

13.12 The Moore–Penrose Pseudoinverse

Chapter 11 solved $\mathbf{Ax} = \mathbf{b}$ twice over: when there is no solution it produced a least-squares solution, and when there are too many it produced the shortest one. It then left an IOU. The two answers “are usually wanted together: given any $\mathbf{A}$ and $\mathbf{b}$, first minimize $\|\mathbf{Ax} - \mathbf{b}\|$, then among the minimizers take the shortest. Chapter 13 packages the combined answer as a single matrix, the **pseudoinverse** $\mathbf{A}^+$, with $\mathbf{x} = \mathbf{A}^+\mathbf{b}$ in all cases at once.” This section builds that matrix and pays the debt.

Throughout, $F = \mathbb{R}$ or $F = \mathbb{C}$, $\mathbf{A} \in M_{m \times n}(F)$, $p = \min(m, n)$, and $r = \text{rank } \mathbf{A}$, with singular values $\sigma_1 \geq \cdots \geq \sigma_p \geq 0$ (Definition 13.9.1), so $\sigma_i > 0$ exactly for $i \leq r$.

13.12.1 One matrix instead of three formulas

Here is the state of play at the end of Chapter 11. To answer “what is the best $\mathbf{x}$?” we had three different recipes, and choosing between them required knowing something about $\mathbf{A}$ first.

- If $\mathbf{A}$ is square and invertible, $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.
- If the columns of $\mathbf{A}$ are independent, $\mathbf{x} = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{b}$ (Corollary 11.4.4).
- If $\mathbf{Ax} = \mathbf{b}$ is consistent but $\mathbf{A}$ has a kernel, $\mathbf{x} = \mathbf{x}_{\min}$, the unique solution orthogonal to $\text{nul}(\mathbf{A})$ (Theorem 11.4.5).

Three recipes, three hypotheses, and a gap: when $\mathbf{A}$ has both a kernel *and* a column space missing $\mathbf{b}$, none of them applies. That is not an exotic case. It is what happens whenever a large table of measurements has both redundant columns and noise.

The goal, stated before we know how to reach it, is a single matrix $\mathbf{A}^+$, depending on $\mathbf{A}$ alone, for which $\mathbf{x} = \mathbf{A}^+\mathbf{b}$ is the right answer in every case. Reading the goal backwards tells us what $\mathbf{A}^+$ must do, and the singular value decomposition tells us how to build it.

13.12.2 Inverting what can be inverted

Why is there no inverse in general? Two obstructions, and they are the two halves of Theorem 11.6.5. The map $\mathbf{x} \mapsto \mathbf{Ax}$ may fail to be injective, because $\text{nul}(\mathbf{A}) \neq \{\mathbf{0}\}$; and it may fail to be surjective, because $\text{col}(\mathbf{A}) \neq F^m$. Neither obstruction is fatal if we are willing to sacrifice part of the domain and part of the codomain.

Restrict the map to $\text{col}(\mathbf{A}^*) = \text{nul}(\mathbf{A})^\perp$, which throws away the kernel. What is left is a bijection onto $\text{col}(\mathbf{A})$, and a bijection has an inverse. Extend that inverse to all of F^m by declaring it $\mathbf{0}$ on $\text{col}(\mathbf{A})^\perp$, where nothing in the image can tell us anything. That is the whole construction, and it is the same trick that turns $\sin$ into $\arcsin$: shrink the domain until the function is injective, shrink the codomain until it is surjective, invert, and say clearly what was given up.

Invert what can be inverted, and send the rest to zero.

The singular value decomposition performs all three operations at once, because it presents $\mathbf{A}$ as a diagonal matrix in orthonormal coordinates, and a diagonal matrix is inverted one entry at a time.

Definition 13.12.1. Moore–Penrose pseudoinverse

Let $\mathbf{A} \in M_{m \times n}(F)$ and let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ be a singular value decomposition (Theorem 13.9.1), so that $\mathbf{U} \in M_m(F)$ and $\mathbf{V} \in M_n(F)$ are unitary and $\mathbf{\Sigma} \in M_{m \times n}(F)$ has the singular values

$\sigma_1, \dots, \sigma_p$ on its diagonal and zeros elsewhere. Let $\Sigma^+ \in M_{n \times m}(F)$ be the matrix with

$$(\Sigma^+)_{ii} = \begin{cases} \sigma_i^{-1}, & \sigma_i > 0, \\ 0, & \sigma_i = 0, \end{cases}$$

for $1 \leq i \leq p$, and all other entries 0. The **Moore–Penrose pseudoinverse** of $\mathbf{A}$ is

$$\mathbf{A}^+ := \mathbf{V}\Sigma^+\mathbf{U}^* \in M_{n \times m}(F).$$

Clause by clause. The matrix Σ^+ is Σ **transposed and then inverted where inversion is legal**: the shape flips from $m \times n$ to $n \times m$, each nonzero diagonal entry is replaced by its reciprocal, and each zero diagonal entry is left alone — **not** replaced by anything, since 0^{-1} does not exist. The outer factors are exchanged and starred, exactly as they would be in $(\mathbf{U}\Sigma\mathbf{V}^*)^{-1} = \mathbf{V}\Sigma^{-1}\mathbf{U}^*$ if Σ were invertible. And the result is $n \times m$, the shape an inverse of $\mathbf{A}$ would have to be.

Well-definedness is a real question here, not a formality. The singular *values* are determined by $\mathbf{A}$, but the unitary factors are not (Theorem 13.9.2), so the recipe above is applied to a choice, and nothing so far says that two choices give the same answer. They do.

Proposition 13.12.1. The pseudoinverse does not depend on the decomposition

Let $\mathbf{A} \in M_{m \times n}(F)$ and let $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*$ be any singular value decomposition (Theorem 13.9.1). Then $\mathbf{X} = \mathbf{V}\Sigma^+\mathbf{U}^*$ is the unique matrix in $M_{n \times m}(F)$ such that

- (i) for every $\mathbf{y} \in \text{col}(\mathbf{A})$, $\mathbf{X}\mathbf{y}$ is the unique vector of $\text{col}(\mathbf{A}^*)$ that $\mathbf{A}$ sends to $\mathbf{y}$;
- (ii) $\mathbf{X}\mathbf{y} = \mathbf{0}$ for every $\mathbf{y} \in \text{col}(\mathbf{A})^\perp$.

In particular $\mathbf{A}^+$ depends on $\mathbf{A}$ alone.

Idea. Conditions (i) and (ii) mention only $\mathbf{A}$, so if one matrix satisfies them, that matrix is determined by $\mathbf{A}$. The work is to check that they can be satisfied at all — which is the “restrict, invert, extend” story of the previous page written out — and that the matrix built from an arbitrary decomposition is the one that satisfies them. For the second part, the singular vectors are orthonormal bases of the subspaces involved, so both conditions can be checked on basis vectors.

Proof. **The conditions determine at most one matrix.** First, $\mathbf{A}$ maps $\text{col}(\mathbf{A}^*)$ bijectively onto $\text{col}(\mathbf{A})$. Injectivity: if $\mathbf{x} \in \text{col}(\mathbf{A}^*)$ and $\mathbf{A}\mathbf{x} = \mathbf{0}$, then $\mathbf{x} \in \text{nul}(\mathbf{A}) \cap \text{nul}(\mathbf{A})^\perp = \{\mathbf{0}\}$, since $\text{col}(\mathbf{A}^*) = \text{nul}(\mathbf{A})^\perp$ by Theorem 11.6.5 (b). Surjectivity: the image is contained in $\text{col}(\mathbf{A})$, and by injectivity it has dimension $\dim \text{col}(\mathbf{A}^*) = \text{rank } \mathbf{A}^* = \text{rank } \mathbf{A} = \dim \text{col}(\mathbf{A})$, the middle equality by Corollary 11.6.6 (a) applied to $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$. So (i) prescribes $\mathbf{X}$ on $\text{col}(\mathbf{A})$ and (ii) prescribes it on $\text{col}(\mathbf{A})^\perp$; since $F^m = \text{col}(\mathbf{A}) \oplus \text{col}(\mathbf{A})^\perp$ by Theorem 11.3.2 (a), a linear map is determined by the two prescriptions.

The matrix $\mathbf{X} = \mathbf{V}\Sigma^+\mathbf{U}^*$ satisfies them. Write $\mathbf{u}_1, \dots, \mathbf{u}_m$ and $\mathbf{v}_1, \dots, \mathbf{v}_n$ for the columns of $\mathbf{U}$ and $\mathbf{V}$. Since rank is unchanged by invertible factors (Theorem 3.5.9), $r = \text{rank } \mathbf{A} = \text{rank } \Sigma$ is the number of nonzero σ_i . By Corollary 13.9.4, $(\mathbf{u}_1, \dots, \mathbf{u}_r)$ is an orthonormal basis of $\text{col}(\mathbf{A})$, the remaining $\mathbf{u}_j$ are an orthonormal basis of $\text{col}(\mathbf{A})^\perp$, and $(\mathbf{v}_1, \dots, \mathbf{v}_r)$ is

an orthonormal basis of $\text{col}(\mathbf{A}^*)$. Multiplying out the definition,

$$\mathbf{X} = \sum_{i=1}^r \sigma_i^{-1} \mathbf{v}_i \mathbf{u}_i^*,$$

so $\mathbf{X}\mathbf{u}_i = \sigma_i^{-1} \mathbf{v}_i$ for $i \leq r$ and $\mathbf{X}\mathbf{u}_j = \mathbf{0}$ for $j > r$. The second statement gives (ii) by linearity.

For (i), let $\mathbf{y} = \sum_{i \leq r} c_i \mathbf{u}_i \in \text{col}(\mathbf{A})$. Then $\mathbf{X}\mathbf{y} = \sum_{i \leq r} c_i \sigma_i^{-1} \mathbf{v}_i \in \text{col}(\mathbf{A}^*)$, and since $\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i$ (the rank-one expansion of Theorem 13.9.3 read one column at a time),

$$\mathbf{A}(\mathbf{X}\mathbf{y}) = \sum_{i \leq r} c_i \sigma_i^{-1} \sigma_i \mathbf{u}_i = \mathbf{y}.$$

By the injectivity proved above, $\mathbf{X}\mathbf{y}$ is the *unique* vector of $\text{col}(\mathbf{A}^*)$ with this property. This proves (i), and with it the proposition. ■

So $\mathbf{A}^+$ is an honest function of $\mathbf{A}$. Five examples, simplest first.

Example. Five pseudoinverses

- (a) **A invertible.** Then $m = n = r$, Σ is invertible and $\Sigma^+ = \Sigma^{-1}$, so $\mathbf{A}^+ = \mathbf{V}\Sigma^{-1}\mathbf{U}^* = (\mathbf{U}\Sigma\mathbf{V}^*)^{-1} = \mathbf{A}^{-1}$. The pseudoinverse extends the inverse, which is the first thing it had to do.
- (b) **A = O** $\in M_{m \times n}(F)$. Every singular value is 0, so $\Sigma^+ = \mathbf{O}$ and $\mathbf{A}^+ = \mathbf{O} \in M_{n \times m}(F)$. The degenerate case is worth keeping in mind: the pseudoinverse of the worst possible matrix is not undefined, it is zero, which is what “invert nothing, kill everything” means.
- (c) **Diagonal.** $\text{diag}(2, 0, 5)^+ = \text{diag}(\frac{1}{2}, 0, \frac{1}{5})$. The zero is left where it is.
- (d) **A single nonzero column** $\mathbf{a} \in M_{n \times 1}(F)$. Here $\sigma_1 = \|\mathbf{a}\|$, $\mathbf{u}_1 = \mathbf{a}/\|\mathbf{a}\|$, and $\mathbf{v}_1 = 1$, so $\mathbf{a}^+ = \mathbf{a}^*/\|\mathbf{a}\|^2$, a row vector. Then $\mathbf{a}^+\mathbf{a} = 1$ but $\mathbf{a}\mathbf{a}^+ = \mathbf{a}\mathbf{a}^*/\|\mathbf{a}\|^2$ is the orthogonal projection onto $\text{span}(\mathbf{a})$.
- (e) **A =** $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. This is $\sigma_1 \mathbf{u}_1 \mathbf{v}_1^*$ with $\sigma_1 = 2$ and $\mathbf{u}_1 = \mathbf{v}_1 = \frac{1}{\sqrt{2}}(1, 1)$, since $2\mathbf{u}_1 \mathbf{v}_1^* = 2 \cdot \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \mathbf{A}$. Hence

$$\mathbf{A}^+ = \frac{1}{2} \mathbf{v}_1 \mathbf{u}_1^* = \frac{1}{4} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

A nonexample by minimal change. Item (a) of the list is often written $\mathbf{A}^+ = (\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^*$, and Theorem 13.12.3 below confirms that formula whenever the columns of $\mathbf{A}$ are independent. Change one column of a 3×2 matrix into a multiple of the other, and the formula stops existing: $\mathbf{A}^* \mathbf{A}$ is then singular (Lemma 11.4.3), so $(\mathbf{A}^* \mathbf{A})^{-1}$ is not a matrix at all. The pseudoinverse survives the change; the formula does not. That is the difference between a definition and a recipe.

Why this definition. It is forced. Any reasonable “inverse” must undo $\mathbf{A}$ where $\mathbf{A}$ can be undone; there is exactly one subspace on which $\mathbf{A}$ is a bijection and which meets $\text{nul}(\mathbf{A})$ only in $\mathbf{0}$ while being determined by $\mathbf{A}$ alone, namely $\text{nul}(\mathbf{A})^\perp$; and setting the answer to $\mathbf{0}$ on $\text{col}(\mathbf{A})^\perp$ is the only choice that keeps the map linear without inventing information. The names are

those of the two people who published the idea, thirty-five years apart: Moore in 1920 through a projection formula, Penrose in 1955 through the four conditions of the next theorem.

Warning. $\mathbf{A}^+\mathbf{A}$ is not the identity, and $\mathbf{A}^+$ is $n \times m$. Two habits from the invertible case have to go. First the shape: if $\mathbf{A}$ is 3×5 then $\mathbf{A}^+$ is 5×3 , so $\mathbf{A}\mathbf{A}^+$ is 3×3 and $\mathbf{A}^+\mathbf{A}$ is 5×5 — they are not even the same size, let alone both $\mathbf{I}$. Second the identity: for $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ of Example (Five pseudoinverses) (e),

$$\mathbf{A}^+\mathbf{A} = \frac{1}{4} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \neq \mathbf{I}_2.$$

The right question is not “is it the identity?” but “the identity on *what?*”, and Theorem 13.12.3 answers it: $\mathbf{A}^+\mathbf{A}$ is the identity on $\text{col}(\mathbf{A}^*)$ and zero on $\text{nul}(\mathbf{A})$.

Quick check. Let $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$. Find $\mathbf{A}^+$, then compute $\mathbf{A}\mathbf{A}^+$ and $\mathbf{A}^+\mathbf{A}$ and identify each as a projection.

Solution. Here $\mathbf{A} = \mathbf{e}_1(1, 1)$ has rank one, so $\sigma_1 = \|\mathbf{e}_1\| \|(1, 1)\| = \sqrt{2}$ with $\mathbf{u}_1 = \mathbf{e}_1$ and $\mathbf{v}_1 = \frac{1}{\sqrt{2}}(1, 1)$, and

$$\mathbf{A}^+ = \sigma_1^{-1} \mathbf{v}_1 \mathbf{u}_1^* = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

Therefore

$$\mathbf{A}\mathbf{A}^+ = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{A}^+\mathbf{A} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Neither is $\mathbf{I}_2$. The first is the orthogonal projection onto $\text{span}(\mathbf{e}_1) = \text{col}(\mathbf{A})$; the second is the orthogonal projection onto $\text{span}((1, 1)) = \text{col}(\mathbf{A}^\top)$, the row space. Each is the identity exactly on the subspace named, and zero on its orthogonal complement — which is what “as close to the identity as the shapes allow” means.

13.12.3 The four Penrose conditions

The definition builds $\mathbf{A}^+$ out of a decomposition. The following theorem characterizes it by four equations that mention nothing but $\mathbf{A}$, $\mathbf{X}$, and the star. It is the second, and better, answer to the well-definedness question, and it is how one *verifies* a candidate pseudoinverse in practice: guess, then check four identities.

Theorem 13.12.2. The Penrose Conditions

Let $\mathbf{A} \in M_{m \times n}(F)$. Then $\mathbf{X} = \mathbf{A}^+$ is the **unique** matrix in $M_{n \times m}(F)$ satisfying all four of

- (P1) $\mathbf{A}\mathbf{X}\mathbf{A} = \mathbf{A}$;
- (P2) $\mathbf{X}\mathbf{A}\mathbf{X} = \mathbf{X}$;
- (P3) $(\mathbf{A}\mathbf{X})^* = \mathbf{A}\mathbf{X}$;
- (P4) $(\mathbf{X}\mathbf{A})^* = \mathbf{X}\mathbf{A}$.

Idea. Existence is a computation with Σ : each condition collapses to an identity between Σ , Σ^+ , and the diagonal matrix of r ones. Uniqueness is a puzzle. Given two solutions $\mathbf{X}_1$ and $\mathbf{X}_2$, the goal is a chain of equalities starting at $\mathbf{X}_1$ and ending at $\mathbf{X}_2$, so the plan is to walk to a symmetric midpoint: get $\mathbf{X}_1 = \mathbf{X}_1 \mathbf{A} \mathbf{X}_2$ by pushing $\mathbf{X}_2$ in from the right, then run the mirror image of the same argument with $\mathbf{X}_2$ as the subject. Two preliminary identities make the walk possible, and both come from starrng one of the conditions — the move that manufactures an $\mathbf{A}^*$ when the hypotheses offer only $\mathbf{A}$.

Proof. Existence. Let $\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^*$ and $\mathbf{X} = \mathbf{V} \Sigma^+ \mathbf{U}^*$. Using $\mathbf{V}^* \mathbf{V} = \mathbf{I}_n$ and $\mathbf{U}^* \mathbf{U} = \mathbf{I}_m$,

$$\mathbf{A} \mathbf{X} = \mathbf{U} \Sigma \Sigma^+ \mathbf{U}^*, \quad \mathbf{X} \mathbf{A} = \mathbf{V} \Sigma^+ \Sigma \mathbf{V}^*.$$

Now $\Sigma \Sigma^+ \in M_m(F)$ is diagonal with $\sigma_i \sigma_i^{-1} = 1$ in the first r diagonal places and 0 elsewhere, and $\Sigma^+ \Sigma \in M_n(F)$ is diagonal with the same pattern. Both are real diagonal, hence Hermitian, so $\mathbf{A} \mathbf{X}$ and $\mathbf{X} \mathbf{A}$ are Hermitian, which is (P3) and (P4). Moreover $\Sigma \Sigma^+ \Sigma = \Sigma$ and $\Sigma^+ \Sigma \Sigma^+ = \Sigma^+$, because multiplying a diagonal matrix by a diagonal matrix of ones and zeros keeps exactly the entries in the positions marked 1, and those are the positions where Σ and Σ^+ are nonzero anyway. Hence

$$\mathbf{A} \mathbf{X} \mathbf{A} = \mathbf{U} \Sigma \Sigma^+ \Sigma \mathbf{V}^* = \mathbf{A}, \quad \mathbf{X} \mathbf{A} \mathbf{X} = \mathbf{V} \Sigma^+ \Sigma \Sigma^+ \mathbf{U}^* = \mathbf{X},$$

which is (P1) and (P2).

Uniqueness. Suppose $\mathbf{X}$ satisfies (P1)–(P4). Two consequences:

$$\mathbf{X} \mathbf{X}^* \mathbf{A}^* = \mathbf{X}, \quad \mathbf{A} \mathbf{A}^* \mathbf{X}^* = \mathbf{A}. \quad (*)$$

For the first, $\mathbf{X} = \mathbf{X} \mathbf{A} \mathbf{X}$ by (P2), and $\mathbf{A} \mathbf{X} = (\mathbf{A} \mathbf{X})^* = \mathbf{X}^* \mathbf{A}^*$ by (P3), so $\mathbf{X} = \mathbf{X}(\mathbf{A} \mathbf{X}) = \mathbf{X} \mathbf{X}^* \mathbf{A}^*$. For the second, $\mathbf{A} = \mathbf{A} \mathbf{X} \mathbf{A}$ by (P1), and $\mathbf{X} \mathbf{A} = (\mathbf{X} \mathbf{A})^* = \mathbf{A}^* \mathbf{X}^*$ by (P4), so $\mathbf{A} = \mathbf{A}(\mathbf{X} \mathbf{A}) = \mathbf{A} \mathbf{A}^* \mathbf{X}^*$. Starrng the second identity of $(*)$ gives the form we also need,

$$\mathbf{X} \mathbf{A} \mathbf{A}^* = \mathbf{A}^*. \quad (**)$$

Now let $\mathbf{X}_1$ and $\mathbf{X}_2$ both satisfy (P1)–(P4). Starting from $\mathbf{X}_1$ and working towards a term containing $\mathbf{X}_2$,

$$\begin{aligned} \mathbf{X}_1 &= \mathbf{X}_1 \mathbf{X}_1^* \mathbf{A}^* = \mathbf{X}_1 \mathbf{X}_1^* (\mathbf{A} \mathbf{X}_2 \mathbf{A})^* \\ &= \mathbf{X}_1 \mathbf{X}_1^* \mathbf{A}^* (\mathbf{A} \mathbf{X}_2)^* = \mathbf{X}_1 \mathbf{X}_1^* \mathbf{A}^* \mathbf{A} \mathbf{X}_2 \\ &= \mathbf{X}_1 \mathbf{A} \mathbf{X}_2. \end{aligned}$$

The first equality is $(*)$ for $\mathbf{X}_1$; the second replaces $\mathbf{A}$ by $\mathbf{A} \mathbf{X}_2 \mathbf{A}$, which is (P1) for $\mathbf{X}_2$; the third reverses the product under the star; the fourth is (P3) for $\mathbf{X}_2$; and the last is $(*)$ for $\mathbf{X}_1$ again, applied to the first three factors.

Now the mirror argument, with $\mathbf{X}_2$ as the subject:

$$\begin{aligned} \mathbf{X}_1 \mathbf{A} \mathbf{X}_2 &= \mathbf{X}_1 \mathbf{A} \mathbf{A}^* \mathbf{X}_2^* \mathbf{X}_2 = \mathbf{A}^* \mathbf{X}_2^* \mathbf{X}_2 \\ &= (\mathbf{X}_2 \mathbf{A})^* \mathbf{X}_2 = \mathbf{X}_2 \mathbf{A} \mathbf{X}_2 = \mathbf{X}_2. \end{aligned}$$

Here the first equality replaces $\mathbf{A}$ by $\mathbf{A} \mathbf{A}^* \mathbf{X}_2^*$, which is $(*)$ for $\mathbf{X}_2$; the second is $(**)$ for $\mathbf{X}_1$; the fourth is (P4) for $\mathbf{X}_2$; and the last is (P2) for $\mathbf{X}_2$.

Hence $\mathbf{X}_1 = \mathbf{X}_2$. Since $\mathbf{A}^+$ satisfies the four conditions, every matrix satisfying them equals $\mathbf{A}^+$. This proves the theorem. ■

Both halves of the argument are the same two moves: star a condition to trade $\mathbf{A}$ for $\mathbf{A}^*$, then use $(*)$ to collapse a triple product. The conditions are also exactly what one should check when an explicit candidate is on the table, and several of the proofs below do nothing else.

Remark. The four conditions are not independent decoration. Dropping (P2) leaves the **generalized inverses**, matrices with $\mathbf{A}\mathbf{X}\mathbf{A} = \mathbf{A}$, of which there are many when $\mathbf{A}$ is singular; dropping (P3) and (P4) leaves more still. Each condition removes freedom, and only all four together remove all of it. Exercise C3 identifies the matrices for which $\mathbf{A}^*$ itself already satisfies them.

13.12.4 What the pseudoinverse does

Theorem 13.12.3. Properties of the pseudoinverse

Let $\mathbf{A} \in M_{m \times n}(F)$. Then:

- (a) $\mathbf{A}\mathbf{A}^+ = P_{\text{col}(\mathbf{A})}$ and $\mathbf{A}^+\mathbf{A} = P_{\text{col}(\mathbf{A}^*)}$, the orthogonal projections onto the column space of $\mathbf{A}$ in F^m and onto the column space of $\mathbf{A}^*$ in F^n ;
- (b) $(\mathbf{A}^+)^+ = \mathbf{A}$;
- (c) $(\mathbf{A}^+)^* = (\mathbf{A}^*)^+$;
- (d) $\text{rank } \mathbf{A}^+ = \text{rank } \mathbf{A}$;
- (e) $\mathbf{A}^+ = \mathbf{A}^{-1}$ if $\mathbf{A}$ is invertible; $\mathbf{A}^+ = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$ if $\text{rank } \mathbf{A} = n$; and $\mathbf{A}^+ = \mathbf{A}^*(\mathbf{A}\mathbf{A}^*)^{-1}$ if $\text{rank } \mathbf{A} = m$.

Idea. Clause (a) is the computation already done in the existence half of Theorem 13.12.2, read through Theorem 11.3.3. Clauses (b) and (c) are one line each, because the Penrose conditions are almost symmetric in $\mathbf{A}$ and $\mathbf{X}$, and starring all four of them produces the four conditions for $\mathbf{A}^*$. Clause (e) is three checks of the same kind: write down the candidate and confirm (P1)–(P4).

Proof.

- (a) With $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*$ and the notation of Proposition 13.12.1, the existence computation gave $\mathbf{A}\mathbf{A}^+ = \mathbf{U}\Sigma\Sigma^+\mathbf{U}^*$, where $\Sigma\Sigma^+$ is diagonal with r ones followed by zeros. Multiplying out,

$$\mathbf{A}\mathbf{A}^+ = \sum_{i=1}^r \mathbf{u}_i \mathbf{u}_i^*.$$

For $\mathbf{y} \in F^m$ this sends $\mathbf{y}$ to $\sum_{i \leq r} \langle \mathbf{y}, \mathbf{u}_i \rangle \mathbf{u}_i$, which by Theorem 11.3.3 (a) is the orthogonal projection of $\mathbf{y}$ onto $\text{span}(\mathbf{u}_1, \dots, \mathbf{u}_r) = \text{col}(\mathbf{A})$ (Corollary 13.9.4). The same computation with $\Sigma^+\Sigma$ gives $\mathbf{A}^+\mathbf{A} = \sum_{i \leq r} \mathbf{v}_i \mathbf{v}_i^* = P_{\text{col}(\mathbf{A}^*)}$.

- (b) The four Penrose conditions for the pair $(\mathbf{A}, \mathbf{X})$ become, when the roles of $\mathbf{A}$ and $\mathbf{X}$ are exchanged, the list (P2), (P1), (P4), (P3) — the same four conditions. So $\mathbf{A}$ satisfies the Penrose conditions for $\mathbf{A}^+$, and [Theorem 13.12.2](#) gives $(\mathbf{A}^+)^+ = \mathbf{A}$.
- (c) Put $\mathbf{X} = (\mathbf{A}^+)^*$ and check the four conditions for $\mathbf{A}^*$. Using $(\mathbf{BC})^* = \mathbf{C}^*\mathbf{B}^*$ throughout,

$$\mathbf{A}^*\mathbf{X}\mathbf{A}^* = (\mathbf{A}\mathbf{A}^+\mathbf{A})^* = \mathbf{A}^*, \quad \mathbf{X}\mathbf{A}^*\mathbf{X} = (\mathbf{A}^+\mathbf{A}\mathbf{A}^+)^* = \mathbf{X},$$

which are (P1) and (P2). Next, $\mathbf{A}^*\mathbf{X} = (\mathbf{A}^+\mathbf{A})^* = \mathbf{A}^+\mathbf{A}$ by (P4) for $\mathbf{A}$, and $\mathbf{A}^+\mathbf{A}$ is Hermitian, so (P3) holds; and $\mathbf{X}\mathbf{A}^* = (\mathbf{A}\mathbf{A}^+)^* = \mathbf{A}\mathbf{A}^+$ by (P3) for $\mathbf{A}$, which is Hermitian, so (P4) holds. By uniqueness, $\mathbf{X} = (\mathbf{A}^+)^*$.

- (d) $\mathbf{A}^+ = \mathbf{V}\Sigma^+\mathbf{U}^*$ with $\mathbf{U}, \mathbf{V}$ invertible, so $\text{rank } \mathbf{A}^+ = \text{rank } \Sigma^+ = r$ by [Theorem 3.5.9](#), the last equality because Σ^+ has exactly r nonzero entries, all in distinct rows and columns.
- (e) The invertible case is [Example \(Five pseudoinverses\)](#) (a). Suppose $\text{rank } \mathbf{A} = n$. Then $\mathbf{A}^*\mathbf{A}$ is invertible ([Corollary 11.4.4](#)), so $\mathbf{X} = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$ is defined, and $((\mathbf{A}^*\mathbf{A})^{-1})^* = (\mathbf{A}^*\mathbf{A})^{-1}$ since $\mathbf{A}^*\mathbf{A}$ is Hermitian. Then

$$\mathbf{X}\mathbf{A} = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{A} = \mathbf{I}_n,$$

which is Hermitian, giving (P4); and (P1), (P2) follow at once, since $\mathbf{A}\mathbf{X}\mathbf{A} = \mathbf{A}\mathbf{I}_n = \mathbf{A}$ and $\mathbf{X}\mathbf{A}\mathbf{X} = \mathbf{I}_n\mathbf{X} = \mathbf{X}$. For (P3), $\mathbf{A}\mathbf{X} = \mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$, whose conjugate transpose is itself. So $\mathbf{X} = \mathbf{A}^+$.

For the last formula, suppose $\text{rank } \mathbf{A} = m$. Then $\text{rank } \mathbf{A}^* = m$ is the number of columns of $\mathbf{A}^*$, so the case just proved applies to $\mathbf{A}^*$ and gives $(\mathbf{A}^+)^+ = (\mathbf{A}\mathbf{A}^*)^{-1}\mathbf{A}$. By (c),

$$\mathbf{A}^+ = ((\mathbf{A}^+)^+)^* = ((\mathbf{A}\mathbf{A}^*)^{-1}\mathbf{A})^* = \mathbf{A}^*(\mathbf{A}\mathbf{A}^*)^{-1},$$

using that $\mathbf{A}\mathbf{A}^*$ is Hermitian, hence so is its inverse. This proves the theorem. ■

Clause (a) is the one to remember, and it answers the question the warning raised. The products $\mathbf{A}\mathbf{A}^+$ and $\mathbf{A}^+\mathbf{A}$ are not identities; they are the identity **restricted to the subspaces where $\mathbf{A}$ is invertible**, and zero on the complements. Clause (e) shows that the three recipes of Chapter 11 were all the pseudoinverse in disguise, each written in the notation that its hypothesis made available.

Warning. $(\mathbf{AB})^+ \neq \mathbf{B}^+\mathbf{A}^+$ in general. The rule $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ does not survive. Take

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

The Quick check above found $\mathbf{A}^+ = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$, and one may confirm it against [Theorem 13.12.2](#): the two products computed there are Hermitian, and $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$, $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$. Meanwhile $\mathbf{B}$ is Hermitian and idempotent, so $\mathbf{X} = \mathbf{B}$ satisfies (P1)–(P4) and $\mathbf{B}^+ = \mathbf{B}$. Now $\mathbf{AB} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \mathbf{B}$, so

$$(\mathbf{AB})^+ = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{B}^+\mathbf{A}^+ = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

The two differ by a factor of 2. The reversal rule does hold under extra hypotheses — for

instance when $\mathbf{A}$ has independent columns and $\mathbf{B}$ has independent rows, which is Exercise C2 — but never assume it.

13.12.5 Least squares and minimum norm, in one matrix

Everything is now in place to pay Chapter 11's debt.

Theorem 13.12.4. The pseudoinverse solves both problems at once

Let $\mathbf{A} \in M_{m \times n}(F)$ and $\mathbf{b} \in F^m$, and put $\mathbf{x}^+ = \mathbf{A}^+\mathbf{b}$. Then:

- (a) $\mathbf{x}^+$ is a least-squares solution of $\mathbf{Ax} = \mathbf{b}$, that is, $\|\mathbf{Ax}^+ - \mathbf{b}\| \leq \|\mathbf{Az} - \mathbf{b}\|$ for every $\mathbf{z} \in F^n$;
- (b) $\|\mathbf{x}^+\| < \|\mathbf{x}\|$ for every other least-squares solution $\mathbf{x}$;
- (c) if $\mathbf{Ax} = \mathbf{b}$ is consistent, $\mathbf{x}^+$ is its minimum-norm solution in the sense of Theorem 11.4.5; if $\mathbf{A}$ is invertible, $\mathbf{x}^+ = \mathbf{A}^{-1}\mathbf{b}$.

Idea. Both halves are already proved; the pseudoinverse only has to be recognized inside them. For (a), $\mathbf{Ax}^+ = \mathbf{AA}^+\mathbf{b}$ is the projection of $\mathbf{b}$ onto $\text{col}(\mathbf{A})$, which is precisely condition (b) of the Least Squares Theorem. For (b), the least-squares solutions are the solutions of one consistent system, so the Minimum-Norm Solution Theorem applies to that system and singles out the solution lying in $\text{nul}(\mathbf{A})^\perp$; and $\mathbf{A}^+\mathbf{b}$ lies there, because every column of $\mathbf{A}^+$ does.

Proof.

- (a) By Theorem 13.12.3 (a), $\mathbf{Ax}^+ = \mathbf{AA}^+\mathbf{b} = P_U\mathbf{b}$ with $U = \text{col}(\mathbf{A})$. This is condition (b) of Theorem 11.4.2, so $\mathbf{x}^+$ is a least-squares solution.
- (b) By Theorem 11.4.2, the least-squares solutions of $\mathbf{Ax} = \mathbf{b}$ are exactly the solutions of the system $\mathbf{Ax} = P_U\mathbf{b}$, which is consistent because $P_U\mathbf{b} \in U = \text{col}(\mathbf{A})$. Apply Theorem 11.4.5 to that system: exactly one of its solutions lies in $\text{nul}(\mathbf{A})^\perp$, and it is strictly shorter than all the others.
It remains to see that $\mathbf{x}^+$ is that one. By (P2), $\mathbf{A}^+ = \mathbf{A}^+\mathbf{AA}^+ = P_{\text{col}(\mathbf{A}^*)}\mathbf{A}^+$ using Theorem 13.12.3 (a), so every column of $\mathbf{A}^+$, and hence every vector $\mathbf{A}^+\mathbf{b}$, lies in $\text{col}(\mathbf{A}^*)$. By Theorem 11.6.5 (b), $\text{col}(\mathbf{A}^*) = \text{nul}(\mathbf{A})^\perp$. So $\mathbf{x}^+$ is the minimum-norm solution of $\mathbf{Ax} = P_U\mathbf{b}$, which by the previous paragraph is what (b) asserts.
- (c) If $\mathbf{Ax} = \mathbf{b}$ is consistent then $\mathbf{b} \in \text{col}(\mathbf{A})$, so $P_U\mathbf{b} = \mathbf{b}$ and the least-squares solutions are the actual solutions; (b) then says $\mathbf{x}^+$ is the shortest of them, which is Theorem 11.4.5's $\mathbf{x}_{\min}$. If $\mathbf{A}$ is invertible, $\mathbf{A}^+ = \mathbf{A}^{-1}$ by Theorem 13.12.3 (e). This proves the theorem. ■

The promise, discharged. Section 4 of Chapter 11 said that the two answers are usually wanted together — first minimize $\|\mathbf{Ax} - \mathbf{b}\|$, then among the minimizers take the shortest — and that Chapter 13 would package the combined answer as a single matrix $\mathbf{A}^+$, with $\mathbf{x} = \mathbf{A}^+\mathbf{b}$

in all cases at once. That is [Theorem 13.12.4](#), word for word: clause (a) does the minimizing, clause (b) takes the shortest, and clause (c) plus [Theorem 13.12.3](#) (e) checks that the three special recipes of Chapter 11 are recovered as special cases. There is no hypothesis on $\mathbf{A}$ anywhere in the statement.

The notation, settled. Chapter 11 wrote $\mathbf{x}_{\min}$ for the minimum-norm solution of a *consistent* system and left the symbol without a formula. From now on it has one, and its meaning extends to every system:

$$\mathbf{x}_{\min} = \mathbf{A}^+ \mathbf{b},$$

the minimum-norm least-squares solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$. For a consistent system this agrees with Chapter 11's $\mathbf{x}_{\min}$, by [Theorem 13.12.4](#) (c); for an inconsistent one it is the shortest of the minimizers of $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$, by clauses (a) and (b). One symbol, one formula, no case distinction.

Example. Chapter 11's minimum-norm solution, recomputed

In Chapter 11 the system

$$x_1 + x_2 + x_3 = 3, \quad x_1 + 2x_2 + 3x_3 = 7$$

was solved for its shortest solution by intersecting the solution line with $\text{nul}(\mathbf{A})^\perp$, as [Theorem 11.4.5](#) prescribes. Recompute $\mathbf{x}_{\min}$ as $\mathbf{A}^+ \mathbf{b}$.

Solution. Here

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix}, \quad \mathbf{b} = (3, 7).$$

The two rows are independent, so $\text{rank } \mathbf{A} = 2 = m$ and [Theorem 13.12.3](#) (e) gives $\mathbf{A}^+ = \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1}$. Now

$$\mathbf{A}\mathbf{A}^\top = \begin{pmatrix} 3 & 6 \\ 6 & 14 \end{pmatrix}, \quad (\mathbf{A}\mathbf{A}^\top)^{-1} = \frac{1}{6} \begin{pmatrix} 14 & -6 \\ -6 & 3 \end{pmatrix},$$

since the determinant is $42 - 36 = 6$. Therefore

$$\mathbf{A}^+ = \frac{1}{6} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 14 & -6 \\ -6 & 3 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 8 & -3 \\ 2 & 0 \\ -4 & 3 \end{pmatrix},$$

and

$$\mathbf{x}_{\min} = \mathbf{A}^+ \mathbf{b} = \frac{1}{6} \begin{pmatrix} 24 - 21 \\ 6 + 0 \\ -12 + 21 \end{pmatrix} = \left(\frac{1}{2}, 1, \frac{3}{2} \right).$$

This is exactly the answer Chapter 11 obtained by hand, and $\|\mathbf{x}_{\min}\|^2 = \frac{1}{4} + 1 + \frac{9}{4} = \frac{7}{2}$ as before. The system is consistent — $\frac{1}{2} + 1 + \frac{3}{2} = 3$ and $\frac{1}{2} + 2 + \frac{9}{2} = 7$ — so no projection was needed; what the pseudoinverse did was replace a geometric argument by one matrix product.

The next example is the case no formula of Chapter 11 reaches: the system is inconsistent **and** underdetermined at the same time.

Example. Inconsistent and underdetermined at once

Find $\mathbf{x}_{\min}$ for

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{b} = (1, 3),$$

and verify directly that it minimizes the residual and is the shortest minimizer.

Solution. The system reads $x_1 + x_2 = 1$ and $x_1 + x_2 = 3$, so it is inconsistent; and $\text{nul}(\mathbf{A}) = \text{span}((1, -1)) \neq \{\mathbf{0}\}$, so the columns are dependent. Neither $\mathbf{A}^{-1}$ nor $(\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^*$ exists. From [Example \(Five pseudoinverses\)](#) (e), $\mathbf{A}^+ = \frac{1}{4} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, so

$$\mathbf{x}_{\min} = \mathbf{A}^+ \mathbf{b} = \frac{1}{4}(1 + 3, 1 + 3) = (1, 1).$$

Check. $\text{col}(\mathbf{A}) = \text{span}((1, 1))$, and the projection of $\mathbf{b} = (1, 3)$ onto it is $\frac{1+3}{2}(1, 1) = (2, 2)$. Indeed $\mathbf{A}(1, 1) = (2, 2)$, so the residual is $\mathbf{b} - (2, 2) = (-1, 1)$, of squared length 2 — the smallest possible. The least-squares solutions are all $\mathbf{x}$ with $x_1 + x_2 = 2$, that is $(1, 1) + t(1, -1)$, and

$$\|(1, 1) + t(1, -1)\|^2 = 2 + 2t^2,$$

which is smallest exactly at $t = 0$. So $(1, 1)$ is the shortest least-squares solution, as [Theorem 13.12.4](#) (b) predicts.

13.12.6 The block criterion, finished

Section 6 tested a block matrix for positive definiteness one block at a time, and had to assume the corner block invertible: the Schur complement $\mathbf{M}/\mathbf{A} = \mathbf{C} - \mathbf{B}^* \mathbf{A}^{-1} \mathbf{B}$ is meaningless otherwise. The pseudoinverse removes that hypothesis, and the promise Section 6 made is paid here.

Proposition 13.12.5. The Semidefinite Block Criterion

Let $\mathbf{A} \in M_k(F)$ and $\mathbf{C} \in M_l(F)$ be self-adjoint and put

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{C} \end{pmatrix} \in M_{k+l}(F).$$

Then $\mathbf{M} \succeq \mathbf{O}$ if and only if $\mathbf{A} \succeq \mathbf{O}$, every column of $\mathbf{B}$ lies in $\text{col}(\mathbf{A})$, and $\mathbf{C} - \mathbf{B}^* \mathbf{A}^+ \mathbf{B} \succeq \mathbf{O}$.

Idea. Section 6 cleared the corner with the shear $\begin{pmatrix} \mathbf{I} & -\mathbf{A}^{-1} \mathbf{B} \\ \mathbf{O} & \mathbf{I} \end{pmatrix}$. Write $\mathbf{A}^+$ in place of $\mathbf{A}^{-1}$ and the same shear still clears it — but only because $\mathbf{A} \mathbf{A}^+$ is the projection onto $\text{col}(\mathbf{A})$, so $\mathbf{A} \mathbf{A}^+ \mathbf{B} = \mathbf{B}$ exactly when the columns of $\mathbf{B}$ are already in $\text{col}(\mathbf{A})$. That condition is not a technicality: it is what the two-sided test forces.

Proof. Suppose first that $\mathbf{M} \succeq \mathbf{O}$. Taking $\mathbf{v} = (\mathbf{x}, \mathbf{0})$ gives $\langle \mathbf{A} \mathbf{x}, \mathbf{x} \rangle \geq 0$, so $\mathbf{A} \succeq \mathbf{O}$. For the

column condition, let $\mathbf{x} \in \text{nul}(\mathbf{A})$ and $\mathbf{y} \in F^l$, and put $\mathbf{v}_t = (t\mathbf{x}, \mathbf{y})$ for $t \in \mathbb{R}$. Then

$$\langle \mathbf{M}\mathbf{v}_t, \mathbf{v}_t \rangle = t^2 \langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle + 2t \Re \langle \mathbf{B}\mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{C}\mathbf{y}, \mathbf{y} \rangle = 2t \Re \langle \mathbf{B}\mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{C}\mathbf{y}, \mathbf{y} \rangle,$$

using $\mathbf{A}\mathbf{x} = \mathbf{0}$. A real affine function of t that is nonnegative for every t has zero slope, so $\Re \langle \mathbf{B}\mathbf{y}, \mathbf{x} \rangle = 0$ for all $\mathbf{y}$; replacing $\mathbf{y}$ by $i\mathbf{y}$ over $\mathbb{C}$ kills the imaginary part too. Hence $\langle \mathbf{B}\mathbf{y}, \mathbf{x} \rangle = 0$ for every $\mathbf{y}$, that is $\mathbf{B}^*\mathbf{x} = \mathbf{0}$. So $\text{nul}(\mathbf{A}) \subseteq \text{nul}(\mathbf{B}^*)$, and taking orthogonal complements with Theorem 11.6.5 gives $\text{col}(\mathbf{B}) \subseteq \text{col}(\mathbf{A})$, since $\mathbf{A}$ is self-adjoint.

Now assume $\mathbf{A} \succeq \mathbf{0}$ and $\text{col}(\mathbf{B}) \subseteq \text{col}(\mathbf{A})$; we show $\mathbf{M} \succeq \mathbf{0}$ is equivalent to $\mathbf{C} - \mathbf{B}^*\mathbf{A}^+\mathbf{B} \succeq \mathbf{0}$, which finishes both directions. By Theorem 13.12.3 (a), $\mathbf{A}\mathbf{A}^+ = P_{\text{col}(\mathbf{A})}$, so $\mathbf{A}\mathbf{A}^+\mathbf{B} = \mathbf{B}$; taking adjoints, and using that $\mathbf{A}$ and $\mathbf{A}^+$ are self-adjoint (Theorem 13.12.3 (c)), also $\mathbf{B}^*\mathbf{A}^+\mathbf{A} = \mathbf{B}^*$. Put

$$\mathbf{S} = \begin{pmatrix} \mathbf{I}_k & -\mathbf{A}^+\mathbf{B} \\ \mathbf{0} & \mathbf{I}_l \end{pmatrix},$$

which is invertible. Multiplying out and canceling with the two identities just recorded,

$$\mathbf{S}^*\mathbf{M}\mathbf{S} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} - \mathbf{B}^*\mathbf{A}^+\mathbf{B} \end{pmatrix}.$$

Since $\mathbf{S}$ is invertible, Proposition 13.1.2 (a) applied to $\mathbf{S}$ and then to $\mathbf{S}^{-1}$ shows that $\mathbf{M} \succeq \mathbf{0}$ if and only if $\mathbf{S}^*\mathbf{M}\mathbf{S} \succeq \mathbf{0}$, and a block diagonal matrix is positive semidefinite exactly when both blocks are. This shows the three conditions together are equivalent to $\mathbf{M} \succeq \mathbf{0}$. ■

The column condition cannot be dropped. With

$$\mathbf{A} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{C} = (1),$$

the choice $\mathbf{B} = (1, 0)^\top$ has its column in $\text{col}(\mathbf{A})$ and gives $\mathbf{C} - \mathbf{B}^*\mathbf{A}^+\mathbf{B} = (\frac{1}{2}) \succeq \mathbf{0}$, so $\mathbf{M} \succeq \mathbf{0}$. The choice $\mathbf{B} = (0, 1)^\top$ does not, and there $\mathbf{M}$ is not positive semidefinite even though $\mathbf{A} \succeq \mathbf{0}$, $\mathbf{C} \succeq \mathbf{0}$ and $\mathbf{B}^*\mathbf{A}^+\mathbf{B} = \mathbf{0}$ makes the third condition hold trivially.

13.12.7 Exercises

13.12.7.1 A. Check your understanding

Exercise. A1

- Define $\mathbf{A}^+$ in terms of a singular value decomposition of $\mathbf{A}$ (Theorem 13.9.1), and say what size it is.
- State the four Penrose conditions, and say what Theorem 13.12.2 asserts about them.
- What are $\mathbf{A}\mathbf{A}^+$ and $\mathbf{A}^+\mathbf{A}$? Under what condition on $\mathbf{A}$ is the second of them $\mathbf{I}_n$?
- Write down $\mathbf{O}^+$ for the zero matrix $\mathbf{O} \in M_{m \times n}(F)$, and $\mathbf{A}^+$ for an invertible $\mathbf{A}$.
- True or false: $(\mathbf{A}\mathbf{B})^+ = \mathbf{B}^+\mathbf{A}^+$ whenever the product $\mathbf{A}\mathbf{B}$ is defined. Justify your answer.

Solution.

- (a) If $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$, then $\mathbf{A}^+ = \mathbf{V}\mathbf{\Sigma}^+\mathbf{U}^*$, where $\mathbf{\Sigma}^+ \in M_{n \times m}(F)$ has σ_i^{-1} in the (i, i) place for each $\sigma_i > 0$ and zeros elsewhere. It is $n \times m$, the transpose of the shape of $\mathbf{A}$.
- (b) $\mathbf{A}\mathbf{X}\mathbf{A} = \mathbf{A}$, $\mathbf{X}\mathbf{A}\mathbf{X} = \mathbf{X}$, $(\mathbf{A}\mathbf{X})^* = \mathbf{A}\mathbf{X}$, $(\mathbf{X}\mathbf{A})^* = \mathbf{X}\mathbf{A}$. The theorem says $\mathbf{A}^+$ satisfies all four and is the only matrix that does; so the four conditions may be used as a definition, and checking them is enough to identify a candidate.
- (c) $\mathbf{A}\mathbf{A}^+ = P_{\text{col}(\mathbf{A})}$ and $\mathbf{A}^+\mathbf{A} = P_{\text{col}(\mathbf{A}^*)}$ (Theorem 13.12.3 (a)). The second is $\mathbf{I}_n$ exactly when $\text{col}(\mathbf{A}^*) = F^n$, that is when $\text{rank } \mathbf{A} = n$, that is when the columns of $\mathbf{A}$ are independent.
- (d) $\mathbf{O}^+ = \mathbf{O} \in M_{n \times m}(F)$; $\mathbf{A}^+ = \mathbf{A}^{-1}$ when $\mathbf{A}$ is invertible.
- (e) False. With $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ the two sides are $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, as computed in the warning above.

13.12.7.2 B. Practice

Exercise. B1: Four pseudoinverses

Compute $\mathbf{A}^+$ for each of the following, stating which method you used.

- (a) $\mathbf{A} = \text{diag}(3, 0, 0, 4) \in M_4(\mathbb{R})$;
- (b) $\mathbf{A} = (1, 2, 2)^\top \in M_{3 \times 1}(\mathbb{R})$;
- (c) $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \in M_2(\mathbb{R})$;
- (d) $\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \in M_{2 \times 3}(\mathbb{R})$.

Solution.

- (a) A diagonal matrix is its own singular value decomposition (Theorem 13.9.1) up to ordering, so invert the nonzero entries in place: $\mathbf{A}^+ = \text{diag}(\frac{1}{3}, 0, 0, \frac{1}{4})$.
- (b) A single nonzero column, so Example (Five pseudoinverses) (d) applies with $\|\mathbf{A}\|^2 = 1 + 4 + 4 = 9$:

$$\mathbf{A}^+ = \frac{1}{9}(1, 2, 2) \in M_{1 \times 3}(\mathbb{R}).$$

- (c) Rank one: $\mathbf{A} = \mathbf{u}\mathbf{v}^\top$ with $\mathbf{u} = (1, 2)$ and $\mathbf{v} = (1, 2)$. Then $\sigma_1 = \|\mathbf{u}\| \|\mathbf{v}\| = 5$ with $\mathbf{u}_1 = \mathbf{u}/\sqrt{5}$ and $\mathbf{v}_1 = \mathbf{v}/\sqrt{5}$, so

$$\mathbf{A}^+ = \frac{1}{5}\mathbf{v}_1\mathbf{u}_1^\top = \frac{1}{25} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}.$$

(Check: $\mathbf{A}\mathbf{A}^+ = \frac{1}{25} \begin{pmatrix} 5 & 10 \\ 10 & 20 \end{pmatrix}$ is Hermitian and idempotent, and $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$.)

- (d) The rows are independent, so $\text{rank } \mathbf{A} = 2 = m$ and Theorem 13.12.3 (e) gives $\mathbf{A}^+ =$

$\mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}$. Here $\mathbf{A}\mathbf{A}^\top = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ with inverse $\frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$, so

$$\mathbf{A}^+ = \frac{1}{3} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}.$$

Exercise. B2: A system that is both inconsistent and underdetermined

Let

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \\ 1 & 2 \end{pmatrix}, \quad \mathbf{b} = (1, 4, 1).$$

Compute $\mathbf{A}^+$ and $\mathbf{x}_{\min} = \mathbf{A}^+\mathbf{b}$. Verify that $\mathbf{A}\mathbf{x}_{\min}$ is the orthogonal projection of $\mathbf{b}$ onto $\text{col}(\mathbf{A})$, and compare $\|\mathbf{x}_{\min}\|$ with the norm of one other least-squares solution.

Solution. Both columns are multiples of $\mathbf{u} = (1, 2, 1)$, so $\mathbf{A} = \mathbf{u}\mathbf{v}^\top$ with $\mathbf{v} = (1, 2)$, and $\text{rank } \mathbf{A} = 1$. Then $\|\mathbf{u}\|^2 = 6$, $\|\mathbf{v}\|^2 = 5$, and $\sigma_1 = \sqrt{30}$, with $\mathbf{u}_1 = \mathbf{u}/\sqrt{6}$ and $\mathbf{v}_1 = \mathbf{v}/\sqrt{5}$. Hence

$$\mathbf{A}^+ = \sigma_1^{-1} \mathbf{v}_1 \mathbf{u}_1^\top = \frac{1}{30} \mathbf{v} \mathbf{u}^\top = \frac{1}{30} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \end{pmatrix}.$$

Therefore

$$\mathbf{x}_{\min} = \mathbf{A}^+\mathbf{b} = \frac{1}{30}(1 + 8 + 1, 2 + 16 + 2) = \left(\frac{1}{3}, \frac{2}{3}\right).$$

Projection. $\text{col}(\mathbf{A}) = \text{span}(\mathbf{u})$ and $\langle \mathbf{b}, \mathbf{u} \rangle = 1 + 8 + 1 = 10$, so $P_{\text{col}(\mathbf{A})}\mathbf{b} = \frac{10}{6}\mathbf{u} = \frac{5}{3}(1, 2, 1)$. And $\mathbf{A}\mathbf{x}_{\min} = \mathbf{u}(\mathbf{v}^\top \mathbf{x}_{\min}) = \mathbf{u}(\frac{1}{3} + \frac{4}{3}) = \frac{5}{3}\mathbf{u}$. They agree.

Shortest. The least-squares solutions are the $\mathbf{x}$ with $\mathbf{v}^\top \mathbf{x} = x_1 + 2x_2 = \frac{5}{3}$. Another one is $(\frac{5}{3}, 0)$, of squared norm $\frac{25}{9}$, against $\|\mathbf{x}_{\min}\|^2 = \frac{1}{9} + \frac{4}{9} = \frac{5}{9}$. The pseudoinverse solution is shorter, as [Theorem 13.12.4](#) (b) requires.

Exercise. B3: Checking a candidate

Let $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $\mathbf{X} = \frac{1}{3} \begin{pmatrix} 2 & 1 & -1 \\ -1 & 1 & 2 \end{pmatrix}$. Verify the four Penrose conditions for $\mathbf{X}$, and hence find the minimum-norm least-squares solution of $\mathbf{A}\mathbf{x} = (1, 2, 3)$.

Solution. Compute the two products first:

$$\mathbf{X}\mathbf{A} = \frac{1}{3} \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} = \mathbf{I}_2, \quad \mathbf{A}\mathbf{X} = \frac{1}{3} \begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}.$$

(P4): $\mathbf{X}\mathbf{A} = \mathbf{I}_2$ is Hermitian. (P3): $\mathbf{A}\mathbf{X}$ is visibly symmetric. (P1): $\mathbf{A}\mathbf{X}\mathbf{A} = \mathbf{A}(\mathbf{X}\mathbf{A}) =$

$\mathbf{A}\mathbf{I}_2 = \mathbf{A}$. (P2): $\mathbf{X}\mathbf{A}\mathbf{X} = \mathbf{I}_2\mathbf{X} = \mathbf{X}$. All four hold, so $\mathbf{X} = \mathbf{A}^+$ by Theorem 13.12.2.

Then

$$\mathbf{x}_{\min} = \mathbf{X}(1, 2, 3) = \frac{1}{3}(2 + 2 - 3, -1 + 2 + 6) = \left(\frac{1}{3}, \frac{7}{3}\right).$$

(The columns of $\mathbf{A}$ are independent, so this is also $(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$, and it is the unique least-squares solution; the normal equations $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \mathbf{x} = (3, 5)$ confirm it.)

13.12.7.3 C. Going deeper

Exercise. C1: Pseudoinverses of $\mathbf{A}^* \mathbf{A}$

Let $\mathbf{A} \in M_{m \times n}(F)$.

- (a) Prove that $(\mathbf{A}^* \mathbf{A})^+ = \mathbf{A}^+ (\mathbf{A}^+)^*$.
- (b) Deduce that $\mathbf{A}^+ = (\mathbf{A}^* \mathbf{A})^+ \mathbf{A}^*$, the formula $(\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^*$ with the inverse replaced by the pseudoinverse.

Hint: use a singular value decomposition and reduce each identity to an identity between diagonal matrices.

Solution. Write $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^*$.

- (a) Then $\mathbf{A}^* \mathbf{A} = \mathbf{V} \mathbf{\Sigma}^* \mathbf{\Sigma} \mathbf{V}^*$, and $\mathbf{\Sigma}^* \mathbf{\Sigma} \in M_n(F)$ is diagonal with entries $\sigma_1^2, \dots, \sigma_r^2$ followed by zeros. That display is a singular value decomposition of $\mathbf{A}^* \mathbf{A}$ in the sense of Theorem 13.9.1 (both outer factors are $\mathbf{V}$, which is unitary, and the middle factor is diagonal with nonnegative entries), so

$$(\mathbf{A}^* \mathbf{A})^+ = \mathbf{V} (\mathbf{\Sigma}^* \mathbf{\Sigma})^+ \mathbf{V}^*,$$

where $(\mathbf{\Sigma}^* \mathbf{\Sigma})^+$ is diagonal with entries σ_i^{-2} for $i \leq r$ and 0 elsewhere. On the other side, $(\mathbf{A}^+)^* = (\mathbf{V} \mathbf{\Sigma}^+ \mathbf{U}^*)^* = \mathbf{U} (\mathbf{\Sigma}^+)^* \mathbf{V}^*$, so

$$\mathbf{A}^+ (\mathbf{A}^+)^* = \mathbf{V} \mathbf{\Sigma}^+ \mathbf{U}^* \mathbf{U} (\mathbf{\Sigma}^+)^* \mathbf{V}^* = \mathbf{V} \mathbf{\Sigma}^+ (\mathbf{\Sigma}^+)^* \mathbf{V}^*.$$

The middle factor $\mathbf{\Sigma}^+ (\mathbf{\Sigma}^+)^*$ is diagonal with entries $\sigma_i^{-1} \sigma_i^{-1} = \sigma_i^{-2}$ for $i \leq r$ and 0 elsewhere, which is $(\mathbf{\Sigma}^* \mathbf{\Sigma})^+$. Hence the two sides agree.

- (b) Using (a) and $\mathbf{A}^* = \mathbf{V} \mathbf{\Sigma}^* \mathbf{U}^*$,

$$(\mathbf{A}^* \mathbf{A})^+ \mathbf{A}^* = \mathbf{V} \mathbf{\Sigma}^+ (\mathbf{\Sigma}^+)^* \mathbf{\Sigma}^* \mathbf{U}^*.$$

Now $(\mathbf{\Sigma}^+)^* \mathbf{\Sigma}^* = (\mathbf{\Sigma} \mathbf{\Sigma}^+)^* = \mathbf{\Sigma} \mathbf{\Sigma}^+$, which is the $m \times m$ diagonal matrix with r ones and then zeros; and $\mathbf{\Sigma}^+$ multiplied by it on the right is $\mathbf{\Sigma}^+$, since $\mathbf{\Sigma}^+$ has nonzero entries only in the first r columns. So the right-hand side is $\mathbf{V} \mathbf{\Sigma}^+ \mathbf{U}^* = \mathbf{A}^+$.

Exercise. C2: When the reversal rule does hold

Let $\mathbf{A} \in M_{m \times n}(F)$ with $\text{rank } \mathbf{A} = n$ and $\mathbf{B} \in M_{n \times q}(F)$ with $\text{rank } \mathbf{B} = n$. Prove that

$$(\mathbf{A}\mathbf{B})^+ = \mathbf{B}^+ \mathbf{A}^+.$$

Hint: write both pseudoinverses with the formulas of Theorem 13.12.3 (e) and verify the Penrose conditions for the product.

Solution. By Theorem 13.12.3 (e), $\mathbf{A}^+ = (\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^*$ and $\mathbf{B}^+ = \mathbf{B}^* (\mathbf{B} \mathbf{B}^*)^{-1}$, the two inverses existing because $\text{rank } \mathbf{A} = n$ and $\text{rank } \mathbf{B}^* = n$. Put

$$\mathbf{X} = \mathbf{B}^+ \mathbf{A}^+ = \mathbf{B}^* (\mathbf{B} \mathbf{B}^*)^{-1} (\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^*.$$

Throughout we use $\mathbf{A}^+ \mathbf{A} = \mathbf{I}_n$ and $\mathbf{B} \mathbf{B}^+ = \mathbf{I}_n$, which the two formulas give at once.

$$(P1): (\mathbf{A} \mathbf{B}) \mathbf{X} (\mathbf{A} \mathbf{B}) = \mathbf{A} (\mathbf{B} \mathbf{B}^+) (\mathbf{A}^+ \mathbf{A}) \mathbf{B} = \mathbf{A} \mathbf{I}_n \mathbf{I}_n \mathbf{B} = \mathbf{A} \mathbf{B}.$$

$$(P2): \mathbf{X} (\mathbf{A} \mathbf{B}) \mathbf{X} = \mathbf{B}^+ (\mathbf{A}^+ \mathbf{A}) (\mathbf{B} \mathbf{B}^+) \mathbf{A}^+ = \mathbf{B}^+ \mathbf{A}^+ = \mathbf{X}.$$

(P3): $(\mathbf{A} \mathbf{B}) \mathbf{X} = \mathbf{A} (\mathbf{B} \mathbf{B}^+) \mathbf{A}^+ = \mathbf{A} \mathbf{A}^+ = \mathbf{A} (\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^*$, which equals its own conjugate transpose, since $\mathbf{A}^* \mathbf{A}$ is Hermitian and so is its inverse.

(P4): $\mathbf{X} (\mathbf{A} \mathbf{B}) = \mathbf{B}^+ (\mathbf{A}^+ \mathbf{A}) \mathbf{B} = \mathbf{B}^+ \mathbf{B} = \mathbf{B}^* (\mathbf{B} \mathbf{B}^*)^{-1} \mathbf{B}$, Hermitian for the same reason.

All four hold, so $\mathbf{X} = (\mathbf{A} \mathbf{B})^+$ by Theorem 13.12.2.

Exercise. C3: When the pseudoinverse is the adjoint

Let $\mathbf{A} \in M_{m \times n}(F)$.

- Prove that $\mathbf{A}^+ = \mathbf{A}^*$ if and only if $\mathbf{A} \mathbf{A}^* \mathbf{A} = \mathbf{A}$.
- Deduce that $\mathbf{A}^+ = \mathbf{A}^*$ if and only if every nonzero singular value of $\mathbf{A}$ equals 1.
- Give an example of such an $\mathbf{A}$ in $M_2(\mathbb{R})$ which is neither unitary nor an orthogonal projection.

Solution.

- ($\Rightarrow$) If $\mathbf{A}^+ = \mathbf{A}^*$, then (P1) reads $\mathbf{A} \mathbf{A}^* \mathbf{A} = \mathbf{A}$.
 ($\Leftarrow$) Suppose $\mathbf{A} \mathbf{A}^* \mathbf{A} = \mathbf{A}$ and put $\mathbf{X} = \mathbf{A}^*$. Then (P1) is the hypothesis; (P2) is $\mathbf{A}^* \mathbf{A} \mathbf{A}^* = (\mathbf{A} \mathbf{A}^* \mathbf{A})^* = \mathbf{A}^*$; (P3) is $(\mathbf{A} \mathbf{A}^*)^* = \mathbf{A} \mathbf{A}^*$, true for every $\mathbf{A}$; and (P4) is $(\mathbf{A}^* \mathbf{A})^* = \mathbf{A}^* \mathbf{A}$, likewise. By Theorem 13.12.2, $\mathbf{A}^+ = \mathbf{A}^*$.
- With $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^*$ we have $\mathbf{A}^* = \mathbf{V} \mathbf{\Sigma}^* \mathbf{U}^*$ and $\mathbf{A}^+ = \mathbf{V} \mathbf{\Sigma}^+ \mathbf{U}^*$. Since $\mathbf{U}$ and $\mathbf{V}$ are invertible, $\mathbf{A}^+ = \mathbf{A}^*$ if and only if $\mathbf{\Sigma}^+ = \mathbf{\Sigma}^*$, that is if and only if $\sigma_i^{-1} = \sigma_i$ for every $i \leq r$, that is if and only if every nonzero singular value is 1 (the equation $\sigma^2 = 1$ with $\sigma > 0$ forces $\sigma = 1$).
- Take

$$\mathbf{A} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

Then $\mathbf{A} \mathbf{A}^\top = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, so $\mathbf{A} \mathbf{A}^\top \mathbf{A} = \mathbf{A}$ and (a) applies. Its singular values are 1 and 0, matching (b). It is not unitary, since it is singular, and not an orthogonal projection, since $\mathbf{A}^\top \neq \mathbf{A}$. Such a matrix maps $\text{col}(\mathbf{A}^*)$ isometrically onto $\text{col}(\mathbf{A})$ and sends $\text{nul}(\mathbf{A})$ to $\mathbf{0}$: a unitary matrix on the part of the space where it is invertible, and nothing at all elsewhere.