Assignment 2

Student ID: u7205329

Exercise 1. We have $X^T := \{ \underset{\sim}{v} \in V : \langle \underset{\sim}{x}, \underset{\sim}{v} \rangle = 0 \text{ for all } \underset{\sim}{x} \in X \}$

1. Let $\underset{\sim}{x} \in X$, $\langle \underset{\sim}{x}, \underset{\sim}{0} \rangle = \langle \underset{\sim}{0}, \underset{\sim}{x} \rangle = 0$ (property of inner product space)
$\Rightarrow \underset{\sim}{0} \in X^T$.

Since X is a subspace of V, $\underset{\sim}{0} \in X$. Hence, $\underset{\sim}{0} \in X \cap X^T$.

Let $\underset{\sim}{w} \in X \cap X^T$, then $\underset{\sim}{w} \in X$ and $\underset{\sim}{w} \in X^T$
this means $\langle \underset{\sim}{w}, \underset{\sim}{w} \rangle = 0$

from the positive definite property of inner product, $\underset{\sim}{w}$ can only be $\underset{\sim}{0}$.

Thus, $X \cap X^T = \{ \underset{\sim}{0} \}$. $\square$

2. We have $X \subseteq Y$, this means if $\underset{\sim}{x} \in X$ then $\underset{\sim}{x} \in Y$. (1)

Let $\underset{\sim}{x} \in X$ and $\underset{\sim}{w} \in Y^T$ (2)
From (1), we know $\underset{\sim}{x} \in Y$ and so $\langle \underset{\sim}{x}, \underset{\sim}{w} \rangle = 0$.

It also follows that $\underset{\sim}{w} \in X^T$ from the definition of X^T. (3)

From (2) and (3), we can conclude that $Y^T \subseteq X^T$. $\square$

Exercise 2.

1. Taking the inner product of $\underset{\sim}{u}$ and $\underset{\sim}{v} - \text{proj}_{\underset{\sim}{u}} \underset{\sim}{v}$, we have:

$$\begin{aligned}
\langle \underset{\sim}{u}, \underset{\sim}{v} - \text{proj}_{\underset{\sim}{u}} \underset{\sim}{v} \rangle &= \left\langle \underset{\sim}{u}, \underset{\sim}{v} - \frac{\langle \underset{\sim}{v}, \underset{\sim}{u} \rangle}{\langle \underset{\sim}{u}, \underset{\sim}{u} \rangle} \underset{\sim}{u} \right\rangle \\
&= \langle \underset{\sim}{u}, \underset{\sim}{v} \rangle - \left\langle \underset{\sim}{u}, \frac{\langle \underset{\sim}{v}, \underset{\sim}{u} \rangle}{\langle \underset{\sim}{u}, \underset{\sim}{u} \rangle} \underset{\sim}{u} \right\rangle \\
&= \langle \underset{\sim}{u}, \underset{\sim}{v} \rangle - \frac{\langle \underset{\sim}{v}, \underset{\sim}{u} \rangle}{\langle \underset{\sim}{u}, \underset{\sim}{u} \rangle} \langle \underset{\sim}{u}, \underset{\sim}{u} \rangle \\
&= \langle \underset{\sim}{u}, \underset{\sim}{v} \rangle - \langle \underset{\sim}{u}, \underset{\sim}{v} \rangle \\
&= 0.
\end{aligned}$$

Since the inner product of $\underset{\sim}{u}$ and $\underset{\sim}{v} - \text{proj}_{\underset{\sim}{u}} (\underset{\sim}{v})$ equals to zero, we can quickly conclude that they are orthogonal.

2. Let $\|\underset{\sim}{x}\| := \sqrt{\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle}$. Prove that $\|\cdot\|$ is a norm.

We want to check if $\|\cdot\|$ satisfies the following properties

1) Positive definite
2) Absolute homogenous
3) Triangle inequality

1. $\|\cdot\|$ satisfies the positive definite property, as shown below

Since $\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle$ is an inner product, we know that

$$\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle > 0 \quad \forall \underset{\sim}{x} \neq 0 \Rightarrow \sqrt{\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle} > 0 \quad \forall \underset{\sim}{x} \neq \underset{\sim}{0}, \text{ and}$$
$$\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle = 0 \Leftrightarrow \underset{\sim}{x} = 0 \Rightarrow \sqrt{\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle} = 0 \Leftrightarrow \underset{\sim}{x} = 0.$$

2. $\|\cdot\|$ satisfies the absolute homogenous property, as shown below

We have $\|\alpha \underset{\sim}{x}\| = \sqrt{\langle \alpha \underset{\sim}{x}, \alpha \underset{\sim}{x} \rangle}$

$$= \sqrt{\alpha^2 \langle \underset{\sim}{x}, \underset{\sim}{x} \rangle}$$

$$= |\alpha| \sqrt{\langle \underset{\sim}{x}, \underset{\sim}{x} \rangle} = |\alpha| \, \|\underset{\sim}{x}\|$$

3. $\|\cdot\|$ satisfies the triangle inequality, as shown below

$$\|\underset{\sim}{x} + \underset{\sim}{y}\|^2 = \langle \underset{\sim}{x} + \underset{\sim}{y}, \underset{\sim}{x} + \underset{\sim}{y} \rangle$$

$$= \langle \underset{\sim}{x}, \underset{\sim}{x} + \underset{\sim}{y} \rangle + \langle \underset{\sim}{y}, \underset{\sim}{x} + \underset{\sim}{y} \rangle$$

$$= \langle \underset{\sim}{x}, \underset{\sim}{x} \rangle + \langle \underset{\sim}{x}, \underset{\sim}{y} \rangle + \langle \underset{\sim}{y}, \underset{\sim}{x} \rangle + \langle \underset{\sim}{y}, \underset{\sim}{y} \rangle$$

$$= \langle \underset{\sim}{x}, \underset{\sim}{x} \rangle + 2 \langle \underset{\sim}{x}, \underset{\sim}{y} \rangle + \langle \underset{\sim}{y}, \underset{\sim}{y} \rangle$$

$$\leq \|\underset{\sim}{x}\|^2 + 2\|\underset{\sim}{x}\| \, \|\underset{\sim}{y}\| + \|\underset{\sim}{y}\|^2$$

Cauchy-Schwartz inequality

$$\Rightarrow \|\underset{\sim}{x} + \underset{\sim}{y}\|^2 \leq (\|\underset{\sim}{x}\| + \|\underset{\sim}{y}\|)^2$$

$$\Rightarrow \|\underset{\sim}{x} + \underset{\sim}{y}\| \leq \|\underset{\sim}{x}\| + \|\underset{\sim}{y}\|$$

Thus, we can conclude that $\|\cdot\|$ is a norm.

Exercise 3.

1. We have $f, g : \mathbb{R}^n \to \mathbb{R}$, $f(\underset{\sim}{x}) = \underset{\sim}{c}^T \underset{\sim}{x}$ and $g(\underset{\sim}{x}) = \sqrt{\underset{\sim}{c}^T \underset{\sim}{x} + \mu^2}$

a) Prove that $\dfrac{d f(\underset{\sim}{x})}{d\underset{\sim}{x}} = \underset{\sim}{c}^T$.

We have
$$\frac{df}{d\underset{\sim}{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{\partial(\underset{\sim}{c}^T \underset{\sim}{x})}{\partial x_1} & \frac{\partial(\underset{\sim}{c}^T \underset{\sim}{x})}{\partial x_2} & \cdots & \frac{\partial(\underset{\sim}{c}^T \underset{\sim}{x})}{\partial x_n} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{\partial(\Sigma c_i x_i)}{\partial x_1} & \frac{\partial(\Sigma c_i x_i)}{\partial x_2} & \cdots & \frac{\partial(\Sigma c_i x_i)}{\partial x_n} \end{bmatrix}$$
$$= \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix} = \underset{\sim}{c}^T .$$

b) Calculate $\dfrac{dg}{d\underset{\sim}{x}}$

We have
$$\frac{dg}{d\underset{\sim}{x}} = \begin{bmatrix} \frac{\partial g}{\partial x_1} & \cdots & \frac{\partial g}{\partial x_n} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{\partial\left(\sqrt{\underset{\sim}{c}^T \underset{\sim}{x} + \mu^2}\right)}{\partial x_1} & \cdots & \frac{\partial\left(\sqrt{\underset{\sim}{c}^T \underset{\sim}{x} + \mu^2}\right)}{\partial x_n} \end{bmatrix}$$

We have for individual x_k
$$\frac{\partial\left(\sqrt{\underset{\sim}{c}^T \underset{\sim}{x} + \mu^2}\right)}{\partial x_k} = \frac{\partial\left(\Sigma c_i x_i + \mu^2\right)^{1/2}}{\partial x_k} = \frac{1}{2}\left(\Sigma c_i x_i + \mu^2\right)^{-1/2} c_k$$
$$= \frac{c_k}{2\sqrt{\Sigma c_i x_i + \mu^2}} = \frac{c_k}{2\sqrt{\underset{\sim}{c}^T \underset{\sim}{x} + \mu^2}}$$

$$\Rightarrow \frac{dg}{d\underset{\sim}{x}} = \begin{bmatrix} \frac{c_1}{2\sqrt{\underset{\sim}{c}^T \underset{\sim}{x} + \mu^2}} & \cdots & \frac{c_n}{2\sqrt{\underset{\sim}{c}^T \underset{\sim}{x} + \mu^2}} \end{bmatrix} = \frac{\underset{\sim}{c}^T}{2\sqrt{\underset{\sim}{c}^T \underset{\sim}{x} + \mu^2}} .$$

2. Prove that the gradient of the regularised least squares error is given by

$$\frac{d\ell(\mathbf{x})}{d\mathbf{x}} = 2\left(\mathbf{x}^T A^T A - \mathbf{b}^T A\right) + 2\lambda \mathbf{x}^T$$

Using the definition of the Euclidean norm, we have

$$\begin{aligned}
\ell(\mathbf{x}) &= \| A\mathbf{x} - \mathbf{b} \|_2^2 + \lambda \| \mathbf{x} \|_2^2 \\
&= (A\mathbf{x} - \mathbf{b})^T (A\mathbf{x} - \mathbf{b}) + \lambda\, \mathbf{x}^T \mathbf{x} \\
&= (A\mathbf{x})^T (A\mathbf{x}) - (A\mathbf{x})^T \mathbf{b} - \mathbf{b}^T (A\mathbf{x}) + \mathbf{b}^T \mathbf{b} + \lambda . \mathbf{x}^T \mathbf{x} \\
&= \underbrace{\mathbf{x}^T A^T A \mathbf{x}}_{A} - \underbrace{2\mathbf{x}^T A^T \mathbf{b}}_{B} + \underbrace{\mathbf{b}^T \mathbf{b}}_{C} + \underbrace{\lambda \mathbf{x}^T \mathbf{x}}_{D}
\end{aligned}$$

Using the sum rule of derivatives, we can find the gradients of each individual parts and add them together later

Part A. We have $A^T A$ is an $n \times n$ symmetric matrix

$$A^T A = \begin{bmatrix} \mathbf{a}_1^T \mathbf{a}_1 & \cdots & \mathbf{a}_1^T \mathbf{a}_n \\ \vdots & & \vdots \\ \mathbf{a}_n^T \mathbf{a}_1 & \cdots & \mathbf{a}_n^T \mathbf{a}_n \end{bmatrix}$$

with $\mathbf{a}_i$, $i = 1, \dots, n$ the column vectors of A

$$\begin{aligned}
\mathbf{x}^T A^T A &= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \mathbf{a}_1^T \mathbf{a}_1 & \cdots & \mathbf{a}_1^T \mathbf{a}_n \\ \vdots & & \vdots \\ \mathbf{a}_n^T \mathbf{a}_1 & \cdots & \mathbf{a}_n^T \mathbf{a}_n \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^n x_i (a_i^T a_1) & \sum_{i=1}^n x_i (a_i^T a_2) & \cdots & \sum_{i=1}^n x_i (a_i^T a_n) \end{bmatrix}
\end{aligned}$$

$$\underset{\sim}{x}^T A^T A \underset{\sim}{x} = \begin{bmatrix} \sum_{i=1}^{n} x_i (a_i^T a_1) & \sum_{i=1}^{n} x_i (a_i^T a_2) & \cdots & \sum_{i=1}^{n} x_i (a_i^T a_n) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= x_1 \cdot \sum_{i=1}^{n} x_i (a_i^T a_1) + x_2 \cdot \sum_{i=1}^{n} x_i (a_i^T a_2) + \cdots + x_n \cdot \sum_{i=1}^{n} x_i (a_i^T a_n)$$

$$\begin{aligned} = \; & x_1^2 (a_1^T a_1) + x_1 x_2 (\underset{\sim}{a_1}^T \underset{\sim}{a_2}) + \cdots + x_1 x_n (a_1^T a_n) \\ + \; & x_2 x_1 (a_2^T a_1) + x_2^2 (a_2^T a_2) + \cdots + x_2 x_n (a_2^T a_n) \\ + \; & \cdots \\ + \; & x_n x_1 (a_n^T a_1) + x_2 x_n (a_n^T a_2) + \cdots + x_n^2 (a_n^T a_n) \end{aligned}$$

$$\frac{\partial (\underset{\sim}{x}^T A^T A \underset{\sim}{x})}{\partial x_1} = 2 x_1 (\underset{\sim}{a_1}^T \underset{\sim}{a_1}) + 2 x_2 (\underset{\sim}{a_1}^T \underset{\sim}{a_2}) + \cdots + 2 x_n (\underset{\sim}{a_1}^T \underset{\sim}{a_n})$$

$$= 2 \sum x_i (a_1^T a_i)$$

$$\frac{d (\underset{\sim}{x}^T A^T A \underset{\sim}{x})}{d \underset{\sim}{x}} = \begin{bmatrix} \frac{\partial (\underset{\sim}{x}^T A^T A x)}{\partial x_1} & \frac{\partial (\underset{\sim}{x}^T A^T A x)}{\partial x_2} & \cdots & \frac{\partial (\underset{\sim}{x}^T A^T A \underset{\sim}{x})}{\partial x_n} \end{bmatrix}$$

$$= \begin{bmatrix} 2 \sum x_i (a_i^T a_1) & 2 \sum x_i (a_i^T a_2) & \cdots & 2 \sum x_i (a_i^T a_n) \end{bmatrix}$$

$$= 2 \begin{bmatrix} \sum x_i (a_i^T a_1) & \sum x_i (a_i^T a_2) & \cdots & \sum x_i (a_i^T a_n) \end{bmatrix}$$

$$= 2 \underset{\sim}{x}^T A^T A .$$

Part B. We have $A^T b = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{k1} \\ a_{12} & a_{22} & & a_{k2} \\ \vdots & & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{kn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} \underset{\sim}{a_1}^T \underset{\sim}{b} \\ \underset{\sim}{a_2}^T \underset{\sim}{b} \\ \vdots \\ \underset{\sim}{a_n}^T \underset{\sim}{b} \end{bmatrix}$

$$\Rightarrow \underset{\sim}{x}^T A^T \underset{\sim}{b} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \underset{\sim}{a_1}^T \underset{\sim}{b} \\ \underset{\sim}{a_2}^T \underset{\sim}{b} \\ \vdots \\ \underset{\sim}{a_n}^T \underset{\sim}{b} \end{bmatrix} = \sum_{i=1}^{n} x_i (\underset{\sim}{a_i}^T \underset{\sim}{b})$$

And so, we have $\dfrac{\partial(2\underset{\sim}{x}^TA^T\underset{\sim}{b})}{\partial x_i} = \dfrac{\partial(2\Sigma x_i(\underset{\sim}{a_i}^T\underset{\sim}{b}))}{\partial x_k} = 2\underset{\sim}{a_k}^T\underset{\sim}{b}$

$$\Rightarrow \frac{d(2\underset{\sim}{x}^TA^T\underset{\sim}{b})}{d\underset{\sim}{x}} = \left[\frac{\partial(2\underset{\sim}{x}^TA^T\underset{\sim}{b})}{\partial x_1} \quad \frac{\partial(2\underset{\sim}{x}^TA^T\underset{\sim}{b})}{\partial x_2} \quad \cdots \quad \frac{\partial(2\underset{\sim}{x}^TA^T\underset{\sim}{b})}{\partial x_n}\right]$$

$$= \left[2(a_1^T\underset{\sim}{b}) \quad 2(\underset{\sim}{a_2}^T\underset{\sim}{b}) \quad \cdots \quad 2(\underset{\sim}{a_n}^T\underset{\sim}{b})\right]$$

$$= 2(A^T\underset{\sim}{b})^T = 2\underset{\sim}{b}^TA.$$

Part C.

$$\frac{d(\underset{\sim}{b}^T\underset{\sim}{b})}{d\underset{\sim}{x}} = \left[\frac{\partial(\underset{\sim}{b}^T\underset{\sim}{b})}{\partial x_1} \quad \frac{\partial(\underset{\sim}{b}^T\underset{\sim}{b})}{\partial x_2} \quad \cdots \quad \frac{\partial(\underset{\sim}{b}^T\underset{\sim}{b})}{\partial x_n}\right]$$

$$= \left[\frac{\partial(\Sigma b_i^2)}{\partial x_1} \quad \frac{\partial(\Sigma b_i^2)}{\partial x_2} \quad \cdots \quad \frac{\partial(\Sigma b_i^2)}{\partial x_n}\right] = \underset{\sim}{0}^T.$$

Part D.

$$\frac{d(\lambda\underset{\sim}{x}^T\underset{\sim}{x})}{d\underset{\sim}{x}} = \left[\frac{\partial(\lambda\underset{\sim}{x}^T\underset{\sim}{x})}{\partial x_1} \quad \frac{\partial(\lambda\underset{\sim}{x}^T\underset{\sim}{x})}{\partial x_2} \quad \cdots \quad \frac{\partial(\lambda\underset{\sim}{x}^T\underset{\sim}{x})}{\partial x_n}\right]$$

$$= \left[\frac{\partial(\lambda\Sigma x_i^2)}{\partial x_1} \quad \frac{\partial(\lambda\Sigma x_i^2)}{\partial x_2} \quad \cdots \quad \frac{\partial(\lambda\Sigma x_i^2)}{\partial x_n}\right]$$

$$= \left[2\lambda x_1 \quad 2\lambda x_2 \quad \cdots \quad 2\lambda x_n\right] = 2\lambda\underset{\sim}{x}^T.$$

Finally,

$$\frac{d\ell(\underset{\sim}{x})}{d\underset{\sim}{x}} = \frac{d}{d\underset{\sim}{x}}(\underset{\sim}{x}^TA^TA\underset{\sim}{x} - 2\underset{\sim}{x}^TA^T\underset{\sim}{b} + \underset{\sim}{b}^T\underset{\sim}{b} + \lambda\underset{\sim}{x}^T\underset{\sim}{x})$$

$$= \frac{d}{d\underset{\sim}{x}}(\underset{\sim}{x}^TA^TA\underset{\sim}{x}) - \frac{d}{d\underset{\sim}{x}}(2\underset{\sim}{x}^TA^T\underset{\sim}{b}) + \frac{d}{d\underset{\sim}{x}}(\underset{\sim}{b}^T\underset{\sim}{b}) + \frac{d}{d\underset{\sim}{x}}(\lambda\underset{\sim}{x}^T\underset{\sim}{x})$$

$$= 2\underset{\sim}{x}^TA^TA - 2\underset{\sim}{b}^TA + \underset{\sim}{0}^T + 2\lambda\underset{\sim}{x}^T.$$

$$= 2(\underset{\sim}{x}^TA^TA - \underset{\sim}{b}^TA) + 2\lambda\underset{\sim}{x}^T. \qquad \square$$