Discussion 14A

1. Pseudoinverses

In this discussion, we will talk about the Moore-Penrose Pseudoinverse, which is a generalization of the idea of matrix inverses that apply to all matrices (regardless of dimension and rank) and allows us to solve a variety of problems we have explored in the past such as least squares and minimum norm problems.

Definition (Compact Moore-Penrose Pseudoinverse):

If the compact SVD of a matrix A is

$$A = U_r \Sigma_r V_r^\top \tag{1}$$

then the compact Moore-Penrose Pseudoinverse of A is

$$A^\dagger = V_r \Sigma_r^{-1} U_r^\top \tag{2}$$

Remember from our knowledge of the compact SVD that U_r and V_r are either square or tall matrices with $r = \text{rank}(A)$ orthonormal columns, and that Σ_r is a square and diagonal matrix with nonzero diagonal entries (and thus has a well-defined inverse Σ_r^{-1} which is also a square diagonal matrix, simply with the diagonal entries being the reciprocal of those in Σ_r).

(a) i. Find expressions for $A^\dagger A$ and $AA^\dagger$ using the SVD definitions for both matrices above and simplify as much as possible without any assumptions about $r = \text{rank}(A)$.

ii. The matrices from the previous part allow us to project a vector onto some specific space. Which space does each matrix ($A^\dagger A$ and $AA^\dagger$) project a vector onto?

(b) Let's examine the case where A is a square matrix $A \in \mathbb{R}^{n \times n}$ with full column rank ($r = \text{rank}(A) = n$) and thus full row rank as well. This means that A is invertible (A^{-1} exists). What are the dimensions of U_r and V_r in this case? How can we simplify $A^\dagger A$ and $AA^\dagger$ in this case? What is the relationship between A^{-1} and $A^\dagger$ in this case?

In the next parts, we will explore how the pseudoinverse relates to least squares and minimum norm problems and their solutions.

(c) In a least squares problem, the standard setup is that we have an overdetermined system that can be represented by the following matrix equation:

$$A\vec{x} = \vec{b} \tag{3}$$

where $A \in \mathbb{R}^{m\times n}$ is usually a tall matrix ($m > n$).

If A is full column rank (which is required for $A^\top A$ to be invertible), the least squares solution (minimizes error) is

$$\vec{x}^* = (A^\top A)^{-1}A^\top\vec{b} \tag{4}$$

Using the SVD of A, simplify the matrix $(A^\top A)^{-1}A^\top$ to be in terms of U_r, Σ_r, and V_r, and relate this to $A^\dagger$, the pseudoinverse of the matrix A. *(HINT: $r = \text{rank}(A) = n$ since A is full column rank so V_r is a square, orthonormal matrix ($V_r^{-1} = V_r^\top$) in this case.)*

(d) In the minimum norm problem, the standard setup is that we have an underdetermined system that can be represented by the same matrix equation:

$$A\vec{x} = \vec{b} \tag{5}$$

this time where $A \in \mathbb{R}^{m\times n}$ is usually a wide matrix ($m < n$).

In this case, if A is full row rank (which is required for $AA^\top$ to be invertible), the minimum norm solution (solution out of the infinite set of solutions that minimizes the norm of the input $\vec{x}$) is

$$\vec{x}^* = A^\top(AA^\top)^{-1}\vec{b} \tag{6}$$

Using the SVD of A, simplify the matrix $A^\top(AA^\top)^{-1}$ to be in terms of U_r, Σ_r, and V_r, and relate this to $A^\dagger$, the pseudoinverse of the matrix A. *(HINT: $r = \text{rank}(A) = m$ since A is full row rank so U_r is a square, orthonormal matrix ($U_r^{-1} = U_r^\top$) in this case.)*

Since the pseudoinverse is actually defined for every matrix (not just those with full column rank or row rank), we can calculate minimum norm solutions for all matrices (in the cases where they are relevant). In general, if we have a system $A\vec{x} = \vec{b}$ as in the previous part, for any matrix A, the minimum norm solution is $\vec{x}^* = A^\dagger\vec{b}$. Let's explore this with a numerical example.

(e) Suppose we have the following system:

$$A\vec{x} = \vec{b} \tag{7}$$

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \\ 0 \end{bmatrix} \tag{8}$$

In this case, we have an infinite number of possibilities for the vector $\vec{x}$ and we want to find the minimum norm solution for $\vec{x}$. A is not full row rank (and is also not full column rank) so the minimum norm solution formula from the previous part will not work. We can still use the pseudoinverse to solve this problem, which exists for all matrices.

i. Calculate the compact SVD $A = U_r \Sigma_r V_r^\top$. *(HINT: $r = \text{rank}(A) = 1$ so you can find this by inspection by thinking about the outer product form.)*

ii. Use the SVD to calculate $A^\dagger$, the pseudoinverse of A.

iii. Find the minimum norm solution $\vec{x}^*$ for the provided system. Does the answer make sense (think about the symmetry of the problem)?

Contributors:

- Nikhil Jain.
- Chancharik Mitra.

