THE PSEUDO-INVERSE OF A MATRIX

OMAR HIJAB

Introduction

The pseudo-inverse of a matrix is widely used in applied settings, most recently in Data Science. However, the pseudo-inverse is apparently not widely appreciated or presented in the typical undergraduate mathematics curriculum.

This note is a presentation of the pseudo-inverse. As the content has been well-understood for the better part of a century, the novelty, if any, is the presentation.

Every (square or non-square) matrix has a pseudo-inverse. The pseudo-inverse, often called the Moore-Penrose inverse [1], [3], [4], is a natural extension of the inverse of a matrix.

The pseudo-inverse of A is sometimes denoted A^+ [5]. However, because the inverse of A is denoted A^{-1}, we find the notation A^- more appropriate.

Let I be the identity matrix. The inverse of A, if it exists, is the unique matrix X satisfying

$$AX = I \qquad \text{and} \qquad XA = I. \tag{1}$$

When this happens, both X and I have the same shape as A.

The residuals are

$$\|AX - I\|^2 \qquad \text{and} \qquad \|XA - I\|^2. \tag{2}$$

These are scalar functions of X. For the residuals to be well-defined, the shape of X must equal the shape of A^t. If A is non-square, the identity matrices I in the residuals are not the same shape.

If A is invertible, the unique minimizer of the residuals is $X = A^{-1}$. Given this, one approach to the pseudo-inverse is to seek a matrix X minimizing the residuals.

Let P_c be the projection onto the column space of A, and P_r the projection onto the row space of A. If A is not full-rank, then one of AX or XA can never equal I. Because of this, another approach to the pseudo-inverse is to modify (1) by seeking X satisfying

$$AX = P_c \qquad \text{and} \qquad XA = P_r. \tag{3}$$

This second approach suggests modifying the residuals to

$$\|AX - P_c\|^2 \qquad \text{and} \qquad \|XA - P_r\|^2, \tag{4}$$

yielding a third approach. In fact, using $P_cA = A$ and $AP_r = A$, it is easy to check that the residuals (4) differ from the residuals (2) by constants.

If A is not full-rank, the equations (3) may have multiple solutions. We say a solution X^* of (3) is *least norm* if $\|X^*\|^2 \le \|X\|^2$ for all solutions X of (3).

Date: January 2026.

Similarly, we say a minimizer X^* for $\|AX - I\|^2$ is *least norm* if $\|X^*\|^2 \le \|X\|^2$ for all minimizers X for $\|AX - I\|^2$, and we say a minimizer X^* for $\|XA - I\|^2$ is *least norm* if $\|X^*\|^2 \le \|X\|^2$ for all minimizers X for $\|XA - I\|^2$.

Statement of Results

If A has d rows and N columns, its *shape* is (d, N).

Theorem 1. *For any matrix A, there is a unique least norm solution A^- of (3), and the shape of A^- equals the shape of A^t. Moreover, A^- is the unique least norm minimizer of $\|AX - I\|^2$ and A^- is the unique least norm minimizer of $\|XA - I\|^2$.*

The matrix A^- is the *pseudo-inverse* of A.

Recall a linear system $Ax = b$ is solvable if b is in the column space of A. When this happens, by applying both sides of $AA^- = P_c$ to b, we obtain a solution $x = A^- b$ of $Ax = b$.

Theorem 2 (Penrose Axioms)**.** *The pseudo-inverse of A is the unique matrix X satisfying*

A. $AXA = A$.
B. $XAX = X$.
C. AX *is symmetric.*
D. XA *is symmetric.*

From these axioms, we conclude *the pseudo-inverse of A^t is the transpose of the pseudo-inverse of A,* the pseudo-inverse of a projection P is P, and the pseudo-inverse of a zero matrix is the transpose of the matrix.

A matrix U is orthogonal if

$$UU^t = I = U^t U,$$

equivalently if $U^{-1} = U^t$. A matrix U is *pseudo-orthogonal* if $U^- = U^t$.

Theorem 3. *A matrix U is pseudo-orthogonal iff UU^t is a projection iff U^tU is a projection.*

In particular, a matrix with orthonormal columns or orthonormal rows is pseudo-orthogonal.

Let $u \cdot v$ be the dot product of vectors u, v. The tensor product of u, v is the matrix $u \otimes v$ satisfying $(u \otimes v)w = (v \cdot w)u$.

Let Q be a symmetric matrix with eigenvalue decomposition

$$Q = \lambda_1 v_1 \otimes v_1 + \lambda_2 v_2 \otimes v_2 + \cdots + \lambda_r v_r \otimes v_r.$$

Here v_1, v_2, …, v_r are orthonormal vectors and λ_1, λ_2, …, λ_r are nonzero.

Theorem 4. *The pseudo-inverse of Q is*

$$Q^- = \frac{1}{\lambda_1} v_1 \otimes v_1 + \frac{1}{\lambda_2} v_2 \otimes v_2 + \cdots + \frac{1}{\lambda_r} v_r \otimes v_r.$$

Let A be a matrix with singular value decomposition

$$A = \sigma_1 u_1 \otimes v_1 + \sigma_2 u_2 \otimes v_2 + \cdots + \sigma_r u_r \otimes v_r.$$

Here u_1, u_2, …, u_r and v_1, v_2, …, v_r are orthonormal vectors, and σ_1, σ_2, …, σ_r are positive.

Theorem 5. *The pseudo-inverse of A is*

$$A^- = \frac{1}{\sigma_1} v_1 \otimes u_1 + \frac{1}{\sigma_2} v_2 \otimes u_2 + \cdots + \frac{1}{\sigma_r} v_r \otimes u_r.$$

Because Theorems 3, 4, 5 are immediate consequences of the Penrose axioms, it is enough to establish Theorems 1 and 2.

An Example

Let

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \end{pmatrix}.$$

Using the Python's `numpy.linalg.pinv`, we have

$$A^- = \frac{1}{150} \begin{pmatrix} -37 & -10 & 17 \\ -20 & -5 & 10 \\ -3 & 0 & 3 \\ 14 & 5 & -4 \\ 31 & 10 & -11 \end{pmatrix}.$$

This can be verified using the Penrose axioms. All numerical values in this section were computed using Python.

If the rows of A are u, v, w, then $w = 2v - u$, so the rank of A is 2. Let a, b, c, d, e be the columns of A. We show

$$f = (-100, -3, 94), \qquad g = (2, 1, 0)$$

span the column space of A.

To this end, recall f and g are linear combinations of a, b, c, d, e iff $Ax = f$ and $Ax = g$ are solvable. Using $x = A^- f$ and $x = A^- g$ to solve $Ax = f$ and $Ax = g$ and writing out the components of x, we obtain

$$50f = 1776a + 985b + 194c - 597d - 1388e,$$
$$50g = -28a - 15b - 2c + 11d + 24e.$$

Similarly, let B be the matrix with columns f and g,

$$B = \begin{pmatrix} -100 & 2 \\ -3 & 1 \\ 94 & 0 \end{pmatrix},$$

and solve $Bx = a$, $Bx = b$, $Bx = c$, $Bx = d$, $Bx = e$ using

$$B^- = \frac{1}{564} \begin{pmatrix} -1 & 2 & 5 \\ 185 & 194 & 203 \end{pmatrix}.$$

Writing out the components of x, we obtain

$$94a = 11f + 597g, \qquad 47b = 6f + 347g, \qquad 94c = 13f + 791g,$$
$$47d = 7f + 444g, \qquad 94e = 15f + 985g.$$

This shows f, g span the column space of A.

Let

$$Q = AA^t = \begin{pmatrix} 55 & 130 & 205 \\ 130 & 330 & 530 \\ 205 & 530 & 855 \end{pmatrix}.$$

Since the column spaces of A and Q agree, Q has two nonzero eigenvalues λ.

To find λ, we restrict Q to its column space by inserting $v = sf + tg$ into $Qv = \lambda v$, leading to

$$Q(sf + tg) = \lambda(sf + tg). \tag{5}$$

Taking the dot product with f then with g leads to the 2×2 homogeneous linear system $(M - \lambda N)\begin{pmatrix} s \\ t \end{pmatrix} = 0$, where

$$M = \begin{pmatrix} f \cdot Qf & f \cdot Qg \\ g \cdot Qf & g \cdot Qg \end{pmatrix} = \begin{pmatrix} 4032830 & 62590 \\ 62590 & 1070 \end{pmatrix},$$

and

$$N = \begin{pmatrix} f \cdot f & f \cdot g \\ g \cdot f & g \cdot g \end{pmatrix} = \begin{pmatrix} 18845 & -203 \\ -203 & 5 \end{pmatrix}.$$

Setting the determinant of $M - \lambda N$ to zero yields

$$\lambda^2 - 1240\lambda + 7500 = 0,$$

which is solved by

$$\lambda_\pm = 620 \pm 10\delta, \qquad \delta = \sqrt{3769}.$$

To solve for (s, t), let

$$S = B^- QB = \begin{pmatrix} 620 & 10 \\ 10\delta^2 & 620 \end{pmatrix}.$$

Let $b = QB\begin{pmatrix} s \\ t \end{pmatrix}$ and $x = \lambda\begin{pmatrix} s \\ t \end{pmatrix}$. Then (5) may be written $Bx = b$, which is solved by $x = B^- b = S\begin{pmatrix} s \\ t \end{pmatrix}$. It follows (5) is implied by

$$S\begin{pmatrix} s \\ t \end{pmatrix} = \lambda\begin{pmatrix} s \\ t \end{pmatrix},$$

which is solved by

$$(s, t) = (1, \pm\delta).$$

Thus the corresponding eigenvectors are $f \pm \delta g$. Let $v_\pm$ be the unit vectors proportional to $f \pm \delta g$.

Since $v_\pm$ correspond to distinct eigenvalues, they are orthonormal. From this, the eigenvalue decomposition of AA^t is

$$AA^t = \lambda_+ v_+ \otimes v_+ + \lambda_- v_- \otimes v_-.$$

Let

$$p = (-72, 11, 94, 177, 260), \qquad q = (4, 3, 2, 1, 0).$$

Then, as before, one checks p, q span the row space of A. Repeating the analysis with $Q = A^t A$, p, q replacing f, g, $C = (p, q)$, and $S = C^- QC$, leads to the same nonzero eigenvalues and the two eigenvectors $p \pm \delta q$. Let $u_\pm$ be the unit vectors proportional to $p \pm \delta q$. Then

$$A^t A = \lambda_+ u_+ \otimes u_+ + \lambda_- u_- \otimes u_-.$$

Let $\sigma_\pm = \sqrt{\lambda_\pm}$. From this, one can show the singular value decompositions of A^t and A are

$$A^t = \sigma_+ u_+ \otimes v_+ + \sigma_- u_- \otimes v_-, \qquad A = \sigma_+ v_+ \otimes u_+ + \sigma_- v_- \otimes u_-.$$

Proofs

The set of matrices with fixed shape is a euclidean vector space with dot product and norm squared

$$A \cdot B = \operatorname{trace}(A^t B), \qquad \|A\|^2 = A \cdot A.$$

Let X_n be a sequence of matrices with fixed shape. We say $X_n \to \infty$ if $\|X_n\| \to \infty$. A scalar-valued function $f(X)$ is *proper* if

$$X_n \to \infty \qquad \Longrightarrow \qquad f(X_n) \to +\infty.$$

If $f(X)$ is continuous and proper, a basic calculus result [2] guarantees the existence of an X^* satisfying $f(X^*) \le f(X)$ for all X. Such an X^* is a *minimizer*.

Lemma 1. *Let A and B be matrices, and suppose X is a matrix with fixed shape such that*

$$f(X) = \|AX - B\|^2$$

is well-defined. Then a minimizer X^* for $f(X)$ exists, and can be chosen so the column space of X^* is in the row space of A.

Proof. It is enough to show $f(X) = \|AX - B\|^2$ is proper on the euclidean space $\mathcal{E}$ of matrices X whose column spaces lie within the row space of A. Argue by contradiction. If $f(X)$ is not proper, there is a sequence of matrices X_n in $\mathcal{E}$, and $X_n \to \infty$ with $f(X_n)$ bounded, hence with $\|AX_n\| \le c$ for some finite c. Let $r_n = \|X_n\|$ and $X'_n = X_n/r_n$. Since $\|X'_n\| = 1$, X'_n is a bounded sequence, hence a subsequence X'_n converges to some X'. By continuity, $\|X'\| = 1$ and X' is in $\mathcal{E}$. Since

$$\|AX'\| = \lim_{n\to\infty} \|AX'_n\| = \lim_{n\to\infty} \frac{\|AX_n\|}{r_n} \le \lim_{n\to\infty} \frac{c}{r_n} = 0,$$

the column space of X' lies in the null space of A. But the column space of X' lies in the row space of A. Since the row and null spaces of A are orthogonal, we must have $X' = 0$, which contradicts $\|X'\| = 1$. Thus $f(X)$ has a minimizer X^* on $\mathcal{E}$. Since any X may be written as $X = P_r X + (I - P_r)X$, and $A(I - P_r)X = 0$, X^* is a minimizer on all matrices X with fixed shape, not just X in $\mathcal{E}$. □

Lemma 2. *Let A be a matrix. A matrix X is a minimizer for $f(X) = \|AX - B\|^2$ iff $A^tAX = A^tB$. In particular, X is a minimizer for $f(X) = \|AX - I\|^2$ iff $AX = P_c$.*

Proof. Since $f(X)$ is quadratic, or by calculus, X is a minimizer iff $(AX - B)\cdot AV = 0$ for all V, which happens iff $A^tAX = A^tB$. Let $P = AX$. When $B = I$, this happens iff $A = P^tA$. Multiplying by X yields $P = P^tP$. It follows P satisfies $P^t = P$ and $P^2 = P$, hence P is a projection. Since $PA = A$, it follows P is the projection onto the column space of A. □

Let X^* be a minimizer for $f(X) = \|AX - B\|^2$. We say X^* is *least norm* if $\|X^*\|^2 \le \|X\|^2$ for all minimizers X for $f(X) = \|AX - B\|^2$.

Lemma 3. *A minimizer X^* for $\|AX - B\|^2$ is least norm iff the column space of X^* is in the row space of A, and there is at most one least norm minimizer for $\|AX - B\|^2$.*

Proof. From Lemma 2, if X_1 and X_2 are minimizers, then $A^tAX_1 = A^tAX_2$, hence $V = X_1 - X_2$ satisfies $A^tAV = 0$. Since the null spaces of A and A^tA agree, this happens iff $AV = 0$. Thus X^* is a least norm minimizer iff $\|X^*\|^2 \le \|X^* + V\|^2$ for all V whose column space lies in the null space of A. Since $\|X^* + V\|^2$ is quadratic in V, or by calculus, X^* is least norm iff $X^* \cdot V = 0$ for all V whose column space lies in the null space of A. Since the row and null spaces of A are orthogonal, this is the same as saying the column space of X^* lies in the row space of A. If X_1 and X_2 are least norm minimizers, then the column space of $V = X_1 - X_2$ lies in both the row and null spaces of A, hence $V = 0$. This establishes uniqueness of the least norm minimizer for $\|AX - B\|^2$. □

Given A, we conclude corresponding to each B, there is a unique least norm minimizer $X^-(B)$ for $\|AX - B\|^2$. Let A_c be the least norm minimizer $X^-(I)$ for $\|AX - I\|^2$.

Lemma 4. *With A fixed, $X^-(B) = A_cB$ for all B.*

Proof. It is enough to show $B \mapsto X^-(B)$ is a linear transformation. But this follows from the linearity in B of the constraints (1) $A^tAX = A^tB$ and (2) the column space of X lying in the row space of A. □

Lemma 5. *The matrix $X = A_c$ is the unique solution of the Penrose axioms.*

Proof. We have to show A_c satisfies the axioms **A**, **B**, **C**, and **D**, and conversely. By Lemma 2, $AA_c = P_c$. Thus AA_c is symmetric and $AA_cA = P_cA = A$, yielding **A** and **C**.

For **B**, choose $B = AA_c$ in $f(X) = \|AX - B\|^2$. Then $A_cB = A_cAA_c$ is the least norm minimizer, and A_c is trivially a minimizer, hence the column space of $V = A_c - A_cAA_c$ lies in the null space of A. Since the column space of A_c lies in the row space of A, $V = 0$, yielding **B**.

Let $P = A_cA$. Then the column space of P lies in the row space of A. By **A**, $A(I - P) = 0$, so the column space of $I - P$ lies in the null space of A. Since the row and null spaces of A are orthogonal,

$$(I - P)x \cdot Py = 0$$

for all x, y, hence

$$P^t - P^tP = P^t(I - P) = 0.$$

Since P^tP is symmetric, P^t hence P is symmetric, obtaining **D.**

For the converse, suppose X satisfies **A**, **B**, **C**, and **D**. Then **A** and **C** imply $A^t(AX - I) = 0$. Thus X is a minimizer for $f(X) = \|AX - I\|^2$. By **D**, XA is symmetric If x is in the null space of A,

$$x \cdot Xy = Ax \cdot XAXy = XAx \cdot Xy = 0.$$

Hence the column space of X is orthogonal to the null space of A, or the column space of X^* is in the row space of A. This shows X is the least norm minimizer $X^-(I)$, hence is A_c. □

Lemma 6. *The least norm minimizer A_c of $\|AX - I\|^2$ equals the least norm minimizer A_r of $\|XA - I\|^2$. Hence A_c is the least norm solution of (3).*

Proof. By transposing, X is a minimizer for $\|XA - I\|^2$ iff X^t is a minimizer for $\|A^tX - I\|^2$. It follows X is a least norm minimizer for $\|XA - I\|^2$ iff X^t is a least norm minimizer for $\|A^tX - I\|^2$. From Lemma 5, it follows X is a least norm minimizer for $\|XA - I\|^2$ iff

A. $A^tX^tA^t = A^t$.
B. $X^tA^tX^t = X^t$.
C. A^tX^t is symmetric.
D. X^tA^t is symmetric.

But these are the Penrose axioms. By uniqueness, a least norm minimizer A_r for $\|XA - I\|^2$ is unique and $A_c = A_r$. As in Lemma 2, one can show $X = A_r$ satisfies $XA = P_r$. Thus $A_c = A_r$ is the least norm solution of (3). □

This establishes Theorems 1 and 2.

References

[1] Bjerhammar, A.: Application of calculus of matrices to method of least squares; with special references to geodetic calculations. Trans. Roy. Inst. Tech. Stockholm **49** (1951)
[2] Hijab, O.: Math for Data Science. Springer, Cham (2025)
[3] Moore, E.H.: On the reciprocal of the general algebraic matrix. Bulletin of the American Mathematical Society **26**, 394–395 (1920)
[4] Penrose, R.: A generalized inverse for matrices. Proceedings of the Cambridge Philosophical Society **51**, 406–413 (1955)
[5] Wikipedia: Moore-penrose inverse. https://en.wikipedia.org/wiki/Moore-Penrose_inverse

Temple University
Email address: hijab@temple.edu