CS 599B: Math for TCS, Spring 2022

Homework 1

Due: 10:00PM, Friday February 25, 2022 on Gradescope.

Instructions. Solutions must be typeset, with LaTeX strongly preferred. You are encouraged to collaborate on the homework problems with each other in small groups (2-4 people). Collaboration may include brainstorming or exploring possible solutions together on a whiteboard, but should not include one person telling the others how to solve a problem. You must also write up the solutions independently (in your own words) and acknowledge your collaborators at the beginning of the first page.

You may freely use without proof any results proved in class, in Mark's lecture notes posted on the class webpage, or in the main body of the texts assigned as reading. Note that this excludes results that appear in the texts as problems and exercises. You may, of course, use such results but you have to prove them first.

Problem 1 (Uncertainty principle, O'Donnell 3.15)**.** For a function $f : \{-1,1\}^n \to \mathbb{R}$, define its *Boolean sparsity* to be $\mathrm{sp}(f) = |\operatorname{supp}(f)| = |\{x \mid f(x) \neq 0\}|$. Similarly, define its *Fourier sparsity* by $\mathrm{sp}(\hat{f}) = |\operatorname{supp}(\hat{f})| = |\{S \mid \hat{f}(S) \neq 0\}|$.

a) Show that for every $S \subseteq [n]$, we have $|\hat{f}(S)| \leq \|f\|_1 := \mathbb{E}_x\left[|f(x)|\right]$.

b) Show that

$$\|\hat{f}\|_2^2 := \sum_{S \subseteq [n]} \hat{f}(S)^2 \leq \mathrm{sp}(\hat{f}) \cdot \|f\|_1^2.$$

c) Show that

$$\|f\|_2^2 := \mathop{\mathbb{E}}_{x \sim \{-1,1\}^n}\left[f(x)^2\right] \geq 2^n \cdot \|f\|_1^2 / \mathrm{sp}(f).$$

d) Deduce the "uncertainty principle": For every nonzero function $f : \{-1,1\}^n \to \mathbb{R}$, we have $\mathrm{sp}(f) \cdot \mathrm{sp}(\hat{f}) \geq 2^n$.

e) When is the uncertainty principle tight? Justify your answer.

Problem 2 (Hardness amplification)**.** A central topic in average-case complexity is *hardness amplification*: Can a function f that is mildly hard on average, e.g., there is no small circuit that computes f correctly for 99.9% of inputs, be generically transformed into a function F that small circuits cannot predict much better than random guessing, e.g., 50.1%? An example of such a result is Yao's XOR lemma, which says that if circuits of size s can compute f with probability at most $1 - \delta$, then circuits of size $\lesssim s(1-2\delta)^k$ can compute $\underbrace{f \oplus \cdots \oplus f}_{k \text{ times}}$ with probability at most $1/2 + (1-2\delta)^k$.

Noise sensitivity turns out to be an important tool for our understanding of hardness amplification for relatively weak classes like **NP**. That is, we would like to understand whether mildly hard functions in **NP** can be transformed into extremely hard functions that are still in **NP**. The XOR lemma doesn't help here, since the XOR of two **NP** languages is not necessarily also in **NP**. In this problem, you will fill in some of the details of Healy, Vadhan, and Viola's exposition of Trevisan's proof of O'Donnell's "Hardness Amplification within **NP**" theorem.

Let $f : \{-1,1\}^n \to \{-1,1\}$ be a δ-hard function, in that for every small circuit T, we have $\Pr_x[T(x) = f(x)] \leq 1 - \delta$. We amplify the hardness of f using functions of the form $(C \circ f^k) : (\{-1,1\}^n)^k \to \{-1,1\}$ defined by $(C \circ f^k)(x^1, \ldots, x^k) = C(f(x^1), \ldots, f(x^k))$, where $C : \{-1,1\}^k \to \{-1,1\}$ is a combining function. The first step is to reduce the study of the computational hardness of $C \circ f^k$ to the information-theoretic hardness of a related object $C \circ g^k$ where g is a *probabilistic function*.

A probabilistic function is a randomized algorithm $h : \{-1,1\}^n \to \{-1,1\}$. It may help to think of h as a function of two inputs: a normal input $x \in \{-1,1\}^n$ and a random input $r \in R$ for some sample space R. So on a fixed input x, the expression $\mathbb{E}[h(x)]$ is shorthand for $\mathbb{E}_{r\sim R}\left[h(x;r)\right]$.

For a probabilistic function $h : \{-1,1\}^n \to \{-1,1\}$, define

$$\mathrm{ExpBias}(h) = \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\left|\mathbb{E}\left[h(x)\right]\right|\right]$$

a) Show that the expected bias of a probabilistic function h characterizes the ability of any (even computationally unbounded) device to compute h:

$$\max_{T:\{-1,1\}^n\to\{-1,1\}} \Pr[T(x) = h(x)] = \frac{1}{2} + \frac{1}{2}\mathrm{ExpBias}(h).$$

If g is a balanced function, i.e., $\mathbb{E}_x\left[\mathbb{E}[g(x)]\right] = 0$ and there exists a set H with $|H| = 2\delta \cdot 2^n$ such that a) $g(x)$ is a uniform bit for $x \in H$ and b) g is deterministic for every $x \notin H$, we say that g is *δ-random*. That is, g is (information-theoretically) hard to predict because there is a 2δ fraction of inputs on which it is completely unpredictable.

Your intuition may suggest that if f is mildly hard, the reason may be different for different circuits. That is, the place where circuit T_1 errs in computing f may be totally different from the place where circuit T_2 errs. Remarkably, Impagliazzo's hardcore lemma says that mildly hard functions "look like" δ-random functions to small circuits – the hardness of every mildly hard function is explained by its being completely unpredictable on a small set of inputs to *every* small circuit. Combining the hardcore lemma with part (a) and a few other ideas, one can show:

Lemma 1 (Informal)**.** *If f is δ-mildly hard for small circuits, then there exists a δ-random function g such that*

$$\Pr[T(x) = (C \circ f^k)(x)] \leq \frac{1}{2} + \frac{1}{2}\mathrm{ExpBias}(C \circ g^k)$$

for every small circuit T.[1]

Thus, in order to show that $C \circ f^k$ is extremely hard, it suffices to show that $C \circ g^k$ has small expected bias for every δ-random g.

b) Define the distribution $E_{2\delta}$ over $\{-1,1\}^k$ as follows. To sample $e \sim E_{2\delta}$, set each e_i independently to -1 with probability 2δ and to $+1$ with probability $1-2\delta$. Show that for every C and every δ-random function g,

$$\mathrm{ExpBias}(C \circ g^k) = \mathop{\mathbb{E}}_{\substack{y \sim \{-1,1\}^k \\ e \sim E_{2\delta}}} \left[\left| \mathop{\mathbb{E}}_{z \sim \{-1,1\}^k} [C(y \oplus (e \wedge z))] \right| \right].$$

The expression looks a bit gnarly, but the following should help to interpret it. To sample the string $y \oplus (e \wedge z)$, first sample $y \sim \{-1,1\}^k$ uniformly. Then choose roughly a 2δ-fraction of the coordinates specified by e, and re-randomize these coordinates according to the string z. Note in particular that the pair $(y, y \oplus (e \wedge z))$ is $(1-2\delta)$-correlated.

c) Show that

$$\mathbf{Stab}_{1-2\delta}(C) = \mathop{\mathbb{E}}_{\substack{y \sim \{-1,1\}^k \\ e \sim E_{2\delta}}} \left[\mathop{\mathbb{E}}_{z,w \sim \{-1,1\}^k} [C(y \oplus (e \wedge z)) \cdot C(y \oplus (e \wedge w))] \right].$$

d) Combine parts (b) and (c) to conclude that

$$\mathrm{ExpBias}(C \circ g^k) \le \sqrt{\mathbf{Stab}_{1-2\delta}(C)}.$$

e) The last piece of the puzzle is to exhibit combining functions C that have very low noise stability. The XOR function (parity) is highly noise sensitive, so it has very low noise stability. Use part (d) and Lemma 1 to (qualitatively) conclude Yao's XOR lemma.

As mentioned before, the XOR lemma is not good enough to obtain hardness amplification within **NP**. To address this problem, we want C to be polynomial-time computable and *monotone*. By composing the Tribes function with a Recursive-Majority-of-3, O'Donnell showed:

Lemma 2. *For every $\delta > 0$ there exists a $k = \mathrm{poly}(1/\delta)$ and a poly-time computable monotone function $C : \{-1,1\}^k \to \{-1,1\}$ such that $\mathbf{Stab}_{1-2\delta}(C) \le 1/\mathrm{poly}(k)$.*

f) Combine part (d), Lemma 1, and Lemma 2 to argue that if a family of functions $f_n : \{-1,1\}^n \to \{-1,1\}$ is in **NP** and $(1/\mathrm{poly}(n))$-hard for small circuits, then there exists a family of functions $F_m : \{-1,1\}^m \to \{-1,1\}$ in **NP** that cannot be computed with probability better than $1/2 + 1/\mathrm{poly}(m)$ by small circuits.

[1] Note that even if we were careful about what "small" means here, statement is not literally true because of various losses in parameters like δ, the circuit size, and the final upper bound on success probability. But let's pretend it's true to simplify the story.

Problem 3 (Bonami variant, O'Donnell 9.34)**.** Prove the following variant of Bonami's Lemma: For every $f : \{-1,1\}^n \to \mathbb{R}$,

$$\mathbb{E}[f(x)^4] \leq \mathrm{sp}(\widehat{f}) \cdot \mathbb{E}[f(x)^2]^2.$$

Thus, not only are low-degree polynomials reasonable, but sparse polynomials are reasonable as well.

Problem 4 (Generalizing KKL)**.**

a) Let $f : \{-1,1\}^n \to \{-1,1\}$ be computed by a decision tree of depth d. Show the following improved version of KKL:

$$\mathbf{MaxInf}[f] \geq \frac{\mathrm{Var}\,[f]}{d}.$$

b) Show that the KKL Theorem is false for real-valued functions $f : \{-1,1\}^n \to [-1,1]$. (O'Donnell 9.20 – look there if you need a hint.)

A great open problem related to these questions the Aaronson-Ambainis conjecture: For any $f : \{-1,1\}^n \to [-1,1]$, we have $\mathbf{MaxInf}[f] \geq \mathrm{poly}(\mathrm{Var}\,[f]\,/\deg(f))$.