CSE 599: Interplay between Convex Optimization and Geometry **Winter 2018**

Homework 2

Lecturer: Yin Tat Lee *Due Date: 2 Mar 2018*

You are allowed to discuss with others. If you discuss with others, please list your collaborators for each question. In any case, you must write your own solutions. This homework is counted 25% of the course.

1. Low Rank Update

1a. (10 Marks) Show that

$$\sigma_{i,S\cup\{i\}} = \begin{cases} \sigma_{i,S} & \text{if } i \in S \\ \frac{1}{1+\frac{1}{\sigma_{i,S}}} & \text{else} \end{cases}.$$

Recall the definition from Section 10.3.2 (Lecture 10).
Hints: $(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$.

1b. (20 Marks) Consider the least square problem

$$\min_x \|Bx - b\|^2 + \|Ex - d\|^2$$

where $B \in \mathbb{R}^{n\times n}$ with $O(n)$ non-zeros and $E \in \mathbb{R}^{r\times n}$ with $O(rn)$ non-zeros. Given the Cholesky decomposition of $B^\top B = LL^\top$ where L has $O(n)$ non-zeros. Show how to find the minimizer x in

$$O(n + nr^2 + r^3) \text{ time.}$$

Hints: Compute the formula of the solution x using the optimality condition of the least square problem. Then use $(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$.
This is how LP solvers solve linear systems by splitting it into sparse rows/cols B and dense rows/cols E.

2. Leverage Score

2a. (10 Marks) Given a matrix A such that $\alpha I \preceq A^\top A \preceq \beta I$. Let a_i be the i^{th} row of A. Show that

$$\frac{1}{\beta}\|a_i\|^2 \le \sigma_i(A) \le \frac{1}{\alpha}\|a_i\|^2$$

2b. (20 Marks, Bonus.) Let $A \in \mathbb{R}^{m\times n}$ be a matrix where each row is a random unit vector and $m = cn\log n$ for some large enough $c > 0$. Use Theorem 10.3.5 to show that

$$\sigma_i(A) = \Theta(\frac{n}{m}).$$

3. Geometric Median This question relies on the following result and proof:
For any convex function f, we define the subgradient ∂f at x by

$$\partial f(x) = \{g \in \mathbb{R}^n \text{ such that } f(y) \ge f(x) + g^\top(y-x) \text{ for all } y\}.$$

Theorem (Mirror Descent). *Given a G-Lipschitz convex function f and an initial point $x^{(1)}$ such that $\left\|x^{(1)} - x^*\right\| \le R$ where x^* is a minimizer of f. Fix an integer T. Consider the algorithm $x^{(t+1)} = x^{(t)} - \frac{R}{G\sqrt{T}}g^{(t)}$ where $g^{(t)} \in \partial f(x^{(t)})$. Then, we have that*

$$f(\frac{1}{T}\sum_{s=1}^{T} x^{(s)}) - f(x^*) \le \frac{GR}{\sqrt{T}}.$$

Proof. Note that

$$f(x^{(s)}) - f(x^*) \leq g^{(s)\top}(x^{(s)} - x^*) = \frac{R}{2G\sqrt{T}} \left\| g^{(s)} \right\|^2 + \frac{G\sqrt{T}}{2R} \left(\left\| x^{(s)} - x^* \right\|^2 - \left\| x^{(s+1)} - x^* \right\|^2 \right).$$

Summing both sides from $s = 1$ to T and using convexity of f, we have

$$\begin{aligned} f(\frac{1}{T}\sum_{s=1}^{T} x^{(s)}) - f(x^*) &\leq \frac{1}{T}\sum_{s=1}^{T} \left(f(x^{(s)}) - f(x^*) \right) \\ &\leq \frac{RG}{2\sqrt{T}} + \frac{G\sqrt{T}}{2RT} \left(\left\| x^{(1)} - x^* \right\|^2 - \left\| x^{(T+1)} - x^* \right\|^2 \right) \leq \frac{GR}{\sqrt{T}}. \end{aligned}$$

□

3a. (10 Marks) Given $a_i \in \mathbb{R}^d$. We define the geometric median problem by

$$f(x) = \sum_{i=1}^{n} \| x - a_i \|_2 .$$

For any x, give a formula of one of the vector $g \in \partial f(x)$. (There may have more than one vector in $\partial f(x)$. Giving any one of them is enough.)

3b. (20 Marks) Show that Mirror descent (mentioned above) can find $x \in \mathbb{R}^d$ such that $f(x) \leq (1+\varepsilon)\min_{x\in\mathbb{R}^d} f(x)$ in $O(\varepsilon^{-2})$ iterations.
Hints: 1) Show that $\left\| x^* - \frac{1}{n}\sum_{i=1}^{n} a_i \right\| \leq \frac{1}{n} f(x^*)$ where x^* is any minimizer of $f(x)$. Using this, show that we have an initial point with distance $R = \frac{1}{n} f(x^*)$ to x^*. 2) Show that f is n-Lipschitz.

3c. (5 Marks) Show that we can find $x \in \mathbb{R}^d$ such that $f(x) \leq (1+\varepsilon)\min_{x\in\mathbb{R}^d} f(x)$ in $O(nd\varepsilon^{-2})$ time.

3d. (10 Marks) Prove the following theorem:

Theorem (Stochastic Mirror Descent)**.** *Given a convex function f and an initial point $x^{(1)}$ such that $\left\| x^{(1)} - x^* \right\| \leq R$ where x^* is a minimizer of f. Fix an integer T. Consider the algorithm $x^{(t+1)} = x^{(t)} - \frac{R}{G\sqrt{T}} g^{(t)}$ where $g^{(t)}$ is random variable such that $\mathbb{E}g^{(t)} \in \partial f(x^{(t)})$ and $\mathbb{E}\left\| g^{(t)} \right\|_2^2 \leq G^2$. Then, we have that*

$$\mathbb{E}f(\frac{1}{T}\sum_{s=1}^{T} x^{(s)}) - f(x^*) \leq \frac{GR}{\sqrt{T}}.$$

3e. (10 Marks) Give a formula of a random vector g_x such that g_x can be computed in $O(d)$ time and that $\mathbb{E}g_x = \partial f(x)$ and $\mathbb{E}\|g_x\|_2^2 \leq n^2$.

3f. (5 Marks) Using 3d and 3e, show that we can find $x \in \mathbb{R}^d$ such that

$$\mathbb{E}f(x) \leq (1+\varepsilon) \min_{x\in\mathbb{R}^d} f(x)$$

in $O(nd + d\varepsilon^{-2})$ time.

3g. (20 Marks, Bonus.) Do whatever you can to improve the running time and explain what is the best possible.