_1(f) = I_2(f) = \cdots = I_n(f)$ (e.g. f is invariant under certain symmetries), then $I_f \gtrsim \log n$.*

Bourgain and Kalai show that under strong symmetry assumptions, the above bound can be improved significantly. For instance if f is a symmetric function, i.e. f's output only depends on the Hamming weight of the input, then $I_f \gtrsim \sqrt{n}$.

A Boolean function $f : \{0,1\}^n \to \{0,1\}$ is called *increasing* (or *monotone*) if $f(x) \leq f(y)$ whenever $x_i \leq y_i$ for all i.

Corollary 3.3. *Let $f : \{0,1\}^n \to \{0,1\}$ increasing balanced function. Then there is a set $J \subseteq [n]$ of size $O_\epsilon(\frac{n}{\log n})$ such that*

$$\mathbb{E}\left[f(x)|x_J = \vec{1}\right] \geq 1-\epsilon.$$

Proof Sketch. Let $i \in [n]$ have the highest influence. Then setting $x_i = 1$ will increase the average of f by at least $\Omega(\frac{\log n}{n})$. Repeat with the new function. □

Conjecture 3.4 (Freidgut)**.** *Let $f : [0,1]^n \to \{0,1\}$ be an increasing function. Then there exists a subset $J \subseteq [n]$ with $|J| = o_\epsilon(n)$ such that*

$$\mathbb{E}\left[f(x)|x_J = \vec{0}\right] \leq \epsilon \quad \textit{or} \quad \mathbb{E}\left[f(x)|x_J = \vec{1}\right] \geq 1-\epsilon.$$

Conjecture 3.5 (Freidgut)**.** *Suppose $f : \{0,1\}^n \to \{0,1\}$ is increasing and $\max_i I_i(f) \leq \frac{c \log n}{n}$, for some constant c. Then there is $J \subseteq [n]$ of size $O_{\epsilon,c}(\log n)$ such that*

$$\mathbb{E}\left[f(x)|x_J = \vec{0}\right] \leq \epsilon \quad \textit{or} \quad \mathbb{E}\left[f(x)|x_J = \vec{1}\right] \geq 1-\epsilon.$$

We can make a similar conjecture for non-monotone functions.

Conjecture 3.6. *Suppose $f : \{0,1\}^n \to \{0,1\}$ satisfies $\max_i I_i(f) \leq \frac{c \log n}{n}$, for some constant c. Then there is $J \subseteq [n]$ of size $O_{\epsilon,c}(\log n)$ and $y \in \{0,1\}^J$ such that*

$$\mathbb{E}\left[f(x)|x_J = y\right] \leq \epsilon \quad \textit{or} \quad \mathbb{E}\left[f(x)|x_J = y\right] \geq 1-\epsilon.$$

The influences of increasing Boolean functions have a very special and useful characterization in terms of f's Fourier coefficients. It is not hard to verify that

$$I_i(f) = \frac{1}{2}\Pr[f(x) \neq f(x+e_i)] = -\frac{1}{2}\mathbb{E} f(x)\chi_{\{i\}}(x) = -\frac{1}{2}\widehat{f}(\{i\}).$$

Using this and the Cauchy-Schwarz inequality, it is easy to get an upper bound on the total influence of increasing functions:

$$I_f = \frac{1}{2}\sum_i |\widehat{f}(\{i\})| \leq \frac{1}{2}\sqrt{n}\sum_i \left(|\widehat{f}(\{i\})|^2\right)^{1/2} \leq \frac{\sqrt{n}}{2}.$$

Note that for non-monotone functions we can have $I_f = n/2$ (e.g. $f = \text{PARITY}$). The above bound is tight since $I_{\text{MAJ}} = \Theta(\sqrt{n})$, where MAJ denotes the majority function:

$$\text{MAJ}(x) := \begin{cases} 1 & \text{if } \sum_i x_i \geq n/2, \\ 0 & \text{otherwise.} \end{cases}$$

4. Chang's Lemma

We move on to a very useful structure result for the large Fourier coefficients of a bounded function. This result has many applications, especially in additive combinatorics. We'll see a proof of this result that uses the Bonami-Beckner inequality.

Consider $f : \{0,1\}^n \to [0,1]$ and let $\alpha = \mathbb{E}[f]$. Note that for any $a \in \{0,1\}^n$, we have

$$\widehat{f}(a) = \mathbb{E}_x f(x)\chi_a(x) \leq \mathbb{E}_x |f(x)||\chi_a(x)| = \alpha.$$

Let A be the set of large coefficients, that is, $A = \{a : |\widehat{f}(a)| \geq \rho\alpha\}$ for a fixed $0 < \rho \leq 1$. It is easy to bound the size of A using Parseval. Note that $\sum_S |\widehat{f}(a)|^2 = \mathbb{E} f(x)^2 \leq \alpha$. Each coefficient in A contributes at least $\rho^2\alpha^2$ to this sum so we must have

$$|A| \leq \frac{1}{\alpha\rho^2}.$$

Chang's Lemma says that the elements of A, viewed as vectors, must live in a small subspace.

Lemma 4.1 (Chang)**.** *Let $f : \{0,1\}^n \to [0,1]$ be such that $\mathbb{E}[f] = \alpha$. Let A be defined as above. Then*

$$\dim(\text{span } A) \leq O\left(\frac{\ln(1/\alpha)}{\rho^2}\right).$$

Proof. Let $r_1, r_2, \dots, r_d$ be independent vectors in A. We want to show that

$$d \leq O\left(\frac{\ln(1/\alpha)}{\rho^2}\right).$$

By a change of coordinates we can assume that $r_i = e_i$ for $1 \leq i \leq d$. That is, without loss of generality, we can assume that the Fourier coefficients corresponding to r_i are $\widehat{f}(\{i\})$. Now define

$$g = \sum_{i=1}^{d} \widehat{f}(\{i\})\chi_{\{i\}}.$$

We will bound $\langle f, g\rangle$ from below and above, and this will result in the desired upper bound on d.

First, by Parseval, we have

$$\mathbb{E}[fg] = \sum_{i=1}^{d} |\widehat{f}(\{i\})|^2 \geq d\rho^2\alpha^2.$$

For the upper bound, we first use Hölder's inequality, and then apply Bonami-Beckner inequality. Let p and q be conjugate exponents with $1 < p \leq 2$. Then,

$$\mathbb{E}[fg] \leq \|f\|_p\|g\|_q \leq \|f\|_p\sqrt{q-1}\|g\|_2 \leq (\mathbb{E}[|f|])^{1/p}\sqrt{q-1}\|g\|_2 = \alpha^{1/p}\sqrt{q-1}\|g\|_2.$$

Since $\mathbb{E}[fg] = \|g\|_2^2$, we get

$$\|g\|_2^2 \leq \alpha^{2/p}(q-1).$$

Now combining the upper and lower bound on $\mathbb{E}[fg]$ we have

$$d\alpha^2\rho^2 \leq \alpha^{2/p}(q-1).$$

Setting $q = \ln \frac{1}{\alpha}$, we conclude

$$d \leq O\left(\frac{\ln(1/\alpha)}{\rho^2}\right).$$

$\square$

Remark 4.2. Alternatively, Chang's Lemma can be proved in a more probabilistic language. Define $g = \sum_{i=1}^{d} \chi_{r_i}$. Again, we want to upper and lower bound $\mathbb{E}[fg]$. By Parseval the lower bound is $\rho\alpha d$. For the upper bound, first note that the linear independence of $r_1, r_2, \ldots, r_d$ translates into probabilistic independence for the corresponding characters. That is, g is the sum of d ± 1 valued i.i.d. random variables when x is chosen uniformly at random. With this view, it is straightforward to upper bound $\mathbb{E}[fg]$ using the concentration of g. $\blacksquare$

School of Computer Science, McGill University, Montréal, Canada
E-mail address: hatami@cs.mcgill.ca