MAS374 Optimization Theory
Homework #5

20150597 Jeonghwan Lee

Department of Mathematical Sciences, KAIST

November 10, 2021

Problem 1 (*Exercise 8.4* in [1]).

(1) Define $f : \mathbb{R}^2 \to [-\infty, +\infty)$ by $f(\alpha, \beta) := \inf \left\{ \alpha d + \frac{\beta^2}{d} : d \in (0, +\infty) \right\}$. Let us consider the following three cases:

- $\alpha < 0$: It holds for any $t \in [1, +\infty)$ that

$$\begin{aligned} f(\alpha, \beta) &\leq \inf \left\{ \alpha d + \frac{\beta^2}{d} : d \in [1, +\infty) \right\} \\ &\leq \inf \left\{ \alpha d + \beta^2 : d \in [1, +\infty) \right\} \\ &\leq \alpha t + \beta^2. \end{aligned} \tag{1.1}$$

  By letting $t \to$ in the inequality (1.1), we have

$$f(\alpha, \beta) \leq \lim_{t \to \infty} \left(\alpha t + \beta^2 \right) = -\infty,$$

  since $\alpha < 0$. Thus, $f(\alpha, \beta) = -\infty$ for every $(\alpha, \beta) \in (-\infty, 0) \times \mathbb{R}$;

- $\alpha = 0$: If $\beta = 0$, then it's clear that $f(0, 0) = 0$ and the optimal value $f(0, 0) = 0$ is attained at every feasible point $d \in (0, +\infty)$. If $\beta \neq 0$, then it holds that

$$f(0, \beta) = \inf \left\{ \frac{\beta^2}{d} : d \in (0, +\infty) \right\} = 0,$$

  and the optimal value $f(0, \beta) = 0$ cannot be attained at any feasible point in $(0, +\infty)$. Therefore, we have $f(0, \beta) = 0$ for any $\beta \in \mathbb{R}$;

- $\alpha > 0$: We may observe that

$$\alpha d + \frac{\beta^2}{d} = \left(\sqrt{\alpha d} - \frac{|\beta|}{\sqrt{d}} \right)^2 + 2 |\beta| \sqrt{\alpha}, \ \forall d \in (0, +\infty). \tag{1.2}$$

  From the identity (1.2), one can see that

$$f(\alpha, \beta) = \inf \left\{ \left(\sqrt{\alpha d} - \frac{|\beta|}{\sqrt{d}} \right)^2 + 2 |\beta| \sqrt{\alpha} : d \in (0, +\infty) \right\} = 2 |\beta| \sqrt{\alpha},$$

  and this optimal value is attained at $d^* = \frac{|\beta|}{\sqrt{\alpha}}$, for every $(\alpha, \beta) \in (0, +\infty) \times \mathbb{R}$.

To sum up, we arrive at

$$f(\alpha,\beta)=\begin{cases}-\infty & \text{if } (\alpha,\beta)\in(-\infty,0)\times\mathbb{R};\\ 2\left|\beta\right|\sqrt{\alpha} & \text{otherwise,}\end{cases}$$

as desired.

(2) To begin with, we define the objective function $g_{\mathbf{z}}(\cdot):(0,+\infty)^m\to\mathbb{R}$ by

$$g_{\mathbf{z}}\left(d_1,d_2,\cdots,d_m\right):=\frac{1}{2}\sum_{i=1}^{m}\left(d_i+\frac{z_i^2}{d_i}\right),\ \forall\mathbf{z}\in\mathbb{R}^m.$$

Then for any $\mathbf{d}=(d_1,d_2,\cdots,d_m)\in(0,+\infty)^m$, we have

$$\begin{aligned}g_{\mathbf{z}}(\mathbf{d})&\geq\frac{1}{2}\sum_{i=1}^{m}\inf\left\{t_i+\frac{z_i^2}{t_i}:t_i\in(0,+\infty)\right\}\\&=\frac{1}{2}\sum_{i=1}^{m}f\left(1,z_i\right)\\&\overset{\text{(a)}}{=}\sum_{i=1}^{m}\left|z_i\right|\\&=\left\|\mathbf{z}\right\|_1,\end{aligned}$$

where the step (a) follows from the part (1), thereby

$$\inf\left\{g_{\mathbf{z}}(\mathbf{d}):\mathbf{d}\in(0,+\infty)^m\right\}\geq\left\|\mathbf{z}\right\|_1. \tag{1.3}$$

On the other hand, let $\mathbf{d}^{(k)}\in(0,+\infty)^m$ be given by

$$d_i^{(k)}:=\begin{cases}\left|z_i\right| & \text{if } i\in\mathcal{S}(\mathbf{z});\\ \frac{1}{k} & \text{otherwise,}\end{cases}$$

where $\mathcal{S}(\mathbf{z}):=\{j\in[m]:z_j\neq0\}$ denotes the support of the vector $\mathbf{z}\in\mathbb{R}^m$. Then,

$$\begin{aligned}g_{\mathbf{z}}\left(\mathbf{d}^{(k)}\right)&=\frac{1}{2}\sum_{i\in\mathcal{S}(\mathbf{z})}2\left|z_i\right|+\frac{1}{2}\sum_{i\in[m]\setminus\mathcal{S}(\mathbf{z})}\frac{1}{k}\\&=\sum_{i\in\mathcal{S}(\mathbf{z})}\left|z_i\right|+\frac{m-\left|\mathcal{S}(\mathbf{z})\right|}{2k}\\&=\sum_{i=1}^{m}\left|z_i\right|+\frac{m-\left|\mathcal{S}(\mathbf{z})\right|}{2k}\\&=\left\|\mathbf{z}\right\|_1+\frac{m-\left|\mathcal{S}(\mathbf{z})\right|}{2k}\end{aligned}$$

for every $k\in\mathbb{N}$. So it follows that

$$\inf\left\{g_{\mathbf{z}}(\mathbf{d}):\mathbf{d}\in(0,+\infty)^m\right\}\leq g_{\mathbf{z}}\left(\mathbf{d}^{(k)}\right)=\left\|\mathbf{z}\right\|_1+\frac{m-\left|\mathcal{S}(\mathbf{z})\right|}{2k} \tag{1.4}$$

for every $k\in\mathbb{N}$. By letting $k\to\infty$ in the bound (1.4), we obtain

$$\inf\left\{g_{\mathbf{z}}(\mathbf{d}):\mathbf{d}\in(0,+\infty)^m\right\}\leq\left\|\mathbf{z}\right\|_1. \tag{1.5}$$

Taking two pieces (1.3) and (1.5) collectively, we may conclude that

$$\inf\left\{g_{\mathbf{z}}(\mathbf{d}) : \mathbf{d} \in (0,+\infty)^m\right\} = \|\mathbf{z}\|_1 \tag{1.6}$$

as desired. Note that the optimization problem

$$p_1^* := \min_{\mathbf{d}\in(0,+\infty)^m} g_{\mathbf{z}}(\mathbf{d}) = \frac{1}{2}\sum_{i=1}^m \left(d_i + \frac{z_i^2}{d_i}\right) \tag{1.7}$$

has the optimal value $p_1^* = \|\mathbf{z}\|_1$ from (1.6). For any $\mathbf{d} \in (0,+\infty)^m$, one can see that

$$g_{\mathbf{z}}(\mathbf{d}) - \|\mathbf{z}\|_1 = \frac{1}{2}\sum_{i=1}^m \frac{1}{d_i}\left(d_i - |z_i|\right)^2 = \frac{1}{2}\left[\sum_{i\in\mathcal{S}(\mathbf{z})} \frac{1}{d_i}\left(d_i - |z_i|\right)^2 + \sum_{i\in[m]\setminus\mathcal{S}(\mathbf{z})} d_i\right]. \tag{1.8}$$

Owing to the identity (1.8), one can make the following conclusion: If $[m]\setminus\mathcal{S}(\mathbf{z}) = \varnothing$, then the optimization problem (1.7) has an optimal solution $\mathbf{d}^* = (|z_1|, |z_2|, \cdots, |z_m|) \in (0,+\infty)^m$. Otherwise, the identity (1.8) yields

$$g_{\mathbf{z}}(\mathbf{d}) - \|\mathbf{z}\|_1 \geq \frac{1}{2}\sum_{i\in[m]\setminus\mathcal{S}(\mathbf{z})} d_i > 0$$

for every $\mathbf{d} \in (0,+\infty)^m$, and this implies that the optimal value $p_1^* = \|\mathbf{z}\|_1$ of the optimization problem (1.7) cannot be attained at any feasible point $\mathbf{d} \in (0,+\infty)^m$ whenever $[m]\setminus\mathcal{S}(\mathbf{z}) \neq \varnothing$. In brief,

(i) $[m]\setminus\mathcal{S}(\mathbf{z}) = \varnothing$: the optimization problem (1.7) attains an optimal solution $\mathbf{d}^* = (|z_1|, |z_2|, \cdots, |z_m|) \in (0,+\infty)^m$;

(ii) Otherwise: the optimization problem (1.7) does not attain any optimal solutions.

(3) Let us define the objective function $h_{\mathbf{z}}(\cdot) : (0,+\infty)^m \to \mathbb{R}$ by

$$h_{\mathbf{z}}(d_1, d_2, \cdots, d_m) := \sum_{i=1}^m \frac{z_i^2}{d_i}.$$

Also, let $\mathcal{X} := \{\mathbf{d} = (d_1, d_2, \cdots, d_m) \in (0,+\infty)^m : \sum_{i=1}^m d_i = 1\}$. For any $\mathbf{d} \in \mathcal{X}$, we obtain from the Cauchy-Schwarz inequality that

$$h_{\mathbf{z}}(\mathbf{d}) = \left(\sum_{i=1}^m d_i\right)\left(\sum_{i=1}^m \frac{z_i^2}{d_i}\right) \geq \left(\sum_{i=1}^m |z_i|\right)^2 = \|\mathbf{z}\|_1^2,$$

thereby it holds that

$$\inf\left\{h_{\mathbf{z}}(\mathbf{d}) : \mathbf{d} \in \mathcal{X}\right\} \geq \|\mathbf{z}\|_1^2. \tag{1.9}$$

On the other hand, we first consider the case where $\mathbf{z} = \mathbf{0} \in \mathbb{R}^m$. Then it's clear that $\|\mathbf{z}\|_1 = 0 = h_{\mathbf{z}}(\mathbf{d})$ for all $\mathbf{d} \in \mathcal{X}$, and we are done! So we may assume that $\mathbf{z} \in \mathbb{R}^m \setminus \{\mathbf{0}\}$. For $k \geq 2$, we define $\mathbf{d}^{(k)} \in (0,+\infty)^m$ by

$$d_i^{(k)} := \begin{cases} \left(1 - \frac{1}{k}\right) \cdot \frac{|z_i|}{\|\mathbf{z}\|_1} & \text{if } i \in \mathcal{S}(\mathbf{z}); \\ \frac{1}{m - |\mathcal{S}(\mathbf{z})|} \cdot \frac{1}{k} & \text{otherwise,} \end{cases}$$

where $\mathcal{S}(\mathbf{z}) := \{j \in [m] : z_j \neq 0\}$ denotes the support of the vector $\mathbf{z} \in \mathbb{R}^m \setminus \{\mathbf{0}\}$. Then,

$$\begin{aligned}
\sum_{i=1}^{m} d_i^{(k)} &= \sum_{i \in \mathcal{S}(\mathbf{z})} d_i^{(k)} + \sum_{i \in [m] \setminus \mathcal{S}(\mathbf{z})} d_i^{(k)} \\
&= \left(1 - \frac{1}{k}\right) \cdot \frac{1}{\|\mathbf{z}\|_1} \sum_{i \in \mathcal{S}(\mathbf{z})} |z_i| + \frac{1}{k} \\
&= \left(1 - \frac{1}{k}\right) + \frac{1}{k} \\
&= 1,
\end{aligned}$$

which ensures that $\mathbf{d}^{(k)} \in \mathcal{X}$ for every $k \geq 2$. Also, one can see that

$$\begin{aligned}
h_{\mathbf{z}}\left(\mathbf{d}^{(k)}\right) &= \sum_{i \in \mathcal{S}(\mathbf{z})} \frac{z_i^2}{d_i^{(k)}} \\
&= \frac{k}{k-1} \cdot \|\mathbf{z}\|_1 \left(\sum_{i \in \mathcal{S}(\mathbf{z})} |z_i| \right) \\
&= \frac{k}{k-1} \cdot \|\mathbf{z}\|_1^2 .
\end{aligned}$$

Therefore, we obtain

$$\inf \{h_{\mathbf{z}}(\mathbf{d}) : \mathbf{d} \in \mathcal{X}\} \leq h_{\mathbf{z}}\left(\mathbf{d}^{(k)}\right) = \frac{k}{k-1} \cdot \|\mathbf{z}\|_1^2 \overset{k \to \infty}{\longrightarrow} \|\mathbf{z}\|_1^2 . \tag{1.10}$$

Taking two pieces (1.9) and (1.10) collectively, one has

$$\inf \{h_{\mathbf{z}}(\mathbf{d}) : \mathbf{d} \in \mathcal{X}\} = \|\mathbf{z}\|_1^2 , \tag{1.11}$$

as desired. Here, we note that the optimization problem

$$p_2^* := \min_{\mathbf{d} \in \mathcal{X}} h_{\mathbf{z}}(\mathbf{d}) = \sum_{i=1}^{m} \frac{z_i^2}{d_i} \tag{1.12}$$

has the optimal value $p_2^* = \|\mathbf{z}\|_1^2$ from (1.11). For any $\mathbf{d} \in \mathcal{X}$, it holds that

$$\begin{aligned}
h_{\mathbf{z}}(\mathbf{d}) - \|\mathbf{z}\|_1^2 &= \left(\sum_{i=1}^{m} \frac{z_i^2}{d_i} \right) \left(\sum_{i=1}^{m} d_i \right) - \left(\sum_{i=1}^{m} |z_i| \right)^2 \\
&= \sum_{1 \leq i < j \leq m} \left(\frac{|z_i|}{\sqrt{d_i}} \cdot \sqrt{d_j} - \frac{|z_j|}{\sqrt{d_j}} \cdot \sqrt{d_i} \right)^2 \\
&= \sum_{1 \leq i < j \leq m} \frac{\left(|z_i|\, d_j - |z_j|\, d_i\right)^2}{d_i d_j}.
\end{aligned} \tag{1.13}$$

Due to the identity (1.13), one can make the following conclusions: If $[m] \setminus \mathcal{S}(\mathbf{z}) = \varnothing$, then the optimization problem (1.12) has an optimal solution $\mathbf{d}^* = \left(\frac{|z_1|}{\|\mathbf{z}\|_1}, \frac{|z_2|}{\|\mathbf{z}\|_1}, \cdots, \frac{|z_m|}{\|\mathbf{z}\|_1} \right) \in \mathcal{X}$. Otherwise, we have $z_i = 0$ for some $i \in [m]$. We claim that the optimization problem (1.12) does not attain any optimal solutions provided that

$[m] \setminus \mathcal{S}(\mathbf{z}) \neq \varnothing$ and $\mathbf{z} \in \mathbb{R}^m \setminus \{\mathbf{0}\}$. Assume towards a contradiction that the optimization problem (1.12) has an optimal solution $\mathbf{d}^* \in \mathcal{X}$. From the identity (1.13), we arrive at

$$\sum_{1 \leq i < j \leq m} \frac{\left(|z_i|\, d_j^* - |z_j|\, d_i^*\right)^2}{d_i^* d_j^*} = 0,$$

which implies that $z_j = 0$ for all $j \in [m]$ and thus we obtain a contradiction! To sum up,

(i) $[m] \setminus \mathcal{S}(\mathbf{z}) = \varnothing$: the optimization problem (1.12) has an optimal solution $\mathbf{d}^* = \left(\frac{|z_1|}{\|\mathbf{z}\|_1}, \frac{|z_2|}{\|\mathbf{z}\|_1}, \cdots, \frac{|z_m|}{\|\mathbf{z}\|_1}\right) \in \mathcal{X}$;

(ii) $[m] \setminus \mathcal{S}(\mathbf{z}) = [m]$, *i.e.*, $\mathbf{z} = \mathbf{0} \in \mathbb{R}^m$: any feasible point in $\mathcal{X}$ is an optimal solution to the optimization problem (1.12);

(iii) Otherwise: the optimal value $p_2^* = \|\mathbf{z}\|_1^2$ of the optimization problem (1.12) cannot be attained at any feasible points,

and this completes our discussion of the part (3).

Problem 2 (*Exercise 8.7* in [1])**.**

(1) We claim that the function $\phi_p(\cdot) : \mathbb{R}^{n \times m} \to \mathbb{R}_+$ is a norm on $\mathbb{R}^{n \times m}$.

Positive definiteness of $\phi_p(\cdot)$: It's clear that $\phi_p\left(\mathbf{O}_{n \times m}\right) = 0$, where $\mathbf{O}_{n \times m} \in \mathbb{R}^{n \times m}$ denotes the $n \times m$ all-zero matrix. Conversely, we assume that

$$\phi_p(\mathbf{X}) = 0 = \max\left\{\left\|\mathbf{X}^\top \mathbf{u}\right\|_p : \mathbf{u} \in \mathbb{S}^{n-1}\right\},$$

where $\mathbb{S}^{n-1} := \{\mathbf{v} \in \mathbb{R}^n : \|\mathbf{v}\|_2 = 1\}$ denotes the unit $(n-1)$-sphere. We claim that $\mathbf{X} = \mathbf{O}_{n \times m}$. To this end, we assume on a contrary that $\mathbf{X} \neq \mathbf{O}_{n \times m}$. Then $\mathbf{x}_i \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ for some $i \in [m]$, thereby

$$\begin{aligned}
0 \geq \left\|\mathbf{X}^\top \cdot \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|_2}\right\|_p &= \|\mathbf{x}_i\|_2^{-1} \cdot \left\|\mathbf{X}^\top \mathbf{x}_i\right\|_p \\
&= \|\mathbf{x}_i\|_2^{-1} \cdot \left\|\begin{bmatrix} \mathbf{x}_1^\top \mathbf{x}_i \\ \mathbf{x}_2^\top \mathbf{x}_i \\ \vdots \\ \mathbf{x}_m^\top \mathbf{x}_i \end{bmatrix}\right\|_p \\
&\geq \|\mathbf{x}_i\|_2^{-1} \cdot \left|\mathbf{x}_i^\top \mathbf{x}_i\right| \\
&= \|\mathbf{x}_i\|_2,
\end{aligned}$$

and this implies $\mathbf{x}_i = \mathbf{0}$, contradiction! So we have $\mathbf{X} = \mathbf{O}_{n \times m}$ and this establishes the positive definiteness of the map $\phi_p(\cdot) : \mathbb{R}^{n \times m} \to \mathbb{R}_+$.

Sub-additivity of $\phi_p(\cdot)$: For any $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n \times m}$, it holds that

$$
\begin{aligned}
\phi_p(\mathbf{X}+\mathbf{Y}) &= \max\left\{ \left\| (\mathbf{X}+\mathbf{Y})^\top \mathbf{u} \right\|_p : \mathbf{u} \in \mathbb{S}^{n-1} \right\} \\
&\overset{\text{(a)}}{\leq} \max\left\{ \left\| \mathbf{X}^\top \mathbf{u} \right\|_p + \left\| \mathbf{Y}^\top \mathbf{u} \right\|_p : \mathbf{u} \in \mathbb{S}^{n-1} \right\} \\
&\leq \max\left\{ \left\| \mathbf{X}^\top \mathbf{u} \right\|_p : \mathbf{u} \in \mathbb{S}^{n-1} \right\} + \max\left\{ \left\| \mathbf{Y}^\top \mathbf{u} \right\|_p : \mathbf{u} \in \mathbb{S}^{n-1} \right\} \\
&= \phi_p(\mathbf{X}) + \phi_p(\mathbf{Y}),
\end{aligned}
$$

where the step (a) follows from the fact that $\left\| \cdot \right\|_p$ is a norm on $\mathbb{R}^m$, and this establishes the sub-additivity of the map $\phi_p(\cdot) : \mathbb{R}^{n \times m} \to \mathbb{R}_+$.

Absolute homogeneity of $\phi_p(\cdot)$: For any $\mathbf{X} \in \mathbb{R}^{n \times m}$ and $\alpha \in \mathbb{R}$, one has

$$
\begin{aligned}
\phi_p(\alpha \mathbf{X}) &= \max\left\{ \left\| (\alpha \mathbf{X})^\top \mathbf{u} \right\|_p : \mathbf{u} \in \mathbb{S}^{n-1} \right\} \\
&= \max\left\{ |\alpha| \left\| \mathbf{X}^\top \mathbf{u} \right\|_p : \mathbf{u} \in \mathbb{S}^{n-1} \right\} \\
&= |\alpha| \max\left\{ \left\| \mathbf{X}^\top \mathbf{u} \right\|_p : \mathbf{u} \in \mathbb{S}^{n-1} \right\} \\
&= |\alpha| \, \phi_p(\mathbf{X}),
\end{aligned}
$$

and this establishes the absolute homogeneity of the map $\phi_p(\cdot) : \mathbb{R}^{n \times m} \to \mathbb{R}_+$.

Taking the above arguments collectively, we deduce that $\phi_p(\cdot) : \mathbb{R}^{n \times m} \to \mathbb{R}_+$ is a norm on $\mathbb{R}^{n \times m}$.

(2) We consider the following optimization problem:

$$
\begin{aligned}
\phi_2(\mathbf{X}) = \max_{\mathbf{u} \in \mathbb{R}^n} \ & \left\| \mathbf{X}^\top \mathbf{u} \right\|_2 \\
\text{subject to } & \left\| \mathbf{u} \right\|_2 = 1.
\end{aligned}
\tag{2.1}
$$

One can observe that the objective function of the optimization problem (2.1) can be expressed as

$$
\begin{aligned}
\left\| \mathbf{X}^\top \mathbf{u} \right\|_2^2 &= \sum_{i=1}^{m} \left(\mathbf{x}_i^\top \mathbf{u} \right)^2 \\
&= \sum_{i=1}^{m} \left(\mathbf{x}_i^\top \mathbf{u} \right)^\top \left(\mathbf{x}_i^\top \mathbf{u} \right) \\
&= \sum_{i=1}^{m} \mathbf{u}^\top \left(\mathbf{x}_i \mathbf{x}_i^\top \right) \mathbf{u} \\
&= \mathbf{u}^\top \left(\sum_{i=1}^{m} \mathbf{x}_i \mathbf{x}_i^\top \right) \mathbf{u} \\
&= \mathbf{u}^\top \left(\mathbf{X} \mathbf{X}^\top \right) \mathbf{u}.
\end{aligned}
\tag{2.2}
$$

At this point, let $\mathbf{X} = \mathbf{U}_r \mathbf{\Sigma} \mathbf{V}_r^\top = \sum_{j=1}^{r} \sigma_j(\mathbf{X}) \mathbf{u}_j \mathbf{v}_j^\top$ be the compact-form SVD of $\mathbf{X}$, where $r := \mathsf{rank}(\mathbf{X}) \leq \min\{m, n\}$, $\mathbf{U}_r := \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \end{bmatrix} \in \mathbb{R}^{n \times r}$, $\mathbf{V}_r := \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r \end{bmatrix} \in \mathbb{R}^{m \times r}$, and

$$
\Sigma := \mathsf{diag}\left(\sigma_1(\mathbf{X}), \sigma_2(\mathbf{X}), \cdots, \sigma_r(\mathbf{X}) \right) \in \mathbb{R}^{r \times r}; \quad \sigma_1(\mathbf{X}) \geq \sigma_2(\mathbf{X}) \geq \cdots \geq \sigma_r(\mathbf{X}) > 0.
$$

Then we have

$$\mathbf{X}\mathbf{X}^\top = \sum_{j=1}^{r} \sigma_j(\mathbf{X})^2 \cdot \mathbf{u}_j \mathbf{u}_j^\top,$$

thereby the equation (2.2) becomes

$$\begin{aligned}
\left\|\mathbf{X}^\top \mathbf{u}\right\|_2^2 &= \mathbf{u}^\top \left(\sum_{j=1}^{r} \sigma_j(\mathbf{X})^2 \cdot \mathbf{u}_j \mathbf{u}_j^\top\right) \mathbf{u} \\
&= \sum_{j=1}^{r} \sigma_j(\mathbf{X})^2 \cdot \left(\mathbf{u}_j^\top \mathbf{u}\right)^2 \\
&\le \sigma_1(\mathbf{X})^2 \left(\sum_{j=1}^{r} \left(\mathbf{u}_j^\top \mathbf{u}\right)^2\right) \\
&= \sigma_1(\mathbf{X})^2 \left(\sum_{j=1}^{n} \left(\mathbf{u}_j^\top \mathbf{u}\right)^2\right) \\
&= \sigma_1(\mathbf{X})^2 \left\{\mathbf{u}^\top \left(\sum_{j=1}^{n} \mathbf{u}_j \mathbf{u}_j^\top\right) \mathbf{u}\right\} \\
&= \sigma_1(\mathbf{X})^2 \left(\mathbf{u}^\top u\right) \\
&= \sigma_1(\mathbf{X})^2,
\end{aligned}$$

where $\{\mathbf{u}_{r+1}, \mathbf{u}_{r+2}, \cdots, \mathbf{u}_n\}$ is an orthonormal basis of $(\mathsf{span}\,(\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r\}))^\perp$, thereby $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$ forms an orthonormal basis for $\mathbb{R}^n$. Therefore, we have

$$\sup\left\{\left\|\mathbf{X}^\top \mathbf{u}\right\|_2 : \mathbf{u} \in \mathbb{S}^{n-1}\right\} \le \sigma_1(\mathbf{X}) = \sigma_{\mathsf{max}}(\mathbf{X}) = \|\mathbf{X}\|_{2\to 2}\,. \tag{2.3}$$

It's straightforward that $\left\|\mathbf{X}^\top \mathbf{u}_1\right\|_2 = \sigma_1(\mathbf{X}) = \sigma_{\mathsf{max}}(\mathbf{X}) = \|\mathbf{X}\|_{2\to 2}$. Thus, the equality holds in the bound (2.3), the optimal value of the optimization problem (2.1) is $\phi_2(\mathbf{X}) = \sigma_1(\mathbf{X}) = \sigma_{\mathsf{max}}(\mathbf{X}) = \|\mathbf{X}\|_{2\to 2}$, and the optimization problem (2.1) has an optimal solution $\mathbf{u}^* = \mathbf{u}_1$, *i.e.*,

$$\mathbf{u}_1 \in \operatorname{argmax}\left\{\left\|\mathbf{X}^\top \mathbf{u}\right\|_2 : \mathbf{u} \in \mathbb{S}^{n-1}\right\},$$

where $\mathbf{u}_1 \in \mathbb{S}^{n-1}$ refers to the first left singular vector of the matrix $\mathbf{X} \in \mathbb{R}^{n\times m}$.

(3) We consider the following optimization problem:

$$\begin{aligned}
\phi_\infty(\mathbf{X}) = \max_{\mathbf{u}\in\mathbb{R}^n}\ & \left\|\mathbf{X}^\top \mathbf{u}\right\|_\infty \\
\text{subject to }\ & \|\mathbf{u}\|_2 = 1.
\end{aligned} \tag{2.4}$$

For any $\mathbf{u} \in \mathbb{S}^{n-1}$, it holds that

$$\begin{aligned}
\left\|\mathbf{X}^\top \mathbf{u}\right\|_\infty &= \max\left\{\left|\mathbf{x}_i^\top \mathbf{u}\right| : i \in [m]\right\} \\
&\overset{\text{(b)}}{\le} \max\left\{\|\mathbf{x}_i\|_2 \cdot \|\mathbf{u}\|_2 : i \in [m]\right\} \\
&= \max\left\{\|\mathbf{x}_i\|_2 : i \in [m]\right\},
\end{aligned}$$

where the step (b) holds due to the Cauchy-Schwarz inequality. Therefore, we arrive at

$$\sup\left\{\left\|\mathbf{X}^{\top}\mathbf{u}\right\|_{\infty}:\mathbf{u}\in\mathbb{S}^{n-1}\right\}\leq\max\left\{\|\mathbf{x}_i\|_2:i\in[m]\right\}. \tag{2.5}$$

On the other hand, let $j^*\in\operatorname{argmax}\left\{\|\mathbf{x}_i\|_2:i\in[m]\right\}$. If $\mathbf{X}=\mathbf{O}_{n\times m}$, it's evident that $\phi_\infty\left(\mathbf{O}_{n\times m}\right)=0$ and any feasible point of the optimization problem (2.4) is its optimal solution. So we now may assume that $\mathbf{X}\in\mathbb{R}^{n\times m}\setminus\{\mathbf{O}_{n\times m}\}$. Then $\mathbf{x}_{j^*}\in\mathbb{R}^n\setminus\{\mathbf{0}\}$ and

$$\left\|\mathbf{X}^{\top}\cdot\frac{\mathbf{x}_{j^*}}{\|\mathbf{x}_{j^*}\|_2}\right\|_{\infty}\geq\left|\mathbf{x}_{j^*}^{\top}\cdot\frac{\mathbf{x}_{j^*}}{\|\mathbf{x}_{j^*}\|_2}\right|=\|\mathbf{x}_{j^*}\|_2=\max\left\{\|\mathbf{x}_i\|_2:i\in[m]\right\},$$

which yields

$$\sup\left\{\left\|\mathbf{X}^{\top}\mathbf{u}\right\|_{\infty}:\mathbf{u}\in\mathbb{S}^{n-1}\right\}\geq\left\|\mathbf{X}^{\top}\cdot\frac{\mathbf{x}_{j^*}}{\|\mathbf{x}_{j^*}\|_2}\right\|_{\infty}\geq\max\left\{\|\mathbf{x}_i\|_2:i\in[m]\right\}. \tag{2.6}$$

Taking two pieces (2.5) and (2.6) collectively, we find that

$$\max\left\{\left\|\mathbf{X}^{\top}\mathbf{u}\right\|_{\infty}:\mathbf{u}\in\mathbb{S}^{n-1}\right\}=\max\left\{\|\mathbf{x}_i\|_2:i\in[m]\right\}, \tag{2.7}$$

and the maximum in the left-hand side of the equation (2.7) is attained at an optimal solution $\mathbf{u}^*=\frac{\mathbf{x}_{j^*}}{\|\mathbf{x}_{j^*}\|_2}$. Hence, the optimal value of the optimization problem (2.4) is $\phi_\infty(\mathbf{X})=\max\left\{\|\mathbf{x}_i\|_2:i\in[m]\right\}$ (notice that this result also holds for the case where $\mathbf{X}=\mathbf{O}_{n\times m}$), and the optimization problem (2.4) has an optimal solution $\mathbf{u}^*=\frac{\mathbf{x}_{j^*}}{\|\mathbf{x}_{j^*}\|_2}$, *i.e.*,

$$\frac{\mathbf{x}_{j^*}}{\|\mathbf{x}_{j^*}\|_2}\in\operatorname{argmax}\left\{\left\|\mathbf{X}^{\top}\mathbf{u}\right\|_{\infty}:\mathbf{u}\in\mathbb{S}^{n-1}\right\},$$

where $j^*\in\operatorname{argmax}\left\{\|\mathbf{x}_i\|_2:i\in[m]\right\}$.

(4) Let $\theta_p(\cdot):\mathbb{R}^{n\times m}\to\mathbb{R}_+$ be defined by

$$\theta_p(\mathbf{X}):=\max\left\{\|\mathbf{X}\mathbf{v}\|_2:\mathbf{v}\in\mathbb{R}^m\text{ such that }\|\mathbf{v}\|_q\leq 1\right\}. \tag{2.8}$$

Note that the map $\theta_p(\cdot):\mathbb{R}^{n\times m}\to\mathbb{R}_+$ is well-defined since the function $\mathbf{v}\in\mathbb{R}^m\mapsto\|\mathbf{X}\mathbf{v}\|_2$ is a continuous function and $\left\{\mathbf{v}\in\mathbb{R}^m:\|\mathbf{v}\|_q\leq 1\right\}$ is a compact subset of $\mathbb{R}^m$. Then we have

$$\begin{aligned}
\phi_p(\mathbf{X})&=\max\left\{\left\|\mathbf{X}^{\top}\mathbf{u}\right\|_p:\mathbf{u}\in\mathbb{S}^{n-1}\right\}\\
&\overset{\text{(c)}}{\leq}\max\left\{\max\left\{\left(\mathbf{X}^{\top}\mathbf{u}\right)^{\top}\mathbf{v}:\mathbf{v}\in\mathbb{R}^m\text{ such that }\|\mathbf{v}\|_q\leq 1\right\}:\mathbf{u}\in\mathbb{S}^{n-1}\right\}\\
&=\max\left\{\max\left\{\mathbf{u}^{\top}\mathbf{X}\mathbf{v}:\mathbf{v}\in\mathbb{R}^m\text{ such that }\|\mathbf{v}\|_q\leq 1\right\}:\mathbf{u}\in\mathbb{S}^{n-1}\right\}\\
&\overset{\text{(d)}}{\leq}\max\left\{\max\left\{\|\mathbf{u}\|_2\cdot\|\mathbf{X}\mathbf{v}\|_2:\mathbf{v}\in\mathbb{R}^m\text{ such that }\|\mathbf{v}\|_q\leq 1\right\}:\mathbf{u}\in\mathbb{S}^{n-1}\right\}\\
&=\max\left\{\cdot\|\mathbf{X}\mathbf{v}\|_2:\mathbf{v}\in\mathbb{R}^m\text{ such that }\|\mathbf{v}\|_q\leq 1\right\}\\
&=\theta_p(\mathbf{X}),
\end{aligned} \tag{2.9}$$

where the step (c) follows from the fact

$$\|\mathbf{x}\|_p=\max\left\{\mathbf{x}^{\top}\mathbf{y}:\mathbf{y}\in\mathbb{R}^n\text{ such that }\|\mathbf{y}\|_q\leq 1\right\} \tag{2.10}$$

for any $\mathbf{x} \in \mathbb{R}^n$ and $p, q \in [1, +\infty]$ satisfying the relation $\frac{1}{p} + \frac{1}{q} = 1$, and the step (d) holds by the Cauchy-Schwarz inequality.

On the other hand, let $\mathbf{v}^* \in \operatorname{argmax}\left\{\|\mathbf{X}\mathbf{v}\|_2 : \mathbf{v} \in \mathbb{R}^m \text{ such that } \|\mathbf{v}\|_q \leq 1\right\}$. Then,

$$\begin{aligned}
\theta_p(\mathbf{X}) &= \|\mathbf{X}\mathbf{v}^*\|_2 \\
&\overset{\text{(e)}}{=} \max\left\{\mathbf{u}^\top \mathbf{X}\mathbf{v}^* : \mathbf{u} \in \mathbb{S}^{n-1}\right\} \\
&= \max\left\{\left(\mathbf{X}^\top \mathbf{u}\right)^\top \mathbf{v}^* : \mathbf{u} \in \mathbb{S}^{n-1}\right\} \\
&\overset{\text{(f)}}{\leq} \max\left\{\left\|\mathbf{X}^\top \mathbf{u}\right\|_p \cdot \|\mathbf{v}^*\|_q : \mathbf{u} \in \mathbb{S}^{n-1}\right\} \\
&\overset{\text{(g)}}{\leq} \max\left\{\left\|\mathbf{X}^\top \mathbf{u}\right\|_p : \mathbf{u} \in \mathbb{S}^{n-1}\right\} \\
&= \phi_p(\mathbf{X}),
\end{aligned} \tag{2.11}$$

where the step (e) makes use of the fact (2.10), the step (f) holds due to the Hölder's inequality, and the step (g) follows from the fact $\|\mathbf{v}^*\|_q \leq 1$. By combining two results (2.9) and (2.11), we arrive at

$$\phi_p(\mathbf{X}) = \theta_p(\mathbf{X}) = \max\left\{\|\mathbf{X}\mathbf{v}\|_2 : \mathbf{v} \in \mathbb{R}^m \text{ such that } \|\mathbf{v}\|_q \leq 1\right\},$$

as desired. This completes the solution to Problem 2.

References

[1] Giuseppe C Calafiore and Laurent El Ghaoui. *Optimization models*. Cambridge university press, 2014.