Linear Algebra II
Pseudo Inverse

Ming-Hsuan Kang

Let T be a linear transformation from V to W. In general, the inverse of T may not exist. However, if we write $V = \ker(T) \oplus V_0$, then $T\big|_{V_0}$ is an isomorphism from V_0 to $\mathrm{Im}(T)$. Let $T_0 : V_0 \mapsto \mathrm{Im}(T)$ defined by $T_0(\vec{v}) = T(\vec{v})$. Then T_0 is invertible. To extend $(T_0)^{-1}$ to a linear transformation on W, let us write $W = \mathrm{Im}(T) \oplus W_0$ for some subspace W_0. Define a linear transformation $T^\dagger : W \mapsto V$ characterized as follows.

- For all $\vec{w} \in \mathrm{Im}(T)$, $T^\dagger(\vec{w}) = (T_0)^{-1}(\vec{w})$.
- For all $\vec{w} \in W_0$, $T^\dagger(\vec{w}) = \vec{0}$.

We call $T^\dagger$ a generalized inverse of T. Note that $T^\dagger$ is not unique and it depends on the choices of V_0 and W_0.

Pseudo Inverse

When V and W are inner product spaces, there are canonical choices of V_0 and W_0, namely

$$V_0 = \ker(T)^{\perp} \quad \text{and} \quad W_0 = \mathrm{Im}(T)^{\perp}.$$

In this case, the corresponding $T^{\dagger}$ is called the (Moore-Penrose) pseudo inverse of T. Let us rewrite the definition of the pseudo inverse.

Definition

Let V and W be two inner product spaces over F. Let T be a linear transformation from V to W. A linear transform $T^{\dagger}$ from W to V is the pseudo inverse of T if the following hold.

- For all $\vec{v} \in \ker(T)^{\perp}$, $T^{\dagger}T(\vec{v}) = \vec{v}$.
- For all $\vec{v} \in \mathrm{Im}(T)^{\perp}$, $T^{\dagger}(v) = \vec{0}$.

From the definition of pseudo inverse, we immediately have the following result.

Theorem

Let $T^\dagger$ be the pseudo inverse of T, then

- $T^\dagger T$ is the orthogonal projection onto $\ker(T)^\perp$.
- $TT^\dagger$ is the orthogonal projection onto $\mathrm{Im}(T)$.

For an $m \times n$ matrix A over F (where $F = \mathbb{R}$ or $\mathbb{C}$), L_A is a linear transformation from F^n to F^m. With respect to the standard inner products on F^n and F^m, there exists the pseudo inverse $L_{A^\dagger}$ of L_A where $A^\dagger$ is an $n \times m$ matrix. In this case, we also say $A^\dagger$ is the pseudo inverse of A.

Pseudo Inverse and SVD

Suppose A is real of rank r. Let $\lambda_1 \geq \cdots \geq \lambda_r > 0$ be the set of positive singular values, $V = (\vec{v}_1 \ \cdots \ \vec{v}_r)$ be the matrix of the first r left singular vectors, and $U = (\vec{u}_1 \ \cdots \ \vec{u}_r)$ be the corresponding matrix of the first r right singular vectors. Recall that we have the compact SVD

$$A = U\Sigma V^t = \sum_{i=1}^{r} \sqrt{\lambda_i}\vec{u}_i\vec{v}_i^t.$$

Here Σ is the diagonal matrix which (i,i)-th entry is $\sqrt{\lambda_i}$ for all i.

Theorem

The pseudo inverse $A = V(\Sigma)^{-1}U^t = \sum_{i=1}^{r} \frac{1}{\sqrt{\lambda_i}}\vec{v}_i\vec{u}_i^t$.

Recall that $\vec{v}_{r+1}, \cdots, \vec{v}_n$ are eigenvectors corresponding to the zero eigenvalue of A^tA and $\vec{u}_1, \cdots, \vec{u}_r$ span the image of L_A. Therefore,

$$\ker(L_A)^\perp = \ker(L_{A^tA})^\perp = \mathrm{span}\{\vec{v}_1, \cdots, \vec{v}_r\}$$

and

$$\mathrm{Im}(L_A)^\perp = \mathrm{span}\{\vec{u}_{r+1}, \cdots, \vec{u}_m\}.$$

Proof

Let $A' = \sum_{i=1}^{r} \frac{1}{\sqrt{\lambda_i}} \vec{v}_i \vec{u}_i^t$. Note that for $i = r+1, \cdots, m$, we have

$$L_{A'}(\vec{u}_i) = \left(\sum_{j=1}^{r} \frac{1}{\sqrt{\lambda_j}} \vec{v}_j \vec{u}_j^t \right) \vec{u}_i = \sum_{j=1}^{r} \frac{1}{\sqrt{\lambda_j}} \vec{v}_j (\vec{u}_j^t \vec{u}_i) = \vec{0}.$$

and for $i = 1, \cdots, r$, we have

$$L_{A'} L_A(\vec{v}_i) = L_{A'}(\sqrt{\lambda_i} \vec{u}_i) = \left(\sum_{j=1}^{r} \frac{1}{\sqrt{\lambda_j}} \vec{v}_j \vec{u}_j^t \right) \left(\sqrt{\lambda_i} \vec{u}_i \right) = \vec{v}_i.$$

Therefore $L_{A'}$ is the pseudo inverse of L_A and $A^\dagger = A'$.

Example

Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$, then $A^t A = \begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix}$ which eigenvalues are 6 and 0.
Let $\vec{v}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\vec{u}_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ be the first left and right singular vector. Then

$$A = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} (\sqrt{6}) \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \end{pmatrix}.$$

Therefore, its pseudo inverse is

$$A^\dagger = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} (\frac{1}{\sqrt{6}}) \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Moreover,

$$A^\dagger A = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad A^\dagger A = \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

which are orthogonal projections.

For a real system of linear equations $A\vec{x} = \vec{b}$, recall that $\vec{x}_0$ is a least square solution if

$$\vec{x}_0 \in \arg\min_{\vec{x}}\{\|A\vec{x} - \vec{b}\|^2\}.$$

Moreover, we have shown that the following are equivalent.

- $\vec{x}_0$ is a least square solution.
- $A\vec{x}_0 = \mathrm{proj}_W(\vec{b})$, where W is the column space of A.
- $A^t A\vec{x}_0 = A^t\vec{b}$.

As usual solutions, least square solutions may not be unique.

Theorem

Let $A^\dagger$ be the pseudo inverse of A. Then $\vec{x}_0 = A^\dagger \vec{b}$ is the unique least square solution of $A\vec{x} = \vec{b}$ of minimal norm. In other words, if $\vec{x}_1$ is another least square solution, then $\|x_0\| < \|x_1\|$.

Proof. Note that the column space of A is indeed the image of L_A. Let $L_{A^\dagger}$ be the pseudo inverse of L_A. Then $L_A L_{A^\dagger}$ is the orthogonal projection onto $\mathrm{Im}(L_A)$. Therefore,

$$A(\vec{x}_0) = AA^\dagger(\vec{b}) = L_A L_{A^\dagger}(\vec{b}) = \mathrm{proj}_W(\vec{b}).$$

We conclude that $\vec{x}_0$ is a least square solution. Since $A(\vec{x}_0 - \vec{x}_1) = \mathrm{proj}_W(\vec{b}) - \mathrm{proj}_W(\vec{b}) = \vec{0}$, we have

$$\vec{x}_1 = \vec{x}_0 + (\vec{x}_0 - \vec{x}_1) \in \ker(L_A)^\perp \oplus \ker(L_A).$$

Thus
$$\|\vec{x}_1\|^2 = \|\vec{x}_0\|^2 + \|\vec{x}_0 - \vec{x}_1\|^2 > \|\vec{x}_0\|.$$

Example

Find the least square solution of minimal norm of $\begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

In the previous example, we have shown that $\frac{1}{6}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ is the pseudo inverse of the coefficient matrix of the linear system. Thus

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

is the least square solution of minimal norm.

Other Inverse

Recall that for a linear transformation T on V, we have

$$V = \ker_\infty(T) \oplus \mathrm{Im}_\infty(T).$$

Since the restriction of T on $\mathrm{Im}_\infty(T)$ is invertible, one can define the so-called Drazin inverse from this decomposition.

Question: Let $A = PJP^{-1}$ be a complex square matrix where J is the Jordan form of A. Can you describe the Drazin inverse of A using its Jordan form?