

# THE PSEUDOINVERSE OF A MATRIX

OMAR HIJAB

## 1. INTRODUCTION.

The pseudoinverse of a matrix is widely used, 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 emphasizing its *conceptual inevitability*.

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

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

The pseudoinverse of  $A$  is sometimes denoted  $A^+$ . 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

$$(1) \quad AX = I = XA.$$

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.

With  $\|X\|^2 = \text{trace}(X^t X)$ , the residuals are

$$(2) \quad \|AX - I\|^2 \quad \text{and} \quad \|XA - I\|^2.$$

These are scalar functions of  $X$ . The residuals are 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 residuals are minimized by  $X = A^{-1}$ . Given this, one approach to the pseudoinverse is to seek  $X$  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 invertible or non-square, 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

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

As before, these equations are defined only if  $X$  and  $A^t$  are the same shape.

This second approach suggests the residuals

$$(3) \quad \|AX - P_c\|^2 \quad \text{and} \quad \|XA - P_r\|^2,$$

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

Let  $f(X)$  be a scalar function of  $X$ . A solution  $X_0$  of  $f(X) = 0$  is *least norm* if  $\|X_0\| \leq \|X\|$  for all solutions  $X$  of  $f(X) = 0$ . A minimizer  $X_0$  of  $f(X)$  is *least norm* if  $\|X_0\| \leq \|X\|$  for all minimizers  $X$  of  $f(X)$ .

## 2. PSEUDOINVERSE.

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

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

**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  $Ax = b$ .

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.**  $U$  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.

Let  $u, v, w$  be vectors. The dot product or inner product of  $v, w$  is  $v \cdot w$ , and the tensor product or outer product of  $u, v$  is the matrix  $u \otimes v$  satisfying  $(u \otimes v)w = u(v \cdot w)$ . Then the transpose  $(u \otimes v)^t$  equals  $v \otimes u$ .

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  $v_1, v_2, \dots, v_r$  are orthonormal vectors and  $\lambda_1, \lambda_2, \dots, \lambda_r$  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  $u_1, u_2, \dots, u_r$  and  $v_1, v_2, \dots, v_r$  are orthonormal vectors, and  $\sigma_1, \sigma_2, \dots, \sigma_r$  are positive. The following implies  $A$  is pseudo-orthogonal iff all  $\sigma_k = 1$ .

**Corollary.** The pseudoinverse of  $A$  is

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

This result is the basis for numerical evaluations of  $A^-$ .

## 3. EXAMPLE.

Let  $A$  be the sequential integer matrix

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

Using the Penrose axioms, one can verify

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

Using  $A^-$ , we compute the eigenvalue decomposition of  $Q = AA^t$ . Conversely, using the eigenvalue decomposition of  $Q$ , we compute  $A^-$ .

Let  $a, b, c, d, e$  be the columns of  $A$ . Since the columns are linear combinations of  $a, b$ , the rank of  $A$  is 2.

Similarly, the rank of any sequential integer matrix  $A$  is 2, so the treatment here can be repeated for  $A$ .

To simplify the computations, instead of  $a, b$ , we use

$$f = (54, 77, 100), \quad g = (0, 1, 2).$$

Using  $A^-$ , we show  $f, g$  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

$$f = t_1a + t_2b + t_3c + t_4d + t_5e.$$

Since  $x = A^-f$ ,  $x = A^-g$  solve  $Ax = f$ ,  $Ax = g$ , we obtain

$$50f = -356a - 155b + 46c + 247d + 448e,$$

$$50g = 8a + 5b + 2c - d - 4e.$$

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

$$B = (f, g) = \begin{pmatrix} 54 & 0 \\ 77 & 1 \\ 100 & 2 \end{pmatrix},$$

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

$$B^- = \frac{1}{324} \begin{pmatrix} 5 & 2 & -1 \\ -277 & -46 & 185 \end{pmatrix}.$$

Writing out the components of  $x$ , we obtain

$$54a = f + 247g, \quad 27b = f + 112g, \quad 18c = f + 67g,$$

$$27d = 2f + 89g, \quad 54e = 5f + 155g.$$

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  $Q$  and  $A$  agree,  $Q$  has two nonzero eigenvalues  $\lambda$ .

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

$$(6) \quad Q(sf + tg) = \lambda(sf + tg).$$

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

$$M = \begin{pmatrix} f \cdot Qf & f \cdot Qg \\ g \cdot Qf & g \cdot Qg \end{pmatrix} = \begin{pmatrix} 22124030 & 360190 \\ 360190 & 5870 \end{pmatrix},$$

and

$$N = \begin{pmatrix} f \cdot f & f \cdot g \\ g \cdot f & g \cdot g \end{pmatrix} = \begin{pmatrix} 18845 & 277 \\ 277 & 5 \end{pmatrix}.$$

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

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

which is solved by

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

To solve for  $(s, t)$ , let

$$S = B^{-1}QB = \begin{pmatrix} 620 & 10 \\ 10\delta^2 & 620 \end{pmatrix}.$$

Then  $\lambda_{\pm}$  are the eigenvalues of  $S$ .

Let  $b = QB \begin{pmatrix} s \\ t \end{pmatrix}$  and  $x = \lambda \begin{pmatrix} s \\ t \end{pmatrix}$ . Then (6) may be written  $Bx = b$ , which is solved by  $x = B^{-1}b = S \begin{pmatrix} s \\ t \end{pmatrix}$ . It follows (6) is implied by

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

which is solved by

$$\lambda = \lambda_{\pm} \quad \text{and} \quad (s, t) = (1, \pm\delta).$$

Thus  $f \pm \delta g$  are eigenvectors of  $AA^t$  corresponding to  $\lambda_{\pm}$ . Since  $\lambda_+ \neq \lambda_-$ ,  $f \pm \delta g$  are an orthogonal basis for the column space of  $A$ .

To obtain a basis for the row space of  $A$ , let

$$p = A^t f, \quad q = A^t g.$$

Since  $f \pm \delta g$  are eigenvectors,

$$A(p \pm \delta q) = \lambda_{\pm}(f \pm \delta g), \quad |p \pm \delta q|^2 = \lambda_{\pm}|f \pm \delta g|^2.$$

It follows  $p \pm \delta q$  are eigenvectors for  $A^t A$  corresponding to  $\lambda_{\pm}$ , hence  $p \pm \delta q$  are an orthogonal basis for the row space.

If  $u_{\pm}$  are the unit vectors positively proportional to  $f \pm \delta g$ ,  $u_{\pm}$  are orthonormal. From this,

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

Let  $\sigma_{\pm} = \sqrt{\lambda_{\pm}}$  and let  $v_{\pm}$  be the unit vectors positively proportional to  $p \pm \delta q$ . Then  $v_{\pm}$  are orthonormal, and

$$v_{\pm} = \frac{1}{\sigma_{\pm}} A^t u_{\pm}.$$

From this,

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

Summarizing,

$$Av_{\pm} = \sigma_{\pm} u_{\pm} \quad \text{and} \quad A^t u_{\pm} = \sigma_{\pm} v_{\pm},$$

which leads to the singular value decompositions

$$A = \sigma_+ u_+ \otimes v_+ + \sigma_- u_- \otimes v_-, \quad A^- = \frac{1}{\sigma_+} v_+ \otimes u_+ + \frac{1}{\sigma_-} v_- \otimes u_-.$$

Inserting  $\sigma_{\pm}$ ,  $u_{\pm}$ ,  $v_{\pm}$  yields (4), (5), and completes the example.

#### 4. PROOF.

The proof of the Theorem is 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

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

If  $X$  is in  $\mathbf{R}^{N \times d}$ , then  $X^t$  is in  $\mathbf{R}^{d \times N}$ , and  $\|X\|^2 = \|X^t\|^2$ .

Throughout  $A$  is in  $\mathbf{R}^{d \times N}$ ,  $B$  is in  $\mathbf{R}^{d \times d}$ , and  $X$  is in  $\mathbf{R}^{N \times d}$ . 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  $\mathcal{E}$ . 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  $\mathcal{E}$ ,

$$\|X_n\| \rightarrow \infty \implies f(X_n) \rightarrow +\infty.$$

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  $r_n = \|X_n\| \rightarrow \infty$  and  $f(X_n)$  bounded, hence with  $\|AX_n\| \leq c$  for some finite  $c$ .

Let  $X'_n = X_n/r_n$ . 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

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

$AX' = 0$ , 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  $X_0 B$  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  $AX = P_c$ ,  $P_r X = X$ , so it is enough to show  $A^t AX = A^t$  iff  $AX = P_c$ .

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

Since  $A = PA$  and  $P = AX$ , the column spaces of  $P$  and  $A$  agree, hence  $P = P_c$ , hence  $AX = P_c$ . Conversely, if  $AX = P_c$ , then  $AXA = P_c A = A$ , so  $A^t(AX) = (AXA)^t = A^t$ .  $\square$

**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.* Write  $\|XA - I\|^2 = \|A^t X^t - I\|^2$  and use Lemma 5. Since the column space of  $A^t$  is the row space of  $A$ ,  $XA = (A^t X^t)^t = P_r^t = P_r$ .  $\square$

**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  $AXA = P_c A = A$ , 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  $P_r X = X$ ,  $P_r P = P$ . Also, 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  $A^t AX = (AXA)^t = A^t$ , hence  $X$  is a minimizer of  $\|AX - I\|^2$ . If  $P = XA$ , **B** and **D** imply  $P$  is a projection, hence  $P = A^t X^t$ .

By **A**,  $A^t = (AXA)^t = PA^t$ . Since  $P = A^t X^t$  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

$$P_r X = P X = X A X = X.$$

This establishes  $X$  is the least norm minimizer of  $\|AX - I\|^2$ .  $\square$

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

*Proof.* Use the transpose invariance of the Penrose axioms,  $\|XA - I\|^2 = \|A^t X^t - I\|^2$ , and Lemma 7.  $\square$

#### REFERENCES

- [1] Hijab O. Math for Data Science. Cham, Switzerland: Springer; 2025.
- [2] Moore EH. On the reciprocal of the general algebraic matrix. Bulletin of the American Mathematical Society. 1920;26:394-5.
- [3] Penrose R. A generalized inverse for matrices. Proceedings of the Cambridge Philosophical Society. 1955;51:406-13.
- [4] Wikipedia. Moore-Penrose\_Inverse. Available from: <https://en.wikipedia.org/wiki/Moore-Penrose-inverse>.

TEMPLE UNIVERSITY

Email address: hijab@temple.edu