L07 Vector projections

1. Projections

   (1) Minimum-distance projection
   S is a subspace of V. For $y \in V$ there exists a unique $\widehat{y} \in S$ such that

   $$\|y - \widehat{y}\|^2 \leq \|y - z\|^2 \text{ for all } z \in S.$$

   This $\widehat{y}$ is called the minimum-distance projection of y onto S.

   (2) Orthogonal projection
   S is a subspace of V. For $y \in V$ there exists a unique $\widehat{y} \in S$ such that

   $$(y - \widehat{y}) \perp S, \text{ i.e., } \langle y - \widehat{y},\, z\rangle = 0 \text{ for all } z \in S.$$

   This $\widehat{y}$ is called the orthogonal projection of y onto S.

   (3) Projection of y onto S
   Minimum-distance projection and orthogonal projection of y onto S are equal, and is called the projection of y onto S denoted as $\pi(y|S)$.

   Proof If $\widehat{y}$ is an orthogonal projection onto S, then $\widehat{y} \in S$ and $(y - \widehat{y}) \perp S$. So for $z \in S$, $(y - \widehat{y}) \perp (\widehat{y} - z)$. Thus by Pythagorean theorem

   $$\|y - z\|^2 = \|y - \widehat{y} + \widehat{y} - z\|^2 = \|y - \widehat{y}\|^2 + \|\widehat{y} - z\|^2 \geq \|y - \widehat{y}\|^2.$$

   Thus $\widehat{y}$ is the minimum-distance projection of y onto S.

2. Projections onto $\mathcal{R}(A)$ and $\mathcal{R}^{\perp}(A)$

   (1) $\pi(y \mid \mathcal{R}(A)) = AA^{+}y$

   Proof With $A \in R^{m\times n}$ $\mathcal{R}(A) \subset R^m$. For $y \in R^m$, let $\widehat{y} = AA^{+}y$. Then

   $$y = AA^{+}y + (I - AA^{+})y = \widehat{y} + (y - \widehat{y})$$

   where $\widehat{y} = AA^{+}y \in \mathcal{R}(AA^{+}) = \mathcal{R}(A)$ and

   $$\begin{aligned} y - \widehat{y} &= (I - AA^{+})y \in \mathcal{R}(I - AA^{+}) = \mathcal{N}^{\perp}(I - AA^{+}) = [\mathcal{N}(I - AA^{+})]^{\perp} \\ &= [\mathcal{R}(AA^{+})]^{\perp} = [\mathcal{R}(A)]^{\perp} = \mathcal{R}^{\perp}(A). \end{aligned}$$

   Thus $(y - \widehat{y}) \perp \mathcal{R}(A)$. Hence $\widehat{y} = \pi(y \mid \mathcal{R}(A))$.

   Comment: AA^{+} is called the projection matrix onto $\mathcal{R}(A)$.

   (2) $\pi(y \mid \mathcal{R}^{\perp}(A)) = (I - AA^{+})y$

   Proof With $A \in R^{m\times n}$ $\mathcal{R}^{\perp}(A) \subset R^m$. For $y \in R^m$, let $\widehat{y} = (I - AA^{+})y$. Then

   $$y = AA^{+}y + (I - AA^{+})y = (y - \widehat{y}) + \widehat{y}$$

   where $\widehat{y} = (I - AA^{+})y \in \mathcal{R}^{\perp}(A)$ and $y - \widehat{y} = AA^{+}y \in \mathcal{R}(A)$. Thus $(y - \widehat{y}) \perp \mathcal{R}^{\perp}(A)$. Hence $\widehat{y} = \pi(y \mid \mathcal{R}^{\perp}(A))$.

   Comment: $I - AA^{+}$ is the projection matrix onto $\mathcal{R}(I - AA^{+}) = \mathcal{R}^{\perp}(A)$.
   $y = \pi(y \mid \mathcal{R}(A)) + \pi(y \mid \mathcal{R}^{\perp}(A))$

3. Projections onto $\mathcal{N}(A)$ and $\mathcal{N}^{\perp}(A)$

(1) $\pi(x|\mathcal{N}(A)) = (I - A^+A)x$

Proof With $A \in R^{m\times n}$ $\mathcal{N}(A) \in R^n$. For $x \in R^n$ let $\widehat{x} = (I - A^+A)x$. Then

$$x = (I - A^+A)x + A^+Ax = \widehat{x} + (x - \widehat{x})$$

where $\widehat{x} \in \mathcal{R}(I - A^+A) = \mathcal{N}(A^+A) = \mathcal{N}(A)$ and

$$x - \widehat{x} = A^+Ax \in \mathcal{R}(A^+A) = \left[\mathcal{N}(A^+A)\right]^{\perp} = [\mathcal{N}(A)]^{\perp} = \mathcal{N}^{\perp}(A).$$

So $(x - \widehat{x}) \perp \mathcal{N}(A)$. Hence $\widehat{x} = \pi(x \mid \mathcal{N}(A))$.

Comment: $I - A^+A$ is the projection matrix onto $\mathcal{N}(A) = \mathcal{R}(I - A^+A)$.

(2) $\pi(x \mid \mathcal{N}^{\perp}(A)) = A^+Ax$.

Proof With $A \in R^{m\times n}$, $\mathcal{N}^{\perp}(A) \subset R^n$. For $x \in R^n$ let $\widehat{x} = A^+Ax$. Then

$$x = (A^+Ax) + (I - A^+A)x = \widehat{x} + (x - \widehat{x})$$

where $\widehat{x} = A^+Ax \in \mathcal{N}^{\perp}(A)$ and $x - \widehat{x} \in \mathcal{N}(A)$. So $(x - \widehat{x}) \perp \mathcal{N}^{\perp}(A)$.
Hence $\widehat{x} = \pi(x \mid \mathcal{N}^{\perp}(A))$.

Comment: A^+A is the projection matrix onto $\mathcal{N}^{\perp}(A) = \mathcal{R}(A')$.
$x = \pi(x \mid \mathcal{N}(A)) + \pi(x \mid \mathcal{N}^{\perp}(A))$.

Ex: One estimate the solution of $Ax = 0$ by x_0. Then the error can be measured by $\min\left\{\|x_0 - x\|^2 : Ax = 0\right\}$. Let $\widehat{x}_0 = \pi(x_0 \mid \mathcal{N}(A))$. This error is

$$\|x_0 - \widehat{x}_0)\|^2 = \|x_0 - \pi(x_0 \mid \mathcal{N}(A))\|^2 = \|\pi(x_0 \mid \mathcal{N}^{\perp}(A)\|^2 = \|A^+Ax_0\|^2 = x_0'A^+Ax_0.$$

With given A and x_0, the problem becomes that of computing for A^+.
If A has full column rank, then A has a left-inverse and A^+ is one of its left-inverse. So $A^+A = I$ and $x_0'A^+Ax_0 = x_0'x_0 = \|x_0\|^2$. Note that under this assumption $\mathcal{N}(A) = \{0\}$ since $\dim[\mathcal{N}(A)] = n - \text{rank}(A) = n - n = 0$.
If A has full row rank, then $A^+ = A'(AA')^{-1}$ and $A^+A = A'(AA')^{-1}A$.
If A has orthonormal rows, then $A^+ = A'$.