CS 225 - Pseudorandomness Prof. Salil Vadhan

Take-Home Final

Harvard SEAS - Fall 2016 *Due: Mon. Dec 12, 2016 (5pm sharp)*

Your take-home final solutions must be typed (in e.g. LaTeX) and submitted electronically to `cs225-hw@seas.harvard.edu`. There are no late days for this problem set. Please name your file `final-`lastname`.*`.

Collaboration or use of references other than the course materials (the text, problem sets, solutions, your notes) is not allowed. You may only discuss these problems with course staff. You may use any result proved in the text, on problem sets, or in sections, as long as you state it clearly. *ed problems are extra credit.

Problem 2.11 (Consequences of Derandomizing prBPP)

Even though **prBPP** is a class of decision problems, it also captures many other types of problems that can be solved by randomized algorithms:

2. (**NP** Search Problems) An **NP** search problem is specified by a polynomial-time verifier V and a polynomial p; the problem is, given an input $x \in \{0,1\}^n$, find a string $y \in \{0,1\}^{p(n)}$ such that $V(x,y) = 1$. Suppose that such a search problem can be solved in probabilistic polynomial time, i.e. there is a probabilistic polynomial-time algorithm A such that for every input $x \in \{0,1\}^n$, outputs $y \in \{0,1\}^{p(n)}$ such that $V(x,y) = 1$ with probability at least $2/3$ over the coin tosses of A. Show that if **prBPP** = **prP**, then there is a deterministic polynomial-time algorithm B such that for every $x \in \{0,1\}^n$, $B(x)$ always outputs $y \in \{0,1\}^{p(n)}$ such that $V(x,y) = 1$. (Hint: consider a promise problem whose instances include pairs (x,r) where r is a prefix of the coin tosses of A, and use it to approximate the Method of Conditional Probabilities.)

4. Use Part 2, the Prime Number Theorem (see Problem 2.4), and the fact that PRIMALITY is in **BPP** (Problem 2.6) to show that if **prBPP** = **prP**, then there is a deterministic polynomial-time algorithm that given a number N, outputs a prime in the interval $[N, 2N)$ for all sufficiently large N.

Problem 6.4 (ℓ_2-extractors)

Call a function $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ a (k,ε) *ℓ_2-extractor* if for every k-source X, $\|\mathrm{Ext}(X,U_d) - U_m\|_2^2 \le \varepsilon/M$

1. Prove that a (k,ε) ℓ_2-extractor is also a $(k,\sqrt{\varepsilon})$ extractor.

2. Show that for every $n,k,m \in \mathbb{N}$ with $m \le n$, $k \ge m/2$, and $\varepsilon > 0$, there exists a (k,ε) ℓ_2-extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $d = \min\{O(n-k+\log(1/\varepsilon)), m/2 + O(\log(n/\varepsilon))\}$. (Hint: Problem 3.4 may be useful.)

3. Show that if $\text{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a $(k,1)$ ℓ_2-extractor, then $d \geq \min\{n-k, m/2\} - O(1)$. (Hint: consider a k-source that is uniform over $\{x : \exists y \ \text{Ext}(x,y) \in T\}$ for an appropriately chosen set T of size $\lfloor M/2D^2 \rfloor$.)

Problem 7.8 (Hitting-Set Generators)

A set $H_m \subset \{0,1\}^m$ is a (t,ε) *hitting set* if for every nonuniform algorithm T running in time t that accepts greater than an ε fraction of m-bit strings, T accepts at least one element of H_m.

1. Show that if, for every m, we can construct an $(m, 1/2)$ hitting set H_m in time $s(m) \geq m$, then $\mathbf{RP} \subset \bigcup_c \mathbf{DTIME}(s(n^c))$. In particular, if $s(m) = \text{poly}(m)$, then $\mathbf{RP} = \mathbf{P}$.

2. Show that if there is a (t,ε) pseudorandom generator $G_m : \{0,1\}^d \to \{0,1\}^m$ computable in time s, then there is a (t,ε) hitting set H_m constructible in time $2^d \cdot s$.

3. Show that if, for every m, we can construct an $(m, 1/2)$ hitting set H_m in time $s(m) = \text{poly}(m)$, then $\mathbf{BPP} = \mathbf{P}$. (Hint: This can be proven in at least two ways: one uses Problem 3.1 and the other uses a variant of Problem 7.1 together with Corollary 7.64. For extra credit, discuss how the efficiency of the derandomization for general $s(m)$ (not necessarily $\text{poly}(m)$) compares between the two approaches.)

4. Define the notion of a (t,k,ε) black-box construction of hitting set-generators (similar to Definition 7.65), and show that, when $t = \infty$, such constructions are equivalent to constructions of *dispersers* (Definition 6.19).

Problem 7.12 (Local List Decoding the Hadamard Code)

For a function $f : \mathbb{Z}_2^m \to \mathbb{Z}_2$, A *parameterized subspace* $x + V$ of $\mathbb{Z}_2^m$ of dimension d is given by a linear map $V : \mathbb{Z}_2^d \to \mathbb{Z}_2^m$ and a shift $x \in \mathbb{Z}_2^m$. (We do not require that the map V be full rank.) We write V for $0 + V$. For a function $f : \mathbb{Z}_2^m \to \mathbb{Z}_2$, we define $f|_{x+V} : \mathbb{Z}_2^d \to \mathbb{Z}_2^m$ by $f|_{x+V}(y) = f(x + V(y))$.

1. Let $c : \mathbb{Z}_2^m \to \mathbb{Z}_2$ be a codeword in the Hadamard code (i.e. a linear function), $r : \mathbb{Z}_2^m \to \mathbb{Z}_2$ a received word, V a parameterized subspace of $\mathbb{Z}_2^m$ of dimension d, and $x \in \mathbb{Z}_2^m$. Show that if $d_H(r|_{x+V}, c|_{x+V}) < 1/2$, then $c(x)$ can be computed from x, V, $c|_V$, and oracle access to r in time $\text{poly}(m, 2^d)$ with 2^d queries to r.

2. Show that for every $m \in \mathbb{N}$ and $\varepsilon > 0$, the Hadamard code of dimension m has a $(1/2-\varepsilon)$ local list-decoding algorithm $(\text{Dec}_1, \text{Dec}_2)$ in which both Dec_1 and Dec_2 run in time $\text{poly}(m, 1/\varepsilon)$, and the list output by Dec_1 has size $O(1/\varepsilon^2)$. (Hint: consider a random parameterized subspace V of dimension $2\log(1/\varepsilon) + O(1)$, and how many choices there are for $c|_V$.)

Problem 7.10(*) (PRGs are Necessary for Derandomization)

1. Call a function $G : \{0,1\}^d \to \{0,1\}^m$ a (t,ℓ,ε) *pseudorandom generator against bounded-nonuniformity algorithms* iff for every probabilistic algorithm T that has a program of length

at most ℓ and that runs in time at most t on inputs of length n, we have

$$|\Pr[T(G(U_d)) = 1] - \Pr[T(U_m) = 1]| \leq \varepsilon.$$

Consider the promise problem Π whose YES instances are truth tables of functions $G : \{0,1\}^d \to \{0,1\}^m$ that are $(m, \log m, 1/m)$ pseudorandom generators against bounded-nonuniformity algorithms, and whose NO instances are truth tables of functions that are not $(m, \log m, 2/m)$ pseudorandom generators against bounded-nonuniformity algorithms. (Here m and d are parameters determined by the input instance G.) Show that Π is in **prBPP**.

2. Using Problem 2.11, show that if **prBPP** = **prP**, then there is a mildly explicit $(m, 1/m)$ pseudorandom generator against uniform algorithms with seed length $O(\log m)$. (See Problem 7.9 for the definition. It turns out that the hypothesis **prBPP** = **prP** here can be weakened to obtain an equivalence between PRGs vs. uniform algorithms and average-case derandomization of **BPP**.)