SEEM 5380 HW1

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

1 Problem 1

1.1

Under the conditions specified in the problem, we cannot arrive at that conclusion. Here is an counter example. Let $m = n = r$, then $U = V = \mathbb{R}^n$. So $\mathcal{M} = \overline{\mathcal{M}} = \mathbb{R}^{n \times n}$.

To make this question more sensible, let us impose an extra condition that $r < \min(m, n) = m$. First the fact that $\mathcal{M} \subset \overline{\mathcal{M}}$ is trivial to verify ($\forall X \in \mathcal{M}$ and $\forall Y \in \overline{\mathcal{M}}^\perp$, $X \cdot Y = 0$, so $\mathcal{M} \subset (\overline{\mathcal{M}}^\perp)^\perp = \overline{\mathcal{M}}$). We focus on why it is a proper subset.

Construct two orthogonal matrices $P \in \mathbb{R}^{m \times r}$, and $Q \in \mathbb{R}^{r \times n}$, s.t., $\mathrm{col}(P) = U$ and $\mathrm{row}(Q) = V$. Then $\forall X \in \mathcal{M}$, it can be decomposed as $X = P\Sigma Q$, where $\Sigma \in \mathbb{R}^{r \times r}$ (This is fairly easy to see after an svd on X). The decomposition is unique. For two decompositions of X, say $X = P\Sigma_1 Q = P\Sigma_2 Q$, then multiply P^T and Q^T at the left and right hand of the equation, we have $\Sigma_1 = \Sigma_2$. So we get an one-to-one mapping from $\mathcal{M}$ to $\mathbb{R}^{r \times r}$. This mapping is also linear (easy to verify). Thus $\dim \mathcal{M} \leq \dim \mathbb{R}^{r \times r} = r^2$. Similar story goes to $\overline{\mathcal{M}}^\perp$ and we have $\dim \overline{\mathcal{M}}^\perp \leq \dim \mathbb{R}^{(m-r)\times(n-r)} = (m-r)(n-r)$. Thus $\dim \mathcal{M} + \dim \overline{\mathcal{M}}^\perp \leq r^2 + (m-r)(n-r) < mn = \dim \mathbb{R}^{m \times n}$. This is a contradiction if $\mathcal{M} = \overline{\mathcal{M}}$ because the sum of $\dim \overline{\mathcal{M}} (= \dim \mathcal{M})$ and $\dim \overline{\mathcal{M}}^\perp$ should be mn then.

1.2

Let orthogonal matrices $P = (P_1\ P_2) \in \mathbb{R}^{m \times m}$ and $Q = (Q_1^T\ Q_2^T)^T \in \mathbb{R}^{n \times n}$, s.t. $\mathrm{col}(P_1) = U, \mathrm{col}(P_1) = U^\perp, \mathrm{row}(Q_1) = V, \mathrm{row}(Q_2) = V^\perp$. Then similar to the last sub-question, $\forall X \in \mathcal{M}, \forall Y \in \overline{\mathcal{M}}^\perp$, we can decompose them as

$$X = [P_1 \quad 0] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q_1 \\ 0 \end{bmatrix}, \quad Y = [0 \quad P_2] \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} 0 \\ Q_2 \end{bmatrix}, \tag{1}$$

where $\Sigma_1 \in \mathbb{R}^{r \times r}$ and $\Sigma_2 \in \mathbb{R}^{(m-r)\times(n-r)}$. So,

$$X + Y = [P_1 \quad P_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}. \tag{2}$$

Thus, $\|X + Y\|_* = \|\Sigma_1 + \Sigma_2\|_* = \|\Sigma_1\|_* + \|\Sigma_2\|_* = \|X\|_* + \|Y\|_*$.

2 Problem 2

2.1

Since $(\mathbb{E}[\|g\|_2])^2 \leq \mathbb{E}[\|g\|_2^2]$, and $\|g\|_2^2$ follows a chi-square distribution, we have $\mathbb{E}[\|g\|_2^2] = n$, and thus $\mathbb{E}[\|g\|_2] \leq \sqrt{n}$.

2.2

$$\begin{aligned} \Pr(\|g\|_2^2 \leq \alpha n) &= \Pr(\exp(t(\alpha n - \|g\|_2^2)) \geq 1) \\ &\leq \mathbb{E}[\exp(t(\alpha n - \|g\|_2^2))] \\ &= \exp(t\alpha n)\mathbb{E}(\exp(-t\|g\|_2^2)) \\ &= \exp(t\alpha n)(1+2t)^{-n/2}. \end{aligned} \tag{3}$$

Note that the inequality comes from the Markov inequality and the last equation is obtained by the moment-generating function of a Chi-squared χ_n^2 random variable being $(1-2t)^{-n/2}$.

2.3

Let $t = (1-\alpha)/(2\alpha)$, and from the inequality from last sub-question, we have

$$\Pr(\|g\|_2^2 \leq \alpha n) \leq \exp((1 - \alpha + \ln \alpha)n/2). \tag{4}$$

3 Problem 3

3.1

Let $f(x) := \|x\|_1$, then the subgradient of f at point $x = (x_1, x_2, \ldots, x_n)$ is

$$\partial f(x) = \{g \mid (g_i = 1 \text{ if } x_i > 0) \text{ and } (g_i = -1 \text{ if } x_i < 0) \text{ and } (-1 \leq g_i \leq 1 \text{ if } x_i = 0)\}. \tag{5}$$

So the optimality condition for the problem is $0 \in \mu\partial f(x) + (x - v)$, which is equivalent to the condition $(v - x)/\mu \in \partial f(x)$.

3.2

From the condition in last sub-question, the the point $x^* = (x_1, x_2, \ldots, x_n)$ which reaches the minimum should satisfy

$$x_i = \begin{cases} v_i - \mu & \text{if } v_i > \mu, \\ v_i + \mu & \text{if } v_i < \mu, \\ 0 & \text{if } -\mu \leq v_i \leq \mu. \end{cases} \tag{6}$$

References

[1] A. M. So. Assignments, 2017. URL `http://www1.se.cuhk.edu.hk/~manchoso/1617/seem5380/`.