MAS374 Optimization theory
Homework #2

20150597 Jeonghwan Lee

Department of Mathematical Sciences, KAIST

September 29, 2021

Problem 1 (*Exercise 4.4* in [1]).

(1) Given any linear subspace $\mathcal{S}$ of $\mathbb{R}^n$, let $\mathcal{P}_{\mathcal{S}}(\cdot) : \mathbb{R}^n \to \mathcal{S}$ denote the orthogonal projection from $\mathbb{R}^n$ onto $\mathcal{S}$. From the definition of $t_i(\mathbf{w})$, it's clear that

$$\mathcal{P}_{\mathcal{L}}(\mathbf{w}) \left(\mathbf{x}^{(i)} \right) = t_i(\mathbf{w}) \cdot \mathbf{w}, \ \forall i \in [m] \text{ and } \mathbf{w} \in \mathbb{R}^n \text{ such that } \|\mathbf{w}\|_2 = 1. \tag{1}$$

Due to the projection theorem (*Theorem 2.2* in [1]), we find that for every $i \in [m]$,

$$\mathbf{x}^{(i)} - \mathcal{P}_{\mathcal{L}}(\mathbf{w}) \left(\mathbf{x}^{(i)} \right) \perp \mathcal{L}(\mathbf{w}) \quad \Rightarrow \quad \mathbf{w}^\top \left\{ \mathbf{x}^{(i)} - \mathcal{P}_{\mathcal{L}}(\mathbf{w}) \left(\mathbf{x}^{(i)} \right) \right\} = 0. \tag{2}$$

Putting (1) into (2), we now have

$$\begin{aligned}
0 &= \mathbf{w}^\top \left\{ \mathbf{x}^{(i)} - \mathcal{P}_{\mathcal{L}}(\mathbf{w}) \left(\mathbf{x}^{(i)} \right) \right\} \\
&= \mathbf{w}^\top \left\{ \mathbf{x}^{(i)} - t_i(\mathbf{w}) \cdot \mathbf{w} \right\} \\
&= \mathbf{w}^\top \mathbf{x}^{(i)} - t_i(\mathbf{w}) \|\mathbf{w}\|_2^2 \\
&\overset{\text{(a)}}{=} \mathbf{w}^\top \mathbf{x}^{(i)} - t_i(\mathbf{w}),
\end{aligned}$$

where the step (a) follows from the fact $\|\mathbf{w}\|_2 = 1$, and thus $t_i(\mathbf{w}) = \mathbf{w}^\top \mathbf{x}^{(i)}$ for every $i \in [m]$.

(2) It's straightforward from (1) that

$$\begin{aligned}
\hat{t}(\mathbf{w}) &= \frac{1}{m} \sum_{i=1}^{m} t_i(\mathbf{w}) \\
&= \frac{1}{m} \sum_{i=1}^{m} \mathbf{w}^\top \mathbf{x}^{(i)} \\
&= \mathbf{w}^\top \left(\frac{1}{m} \sum_{i=1}^{m} \mathbf{x}^{(i)} \right) \\
&= \mathbf{w}^\top \hat{\mathbf{x}},
\end{aligned}$$

where $\hat{\mathbf{x}} := \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}^{(i)}$ is the sample mean of the data points $\left\{\mathbf{x}^{(i)} : i \in [m]\right\}$. The current problem assumes that the function $\hat{t}(\cdot) : \mathbb{S}^{n-1} \to \mathbb{R}$ is a constant function, where $\mathbb{S}^{n-1}$ refers to the $(n-1)$-dimensional unit sphere. If $\hat{\mathbf{x}} \neq \mathbf{0}$, then we have

$$\hat{t}\left(-\frac{\hat{\mathbf{x}}}{\|\hat{\mathbf{x}}\|_2}\right) = -\|\hat{\mathbf{x}}\|_2 \neq \|\hat{\mathbf{x}}\|_2 = \hat{t}\left(\frac{\hat{\mathbf{x}}}{\|\hat{\mathbf{x}}\|_2}\right),$$

which contradicts that assumption. So one can conclude that $\hat{\mathbf{x}} = \mathbf{0}$.

(3) To begin with, let us recall the definition of the sample covariance matrix $\boldsymbol{\Sigma} \in \mathcal{S}^n$, where $\mathcal{S}^n$ denotes the set of all $n \times n$ real symmetric matrices:

$$\boldsymbol{\Sigma} := \frac{1}{m}\sum_{i=1}^{m}\left(\mathbf{x}^{(i)} - \hat{\mathbf{x}}\right)\left(\mathbf{x}^{(i)} - \hat{\mathbf{x}}\right)^{\top}.$$

By the spectral theorem, the $n \times n$ real symmetric matrix $\boldsymbol{\Sigma}$ admits the spectral decomposition

$$\boldsymbol{\Sigma} = \mathbf{U}\mathbf{D}\mathbf{U}^{\top}, \tag{3}$$

where $\mathbf{U} \in \mathcal{O}(n)$ and $\mathbf{D} := \mathsf{diag}\left(\lambda_1(\boldsymbol{\Sigma}), \lambda_2(\boldsymbol{\Sigma}), \cdots, \lambda_n(\boldsymbol{\Sigma})\right) \in \mathbb{R}^{n\times n}$. Here, $\mathcal{O}(n)$ refers to the orthogonal group in dimension n. Now we observe that for any $\mathbf{w} \in \mathbb{S}^{n-1}$,

$$\begin{aligned}
\mathbf{w}^{\top}\boldsymbol{\Sigma}\mathbf{w} &= \frac{1}{m}\sum_{i=1}^{m}\left\{\mathbf{w}^{\top}\left(\mathbf{x}^{(i)} - \hat{\mathbf{x}}\right)\right\}^2 \\
&\stackrel{\text{(b)}}{=} \frac{1}{m}\sum_{i=1}^{m}\left\{\mathbf{w}^{\top}\mathbf{x}^{(i)}\right\}^2 \\
&\stackrel{\text{(c)}}{=} \frac{1}{m}\sum_{i=1}^{m}\left\{t_i(\mathbf{w})\right\}^2 \\
&= \sigma^2(\omega),
\end{aligned}$$

where the step (b) and the step (c) follows from part (2) and part (1), respectively. Since the current problem assumes that the function $\sigma^2(\cdot) : \mathbb{S}^{n-1} \to \mathbb{R}$ is constant, one can conclude that the quadratic form

$$\mathbf{w} \in \mathbb{S}^{n-1} \mapsto \mathbf{w}^{\top}\boldsymbol{\Sigma}\mathbf{w} \in \mathbb{R}$$

is also a constant function. This implies

$$\lambda_1(\boldsymbol{\Sigma}) = \lambda_2(\boldsymbol{\Sigma}) = \cdots = \lambda_n(\boldsymbol{\Sigma}) = \sigma^2. \tag{4}$$

Taking two pieces (3) and (4) collectively, we establish

$$\boldsymbol{\Sigma} = \mathbf{U}\left(\sigma^2 \cdot \mathbf{I}_n\right)\mathbf{U}^{\top} \stackrel{\text{(d)}}{=} \sigma^2 \cdot \mathbf{I}_n,$$

where the step (d) holds since $\mathbf{U}$ is an $n \times n$ orthogonal matrix.

Problem 2 (*Exercise 5.1* in [1]).

(1) It's straightforward to realize that all row vectors as well as column vectors of $\mathbf{A}$ are orthonormal. Moreover, it's clear that

$$\mathbf{A}^{\top}\mathbf{A} = \mathbf{A}\mathbf{A}^{\top} = \mathbf{I}_3,$$

thereby $\mathbf{A} \in \mathcal{O}(3)$.

(2) Since $\mathbf{A}$ is a 3×3 real symmetric matrix, it admits the spectral decomposition which plays a role as a singular value decomposition (SVD for brevity) of $\mathbf{A}$. So it suffices to find its spectral decomposition. To this end, we first compute the eigenvalues of $\mathbf{A}$. The characteristic polynomial of $\mathbf{A}$ is given by

$$\mathsf{ch}_{\mathbf{A}}(x) := \mathsf{det}\,(x\mathbf{I}_3 - \mathbf{A}) = (x-1)(x+1)^2.$$

Therefore, we obtain

$$\lambda_1(\mathbf{A}) = 1 \quad \text{and} \quad \lambda_2(\mathbf{A}) = \lambda_3(\mathbf{A}) = -1.$$

Involving some straightforward computations, one can reveal that

$$\mathcal{E}(1) := \mathcal{N}\,(\mathbf{A} - \mathbf{I}_3) = \mathsf{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\} \quad \text{and} \quad \mathcal{E}(-1) := \mathcal{N}\,(\mathbf{A} - \mathbf{I}_3) = \left\{ \mathbf{x} \in \mathbb{R}^3 : \mathbf{1}_3^\top \mathbf{x} = 0 \right\}, \tag{5}$$

where $\mathbf{1}_3 := \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\top$. Here, $\mathcal{E}(1)$ and $\mathcal{E}(-1)$ stand for the eigen-spaces of $\mathbf{A}$ associated to its eigenvalues 1 and -1, respectively. It's clear that $\mathcal{E}(1) \perp \mathcal{E}(-1)$. Consider

$$\mathbf{v}_1 := \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3, \quad \mathbf{v}_2 := \mathbf{e}_1 - \mathbf{e}_3, \quad \text{and} \quad \mathbf{v}_3 := \mathbf{e}_2 - \mathbf{e}_3,$$

where $\mathbf{e}_i$ denotes the i-th unit vector in Euclidean spaces. Then, $\{\mathbf{v}_1\}$ and $\{\mathbf{v}_2, \mathbf{v}_3\}$ form bases of $\mathcal{E}(1)$ and $\mathcal{E}(-1)$, respectively. By employing the Gram-Schmidt orthonormalization process, we would like to obtain orthonormal bases $\{\mathbf{u}_1\}$ and $\{\mathbf{u}_2, \mathbf{u}_3\}$ for $\mathcal{E}(1)$ and $\mathcal{E}(-1)$, respectively:

$$\begin{aligned}
\mathbf{u}_1 &:= \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|_2} = \frac{1}{\sqrt{3}}\,(\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3)\,; \\
\mathbf{u}_2 &:= \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|_2} = \frac{1}{2}\,(\mathbf{e}_1 - \mathbf{e}_3)\,; \\
\mathbf{u}_3 &:= \frac{\mathbf{v}_3 - \frac{\langle \mathbf{v}_2, \mathbf{v}_3 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2}{\left\| \mathbf{v}_3 - \frac{\langle \mathbf{v}_2, \mathbf{v}_3 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 \right\|_2} = \frac{-\frac{1}{2}\mathbf{e}_1 + \mathbf{e}_2 - \frac{1}{2}\mathbf{e}_3}{\left\| -\frac{1}{2}\mathbf{e}_1 + \mathbf{e}_2 - \frac{1}{2}\mathbf{e}_3 \right\|_2} = \frac{1}{\sqrt{6}}\,(-\mathbf{e}_1 + 2\mathbf{e}_2 - \mathbf{e}_3)\,.
\end{aligned}$$

Since $\mathcal{E}(1) \perp \mathcal{E}(-1)$, we find that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ forms an orthonormal basis for $\mathbb{R}^3$. Let $\mathbf{U} := \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} \in \mathcal{O}_3$. Then we have

$$\mathbf{A}\mathbf{U} = \begin{bmatrix} \mathbf{A}\mathbf{u}_1 & \mathbf{A}\mathbf{u}_2 & \mathbf{A}\mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} \lambda_1(\mathbf{A})\mathbf{u}_1 & \lambda_2(\mathbf{A})\mathbf{u}_2 & \lambda_3(\mathbf{A})\mathbf{u}_3 \end{bmatrix} = \mathbf{U}\boldsymbol{\Sigma}, \tag{6}$$

where $\boldsymbol{\Sigma} := \mathsf{diag}\,(\lambda_1(\mathbf{A}), \lambda_2(\mathbf{A}), \lambda_3(\mathbf{A})) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \in \mathbb{R}^{3\times 3}$. So we arrive at the following spectral decomposition of $\mathbf{A}$:

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^\top = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \end{bmatrix}^\top,$$

and this gives us an SVD of $\mathbf{A}$ as well.

Problem 3 (*Exercise 5.3* in [1]).

(1) Let $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top \in \mathbb{R}^{n\times m}$ be a singular value decomposition of $\mathbf{A}$, where $\mathbf{U} := \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix} \in \mathcal{O}(n)$, $\mathbf{V} := \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \end{bmatrix} \in \mathcal{O}(m)$, and

$$\boldsymbol{\Sigma} := \begin{bmatrix} \mathsf{diag}\left(\sigma_1, \sigma_2, \cdots, \sigma_m\right) \\ \mathbf{O}_{(n-m)\times m} \end{bmatrix} \in \mathbb{R}^{n\times m}.$$

Here, $\mathcal{O}(d)$ denotes the orthogonal group in dimension d and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0$ are the singular values of $\mathbf{A}$. Now we consider the matrix $\tilde{\mathbf{A}} := \begin{bmatrix} \mathbf{A} \\ \mathbf{I}_m \end{bmatrix} \in \mathbb{R}^{(n+m)\times m}$. Then we have

$$\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A}^\top & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{A} \\ \mathbf{I}_m \end{bmatrix} = \mathbf{A}^\top \mathbf{A} + \mathbf{I}_m \in \mathbb{R}^{m\times m}. \tag{7}$$

Let $\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \cdots \geq \tilde{\sigma}_m \geq 0$ denote the singular values of $\tilde{\mathbf{A}}$. Then we obtain for every $i \in [m]$,

$$\begin{aligned} \tilde{\sigma}_i^2 &= \lambda_i\left(\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}}\right) \\ &\overset{\text{(a)}}{=} \lambda_i\left(\mathbf{A}^\top \mathbf{A} + \mathbf{I}_m\right) \\ &= \lambda_i\left\{\mathbf{V}\left(\boldsymbol{\Sigma}^\top \boldsymbol{\Sigma} + \mathbf{I}_m\right)\mathbf{V}^\top\right\} \\ &\overset{\text{(b)}}{=} \lambda_i\left(\boldsymbol{\Sigma}^\top \boldsymbol{\Sigma} + \mathbf{I}_m\right) \\ &= \lambda_i\left(\mathsf{diag}\left(1+\sigma_1^2, 1+\sigma_2^2, \cdots, 1+\sigma_m^2\right)\right) \\ &= 1 + \sigma_i^2, \end{aligned}$$

as desired, where the step (a) follows from the identity (7), and the step (b) holds by the facts that $\mathbf{V} \in \mathcal{O}(m)$ and the similar matrices have the same eigenvalues. So we arrive at $\tilde{\sigma}_i = \sqrt{1+\sigma_i^2}$ for every $i \in [m]$.

(b) We first observe that

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top \\ \mathbf{V}\mathbf{V}^\top \end{bmatrix} = \begin{bmatrix} \mathbf{U} & \mathbf{O}_{n\times m} \\ \mathbf{O}_{m\times n} & \mathbf{V} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma} \\ \mathbf{I}_m \end{bmatrix} \mathbf{V}^\top \tag{8}$$

It is clear from the fact $\mathbf{U} \in \mathcal{O}(n)$ & $\mathbf{V} \in \mathcal{O}(m)$ that

$$\begin{bmatrix} \mathbf{U} & \mathbf{O}_{n\times m} \\ \mathbf{O}_{m\times n} & \mathbf{V} \end{bmatrix} \in \mathcal{O}(n+m). \tag{9}$$

From (8), we may observe that it suffices to find an SVD of the matrix $\begin{bmatrix} \boldsymbol{\Sigma} \\ \mathbf{I}_m \end{bmatrix} \in \mathbb{R}^{(n+m)\times m}$. Given any $d \in \mathbb{N}$, let $\mathbf{e}_i^{(d)} \in \mathbb{R}^d$ denote the i-th unit vector in the d-dimensional Euclidean space $\mathbb{R}^d$. Set

$$\begin{aligned} \mathbf{w}_i &:= \frac{\sigma_i}{\tilde{\sigma}_i}\mathbf{e}_i^{(n+m)} + \frac{1}{\tilde{\sigma}_i}\mathbf{e}_{n+i}^{(n+m)} \text{ for } i \in [m]; \\ \mathbf{w}_{m+i} &:= -\frac{1}{\tilde{\sigma}_i}\mathbf{e}_i^{(n+m)} + \frac{\sigma_i}{\tilde{\sigma}_i}\mathbf{e}_{n+i}^{(n+m)} \text{ for } i \in [m]; \\ \mathbf{w}_{2m+i} &:= \mathbf{e}_{m+i}^{(n+m)} \text{ for } i \in [n-m]. \end{aligned} \tag{10}$$

Here, $[d] := \{1, 2, \cdots, d\}$ for all $d \in \mathbb{N}$. Then, it is straightforward to reveal that $\{\mathbf{w}_i \in \mathbb{R}^{n+m} : i \in [n+m]\}$ forms an orthonormal basis for $\mathbb{R}^{n+m}$. Also since

$$\mathbf{w}_i \left(\mathbf{e}_i^{(m)}\right)^\top = \frac{\sigma_i}{\tilde{\sigma}_i} \mathbf{e}_i^{(n+m)} \left(\mathbf{e}_i^{(m)}\right)^\top + \frac{1}{\tilde{\sigma}_i} \mathbf{e}_{n+i}^{(n+m)} \left(\mathbf{e}_i^{(m)}\right)^\top$$

for every $i \in [m]$, we have

$$\begin{bmatrix} \mathbf{\Sigma} \\ \mathbf{I}_m \end{bmatrix} = \sum_{i=1}^{m} \tilde{\sigma}_i \mathbf{w}_i \left(\mathbf{e}_i^{(m)}\right)^\top = \mathbf{W}\tilde{\mathbf{\Sigma}}\mathbf{I}_m^\top, \tag{11}$$

where $\mathbf{W} := \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots \mathbf{w}_{n+m} \end{bmatrix} \in \mathcal{O}(n+m)$ and $\tilde{\mathbf{\Sigma}} := \begin{bmatrix} \mathsf{diag}\left(\tilde{\sigma}_1, \tilde{\sigma}_2, \cdots, \tilde{\sigma}_m\right) \\ \mathbf{O}_{n \times m} \end{bmatrix} \in \mathbb{R}^{(n+m) \times m}$. Therefore, the equation (11) gives an SVD of $\begin{bmatrix} \mathbf{\Sigma} \\ \mathbf{I}_m \end{bmatrix}$. So by substituting (11) into the equation (8), we now obtain

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{U} & \mathbf{O}_{n \times m} \\ \mathbf{O}_{m \times n} & \mathbf{V} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma} \\ \mathbf{I}_m \end{bmatrix} \mathbf{V}^\top = \begin{bmatrix} \mathbf{U} & \mathbf{O}_{n \times m} \\ \mathbf{O}_{m \times n} & \mathbf{V} \end{bmatrix} \mathbf{W}\tilde{\mathbf{\Sigma}}\mathbf{I}_m^\top \mathbf{V}^\top = \underbrace{\left(\begin{bmatrix} \mathbf{U} & \mathbf{O}_{n \times m} \\ \mathbf{O}_{m \times n} & \mathbf{V} \end{bmatrix} \mathbf{W} \right)}_{=: \tilde{\mathbf{U}}} \tilde{\mathbf{\Sigma}}\mathbf{V}^\top. \tag{12}$$

Putting $\tilde{\mathbf{U}} := \begin{bmatrix} \mathbf{U} & \mathbf{O}_{n \times m} \\ \mathbf{O}_{m \times n} & \mathbf{V} \end{bmatrix} \mathbf{W} \in \mathcal{O}(n+m)$ and $\tilde{\mathbf{V}} := \mathbf{V} \in \mathcal{O}(m)$, the equation (12) provides an SVD of $\tilde{\mathbf{A}}$,

$$\tilde{\mathbf{A}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^\top.$$

This completes our explicit derivation of an SVD of the matrix $\tilde{\mathbf{A}} \in \mathbb{R}^{(n+m) \times m}$.

References

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