CAS CS 591 B: Communication Complexity

Prof. Mark Bun Fall 2019

Lecture Notes 18:

Pattern Matrix Method

Reading.

- Sherstov, The Pattern Matrix Method

We'll continue our discussion of lifting theorems by studying Sherstov's Pattern Matrix Method, which is a technique for lifting lower bounds on the approximability of boolean functions by polynomials to communication lower bounds.

Definition 1. Let $\varepsilon > 0$ and let $f : \{-1,1\}^n \to \{-1,1\}$. A real polynomial $p : \{-1,1\}^n \to \mathbb{R}$ ε-approximates f if $|p(x) - f(x)| \leq \varepsilon$ for all $x \in \{-1,1\}^n$. The ε-approximate degree of f is the least degree of a polynomial p which approximates f, and is denoted $\mathrm{adeg}_\varepsilon(f)$.

Approximate degree itself is versatile for proving lower bounds in query complexity. For example, it lower bounds randomized query (decision tree) complexity.

Proposition 2. *Let $f : \{-1,1\}^n \to \{-1,1\}$. Then for every $\varepsilon > 0$, $\mathbf{BPP^{dt}}_\varepsilon(f) \geq \mathrm{adeg}_{2\varepsilon}(f)$.*

Proof. Let $\mathcal{T}$ be a randomized decision tree of depth d computing f to error ε. We will use $\mathcal{T}$ to construct a (2ε)-approximating polynomial for f of degree d.

First observe that if T is a deterministic decision tree of depth d, then $T(x)$ can be written exactly as a degree-d real polynomial p_T. Now define

$$p(x) = \mathbb{E}_{T \leftarrow \mathcal{T}}[p_T(x)]$$

which is also a degree-d-polynomial and satisfies $p(x) \in [-1,1]$ for all x. Moreover, $p(x)$ has the property that if $f(x) = 1$ then

$$p(x) = \Pr[p_T(x) = 1] - \Pr[p_T(x) = -1] \geq 1 - 2\varepsilon,$$

and similarly if $f(x) = -1$ then $p(x) \leq -1 + 2\varepsilon$. Hence p is a 2ε-approximating polynomial for f. □

Although we won't really talk about it, much of the power of approximate degree comes from the fact that it also lower bounds *quantum* query complexity.

Example 3. Let XOR_n be the parity function on n bits. Then for every $\varepsilon \in (0,1)$ we have $\mathrm{adeg}_\varepsilon(\mathrm{XOR}_n) = n$.

Proof. Every boolean function is exactly computed by a polynomial of degree n. One can see this by considering the polynomial obtained from a depth-n deterministic decision tree. To see that degree n is necessary, suppose $p : \{-1,1\}^n \to \mathbb{R}$ is any polynomial which agrees with XOR_n in sign.

(An approximating polynomial with $\varepsilon < 1$ certainly does this). Then consider the symmetrization $P : \{0, \ldots, n\} \to R$ of p defined by

$$P(k) = \mathbb{E}_{|x|=k} p(x).$$

One can show that $\deg P \leq \deg p$. Moreover, $P(k) < 0$ whenever k is even and $P(k) > 0$ whenever k is odd. So P has at least n sign changes, and therefore $\deg p \geq \deg P \geq n$. □

Example 4. $\widetilde{\deg}_{1/3}(\mathrm{OR}_n) = \Theta(\sqrt{n})$. For any $\varepsilon \in (0,1)$, we have $\widetilde{\deg}_{\varepsilon}(\mathrm{AND}_{n^{1/3}} \circ \mathrm{OR}_{n^{2/3}}) \geq \Omega(n^{1/3})$.

The Pattern Matrix Method applies to composed functions of the form $F = f \circ g_m^n$ where $g_m : \{-1,1\}^m \times ([m] \times \{-1,1\})$ is a variant of the indexing gadget defined by

$$g(x, (i, w)) = x_i \oplus w.$$

It is quite a versatile technique with several different specific formulations (lifting different variants of approximate degree to different measures of communication complexity). Here we will only state two.

Theorem 5. *Let $F = f \circ g_m^n$. Then For every $\delta > 0$,*

$$\mathrm{disc}(F) \leq \delta + m^{-\widetilde{\deg}_{1-\delta}(f)/2}.$$

To see how such a theorem can be used to prove communication lower bounds, suppose we were able to show that

$$\widetilde{\deg}_{1-2^{-c}}(f) \geq c$$

for some function f. Then taking $m = 4$ and $\delta = 2^{-c}$ we would have

$$\mathrm{disc}(F) \leq 2^{-c} + 4^{-c/2} = 2^{-c+1}.$$

Hence $\mathbf{PP}^{\mathbf{cc}}(F) \geq \Omega(\log(1/\mathrm{disc}(F))) \geq \Omega(c)$.

Here is another formulation. For a sign matrix S and $\varepsilon > 0$, let $\mathrm{arank}_{\varepsilon}(S)$ be the minimum rank of a real matrix R such that $|R[x,y] - S[x,y]| \leq \varepsilon$ for every x, y. By similar arguments to what we used to show that rank lower bounds $\mathbf{P}^{\mathbf{cc}}$ and sign rank lower bounds $\mathbf{UPP}^{\mathbf{cc}}$, approximate rank gives a lower bound on randomized communication complexity:

$$\mathbf{BPP}^{\mathbf{cc}}{}_{1/3}(F) \geq \Omega(\log \mathrm{arank}_{1/3}(S_F)) - O(\log n).$$

(The additive loss of $O(\log n)$ comes from using Newman's theorem to convert a protocol for F to a private-coin protocol.) Actually, approximate rank gives the same kind of lower bound for quantum communication complexity as well, and the Pattern Matrix Method is the only known tool for lifting a generic query complexity measure (such as approximate degree) to quantum communication complexity.

Theorem 6. *Let $0 < \delta < \varepsilon$ and let $F = f \circ g_m^n$. Then*

$$\mathrm{arank}_{\delta}(F) \geq \left(\frac{\varepsilon - \delta}{1 + \delta}\right)^2 \cdot m^{\widetilde{\deg}_{\varepsilon}(f)}$$

1 Dual Formulation of Approximate Degree

The proof strategy underlying Theorems 5 and 6 is totally different from what we used to prove the deterministic lifting theorem. Recall that in the deterministic lifting theorem, we took a protocol Π for the composed function F and used it to construct a decision tree for f. Here, we take the opposite approach. We will argue that if f is hard to approximate by low degree polynomials, there is a certain object ψ which serves as a witness to this fact. We will then use ψ to construct a distribution under which F has high communication complexity.

The following theorem produces the required ψ through an equivalent characterization of approximate degree.

Theorem 7. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a boolean function. Then $\widetilde{\deg}_\varepsilon(f) > d$ if and only if there exists a function $\psi : \{-1,1\}^n \to \mathbb{R}$ such that*

1. $\langle f, \psi \rangle := \sum_{x \in \{-1,1\}^n} f(x)\psi(x) > \varepsilon$
2. $\|\psi\|_1 := \sum_{x \in \{-1,1\}^n} |\psi(x)| = 1$
3. $\hat{\psi}(S) = 2^{-n} \sum_{x \in \{-1,1\}^n} \psi(x)\chi_S(x) = 0$ *for every* $S \subseteq [n]$ *with* $|S| \leq d$. *Here* $\chi_S(x) = \prod_{i \in S} x_i$.

One way to prove this theorem is using linear programming duality. The question "what is the least ε for which there exists a degree-d approximating polynomial for f" is captured by the LP

$$\begin{aligned} \min_{p,\varepsilon} \quad & \varepsilon \\ \text{s.t.} \quad & |p(x) - f(x)| \leq \varepsilon \quad \forall x \in \{-1,1\}^n \end{aligned}$$

This is a linear program in the variables c_S in the representation $p(x) = \sum_{|S| \leq d} c_S \chi_S(x)$ and the parameter ε. One can take the dual of this LP and obtain

$$\begin{aligned} \max_{\psi} \quad & \psi(x)f(x) \\ \text{s.t.} \quad & \sum_{x \in \{-1,1\}^n} |\psi(x)| = 1 \\ & \sum_{x \in \{-1,1\}^n} \psi(x)\chi_S(x) = 0 \quad \forall |S| \leq d. \end{aligned}$$

By strong LP duality, the dual has value greater than ε iff the primal has value greater than ε. So there exists a function ψ with the properties listed in Theorem 7 iff f cannot be approximated to error ε by degree-d polynomials.

To gain intuition for this characterization, it is helpful to work out by hand the "easy" direction of the equivalence, which is that the existence of such a function ψ implies that $\widetilde{\deg}_\varepsilon(f) > d$. To see this, suppose for the sake of contradiction that such a ψ existed as well as an approximating polynomial p for f of degree d. Then we would have

$$\langle f - p, \psi \rangle = \langle f, \psi \rangle - \langle p, \psi \rangle > \varepsilon$$

using properties 1 and 3 of ψ together with the fact that p is a linear combination of degree-d monomials. On the other hand,

$$|\langle f - p, \psi \rangle| \leq \sum_{x \in \{-1,1\}^n} |f(x) - p(x)| \cdot |\psi(x)| \leq \max_x |f(x) - p(x)| \cdot \|\psi\|_1 \leq \varepsilon$$

using the fact that p is an approximating polynomial together with property 2 of ψ. This is a contradiction.