A Theorist's Toolkit **CMU 15-751*, Fall 2016**

TAKE-HOME FINAL

Due: 11:59pm, Sunday December 11, email the pdf to toolkit2016homework@gmail.com

---

Take-home final policy: You must complete the take-home final by yourself. You may not discuss any of the problems with any other person. You may only consult your own class notes. You may ask for problem clarifications on Piazza (privately) as needed. As usual, LaTeX typesetting with pdf output is mandatory.

Solve 5 of the 9 problems, including at least 2 problems from the last 4 problems.

---

1. Show that
$$\sum_{t=1}^{n}(1+1/t)^t = en \pm O(\log n).$$

2. A *random 3SAT instance* $\mathcal{I}$ with n variables and $m = m(n)$ constraints is defined as follows: There are n Boolean variables $x_1, \dots, x_n$, and m constraints are randomly chosen as follows: Each constraint is (independently) is chosen of the form $x_i^{b_1} \vee x_j^{b_2} \vee x_k^{b_3}$, where: (i, j, k) is a uniformly random triple of distinct indices in $[n]$; $b_1, b_2, b_3 \in \{0, 1\}$ are chosen uniformly and independently; and, "x^1" just signifies the variables x, whereas "x^0" signifies the logical negation of x.

   Show that for all $\varepsilon > 0$, there is a large enough constant $C = C(\varepsilon)$ such that, when $m \geq Cn$,
$$\mathrm{Opt}(\mathcal{I}) \leq 7/8 + \varepsilon \quad \text{except with probability at most } 2^{-n}.$$

3. (a) Show that if $G = (V, E)$ is a connected undirected graph whose normalized Laplacian L has an eigenvalue of 2, then G is bipartite. (Recall $L = I - A$, where A is the normalized adjacency matrix.)
   (b) Give an explicit counterexample showing the statement is false if we drop the hypothesis that G is connected.

4. In the *Set-Cover* problem, the input consists of a *ground set* E of n elements, a collection of m subsets $S_1, \dots, S_m \subseteq E$, and associated nonnegative *costs* $c_1, \dots, c_m$ for the subsets. The task is to find a minimum-cost collection of subsets S_i whose union is all of E.
   (a) Formulate the Set-Cover problem exactly as an integer linear program. Then indicate how to relax it to a linear program.
   (b) Derive the *dual* linear program.
   (c) Give an English-language description/interpretation of the optimization problem defined by the dual linear program.

5. (a) For $n \times n$ matrices M and N, let $\langle M, N\rangle_F = \sum_{i,j\in[n]} M[i,j] \cdot N[i,j]$. (Equivalently, $\langle M, N\rangle_F = \mathrm{Tr}(M^T N)$.) Let S^n denote the set of $n \times n$ symmetric matrices and let S^n_+ denote the set of $n \times n$ symmetric positive-semidefinite matrices. Define
$$\widetilde{S} := \{M \in S^n \mid \forall N \in S^n_+,\ \langle M, N\rangle_F \geq 0\}.$$
   Prove that $\widetilde{S} = S^n_+$.

(b) Consider a SDP of the form below where C, A_i are symmetric $n \times n$ matrices and we are optimizing over $X \in S^n_+$:

$$\begin{aligned} &\max \quad \langle C, X\rangle_F \\ &\langle A_i, X\rangle_F = b_i \quad i = 1, 2, \ldots, m \\ &X \succcurlyeq 0 \end{aligned}$$

Consider the "dual" SDP

$$\begin{aligned} \min \quad &\sum_{i=1}^{m} b_i y_i \\ &\sum_{i=1}^{m} y_i A_i \succcurlyeq C \end{aligned}$$

(where $A \succcurlyeq B$ means $A - B$ is positive semidefinite).

Show that if X and y are any feasible solutions to the above SDPs, then $\langle C, X\rangle_F \le \sum_{i=1}^{m} b_i y_i$.

(c) Consider the following SDP:

$$\begin{aligned} &\max_{X \in S^3} \quad -X_{11} \\ &X_{22} = 0 \\ &X_{11} + 2X_{23} = 2 \\ &X \succcurlyeq 0 \end{aligned}$$

i. What is its optimum value? Why?

ii. Write down its dual as in part (b). What is the optimum value of the dual?

6. Consider the Reed-Solomon code RS$[n, k]$ over a field $\mathbb{F}_q$ with evaluation points $\{a_1, a_2, \ldots, a_n\}$. Assume n is even for convenience. When the number of errors can be as high as $n/2$, one cannot uniquely recover the message polynomial, as the received word might agree with one polynomial in the first half and another polynomial in the second half.

Suppose $r \in \mathbb{F}_q^n$ is a noisy received word such that for each $i \in \{1, 2, \ldots, n\}$, $r_i \in \{p_1(a_i), p_2(a_i)\}$, and $|\{i : r_i = p_1(a_i)\}| \ge n/2$ as well as $|\{i : r_i = p_2(a_i)\}| \ge n/2$. Assume that $k \le n/4$.

(a) Show that p_1 and p_2 are the only polynomials of degree $< k$ that agree with r on at least $n/2$ positions.

(b) Give a polynomial time algorithm that given as input an r satisfying the above promise (for a unknown pair of polynomials p_1, p_2), finds the polynomials p_1, p_2.

Hint: For each i, $(r_i - p_1(a_i))(r_i - p_2(a_i)) = 0$. Use this to interpolate a polynomial in $\mathbb{F}_q[X, Y]$ that is quadratic in Y and vanishes on all (a_i, r_i).

7. (a) Let $f \in \mathbb{F}[X_1, X_2, \ldots, X_n]$ be a polynomial over a field $\mathbb{F}$ with total degree n such that the monomial $X_1 X_2 \cdots X_n$ has a nonzero coefficient in f. (Note that f is **not** assumed to be multilinear, and might have degree bigger than 1 in the X_i's.) Let $S_1, S_2, \ldots, S_n$ be arbitrary subsets of $\mathbb{F}$ with $|S_i| = 2$ for $1 \le i \le n$. Prove that $f(a) \ne 0$ for some $a \in S_1 \times S_2 \times \cdots \times S_n$.

Hint: Show that there is a nonzero multilinear polynomial g that agrees with f on $S_1 \times S_2 \times \cdots \times S_n$.

(b) Suppose $G = (V, E)$ is a 5-regular graph (i.e., every vertex has degree exactly 5). Prove that G has a subgraph H that is 3-regular.

Hint: Apply part (a) to the following polynomial in variables X_e, $e \in E$, over the field $\mathbb{F}_3$:

$$\prod_{v\in V}\Big(1 - \big(\sum_{e\in\Gamma(v)} X_e\big)^2\Big) - \prod_{e\in E}(1 - X_e)\,,$$

where $\Gamma(v)$ denotes the set of (five) edges incident on v in G.

8. The *inner product mod 2* function, $\mathrm{IP}_{2n} : \{0,1\}^{2n} \to \{0,1\}$ is defined by $\mathrm{IP}_{2n}(x,y) = \sum_{i=1}^{n} x_i y_i \pmod 2$. The goal of this problem is to show a linear lower bound on its randomized communication complexity.

(a) We wish to show that any (public-coins) randomized communication protocol with error at most $1/4$ must use at least $n/2 - 1$ bits of communication. Show that it suffices to prove the below statement:

"Let $\mathcal{R}_1 \times \mathcal{S}_1, \dots, \mathcal{R}_{2^c} \times \mathcal{S}_{2^c}$ be a partition of $\{-1,+1\}^n \times 2^{[n]}$ into combinatorial rectangles[1], and let $z_1, \dots, z_{2^c} \in \{-1,+1\}$. Suppose[2]

$$\mathop{\mathbf{E}}_{\substack{\boldsymbol{x}\sim\{-1,+1\}^n \text{ uniformly} \\ \boldsymbol{S}\sim 2^{[n]} \text{ uniformly}}}\left[\left(\sum_{i=1}^{2^c} 1_{\mathcal{R}_i}(\boldsymbol{x}) 1_{\mathcal{S}_i}(\boldsymbol{S}) z_i\right)\chi_{\boldsymbol{S}}(\boldsymbol{x})\right] \geq 1/2.$$

Then $c \geq n/2 - 1$."

(Hint: Yao's Principle is involved here.)

(b) Prove the above statement using Fourier analysis of Boolean functions. You will need Parseval's identity (as well as Cauchy–Schwarz in the form $\mathbf{E}[\boldsymbol{Z}] \leq \sqrt{\mathbf{E}[\boldsymbol{Z}^2]}$).

9. Let us make the following strong hardness assumption: $\forall \varepsilon > 0$, there exist constants k, C depending only on ε, such that determining if an input k-SAT instance on n variables and at most Cn clauses is satisfiable requires at least $2^{(1-\varepsilon)n}$ time in the worst case.

Consider the following "disjoint sets" problem: We are given a set family $\mathcal{F}$ of N subsets of a universe U, $|U| \leq O(\log N)$. The goal is to determine if there exist $S \neq T \in \mathcal{F}$ such that $S \cap T = \emptyset$.

Prove that, under the above hardness assumption, for any constant $\delta > 0$, there is no $O(N^{2-\delta})$ time algorithm for the disjoint sets problem.

Hint: Split the variables of the SAT instance into two halves, and with each of the $2^{n/2}$ partial assignments to each half, associate an appropriate subset of clauses.

[1] The notation $2^{[n]}$ means the collection of all subsets of $[n]$.

[2] Using the notation $\chi_S(x) = \prod_{i\in S} x_i$ and $1_A(b) = \begin{cases} 1 & \text{if } b \in A, \\ 0 & \text{else.} \end{cases}$.