MAS374 Optimization Theory
Homework #6

20150597 Jeonghwan Lee

Department of Mathematical Sciences, KAIST

November 17, 2021

Problem 1 (*Exercise 9.3* in [1])**.**

(1) Given any two distinct points $\mathbf{p} \neq \mathbf{q} \in \mathbb{R}^n$, we denote the closed line segment whose endpoints are $\mathbf{p}$ and $\mathbf{q}$ by

$$\mathcal{L}(\mathbf{p};\mathbf{q}) := \{(1-\theta)\,\mathbf{p} + \theta\mathbf{q} : \theta \in [0,1]\}.$$

Then,

$$\begin{aligned} D_* &:= (\text{the minimum distance from a point } \mathbf{a} \in \mathbb{R}^n \text{ to the line segment } \mathcal{L}(\mathbf{p};\mathbf{q})) \\ &= \min\{\|(1-\lambda)\,\mathbf{p} + \lambda\mathbf{q} - \mathbf{a}\|_2 : \lambda \in [0,1]\} \\ &= \min\{\|\lambda\,(\mathbf{q}-\mathbf{p}) + (\mathbf{p}-\mathbf{a})\|_2 : \lambda \in [0,1]\}. \end{aligned} \tag{1.1}$$

So it suffices to choose $\mathbf{c}, \mathbf{d} \in \mathbb{R}^n$ by $\mathbf{c} := \mathbf{q} - \mathbf{p}$ and $\mathbf{d} := \mathbf{p} - \mathbf{a}$. Note that

$$\begin{aligned} D_* &= \min\{\|\lambda\,\{(\mathbf{q}-\mathbf{a}) - (\mathbf{p}-\mathbf{a})\} + (\mathbf{p}-\mathbf{a})\|_2 : \lambda \in [0,1]\} \\ &= (\text{the minimum distance from a point } \mathbf{0} \in \mathbb{R}^n \text{ to the line segment } \mathcal{L}(\mathbf{p}-\mathbf{a};\mathbf{q}-\mathbf{a})). \end{aligned}$$

In words, the current scenario is completely equivalent with the circumstance occurred by translating three points $\mathbf{p}, \mathbf{q}, \mathbf{a} \in \mathbb{R}^n$ by $-\mathbf{a} \in \mathbb{R}^n$. So one can always assume that $\mathbf{a} = \mathbf{0}$ without loss of generality!

(2) Hereafter, we assume that $\mathbf{a} = \mathbf{0}$ without loss of generality. Then, one can interpret the optimization problem (1.1) in the following equivalent way:

$$D_*^2 = \min\{f_0(\theta) : \theta \in [0,1]\}, \tag{1.2}$$

where $f_0(\theta) := \|\theta\,(\mathbf{q}-\mathbf{p}) + \mathbf{p}\|_2^2$ for $\theta \in \mathbb{R}$. Doing some straightforward algebra, we arrive at

$$f_0(\theta) = \|\mathbf{p}-\mathbf{q}\|_2^2 \left\{\theta - \frac{\mathbf{p}^\top(\mathbf{p}-\mathbf{q})}{\|\mathbf{p}-\mathbf{q}\|_2^2}\right\}^2 + \mathbf{q}^\top\mathbf{q} - \frac{\left\{\mathbf{q}^\top(\mathbf{p}-\mathbf{q})\right\}^2}{\|\mathbf{p}-\mathbf{q}\|_2^2},\ \forall \theta \in \mathbb{R}.$$

Then it's clear that

$$\operatorname{argmin}\{f_0(\theta) : \theta \in \mathbb{R}\} = \{\theta^*\} = \left\{\frac{\mathbf{p}^\top(\mathbf{p}-\mathbf{q})}{\|\mathbf{p}-\mathbf{q}\|_2^2}\right\}.$$

So in order to solve the optimization problem (1.2), we consider the following three cases:

(i) $\theta^* > 1$: One can easily see that $\theta^* > 1$ if and only if $\mathbf{p}^\top \mathbf{q} > \mathbf{q}^\top \mathbf{q}$ and

$$\operatorname{argmin}\{f_0(\theta) : \theta \in [0,1]\} = \{1\}. \tag{1.3}$$

Therefore, we have

$$D_*^2 = f_0(1) = \mathbf{q}^\top \mathbf{q}, \tag{1.4}$$

provided that $\mathbf{p}^\top \mathbf{q} > \mathbf{q}^\top \mathbf{q}$;

(ii) $\theta^* < 0$: One can easily observe that $\theta^* < 0$ if and only if $\mathbf{p}^\top \mathbf{q} > \mathbf{p}^\top \mathbf{p}$ and

$$\operatorname{argmin}\{f_0(\theta) : \theta \in [0,1]\} = \{0\}. \tag{1.5}$$

Therefore, we get

$$D_*^2 = f_0(0) = \mathbf{p}^\top \mathbf{p}, \tag{1.6}$$

provided that $\mathbf{p}^\top \mathbf{q} > \mathbf{p}^\top \mathbf{p}$;

(iii) $0 \le \theta^* \le 1$: One can easily recognize that $0 \le \theta^* \le 1$ if and only if $\mathbf{p}^\top \mathbf{q} \le \min\left\{\mathbf{p}^\top \mathbf{p}, \mathbf{q}^\top \mathbf{q}\right\}$ and

$$\operatorname{argmin}\{f_0(\theta) : \theta \in [0,1]\} = \{\theta^*\} = \left\{\frac{\mathbf{p}^\top(\mathbf{p}-\mathbf{q})}{\|\mathbf{p}-\mathbf{q}\|_2^2}\right\}. \tag{1.7}$$

Therefore, we get

$$D_*^2 = f_0(\theta^*) = \mathbf{q}^\top \mathbf{q} - \frac{\left\{\mathbf{q}^\top(\mathbf{p}-\mathbf{q})\right\}^2}{\|\mathbf{p}-\mathbf{q}\|_2^2}, \tag{1.8}$$

provided that $\mathbf{p}^\top \mathbf{q} \le \min\left\{\mathbf{p}^\top \mathbf{p}, \mathbf{q}^\top \mathbf{q}\right\}$.

Taking three pieces (1.4), (1.6), and (1.8) collectively, one has

$$D_*^2 = \begin{cases} \mathbf{q}^\top \mathbf{q} - \frac{\left\{\mathbf{q}^\top(\mathbf{p}-\mathbf{q})\right\}^2}{\|\mathbf{p}-\mathbf{q}\|_2^2} & \text{if } \mathbf{p}^\top \mathbf{q} \le \min\left\{\mathbf{p}^\top \mathbf{p}, \mathbf{q}^\top \mathbf{q}\right\}; \\ \mathbf{q}^\top \mathbf{q} & \text{if } \mathbf{p}^\top \mathbf{q} > \mathbf{q}^\top \mathbf{q}; \\ \mathbf{p}^\top \mathbf{p} & \text{if } \mathbf{p}^\top \mathbf{q} > \mathbf{p}^\top \mathbf{p}. \end{cases} \tag{1.9}$$

(3) Lastly, we would like to give you some geometric interpretations of the result in part (2).

Case #1. $\theta^* > 1$**:** This case corresponds to the case where $\mathbf{p}^\top \mathbf{q} > \mathbf{q}^\top \mathbf{q}$. For this case, the closed point on the line segment $\mathcal{L}(\mathbf{p};\mathbf{q})$ from the minimum-distance point of the affine line $\{(1-\theta)\mathbf{p} + \theta\mathbf{q} : \theta \in \mathbb{R}\}$ from $\mathbf{0} \in \mathbb{R}^n$, $(1-\theta^*)\mathbf{p} + \theta^*\mathbf{q}$, is precisely $\mathbf{q}$. By considering the Pythagorean theorem, the closest point on the line segment $\mathcal{L}(\mathbf{p};\mathbf{q})$ from $\mathbf{0} \in \mathbb{R}^n$ becomes $\mathbf{q}$. See Figure 1 for detailed visualization.

Case #2. $\theta^* < 0$**:** This case corresponds to the case where $\mathbf{p}^\top \mathbf{q} > \mathbf{p}^\top \mathbf{p}$. For this case, the closed point on the line segment $\mathcal{L}(\mathbf{p};\mathbf{q})$ from the minimum-distance point of the affine line $\{(1-\theta)\mathbf{p} + \theta\mathbf{q} : \theta \in \mathbb{R}\}$ from $\mathbf{0} \in \mathbb{R}^n$, $(1-\theta^*)\mathbf{p} + \theta^*\mathbf{q}$, is precisely $\mathbf{p}$. By considering the Pythagorean theorem, the closest point on the line segment $\mathcal{L}(\mathbf{p};\mathbf{q})$ from $\mathbf{0} \in \mathbb{R}^n$ becomes $\mathbf{p}$. See Figure 2 for detailed visualization.

Case #3. $0 \le \theta^* \le 1$**:** This case corresponds to the case where $\mathbf{p}^\top \mathbf{q} \le \min\left\{\mathbf{p}^\top \mathbf{p}, \mathbf{q}^\top \mathbf{q}\right\}$. For this case, the minimum-distance point of the affine line $\{(1-\theta)\mathbf{p} + \theta\mathbf{q} : \theta \in \mathbb{R}\}$ from $\mathbf{0} \in \mathbb{R}^n$, $(1-\theta^*)\mathbf{p} + \theta^*\mathbf{q}$, lies on the line segment $\mathcal{L}(\mathbf{p};\mathbf{q})$. So the closest point on the line segment $\mathcal{L}(\mathbf{p};\mathbf{q})$ from $\mathbf{0} \in \mathbb{R}^n$ is $(1-\theta^*)\mathbf{p} + \theta^*\mathbf{q}$. See Figure 3 for detailed visualization.

This completes the geometric interpretation of the result in the part (2).

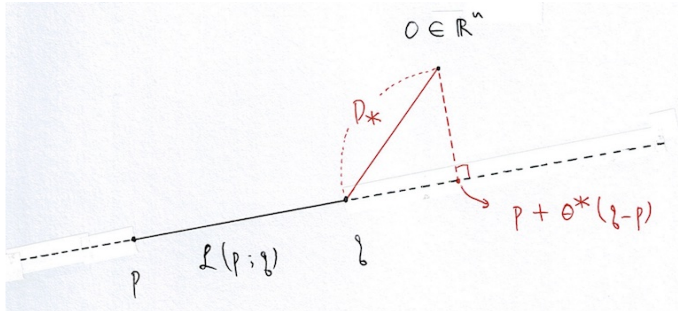

Figure 1: The case for which $\theta^* > 1 \Leftrightarrow \mathbf{p}^\top \mathbf{q} > \mathbf{q}^\top \mathbf{q}$.

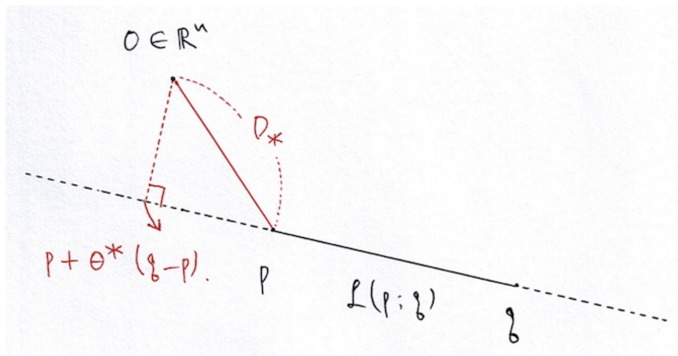

Figure 2: The case for which $\theta^* < 0 \Leftrightarrow \mathbf{p}^\top \mathbf{q} > \mathbf{p}^\top \mathbf{p}$.

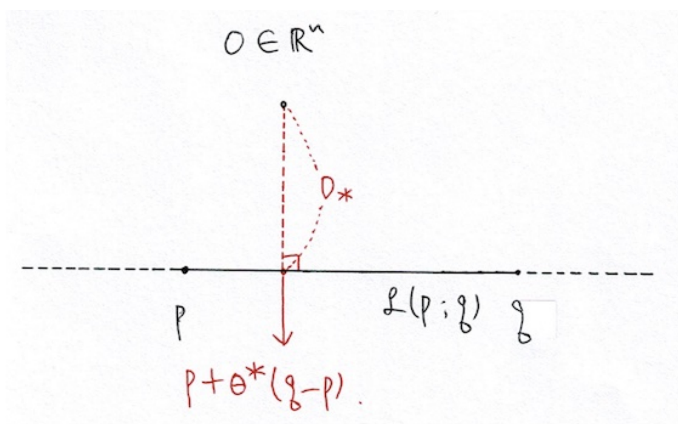

Figure 3: The case for which $0 \leq \theta^* \leq 1 \Leftrightarrow \mathbf{p}^\top \mathbf{q} \leq \min\left\{\mathbf{p}^\top \mathbf{p}, \mathbf{q}^\top \mathbf{q}\right\}$.

Problem 2 (*Exercise 9.9* in [1]).

(1) Let $\mathbf{A} := \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_n^\top \end{bmatrix} \in \mathbb{R}_{++}^{n \times n}$ and $f(\cdot) : \mathcal{S} := \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \succeq \mathbf{0} \text{ and } \mathbf{1}_n^\top \mathbf{x} = 1\right\} \to \mathbb{R}_{++}$, where

$$f(\mathbf{x}) := \min\left\{\frac{\mathbf{a}_i^\top \mathbf{x}}{x_i} : i \in [n]\right\},$$

where $\mathbf{1}_n \in \mathbb{R}^n$ denotes the n-dimensional all-one vector. Here, we adopt the convention that $\frac{\mathbf{a}_i^\top \mathbf{x}}{x_i} := +\infty$ if $x_i = 0$. Since $\mathcal{S} \subseteq \mathbb{R}_+^n$ and $f(\mathbf{x}) > 0$ for every $\mathbf{x} \in \mathcal{S}$, it's evident that $f(\mathbf{x}) \cdot \mathbf{x} \succeq \mathbf{0}$, *i.e.*, $f(\mathbf{x}) \cdot \mathbf{x} \in \mathbb{R}_+^n$. In order to establish $\mathbf{A}\mathbf{x} \succeq f(\mathbf{x}) \cdot \mathbf{x}$, it suffices to show that

$$[\mathbf{A}\mathbf{x}]_i - f(\mathbf{x}) \cdot x_i \geq 0, \ \forall i \in [n]. \tag{2.1}$$

Indeed, this result holds since

$$[\mathbf{A}\mathbf{x}]_i - f(\mathbf{x}) \cdot x_i = \mathbf{a}_i^\top \mathbf{x} - f(\mathbf{x}) \cdot x_i = \begin{cases} \mathbf{a}_i^\top \mathbf{x} & \text{if } x_i = 0; \\ x_i \left\{ \frac{\mathbf{a}_i^\top \mathbf{x}}{x_i} - f(\mathbf{x}) \right\} & \text{otherwise.} \end{cases} \geq 0$$

for every $i \in [n]$ and $\mathbf{x} \in \mathcal{S}$. This completes the proof of

$$\mathbf{A}\mathbf{x} \succeq f(\mathbf{x}) \cdot \mathbf{x} \succeq \mathbf{0}, \ \forall \left(\mathbf{x}, \mathbf{A}\right) \in \mathcal{S} \times \mathbb{R}_{++}^{n \times n}. \tag{2.2}$$

(2) Let $\mathbf{w} \in \mathbb{R}_{++}^n$ be a left eigenvector of $\mathbf{A}$ corresponding to a dominant eigenvalue $\lambda = \rho(\mathbf{A}) > 0$, *i.e.*, $\mathbf{w}^\top \mathbf{A} = \lambda \mathbf{w}^\top$. Since $\mathbf{A}\mathbf{x} - f(\mathbf{x}) \cdot \mathbf{x} \in \mathbb{R}_+^n$ for all $\mathbf{x} \in \mathcal{S}$, one has

$$\mathbf{w}^\top \left\{ \mathbf{A}\mathbf{x} - f(\mathbf{x}) \cdot \mathbf{x} \right\} = \left(\mathbf{w}^\top \mathbf{A} \right) \mathbf{x} - f(\mathbf{x}) \left(\mathbf{w}^\top \mathbf{x} \right) = \left\{ \lambda - f(\mathbf{x}) \right\} \left(\mathbf{w}^\top \mathbf{x} \right) \geq 0$$

for every $\mathbf{x} \in \mathcal{S}$. Since $\mathbf{w}^\top \mathbf{x} > 0$ for every $\mathbf{x} \in \mathcal{S}$, we obtain $f(\mathbf{x}) \leq \lambda$ for all $\mathbf{x} \in \mathcal{S}$, *i.e.*,

$$\sup \left\{ f(\mathbf{x}) : \mathbf{x} \in \mathcal{S} \right\} \leq \lambda. \tag{2.3}$$

(3) Let $\mathbf{v} \in \mathbb{R}_{++}^n$ be a right eigenvector of $\mathbf{A}$ corresponding to a dominant eigenvalue $\lambda = \rho(\mathbf{A}) > 0$, *i.e.*, $\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$. Then it's clear that $\mathbf{a}_i^\top \mathbf{v} = \lambda v_i$ for all $i \in [n]$. Letting $\tilde{\mathbf{v}} := \frac{\mathbf{v}}{\|\mathbf{v}\|_1}$, it's clear that $\tilde{\mathbf{v}} \in \mathcal{S}$ and

$$f\left(\tilde{\mathbf{v}}\right) = \max \left\{ \frac{\mathbf{a}_i^\top \tilde{\mathbf{v}}}{\tilde{v}_i} : i \in [n] \right\} = \max \left\{ \frac{\mathbf{a}_i^\top \mathbf{v}}{v_i} : i \in [n] \right\} = \lambda.$$

So we arrive at

$$\sup \left\{ f(\mathbf{x}) : \mathbf{x} \in \mathcal{S} \right\} \overset{\text{(a)}}{\leq} \lambda = f\left(\tilde{\mathbf{v}}\right) \leq \sup \left\{ f(\mathbf{x}) : \mathbf{x} \in \mathcal{S} \right\},$$

where the step (a) follows from the inequality (2.3). This yields

$$\lambda = \max \left\{ f(\mathbf{x}) : \mathbf{x} \in \mathcal{S} \right\} \quad \text{and} \quad \tilde{\mathbf{v}} \in \operatorname{argmax} \left\{ f(\mathbf{x}) : \mathbf{x} \in \mathcal{S} \right\},$$

as desired.

References

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