Special Topics in Complexity Theory, Fall 2017. Instructor: Emanuele Viola

1 Lecture 6-7, Scribe: Willy Quach

In these lectures, we introduce k-wise indistinguishability and link this notion to the approximate degree of a function. Then, we study the approximate degree of some functions, namely, the AND function and the AND-OR function. For the latter function we begin to see a proof that is different (either in substance or language) from the proofs in the literature. We begin with some LaTeXtips.

1.1 Some LaTeX tips.

- Mind the punctuation. Treat equations as part of the phrases; add commas, colons, etc accordingly.
- In math mode, it is usually better to use \ell (ℓ) rather than regular l. The latter can be confused with 1.
- Align equations with \begin{align} $\cdots$ \end{align} with the alignment character &.
- For set inclusion, use \subset ($\subset$) only if you mean *proper* inclusion (which is uncommon). Otherwise use \subseteq ($\subseteq$). (Not everybody agrees with this, but this seems the most natural convention by analogy with $<$ and $\leq$.)

1.2 Introducing k-wise indistinguishability.

We studied previously the following questions:

- What is the minimum k such that any k-wise independent distribution P over $\{0,1\}^n$ fools AC^0 (*i.e.* $\mathbb{E}C(P) \approx \mathbb{E}C(U)$ for all $poly(n)$-size circuits C with constant depth)?

  We saw that $k = \log^{\mathcal{O}(d)}(s/\epsilon)$ is enough.

- What is the minimum k such that P fools the AND function?

  Taking $k = \mathcal{O}(1)$ for $\epsilon = \mathcal{O}(1)$ suffices (more precisely we saw that k-wise independence fools the AND function with $\epsilon = 2^{-\Omega(k)}$).

Consider now P and Q two distributions over $\{0,1\}^n$ that are *k-wise indistinguishable*, that is, any projection over k bits of P and Q have the same distribution. We can ask similar questions:

- What is the minimum k such that AC^0 cannot distinguish P and Q (*i.e.* $\mathbb{E}C(P) \approx \mathbb{E}C(Q)$ for all $poly(n)$-size circuits C with constant depth)?

  It turns out this requires $k \geq n^{1-o(1)}$: there are some distributions that are almost always distinguishable in this regime. (Whether $k = \Omega(n)$ is necessary or not is an open question.)

  Also, $k = n\left(1 - \frac{1}{polylog(n)}\right)$ suffices to fool AC^0 (in which case ϵ is essentially exponentially small).

- What is the minimum k such that the AND function (on n bits) cannot distinguish P and Q?

  It turns out that $k = \Theta(\sqrt{n})$ is necessary and sufficient. More precisely:

  - There exists some P, Q over $\{0,1\}^n$ that are $c\sqrt{n}$-wise indistinguishable for some constant c, but such that:

$$\left|\Pr_P[AND(P) = 1] - \Pr_Q[AND(Q) = 1]\right| \geq 0.99\,;$$

  - For all P, Q that are $c'\sqrt{n}$-wise indistinguishable for some bigger constant c', we have:

$$\left|\Pr_P[AND(P) = 1] - \Pr_Q[AND(Q) = 1]\right| \leq 0.01\,.$$

1.3 Duality.

Those question are actually equivalent to ones related about approximation by real-valued polynomials:

Theorem 1. Let $f : \{0,1\}^n \to \{0,1\}$ be a function, and k an integer. Then:

$$\max_{P,Q\ k\text{-wise indist.}} |\mathbb{E}f(P) - \mathbb{E}f(Q)| = \min\{\,\epsilon \mid \exists g \in \mathbb{R}_k[X] : \forall x, |f(x) - g(x)| \leq \epsilon\}.$$

Here $\mathbb{R}_k[X]$ denotes degree-k real polynomials. We will denote the right-hand side $\epsilon_k(f)$.

Some examples:

- $f = 1$: then $\mathbb{E}f(P) = 1$ for all distribution P, so that both sides of the equality are 0.
- $f(x) = \sum_i x_i \bmod 2$ the parity function on n bits.

  Then for $k = n - 1$, the left-hand side is at least $1/2$: take P to be uniform; and Q to be uniform on $n - 1$ bits, defining the nth bit to be $Q_n = \sum_{i<n} Q_i \bmod 2$ to be the parity of the first $n - 1$ bits. Then $\mathbb{E}f(P) = 1/2$ but $\mathbb{E}f(Q) = 0$.

  Furthermore, we have:

  Claim 2. $\epsilon_{n-1}(\text{Parity}) \geq 1/2$.

  Proof. Suppose by contradiction that some polynomial g has degree k and approximates Parity by $\epsilon < 1/2$.

  The key ingredient is to *symmetrize* a polynomial p, by letting

  $$p^{sym}(x) := \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} f(\pi x),$$

  where π ranges over permutations. Note that $p^{sym}(x)$ only depends on $\|x\| = \sum_i x_i$.

  Now we claim that there is a *univariate* polynomial p' also of degree k such that

  $$p'(\sum x_i) = p^{sym}(x_1, x_2, \ldots, x_n)$$

  for every x.

  To illustrate, let M be a monomial of p. For instance if $M = X_1$, then $p'(i) = i/n$, where i is the Hamming weight of the input. (For this we think of the input as being $\in \{0, 1\}$. Similar calculations can be done for $\in \{-1, -1\}$.)

  If $M = X_1 X_2$, then $p'(i) = \frac{i}{n} \cdot \frac{i-1}{n}$ which is quadratic in i.

  And so on.

  More generally $p^{sym}(X_1, \ldots, X_n)$ is a symmetric polynomial. As $\{(\sum_j X_j)^\ell\}_{\ell \leq k}$ form a basis of symmetric polynomials of degree k, p^{sym} can be written as a linear combination in this basis. Now note that $\{(\sum_j X_j)^\ell(x)\}_{\ell \leq k}$

only depends on $\|x\|$; substituting $i = \sum_j X_j$ gives that p' is of degree $\leq k$ in i.

(Note that the degree of p' can be strictly less than the degree of p (*e.g.* for $p(X_1, X_2) = X_1 - X_2$: we have $p^{sym} = p' = 0$).)

Then, applying symmetrization on g, if g is a real polynomial ϵ-close to Parity (in ℓ_∞ norm), then g' is also ϵ-close to Parity' (as a convex combination of close values).

Finally, remark that for every integer $k \in \{0, \ldots, \lfloor n/2 \rfloor\}$, we have: $\text{Parity}'(2k) = 0$ and $\text{Parity}'(2k+1) = 1$. In particular, as $\epsilon < 1/2$, $g' - 1/2$ must have at least n zeroes, and must therefore be zero, which is a contradiction.

□

We will now focus on proving the theorem.

Note that one direction is easy: if a function $f$is closely approximated by a polynomial g of degree k, it cannot distinguish two k-wise indistinguishable distributions P and Q:

$$\begin{aligned}\mathbb{E}[f(P)] &= \mathbb{E}[g(P)] \pm \epsilon \\ &\overset{(*)}{=} \mathbb{E}[g(Q)] \pm \epsilon \\ &= \mathbb{E}[f(Q)] \pm 2\epsilon\,,\end{aligned}$$

where $(*)$ comes from the fact that P and Q are k-wise indistinguishable.

The general proof goes by a Linear Programming Duality (aka finite-dimensional Hahn-Banach theorem, min-max, etc.). This states that:

If $A \in \mathbb{R}^{n \times m}$, $x \in \mathbb{R}^m$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}^m$, then:

$$\left|\begin{array}{rcl} \min\langle c, x\rangle & = & \sum_{i \leq m} c_i x_i \\ \text{subject to:} \quad Ax & = & b \\ x & \geq & 0 \end{array}\right| = \left|\begin{array}{rl} & \max\langle b, y\rangle \\ \text{subject to:} & A^T y \leq c \end{array}\right|$$

We can now prove the theorem:

Proof. The proof will consist in rewriting the sides of the equality in the theorem as outputs of a Linear Program. Let us focus on the left side of the equality: $\max_{P,Q\, k\text{-wise indist.}} |\mathbb{E}f(P) - \mathbb{E}f(Q)|$.

We will introduce 2^{n+1} variables, namely P_x and Q_x for every $x \in \{0,1\}^n$, which will represent $\Pr[D = x]$ for $D = P, Q$.

We will also use the following, which can be proved similarly to the Vazirani XOR Lemma:

Claim 3. Two distributions P and Q are k-wise indistinguishable if and only if: $\forall S \subseteq \{1, \ldots, n\}$ with $|S| \leq k$, $\sum_x P_x \chi_S(x) - \sum_x Q_x \chi_S(x) = 0$, where $\chi_S(X) = \prod_S X_i$ is the Fourier basis of boolean functions.

The quantity $\max_{P,Q\, k\text{-wise indist.}} |\mathbb{E}f(P) - \mathbb{E}f(Q)|$ can then be rewritten:

$$-\min \textstyle\sum_x P_x f(x) - \sum_x Q_x f(x)$$

$$\begin{aligned} \text{subject to:} \quad \textstyle\sum_x P_x &= 1 \\ \textstyle\sum_x Q_x &= 1 \\ \forall S \subseteq \{1,\ldots,n\} \text{ s.t. } |S| \leq k, \textstyle\sum_x (P_x - Q_x)\chi_S(x) &= 0 \end{aligned}$$

Following the syntax of LP Duality stated above, we have:

$c^T = \overbrace{\cdots f(x) \cdots}^{2^n} \overbrace{\cdots\cdots - f(x) \cdots}^{2^n} \in \mathbb{R}^{2n}$, (where x goes over $\{0,1\}^n$),

$x^T = \overbrace{\cdots P_x \cdots}^{2^n} \overbrace{\cdots\cdots Q_x \cdots}^{2^n} \in \mathbb{R}^{2n}$,

$b^T = 11\overbrace{0\cdots0}^{\#S}$,

$$A = \begin{pmatrix} \overbrace{1\cdots\cdots1}^{2^n} & \overbrace{0\cdots\cdots0}^{2^n} \\ 0\cdots\cdots0 & 1\cdots\cdots1 \\ \cdots\cdots & \cdots\cdots \\ \vdots\cdots\cdots\vdots & \vdots\cdots\cdots\vdots \\ \cdots\chi_S(x)\cdots & \cdots -\chi_S(x)\cdots \\ \vdots\cdots\cdots\vdots & \vdots\cdots\cdots\vdots \\ \cdots\cdots & \cdots\cdots \end{pmatrix},$$

where the rows of A except the first two correspond to some $S \subseteq \{1, \ldots, n\}$ such that $|S| \leq k$.

We apply LP duality. We shall denote the new set of variables by

$y^T = d\, d' \overbrace{\cdots d_S \cdots}^{\#S}$.

We have the following program:

$$-\max d + d'$$

$$\begin{array}{rll} \text{subject to:} & \forall x, d + \sum_x d_S \chi_S(x) & \leq f(x) \\ & \forall x, d' - \sum_x d_S \chi_S(x) & \leq -f(x) \end{array}$$

Writing $d' = -d - \epsilon$, the objective becomes to minimize ϵ, while the second set of constraints can be rewritten:

$$\forall x, d + \epsilon + \sum_S d_S \chi_S(x) \geq f(x)\,.$$

The expression $d + \sum_S d_S \chi_S(X)$ is an arbitrary degree-k polynomial which we denote by $g(X)$. So our constrains become

$$\begin{aligned} g(x) &\leq f(x) \\ g(x) + \epsilon &\geq f(x). \end{aligned}$$

Where g ranges over all degree-k polynomials, and we are trying to minimize ϵ. Because g is always below f, but when you add ϵ it becomes bigger, g is always within ϵ of f. □

1.4 Approximate Degree of AND.

Let us now study the AND function on n bits. Let us denote $d_\epsilon(f)$ the minimal degree of a polynomial approximating f with error ϵ.

We will show that $d_{1/3}(\text{AND}) = \Theta(\sqrt{n})$.

Let us first show the upper bound:

Claim 4. We have:

$$d_{1/3}(\text{AND}) = \mathcal{O}(\sqrt{n}).$$

To prove this claim, we will consider a special family of polynomials:

Definition 5. (Chebychev polynomials of the first kind.)

The Chebychev polynomials (of the first kind) are a family $\{T_k\}_{k \in \mathbb{N}}$ of polynomials defined inductively as:

- $T_0(X) := 1$,
- $T_1(X) := X$,
- $\forall k \geq 1, T_{k+1}(X) := 2XT_k - T_{k-1}$.

Those polynomials satisfy some useful properties:

1. $\forall x \in [-1,1], T_k(x) = \cos(k \arccos(x))$,
2. $\forall x \in [-1,1], \forall k, |T_k(x)| \leq 1$,
3. $\forall x$ such that $|x| \geq 1, |T_k'(x)| \geq k^2$,
4. $\forall k, T_k(1) = 1$.

Property 2 follows from 1, and property 4 follows from a direct induction. For a nice picture of these polynomials you should have come to class (or I guess you can check wikipedia). We can now prove our upper bound:

Proof. Proof of Claim:

We construct a univariate polynomial $p : \{0, 1, \ldots, n\} \to \mathbb{R}$ such that:

- $\deg p = \mathcal{O}(\sqrt{n})$;
- $\forall i < n, |P(i)| \leq 1/3$;
- $|P(n) - 1| \leq 1/3$.

In other words, p will be close to 0 on $[0, n-1]$, and close to 1 on n. Then, we can naturally define the polynomial for the AND function on n bits to be $q(X_1, \ldots, X_n) := p(\sum_i X_i)$, which also has degree $\mathcal{O}(\sqrt{n})$. Indeed, we want q to be close to 0 if X has Hamming weight less than n, while being close to 1 on X of Hamming weight n (by definition of AND). This will conclude the proof.

Let us define p as follows:

$$\forall i \leq n, \quad p(i) := \frac{T_k\left(\frac{i}{n-1}\right)}{T_k\left(\frac{n}{n-1}\right)}.$$

Intuitively, this uses the fact that Chebychev polynomials are bounded in $[-1, 1]$ (Property 2.) and then increase very fast (Property 3.).

More precisely, we have:

- $p(n) = 1$ by construction;

- for $i < n$, we have:

  $T_k\left(\frac{i}{n-1}\right) \leq 1$ by Property 2.;

  $T_k\left(\frac{n}{n-1}\right) = T_k\left(1 + \frac{1}{n-1}\right) \geq 1 + \frac{k^2}{n-1}$ by Property 3. and 4., and therefore for some $k = \mathcal{O}(\sqrt{n})$, we have: $T_k\left(\frac{n}{n-1}\right) \geq 3$.

□

Let us now prove the corresponding lower bound:

Claim 6. We have:

$$d_{1/3}(\text{AND}) = \Omega(\sqrt{n}).$$

Proof. Let p be a polynomial that approximates the AND function with error $1/3$. Consider the univariate symmetrization p' of p.

We have the following result from approximation theory:

Theorem 7. Let q be a real univariate polynomial such that:

1. $\forall i \in \{0, \ldots, n\}, |q(i)| \leq \mathcal{O}(1)$;
2. $q'(x) \geq \Omega(1)$ for some $x \in [0, n]$.

   Then $\deg q = \Omega(\sqrt{n})$.

To prove our claim, it is therefore sufficient to check that p' satisfies conditions 1. and 2., as we saw that $\deg p \geq \deg p'$:

1. We have: $\forall i \in \{0, \ldots, n\}, |p'(i)| \leq 1 + 1/3$ by assumption on p;
2. We have $p'(n-1) \leq 1/3$ and $p'(n) \geq 2/3$ (by assumption), so that the mean value theorem gives some x such that $p'(x) \geq \Omega(1)$.

This concludes the proof. □

1.5 Approximate Degree of AND-OR.

Consider the AND function on R bits and the OR function on N bits. Let AND-OR: $\{0,1\}^{R\times N} \to \{0,1\}$ be their composition (which outputs the AND of the R outputs of the OR function on N-bits (disjoint) blocks).

It is known that $d_{1/3}(\text{AND-OR}) = \Theta(\sqrt{RN})$. To prove the upper bound, we will need a technique to compose approximating polynomials which we will discuss later.

Now we focus on the lower bound. This lower bound was recently proved independently by Sherstov and by Bun and Thaler. We present a proof that is different (either in substance or in language) and which we find more intuitive. Our proof replaces the "dual block method" with the following lemma.

Lemma 8. Suppose that

distributions A^0, A^1 over $\{0,1\}^{n_A}$ are k_A-wise indistinguishable distributions; and

distributions B^0, B^1 over $\{0,1\}^{n_B}$ are k_B-wise indistinguishable distributions.

Define C^0, C^1 over $\{0,1\}^{n_A \cdot n_B}$ as follows: C^b: draw a sample $x \in \{0,1\}^{n_A}$ from A^b, and replace each bit x_i by a sample of B^{x_i} (independently).

Then C^0 and C^1 are $k_A \cdot k_B$-wise indistinguishable.

Proof. Consider any set $S \subseteq \{1, \ldots, n_A \cdot n_B\}$ of $k_A \cdot k_B$ bit positions; let us show that they have the same distribution in C^0 and C^1.

View the $n_A \cdot n_B$ as n_A *blocks* of n_B bits. Call a block K of n_B bits *heavy* if $|S \cap K| \geq k_B$; call the other blocks *light*.

There are at most k_A heavy blocks by assumption, so that the distribution of the (entire) heavy blocks are the same in C^0 and C^1 by k_A-wise indistinguishability of A^0 and A^1.

Furthermore, conditioned on any outcome for the A^b samples in C^b, the light blocks have the same distribution in both C^0 and C^1 by k_B-wise indistinguishability of B^0 and B^1.

Therefore C^0 and C^1 are $k_A \cdot k_B$-wise indistinguishable. □