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 hopes to remedy this lapse. 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

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

If A has d rows and N columns, the *shape* of A is (d, N). When (1) 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. A solution X^* of (3) is *least norm* if $\|X^*\| \le \|X\|$ for all X satisfying either of the equations in (3).

Date: January 2026.

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

Theorem 1. *There is a unique least norm solution of* (3)*, there is a unique least norm minimizer of* $\|AX - I\|^2$*, and there is a unique least norm minimizer of* $\|XA - I\|^2$*, and the three matrices agree.*

The *pseudo-inverse* A^- of A is the matrix in Theorem 1. The shape of A^- equals the shape of A^t.

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.*

The axioms imply $(A^-)^t = (A^t)^-$, and $P^- = P$ if P is a projection.

A matrix U is orthogonal if

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

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.*

Examples are matrices with orthonormal or zero columns and matrices with orthonormal or zero rows. In particular, a zero matrix is pseudo-orthogonal.

Theorem 4. $P_c = AA^-$ *and* $P_r = A^-A$*.*

Recall a linear system $Ax = b$ is solvable if b is in the column space of A. By Theorem 4, this happens iff $AA^-b = b$, exhibiting a solution $x = A^-b$ of $Ax = b$.

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 of rank r and 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 $v_1, v_2, \ldots, v_r$ are orthonormal vectors and $\lambda_1, \lambda_2, \ldots, \lambda_r$ are nonzero.

Theorem 5. *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 of rank r and 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 $u_1, u_2, \ldots, u_r$ and $v_1, v_2, \ldots, v_r$ are orthonormal vectors, and $\sigma_1, \sigma_2, \ldots, \sigma_r$ are positive.

Theorem 6. *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, 6 are consequences of Theorem 2, it is enough to establish the first two theorems.

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 Penrose axioms, one verifies

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

If the rows of A are u, v, w, then $w = 2v - u$, so the rank of A is $r = 2$. 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 $x = (t_1, t_2, t_3, t_4, t_5)$ solves $Ax = f$ iff

$$f = t_1 a + t_2 b + t_3 c + t_4 d + t_5 e.$$

Using $x = A^- f$ and $x = A^- g$ to solve $Ax = f$ and $Ax = g$ and writing out the components of x, we obtain

$$\begin{aligned} 50f &= 1776a + 985b + 194c - 597d - 1388e, \\ 50g &= -28a - 15b - 2c + 11d + 24e. \end{aligned}$$

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

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

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

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

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 $u = sf + tg$ into $Qu = \lambda u$, 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 of Q are $f \pm \delta g$. Let $u_\pm$ be the unit vectors proportional to $f \pm \delta g$.

Since $u_\pm$ correspond to distinct eigenvalues, they are orthogonal. Hence $u_\pm$ form an orthonormal basis for the column space of A, From this, the eigenvalue decomposition of AA^t is

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

Let $p = (-72, 11, 94, 177, 260)$, $q = (4, 3, 2, 1, 0)$. Then, as before, one checks p, q span the row space of A. Repeating the analysis with

$$Q = A^t A = \begin{pmatrix} 158 & 176 & 194 & 212 & 230 \\ 176 & 197 & 218 & 239 & 260 \\ 194 & 218 & 242 & 266 & 290 \\ 212 & 239 & 266 & 293 & 320 \\ 230 & 260 & 290 & 320 & 350 \end{pmatrix}, \qquad B = (p, q) = \begin{pmatrix} -72 & 4 \\ 11 & 3 \\ 94 & 2 \\ 177 & 1 \\ 260 & 0 \end{pmatrix},$$

and $S = B^- QB$, leads to the same nonzero eigenvalues and the two eigenvectors $p \pm \delta q$ of Q. Let $v_\pm$ be the unit vectors proportional to $p \pm \delta q$. Then $v_\pm$ form an orthonormal basis for the row space of A, and

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

Let $\sigma_\pm = \sqrt{\lambda_\pm}$. From this, it follows the singular value decomposition of A is

$$A = \sigma_+ u_+ \otimes v_+ + \sigma_- u_- \otimes v_-.$$

Proofs

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

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

Although everything works just as well for complex matrices, we assume throughout our matrices are real.

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 and X be matrices with compatible shapes, in the sense*

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

is well-defined. Then a minimizer X^* of $f(X)$ exists, and can be chosen to satisfy $P_r X^* = X^*$.

Proof. We show $f(X) = \|AX - B\|^2$ is proper on the euclidean space $\mathcal{E}$ of matrices X satisfying $P_r X = X$. 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 $P_r X' = X'$. Since

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

$AX' = 0$. Since the row and null spaces of A are orthogonal, $P_r X' = 0$, hence $X' = 0$, 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 $\mathcal{E}$. $\square$

A minimizer X^* of $\|AX - B\|^2$ is *least norm* if $\|X^*\| \le \|X\|$ for all minimizers X of $\|AX - B\|^2$.

Lemma 2. *X is a least norm minimizer of $\|AX - B\|^2$ iff $A^t AX = A^t B$ and $P_r X = X$. In particular, X is a least norm minimizer of $\|AX - I\|^2$ iff $AX = P_c$ and $P_r X = X$.*

Proof. Since $\|AX - B\|^2$ is quadratic, X is a minimizer of $\|AX - B\|^2$ iff $(AX - B) \cdot AV = 0$ for all V, which happens iff $A^t AX = A^t B$. If X_1 and X_2 are minimizers of $\|AX - B\|^2$, then $V = X_1 - X_2$ satisfies $A^t AV = 0$. Since the null spaces of A and $A^t A$ agree, V satisfies $AV = 0$, or $P_r V = 0$. Thus a minimizer X of $\|AX - B\|^2$ is least norm iff X minimizes $\|X + V\|^2$ over V satisfying $P_r V = 0$. Since $\|X + V\|^2$ is quadratic, X is least norm iff $X \cdot V = 0$ for all V satisfying $P_r V = 0$. But $A(I - P_r)V = 0$ for all V, hence $(I - P_r)X \cdot V = X \cdot (I - P_r)V = 0$ for all V, so X is least norm iff $P_r X = X$.

Specializing to $B = I$, if X is a minimizer of $\|AX - I\|^2$, then $A^t AX = A^t$. If $P = AX$, this is equivalent to $A = P^t A$ and $P_r X = X$. Multiplying the first equality by X yields $P = P^tP$. It follows P satisfies $P^t = P$ and $P^2 = P$, hence P is a projection. From $PA = A$, it follows $P = P_c$. Conversely, if $AX = P_c$, then $AXA = P_cA = A$, so $A^t = (AXA)^t = A^t(AX)$, so X is a minimizer of $\|AX - I\|^2$. $\square$

Lemma 3. *There is a unique least norm minimizer of $\|AX - B\|^2$. Moreover, if X is the least norm minimizer of $\|AX - I\|^2$, then XB is the least norm minimizer of $\|AX - B\|^2$.*

Proof. If X_1 and X_2 are least norm minimizers of $\|AX - B\|^2$, then $V = X_1 - X_2$ satisfies $A^tAV = 0$ and $P_rV = V$. From this, $AV = 0$, or $P_rV = 0$, hence $V = 0$. This establishes uniqueness. If X is the least norm minimizer of $\|AX - I\|^2$, then $A^tAX = A^t$ and $P_rX = X$. Right-multiplying by B yields $A^tA(XB) = A^tB$ and $P_r(XB) = XB$, hence XB is the least norm minimizer of $\|AX - B\|^2$. $\square$

Lemma 4. *X is the least norm minimizer of $\|AX - I\|^2$ iff X satisfies the Penrose axioms.*

Proof. If X is the least norm minimizer of $\|AX - I\|^2$, by Lemma 2, $AX = P_c$ and $P_rX = X$. Thus $AXA = P_cA = A$, yielding **A**. Since $AX = P_c$ is symmetric, this yields **C**. From Lemma 3, XP_c is the least norm minimizer of $\|AX - P_c\|^2$. Since $\|AX - I\|^2$ and $\|AX - P_c\|^2$ differ by a constant, they have the same least norm minimizer, hence $XP_c = X$, or $XAX = X$, yielding **B**. Let $P = XA$. Then $P_rP = P$. By **A**, $A(I - P) = 0$, so $P_r(I - P) = 0$. This shows $P = P_r$, hence P is symmetric, obtaining **D.**

Conversely, assume X satisfies the Penrose axioms. Then **A** and **C** imply $A^tAX = (AXA)^t = A^t$, thus X is a minimizer of $\|AX - I\|^2$. If $P = XA$, **B** and **D** imply P is a projection, hence $P = A^tX^t$. Thus the column space of P lies in the column space of A^t. By **A**, $A^t = (AXA)^t = PA^t$, so the column space of A^t lies in the column space of P. Hence the column spaces agree. Since the column space of A^t equals the row space of A, $P = P_r$, hence $P_rX = PX = XAX = X$. This shows X is the least norm minimizer of $\|AX - I\|^2$. $\square$

Lemma 5. *The least norm minimizer of $\|AX - I\|^2$ equals the least norm minimizer of $\|XA - I\|^2$.*

Proof. By transposing, X is a minimizer of $\|XA - I\|^2$ iff X^t is a minimizer of $\|A^tX - I\|^2$. It follows X is a least norm minimizer of $\|XA - I\|^2$ iff X^t is a least norm minimizer of $\|A^tX - I\|^2$. From Lemma 4, it follows X is a least norm minimizer of $\|XA - I\|^2$ iff X^t satisfies the Penrose axioms for A^t, or

- **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.

Since this is equivalent to X satisfying the Penrose axioms for A, by uniqueness, a least norm minimizer of $\|XA - I\|^2$ must equal the least norm minimizer of $\|AX - I\|^2$. $\square$

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

Proof. As before, a solution X of (3) is least norm iff $P_r X = X$ and $XP_c = X$. Let X be the least norm minimizer of $\|AX - I\|^2$. From Lemma 2, $AX = P_c$ and $P_r X = X$. Since X is also the least norm minimizer of $\|XA - I\|^2$, X^t is the least norm minimizer of $\|A^t X - I\|^2$. Since the row and column space projections are interchanged by taking transpose, we have $A^t X^t = P_r$ and $P_c X^t = X^t$. Hence $XA = P_r$ and $XP_c = X$. Thus X is a least norm solution of (3). The steps are reversible, so the converse holds. □

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