Very short notes on the pseudo-inverse matrix

Herman Jaramillo Villegas
Universidad de Medellín

February 26, 2024

1 Theory

The book Machine Learning for Science and Engineering [1] has a detailed explanation of the pseudo-inverse of a matrix A. Here we make a great simplification with the purpose of only get the idea about what it is, with no computations.

We would like to find the inverse of any matrix A (in $\mathbb{C}^{m\times n}$) but this is only true for a very tiny set of matrices. Can we find something that approaches the inverse? What would be the best solution for the system $Ax = b$, in terms of a matrix A^+ (which we would call the pseudo-inverse). A good solution is that of the least squares problem. That is, the solution x such that $\|Ax - b\|^2$ is the smallest possible. Such a matrix is the pseudo-inverse.

We list a few cases on how to solve $Ax = b$.

- If $b \in \mathcal{R}(A)$ then the solution exists. We branch here two cases:
  - If $\mathcal{N} = \{0\}$ then A^{-1} exists and $x = A^{-1}b$ is the solution. The pseudo-inverse is given by $A^+ = A^{-1}$.
  - If $\mathcal{N} \neq \{0\}$ the solution is not unique, we should find the min $\|x\|$ such with $Ax = b$. $x = A^+b$.

    For example, find the pseudoinverse of $A = (2, 3)$. We know that the range of the matrix is the whole real line $\mathbb{R}$. That is either column vector 2 or the column vector 3 generate any real number. Let us pick, for example, $b = 5$ and solve $Ax = b$.

$$2x + 3y = 5.$$

This equation has an infinite number of solutions. The null space is given by any vector multiple of $x = -3, y = 2$. That is

$$\mathcal{N} = \alpha \begin{pmatrix} -3 \\ 2 \end{pmatrix}.$$

Figure 1 illustrates the null space, the solution space and the solution x which happens to be the smallest possible solution. It is easy to see that

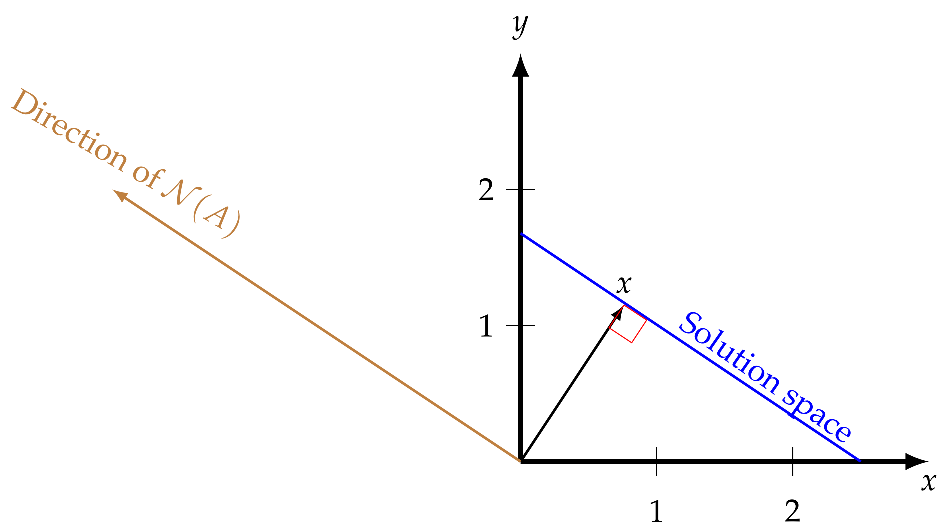

Figure 1: The solution space for the example with $A = (2, 3)$ and vector $b = 5$ is represented by the blue line. The null space $\mathcal{N}(A)$ is represented by the brown line. The smallest vector in the solution space is perpendicular to the null space and represents the solution to the least-squares problem.

the solution x is given by

$$\begin{pmatrix} 0.76923 \\ 1.15385 \end{pmatrix}.$$

We leave the proof of this to the reader. We also ask the reader to find the matrix A^+ such that $A^+ x = b$.

- If $b \notin \mathcal{R}(A)$ the solution does not exist. We should project b into $\mathcal{R}(A)$ by solving $A^T A x = A^T b$. This provides two cases:

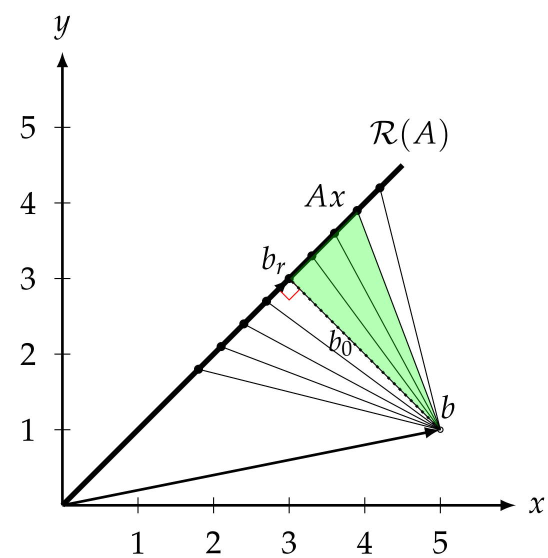

Figure 2: The green right triangle has side lengths $\|b_0\| = \|b - b_r\|$, $\|Ax - b\|$ (hypotenuse), and $\|Ax - b_r\|$ along the space $\mathcal{R}(A)$. Vector b_0 is the shortest of all vectors from b to $\mathcal{R}(A)$.

- If $\mathcal{N} = \{0\}$ the solution is unique and it is the least square solution. That is, $A^T Ax = A^T b$ and

$$x = (A^T A)^{-1} A^T b.$$

  The pseudo-inverse in this case is $A^+ = (A^T A)^{-1} A^T$.

- If $\mathcal{N} \neq \{0\}$ the solution is not unique. Solve the Regularized Least Square problem $(A^T A + \lambda I)x = A^T b$, that is,

$$x = (A^T A + \lambda I)^{-1} A^T b$$

  The pseudo-inverse is the matrix

$$A^+ = \lim_{\lambda \to 0} (A^T A + \lambda I)^{-1} A^T.$$

  Figure 3 shows the interpretation of the pseudo-inverse solution when $b \notin \mathcal{R}$ and $\mathcal{N} \neq \{0\}$.

Figure 4 illustrates this.

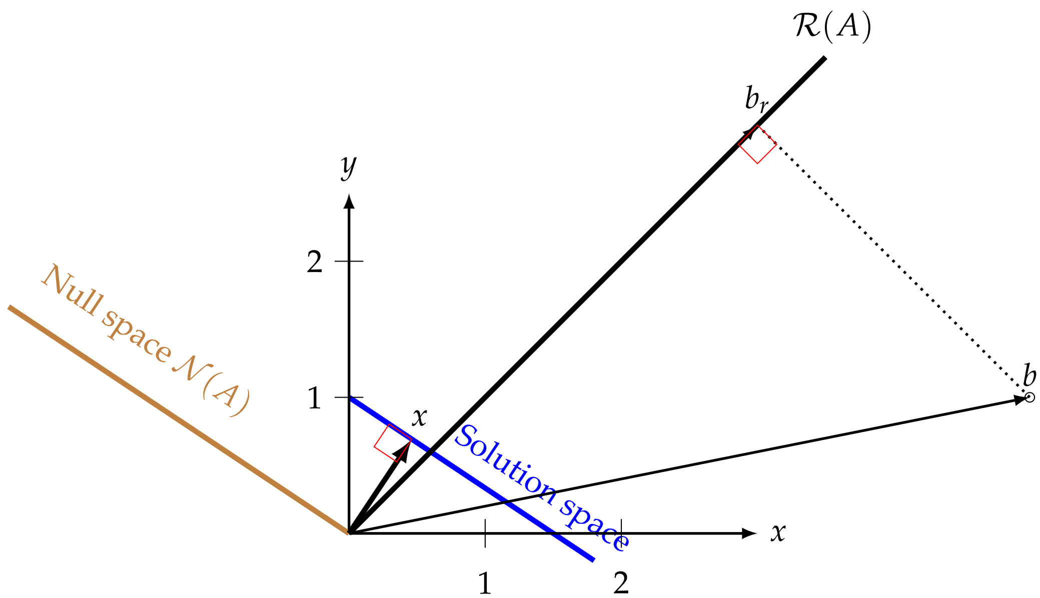

Figure 3: Geometrical interpretation of the pseudo-inverse: Point b is not in $\mathcal{R}(A)$, so $Ax = b$ does not have a solution. We project b into $\mathcal{R}(A)$ and call the projected vector b_r. The space solution of $Ax = b_r$ is the blue line. The null space is parallel to the space solution. The solution x is the shortest vector from the origin to the solution space.

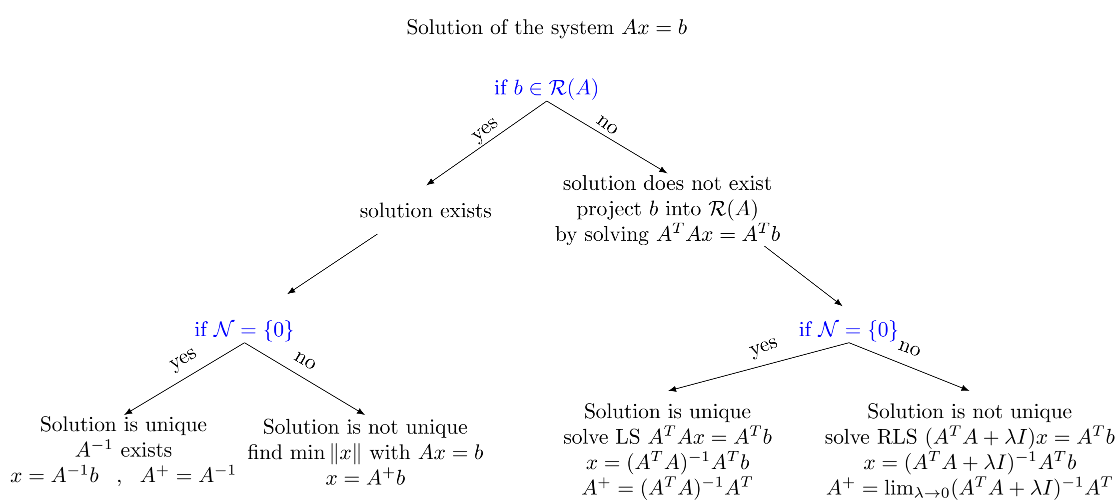

Figure 4: This two-branch binary decision tree shows the general framework to solve $Ax = b$ for $A \in \mathbb{R}^{m\times n}$. The blue text represents the nodes of the tree. The pseudo-inverse A^+ solves all cases but could be found differently according to each case. The case of $b \in \mathbb{R}$ and $\mathcal{N} \neq \{0\}$ is not explicitly solved. Here, the pseudo-inverse A^+ could be found by any of the methods indicated in the text and the references listed for it.

Figure 2 shows a summary of the whole process of finding the pseudo-solution and the pseudo-inverse.

1.1 Pseudo-inverse algorithm

There are several ways to compute a pseudo-inverse ([2]). A particular algorithm to find a pseudo-inverse is easily formulated from the SVD of A. We begin with

$$A = U\Sigma V^* ,$$

with A as an $m \times n$ matrix, U as an $m \times m$ matrix, Σ as an $m \times n$ matrix, and V as an $n \times n$ matrix. The inverse of U is U^*, and the inverse of V is V^*. We do not have an inverse for Σ, which may not even be a square matrix. In general, we can write

$$\Sigma = \left(\begin{array}{c|c} \Sigma_r & 0 \\ \hline 0 & 0 \end{array} \right) .$$

The obvious way to construct the inverse of Σ is by first taking the transpose of Σ, followed by inverting all nonzero diagonal elements (singular values) in Σ_r, and zeroing-out those entries which are already 0. That is, assuming that there are $r \leq \min\{m, n\}$ nonzero singular values σ_i :

$$\Sigma^+ = \begin{pmatrix} \sigma_1^{-1} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \sigma_2^{-1} & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_r^{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix} .$$

We define the ij-component of the Σ^+ matrix as follows:

$$\sigma_{ij}^+ = \begin{cases} \delta_{ij}\sigma_i^{-1} & i \leq r \\ 0 & i > r . \end{cases} \tag{1}$$

The matrix

$$A^+ = V\Sigma^+ U^* \tag{2}$$

is the pseudo-inverse of A.

A^+ solves the least-squares problem. That is, $x = A^+ b$ is such that $\|Ax - b\|^2$ is minimum, and, in addition, x is normal to the null space of A. In other words, $\langle x, y \rangle = 0$, for all y such that $Ay = 0$.

The pseudo-solution of the problem $Ax = b$ is given by

$$x^+ = A^+ b = V\Sigma^+ U^* b.$$

To summarize, we have shown that, unlike the matrix-inverse A^{-1} that only exists for square and non-singular matrices, the pseudo-inverse A^+ exists and is unique for all matrices A whose entries are real or complex numbers. This pseudo-inverse A^+ can be computed readily by using the singular value decomposition of A.

Problems 1.1

1. *Show that the smallest vector such that* $2x + 3y = 5$ *is*

$$u = \begin{pmatrix} 0.76923 \\ 1.15385 \end{pmatrix}$$

2. *Find the pseudo-inverse of the matrix* $A = \begin{pmatrix} 2 & 3 \end{pmatrix}$. ***Hint:*** *Recall that the pseudoinverse* A^+ *should satisfy* $A^+ b = u$. b *can be any non-zero number in* $\mathbb{R}$, *you can choose* $b = 5$ *to fit the equation on the previous point but that is not necessary.*

References

[1] H. Jaramillo and A. Rüger. *Machine Learning for Science and Engineering. Volume I: Fundamentals*. Society of Exploration Geophysicists, accepted for publication.

[2] James P. Keener. *Principles of Applied Mathematics: Transformation and Approximation*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1988.