The Moore-Penrose pseudo-inverse: theory, applications, and a generalization

Ben Wilson

Dr. Catherine Kublik

Definition/Introduction

The only matrices with inverses are square and nonsingular. It is however possible to generalize the notion of inverse to square-singular matrices and rectangular matrices. The Moore-Penrose pseudo-inverse is the most common generalized inverse. For the sake of simplicity, we will use real valued matrices.

Theorem 1: *Let* $A \in \mathcal{M}_{n,m}$, then there exists a unique matrix, $B \in \mathcal{M}_{mn}$, which satisfies the four Moore-Penrose conditions

1. $ABA = A$
2. $BAB = B$
3. $BA = (BA)^T$
4. $AB = (AB)^T$

We define the Moore-Penrose pseudo-inverse denoted as $A^\dagger$, as the unique matrix B.

Properties

Prop 1: *Let* $A \in \mathcal{M}_{n,m}, rank(A) = r, A$ can be written as $A = FR$, where $F \in \mathcal{M}_{n,r}, R \in \mathcal{M}_{r,m}$, and $rank(F) = rank(R) = r. F$ is constructed using the r linearly independent columns of A. Then,

$$A^\dagger = R^T(RR^T)^{-1}(F^TF)^{-1}F^T$$

Corollary 1: If A has full row rank then,

$$A^\dagger = A^T(AA^T)^{-1}$$

Similarly, if A has full column rank then,

$$A^\dagger = (A^TA)^{-1}A^T$$

Computation

The most standard method to compute the Moore-Penrose pseudo-inverse is the SVD decomposition. $A \in \mathcal{M}_{n,m}, rank(A) = r,$ then there exist U, V unitary matrices and Σ diagonal, such that $A = U\Sigma V^T$. To calculate $A^\dagger$, one must calculate $\Sigma^\dagger$. $\Sigma^\dagger$ is a diagonal matrix, where the diagonal is the reciprocal of the elements in the diagonal of Σ. Note that $\Sigma^\dagger\Sigma$ is a diagonal matrix with the first r diagonal values being 1 and the remaining are 0. So

$$A^\dagger = V\Sigma^\dagger U^T$$

Left and Right Inverse

If there is a left inverse for a matrix, than $A \in \mathcal{M}_{n,m}$ must have full column rank and $null(A) = \{0\}$, so A^TA has full rank, and is invertible. So $(A^TA)^{-1}A^TA = I$, thus one left inverse is

$$A^{-1}_{left} = (A^TA)^{-1}A^T = A^\dagger$$

Similarly, a right inverse occurs when A has full row rank and $null(A^T) = \{0\}$. So one right inverse is

$$A^{-1}_{right} = (A^TA)^{-1}A^T = A^\dagger$$

So if A has full column rank then $A^\dagger$ is a left inverse, and if A has full row rank than $A^\dagger$ is a right inverse. In general, $A^\dagger$ is neither a left or right inverse. Figure 1 shows the four subspaces of A. We can create a linear bijective function

$$T: C(A^T) \longrightarrow C(A): x \in C(A^T), T(x) = Ax$$

So then,

$$T^{-1}: C(A) \longrightarrow C(A^T): T(Ax) = A^\dagger Ax = x, \forall x \in C(A^T)$$

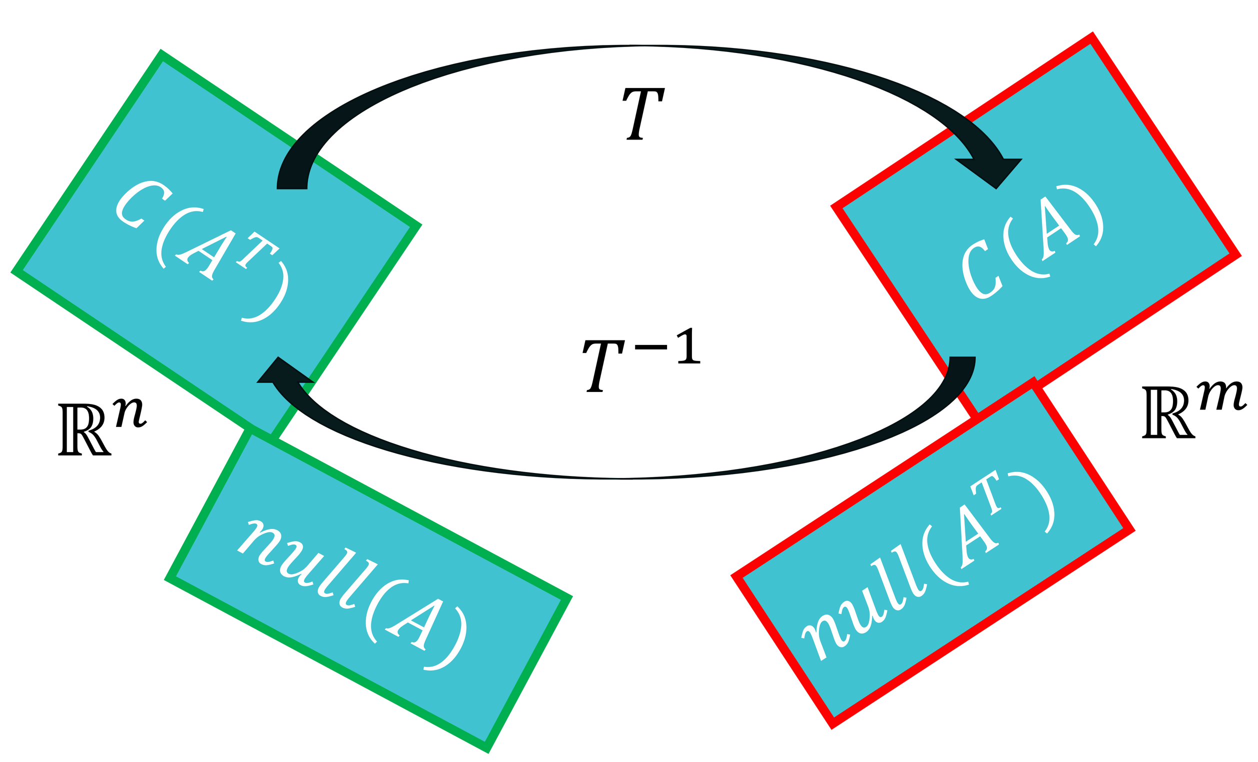

Fig. 1: Subspaces of A

Least Squares Problems

We look for the best solution to $Ax = b$, using $\min \parallel Ax - b \parallel, A \in \mathcal{M}_{n,m}, x \in \mathbb{R}^n$, where $\parallel \parallel$ is derived from the standard inner product on $\mathbb{R}^n$.

Theorem 2: $x_0 = A^\dagger b$ is the best approximate solution of $Ax = b$. With the standard inner product on $\mathbb{R}^n$.

The normal equations for standard inner product on $\mathbb{R}^n$ is

$$A^TAx = A^Tb$$

The least-squares solution satisfies the normal equation.

A Generalization of Least Squares

We generalize the least square to a general norm on $\mathbb{R}^n$ derived from a general inner product.

Theorem 3: $\langle x, y \rangle$ is an inner product on $\mathbb{R}^n$ if and only if $\langle x, y \rangle = x^TCy$, where C is a symmetric positive definite matrix.

A generalized normal equation can be found,

$$A^TCAx = A^TCb$$

The least-squares solution to our generalized least-squares problem now satisfies the generalized normal equations

Note that when $C = I_n$ The generalized normal equations reduce to the normal equations for the standard least-square problem.

From the generalized normal equations, we can see that if A has full column rank, then A^TCA is invertible and thus the solution of the generalized least square is, $(A^TCA)^{-1}A^TCb$.

Conjecture: When A has full column rank the generalized pseudo-inverse is,

$$A^\dagger = (A^TCA)^{-1}A^TC$$

Note than when $C = I_n$, we recover the standard formula for $A^\dagger$ in the case where A has full column rank.