L17 Least square solutions

1. Least square solutions

   (1) Definition of least square solutions
   Regardless if $Ax = b$ is consistent or inconsistent, define

   $$\widehat{x} \text{ is a least square solution (LSS) to } Ax = b \overset{def}{\iff} \|b - A\widehat{x}\|^2 \le \|b - Ax\|^2 \text{ for all } x.$$

   Then the collection of all LSSs to $Ax = b$ is $A^+b + \mathcal{N}(A)$.

   Proof: $\widehat{x}$ is a LSS to $Ax = b \iff \|b - A\widehat{x}\|^2 \le \|b - Ax\|^2$ for all $x \in C^n$
   $\iff A\widehat{x} = \pi(b|\mathcal{R}(A)) = AA^+b$
   $\iff A(\widehat{x} - A^+b) = 0 \iff \widehat{x} - A^+b \in \mathcal{N}(A)$
   $\iff \widehat{x} \in A^+b + \mathcal{N}(A)$.

   Comment: $A^+b + \mathcal{N}(A) = A^+b + \mathcal{R}(I - A^+A) = \{A^+b + (I - A^+A)h : \, h \in C^n\}$.

   (2) A relation
   Suppose $Ax = b$ is consistent. Then

   $$\widehat{x} \text{ is a LSS to } Ax = b \iff \widehat{x} \text{ is an ordinary solution to } Ax = b.$$

   Pf: $Ax = b$ is consistent$\Longrightarrow \exists\, x_0$ such that $Ax_0 = b$.

   ⇒: If $\widehat{x}$ is a LSS to $Ax = b$, then $\|b - A\widehat{x}\|^2 \le \|b - Ax_0\|^2 = 0$. So $A\widehat{x} = b$.
   Thus $\widehat{x}$ is an ordinary solution to $Ax = b$.

   ⇐: If $\widehat{x}$ is an ordinary solution to $Ax = b$, $\|b - A\widehat{x}\|^2 = 0 \le \|b - Ax\|^2$ for all x,
   So $\widehat{x}$ is a LSS to $Ax = b$.

   (3) The collection of all solutions
   to consistent $Ax = b$ is $A^+b + \mathcal{N}(A) = A^+b + \mathcal{R}(I - A^+A) = \{A^+b + (I - A^+A)h : \, h \in C^n\}$.

Ex1: For regression $y = X\beta + \epsilon$ with $E(\epsilon) = 0$, $E(y) = X\beta \in \mathcal{R}(X)$. Define

$$\widehat{\beta} \text{ is a least square estimator LSE for } \beta \iff \|y - X\widehat{\beta}\|^2 \le \|y - X\beta\|^2 \text{ for all } \beta.$$

Let LSE(β) be the collection of all LSEs of β. Then

$$\widehat{\beta} \in \text{LSE}(\beta) \iff \widehat{\beta} \text{ is a LSS to } X\beta = y.$$

So LSE(β) $= X^+y + \mathcal{N}(X)$.

2. Minimum norm LSS

   (1) Other expressions for the collection of all LSSs
   Let x_0 be a LSS to $Ax = b$. Then $\quad A^+b + \mathcal{N}(A) = x_0 + \mathcal{N}(A)$

   Proof Note that x_0 is a LSS to $Ax = b \iff Ax_0 = \pi(b|\mathcal{R}(A)) \iff Ax_0 = AA^+b$.

   ⊂: $\widehat{x} \in A^+b + \mathcal{N}(A) \Longrightarrow \widehat{x} = A^+b + z_1$ where $z_1 \in \mathcal{N}(A)$.
   So $\widehat{x} = x_0 + (A^+b - x_0 + z_1)$.
   But $A(A^+b - x_0 + z_n) = (AA^+b - Ax_0) + Az_1 = 0 + 0 = 0$.
   Thus $A^+b - x_0 + z_1 \in \mathcal{N}(A)$. Hence $\widehat{x} \in x_0 + \mathcal{N}(A)$.

$\supset$**:** $\widehat{x} \in x_0 + \mathcal{N}(A) \Longrightarrow \widehat{x} = x_0 + z_2$ where $z_2 \in \mathcal{N}(A)$.
So $\widehat{x} = A^+b + (x_0 - A^+b + z_2)$.
But $A(x_0 - A^+b + z_2) = (Ax_0 - AA^+b) + Az_2 = 0 + 0 = 0$.
Thus $x_0 - A^+b + z_2 \in \mathcal{N}(A)$. Hence $\widehat{x} \in A^+b + \mathcal{N}(A)$.

Comment: In terms of expressing the collection of all LSSs to $Ax = b$, the role of A^+b can be replaced by any LSS x_0 to $Ax = b$. But A^+b is special in the sense of (2) below.

(2) Minimum norm LSS
For $Ax = b$, A^+b is the minimum norm LSS.

Proof First $A^+b = A^+b + 0 \in A^+b + \mathcal{N}(A)$. So A^+b is a LSS to $Ax = b$.
Secondly for LSS $\widehat{x}$ to $Ax = b$, $\widehat{x} \in A^+b + \mathcal{N}(A)$. So $\widehat{x} = A^+b + z$ where $z \in \mathcal{N}(A)$.
But $A^+b \in \mathcal{R}(A^+) = \mathcal{N}^\perp((A^+)^*) = \mathcal{N}^\perp(A)$. Thus $A^+b \perp z$.
By Pythagorean theorem

$$\|\widehat{x}\|^2 = \|A^+b + z\|^2 = \|A^+b\|^2 + \|z\|^2 \geq \|A^+b\|^2.$$

So A^+b is the minimum norm LSS to $Ax = b$.

2. Linear space restricted LSSs

(1) Definition and iff conditions
$A \in C^{m\times n}$ and $\mathcal{S}$ is a subspace in C^n.
$\widehat{x}$ is a LSS to $Ax = b$ restricted by $x \in \mathcal{S}$

$$\overset{def}{\Longleftrightarrow} \widehat{x} \in \mathcal{S} \text{ and } \|b - A\widehat{x}\|^2 \leq \|b - Ax\|^2 \text{ for all } x \in \mathcal{S}$$
$$\Longleftrightarrow \widehat{x} \in \mathcal{S} \text{ and } A\widehat{x} = \pi(b \mid A\mathcal{S}).$$

Comment: $A\mathcal{S} = \{As : s \in \mathcal{S}\}$ is a linear space.

(2) LSSs to $Ax = b$ under $Bx = 0$
The collection of all LSSs to $Ax = b$ under $Bx = 0$ is $[A(I - B^+B)]^+b + \mathcal{N}(A) \cap \mathcal{N}(B)$.

Proof $Bx = 0 \Longleftrightarrow x \in \mathcal{N}(B) = \mathcal{R}(I - B^+B) = \{(I - B^+B)h : h \in C^n\} = \mathcal{S}$.
So $A\mathcal{S} = A\mathcal{R}(I - B^+B) = \mathcal{R}(A(I - B^+B))$. Thus

$$\begin{aligned}
& \widehat{x} \text{ is a LSS to } Ax = b \text{ under } Bx = 0 \\
\Longleftrightarrow\ & \widehat{x} \in \mathcal{N}(B) \text{ and } A\widehat{x} = \pi(y|\mathcal{R}(A(I - B^+B))) = A[A(I - B^+B)]^+b \\
\Longleftrightarrow\ & \widehat{x} \in \mathcal{N}(B) \text{ and } A\{\widehat{x} - [A(I - B^+B)]^+b\} = 0 \\
\Longleftrightarrow\ & \widehat{x} \in \mathcal{N}(B) \text{ and } \widehat{x} - [A(I - B^+B)]^+b \in \mathcal{N}(A) \\
\Longleftrightarrow\ & \widehat{x} \in \mathcal{N}(B) \text{ and } \widehat{x} \in [A(I - B^+B)]^+b + \mathcal{N}(A) \\
\Longleftrightarrow\ & \widehat{x} \in \{[A(I - B^+B)]^+b + \mathcal{N}(A)\} \cap \mathcal{N}(B) \\
\Longleftrightarrow\ & \widehat{x} \in [A(I - B^+B)]^+b + \mathcal{N}(A) \cap \mathcal{N}(B).
\end{aligned}$$

Ex2: With $B = 0$, $\mathcal{N}(B) = \mathcal{R}(I - B^+B) = \mathcal{R}(I_n) = C^n$.
With $B = 0$, there is no restriction on x, $A(I - B^+B) = A$ and $\mathcal{N}(A) \cap \mathcal{N}(B) = \mathcal{N}(A)$.
Thus $[A(I - B^+B)]^+b + \mathcal{N}(A) \cap \mathcal{N}(B) = A^+b + \mathcal{N}(A)$.

L18: Matrix spaces

1. $\mathcal{R}(A,\, B)$

   (1) A linear transformation
   With $A \in C^{m\times p}$ and $B \in C^{q\times n}$, $Y = f(X) = AXB$ is a linear transformation from $X \in C^{p\times q}$ to $Y \in C^{m\times n}$.
   Proof $f(\alpha X + \beta Z) = A(\alpha X + \beta Z)B = \alpha(AXB) + \beta(AZB) = \alpha f(X) + \beta f(Z)$.

   (2) Range $\mathcal{R}(A,\, B)$
   The range of the LT in (1)

   $$\mathcal{R}(A,\, B) = \{AXB \in C^{m\times n} :\, X \in C^{p\times q}\}$$

   is a subspace of $C^{m\times n}$.

   Proof $Y_1,\, Y_2 \in \mathcal{R}(A,\, B) \Longrightarrow Y_1 = AX_1B$ and $Y_2 = AX_2B$
   $\Longrightarrow \alpha Y_1 + \beta Y_2 = A(\alpha X_1 + \beta X_2)B \in \mathcal{R}(A,\, B)$. So $\mathcal{R}(A,\, B)$ is a subspace of $C^{m\times n}$.

   (3) $\mathcal{R}(A,\, B) = \mathcal{R}(A,\, I_n) \cap \mathcal{R}(I_m,\, B)$.
   Proof $\subset$**:** $Y \in \mathcal{R}(A,\, B) \Longrightarrow Y = AXB = I_m(AX)B \in \mathcal{R}(I_m\, B)$
   $Y \in \mathcal{R}(A,\, B) \Longrightarrow Y = AXB = A(XB)I_n \in \mathcal{R}(A,\, I_n)$.
   So $Y \in \mathcal{R}(A,\, I_n) \cap \mathcal{R}(I_m,\, B)$.
   $\supset$**:** $Y \in \mathcal{R}(A,\, I_n) \cap \mathcal{R}(I_m,\, B) \Longrightarrow Y = AX_1I_n = I_mX_2B$.
   So $Y = AA^+AX_1I_n = AA^+I_mX_2B \in \mathcal{R}(A,\, B)$.
   Ex1: With $A \in C^{m\times n}$, $\mathcal{R}(A) = \{Ax :\, x \in C^n\} = \{Ax1 :\, x \in C^{n\times 1}\} = \mathcal{R}(A,\, 1)$.
   Thus vector space $\mathcal{R}(A)$ is a special case of $\mathcal{R}(A,\, B)$.

2. $\mathcal{N}(A,\, B)$

   (1) Kernel $\mathcal{N}(A,\, B)$.
   For the LT in (1) of 1, the Kernel

   $$\mathcal{N}(A,\, B) = \{X \in C^{p\times q} :\, AXB = 0\}$$

   is closed under LCs and hence is a subspace of $C^{p\times q}$, also called the null space associated with A and B.

   (2) $\mathcal{N}(A,\, B) = \mathcal{N}(A,\, I_q) + \mathcal{N}(I_p,\, B)$
   Proof $\subset$**:** If $X \in \mathcal{N}(A,\, B)$, then $AXB = 0$. So

   $$X = (I - A^+A)X + A^+AX(I - BB^+).$$

   Here $(I - A^+A)X \in \mathcal{N}(A,\, I_q)$ since $A[(I - A^+A)X]I_q = 0$; and
   $A^+AX(I - BB^+) \in \mathcal{N}(I_p,\, B)$ since $I_p[A^+AX(I - BB^+)]B = 0$.
   Thus $X \in \mathcal{N}(A,\, I_q) + \mathcal{N}(I_p,\, B)$.
   $\supset$**:** If $X \in \mathcal{N}(A,\, I_q) + \mathcal{N}(I_p,\, B)$, then $X = X_1 + X_2$ where $X_1 \in \mathcal{N}(A,\, I_q)$ and $X_2 \in \mathcal{N}(I_p,\, B)$.
   Thus $AXB = A(X_1B)I_q + I_p(AX_2)B = 0 + 0 = 0$. So $X \in \mathcal{N}(A,\, B)$.
   Ex2: With $A \in C^{m\times n}$,
   $\mathcal{N}(A) = \{x \in C^n :\, Ax = 0\} = \{x \in C^{n\times 1} :\, Ax1 = 0\} = \mathcal{N}(A,\, 1)$.
   Thus vector space $\mathcal{N}(A)$ is a special case of $\mathcal{N}(A,\, B)$.

3. Equivalent expressions

(1) In $\mathcal{R}(A,\, B)$

(i(A can be replaced by AA^-, AA^+, AA^*, $(A^+)^*$

(ii) B can be replaced by B^-B, B^+B, B^*B, $(B^+)^*$

(iii) A and B can be simultaneously replaced

Proof (i) We show $\mathcal{R}(A,\, B) = \mathcal{R}(AA^-,\, B)$.
$\subset$: $Y \in \mathcal{R}(A,\, B) \Longrightarrow Y = AXB = AA^-(AX)B \in \mathcal{R}(AA^-,\, B)$.
$\supset$: $Y \in \mathcal{R}(AA^-,\, B) \Longrightarrow Y = AA^+XB \in \mathcal{R}(A,\, B)$.

(ii) We show $\mathcal{R}(A,\, B) = \mathcal{R}(A,\, B^-B)$.
$\subset$: $Y \in \mathcal{R}(A,\, B) \Longrightarrow Y = AXB = A(XB)B^-B \in \mathcal{R}(A,\, B^-B)$.
$\supset$: $Y \in \mathcal{R}(A,\, B^-B) \Longrightarrow Y = A(XB^-)B \in \mathcal{R}(A,\, B)$.

(iii) We show $\mathcal{R}(A,\, B) = \mathcal{R}(AA^-,\, B^-B)$
$\mathcal{R}(A,\, B) = \mathcal{R}(AA^-,\, B) = \mathcal{R}(AA^-,\, B^-B)$.

(2) In $\mathcal{N}(A,\, B)$

(i(A can be replaced by A^-A, A^+A, A^*A, $(A^+)^*$

(ii) B can be replaced by BB^-, BB^+, BB^*, $(B^+)^*$

(iii) A and B can be simultaneously replaced

Proof (i) We show $\mathcal{N}(A,\, B) = \mathcal{R}(A^-A,\, B)$.
$\subset$: $X \in \mathcal{N}(A,\, B) \Longrightarrow AXB = 0 \Longrightarrow A^-AXB = 0 \Longrightarrow X \in \mathcal{N}(A^-A,\, B)$.
$\supset$: $X \in \mathcal{N}(A^-A,\, B) \Longrightarrow A^-AXB = 0 \Longrightarrow AXB = 0 \Longrightarrow X \in \mathcal{N}(A,\, B)$.

(ii) We show $\mathcal{N}(A,\, B) = \mathcal{N}(A,\, BB^-)$.
$\subset$: $X \in \mathcal{N}(A,\, B) \Longrightarrow AXB = 0 \Longrightarrow A(XB)B^- = 0 \Longrightarrow X \in \mathcal{N}(A,\, BB^-B)$.
$\supset$: $X \in \mathcal{N}(A,\, BB^-) \Longrightarrow AXBB^- = 0 \Longrightarrow AXB = 0 \Longrightarrow X \in \mathcal{N}(A,\, B)$.

(iii) We show $\mathcal{N}(A,\, B) = \mathcal{R}(A^-A,\, BB^-)$
$\mathcal{N}(A,\, B) = \mathcal{N}(A^-A,\, B) = \mathcal{N}(A^-A,\, BB^-)$.

Ex3: $\mathcal{R}(AD_1,\, D_2B) \subset \mathcal{R}(A,\, B)$ since
$Y \in \mathcal{R}(AD_1,\, D_2B) \Longrightarrow Y = A(D_1XD_2)B \in \mathcal{R}(A,\, B)$.
Thus $\mathcal{R}(A,\, B) = \mathcal{R}(AA^-A,\, BB^-B) \subset \mathcal{R}(AA^-,\, B^-B) \subset \mathcal{R}(A,\, B)$.
So $\mathcal{R}(A,\, B) = \mathcal{R}(AA^-,\, B^-B)$.

Ex4: $\mathcal{N}(A,\, B) \subset \mathcal{N}(D_1A,\, BD_2)$ since
$X \in \mathcal{N}(A,\, B) \Longrightarrow AXB = 0 \Longrightarrow D_1AXBD_2 = 0 \Longrightarrow X \in \mathcal{N}(D_1A,\, BD_2)$.
Thus $\mathcal{N}(A,\, B) \subset \mathcal{N}(A^-A,\, BB^-) \subset \mathcal{N}(AA^-A,\, BB^-B) = \mathcal{N}(A,\, B)$.
So $\mathcal{N}(A,\, B) = \mathcal{N}(A^-A,\, BB^-)$.

Ex5: $\mathcal{R}(A) = \mathcal{R}(A,\, 1) = \mathcal{R}((A^+)^*,\, 1) = \mathcal{R}((A^+)^*)$.
$\mathcal{N}(A) = \mathcal{N}(A,\, 1) = \mathcal{N}(A^*A,\, 1) = \mathcal{N}(A^*A)$