Lecture 4

Lecturer: Irit Dinur *Scribe: Cheng Chen*

1 Recap

Last time, we considered the function where no variable has large influence. That is to find balanced Boolean function that minimizes $\max_i \mathrm{Inf}_i(f)$. It turns out that TRIBES achieves the best possible bound, $O(\log n/n)$. This result is known as KKL theorem, and we will finish the proof in this lecture. Before go into the proof, we summarize some results below:

Let $\rho \in [0, 1]$, for any fixed $x \in \{-1, 1\}$, we define the ρ-correlated bit y to x, denoted by $y \sim N_\rho(x)$, that is with probability ρ we have $y = x$ and with probability $1 - \rho$ we have y uniformly random. Usually, we also define $\delta = (1 - \rho)/2$, then $y = x$ with probability $1 - \delta$ and $y = -x$ with probability δ. Furthermore, we say a n-bit string y is ρ-correlated to n-bit string x if each bit of y is independently ρ-correlated to the corresponding bit of x.

For any fixed $\rho \in [0, 1]$, the ρ-noisy hypercube graph is a weighted complete graph, where the nodes are all the n-bit strings and the weight of an edge (x, y) is equal to the probability of taking x, flipping each bit with probability δ and reaching y. Define T_ρ to be

$$T_\rho = \begin{pmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{pmatrix}$$

Then the adjacency matrix of the ρ-noisy hypercube graph is the tensor product $T_\rho^{\otimes n}$. It is easy to verify that all the eigenvectors are the Fourier bases χ_S with corresponding eigenvalues $\rho^{|S|}$. Further, we define the ρ-noisy stability of f to be $\mathrm{Stab}_\rho(f) = \langle f, T_\rho f\rangle = \sum_S \rho^{|S|} \hat{f}_S^2$.

Let $A \subseteq \{-1, 1\}^n$ of size $\alpha 2^n$ and $\mathbf{1}_A(x) : \{-1, 1\}^n \to \{0, 1\}$ be the indicator function. Small Set Expansion theorem states that $\mathrm{Stab}_\rho(\mathbf{1}_A) \le \alpha^{2/(1+\rho)}$. Note that since $\mathbf{1}_A$ is 0-1 function, $\mathrm{Stab}_\rho(\mathbf{1}_A) = \mathbb{E}[\mathbf{1}_A(x)\mathbf{1}_A(y)] = \Pr[x \in A \wedge y \in A] = \alpha \Pr[y \in A \mid x \in A]$. So $\mathrm{Stab}_\rho(\mathbf{1}_A)$ is the probability that a random walk starting at a point $x \in A$ remains in A, normalized by the density of A. And Small Set Expansion is equivalently stated as $Pr[y \in A \mid x \in A] \le \alpha^{(1-\rho)/(1+\rho)}$. In particular, when α is small, this probability is is very small, i.e. the noisy hypercube graph is a good expander.

To prove Small Set Expansion theorem, we need the $(4, 2)$-hypercontractivity lemma due to Bonami, which states that the 4-norm of $T_{1/\sqrt{3}}f$ is at most the 2-norm of f. We will state both theorems explicitly in the next section.

2 Proof of Small Set Expansion theorem and KKL theorem

Definition 1. *(p-norm) Let $p \ge 1$ and $f : \{-1, 1\}^n \to \mathbb{R}$, define the p-norm of f be $\mathbb{E}_x[|f(x)|^p]^{1/p}$.*

Lemma 2. *(Hypercontractivity) Let $f : \{-1, 1\}^n \to \mathbb{R}$ and $\rho = 1/\sqrt{3}$. Then $\|T_\rho f\|_4 \le \|f\|_2$.*

Theorem 3. *(Small Set Expansion) Let $A \subseteq \{-1,1\}^n$ with density α and $f = \mathbf{1}_A$ be the indicator function. Then $Stab_{1/3}(f) \leq \alpha^{3/2}$.*

Proof. Using the Fourier expansion of the noise operator, we have

$$\langle f, T_{1/3} f \rangle = \sum_S \hat{f}_S^2 / 3^{|S|} = \langle T_{1/\sqrt{3}} f, T_{1/\sqrt{3}} f \rangle = \|T_{1/\sqrt{3}} f\|_2^2 \tag{1}$$

By Holder's Inequality that $\mathbb{E}_x[f(x)g(x)] \leq \|f\|_p \cdot \|g\|_q$ for any $p, q \geq 1$ and $1/p + 1/q = 1$, we have

$$\langle f, T_{1/3} f \rangle \leq \|f\|_{4/3} \cdot \|T_{1/3} f\|_4 = \|f\|_{4/3} \cdot \|T_{1/\sqrt{3}}(T_{1/\sqrt{3}} f)\|_4 \leq \|f\|_{4/3} \cdot \|T_{1/\sqrt{3}} f\|_2 \tag{2}$$

where we also use Bonami's hypercontractivity lemma for $T_{1/\sqrt{3}} f$ in the second inequality. Combine (1) and (2), and note that $f(x)^{4/3} = f(x)$ since f is 0-1 function, we have

$$\|T_{1/\sqrt{3}} f\|_2 \leq \|f\|_{4/3} = \mathbb{E}[f(x)^{4/3}]^{3/4} = \alpha^{3/4}$$

Therefore we have $\mathrm{Stab}_{1/3}(f) = \langle f, T_{1/3} f \rangle = \|T_{1/\sqrt{3}} f\|_2^2 \leq \alpha^{3/2}$. $\square$

Theorem 4. *(KKL) Let $f : \{-1,1\}^n \to \{-1,1\}$ with $\mathbb{E}[f] = 0$. Then there exists $i \in [n]$ such that $Inf_i(f) = \Omega(\log n / n)$.*

Proof. The approach is to prove that either the total influence $\mathrm{Inf}(f)$ is large, say $\Omega(\log n)$, and therefore the theorem follows, or there exists some i with $\mathrm{Inf}_i(f) \geq 1/\sqrt{n}$. Recall the derivative operator $(D_i f)(x) = (f(x) - f(x^{\oplus i}))/2$. The idea is to apply the Small Set Expansion theorem to $D_i f$ and sum over all $i \in [n]$. Although the Small Set Expansion theorem originally states for 0-1 functions, but what actually used in the proof is that $f(x)^{4/3} = |f(x)|$, which is also the case for $D_i f$. Note that the (absolute value) density of $D_i f$ is $\mathrm{Inf}_i(f)$, then we have

$$\sum_{i=1}^n \langle D_i f, T_{1/3}(D_i f) \rangle \leq \sum_{i=1}^n \mathrm{Inf}_i(f) \sqrt{\mathrm{Inf}_i(f)} \leq \mathrm{Inf}(f) \sqrt{\max\{\mathrm{Inf}_i(f)\}} \tag{3}$$

On the other hand, we have

$$\begin{aligned}
\sum_{i=1}^n \langle D_i f, T_{1/3}(D_i f) \rangle &= \sum_{i=1}^n \sum_{S \ni i} \hat{f}_S^2 / 3^{|S|} = \sum_{|S| \geq 1} |S| \hat{f}_S^2 / 3^{|S|} \\
&\geq \sum_{1 \leq |S| \leq 2\mathrm{Inf}(f)} |S| \hat{f}_S^2 / 3^{|S|} \\
&\geq \sum_{1 \leq |S| \leq 2\mathrm{Inf}(f)} 2\mathrm{Inf}(f) \hat{f}_S^2 / 3^{2\mathrm{Inf}(f)} \qquad (4) \\
&\geq \mathrm{Inf}(f) / 3^{2\mathrm{Inf}(f)} \qquad (5)
\end{aligned}$$

where (4) uses the fact that $x/3^x$ is a decreasing function when $x \geq 1$, and (5) uses Markov's inequality $\sum_{|S| \geq 2\mathrm{Inf}(f)} \hat{f}_S^2 \leq 1/2$ along with the assumption that f is balanced. Combine (3) and (5), we have $\max\{\mathrm{Inf}_i(f)\} \geq 1/3^{4\mathrm{Inf}(f)}$. That is that either $\mathrm{Inf}(f) \geq \log n / 8 \log 3$ or $\max\{\mathrm{Inf}_i(f)\} \geq 1/\sqrt{n}$. $\square$

We also note that there is an interesting corollary of KKL theorem, though not mentioned in the lecture, that a $O(1/\log n)$ fraction of voters can collaboratively bias the outcome of an almost balanced voting scheme in their favor with probability 99%.

Theorem 5. *Let $f : \{-1,1\}^n \to \{-1,1\}$ is almost balanced ($Var[f] \geq \Omega(1)$), then for all $\epsilon > 0$, there exists a coalition $J \subseteq [n]$ of size at most $O(\log(1/\epsilon))n/\log n$ such that $Inf_J(f) \geq 1-\epsilon$.*

Here Inf_J of coalition J is a generalization of influence of i-th bit. For formal definition and properties, see Section 2 and 3 in [1]. For a proof for this theorem, see Corollary 4.5 in [1].

Definition 6. *(Juntas) Let $f : \{-1,1\}^n \to \{-1,1\}$. Then f is a juntas function on J iff there exists some function $g : \{-1,1\}^{|J|} \to \{-1,1\}$ such that $f(x_1,\ldots,x_n) = g(x_J)$.*

We also list two related result: Friedgut's Juntas Theorem says that Boolean function with small total influence are "close" to juntas function, and FKN theorem says that Boolean functions with big level-1 degree are close to $\pm\mathrm{DICT}_i$.

Theorem 7. *(Friedgut's Junta Theorem) Let $f : \{-1,1\}^n \to \{-1,1\}$. Then for every $0 < \epsilon < 1$, there exists some juntas function g on J with $|J| = 2^{O(Inf(f)/\epsilon)}$ such that $\|f-g\|_2 \leq \epsilon$.*

Theorem 8. *(Friedgut-Kalai-Naor) Let $f : \{-1,1\}^n \to \{-1,1\}$ and $W^1(f) = \sum_{|S|=1}\hat{f}_S^2$. If $W^1(f) \geq 1-\epsilon$, then there exists some $i \in [n]$ such that $\|f - DICT_i\|_2 \leq O(\epsilon)$ or $\|f + DICT_i\|_2 \leq O(\epsilon)$. Specially, $W^1(f) = 1$ iff $f = \pm DICT_i$.*

3 Dictatorship versus Quasirandomness Test

Before stating the definition of quasirandom function, we briefly describe the motivations. Functions with quasirandom property can be used to prove inapproximability of constraint satisfaction problems. The Boolean hypercube serves as a gadget in inapproximate reductions. Concretely, we use the "long code" to encode elements in $[n]$ to functions from $\{-1,1\}^n$ to $\{-1,1\}$ by mapping i to $\chi_{\{i\}}$, and decoded by most influential i. We'd like to use some noisy version of influence in order to avoid the case that f is very spiky. And at last, we want to distinguish dictator functions from functions that each coordinate has small noisy influence.

Definition 9. *(noisy influence) Let $f : \{-1,1\}^n \to \mathbb{R}$ and $\delta \in [0,1]$. The i-th δ-noisy influence of f is $Inf_i^{(\delta)}(f) = Inf_i\left(\frac{T_{1-\delta}f}{1-\delta}\right) = \sum_{S\ni i}(1-\delta)^{|S|-1}\hat{f}_S^2$.*

Definition 10. *(quasirandom) Let $f : \{-1,1\}^n \to \mathbb{R}$ and $\epsilon, \delta \in [0,1]$. We say that f is (ϵ,δ)-quasirandom if $Inf_i^{(\delta)}(f) \leq \epsilon$ for all $i \in [n]$.*

Here δ is a parameter. When $\delta = 0$, we have $\mathrm{Inf}_i^{(\delta)}(f) = \mathrm{Inf}_i(f)$. When $\delta = 1$, we have $\mathrm{Inf}_i^{(\delta)}(f) = \hat{f}_{\{i\}}^2$. For intermediate δ, we can think of it as a interpolation. Observe that dictator function $DICT_i$ has $\mathrm{Inf}_i^{(\delta)}(\chi_{\{i\}}) = 1$, therefore it is $(1,0)$-quasirandom. For parity function $\chi_{[n]}$ and positive δ, we have $\mathrm{Inf}_i^{(\delta)}(f) = (1-\delta)^{n-1}\hat{f}_{[n]}^2 \ll 1$ for all $i \in [n]$. The next fact states that even functions far from being quasirandom can only have a small number of variables with large noisy influence.

Fact 11. *Let $f : \{-1,1\}^n \to \{-1,1\}$, and let $J = \{i \in [n] : Inf_i^{(\delta)}(f) \geq \epsilon\}$ be the set of coordinates with large noisy influences. Then $|J| \leq 1/\epsilon\delta$.*

Proof.

$$\epsilon|J| \leq \sum_{i\in J} \mathrm{Inf}_i^{(\delta)}(f) \leq \sum_{i=1}^{n} \mathrm{Inf}_i^{(\delta)}(f) = \sum_{|S|\geq 1} |S|(1-\delta)^{|S|-1} \hat{f}_S^2$$

It remains to prove that $|S|(1-\delta)^{|S|-1} \leq 1/\delta$ for all $S \subseteq [n]$. Note that $(1-\delta)^{|S|-1} \leq (1-\delta)^{i-1}$ for all $i \leq |S|$, therefore

$$|S|(1-\delta)^{|S|-1} \leq \sum_{i=1}^{|S|}(1-\delta)^{i-1} \leq \sum_{i\geq 1}(1-\delta)^{i-1} = 1/\delta$$

□

Definition 12. *(DICT vs QRAND) Let $0 \leq s < c \leq 1$ and $q \geq 1$. A q-good (c,s) dictator versus quasirandom test is a non-adaptive randomized algorithm that makes q non-adaptive queries to a function $f : \{-1,1\}^n \to \{-1,1\}$ and output accepts or rejects. And the test satisfies*

1. *(completeness) If $f = DICT_i$ for some i, $\Pr[accepts] \geq c - o_\epsilon(1)$.*
2. *(soundness) If f is (ϵ,δ)-quasirandom, $\Pr[accepts] \leq s + o_\epsilon(1)$.*

Consider the BLR test, that is we uniformly random sample $x, y \in \{-1,1\}^n$, and accept f iff $f(x)f(y)f(xy) = 1$. Here xy is computed pointwise. This is a linearity test and accepts any dictator function with probability 1. However, it is not a good dictator versus quasirandom test since it also accepts parity function with probability 1. [2] introduces a way to fix this. As we have seen, BLR test can distinguish linear function while NAE test can distinguish dictator and anti-dictator function, both of them use 3 non adaptive queries. Combine them together, we have a 6 non adaptive queries test for dictator functions. Furthermore, we can randomly perform one of the two tests, each with probability 1/2, and reduce the query complexity to 3 while only incurring a constant factor in the rejection probability.

The test above is to identify dictator functions. We will look at tests that specially distinguishes dictator functions and quasirandom functions below. We will consider Hastad's 3-query $3\mathrm{XOR}_\delta$ test, which is BLR test plus some tweaks.

Algorithm 1: Hastad's 3-query $3\mathrm{XOR}_\delta$ test

1. Pick $x, y \in \{-1,1\}^n$ uniformly and independently;
2. Pick $z \sim N_{1-\delta}(xy)$, that is z_i is independently chosen to be $x_i y_i$ with probability $1-\delta/2$ and $-x_i y_i$ with probability $\delta/2$;
3. Query f on x, y and z;
4. Accept iff $f(x)f(y)f(z) = 1$;

Claim 13. *(Hastad) Let $\delta = \epsilon$ and f be balanced, the test above is $(1, 1/2)$ dictator versus (ϵ,ϵ)-quasirandom test.*

Proof. First, it is easy to check that dictator functions pass with probability $1-\epsilon/2$.

For soundness part, we have

$$\begin{aligned}
Pr[\text{3XOR}_\epsilon \text{ accepts } f] &= \frac{1}{2}+\frac{1}{2}\mathbb{E}_{x,y,z}[f(x)f(y)f(z)] \\
&= \frac{1}{2}+\frac{1}{2}\mathbb{E}_{x,y}[f(x)f(y)(T_{1-\epsilon}f)(xy)] \\
&= \frac{1}{2}+\frac{1}{2}\mathbb{E}_{u}[\mathbb{E}_x[f(x)f(xu)](T_{1-\epsilon}f)(u)] \\
&= \frac{1}{2}+\frac{1}{2}\mathbb{E}_{u}[(f*f)(u)(T_{1-\epsilon}f)(xy)] \\
&= \frac{1}{2}+\frac{1}{2}\sum_S (1-\epsilon)^{|S|}\hat{f}_S^3 \leq \frac{1}{2}+\frac{1}{2}\max_S\{(1-\epsilon)^{|S|}\hat{f}_S\}
\end{aligned}$$

We then prove that $(1-\epsilon)^{|S|}\hat{f}_S \leq \sqrt{\epsilon}$ for all S if f is (ϵ,ϵ)-quasirandom, from which the claim follows. Suppose for contradiction not, then $\sqrt{\epsilon} < (1-\epsilon)^{|S|}\hat{f}_S$. So $\epsilon < (1-\epsilon)^{2|S|}\hat{f}_S^2 \leq (1-\epsilon)^{|S|-1}\hat{f}_S^2 \leq \text{Inf}_i^{(1-\epsilon)}(f)$ for all $i \in S$, which is contradiction when $S \neq \emptyset$. Therefore we conclude the claim. $\square$

References

[1] Ryan O'Donnell, Karl Wimmer: *Influences, Coalitions, and the Tribes function.* CMU 18-859S 2007 Lecture 14

[2] Ryan O'Donnell, Li-Yang Tan: *Analysis of Boolean Functions.* Barbados Workshop on Computational Complexity 2012