STAT 510 Homework 2

Yifan Zhu

January 23, 2017

1. $\boldsymbol{z} \in \mathcal{C}(\boldsymbol{X}) \Rightarrow \boldsymbol{z} = \boldsymbol{X}\boldsymbol{b}$ for some $\boldsymbol{b}$. Hence

$$
\begin{aligned}
&(\boldsymbol{y} - \boldsymbol{P_X}\boldsymbol{y})^T(\boldsymbol{P_X}\boldsymbol{y} - \boldsymbol{z}) \\
=&(\boldsymbol{y} - \boldsymbol{P_X}\boldsymbol{y})^T(\boldsymbol{P_X}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) \\
=&\boldsymbol{y}^T(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{P_X}\boldsymbol{y} - \boldsymbol{y}^T(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{X}\boldsymbol{b} \\
=&\boldsymbol{y}^T(\boldsymbol{P_X} - \boldsymbol{P_X})\boldsymbol{y} - \boldsymbol{y}^T(\boldsymbol{X} - \boldsymbol{X})\boldsymbol{b} \\
=&\boldsymbol{0}
\end{aligned}
$$

We also have $\boldsymbol{z} \neq \boldsymbol{P_X}\boldsymbol{y} \Rightarrow \boldsymbol{P_X}\boldsymbol{y} - \boldsymbol{z} \neq 0$. Thus we have

$$\|\boldsymbol{y} - \boldsymbol{z}\|^2 = \|\boldsymbol{y} - \boldsymbol{P_X} + \boldsymbol{P_X} - \boldsymbol{z}\|^2 > \|\boldsymbol{y} - \boldsymbol{P_X}\boldsymbol{y}\|^2$$

2. For projection matrix $\boldsymbol{P_X}$ we have $\boldsymbol{P_X}\boldsymbol{X} = \boldsymbol{X}$. Let $\boldsymbol{X} = \begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_p\end{bmatrix} = [x_{ij}]_{n\times p}$ and $\boldsymbol{P_X} = \begin{bmatrix}\boldsymbol{\epsilon}_1 & \boldsymbol{\epsilon}_2 & \cdots & \boldsymbol{\epsilon}_n\end{bmatrix}$. Hence we have

$$\begin{bmatrix}\boldsymbol{\epsilon}_1 & \boldsymbol{\epsilon}_2 & \cdots & \boldsymbol{\epsilon}_n\end{bmatrix}\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn}\end{bmatrix} = \begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_p\end{bmatrix}$$

Thus

$$\boldsymbol{x}_j = \sum_{i=1}^{n} x_{ij}\boldsymbol{\epsilon}_i \Rightarrow \mathcal{C}(\boldsymbol{X}) \subset \mathcal{C}(\boldsymbol{P_X})$$

We also have

$$\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T = \boldsymbol{P_X}$$

Let $(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T = [a_{ij}]_{p\times n}$, thus

$$\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_p\end{bmatrix}\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\epsilon}_1 & \boldsymbol{\epsilon}_2 & \cdots & \boldsymbol{\epsilon}_n\end{bmatrix}$$

Thus

$$\boldsymbol{\epsilon}_j = \sum_{i=1}^{p} a_{ij}\boldsymbol{x}_i \Rightarrow \mathcal{C}(\boldsymbol{P_X}) \subset \mathcal{C}(\boldsymbol{X})$$

Hence we have $\mathcal{C}(\boldsymbol{P_X}) = \mathcal{C}(\boldsymbol{X})$.

3.

$$\begin{aligned}
&\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{y}\\
=&\boldsymbol{X}^T\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y}\\
=&\boldsymbol{X}^T\boldsymbol{P}_{\boldsymbol{X}}^T\boldsymbol{y}\\
=&(\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X})^T\boldsymbol{y}\\
=&\boldsymbol{X}^T\boldsymbol{y}
\end{aligned}$$

Hence $(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{y}$ is a solution of $\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}^T\boldsymbol{y}$.

4. (a) $\boldsymbol{C}\hat{\boldsymbol{\beta}} = C(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}) = \boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{\epsilon} = \boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{\epsilon} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{\epsilon} = \boldsymbol{C}\boldsymbol{\beta} + \boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{\epsilon}$.
$\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, thus $\boldsymbol{C}\boldsymbol{\beta} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\begin{aligned}
&\boldsymbol{\mu} = \boldsymbol{C}\boldsymbol{\beta}\\
&\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}\sigma^2\boldsymbol{I}\boldsymbol{P}_{\boldsymbol{X}}^T\boldsymbol{A}^T = \boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{A}^T = \boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{A}^T = \boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{C}^T
\end{aligned}$$

(b) Let $\boldsymbol{G} = (\boldsymbol{X}^T\boldsymbol{X})^-$ be one of the generalized inverse of $\boldsymbol{X}^T\boldsymbol{X}$ and $\boldsymbol{G}^T$ be its transpose. Thus

$$Var(\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{y}) = \boldsymbol{C}\boldsymbol{G}\boldsymbol{X}^T\sigma^2\boldsymbol{I}\boldsymbol{X}\boldsymbol{G}^T\boldsymbol{C}^T = \boldsymbol{C}\boldsymbol{G}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{G}^T\boldsymbol{C}^T$$