Analysis of Boolean Functions April 30 2013

Lecture 9

Lecturer: Irit Dinur *Scribe: Mohamamd Bavarian*

1 Introduction

Today is the last lecture. So let's start with some reviewing of the basics and the highlights of the course. Given any function $f : \{\pm 1\}^n \to \mathbb{R}$ we defined the Fourier transform as the expansion of the function in terms of the character functions $\chi_S(S) = \prod_{i \in S} x_i$. We saw that many combinatorial properties of subsets of Boolean hypercube, such as the edge-expansion, have analytic analogues in terms of the influences of the function. As you perhaps should recall these were defines by

$$\inf_i(f) = \sum_{S \ni i} \widehat{f}(S)^2 \qquad I[f] = \sum_{i=1}^{n} \inf_i[f] = \sum_{S \subseteq [n]} |S| \widehat{f}(S)^2 \,.$$

One of the most important tools in our study was the noise operator T_ρ defined by

$$T_\rho f(x) = \sum_{\alpha} \widehat{f}(S) \chi_\alpha(x) \,.$$

Many of the *structural theorems* for functions of low average-sensitivity that we saw, such as FKN, KKL and Friedgut's theorem, used hypercontractvity of this operator at their core. We also introduced some of the applications of Boolean Fourier analysis to complexity theory via dictatorship vs quasi randomness test which is an important step in almost all optimal inapproximibility results. Related to this we saw the invariance principle of Mossel et al a variant of Berry-Essen theorem for low-degree polynomials. According to the invariance principle, for any τ-quasi-random low-degree polynomials $p(X)$ of degree d we have

$$\forall t \in \mathbb{R} : \quad \left| \Pr_{X \in G^n}[P(X) \leq t] - \Pr_{A \in \{\pm 1\}^n}[P(A) \leq t] \right| = O\left(C(\tau, d) \|P\|_2\right) \,.$$

Today's topic is about *derandomization* of the above results. More specifically, we will discuss the work of Barak et al [BGH$^+$12] giving the example of a graph with poly(n) vertices and the top (adjacency) eigenvalues resembling that of noisy hypercube graph. As the noisy hypercube was the main gadget used in UG-hardness results such as work of Khot et al on UG-hardness of MAX-cut [KKMO07], this gives much more efficient reductions.

However, there was another motivation for the result of Barak et al which we shall focus on on this lecture. For explaining this motivation, first we need to recall the notion of *small-set expansion*. The prototypical example of small-set expanders is Boolean hypercube and its noisy version for different parameter regimes. Small set expansion is a very natural combinatorial property of a graph so it might be interesting to study the complexity of deciding whether a graph is small-set expander or not without any further motivation. However, as it turns out the concept is also closely to one of the most important outstanding conjectures in hardness of approximation which is the Unique Games Conjecture. The closely related, small-set expansion hypothesis, defined and investigated by Steurer and Raghavendra [RS10], is as follows.

Conjecture 1 (SSE hypothesis) *For any $\epsilon > 0$ there exists a $k > 0$ so that given a (d-regular) graph $G = (V, E)$ it is NP-hard to distinguish from these two cases,*

1. *Yes instance: There exists $S \subseteq V$ of size $\leq |V|/k$ such that $E(S, S^c) \leq \epsilon d|S|$.*
2. *No instance: For any set $S \subseteq V$ of size $\leq |V|/k$ we have $E(S, S^c) \geq (1 - \epsilon)d|S|$.*

The No instances of the above problem is what is usually called (k, ϵ)-small set expander. More formally,

Definition 2 *A graph $G = (V, E)$ is called a (k, ϵ) small-set expander if for any set $S \subseteq V$ with $\sum_{v \in S} d_v \leq \frac{2|E|}{k}$ we have $E(S, S^c) \geq (1 - \epsilon) \sum_{v \in S} d_v$.*

In a surprising development, Arora et al [ABS10] gave a subexponential for the SSE problem. The main observation that allowed Arora et al to achieve their algorithm was that any small-set expander graph has at most n^ϵ eigenvalues greater than $1 - \epsilon$. Using the above observation combined with an algorithm dubbed as "subspace enumeration" (see Kolla et al [KT08, Kol10]) this gave the sub exponential algorithm for unique games. The runtime of their algorithm, however, crucially, depended on the upper bound on the number of n^ϵ for the number of number of eigenvalues larger than $1 - \epsilon$ for small-set expanders. Indeed at the time, the worst case lower bound in the number of such high eigenvalues was poly-logarithmic in n as opposed to n^ϵ which was achieved by noisy Boolean hypercube. Had this been always the case it would have meant that the algorithm of Arora et al might actually run in quasi-polynomial time on any SSE instance which would assuming exponential time hypothesis disprove the SSE conjecture (and most likely, the unique games conjecture). The derandomization of noisy hypercube by Barak et al quashed those hopes by providing an example of a graph with much fewer vertices than noisy Boolean hypercube with the same structure of top eigenvalues. This meant that lemma of Arora et al was tight and hence that line of algorithmic attack, at least without further insight, would not be able to give a quasi polynomial time algorithm for UG or SSE. The theorem of Barak et al is the following,

Theorem 3 *For every constant $\epsilon > 0$, there is an n-vertex small-set expander graph with $2^{(\log n)^{\Omega(1)}}$ eigenvectors with eigenvalues greater than $1 - \epsilon$.*

2 Construction of the short code

To appreciate the short code, let's first review the long code and its useful properties in proving hardness of approximation of results. The long code is $(2^{2^k}, k, 1/2)_2$ code with messages given by $v \in \mathbb{F}_2^k$ and encoding of v by a list of of size 2^{2^k} of evaluations all possible functions $f : \mathbb{F}_2^k \to \mathbb{F}_2$ on v. Let $N = 2^k$ and identify $\mathbb{F}_2^k$ with $[N]$ to see that 2^N functions $f : [N] \to \mathbb{F}_2$ now naturally correspond to vertices of Boolean hypercube $\{0, 1\}^N$. Notice that the space of all function $f : \mathbb{F}_2^k \to \mathbb{F}_2$ is a group via addition. The noisy Boolean hypercube denoted by $H_{N,\epsilon}$ in this setting can be seen as the Cayley graph associated with this group with the (weighted) set of generators as

$$\forall f : \mathbb{F}_2^k \to \mathbb{F}_2 \,,\, \mathrm{wt}(f) := (1 - 2\epsilon)^{|S_f|} \qquad S_f := \{v \in \mathbb{F}_2^k : \, f(v) = 1\} \,.$$

This weight structure exactly determines the the adjacency matrix of corresponding Cayley graph $H_{N,\epsilon}$ by associate the weight of an edge $e = (g_1, g_2)$ with $\mathrm{wt}(g_1 - g_2)$. (Notice that characteristic is 2 here so $g_1 - g_2 = g_2 - g_1$) The main property of $H_{N,\epsilon}$ and the long code that we will try to preserve is the local testability and decidability of this graph which we encountered in the context *dictator vs quasi-randomness* test. What was important there was that the N eigenfunction corresponding to dictators corresponded to very low-value cuts in $H_{N,\epsilon}$ whereas any cut that was sufficiently pseudorandom (i.e. the Fourier mass of its characteristic function was not concentrated on low degrees) had much higher expansion.
We plan to derandomize $H_{N,\epsilon}$ by taking a small pseudorandom subspace of $\{0, 1\}^N$. The idea of derandomization is to take the following the "short code" approach by encoding a $v \in \mathbb{F}_2^k$ as follows

$$E : v \to E(v) = (f(v))_{f \in P_d} \,.$$

Where P_d here the set of all polynomials of degree d for some constant $d = O(1)$. So our short code will be a $(2^{\sum_{i \leq d} \binom{k}{i}}, k, 1/2)$ code. Since the set of polynomials of degree at most d is a subspace we see that our short code $\mathcal{D}$ can be seen as a subspace of $\{0, 1\}^N$.
In order to fully specify our derandomization of $H_{N,\epsilon}$ we now must describe the edge structure of $\mathcal{D}$ to turn it into a Cayley graph. The edge structure of $\mathcal{D}$ would be very simple. For a $f \in \mathcal{D}$, we put an edge among

f and g with $f-g$ is a codeword of minimum weight in $\mathcal{D}$, i.e. of the weight 2^{-d}. This completes the description derandomized noisy cube of our $G = \mathrm{Cay}(\mathcal{D})$. A useful view of this graph which is important in analysis of eigenvalues of G is by considering the canonical tester $\mathcal{T}$ for membership in the code $\mathcal{C} = \mathcal{D}^{\perp}$ which is dual of $\mathcal{D}$. Given an input α, the canonical tester $\mathcal{T}$ picks an element of $q \in \mathcal{D}$ of minimum weight weight uniformly at random and rejects α if and only if $\langle \alpha, q \rangle = 1$. The soundness of this test $\mathcal{T}$ and its smoothness properties, i.e. the fact that $\mathcal{T}$ is pairwise independent, will be very important in the analysis of Barak et al of the eigenvalue structure of G. More concretely, they will use the following theorem of Bhattacharyya et al.

Theorem 4 ([BKS$^+$10]) *There exists a constant $\eta_0 > 0$ such that for all n, d and $k < \eta_0 2^d$ the tester $\mathcal{T}$ describes above has soundness $s(k) \geq \frac{k}{2} 2^{-k}$ where the soundness of $\mathcal{T}$ defined as,*

$$s(k) := \min_{\substack{\alpha \in \mathbb{F}_2^N \\ \Delta(\alpha, \mathcal{C}) \geq k}} \Pr_{q \sim \mathcal{T}} [\langle \alpha, q \rangle = 1] \, .$$

The above lower bound on the soundness of $\mathcal{T}$ is important because of the following theorem of Barak et al.

Theorem 5 (SSE for $\mathrm{Cay}(\mathcal{C}, \mathcal{T})$) *Let $\mathcal{C}$ be an $[N, K, D]_2$ linear code that has canonical tester $\mathcal{T}$ with query complexity ϵN and soundness curve $s(\cdot)$ and $k < D/5$. The graph $G = \mathrm{Cay}(\mathcal{C}^{\perp}, \mathcal{T})$ has 2^{N-K} vertices and at least $N/2$ eigenvalues larger than $1 - 4\epsilon$. All subsets S of G will then satisfy*

$$\Phi(S) \geq 2s(k) - 3^k \sqrt{\mu(S)} \, .$$

Notice that in above theorem implies that the guarantee for the expansion improves as size of S gets smaller. By picking k appropriately in theorem 5 and using results from [BKS$^+$10] on soundness of canonical Reed-Muller test , Barak et proves the theorem 3. In the remainder of the lecture, we shall develop the necessary tools to prove theorem 5.

3 Fourier analysis on the short code

Let $G = \mathrm{Cay}(\mathcal{D}, \mathcal{T})$ be the short code. We want to develop Fourier analysis for functions $f : G \to \mathbb{R}$. Let $\{\chi_\alpha\}$ be the set of eigenvectors of adjacency matrix of G. It is well known that these form an complete orthonormal set of functions over the vector-space of real-valued functions over G as

$$\mathbb{E}_{x \in G}[\chi_\alpha(x)\chi_\beta(x)] = \delta_{\alpha\beta} \, .$$

So the expansion of a function in terms of χ_α is going to play the role of Fourier characters in this setting. Now to make this a full-fledged Fourier transform we need a notion of "weight" for various characters and we shall related this notion of weight to the eigenvalue λ_α corresponding to χ_α. The main observation is the following easy proposition,

Proposition 6 *Every Fourier character χ_α where $\alpha \in \mathbb{F}_2^N$ induces an eigenvector to $G := \mathrm{Cay}(\mathcal{D}, \mathcal{T})$. Any two character α and β induce the same character if and only if $\alpha - \beta \in \mathcal{C} = \mathcal{D}^{\perp}$.*

So we can identify the set of eigenfunction of G with $\mathbb{F}_2^N / \mathcal{C}$. Then its natural to define Fourier weight as follows

$$\deg(\chi_\alpha) = \min_{c \in \mathcal{C}} \mathrm{wt}(\alpha + c) = \Delta(\alpha, \mathcal{C}) \, .$$

Finally the following proposition relates the eigenvalue λ_α to soundness of the tester.

Proposition 7 *For any $\alpha \in \mathbb{F}_2^N$, $\lambda_\alpha = 1 - 2s(\alpha)$.*

Proof The edges of G correspond to picking random q according to the tester $\mathcal{T}$. So we see for any $v \in G$ we have

$$\lambda_\alpha \chi_\alpha(v) = \mathbb{E}_{q\sim\mathcal{T}}[\chi_\alpha(q+v)] = \mathbb{E}_{q\sim\mathcal{T}}[(-1)^{\alpha\cdot q}]\chi_\alpha(v) = \chi_\alpha(v)(1 - 2\Pr[\alpha \cdot q = 1]) .$$

So $\lambda_\alpha = 1 - 2s(\alpha)$. ■

Hence, we can see the significance of result of Bhattacharyya et al for the analysis of the short code; the soundness of the canonical test for $\mathcal{C}$ translates to the fact high degree characters α have eigenvalue bounded away from 1. This is analogous to the case of noisy hypercube where the eigenvalue was proportional to $(1-2\epsilon)^{|\alpha|}$.
Now having established the proper setting for Fourier analysis over the short code we shall go ahead and prove the theorem 5.
Proof Consider the N dictator characters χ_i. Let $p_i = \Pr_{q\sim\mathcal{T}}[q_i = 1]$. We know that $\mathbb{E}_{i\in[N]}[q_i] = \epsilon$ because the test has ϵN query complexity. Hence by Markov inequality for at least $N/2$ indices $p_i \leq 2\epsilon$. Since $\lambda_i = \mathbb{E}[(-1)^{q_i}] = 1 - 2p_i \geq 1 - 4\epsilon$, we can deduce the first claim.
Let $S \subseteq V(G)$ and consider the function $f : G \to \{0,1\}$ to be the characteristic function of S. We have $\Phi(S) = 1 - \Pr_{x\sim y}[y \in S | x \in S]$. Hence, we have

$$\mu(S)(1-\Phi(S)) = \mathbb{E}_{x\in S, y\in N(S)}[f(x)f(y)] = \mathbb{E}_{q\sim T,x}[f(x)f(x+q)]$$

$$\begin{aligned} &= \sum_{\alpha\in\mathbb{F}_2^n/\mathcal{C}} \widehat{f}_\alpha^2 \lambda_\alpha \leq \sum_{\mathrm{wt}(\alpha)\leq k} \widehat{f}_\alpha^2 \lambda_\alpha + (1-2s(k))\mathbb{E}[f^2] \\ &= \mu(S)(1-2s(k)) + \|f^{\leq k}\|_2^2 . \end{aligned}$$

This means $\Phi(S) \leq 2s(k) - \frac{1}{\mu(S)}\|f^{\leq k}\|_2^2$. Now we want to prove an upper bound on $\|f^{\leq k}\|_2^2$. This is the setting of $2 \to 4/3$ bound in Boolean Fourier analysis. We claim in this case $\|f^{\leq k}\|_2^2 \leq 3^k \|f\|_{4/3}^2$ also follows if we have the $4 \to 2$ inequality $\|f^{\leq k}\|_4 \leq \sqrt{3}^k \|f\|_2$. (This is because Holder inequality argument for $2 \to 4$ bound to $4/3 \to 2$ bound is generic and applies to any probability space.) Now the $4 \to 2$ bound itself follows from the same bound on Boolean hypercube by following operation: Define $g : \{0,1\}^n \to \mathbb{R}$ by simply setting

$$g(x) = \sum_{\alpha\in A} \widehat{f}_\alpha \chi_\alpha(x) ,$$

where A is the set of minimal weight representatives of the cosets $\alpha \in \mathbb{F}_2^n/C$ for $\deg(\alpha) \leq k$. Now applying our inequality to g we have

$$\|g\|_4 \leq \sqrt{3}^k \|g\|_2 \leq \sqrt{3}^k \|f^{\leq k}\|_2 \leq \sqrt{3}^k \|f\|_2 .$$

Now the crucial point here is that $\mathbb{E}[g^4] = \mathbb{E}[(f^{\leq k})^4]$. To see that, first notice that degree of g^4 is $4k < D$. Now, since distance of $\mathcal{C}$ is D, its dual the coordinates of the codewords of $\mathcal{C}$'s dual $\mathcal{D}$ has D-wise independence. This means all monomials of degree $\leq D$ have exactly expectation as if they were unbiased Bernoulli RV just as in Boolean hypercube. This indeed proves $\|g\|_4 = \|f^{\leq k}\|_4$ which finishes the proof. ■

4 Conclusion

The paper of Barak et al contains many more interesting results on the short code. Chief among these is the derandomization of "majority is stablest" theorem and the invariance principle. This essentially means that the short code can replace the long code on the UG-hardness of max-cut. Given that we saw that the small-set expansion property also derandomized in this case, it is natural to ask what other theorems in Boolean Fourier analysis can be derandomized? What about p-biased Fourier analysis which also has many applications in hardness of approximation. Maybe one good place to start could be the Friedgut-Kalai-Naor theorem.

References

[ABS10] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. In *Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 563–572. IEEE Computer Society, 2010.

[BGH+12] Boaz Barak, Parikshit Gopalan, Johan Hastad, Raghu Meka, Prasad Raghavendra, and David Steurer. Making the long code shorter. In *Proceedings of the 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 370–379, 2012.

[BKS+10] Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. Property testing. chapter Optimal Testing of Reed-Muller Code, pages 269–275. 2010.

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM J. Comput.*, 37(1):319–357, April 2007.

[Kol10] Alexandra Kolla. Spectral algorithms for unique games. In *Computational Complexity (CCC), 2010 IEEE 25th Annual Conference on*, pages 122–130. IEEE, 2010.

[KT08] Alexandra Kolla and Madhur Tulsiani. Playing random and expanding unique games. *Unpublished manuscript*, 47, 2008.

[RS10] Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the 42nd ACM symposium on Theory of computing*, pages 755–764. ACM, 2010.