Toolkit for TCS	**Monsoon 2018**
HW 2	
Out: September 25, 2018	**Due:** October 16, 2018

(1) Please take time to write clear and concise solutions. You are STRONGLY *encouraged to submit LaTeXed solutions by email. (2) Collaboration is OK, but please write your answers yourself, and include in your answers the names of* EVERYONE *you collaborated with and* ALL *references other than class notes you consulted.*

1. (6 points) **[The Lindeberg replacement trick, or "hybrid method"]** Let X be a random variable with mean 0, and variance 1 such that $\mathbf{E}\left[|X|^3\right] \le \gamma$ for some fixed $\gamma > 0$, and let $X_1, X_2, \ldots, X_n$ be identically distributed independent copies of X. Let $Y_1, Y_2, \ldots, Y_n$ be identically distributed independent copies of a standard normal variable (mean 0 and variance 1). Let f be a thrice differentiable function such that there exists a constant M so that for all $x \in \mathbb{R}$, $|f'(x)|, |f''(x)|, |f'''(x)| \le M$.

   For $0 \le k \le n$, define

$$T_k := \frac{1}{\sqrt{n}}\left(\sum_{i=1}^{k-1} Y_i + \sum_{i=k+1}^{n} X_i\right), \tag{1}$$

$$S_k := T_k + \frac{Y_k}{\sqrt{n}}. \tag{2}$$

   (Note that $S_{k-1} = T_k + X_k/\sqrt{n}$.)

   (a) (2 points) Show that for all $1 \le k \le n$

$$\left|f(S_k) - f(S_{k-1}) - \frac{f'(T_k)}{\sqrt{n}}(Y_k - X_k) - \frac{f''(T_k)}{2n}(Y_k^2 - X_k^2)\right| \le \frac{M}{6n^{3/2}}(|Y_k|^3 + |X|^3),$$

   (b) (4 points) Use the above to prove that that there exists a constant c (depending only upon M and γ) such that for $1 \le k \le n$,

$$|\mathbf{E}\left[f(S_k) - f(S_{k-1})\right]| \le \frac{c}{n^{3/2}},$$

   and hence conclude that

$$|\mathbf{E}\left[f(S_n)\right] - \mathbf{E}\left[f(S_0)\right]| \le \frac{c}{\sqrt{n}}.$$

   Also obtain an upper bound on c in terms of M and γ.

2. (9 points) **[KKT conditions]** Let $f, g_1, g_2, \ldots, g_k : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable functions. Consider the following program:

$$\begin{aligned} \min \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0,\ 1 \le i \le k \\ & h_i^T x = l_i,\ 1 \le i \le m \end{aligned}$$

   Let x^* be a feasible optimum for the above program, and let T be the set of inequalities of the program that are tight at x^*.

   (a) (2 points) Show that for x^* to be a feasible optimum, it is necessary that

$$\neg\exists v \in \mathbb{R}^n, v \ne 0, v^T h_i = 0 \text{ for } 1 \le i \le m,\ v^T \nabla g_i(x^*) < 0 \text{ for } i \in T, \text{ and } v^T \nabla f(x^*) < 0.$$

   (b) (4 points) Show that there exists a matrix A and a vector b such that the condition in the previous item is equivalent to the infeasibility of

$$Ax \le 0 \text{ and } b^T x > 0.$$

(c) (2 points) Use A and b as constructed in the previous part to show that for x^* to be a feasible optimum it is necessary that there exist non-negative $\mu_0, \mu_1, \ldots, \mu_k$ not all zero and reals $\nu_1, \nu_2, \ldots, \nu_m$ such that

$$\mu_0 \nabla f(x^*) + \sum_{i=1}^{k} \mu_i \nabla g_i(x^*) + \sum_{i=1}^{m} \nu_i h_i = 0.$$

(d) (1 point) Assume now that the vectors $\{\nabla g_i(x^*)\}_{i \in T}$ and $\{h_i\}_{1 \leq i \leq m}$ (i.e., the gradients of all the tight constraints) are linearly independent. Show then that for x^* to be a feasible optimum it is necessary that there exist non-negative $\mu_1, \ldots, \mu_k$ and reals $\nu_1, \nu_2, \ldots, \nu_m$ such that

$$-\nabla f(x^*) = \sum_{i=1}^{k} \mu_i \nabla g_i(x^*) + \sum_{i=1}^{m} \nu_i h_i,$$

and that $\mu_i g_i(x^*) = 0$ for $1 \leq i \leq k$.

3. (6 points) Recall that $\langle A, B \rangle := \sum_{1 \leq i,j \leq n} A_{ij} B_{ij}$ is the natural inner product on the space of $n \times n$ matrices obtained by considering them as vectors in $\mathbb{R}^{n^2}$.

Suppose now that we are given $n \times n$ symmetric matrices $A_1, A_2, \ldots, A_m$ and real numbers $b_1, b_2, \ldots, b_m$. Consider the following system, which we call $\mathcal{P}$, of inequalities in terms of a variable matrix X:

$$\langle A_i, X \rangle \leq b_i, \; 1 \leq i \leq m,$$
$$X \text{ is symmetric and } X \succeq 0.$$

Consider also the following system, which we call $\mathcal{D}$ where the variables form a vector $y \in \mathbb{R}^m$:

$$\sum_{i=1}^{m} y_i A_i \succeq 0, \text{ and}$$
$$y^T b = \sum_{i=1}^{m} y_i b_i < 0.$$

Show that exactly one of $\mathcal{P}$ and $\mathcal{D}$ is feasible.

Hint: It might help to consider the set $S := \{(\langle A_1, X\rangle, \langle A_2, X\rangle, \ldots, \langle A_m, X\rangle) \mid X \text{ is PSD and symmetric}\}$.

4. (8 points) **[(Modified) Problem 15.5 in V. Vazirani's *Approximation Algorithms*]** Let $G = (V, E)$ be an undirected, weighted graph with positive weights $(w_e)_{e \in E}$.

(a) (0 points) By introducing an integral variable x_e for each edge $e \in E$, write an integer linear program whose optimal value is the maximum weight matching in G.

(b) (4 points) Consider the natural linear programming relaxation of the integer program in the previous part. Show that its optimal value is no more than twice the optimal value of the integer program (in other words, the integrality gap of the relaxation lies between 1/2 and 1).

(c) (4 points) Suppose now that the graph is bipartite. Show that the optimum values of the relaxation and the integer program are the same (in other words, the integrality gap of the relaxation is 1).

5. (8 points) **[From Satish Rao's homework assignments]** In the experts framework, we compared an online algorithm that was allowed to pick among n (expert) strategies, with the best choice of a single expert in hindsight. This can be seen as bound on a quantity known as *expected regret*. This is defined as how much worse the online player does, in expectation, than the best offline player; i.e., if the best expert suffers a loss of $L^\star$ while the expected loss of the online player is $\mathbf{E}[L]$, then the regret is $\mathbf{E}[L] - L^\star$.

Assume that the game lasts for T rounds, and the loss in each round is in the interval $[0, 1]$.

(a) (4 points) Show that when T is large enough (compared to $\log n$) the expected regret suffered by the experts algorithm (with appropriate parameters) compared to the best expert is $O\left(\sqrt{T \cdot \log n}\right)$.

(b) (4 points) How important was it that the online player was allowed to switch between experts, while the offline player had to stick to a single expert? To answer let us consider the regret suffered by the online player compared to an offline player who is allowed to switch between experts at every step. Construct an example where the loss $L^\star$ suffered by the offline player is 0, while the regret of the online player is at least $T \cdot \left(1 - \frac{1}{n}\right)$.