L08 Projection matrix and least square solutions

1. Projection matrix

   (1) Definition
   $P \in R^{n\times n}$ is a projection matrix if $\pi(x \mid \mathcal{R}(P)) = Px$ for all $x \in R^n$.

   (2) Sufficient and necessary condition
   P is a projection matrix $\Longleftrightarrow P' = P = P^2$.

   Proof. **$\Rightarrow$:** If P is a projection matrix, then $\pi(x \mid \mathcal{R}(P)) = Px$ for all x.
   But $\pi(x \mid \mathcal{R}(P)) = PP^+x$. So $PP^+x = Px$ for all x. Thus $P = PP^+$.
   Hence $P' = P = P^2$.
   $\Leftarrow$: If $P' = P = P^2$, then $P^+ = P$ and $PP^+ = PP = P$.
   So $\pi(x \mid \mathcal{R}(P)) = PP^+x = PPx = Px$ for all x.
   Thus P is a projection matrix.

   (3) Properties
   If p is a projection matrix, then Px is the projection of x onto
   $$\mathcal{R}(P) = \mathcal{N}^\perp(P) = \mathcal{N}(I-P) = \mathcal{R}^\perp(I-P).$$
   Proof When $P' = P = P^2$, $(I-P)' = I-P = (I-P)^2$. So $\mathcal{R}(P) = \mathcal{N}^\perp(P)$ and $\mathcal{R}(P) = \mathcal{N}(I-P) = \mathcal{R}^\perp(I-P)$.

   (4) Comments
   AA^+, A^+A, $I-AA^+$ and $I-A^+A$ are four real symmetric and idempotent matrices that produce projection matrices onto eight spaces: $\mathcal{R}(A)$, $\mathcal{R}(A')$, $\mathcal{R}^\perp(A)$, $\mathcal{R}^\perp(A')$, $\mathcal{N}(A)$, $\mathcal{N}(A')$, $\mathcal{N}^\perp(A)$ and $\mathcal{N}^\perp(A')$.

   Ex1: A^+A is the projection matrix onto four spaces. Find these four spaces.
   $\mathcal{R}(A^+A) = \mathcal{R}\left((A')(A')^+\right) = \mathcal{R}(A')$; $\mathcal{R}(A^+A) = \mathcal{N}^\perp(A^+A) = \mathcal{N}^\perp(A)$;
   $\mathcal{R}(A^+A) = \mathcal{N}(I-A^+A) = \mathcal{R}^\perp(I-A^+A)$. Thus the four spaces are
   $$\mathcal{R}(A') = \mathcal{N}^\perp(A) = \mathcal{R}^\perp(I-A^+A) = \mathcal{N}(I-A^+A).$$

   Ex2: Find projection matrix onto $\pi(x|\mathcal{N}(A))$.
   $\mathcal{N}(A) = \mathcal{N}(A^+A) = \mathcal{R}(I-A^+A)$. So the projection matrix onto $\mathcal{N}(A)$ is $I-A^+A$.

   Ex3: Find projection matrix onto $A\mathcal{R}(B)$.
   $A\mathcal{R}(B) = \{Ay : y \in \mathcal{R}(B)\} = \{ABx : x\} = \mathcal{R}(AB)$.
   Thus the projection matrix onto $A\mathcal{R}(B) = \mathcal{R}(AB)$ is $(AB)(AB)^+$.

   Ex4: Find projection matrix onto $\pi(x|A\mathcal{N}(B))$.
   $A\mathcal{N}(B) = A\mathcal{R}(I-B^+B) = \mathcal{R}\left(A(I-B^+B)\right)$. So the projection matrix onto $A\mathcal{N}(B) = \mathcal{R}\left(A(I-B^+B)\right)$ is $[A(I-B^+B)][A(I-B^+B)]^+$.

2. Least square solutions

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

   (2) Traditional solutions
   $Ax = b$ is consistent $\overset{def}{\Longleftrightarrow} \exists\, x_0$ such that $Ax_0 = b$
   Suppose $Ax = b$ is consistent. Then $\widehat{x}$ is a solution to $Ax = b \overset{def}{\Longleftrightarrow} A\widehat{x} = b$.

(3) Relations

Suppose $Ax = b$ is consistent with a solution x_0. Then

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

Proof $\Rightarrow$**:** If $\widehat{x}$ is a LSS to $Ax = b$, then $\|b - A\widehat{x}\|^2 \leq \|b - Ax_0\|^2 = 0$.
So $b - A\widehat{x} = 0$, i.e., $A\widehat{x} = b$. Thus $\widehat{x}$ is a solution to $Ax = b$.

$\Leftarrow$**:** If $\widehat{x}$ is a solution to $Ax = b$, then $\|b - A\widehat{x}\|^2 = 0 \leq \|b - Ax\|^2$ for all x.
Thus $\widehat{x}$ is a LSS to $Ax = b$.

Comment: Any conclusions on LSS will hold for solutions for a consistent equation.

.

(4) The collection of all LSSs

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

Proof $\widehat{x} \in \text{LSS}(Ax = b) \iff \|A\widehat{x} - b\|^2 \leq \|Ax - b\|^2$ for all x

$\iff A\widehat{x} = \pi(b|\mathcal{R}(A)) = AA^+b \Longleftrightarrow A(\widehat{x} - A^+b) = 0$

$\iff \widehat{x} - A^+b \in \mathcal{N}(A) \Longleftrightarrow \widehat{x} \in A^+b + \mathcal{N}(A)$.

Comment: For consistent $Ax = b$, the collection of all solutions is $A^+b + \mathcal{N}(A)$.

3. Structure of the LSS$(Ax = b)$

(1) An affine set

Set $\mathcal{A}$ in LS V is an affine set $\overset{def}{\Longleftrightarrow}$ $\mathcal{A}$ is closed under affine combination

$\overset{def}{\Longleftrightarrow}$ $\alpha x + (1 - \alpha)y \in \mathcal{A}$ for all x, $y \in \mathcal{A}$ and $\alpha \in R$

Comment: A convex combination is an affine combination.
An affine set is a convex set.

(2) A linear space

Suppose $\mathcal{A}$ is an affine set. Then $\mathcal{A}$ is a space $\Longleftrightarrow 0 \in \mathcal{A}$.

Proof $\Rightarrow$**:** Trivial

$\Leftarrow$**:** $x \in \mathcal{A} \Longrightarrow \alpha x = \alpha x + (1 - \alpha)0 \in \mathcal{A}$.
x, $y \in \mathcal{A} \Longrightarrow x + y = 2\left(\frac{1}{2}x + \frac{1}{2}y\right) \in \mathcal{A}$. So $\mathcal{A}$ is a subspace.

(3) Relation

$\mathcal{A}$ is an affine set $\Longleftrightarrow A = x_0 + S$ where S is a LS

Proof $\Rightarrow$**:** With $x_0 \in \mathcal{A}$, $\mathcal{A} = x_0 + (\mathcal{A} - x_0)$. We show that $S = \mathcal{A} - x_0$ is a space. First $0 = x_0 - x_0 \in \mathcal{A} - x_0 = S$.
Secondly with $x - x_0$, $y - x_0 \in \mathcal{A} - x_0 = S$,

$$\alpha(x - x_0) + (1 - \alpha(y - x_0) = (\alpha x + (1 - \alpha)y) - x_0 \in \mathcal{A} - x_0 = S.$$

So S is an affine set containing 0 and hence is a subspace.

$\Leftarrow$**:** For $x_0 + s_1$, $x_0 + s_2 \in x_0 + S = \mathcal{A}$,

$$\alpha(x_0 + s_1) + (1 - \alpha)(x_0 + s_2) = x_0 + [\alpha s_1 + (1 - \alpha)s_2] \in x_0 + S = \mathcal{A}.$$

So $\mathcal{A}$ is an affine set.

(4) Structure of LSS$(Ax = b)$

LSS$(Ax = b) = A^+b + \mathcal{N}(A)$ is an affine set where $A^+b \in \text{LSS}(Ax = b)$.

Proof LSS$(Ax = b)$ has $x_0 + S$ form.

L09 Restricted least square solutions

1. $(ABC)^+$

   (1) $(ABC)^+ = C^+C(ABC)^+AA^+$

   Proof Let $M = ABC$ and $G = C^+C(ABC)AA^+$. Then

   $$\begin{aligned} \text{(i) } MGM &= (ABC)[C^+C(ABC)^+AA^+](ABC) = (ABC)(ABC)^+(ABC) \\ &= ABC = M \\ \text{(ii) } GMG &= [C^+C(ABC)^+AA^+](ABC)[C^+C(ABC)^+AA^+] \\ &= C^+C(ABC)^+(ABC)(ABC)^+AA^+ = C^+C(ABC)^+AA^+ \\ &= G \end{aligned}$$

   (iii) $MG = (ABC)[C^+C(ABC)^+AA^+] = (ABC)(ABC)^+AA^+$.
   But $AA^+(ABC)(ABC)^+ = MM^+ \Longrightarrow (ABC)(ABC)^+AA^+ = MM^+$.
   Thus $MG = MM^+$ is symmetric.

   (iv) $GM = [C^+C(ABC)^+AA^+](ABC) = C^+C(ABC)^+(ABC)$.
   But $(ABC)^+(ABC)C^+C = M^+M \Longrightarrow C^+C(ABC)^+(ABC) = M^+M$.
   Thus $GM = M^+M$ is symmetric.

   Hence $G = M^+$.

   (2) $(ABC)^+ = C^+C(ABC)^+$ and $(ABC)^+ = (ABC)^+AA^+$

   Proof $(ABC)^+ = [I(AB)C]^+ = C^+C(ABC)^+II^+ = C^+C(ABC)^+$.
   $(ABC)^+ = [A(BC)I]^+ = I^+I(ABC)^+AA^+ = (ABC)^+AA^+$.

   (3) $(AB)^+ = B^+B(AB)^+AA^+ = B^+B(AB)^+ = (AB)^+AA^+$

   Proof $(AB)^+ = (AIB)^+$; $(AB)^+ = (IAB)^+$; $(AB)^+ = (ABI)^+$.

   (4) The projection matrix onto

   $$A\mathcal{N}(A) = A\mathcal{R}(I - B^+B) = \mathcal{R}[A(I - B^+B)] \text{ is}$$

   $[A(I - B^+B)][A(I - B^+B)]^+ = A[A(I - B^+B)]^+$.

2. Restricted least square solutions
   $\mathcal{D}$ is a closed convex set in LS V.

   (1) Definition: Restricted least square solutions

   $$\begin{aligned} &\widehat{x} \text{ is a restricted least square soution to } Ax = b \text{ under } x \in \mathcal{D} \\ \overset{def}{\Longleftrightarrow}\ &\widehat{x} \in \mathcal{D} \text{ and } \|b - A\widehat{x}\|^2 \le \|b - Ax\|^2 \text{ for all } x \in \mathcal{D} \end{aligned}$$

   (2) Definition: Projection onto closed convex set
   For $x \in V$ there exists a unique $\widehat{x} \in \mathcal{D}$ such that

   $$\|x - \widehat{x}\|^2 \le \|x - y\|^2 \text{ for all } y \in \mathcal{D}.$$

   This $\widehat{x}$ is called the projection of x onto $\mathcal{D}$ and is denoted as $\pi(x \mid \mathcal{D})$.

   (3) Theorem: A relation
   $\widehat{x}$ is a LSS to $Ax = b$ under the restriction $x \in \mathcal{D} \Longleftrightarrow A\widehat{x} = \pi(b \mid A\mathcal{D})$.

   Proof

   $$\begin{aligned} &\widehat{x} \text{ is a restricted LSS to } Ax = b \text{ under } x \in \mathcal{D} \\ \overset{def}{\Longleftrightarrow}\ &\widehat{x} \in \mathcal{D} \text{ and } \|b - A\widehat{x}\|^2 \le \|b - Ax\|^2 \text{ for all } x \in \mathcal{D} \\ \Longleftrightarrow\ &A\widehat{x} = \pi(b \mid A\mathcal{D}). \end{aligned}$$

(4) Example: LSS under $Bx = 0$

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

Proof $\quad \widehat{x}$ is a LSS to $Ax = b$ under consistent $Bx = 0$

$\iff \widehat{x}$ is a LSS to $Ax = b$ under $x \in \mathcal{N}(B)$

$\iff A\widehat{x} = \pi(b \mid A\mathcal{N}(A)) = A[A(I - B^+B)]^+b$

$\iff A\{\widehat{x} - [A(I - B^+B)]^+b\} = 0 \iff \widehat{x} \in [A(I - B^+B)]^+b + \mathcal{N}(A)$.

3. Restricted LSS under affine set restriction

(1) Projection onto an affine set

$\mathcal{A} = x_0 + S$ is an affine set in V where S is a subspace of V. For $x \in V$,

$$\pi(x \mid \mathcal{A}) = x_0 + \pi(x - x_0 \mid S).$$

Proof By the definition of $\pi(x \mid \mathcal{A})$,

$$\widehat{x} = \pi(x \mid \mathcal{A}) \iff \widehat{x} \in \mathcal{A} \text{ and } \|x - \widehat{x}\|^2 \le \|x - y\|^2 \text{ for all } y \in \mathcal{A}.$$

$\widehat{x} \in \mathcal{A} = x_0 + S \Longrightarrow \widehat{x} = x_0 + s_*$ where $s_* \in S$.

$y \in \mathcal{A} = x_0 + S \Longrightarrow y = x_0 + s$ for all $s \in S$. So

$$\begin{aligned} \widehat{x} = \pi(x \mid \mathcal{A}) \quad &\iff \quad \widehat{x} = x_0 + s_* \text{ where } x_* \in S \text{ and} \\ &\qquad\ \|x - (x_0 + s_*)\|^2 \le \|x - (x_0 + s)\|^2 \text{ for all } s \in S \\ &\iff \quad \widehat{x} = x_0 + s_* \text{ where } s_* \in S \text{ and} \\ &\qquad\ \|(x - x_0) - s_*\|^2 \le \|(x - x_0) - s\|^2 \text{ for all } s \in S \\ &\iff \quad \widehat{x} = x_0 + s_* \text{ where } s_* = \pi(x - x_0 \mid S) \\ &\iff \quad \widehat{x} = x_0 + \pi(x - x_0 \mid S). \end{aligned}$$

(2) LSS under affine set restriction

$\widehat{x}$ is a LSS to $Ax = b$ under $x \in x_0 + S \iff A\widehat{x} = Ax_0 + (b - Ax_0 \mid AS)$

Proof $\quad \widehat{x}$ is a LSS to $Ax = b$ under $x \in \mathcal{D} = \mathcal{A} = x_0 + S$

$\iff A\widehat{x} = \pi(b \mid A\mathcal{D}) = \pi(b \mid Ax_0 + AS) = Ax_0 + \pi(b - Ax_0 \mid AS)$.

(3) Theorem

The collection of all LSSs to $Ax = b$ under $x \in x_0 + \mathcal{R}(B)$ is $x_0 + B(AB)^+(b - Ax_0) + \mathcal{N}(A)$.

Proof $\quad \widehat{x}$ is a LSS to $Ax = b$ under $x \in x_0 + S = x_0 + \mathcal{R}(B)$

$\iff A\widehat{x} = Ax_0 + \pi(b - Ax_0 \mid \mathrm{AR}(B)) = Ax_0 + (AB)(AB)^+(b - Ax_0)$

$\iff A[\widehat{x} - x_0 - B(AB)^+(b - Ax_0)] = 0$

$\iff \widehat{x} - x_0 - B(AB)^+(b - Ax_0) \in \mathcal{N}(A)$

$\iff \widehat{x} \in x_0 + B(AB)^+(b - Ax_0) + \mathcal{N}(A)$.

(4) Example: The collection of all LSSs to $Ax = b$ under consistent $Bx = c$ is

$$B^+c + [A(I - B^+B)]^+(b - AB^+c) + \mathcal{N}(A).$$

Proof For consistent $Bx = c$, $Bx = c \iff x \in \mathcal{D} = B^+c + \mathcal{N}(B) = B^+c + \mathcal{R}(I - B^+B)$.

So $\quad \widehat{x}$ is a LSS to $Ax = b$ under consistent $Bx = c$

$\iff \widehat{x} \in B^+c + (I - B^+B)[A(I - B^+B)]^+(b - AB^+c) + \mathcal{N}(A)$

$\iff \widehat{x} \in B^+c + [A(I - B^+B)]^+(b - AB^+c) + \mathcal{N}(A)$.

Ex: With $c = 0$, the collection of LSSs to $Ax = b$ under $Bx = 0$ is

$$[A(I - B^+B)]b + \mathcal{N}(A).$$