COS 522 Complexity — Homework 4.

Boaz Barak

Total of 120 points. Due April 10th, 2006.

Exercise 1 (20 points)**.** Suppose that there exists a polynomial-time algorithm G and a constant $c > 0$ such that for any s, and any circuit C of size $\leq s$, if x is chosen at random from $\{0,1\}^{c \log s}$ then

$$|\Pr[C(G(1^s, x)) = 1] - \Pr[C(U_s) = 1]| < \frac{1}{10}$$

(where if C is takes $n \leq s$ bits as input, then by $C(y)$ we mean apply C to the first n bits of y.)

Prove that there exists a function $f \in \mathbf{E} = \mathbf{DTIME}(2^{O(n)})$ (with one bit of output) such that f is not computable by $2^{n/\log n}$-size circuits.

Exercise 2 (20 points)**.** For X a random variable over $\{0,1\}^n$, we define $H_\infty(X)$ (called the *min-entropy* of X) to be the smallest number k such that $\Pr[X = x] \leq 2^{-k}$ for every $x \in \{0,1\}^n$. We define $H_2(X)$ (called the *two entropy* of X) to be $\log(1/cp(X))$ where $cp(X)$ is the *collision probability* of X. That is, $cp(X) = \Pr[X = X'] = \sum_{x \in \{0,1\}^n} (\Pr[X = x])^2$ where X, X' are two independent copies of X. Note that we can think of X as a vector of 2^n non-negative numbers summing to one, in which case $cp(X)$ is equal to $\|x\|_2^2$. We say that X is a *convex combination* of distributions $X_1, \ldots, X_N$ if there are non-negative numbers $\alpha_1, \ldots, \alpha_N$ such that $\sum_{i=1}^N \alpha_i = 1$ and $X = \sum_i \alpha_i X_i$ (where this summation is in vector notation, alternatively one can think of choosing a random element from X as first choosing i with probability α_i and then choosing a random element from X_i).

1. Prove that $H_\infty(X) \leq H_2(X)$.

2. Prove that $H_2(X) = n$ iff X is distributed according to the uniform distribution on $\{0,1\}^n$.

3. Prove that $H_2(X) = n$ iff for every non zero vector $r \in \{0,1\}^n$, $\Pr[< X, r >= 0 \pmod 2] = \frac{1}{2}$. See footnote for hint[1]

4. Let k be a whole number in $[n]$. Prove that every X with $H_\infty(X) \geq k$ is a convex combination of distributions $X_1, \ldots, X_N$ where each X_i is the uniform distribution over some set $S_i \subseteq \{0,1\}^n$ with $|S_i| \geq 2^k$. (For partial credit, prove that X is of statistical distance less than 1/100 to a distribution that is such a convex combination.)

Exercise 3 (20 points + 5 points bonus)**.** For a subset $C \subseteq \{0,1\}^n$, we say that C is a *good code* if $|C| \geq 2^{n/100}$ and $\mathsf{mindist}(C) \geq n/100$ where

$$\mathsf{mindist}(C) = \min_{x \neq x' \in C} \left|\{i \in [n] : x_i \neq x'_i\}\right|$$

[1] **Hint:** (this is not the only way to do this) use the fact that the norm two of a vector is the same if the vector is expressed under a different orthonormal basis, and consider the vector X represented in the basis $\{Z^r\}_{r \in \{0,1\}^n}$ where the x^{th} coordinate of Z_r is $+2^{n/2}$ if $< x, r >= 0 \pmod 2$ and $-2^{n/2}$ otherwise.

1. Prove that if C is a linear subspace then $\mathsf{mindist}(C)$ to $\min_{0^n \neq x \in C} |\{i \in [n] : x_i = 1\}|$.

2. Prove using the probabilistic method that there exists a good code C that is a linear subspace (that is, it satisfies that if $x, x' \in C$ then $x \oplus x' \in C$).

3. Prove that there exists no good code C with $\mathsf{mindist}(C) \geq 0.51n$. See footnote for hint[2]. For 5 bonus points, prove that there exists no good code C with $\mathsf{mindist}(C) \geq \frac{n}{2} - \sqrt{n}$.

Exercise 4 (20 points + 15 points bonus)**.** We can define a subspace $C \subseteq \{0,1\}^n$ of dimension $\geq d$ by specifying a set of $k = n - d$ linear equations that this set satisfies. That is, each equation stipulates that the sum (mod 2) of some variables is equal to 0. We can also denote this in a bipartite graph $G = (X, Y, E)$ where $|X| = n, |Y| = k$ and for every $j \in [k]$ the neighbors of $y_j \in Y$ correspond to the variables appearing in the j^{th} equation. We'll restrict ourselves into graphs where each $x_i \in X$ is connected to at most 10 elements of Y (i.e., x_i appears in at most 10 equations).

1. Choose G with $|X| = n$ and $|Y| = k = 0.9n$ at random by choosing 10 random neighbors in Y for each $x \in X$. Prove that with probability > 0.9 it holds that for every set $S \subseteq X$ with $|S| \leq n/30$, it holds that $\Gamma(X) \geq 9|S|$. We call this condition (*)

2. Prove that if such a graph G satisfies the condition (*) then the corresponding code is good. See footnote for hint[3]

3. (15 points bonus) Find an efficient algorithm to decode this code. That is, show a polynomial-time algorithm A that given G satisfying (*) with corresponding code C and given y such that there exists some $x \in C$ with Hamming distance of x and y less than $n/1000$, manages to find this vector x. (Although this can be solved without this, you can use also a probabilistic algorithm if you like.) See footnote for hint[4]

Exercise 5 (20 points)**.** 1. Let $\mathsf{3SAT}_{10}$ be the variant of $\mathsf{3SAT}$ where the formula is restricted to have the condition that each variable does not appear in more than 10 clauses. Prove that $\mathsf{3SAT}_{10}$ is NP-complete.

2. Suppose that there's a polynomial-time algorithm A that on input a $\mathsf{3SAT}_{10}$ formula ϕ, outputs 1 if ϕ is satisfiable and outputs 0 if for any assignment x for ϕ, at least a 1/1000 fraction of the clauses are not satisfied by x. (There's no guarantee what A does on formulas that are not fully but are 999/1000 satisfiable). Prove if this is the case then is a polynomial-time algorithm B that on input a standard $3SAT$ formula ψ (possibly with each variable appearing in many clauses) outputs 1 if ψ is satisfiable and outputs 0 for any assignment y for ψ, at least 0.9 fraction of the clauses are not satisfied by y. (you can use a probabilistic algorithm B if you like, although it can be done without this.) See footnote for hint[5]

[2]**Hint:** Think of the codewords as vectors in $\mathbb{R}^n$ with $+1$ representing zero and -1 representing one. Then, use the fact that the distance is related to the inner product of such vectors.

[3]**Hint:** note that if G satisfies this condition then for any such S there exist many $y \in \Gamma(S)$ that are connected to exactly one element of S.

[4]**Hint:** For starters try to find an algorithm that transforms y that is of distance $d \leq n/1000$ to the code into y' that is of distance $0.9d$ to the code.

[5]**Hint:** use expander graphs.