SEEM 5380 Final Examination

Zhizhong Li
May 2017
Information Engineering, CUHK
`1155070507, lz015@ie.cuhk.edu.hk`

1 Problem 1

Both the Frobenius norm and the nuclear norm are invariant under left or right multiplication by a orthogonal matrix. Thus the problem is can be rewritten as

$$\begin{aligned}\operatorname{prox}_{\|\cdot\|_*}(X) &= \operatorname*{argmin}_{Z\in\mathbb{R}^{m\times n}} \frac{1}{2}\|X-Z\|_F^2 + \|Z\|_* \\ &= \operatorname*{argmin}_{Z\in\mathbb{R}^{m\times n}} \frac{1}{2}\|U^TXV - U^TZV\|_F^2 + \|U^TZV\|_* \\ &= \operatorname*{argmin}_{Y\in\mathbb{R}^{m\times n}} \frac{1}{2}\|\Sigma - Y\|_F^2 + \|Y\|_*,\end{aligned} \tag{1}$$

where $Y := U^TZV$. We only need to prove that $Y = \Sigma_1$ is the minimum point in the simplified formulation. Since the problem is convex (because it is the sum of two norms), it suffices to show that the first order optimality condition holds at point $Y = \Sigma_1$,

$$0 \in \partial_Y \left(\frac{1}{2}\|\Sigma - Y\|_F^2 + \|Y\|_* \right)\bigg|_{Y=\Sigma_1}. \tag{2}$$

To simplify notation, denote $\sigma_i := \sigma_i(X), i = 1, 2, \ldots, m$, and suppose the first r eigenvalues are greater than 1, *i.e.*

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 1 \geq \sigma_{r+1} \geq \cdots \geq \sigma_m \geq 0. \tag{3}$$

Then we can write Σ and Σ_1 as

$$\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_m), \tag{4}$$

and

$$\Sigma_1 = \operatorname{diag}(\sigma_1 - 1, \ldots, \sigma_r - 1, 0, \ldots, 0). \tag{5}$$

Denote $W_1 = \operatorname{diag}(\sigma_{r+1}, \ldots, \sigma_m)$. Then the subgradient at Σ_1 is

$$\begin{aligned}&\partial_Y \left(\frac{1}{2}\|\Sigma - Y\|_F^2 + \|Y\|_* \right)\bigg|_{Y=\Sigma_1} \\ &= \left\{ \Sigma_1 - \Sigma + \begin{bmatrix} I_r & 0 \\ 0 & W \end{bmatrix} : \|W\| \leq 1 \right\} \\ &= \left\{ \begin{bmatrix} -I_r & 0 \\ 0 & -W_1 \end{bmatrix} + \begin{bmatrix} I_r & 0 \\ 0 & W \end{bmatrix} : \|W\| \leq 1 \right\} \\ &= \left\{ \begin{bmatrix} 0 & 0 \\ 0 & W - W_1 \end{bmatrix} : \|W\| \leq 1 \right\}\end{aligned} \tag{6}$$

where $\|W\|$ is the spectral norm, and we used the result from HW2, Problem 2 to compute the subgradient of nuclear norm. Since $\|W_1\| = \sigma_{r+1} \leq 1$, by simply making $W = W_1$, we know 0 belongs to the subgradient of the objective, as desired.

2 Problem 2

2.1

$$\begin{aligned}
d(X^k, \mathcal{X}) &= \left\| \begin{bmatrix} 2\delta_k^2 & \delta_k \\ \delta_k & \delta_k^2 \end{bmatrix} \right\|_F \\
&= \sqrt{2\delta_k^2 + 5\delta_k^4} \\
&= \sqrt{2}\delta_k \sqrt{1 + \frac{5}{2}\delta_k^2} \\
&= \sqrt{2}\delta_k \left(1 + \frac{5}{4}\delta_k^2 + O(\delta_k^4)\right) \\
&= \Theta(\delta_k).
\end{aligned} \tag{7}$$

2.2

In HW2, we have computed the gradient of $h(\mathcal{A}(X))$ as

$$\nabla_X h(\mathcal{A}(X)) = \begin{bmatrix} \frac{3}{2}X_{11} - 2X_{22} - \frac{5}{2} & 0 \\ 0 & -2X_{11} + 3X_{22} + 1 \end{bmatrix}. \tag{8}$$

So

$$\nabla_X h(\mathcal{A}(X^k)) = \begin{bmatrix} \delta_k^2 - 1 & 0 \\ 0 & -\delta_k^2 - 1 \end{bmatrix}. \tag{9}$$

And we have

$$\begin{aligned}
X^k - \nabla_X h(\mathcal{A}(X^k)) &= \begin{bmatrix} \delta_k^2 + 2 & \delta_k \\ \delta_k & 2\delta_k^2 + 1 \end{bmatrix} \\
&= \begin{bmatrix} \frac{-\delta_k}{\sqrt{1+\delta_k^2}} & \frac{1}{\sqrt{1+\delta_k^2}} \\ \frac{1}{\sqrt{1+\delta_k^2}} & \frac{\delta_k}{\sqrt{1+\delta_k^2}} \end{bmatrix} \begin{bmatrix} 1 + \delta_k^2 & 0 \\ 0 & 2 + 2\delta_k^2 \end{bmatrix} \begin{bmatrix} \frac{-\delta_k}{\sqrt{1+\delta_k^2}} & \frac{1}{\sqrt{1+\delta_k^2}} \\ \frac{1}{\sqrt{1+\delta_k^2}} & \frac{\delta_k}{\sqrt{1+\delta_k^2}} \end{bmatrix}.
\end{aligned} \tag{10}$$

So,

$$\begin{aligned}
\mathrm{prox}_{\|\cdot\|_*}(X^k - \nabla_X h(\mathcal{A}(X^k))) &= \frac{1}{1+\delta_k^2} \begin{bmatrix} -\delta_k & 1 \\ 1 & \delta_k \end{bmatrix} \begin{bmatrix} \delta_k^2 & 0 \\ 0 & 1 + 2\delta_k^2 \end{bmatrix} \begin{bmatrix} -\delta_k & 1 \\ 1 & \delta_k \end{bmatrix} \\
&= \begin{bmatrix} 1 + \delta_k^2 & \delta_k \\ \delta_k & 2\delta_k^2 \end{bmatrix}.
\end{aligned} \tag{11}$$

So

$$E(X^k) = \mathrm{prox}_{\|\cdot\|_*}(X^k - \nabla_X h(\mathcal{A}(X^k))) - X^k = \begin{bmatrix} -\delta_k^2 & 0 \\ 0 & \delta_k^2 \end{bmatrix}. \tag{12}$$

2.3

Since

$$\|E(X^k)\|_F = \sqrt{2}\delta_k^2 = \Theta(\delta_k^2) \tag{13}$$

is a higher order infinitesimal than $d(X^k, \mathcal{X}) = \Theta(\delta_k)$, so there does not exist a constant $\mu > 0$, such that $d(X^k, \mathcal{X}) < \mu \|E(X^k)\|_F$ holds for all k.

3 Problem 3

First observe that the right hand side of the condition (b) is 1-homogeneous. This motivate us to consider if we can reduce the problem to the case when $\|\Delta\|_2 = 1$, since the absolute length is unimportant here. However, we need to deal with the left hand side because Δ appears in the term $\nabla\mathcal{L}(\theta^* + \Delta)$. This is easy to handle due to the convexity of $\mathcal{L}$. Since $(\nabla\mathcal{L}(\theta^* + \Delta))^T\Delta/\|\Delta\|_2$ is the directional derivative of $\mathcal{L}$ at $(\theta^* + \Delta)$ along direction Δ. Restrict $\mathcal{L}$ on line $t \mapsto \theta^* + t\Delta$, function $f(t) := \mathcal{L}(\theta^* + t\Delta)$ is convex and its derivative is increasing. Thus for $\|\Delta\|_2 \geq 1$,

$$(\nabla\mathcal{L}(\theta^* + \Delta))^T \frac{\Delta}{\|\Delta\|_2} = \left.\frac{df(t)}{dt}\right|_{t=1} \geq \left.\frac{df(t)}{dt}\right|_{t=\frac{1}{\|\Delta\|_2}} = (\nabla\mathcal{L}(\theta^* + \Delta/\|\Delta\|_2))^T \frac{\Delta}{\|\Delta\|_2}. \tag{14}$$

So

$$(\nabla\mathcal{L}(\theta^* + \Delta) - \nabla\mathcal{L}(\theta^*))^T\Delta \geq (\nabla\mathcal{L}(\theta^* + \Delta/\|\Delta\|_2) - \nabla\mathcal{L}(\theta^*))^T\Delta. \tag{15}$$

We then turn the condition (b) to a homogeneous form

$$(\nabla\mathcal{L}(\theta^* + \Delta/\|\Delta\|_2) - \nabla\mathcal{L}(\theta^*))^T\Delta \geq \alpha_2\|\Delta\|_2 - \tau_2\sqrt{\frac{\log d}{n}}\|\Delta\|_1. \tag{16}$$

So we can assume that $\|\Delta\|_2 = 1$ because it will automatically be true for $\|\Delta\|_2 > 1$. Use condition (a) when $\|\Delta\|_2 = 1$, we have

$$(\nabla\mathcal{L}(\theta^* + \Delta) - \nabla\mathcal{L}(\theta^*))^T\Delta \geq \alpha_1 - \tau_1\frac{\log d}{n}\|\Delta\|_1^2. \tag{17}$$

we only need to verify that

$$\alpha_1 - \tau_1\frac{\log d}{n}\|\Delta\|_1^2 \geq \alpha_2 - \sqrt{\frac{\log d}{n}}\|\Delta\|_1. \tag{18}$$

This is equivalent to

$$\tau_1\frac{\log d}{n}\|\Delta\|_1^2 \leq \sqrt{\frac{\log d}{n}}\|\Delta\|_1, \tag{19}$$

and equivalent to

$$\tau_1\sqrt{\frac{\log d}{n}}\|\Delta\|_1 \leq 1, \tag{20}$$

This holds because

$$\tau_1\sqrt{\frac{\log d}{n}}\|\Delta\|_1 \leq \tau_1\sqrt{\frac{\log d}{4R^2\tau_1^2\log d}}2R = 1. \tag{21}$$

4 Problem 4

The proof can be found in page 19 of [1]. I do not want to copy & paste so leave the following un-proved status as it is. This is unfortunate, been too confident to solve this in the deadline day...

Note: the following is fruitless.

The update rule of GPM is,

$$\begin{aligned} w^k &\leftarrow z^k + \frac{\alpha}{n} C z^k, \\ z^{k+1} &\leftarrow \frac{w^k}{|w^k|}. \end{aligned} \tag{22}$$

So we have

$$\begin{aligned} f(z^{k+1}) - f(z^k) &= (z^{k+1})^H C z^{k+1} - (z^k)^H C z^k \\ &= |(z^{k+1})^H z^*|^2 - |(z^k)^H z^*|^2 + (z^{k+1})^H \Delta z^{k+1} - (z^k)^H \Delta z^k \\ &= (z^{k+1})^H \Delta z^{k+1} - (z^k)^H \Delta z^k \\ &= (z^{k+1})^H \left(\Delta + \frac{n}{\alpha} I\right) z^{k+1} - (z^k)^H \left(\Delta + \frac{n}{\alpha} I\right) z^k \\ &= (z^{k+1} - z^k)^H \left(\Delta + \frac{n}{\alpha} I\right) (z^{k+1} - z^k) + \\ &\quad (z^{k+1} - z^k)^H \left(\Delta + \frac{n}{\alpha} I\right) z^k + (z^k)^H \left(\Delta + \frac{n}{\alpha} I\right) (z^{k+1} - z^k) \\ &= (z^{k+1} - z^k)^H \left(\Delta + \frac{n}{\alpha} I\right) (z^{k+1} - z^k) + \\ &\quad \mathrm{Re}\left((z^{k+1} - z^k)^H \left(\Delta + \frac{n}{\alpha} I\right) z^k \right). \end{aligned} \tag{23}$$

We only need to show

$$\mathrm{Re}\left((z^{k+1} - z^k)^H \left(\Delta + \frac{n}{\alpha} I\right) z^k \right) \geq 0. \tag{24}$$

Since

$$z^{k+1} = \left(I + \frac{\alpha}{n} C\right) \frac{z^k}{|w^k|} = \left(I + \frac{\alpha}{n} \Delta\right) \frac{z^k}{|w^k|} + \frac{\alpha (z^*)^H z^k}{n |w^k|} z^* \tag{25}$$

References

[1] H. Liu, M.-C. Yue, and A. M.-C. So. On the estimation performance and convergence rate of the generalized power method for phase synchronization. *arXiv preprint arXiv:1603.00211*, 2016.

[2] A. M. So. SEEM 5380 Course, 2017. URL `http://www1.se.cuhk.edu.hk/~manchoso/1617/seem5380/`.