# MAS374 Optimization theory
# Homework #1

20150597 Jeonghwan Lee

Department of Mathematical Sciences, KAIST

September 14, 2021

**Problem 1** (*Exercise 2.2* in [1]).

(1) Set $\mathbf{x}_0 := (1, 0, 0)$. Then it's clear that

$$\mathbf{x}_0 \in \mathcal{P} := \left\{ \mathbf{x} \in \mathbb{R}^3 : x_1 + 2x_2 + 3x_3 = 1 \right\} \subseteq \mathbb{R}^3.$$

In order to show that $\mathcal{P}$ is an affine space of dimension 2, it suffices to prove that $\mathcal{P} - \mathbf{x}_0 := \left\{ \mathbf{x} - \mathbf{x}_0 : \mathbf{x} \in \mathbb{R}^3 \right\}$ is a linear subspace of $\mathbb{R}^3$ of dimension 2. Note that $\mathcal{P} - \mathbf{x}_0 = \left\{ \mathbf{x} \in \mathbb{R}^3 : x_1 + 2x_2 + 3x_3 = 0 \right\}$. We define a function $f : \mathbb{R}^3 \to \mathbb{R}$ by

$$f(\mathbf{x}) = f(x_1, x_2, x_3) := x_1 + 2x_2 + 3x_3 = (1, 2, 3)^\top \mathbf{x}.$$

Then it's evident that $f : \mathbb{R}^3 \to \mathbb{R}$ is a non-zero linear functional on $\mathbb{R}^3$, *i.e.*, a linear functional on $\mathbb{R}^3$ with $\mathsf{im}(f) = \mathbb{R}$. From $\mathcal{P} - \mathbf{x}_0 = \mathsf{ker}(f)$, we obtain

$$\begin{aligned} \dim_{\mathbb{R}} (\mathcal{P} - \mathbf{x}_0) &= \dim_{\mathbb{R}} (\mathsf{ker}(f)) \\ &\overset{\text{(a)}}{=} \dim_{\mathbb{R}} \left( \mathbb{R}^3 \right) - \dim_{\mathbb{R}} (\mathsf{im}(f)) \\ &= 2, \end{aligned}$$

as desired, where the step (a) follows from the rank-nullity theorem. This confirms that $\mathcal{P}$ is an affine space over $\mathbb{R}$ of dimension 2.

I would like to provide some additional remark. Suppose $\mathbf{x} = (x_1, x_2, x_3) \in \mathcal{P} - \mathbf{x}_0$. Then one can obtain the relation $x_1 = -2x_2 - 3x_3$, thereby we have

$$\mathbf{x} = (x_1, x_2, x_3) = (-2x_2 - 3x_3, x_2, x_3) = x_2 (-2, 1, 0) + x_3 (-3, 0, 1).$$

So it can be easily seen that $\{(-2, 1, 0), (-3, 0, 1)\}$ forms a basis for the linear subspace $\mathcal{P} - \mathbf{x}_0$ of $\mathbb{R}^3$, and this gives the representation

$$\mathcal{P} = (1, 0, 0) + \mathsf{span} (\{(-2, 1, 0), (-3, 0, 1)\}).$$

(2) We first recall the following generalized result regarding the distance of a point from a hyperplane in the $n$-dimensional Euclidean space $\mathbb{R}^n$:

**Lemma 1.** *Let* $\mathbf{x}_0 \in \mathbb{R}^n$ *and*

$$\mathcal{H}(\mathbf{a}; b) := \left\{ \mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = b \right\},$$

*where* $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ *and* $b \in \mathbb{R}$. *Then, the minimum Euclidean distance from* $\mathbf{x}_0$ *to the hyperplane* $\mathcal{H}(\mathbf{a}; b)$ *is given by*

$$\mathsf{dist}\left(\mathbf{x}_0, \mathcal{H}(\mathbf{a}; b)\right) := \inf\left\{ \|\mathbf{x}_0 - \mathbf{x}\|_2 : \mathbf{x} \in \mathcal{H}(\mathbf{a}; b) \right\} = \frac{\left| b - \mathbf{a}^\top \mathbf{x}_0 \right|}{\|\mathbf{a}\|_2}, \tag{1}$$

*and the point that achieves the minimum distance is*

$$\operatorname{argmin}\left\{ \|\mathbf{x}_0 - \mathbf{x}\|_2 : \mathbf{x} \in \mathcal{H}(\mathbf{a}; b) \right\} = \left\{ \mathbf{x}_0 + \frac{b - \mathbf{a}^\top \mathbf{x}_0}{\|\mathbf{a}\|_2^2} \cdot \mathbf{a} \right\}. \tag{2}$$

*Proof of Lemma 1.* Let us consider the straight line $\mathcal{L} := \{\mathbf{x}_0 + t\mathbf{a} : t \in \mathbb{R}\}$ in $\mathbb{R}^n$, and its intersection with the hyperplane $\mathcal{H}(\mathbf{a}; b)$:

$$\mathcal{L} \cap \mathcal{H}(\mathbf{a}; b) = \{\mathbf{x}^*\},$$

where $\mathbf{x}^* = \mathbf{x}_0 + t^*\mathbf{a}$ for some $t^* \in \mathbb{R}$. Here, we note that the scalar $t^* \in \mathbb{R}$ can be computed explicitly since $\mathbf{x}^* = \mathbf{x}_0 + t^*\mathbf{a} \in \mathcal{H}(\mathbf{a}; b)$: one can observe that

$$b = \mathbf{a}^\top \mathbf{x}^* = \mathbf{a}^\top \left(\mathbf{x}_0 + t^*\mathbf{a}\right) = \mathbf{a}^\top \mathbf{x}_0 + t^* \|\mathbf{a}\|_2^2,$$

and this yields

$$t^* = \frac{b - \mathbf{a}^\top \mathbf{x}_0}{\|\mathbf{a}\|_2^2}.$$

Now, we claim that $\mathbf{x}^* \in \operatorname{argmin}\left\{ \|\mathbf{x}_0 - \mathbf{x}\|_2 : \mathbf{x} \in \mathcal{H}(\mathbf{a}; b) \right\}$. Choose any $\mathbf{y} \in \mathcal{H}(\mathbf{a}; b)$. From

$$\mathbf{a}^\top \mathbf{y} = b = \mathbf{a}^\top \mathbf{x}^*,$$

we see that

$$0 = \mathbf{a}^\top \left(\mathbf{x}^* - \mathbf{y}\right) \quad \Rightarrow \quad \mathbf{x}^* - \mathbf{y} \in \left(\{\mathbf{a}\}\right)^\perp. \tag{3}$$

Since $\mathbf{x}_0 - \mathbf{x}^* = -t^*\mathbf{a} \in \mathsf{span}\left(\{\mathbf{a}\}\right)$, it follows from (3) that

$$\langle \mathbf{x}_0 - \mathbf{x}^*, \mathbf{x}^* - \mathbf{y} \rangle = \left(\mathbf{x}_0 - \mathbf{x}^*\right)^\top \left(\mathbf{x}^* - \mathbf{y}\right) = 0. \tag{4}$$

Hence, it can be shown that

$$\begin{aligned} \|\mathbf{x}_0 - \mathbf{y}\|_2^2 &= \|(\mathbf{x}_0 - \mathbf{x}^*) + (\mathbf{x}^* - \mathbf{y})\|_2^2 \\ &= \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 + \|\mathbf{x}^* - \mathbf{y}\|_2^2 + 2\langle \mathbf{x}_0 - \mathbf{x}^*, \mathbf{x}^* - \mathbf{y} \rangle \\ &\overset{\text{(b)}}{=} \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 + \|\mathbf{x}^* - \mathbf{y}\|_2^2 \\ &\geq \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2, \end{aligned} \tag{5}$$

where the step (b) makes use of the equation (4). Note that the equality in (5) holds if and only if $\mathbf{y} = \mathbf{x}^*$. Therefore we may conclude that

$$\mathbf{x}^* = \mathbf{x}_0 + t^*\mathbf{a} \in \operatorname{argmin}\left\{ \|\mathbf{x}_0 - \mathbf{x}\|_2 : \mathbf{x} \in \mathcal{H}(\mathbf{a}; b) \right\},$$

and as a result, we obtain

$$\begin{aligned}\mathsf{dist}\left(\mathbf{x}_0, \mathcal{H}(\mathbf{a}; b)\right) &= \inf\left\{\|\mathbf{x}_0 - \mathbf{x}\|_2 : \mathbf{x} \in \mathcal{H}(\mathbf{a}; b)\right\}\\ &= \|\mathbf{x}_0 - \mathbf{x}^*\|_2\\ &= |t^*|\,\|\mathbf{a}\|_2\\ &= \frac{\left|b - \mathbf{a}^\top \mathbf{x}_0\right|}{\|\mathbf{a}\|_2},\end{aligned}$$

as desired.

$\square$

Finally, we compute the minimum Euclidean distance from $\mathbf{0} \in \mathbb{R}^3$ to $\mathcal{P}$ and the point that attains the minimum distance by applying Lemma 1 for $n = 3$, $\mathbf{x}_0 = \mathbf{0} \in \mathbb{R}^3$, $\mathbf{a} = (1, 2, 3) \in \mathbb{R}^3$, and $b = 1 \in \mathbb{R}$: the minimum Euclidean distance from $\mathbf{0} \in \mathbb{R}^3$ to $\mathcal{P}$ is given by

$$\mathsf{dist}\left(\mathbf{0}, \mathcal{P}\right) = \frac{\left|b - \mathbf{a}^\top \mathbf{x}_0\right|}{\|\mathbf{a}\|_2} = \frac{1}{\sqrt{14}},$$

and the point that achieves the minimum distance is

$$\mathbf{x}^* = \mathbf{x}_0 + \frac{b - \mathbf{a}^\top \mathbf{x}_0}{\|\mathbf{a}\|_2^2} \cdot \mathbf{a} = \left(\frac{1}{14}, \frac{1}{7}, \frac{3}{14}\right).$$

**Problem 2** (*Exercise 2.7* in [1]).

We will prove that for any $p, q \in [1, +\infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$, it holds that

$$\left|\mathbf{x}^\top \mathbf{y}\right| \overset{\text{(a)}}{\leq} \sum_{k=1}^{n} |x_k y_k| \overset{\text{(b)}}{\leq} \|\mathbf{x}\|_p \|\mathbf{y}\|_q\,,\ \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n. \tag{6}$$

The cases for which either $\mathbf{x} = \mathbf{0}$ or $\mathbf{y} = \mathbf{0}$ are trivial. So from now on, we may assume that $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$. The first inequality (a) immediately follows from the triangle inequality.

*Case #1.* $(p, q) = (1, +\infty)$: The second inequality (b) holds since

$$\sum_{k=1}^{n} |x_k y_k| \leq \sum_{k=1}^{n} |x_k|\,\|\mathbf{y}\|_\infty = \|\mathbf{x}\|_1 \|\mathbf{y}\|_\infty\,.$$

*Case #2.* $(p, q) = (+\infty, 1)$: The second inequality (b) follows because

$$\sum_{k=1}^{n} |x_k y_k| \leq \sum_{k=1}^{n} \|\mathbf{x}\|_\infty |y_k| = \|\mathbf{x}\|_\infty \|\mathbf{y}\|_1\,.$$

*Case #3.* $p, q \in (1, +\infty)$: Consider the normalized vectors $\mathbf{u} := \frac{\mathbf{x}}{\|\mathbf{x}\|_p}$ and $\mathbf{v} := \frac{\mathbf{y}}{\|\mathbf{y}\|_q}$. Then we have

$$\|\mathbf{u}\|_p = \|\mathbf{v}\|_q = 1, \tag{7}$$

and

$$\begin{aligned}
\sum_{k=1}^{n} |x_k y_k| &= \|\mathbf{x}\|_p \|\mathbf{y}\|_q \left( \sum_{k=1}^{n} |u_k v_k| \right) \\
&\overset{\text{(c)}}{\leq} \|\mathbf{x}\|_p \|\mathbf{y}\|_q \left\{ \sum_{k=1}^{n} \left( \frac{1}{p} |u_k|^p + \frac{1}{q} |v_k|^q \right) \right\} \\
&= \|\mathbf{x}\|_p \|\mathbf{y}\|_q \left( \frac{1}{p} \|\mathbf{u}\|_p + \frac{1}{q} \|\mathbf{v}\|_q \right) \\
&\overset{\text{(d)}}{=} \|\mathbf{x}\|_p \|\mathbf{y}\|_q ,
\end{aligned} \tag{8}$$

where the step (c) follows from the Young's inequality: if $p, q \in (1, +\infty)$ and $a, b \in [0, +\infty)$, then

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}, \tag{9}$$

and the step (d) is due to the fact (7) together with $\frac{1}{p} + \frac{1}{q} = 1$. We finish the proof by establishing the Young's inequality (9). From the concavity of the function $x \in (0, +\infty) \mapsto \log x \in \mathbb{R}$, we obtain

$$\log(ab) = \frac{1}{p} \log(a^p) + \frac{1}{q} \log(b^q) \leq \log\left( \frac{a^p}{p} + \frac{b^q}{q} \right),$$

and the non-decreasing property of the function $x \in (0, +\infty) \mapsto \log x \in \mathbb{R}$ yields the desired result.

**Problem 3** (*Exercise 3.1* in [1]).

(i) We first recall that for each $(i, j) \in [k] \times [n]$,

$$[\mathcal{J}_h(\mathbf{x})]_{ij} = D_j h_i(\mathbf{x}) := \frac{\partial h_i}{\partial x_j}(\mathbf{x}), \tag{10}$$

where $\mathcal{J}_h(\mathbf{x}) \in \mathbb{R}^{k \times n}$ is the Jacobian matrix of $h : \mathbb{R}^n \times \mathbb{R}^k$ at $\mathbf{x} \in \mathbb{R}^n$. Since

$$h_i(\mathbf{x}) = (f_i \circ g)(\mathbf{x}) = f_i(g_1(\mathbf{x}), g_2(\mathbf{x}), \cdots, g_m(\mathbf{x})), \ \forall \mathbf{x} \in \mathbb{R}^n,$$

the chain rule gives

$$\begin{aligned}
\frac{\partial h_i}{\partial x_j}(\mathbf{x}) &= \sum_{k=1}^{m} D_k f_i(g_1(\mathbf{x}), g_2(\mathbf{x}), \cdots, g_m(\mathbf{x})) \cdot D_j g_k(\mathbf{x}) \\
&= \sum_{k=1}^{m} [\mathcal{J}_f(g(\mathbf{x}))]_{ik} [\mathcal{J}_g(\mathbf{x})]_{kj} \\
&= [\mathcal{J}_f(g(\mathbf{x})) \cdot \mathcal{J}_g(\mathbf{x})]_{ij} .
\end{aligned} \tag{11}$$

Taking two pieces (10) and (11) collectively yields the desired result, known as the *general chain rule*:

$$\mathcal{J}_h(\mathbf{x}) = \mathcal{J}_f(g(\mathbf{x})) \cdot \mathcal{J}_g(\mathbf{x}), \ \forall \mathbf{x} \in \mathbb{R}^n. \tag{12}$$

(ii) By (i), it suffices to show that

$$\mathcal{J}_g(\mathbf{x}) = \mathbf{A} = [A_{ij}]_{(i,j) \in [m] \times [n]} \in \mathbb{R}^{m \times n}, \ \forall \mathbf{x} \in \mathbb{R}^n. \tag{13}$$

From

$$\begin{aligned}
g(\mathbf{x}) &= (g_1(\mathbf{x}), g_2(\mathbf{x}), \cdots, g_m(\mathbf{x})) \\
&= \mathbf{A}\mathbf{x} + \mathbf{b} \\
&= \left( \sum_{k=1}^{n} A_{1k}x_k + b_1, \sum_{k=1}^{n} A_{2k}x_k + b_2, \cdots, \sum_{k=1}^{n} A_{mk}x_k + b_m \right),
\end{aligned}$$

we have for every $(i, j) \in [m] \times [n]$,

$$[\mathcal{J}_g(\mathbf{x})]_{ij} = D_j g_i(\mathbf{x}) = \frac{\partial}{\partial x_j} \left( \sum_{k=1}^{n} A_{ik}x_k + b_i \right) = A_{ij},$$

which implies (13) as desired.

(iii) For any real-valued differentiable function $\varphi : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$,

$$\nabla\varphi(\mathbf{x}) = \mathcal{J}_\varphi(\mathbf{x})^\top = \begin{bmatrix} D_1\varphi(\mathbf{x}) & D_2\varphi(\mathbf{x}) & \cdots & D_n\varphi(\mathbf{x}) \end{bmatrix}^\top \in \mathbb{R}^{n\times 1}.$$

Consider the affine function $g : \mathbb{R}^n \to \mathbb{R}^m$ defined by $g(\mathbf{x}) := \mathbf{A}\mathbf{x} + \mathbf{b}$, where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^{m\times 1}$, and a differentiable function $f : \mathbb{R}^m \to \mathbb{R}$. Then by (ii), we have

$$\nabla h(\mathbf{x}) = \mathcal{J}_h(\mathbf{x})^\top = \{\mathcal{J}_f(g(\mathbf{x}) \cdot \mathbf{A}\}^\top = \mathbf{A}^\top \mathcal{J}_f(g(\mathbf{x})^\top = \mathbf{A}^\top \nabla f\left(g(\mathbf{x})\right), \ \forall \mathbf{x} \in \mathbb{R}^n. \tag{14}$$

Finally, we evaluate the Hessian $\nabla^2 h(\mathbf{x}) = \left[\frac{\partial^2 h}{\partial x_j \partial x_i}(\mathbf{x})\right]_{(i,j)\in[n]\times[n]} \in \mathbb{R}^{n\times n}$ of $h : \mathbb{R}^n \to \mathbb{R}$. Hereafter, we assume that $f : \mathbb{R}^m \to \mathbb{R}$ is a twice differentiable function. The $(i, j)$-th entry of the Hessian $\nabla^2 h(\mathbf{x})$ of the function $h : \mathbb{R}^n \to \mathbb{R}$ at $\mathbf{x} \in \mathbb{R}^n$ is given by

$$\begin{aligned}
\frac{\partial^2 h}{\partial x_j \partial x_i}(\mathbf{x}) &= \frac{\partial}{\partial x_j} \left( \frac{\partial h}{\partial x_i}(\mathbf{x}) \right) \\
&\overset{\text{(a)}}{=} \frac{\partial}{\partial x_j} \left( \sum_{k=1}^{m} A_{ki} \left(D_k f\right)(g(\mathbf{x})) \right) \\
&= \sum_{k=1}^{m} A_{ki} \cdot \frac{\partial}{\partial x_j} \left(D_k f\right)(g(\mathbf{x})) \\
&\overset{\text{(b)}}{=} \sum_{k=1}^{m} A_{ki} \cdot \left\{ \sum_{l=1}^{m} \left(D_l D_k f\right)(g(\mathbf{x})) \cdot D_j g_l(\mathbf{x}) \right\} \\
&\overset{\text{(c)}}{=} \sum_{k=1}^{m} \sum_{l=1}^{m} A_{ki} \left[\left(\nabla^2 f\right)(g(\mathbf{x}))\right]_{kl} A_{lj} \\
&= \left[\mathbf{A}^\top \left(\nabla^2 f\right)(g(\mathbf{x}))\mathbf{A}\right]_{ij},
\end{aligned} \tag{15}$$

where the steps (a)–(c) hold due to the following reasons:

(a) the equation (14);

(b) the general chain rule (i);

(c) the equation (13).

The equation (15) completes the proof of the fact $\nabla^2 h(\mathbf{x}) = \mathbf{A}^\top \left(\nabla^2 f\right)(g(\mathbf{x}))\mathbf{A}$, $\forall \mathbf{x} \in \mathbb{R}^n$.

**Problem 4** (*Exercise 3.7* in [1]).

(i) It's clear from the definition of null-space that

$$\mathcal{N}(\mathbf{A}) \leq \mathcal{N}\left(\mathbf{A}^\top \mathbf{A}\right). \tag{16}$$

Take any $\mathbf{x} \in \mathcal{N}\left(\mathbf{A}^\top \mathbf{A}\right)$. Then we have

$$0 = \mathbf{x}^\top \cdot \mathbf{0} = \mathbf{x}^\top \left(\mathbf{A}^\top \mathbf{A}\mathbf{x}\right) = (\mathbf{A}\mathbf{x})^\top (\mathbf{A}\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|_2^2,$$

and this implies $\mathbf{A}\mathbf{x} = \mathbf{0} \in \mathbb{R}^m$. Combining this conclusion together with the fact (16), we arrive at

$$\mathcal{N}(\mathbf{A}) = \mathcal{N}\left(\mathbf{A}^\top \mathbf{A}\right). \tag{17}$$

(ii) To begin with, one can recognize that for any $\mathbf{M} \in \mathbb{R}^{m \times n}$,

$$\begin{aligned} \mathcal{R}\left(\mathbf{M}^\top\right)^\perp &= \left\{\mathbf{x} \in \mathbb{R}^n : \left\langle \mathbf{x}, \mathbf{M}^\top \mathbf{y}\right\rangle = 0,\ \forall \mathbf{y} \in \mathbb{R}^m\right\} \\ &= \left\{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{M}\mathbf{x}, \mathbf{y}\rangle = 0,\ \forall \mathbf{y} \in \mathbb{R}^m\right\} \\ &= \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{M}\mathbf{x} = \mathbf{0}\right\} \\ &= \mathcal{N}(\mathbf{M}). \end{aligned} \tag{18}$$

Therefore, we find that

$$\mathcal{R}\left(\mathbf{A}^\top\right) \overset{\text{(a)}}{=} \mathcal{N}(\mathbf{A})^\perp \overset{\text{(b)}}{=} \mathcal{N}\left(\mathbf{A}^\top \mathbf{A}\right)^\perp \overset{\text{(c)}}{=} \left\{\mathcal{R}\left(\left(\mathbf{A}^\top \mathbf{A}\right)^\top\right)^\perp\right\}^\perp = \mathcal{R}\left(\mathbf{A}^\top \mathbf{A}\right),$$

where the step (a) and (c) follow from the fact (18), and the step (b) is owing to the fact (17). This completes the proof of desired results.

## References

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