# Homework 1

20213073 Donggyu Kim

March 23, 2021

1. (a) $\boldsymbol{A}_\Phi = \begin{bmatrix} 2 & 2 & -1 \\ 0 & 3 & -4 \\ 4 & 4 & 3 \\ 1 & 0 & 0 \end{bmatrix}$

   (b) $\text{rank}(\boldsymbol{A}_\Phi) = 3$. The columns of $\boldsymbol{A}_\Phi$ are linearly independent.

   (c) The kernel and image of $\Phi$ are as follows:

$$
\begin{aligned}
\ker(\Phi) &= \{\boldsymbol{x} \in \mathbb{R}^3 \mid \Phi(\boldsymbol{x}) = \boldsymbol{A}_\Phi \boldsymbol{x} = 0\} \\
&= \{\boldsymbol{0}\} \\
\text{Im}(\Phi) &= \{\boldsymbol{y} \in \mathbb{R}^4 \mid \exists \boldsymbol{x} \in \mathbb{R}^3 \text{ s.t. } \Phi(\boldsymbol{x}) = \boldsymbol{y}\} \\
&= \left\{ \alpha \begin{bmatrix} 2 \\ 0 \\ 4 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} 2 \\ 3 \\ 4 \\ 0 \end{bmatrix} + \gamma \begin{bmatrix} -1 \\ -4 \\ 3 \\ 0 \end{bmatrix} \middle| \, \alpha, \beta, \gamma \in \mathbb{R} \right\}
\end{aligned}
$$

   Thus, $\dim(\ker(\Phi)) = 0$ and $\dim(\text{Im}(\Phi)) = 3$.

2. *Proof.*

$$
\begin{aligned}
\|\boldsymbol{x}\| &= \frac{1}{2}(\|\boldsymbol{x}\| + \|\boldsymbol{x}\|) \\
&= \frac{1}{2}(\|\boldsymbol{x}\| + \|-\boldsymbol{x}\|) && (\because \text{Homogeneity: } \|(-1) \cdot \boldsymbol{x}\| = |-1| \cdot \|\boldsymbol{x}\| = \|\boldsymbol{x}\|) \\
&\geq \frac{1}{2}\|\boldsymbol{x} + (-\boldsymbol{x})\| && (\because \text{Subadditivity}) \\
&= \frac{1}{2}\|\boldsymbol{0}\| \\
&= 0
\end{aligned}
$$

□

3. (a) *Proof.*
   When $\boldsymbol{x} = \boldsymbol{0}$, it satisfies the inequality $\|\boldsymbol{x}\|_p = \|\boldsymbol{x}\|_r$.
   So let's think about the case when $\boldsymbol{x} \neq 0$. Then, division by $\|\boldsymbol{x}\|_p$ is valid.

$$
\sum_{i=1}^{n} \left( \frac{|x_i|}{\|\boldsymbol{x}\|_p} \right)^p = \frac{1}{\|\boldsymbol{x}\|_p^p} \sum_{i=1}^{n} |x_i|^p = 1
$$

This implies that,

$$0 \le \frac{|x_i|}{\|\boldsymbol{x}\|_p} \le 1$$

for all $i = 1, 2, \ldots, n$.
Since $0 < r < p$,

$$\left(\frac{|x_i|}{\|\boldsymbol{x}\|_p}\right)^p \le \left(\frac{|x_i|}{\|\boldsymbol{x}\|_p}\right)^r$$

Therefore,

$$\begin{aligned} 1 = \sum_{i=1}^{n} \left(\frac{|x_i|}{\|\boldsymbol{x}\|_p}\right)^p \le \sum_{i=1}^{n} \left(\frac{|x_i|}{\|\boldsymbol{x}\|_p}\right)^r &= \left(\frac{\|\boldsymbol{x}\|_r}{\|\boldsymbol{x}\|_p}\right)^r \\ 1 &\le \frac{\|\boldsymbol{x}\|_r}{\|\boldsymbol{x}\|_p} \\ \therefore \|\boldsymbol{x}\|_p &\le \|\boldsymbol{x}\|_r \end{aligned}$$

□

(b) *Proof.*
Let $y = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^\top = \begin{bmatrix} |x_1|^r & |x_2|^r & \cdots & |x_n|^r \end{bmatrix}^\top$.

Since $\frac{p}{r}, \frac{p}{p-r} \in (1, \infty)$ and $\frac{r}{p} + \frac{p-r}{p} = 1$, we can use Hölder's inequality as below.

$$\begin{aligned} \sum_{i=1}^{n} |x_i|^r &= \sum_{i=1}^{n} |1 \cdot y_i| \\ &\le \|\mathbf{1}\|_{\frac{p}{p-r}} \|\boldsymbol{y}\|_{\frac{p}{r}} \end{aligned}$$

By the definition of $l_p$ norm, we get

$$\begin{aligned} \|\mathbf{1}\|_{\frac{p}{p-r}} \|\boldsymbol{y}\|_{\frac{p}{r}} &= n^{\frac{p-r}{p}} \left(\sum_{i=1}^{n} |y_i|^{\frac{p}{r}}\right)^{\frac{r}{p}} \\ &= n^{\frac{p-r}{p}} \left(\sum_{i=1}^{n} ||x_i|^r|^{\frac{p}{r}}\right)^{\frac{r}{p}} \\ &= n^{\frac{p-r}{p}} \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{r}{p}} \end{aligned}$$

The proof ends by taking $r$-th root.

$$\begin{aligned} \|\boldsymbol{x}\|_r &= \left(\sum_{i=1}^{n} |x_i|^r\right)^{\frac{1}{r}} \\ &\le n^{\frac{p-r}{pr}} \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \\ &= n^{\frac{1}{r}-\frac{1}{p}} \|\boldsymbol{x}\|_p \end{aligned}$$

□

4. *Proof.*

Before the start, we can see that $\forall a \in \mathbb{R}$,

$$\frac{\|\boldsymbol{A}(a\boldsymbol{x})\|_2}{\|a\boldsymbol{x}\|_2} = \frac{|a|\|\boldsymbol{A}\boldsymbol{x}\|_2}{|a|\|\boldsymbol{x}\|_2} = \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}$$

Therefore, without loss of generality, we can assume $\|\boldsymbol{x}\|_2 = 1$.

Let the singular value decomposition of $\boldsymbol{A}$ as follows:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$$

Then the square of the fraction becomes

$$\begin{aligned}
\|\boldsymbol{A}\boldsymbol{x}\|_2^2 &= (\boldsymbol{A}\boldsymbol{x})^\top(\boldsymbol{A}\boldsymbol{x}) \\
&= \boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x} \\
&= \boldsymbol{x}^\top\boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\boldsymbol{x} \\
&= \boldsymbol{x}^\top\boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}\boldsymbol{V}^\top\boldsymbol{x}
\end{aligned}$$

We can rewrite this in sigma notation

$$\begin{aligned}
\boldsymbol{x}^\top\boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}\boldsymbol{V}^\top\boldsymbol{x} &= \begin{bmatrix}\cdots \boldsymbol{x}^\top\boldsymbol{v}_i \cdots\end{bmatrix} \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix} \begin{bmatrix} \vdots \\ \boldsymbol{v}_i^\top\boldsymbol{x} \\ \vdots \end{bmatrix} \\
&= \sum_{i=1}^n \sigma_i^2(\boldsymbol{v}_i^\top\boldsymbol{x})^2
\end{aligned}$$

where $\{\boldsymbol{v}_i\}$ are the column vectors of $\boldsymbol{V}$, and $\{\sigma_i\}$ are the singular values of $\boldsymbol{A}$.

Now let's use the fact that $\boldsymbol{V}$ is an $n \times n$ orthogonal matrix. This means that its column vectors can span $\mathbb{R}^n$, colspace($\boldsymbol{V}$)=$\mathbb{R}^n$. Therefore, $\boldsymbol{x}$ can be represented as the linear combination of them.

$$\boldsymbol{x} = \sum_{i=1}^n c_i\boldsymbol{v}_i$$

By substituting this, we get

$$\begin{aligned}
\sum_{i=1}^n \sigma_i^2(\boldsymbol{v}_i^\top\boldsymbol{x})^2 &= \sum_{i=1}^n \sigma_i^2\left(\boldsymbol{v}_i^\top\sum_{j=1}^n c_j\boldsymbol{v}_j\right)^2 \\
&= \sum_{i=1}^n\sum_{j=1}^n(\sigma_i c_j\boldsymbol{v}_i^\top\boldsymbol{v}_j)^2 \\
&= \sum_{i=1}^n(\sigma_i c_i\boldsymbol{v}_i^\top\boldsymbol{v}_i)^2 && (\because \text{Orthogonality: } \boldsymbol{v}_i^\top\boldsymbol{v}_j = 0 \text{ for } i \neq j) \\
&= \sum_{i=1}^n \sigma_i^2 c_i^2 && (\because \text{Orthonormality: } \|\boldsymbol{v}_i\|_2 = \boldsymbol{v}_i^\top\boldsymbol{v}_i = 1)
\end{aligned}$$

Recall that we assumed $\|\boldsymbol{x}\|_2 = 1$. This gives an inequality with respect to $\{c_i\}$.

$$\begin{aligned}
\|\boldsymbol{x}\|_2 &= \boldsymbol{x}^\top \boldsymbol{x} \\
&= \left(\sum_{i=1}^{n} c_i \boldsymbol{v}_i^\top\right)\left(\sum_{i=1}^{n} c_i \boldsymbol{v}_i\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \boldsymbol{v}_i^\top \boldsymbol{v}_j \\
&= \sum_{i=1}^{n} c_i^2 && (\because \text{Orthonormality}) \\
&= 1
\end{aligned}$$

Thus, we finally get

$$\|\boldsymbol{A}\boldsymbol{x}\|_2^2 = \sum_{i=1}^{n} \sigma_i^2 c_i^2 \leq \sum_{i=1}^{n} \sigma_1^2 c_i^2 = \sigma_1^2$$

where $\sigma_1$ is the largest singular value of $\boldsymbol{A}$.

$$\therefore \max_{\boldsymbol{x}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = \max_{\|\boldsymbol{x}\|_2=1} \|\boldsymbol{A}\boldsymbol{x}\|_2 = \sigma_1$$

□

5. First, we have to get the eigenvalues of $\boldsymbol{A}^\top \boldsymbol{A}$.

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}$$

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Sigma}^\top \boldsymbol{U}^\top \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top = \boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^\top = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$$

$$\begin{aligned}
\det(\boldsymbol{A}^\top \boldsymbol{A} - \lambda \boldsymbol{I}) &= \begin{vmatrix} 5-\lambda & 3 \\ 3 & 5-\lambda \end{vmatrix} \\
&= (5-\lambda)^2 - 9 \\
&= \lambda^2 - 10\lambda + 16 \\
&= (\lambda - 2)(\lambda - 8) = 0 \quad \Rightarrow \lambda = 2, 8
\end{aligned}$$

Then, let's find the eigenspace for each eigenvalue.

$$E_2 = \left\{ \boldsymbol{x} \in \mathbb{R}^2 \mid \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \boldsymbol{x} = 0 \right\} = \left\{ \alpha \begin{bmatrix} 1 \\ -1 \end{bmatrix} \mid \alpha \in \mathbb{R} \right\}$$

$$E_8 = \left\{ \boldsymbol{x} \in \mathbb{R}^2 \mid \begin{bmatrix} -3 & 3 \\ 3 & -3 \end{bmatrix} \boldsymbol{x} = 0 \right\} = \left\{ \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix} \mid \alpha \in \mathbb{R} \right\}$$

Since $\boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^\top \boldsymbol{V} = \boldsymbol{V}\boldsymbol{\Sigma}^2$,

$$\boldsymbol{\Sigma}^2 = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix} \Rightarrow \boldsymbol{\Sigma} = \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}$$

$$\boldsymbol{V} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

$\boldsymbol{U}$ can be found in the same way, but we do not have to do this again. Since we already computed $\boldsymbol{V}$ and $\boldsymbol{\Sigma}$, let's just use them.

$$\begin{aligned}\boldsymbol{U} &= \boldsymbol{A}\boldsymbol{V}\boldsymbol{\Sigma}^{-1} \\ &= \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{2\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\end{aligned}$$

Finally we found the singular value decomposition of $\boldsymbol{A}$.

$$\begin{aligned}\boldsymbol{A} &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top \\ &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \\ &= \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}\end{aligned}$$

**IMPORTANT**: Numerator layout and denominator layout were mixed together in the lecture. For the problem 6 and 7, I will use *denominator* layout. For example, $\frac{\partial y}{\partial \boldsymbol{x}} \in \mathbb{R}^n$ and $\frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A}^\top$. You can get the answer for numerator layout by taking a transpose to the answer from denominator layout.

6. Let's compute $\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}$ first. Since $\boldsymbol{h} : \mathbb{R}^D \to \mathbb{R}^D$, the result should be in $\mathbb{R}^{D\times D}$.

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{h}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial (\boldsymbol{x} - \boldsymbol{\mu})}{\partial \boldsymbol{x}} = \boldsymbol{I}_D \in \mathbb{R}^{D\times D}$$

Next, let's compute $\frac{\partial z}{\partial \boldsymbol{y}}$. Since $g : \mathbb{R}^D \to \mathbb{R}$, the result should be in $\mathbb{R}^D$.

$$\frac{\partial z}{\partial \boldsymbol{y}} = \frac{\partial g(\boldsymbol{y})}{\partial \boldsymbol{y}} = \frac{\partial (\boldsymbol{y}^\top \boldsymbol{S}^{-1}\boldsymbol{y})}{\partial \boldsymbol{y}} = \underbrace{(\boldsymbol{S}^{-1} + \boldsymbol{S}^{-\top})}_{\mathbb{R}^{D\times D}} \underbrace{\boldsymbol{y}}_{\mathbb{R}^D} \in \mathbb{R}^D$$

Now, let's compute $\frac{\partial f}{\partial z}$. We know that the result is in $\mathbb{R}$.

$$\frac{\partial f}{\partial z} = \frac{\partial \left(\exp\left(-\frac{1}{2}z\right)\right)}{\partial z} = -\frac{1}{2}\exp\left(-\frac{1}{2}z\right) \in \mathbb{R}$$

By using the chain rule - in proper order that the result of $\frac{\partial f}{\partial \boldsymbol{x}}$ is in $\mathbb{R}^D$ - we can get

$$\begin{aligned}\frac{\partial f}{\partial \boldsymbol{x}} &= \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\frac{\partial z}{\partial \boldsymbol{y}}\frac{\partial f}{\partial z} \\ &= \boldsymbol{I}_D(\boldsymbol{S}^{-1} + \boldsymbol{S}^{-\top})\boldsymbol{y}\left(-\frac{1}{2}\exp\left(-\frac{1}{2}z\right)\right) \\ &= -\frac{1}{2}\exp\left(-\frac{1}{2}z\right)(\boldsymbol{S}^{-1} + \boldsymbol{S}^{-\top})\boldsymbol{y}\end{aligned}$$

After some substitution, the final result is as below.

$$\begin{aligned}\frac{\partial f}{\partial \boldsymbol{x}} &= -\frac{1}{2}\exp\left(-\frac{1}{2}(\boldsymbol{y}^\top \boldsymbol{S}^{-1}\boldsymbol{y})\right)(\boldsymbol{S}^{-1} + \boldsymbol{S}^{-\top})\boldsymbol{y} \\ &= -\frac{1}{2}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{S}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)(\boldsymbol{S}^{-1} + \boldsymbol{S}^{-\top})(\boldsymbol{x}-\boldsymbol{\mu}) \in \mathbb{R}^D\end{aligned}$$

7. Let's compute $\frac{\partial z}{\partial \boldsymbol{x}}$ first. Since $\boldsymbol{x} \in \mathbb{R}^D, \boldsymbol{z} \in \mathbb{R}^E$, the result should be in $\mathbb{R}^{D\times E}$.

$$\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = \frac{\partial(\boldsymbol{Ax}+\boldsymbol{b})}{\partial \boldsymbol{x}} = \boldsymbol{A}^\top \in \mathbb{R}^{D\times E}$$

Let's compute $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}}$ now. Since $\boldsymbol{f} : \mathbb{R}^E \to \mathbb{R}^E$, the result should be in $\mathbb{R}^{E\times E}$.

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}} = \frac{\partial(\sin(\boldsymbol{z}))}{\partial \boldsymbol{z}} = \mathrm{diag}(\cos(z_1), \cos(z_2), \ldots, \cos(z_E)) \in \mathbb{R}^{E\times E}$$

where $\boldsymbol{z} = \begin{bmatrix} z_1 & z_2 & \cdots & z_E \end{bmatrix}^\top$.

By using the chain rule - in proper order that the result of $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}$ is in $\mathbb{R}^{D\times E}$ - we get

$$\begin{aligned}\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} &= \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}} \\ &= \boldsymbol{A}^\top \mathrm{diag}(\cos(z_1), \cos(z_2), \ldots, \cos(z_E)) \in \mathbb{R}^{D\times E}\end{aligned}$$