COS 598 Week 8
Degree/discrepancy Theorem.
Group representations

Lecturers: Moritz Hardt/Boaz Barak
Scribe: Simion Filip

April 4, 2008

1 Degree/discrepancy theorem

The ideas presented in this section are due to [She07]. The heuristic idea of the construction is as follows. We start with some Boolean function

$$f : \{-1, 1\}^n \to \{-1, 1\} \text{ of "high degree"}$$

We then produce a pair

$$\begin{cases} \text{"Nice" distribution} \mu \text{ on } \{-1,1\}^n \\ \text{"Useful" matrix } M \end{cases}$$

such that $\mathrm{disc}_\mu(M)$ is *low*. This can turn out to be useful in proving lower bounds. To begin with, we need the following

Definition 1.1 (Threshold degree)**.** We will say that the threshold degree of f is at most d (and write $thr(f) \leq d$) if there exists a degree d polynomial P (with real coefficients, but can in fact assume integral) such that $\forall x \in \{-1, 1\}^n$, $f(x) = sgn(P(x))$.

With this definition, we are ready to state the first ingredient of the construction.

Theorem 1.2. *If $thr(f) = d$ then there exists a distribution μ such that*

$$\forall P \in \mathbb{R}[x], \textit{with deg } P < d,\ \mathbb{E}_\mu(f \cdot P) = 0$$

Sketch. We shall make use of Farkas' Lemma, a very nice discussion of which is provided in [Tao]. The statement of the lemma is simple: If we have a system of linear inequalities $P_i(x) \geq 0$, then either it has a solution, or it implies a contradiction, i.e. there exist $q_i \geq 0$ such that $\sum q_i P_i(x) = -1$. In our case, we have three types of conditions

$$\begin{cases} \mathbb{E}_\mu(f \cdot x^\alpha) = 0, \text{ where } x^\alpha \text{ is any monomial of degree strictly less than } d \\ \mu(v) \geq 0, \text{ where v is any vector on n bits} \\ \mathbb{E}_\mu(1) - 1 \geq 0 \end{cases}$$

Remark that $A = 0$ is equivalent to $A \geq 0$ and $-A \geq 0$. A little explanation of the conditions is in order. The middle condition forces μ to be a positive function on the hypercube, while the last one forces it to be non-trivial. Once we get such a non-trivial μ satisfying the above, we can just renormalize, since everything is homogenous in μ. So, we can assume, by Farkas' lemma, that there is a contradiction in the system. Let us write it, using linearity of expectation and assembling the linear combinations of the first type of conditions into a single polynomial

$$\mathbb{E}_\mu(f \cdot P) + \sum_{v \in \{-1,1\}^n} q_v \mu(v) + Q(\mathbb{E}_\mu(1) - 1) = -1 \Leftrightarrow$$

$$\sum_{v \in \{-1,1\}^n} \mu(v) f(v) P(v) + \sum_{v \in \{-1,1\}^n} q_v \mu(v) + Q[\sum_{v \in \{-1,1\}^n} \mu(v)] - Q = -1$$

Recall that in our considerations, the variables are $\mu(v)$. So, the only chance is that $Q = 1$, and all the other positions cancel, i.e.

$$f(v)P(v) + q_v + 1 = 0$$

But using that $q_v \geq 0$, we have that $f(v)P(v) < 0, \forall v \in \{-1,1\}^n$. So, $-P$ gives a contradiction to the assumption on the threshold degree of f □

With this, we have accomplished the first part of our program - obtaining a "nice" distribution. Now, we go on an construct the matrix. Again, we are given a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ and some $N \geq n$. Consider the matrix M with 2^N rows and $\binom{N}{n}$ columns. The rows are indexed by vertices in the corresponding hypercube, while the columns by subsets on n elements of $[N]$. If we are given a vertex of the hypercube X and a subset S, $M_{X,S} := f(X|S)$, i.e. the corresponding entry of M is given by the value of f when projecting the vertex on the subset given by S.

Another, more fruitful way to interpret this is as follows. We have a row and a columns player. They both know the function f, but the row player knows the N-vector X, while the column player knows the subset of X on which f has to be evaluated. Bounds on the discrepancy of M would give bounds on the communication complexity of this game. The main theorem in this sense is

Theorem 1.3 ([She07]). *If f has threshold degree at least d, then there exists a distribution λ such that*

$$disc_\lambda(M) \leq \left(\frac{O(n^2)}{Nd}\right)^d$$

Before we embark on the proof of this theorem, we make a few remarks on the method and some preliminary assumptions. First of all, recall that the discrepancy with respect to a measure λ was defined as

$$\mathrm{disc}_\lambda(M) = \max_{\text{Rectangles R}} \sum_R \lambda(X,S) M_{X,S}$$

It can be shown (by a simply probabilistic argument) that

$$\begin{aligned} &\exists \alpha_X \in \{-1,1\}, \beta_S \in \{-1,1\} \text{ such that} \\ &\mathrm{disc}_\lambda(M) \le \sum_{X,S} \alpha_X \beta_S \cdot \lambda(X,S) M_{X,S} \end{aligned}$$

In fact, the above inequality can be reversed up to some constants. The plan of the proof then is to first produce a measure λ on our big matrix M, and to estimate $\mathrm{disc}_\lambda(M)^2$ instead of just $\mathrm{disc}_\lambda(M)$. We will get some cancelations after the expansion of the squared quantity.

Proof. From the fact that $trh(f) \ge d$, we deduce that there exists a measure μ as in theorem 1.2. Define then the measure λ by

$$\lambda(X,S) := \mu(X|S) \cdot 2^{-N+n} \cdot \binom{N}{n}^{-1}$$

The additional factors are for normalization purposes. In the following chain of inequalities, all the expectations are taken with respect to the *uniform* distribution of X and S:

$$\begin{aligned} \mathrm{disc}_\lambda(M)^2 &= \left(\sum_{X,S} \alpha_X \beta_S \lambda(X,S) f(X,S) \right)^2 \\ &= \left(2^n \mathop{\mathbb{E}}_{X\sim U} \mathop{\mathbb{E}}_{S\sim U} \alpha_X \beta_S \cdot \mu(X|S) f(X|S) \right)^2 \\ &= \left(2^n \mathop{\mathbb{E}}_{X\sim U} \alpha_X \mathop{\mathbb{E}}_{S\sim U} \beta_S \cdot \mu(X|S) f(X|S) \right)^2 \\ &\le 4^n \underbrace{\mathbb{E}\alpha_X^2 \left(\mathbb{E}\beta_S \cdot \mu(X|S) f(X|S) \right)^2}_{\text{by Cauchy-Schwartz}} \\ &= 4^n \underbrace{\mathop{\mathbb{E}}_{S,T} \beta_S \beta_T \mathop{\mathbb{E}}_{X} \mu(X|S) f(X|S) \mu(X|T) f(X|T)}_{\text{using } \alpha_X^2 = 1 \text{ and expanding}} \\ &\le 4^n \mathop{\mathbb{E}}_{S,T} | \underbrace{\mathop{\mathbb{E}}_{X} \mu(X|S) f(X|S) \mu(X|T) f(X|T)}_{\Gamma(S,T)} | \end{aligned}$$

To sum up, we have obtained

$$\mathrm{disc}_\lambda(M)^2 \le 4^n \sum_{k=0}^{n} \mathbb{P}(|S \cap T| = k) \cdot \max_{|S\cap T|=k} |\Gamma(S,T)|$$

Now, the necessary step is given by the following

Claim. *If* $|S \cap T| < d$ *then* $\Gamma(S,T) = 0$

Proof. Without loss of generality, let us assume that $S = \{1, \dots, n\}$ and $T = \{1, \dots, k\} \cup \{n+1, \dots, 2n-k\}$. Then, the quantity to be considered is

$$\mathop{\mathbb{E}}_{x_1,\dots x_n} \mu(x_1, \dots, x_n) f(x_1, \dots, x_n) \cdot$$
$$\cdot \underbrace{\mathop{\mathbb{E}}_{x_{n+1},\dots,x_{2n-k}} \mu(x_1, \dots, x_k, \dots, x_{n+1}, \dots, x_{2n-k}) f(x_1, \dots, x_k, \dots, x_{n+1}, \dots, x_{2n-k})}_{P(x_1,\dots,x_k),\text{ multilinear, of degree } \leq k}$$

We can thus apply Theorem 1.2 to the second factor and conclude that the expression indeed vanishes. ☐

To conclude, we just need to estimate the choices of some sets configurations. When $|S \cap T| > d$, the expression can be estimated (using that the expectations involved are ≤ 1)

$$2^{-n} \mathop{\mathbb{E}}_{\mu}(f \cdot 2^{-n+k} \mathop{\mathbb{E}}_{\mu}(f)) \leq 2^{-2n+k} = 4^{-n} \cdot 2^k$$

We also use

$$\begin{aligned} \mathbb{P}(|S \cap T| = k) &= \underbrace{\binom{n}{k}}_{\cap \text{ with } S} \underbrace{\binom{N-n}{n-k}}_{\text{outside } S} \underbrace{\binom{N}{n}^{-1}}_{normalization} \\ &\leq \left(\frac{en^2}{Nk}\right)^k \end{aligned}$$

In conclusion,

$$\mathrm{disc}_\lambda(M) \leq \sum_{k \geq d}^{n} \left(\frac{en^2}{Nk}\right)^k \cdot 2^k \leq \left(\frac{O(n^2)}{Nd}\right)^d$$

as promised. ☐

1.1 Applications

Here, we sketch a few an application of the developed method. One can use the theorem proved to separate $\mathtt{AC}^0$ from Majority of Threshold circuits. The idea is to exhibit a function that is computed by a depth 3 circuit, but has exponentially small discrepancy. Combined with a result of Nisan who shows that Majority of Threshold circuits can compute only functions of high discrepancy or has to use a large number of threshold gates, this gives the separation. The result of Nisan has behind the idea that there exists an efficient communication protocol for Majority of Threshold functions.

Proceeding to the construction, one defines the Minsky-Papert function on $n = 4m^3$ variables.

$$\mathrm{MP}_m(x) := \bigvee_{i=1}^{m} \bigwedge_{j=1}^{4m^2} x_{i,j}$$

It is a result of Minsky and Papert that MP_m has threshold degree m. Using Theorem 1.3, we see that this function has discrepancy $\exp(-\Omega(n^{1/5}))$ with respect to a given distribution. To see this, just apply the theorem with $N \gg n^2$.

Another application of this method was found recently by Lee-Shraibman and Chattopadhyay-Ada in proving a lower bound on the Disjointness predicate in the multiparty number-on-the-forehead model.

Many thanks to Moritz for help with expanding this section

2 Group Representations

This section is intended to provide a brief introduction to the main facts of the representation theory of finite groups. The main idea is to study a group by considering its possible actions on a simple object, such as a finite dimensional vector space.

2.1 Abelian groups

Let us first consider the case of an abelian group G with n elements.

Definition 2.1. A map $\rho : G \to \mathbb{C}$ is said to be a representation of G if and only if it is a group homomorphism, i.e.

$$\rho(xy) = \rho(x)\rho(y)$$

Remark 2.2. *For this to hold, it is clear that $\rho(id) = 1$, and moreover, since the group is finite, it must be that $\forall g \in G, |\rho(g)| = 1$. Moreover, $\rho(g^{-1}) = \overline{\rho(g)}$. Another way to view a representation in this case is by considering ρ as a function on the group. We then have the following inner product*

$$\langle \rho, \mu \rangle := \mathop{\mathbb{E}}_{g \in G} \rho(g)\overline{\mu(g)}$$

The main result concerning abelian groups is the following

Theorem 2.3. *The set of representations of G, viewed as a subset of the vector space of functions $G \to \mathbb{C}$, is an orthonormal basis.*

Proof. The first part is to show that if $|G| = n$, then we have at least n distinct representations. We will then show that any two representations are orthonormal (if distinct), and this will conclude the proof, sice the dimension of the space of functions on G is n.

If $G = \mathbb{Z}/(m)$ a cyclic abelian group, then we fix a primitive m^{th} root of unity ω_m and define the m representations by $\rho_i(j) = \omega_m^{i \cdot j}$, where $i = 0 \ldots m-1$.

We now remark that if we have two abelian groups G_1, G_2, and representations ρ, μ of each, one can have a representation of $G_1 \times G_2$ given by $\gamma(g_1 \times g_2) = \rho(g_1)\mu(g_2)$. It is easy then to see that the above construction extends to arbitrary products of cyclic groups, and using the classification theorem for finite abelian groups which states that all of them are products of cyclic groups, we are done with the first part.

Orthogonality follows from the following chain of identities

$$\begin{aligned}\langle \rho, \mu \rangle &:= \mathop{\mathbb{E}}_{g \in G} \rho(g)\overline{\mu(g)} = \mathop{\mathbb{E}}_{g \in G} \rho(g_0 g)\overline{\mu(g_0 g)} \\ &= \mathop{\mathbb{E}}_{g \in G} \rho(g_0)\overline{\mu(g_0)}\rho(g)\overline{\mu(g)} \\ &= \frac{\rho(g_0)}{\mu(g_0)} \mathop{\mathbb{E}}_{g \in G} \rho(g)\overline{\mu(g)} = \frac{\rho(g_0)}{\mu(g_0)} \langle \rho, \mu \rangle\end{aligned}$$

So, if ρ and μ are distinct, it is clear that it must be that $\langle \rho, \mu \rangle = 0$ □

Another useful operation on representations of abelian groups is the convolution. Formally, it is given by

$$(\rho * \mu)(g) := \sum_{hh'=g} \rho(h)\mu(h')$$

Apriori, it is not clear that this is associative and even why it would make sense to study it. But a better interpretation of it is to view a representation as an element of the algebra $\mathbb{C}[G]$, i.e. the algebra generated freely by the elements of G, with the relations among them:

$$\rho = \sum \rho(g) \cdot g \in \mathbb{C}[G]$$

Then, convolution is just multiplication in this algebra. Useful results about convolutions are

Theorem 2.4. *For two distinct representations ρ, μ, we have that*

$$\rho * \rho = n\rho \text{ and } \rho * \mu = 0$$

Proof. The first statement is trivial:

$$(\rho * \rho)(g) = \sum_{hh'=g} \rho(h)\rho(h') = \sum_{hh'=g} \rho(g) = n\rho(g)$$

For the second one, write $\rho(h) = \rho(g/h') = \rho(g) \cdot \rho^{-1}(h')$, where $hh' = g$.

$$(\rho * \mu)(g) = \sum_{hh'=g} \rho(h)\mu(h') = \sum_{hh'=h} \rho(g)\rho^{-1}(h')\mu(h') = \rho(g)\langle \rho^{-1}, \mu \rangle = 0$$

□

Corollary. *If we write two functions on G in terms of the representations: $F = \sum a_i \rho_i, H = \sum b_i \rho_i$, then*

$$F * H = \frac{1}{n} \sum a_i b_i \rho_i$$

2.2 General finite groups

We would like to generalize our notion of representation to general non-abelian groups. Of course, because of the lack of commutativity, to give ourselves more freedom, we should consider $\mathrm{GL}_n(\mathbb{C})$. In it's outmost generality, we can consider

Definition 2.5. An n-dimensional representation of a group G is a homomorphism $\rho : G \to \mathrm{GL}_n(\mathbb{C})$.

It turns out to be extremely useful to have an invariant (with respect to the action of G) inner product on $\mathbb{C}^n$. Since the group is finite, we can pick any inner product and then just average over the whole group. This allows us to replace in the above definition GL by U, that is, we can assume the representation is given by *unitary* matrices.

There are several important notions regarding representations that must be discussed. First, note that if we have two representations $\rho : G \to \mathrm{U}(V)$ and $\mu : G \to \mathrm{U}(W)$, then we can form their direct sum

$$\rho \oplus \mu : \mathrm{U}(V \oplus W)$$

Moreover, when we have another group H with representation $\gamma : H \to \mathrm{U}(T)$, we can form their tensor product

$$\rho \otimes \gamma : G \times H \to \mathrm{U}(V \otimes T)$$

Example 2.6 (The regular representation)**.** Every group acts on itself, so we can consider a $n = |G|$-dimensional vector space V whose coordinates are indexed by the elements of the group, and where G acts by permutation. This gives us a representation $REG(g)$, where the image are some permutation matrices.

As we saw above, having a few representations allows one to construct many others. One of the main goals of representation theory is to understand what are *all* the possible representations of a given group G. It turns out that all of them are direct sums of a finite number of building blocks.

Definition 2.7. A representation ρ on a vector space W is said to be *reducible* if it is the direct sum of two other representations. Equivalently, there exists a subspace $V \subset W$ such that $\rho(G)V = V$. Then $W = V \oplus V^{\perp}$. If it is not reducible, we call it *irreducible.*

It is clear from the above definition that it is enough to know all the irreducible representations and to find out a way to decompose a representation into irreducible ones. We shall need a few more definitions:

Definition 2.8. A morphism of representations $\rho : G \to \mathrm{U}(V), \mu : G \to \mathrm{U}(W)$ is a map between vector spaces $T : V \to W$ such that for any $g \in G$, $T \cdot \rho(g) = \mu(g) \cdot T$.

Definition 2.9. The character of a representation ρ is the map $\chi_\rho : G \to \mathbb{C}$ given by $\chi_\rho(g) = \operatorname{tr} \rho(g)$.

The main theorem we shall prove states that given two distinct irreducible representations, their characters are orthogonal. This parallels the corresponding result on commutative groups. Moreover, the reader can check (although it will also follow from our proof) that the irreducible representations of an abelian group are all 1-dimensional. This follows because a set of matrices that commute can be simultaneously diagonalized. Let us now state the main orthogonality result:

Theorem 2.10 (Orthogonality of characters)**.** *Suppose that ρ and μ are two distinct irreducible representations of G. Then*

$$\langle \chi_\rho, \chi_\mu \rangle = 0$$

Let us first deduce a few important consequences of the above theorem. First, it is clear that any representation decomposes as the direct sum of finitely many irreducible ones, and this decomposition if unique up to the order of factors. Suppose we thus have $\rho = \rho_1 \oplus \ldots \oplus \rho_k$, where all the ρ_i's are irreducible. If we are given another irreducible representation μ, then

$$\langle \chi_\rho, \chi_\mu \rangle = \text{ \# of } \rho_i\text{'s isomorphic to } \mu$$

The main consequence of this is that the regular representation contains *all* representations of G, and with multiplicities we can explicitly give:

$$\begin{aligned}\langle \chi_{REG}, \chi_\mu \rangle &= \sum \frac{1}{n} \operatorname{tr}(REG(g)) \operatorname{tr}(\mu(g)) = \\ &= \text{(using that } \operatorname{tr}(REG(g)) = n\delta_{g,\mathrm{id}}) \\ &= \frac{1}{n} \operatorname{tr}(REG(1)) \cdot \dim \mu \\ &= \dim \mu\end{aligned}$$

This allows us to conclude that

$$REG = \bigoplus_{\rho_i \text{ irreducible rep.}} \rho_i^{\dim \rho_i}$$

This shows that *all* irreducible representations of G are contained with corresponding multiplicity in REG. In particular,

$$|G| = \sum_{\text{all irreducible } \rho_i} |\dim \rho_i|^2$$

. Before we proceed to the proof of the orthogonality theorem, we need a small

Lemma 2.11.

$$\frac{1}{n} \sum_{g \in G} \operatorname{tr}(\rho(g)) \neq 0 \Leftrightarrow \exists v \textit{ such that } \rho(g)v = v, \forall g \in G$$

Proof. Define $R := \frac{1}{n}\sum_{g\in G}\rho(g)$. It follows then that $R^2 = R$, since $\forall g_0 \in G$,

$$\rho(g_0)\sum_{g\in G}\rho(g) = \sum_{g\in G}\rho(g)$$

It is clear that $\operatorname{tr} R \neq 0$ if and only if R has an eigenvalue different from 0 as well. Since R is a projection, we deduce it's only non-zero eigenvalue can be 1. So, let $Rv = v$.

$$\|v\| = \|Rv\| \leq \frac{1}{n}\sum \|\rho(g)v\| = \ v\|$$

Hence in the above inequalities we have equalities everywhere, and this means all $\rho(g)v$ and v are proportional, with constant of proportionality 1, thus $\rho(g)v = v, \forall g$. □

There are a few more remarks before we prove the orthogonality result. The first result is known as **Schur's lemma**. We defined a morphism of representations above. Schur's lemma states that a morphism between two *irreducible* representations is either trivial (sending everything to 0), or provides an isomorphism between the representations. The proof is simple - if the morphism is $L : V \to W$, then both $\ker L$ and $\operatorname{im} L$ are invariant (w.r.t. the corresponding representations) subspaces. Since the representations are irreducible, this means that each is either 0, or the whole space. $\ker L = V \Leftrightarrow \operatorname{im} L = 0$, and is equivalent to L being the 0-morphism. The other possible situation is $\ker L = 0, \operatorname{im} L = W$, yielding that L is indeed an isomorphism.

Proof of Theorem 2.10 . Consider the quantity

$$\frac{1}{n}\sum_{g\in G}\operatorname{tr}(\rho(g))\overline{\operatorname{tr}(\mu(g))} = \frac{1}{n}\sum_{g\in G}\operatorname{tr}(\rho(g))\operatorname{tr}(\mu(g^{-1})) = \frac{1}{n}\sum_{g\in G}\operatorname{tr}\left[\rho(g)\otimes\mu(g^{-1})\right]$$

The last equality comes from the fact that if we have two operators A, B on V, W respectively, with eigenvalues λ_i, μ_j with eigenvectors v_i, w_j, then the eigenvalues of $A \otimes B$ on $V \otimes W$ are $\lambda_i\mu_j$, with eigenvectors $v_i \otimes w_j$. We claim that if the above expression is different from 0, we shall exhibit a non-trivial morphism between ρ and μ. By Schur's lemma, it will be an isomorphism.

By the previous lemma, there exists an $0 \neq L \in V \otimes W$ such that

$$(\rho(g) \otimes \mu(g^{-1}))L = L, \forall g \in G$$

We are almost done, once we prove the following claim. Given two vectors spaces V, W, with fixed basis $u_i, i = 1, \ldots, n, v_j, j = 1, \ldots, m$, identify the space $V \otimes W$ with the space of $n \times m$ matrices with basis element $u_i \otimes v_j$ sent to the matrix having 1 in row i, column j, and 0 everywhere else. The claim is then that if we are given A, B matrices acting on V, respectively W, then $\forall L \in V \otimes W$, $(A \otimes B)L = ALB$. Notice that by linearity of both sides in L, it is enough to prove this for $L = u_i \otimes v_j$. But then,

$$(A \otimes B)u_i \otimes v_j = (Au_i) \otimes (Bv_j) = (\sum_k A_{k,i} u_k) \otimes (\sum_l B_{l,j} v_l)$$
$$= \sum_{k,l} A_{k,i} B_{l,j} u_k \otimes v_l$$

However, it is easy to see that if $\triangle_{i,j}$ has 1 in the cell (i, j) and 0 everywhere else, then $(A\triangle_{i,j}B)_{k,l} = A_{k,i}B_{l,j}$, as we claimed. Putting these together, we can apply the claim to the identity $(\rho(g) \otimes \mu(g^{-1}))L = L$ and conclude that, in matrix form, $\rho(g)L\mu(g)^{-1} = L$, or equivalently $\rho(g)L = L\mu(g)$. By Schur's lemma, this yields an isomorphism.

□

References

[She07] Alexander A. Sherstov. Separating AC0 from depth-2 majority circuits. In *STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 294–301, New York, NY, USA, 2007. ACM.

[Tao] Terrence Tao. http://terrytao.wordpress.com/2007/11/30/the-hahn-banach-theorem-mengers-theorem-and-hellys-theorem/.