SEEM 5380: Optim. Methods for High–Dim. Statistics	2016–17 Second Term

Take–Home Final Examination

Release Date: April 24, 2017 — Due: 5pm, May 10, 2017

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (20pts). Let $X \in \mathbb{R}^{m\times n}$ be given with $m \le n$. Consider

$$\mathrm{prox}_{\|\cdot\|_*}(X) = \arg\min_{Z\in\mathbb{R}^{m\times n}} \left\{ \frac{1}{2}\|X - Z\|_F^2 + \|Z\|_* \right\}.$$

Suppose that $X = U\Sigma V^T$ is a singular value decomposition of X, where $\Sigma = \mathrm{Diag}\left(\{\sigma_i(X)\}_{i=1}^m\right)$. Show that

$$\mathrm{prox}_{\|\cdot\|_*}(X) = U\Sigma_1 V^T,$$

where

$$\Sigma_1 = \mathrm{Diag}\left(\{\max\{0, \sigma_i(X) - 1\}\}_{i=1}^m\right).$$

Problem 2 (25pts). This problem is a continuation of Problem 3 of Homework 2. Recall that the optimal solution set is given by

$$\mathcal{X} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\}.$$

Let $\{\delta_k\}_{k\ge 0}$ be a sequence satisfying $\delta_k \searrow 0$. Define

$$X^k = \begin{bmatrix} 1 + 2\delta_k^2 & \delta_k \\ \delta_k & \delta_k^2 \end{bmatrix} \quad \text{for } k = 0, 1, \ldots.$$

(a) **(5pts).** Show that $d(X^k, \mathcal{X}) = \Theta(\delta_k)$.

(b) **(15pts).** Using the result of Problem 1, show that

$$E(X^k) \triangleq \mathrm{prox}_{\|\cdot\|_*}(X^k - \nabla(h(\mathcal{A}(X^k)))) - X^k = \begin{bmatrix} -\delta_k^2 & 0 \\ 0 & \delta_k^2 \end{bmatrix}.$$

(c) **(5pts).** Using the results in (a) and (b), show that there does not exist a constant $\mu > 0$ satisfying $d(X^k, \mathcal{X}) \le \mu \cdot \|E(X^k)\|_F$ for all $k \ge 0$. In other words, the local error bound fails to hold for the optimization problem in Problem 3 of Homework 2.

Problem 3 (25pts). Let $\mathcal{L} : \mathbb{R}^d \to \mathbb{R}$ be a smooth loss function. Recall that $\mathcal{L}$ satisfies the restricted strong convexity property at $\theta^* \in \mathbb{R}^d$ if

$$(\nabla\mathcal{L}(\theta^* + \Delta) - \nabla\mathcal{L}(\theta^*))^T\Delta \ge \begin{cases} \alpha_1\|\Delta\|_2^2 - \tau_1\dfrac{\log d}{n}\|\Delta\|_1^2 & \text{if } \|\Delta\|_2 \le 1, \quad (a) \\[2ex] \alpha_2\|\Delta\|_2 - \tau_2\sqrt{\dfrac{\log d}{n}}\|\Delta\|_1 & \text{if } \|\Delta\|_2 \ge 1 \quad (b) \end{cases}$$

for some constants $\alpha_1, \alpha_2 > 0$ and $\tau_1, \tau_2 \geq 0$. Suppose that (i) $\mathcal{L}$ is convex, (ii) $\|\Delta\|_1 \leq 2R$, (iii) $n \geq 4R^2\tau_1^2 \log d$, and (iv) condition (a) holds. Show that condition (b) also holds with $\alpha_2 = \alpha_1$ and $\tau_2 = 1$.

Problem 4 (20pts). Recall the generalized power method (GPM) for solving the following phase synchronization problem:

$$\hat{z} \in \arg\max_{z \in \mathbb{T}^n} \left\{ f(z) \triangleq z^H C z \right\},$$

where $C = (z^\star)(z^\star)^H + \Delta$ and $\mathbb{T}^n = \{z \in \mathbb{C}^n : |z_j| = 1 \text{ for } j = 1, \ldots, n\}$. Show that the iterates $\{z^k\}_{k \geq 0}$ generated by the GPM satisfies

$$f(z^{k+1}) - f(z^k) \geq \lambda_{\min}\left(\Delta + \frac{n}{\alpha} I\right) \|z^{k+1} - z^k\|_2^2,$$

where $\lambda_{\min}\left(\Delta + \frac{n}{\alpha} I\right)$ is, as usual, the smallest eigenvalue of $\Delta + \frac{n}{\alpha} I$ and $\alpha > 0$ is the step size in the GPM.