THE PSEUDOINVERSE OF A MATRIX

OMAR HIJAB

Introduction

The pseudoinverse of a matrix is widely used, most recently in Data Science, and its numerical evaluation has been thoroughly studied over the years.

What is apparently lacking, in the typical undergraduate mathematics curriculum, is a clear and short presentation of its *conceptual inevitability*.

This note hopes to remedy this lapse. As the pseudoinverse has been studied for the better part of a century, the novelty, if any, is the presentation.

Every (square or non-square) matrix has a pseudoinverse. The pseudoinverse, sometimes called a generalized inverse or Moore-Penrose inverse [2], [3], is a natural extension of the inverse of a matrix.

The pseudoinverse of A is sometimes denoted A^+ [4]. 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

$$AX = I = XA. \tag{1}$$

Matrices are the same shape if they have the same number of rows and columns. When (1) happens, A, X, and I are the same shape.

The residuals are

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

These are scalar functions of X. The residuals are well-defined only if X and A^t are the same shape. 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 pseudoinverse is to seek a matrix X minimizing the residuals.

Let P_c be the projection onto the column space of A, and let P_r be 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 pseudoinverse is to modify (1) by seeking X satisfying

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

This second approach suggests modifying the residuals to

$$\|AX - P_c\|^2 \qquad \text{and} \qquad \|XA - P_r\|^2, \tag{3}$$

yielding a third approach. In fact, using $P_cA = A$ and $AP_r = A$, it is easy to check that the residuals (3) differ from the residuals (2) by constants.

Date: January 2026.

A solution X_0 of $f(X) = 0$ is *least norm* if $\|X_0\| \le \|X\|$ for all solutions X of $f(X) = 0$. A minimizer X_0 of $f(X)$ is *least norm* if $\|X_0\| \le \|X\|$ for all minimizers X of $f(X)$.

The *pseudoinverse* A^- of A is the matrix X in the following Theorem. In particular, A^- and A^t are the same shape.

Theorem. *There is a unique least norm solution of $AX = P_c$, a unique least norm solution of $XA = P_r$, a unique least norm minimizer of $\|AX - I\|^2$, a unique least norm minimizer of $\|XA - I\|^2$, a unique solution of the Penrose axioms*

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

and the five matrices are equal.

Immediate consequences are $(A^-)^t = (A^t)^-$, 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 $U^- = U^t$.

Corollary. *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.

Corollary. *$P_c = AA^-$ and $P_r = A^-A$.*

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

Let u, v, w be vectors. The dot product is denoted $v \cdot w$, and 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.

Corollary. *The pseudoinverse 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.

Corollary. *The pseudoinverse of A is*

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

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
$$A^- = \frac{1}{150} \begin{pmatrix} -37 & -10 & 17 \\ -20 & -5 & 10 \\ -3 & 0 & 3 \\ 14 & 5 & -4 \\ 31 & 10 & -11 \end{pmatrix}.$$
All computations are done using the Python library `numpy`.

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
$$Q(sf + tg) = \lambda(sf + tg). \tag{4}$$

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 (4) may be written $Bx = b$, which is solved by $x = B^- b = S\begin{pmatrix} s \\ t \end{pmatrix}$. It follows (4) 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, using A^t instead of A, 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 S, 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

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

Proof of the Theorem

The proof is presented as a series of Lemmas, some of whose proofs are left to the reader. The euclidean space $\mathbf{R}^{N\times d}$ is the set of real matrices X with N rows and d columns. The dot product and norm squared in $\mathbf{R}^{N\times d}$ are

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

If X is in $\mathbf{R}^{N\times d}$, then X^t is in $\mathbf{R}^{d\times N}$, and $\|X\|^2 = \|X^t\|^2$. Let A be in $\mathbf{R}^{d\times N}$ and let B be in $\mathbf{R}^{d\times d}$. Then $AX - B$ is in $\mathbf{R}^{d\times d}$.

Since the row and null spaces of A are orthogonal, $AV = 0$ iff $P_r V = 0$. Since $V \cdot A^t AV = AV \cdot AV = \|AV\|^2$, $AV = 0$ iff $A^t AV = 0$.

Lemma 1. *$\|X\|^2 \leq \|X + V\|^2$ for all V in $\mathbf{R}^{N\times d}$ satisfying $AV = 0$ iff $P_r X = X$.*

Let $f(X)$ be a scalar function defined on a euclidean space $\mathcal{E}$ of matrices. If X_0 in $\mathcal{E}$ satisfies $f(X_0) \leq f(X)$ for all X in $\mathcal{E}$, then X_0 is a *minimizer of $f(X)$ on $\mathcal{E}$.*

A function $f(X)$ is *proper on $\mathcal{E}$* if for every sequence X_n in $\mathcal{E}$,

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

If $f(X)$ is continuous and proper on $\mathcal{E}$, a basic calculus result [1] guarantees the existence of a minimizer of $f(X)$ on $\mathcal{E}$. When $\mathcal{E}$ is not specified, $\mathcal{E} = \mathbf{R}^{N\times d}$.

Lemma 2. *There is a minimizer X_0 of $\|AX - B\|^2$ satisfying $P_r X_0 = X_0$.*

Proof. Let $\mathcal{E}$ be the subspace of $\mathbf{R}^{N\times d}$ of matrices X satisfying $P_r X = X$. We show $f(X) = \|AX - B\|^2$ is proper on $\mathcal{E}$, arguing by contradiction.

If $f(X)$ is not proper on $\mathcal{E}$, there is a sequence of matrices X_n in $\mathcal{E}$ with $f(X_n)$ bounded and $X_n \to \infty$, hence with $\|AX_n\| \leq c$ 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

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

$AX' = 0$, hence $P_r X' = 0$, hence $X' = 0$, which contradicts $\|X'\| = 1$.

Thus there is a minimizer X_0 of $f(X)$ on $\mathcal{E}$. For X in $\mathbf{R}^{N\times d}$, $P_r X$ is in $\mathcal{E}$, and $f(X) = f(P_r X)$, hence X_0 is a minimizer of $f(X)$. $\square$

Lemma 3. *There is a unique least norm minimizer X of $\|AX - B\|^2$, and the least norm minimizer X of $\|AX - B\|^2$ equals the unique solution of the pair $A^t AX = A^t B$, $P_r X = X$.*

Proof. Since $f(X) = \|AX - B\|^2$ is convex, X is a minimizer of $f(X)$ iff the derivative of $f(X + tV)$ at $t = 0$ vanishes for all V. Writing this out, $(AX - B) \cdot AV = 0$ for all V, which happens iff $A^t AX = A^t B$.

By Lemma 2, there is a minimizer X satisfying $P_r X = X$. By Lemma 1, a minimizer X is least norm iff $P_r X = X$, hence there is a least norm minimizer.

If X and X_0 are minimizers, then $V = X - X_0$ satisfies $A^t AV = 0$, hence $P_r V = 0$. If X and X_0 are least norm minimizers, then $P_r V = V$, so $V = 0$. This establishes the uniqueness of the least norm minimizer.

Since the steps are reversible, the solution of the pair $A^t AX = A^t B$, $P_r X = X$ is unique. $\square$

Lemma 4. *If X_0 is the least norm minimizer of $\|AX - I\|^2$, then X_0B is the least norm minimizer of $\|AX - B\|^2$.*

Lemma 5. *The least norm minimizer X of $\|AX - I\|^2$ equals the unique least norm solution of $AX = P_c$.*

Proof. X is a least norm solution of $AX = P_c$ iff X is a solution of the pair $AX = P_c$, $P_rX = X$, so it is enough to show $A^tAX = A^t$ iff $AX = P_c$.

If $P = AX$ and $A^tAX = A^t$, then $A = P^tA$. Right-multiplying by X yields $P = P^tP$. It follows P satisfies $P^t = P$ and $P^2 = P$, hence P is a projection.

Since $A = PA$ and $P = AX$, $P = P_c$, hence $AX = P_c$. Conversely, if $AX = P_c$, then $AXA = P_cA = A$, so $A^t(AX) = (AXA)^t = A^t$. □

Lemma 6. *There is a unique least norm minimizer X of $\|XA - I\|^2$, and the least norm minimizer X of $\|XA - I\|^2$ equals the unique least norm solution of $XA = P_r$.*

Proof. Replace A by A^t in Lemma 5. □

Lemma 7. *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 $\|AX - I\|^2$, then $AX = P_c$. Thus $AXA = P_cA = A$, yielding **A**. Since P_c is symmetric, we also have **C**. From Lemma 4, 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, their minimizers agree, hence they have the same least norm minimizer. This shows $XP_c = X$, or $XAX = X$, yielding **B**. Let $P = XA$. Since $P_rX = X$, $P_rP = P$. 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 $A^tAX = (AXA)^t = A^t$, hence X is a minimizer of $\|AX - I\|^2$. If $P = XA$, **B** and **D** imply P is a projection, hence $P = A^tX^t$. By **A**, $A^t = (AXA)^t = PA^t$. Since $P = A^tX^t$ and $A^t = PA^t$, the column spaces of P and A^t agree, hence the column space of P equals the row space of A. Thus $P = P_r$, hence $P_rX = PX = XAX = X$. Hence X is the least norm minimizer of $\|AX - I\|^2$. □

Lemma 8. *The least norm minimizer of $\|XA - I\|^2$ equals the least norm minimizer of $\|AX - I\|^2$.*

Proof. If X is the least norm minimizer of $\|XA - I\|^2$, X^t is the least norm minimizer of $\|A^tX - I\|^2$, hence X^t satisfies the Penrose axioms for A^t. But X^t satisfies the Penrose axioms for A^t iff X satisfies the Penrose axioms for A. Thus X is the least norm minimizer of $\|AX - I\|^2$. □

References

[1] Hijab, O.: Math for Data Science. Springer, Cham (2025)

[2] Moore, E.H.: On the reciprocal of the general algebraic matrix. Bulletin of the American Mathematical Society **26**, 394–395 (1920)

[3] Penrose, R.: A generalized inverse for matrices. Proceedings of the Cambridge Philosophical Society **51**, 406–413 (1955)

[4] Wikipedia: Moore-Penrose inverse. https://en.wikipedia.org/wiki/Moore-Penrose_inverse

Temple University

Email address: `hijab@temple.edu`