Example 164. Show that the eigenvalues of $A^T A$ are all nonnegative.

Proof. Suppose that λ is an eigenvalue of $A^T A$. Then $A^T A\boldsymbol{v} = \lambda\boldsymbol{v}$ (where $\boldsymbol{v}$ is a λ-eigenvector).
It follows that $\underbrace{\boldsymbol{v}^T A^T A\boldsymbol{v}}_{= \|A\boldsymbol{v}\|^2 \geqslant 0} = \lambda\boldsymbol{v}^T\boldsymbol{v} = \lambda\|\boldsymbol{v}\|^2$. Finally, $\lambda\|\boldsymbol{v}\|^2 \geqslant 0$ implies that $\lambda \geqslant 0$. □

The **pseudoinverse** of an $m \times n$ matrix A is the matrix A^+ such that the system $A\boldsymbol{x} = \boldsymbol{b}$ has "optimal" solution $\boldsymbol{x} = A^+\boldsymbol{b}$.

Here, "optimal" means that $\boldsymbol{x}$ is the smallest least squares solution.
In particular:

- If $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution, then $\boldsymbol{x} = A^+\boldsymbol{b}$ is that solution.
- If $A\boldsymbol{x} = \boldsymbol{b}$ has many solutions, then $\boldsymbol{x} = A^+\boldsymbol{b}$ is the one of smallest norm (the "optimal" one; and there is indeed only one such optimal solution).
- If $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent but has a unique least squares solution, then $\boldsymbol{x} = A^+\boldsymbol{b}$ is that least squares solution.
- If $A\boldsymbol{x} = \boldsymbol{b}$ has many least squares solutions, then $\boldsymbol{x} = A^+\boldsymbol{b}$ is the one with smallest norm.

When there is a unique (least squares) solution, we know how to find the pseudoinverse:

- If A is invertible, then $A^+ = A^{-1}$.
- If A has full column rank, then $A^+ = (A^T A)^{-1} A^T$.
  Recall. If $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent, a least squares solution can be determined by solving $A^T A\boldsymbol{x} = A^T\boldsymbol{b}$. If A has full column rank (i.e. the columns of A are independent; in this context, the typical case), then $\boldsymbol{x} = (A^T A)^{-1} A^T\boldsymbol{b}$ is the **unique** least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$.

Example 165.

(a) What is the pseudoinverse of $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \end{bmatrix}$?

(b) What is the pseudoinverse of $\Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}$?

(c) What is the pseudoinverse of $\Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$?

(d) In each case, compute $\Sigma^+\Sigma$ and $\Sigma\Sigma^+$.

Armin Straub
straub@southalabama.edu

Solution.

(a) Recall that, if A has full column rank, then $A^+ = (A^T A)^{-1} A^T$.

Here, $\Sigma^T\Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}$, so that $\Sigma^+ = (\Sigma^T\Sigma)^{-1}\Sigma^T = \begin{bmatrix} 1/4 & \\ & 1/9 \end{bmatrix}\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \end{bmatrix}$.

Alternative. Let us think about the optimal solution to $\Sigma\boldsymbol{x} = \boldsymbol{b}$, that is, $\begin{bmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$.

The (unique) least squares solution is $\boldsymbol{x} = \begin{bmatrix} b_1/2 \\ b_2/3 \end{bmatrix}$. (Review if this is not obvious!)

Since $\begin{bmatrix} b_1/2 \\ b_2/3 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \end{bmatrix}\boldsymbol{b}$, we conclude that $\Sigma^+ = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \end{bmatrix}$.

(b) Let us think about the smallest norm ("optimal") solution to $\Sigma\boldsymbol{x} = \boldsymbol{b}$, that is, $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$.

The general solution is $\boldsymbol{x} = \begin{bmatrix} b_1/2 \\ b_2/3 \\ t \end{bmatrix}$, where t is a free parameter.

Clearly, the smallest norm solution is $\begin{bmatrix} b_1/2 \\ b_2/3 \\ 0 \end{bmatrix}$.

Since $\begin{bmatrix} b_1/2 \\ b_2/3 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \\ 0 & 0 \end{bmatrix}\boldsymbol{b}$, we conclude that $\Sigma^+ = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \\ 0 & 0 \end{bmatrix}$.

(c) Now, $\Sigma\boldsymbol{x} = \boldsymbol{b}$, that is, $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ has no solution (unless $b_2 = 0$).

We therefore need to think about least squares solutions.

The general least squares solution (why?!) is $\boldsymbol{x} = \begin{bmatrix} b_1/2 \\ s \\ t \end{bmatrix}$, where s, t are free parameters.

Clearly, the smallest norm least squares solution is $\begin{bmatrix} b_1/2 \\ 0 \\ 0 \end{bmatrix}$.

Since $\begin{bmatrix} b_1/2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}\boldsymbol{b}$, we conclude that $\Sigma^+ = \begin{bmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.

(d) Firstly, $\Sigma^+\Sigma = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\Sigma\Sigma^+ = \begin{bmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

Secondly, $\Sigma^+\Sigma = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and $\Sigma\Sigma^+ = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

[Note how the pseudoinverse tries to behave like the regular inverse. But since Σ has only 2 columns, $\Sigma^+\Sigma$ and $\Sigma\Sigma^+$ can have rank at most 2 (so cannot be the full 3×3 identity).]

Thirdly, $\Sigma^+\Sigma = \begin{bmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and $\Sigma\Sigma^+ = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

[Here, Σ has rank 1, so that $\Sigma^+\Sigma$ and $\Sigma\Sigma^+$ can have rank at most 1.]

In general. Proceeding, as in this example, we find that the pseudoinverse of any $m\times n$ diagonal matrix Σ is the $n\times m$ (transposed dimensions!) diagonal matrix whose nonzero entries are the inverses of the entries of Σ.

Comment. Observe that, in all three cases, $\Sigma^{++} = \Sigma$.

Comment. Note that $\begin{bmatrix} 1 & 0 \\ 0 & \varepsilon \end{bmatrix}^+ = \begin{bmatrix} 1 & 0 \\ 0 & \varepsilon^{-1} \end{bmatrix}$ for small $\varepsilon \neq 0$, while $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}^+ = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. This shows that the pseudoinverse is not a continuous operation.

Armin Straub
straub@southalabama.edu

It turns out that the pseudoinverse A^+ can be easily obtained from the SVD of A:

Theorem 166. The **pseudoinverse** of an $m \times n$ matrix A with SVD $A = U\Sigma V^T$ is

$$A^+ = V\Sigma^+ U^T,$$

where Σ^+, the pseudoinverse of Σ, is the $n \times m$ diagonal matrix, whose nonzero entries are the inverses of the entries of Σ.

Proof. The equation $A\boldsymbol{x} = \boldsymbol{b}$ is equivalent to $U\Sigma V^T\boldsymbol{x} = \boldsymbol{b}$ and, thus, $\Sigma V^T\boldsymbol{x} = U^T\boldsymbol{b}$.
Write $\boldsymbol{y} = V^T\boldsymbol{x}$ and note that $\boldsymbol{y}$ and $\boldsymbol{x}$ have the same norm (why?!).
We already know that the equation $\Sigma\boldsymbol{y} = U^T\boldsymbol{b}$ has optimal solution $\boldsymbol{y} = \Sigma^+ U^T\boldsymbol{b}$.
Since $\boldsymbol{y}$ and $\boldsymbol{x}$ have the same norm, it follows that $\boldsymbol{x} = V\boldsymbol{y} = V\Sigma^+ U^T\boldsymbol{b}$ is the optimal solution to $A\boldsymbol{x} = \boldsymbol{b}$.
Hence, $A^+ = V\Sigma^+ U^T$. □

Lemma 167. The pseudoinverse of A^+ is $A^{++} = A$.

Proof. Starting with the SVD $A = U\Sigma V^T$, we have $A^+ = V\Sigma^+ U^T$, which is the SVD of A^+.
Therefore, $A^{++} = U\Sigma^{++} V^T$. The claim thus follows from $\Sigma^{++} = \Sigma$. □

Example 168. Determine the pseudoinverse of $A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$ in two ways.

First, using the SVD and, second, using the fact that A has full column rank.

Solution. (SVD) We have computed the SVD of this matrix before.

Since $A = U\Sigma V^T$ with $U = \begin{bmatrix} -2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \\ -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix}$, $\Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$, $V = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$,

the pseudoinverse is $A^+ = V\Sigma^+ U^T$ where $\Sigma^+ = \begin{bmatrix} 1/\sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

Multiplying these matrices, $A^+ = \frac{1}{3}\begin{bmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \end{bmatrix}$.

Comment. For many applications, it may be neither necessary nor helpful to multiply V, Σ^+, U^T.

Solution. (full column rank) Since A clearly has full column rank, we also have $A^+ = (A^T A)^{-1} A^T$.
Indeed, $A^+ = (A^T A)^{-1} A^T = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \end{bmatrix}$.

Example 169. What is the pseudoinverse of $A = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$?

Solution. Recall (or compute) that $A = U\Sigma V^T$ with $U = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$, $\Sigma = \begin{bmatrix} \sqrt{10} & \\ & 0 \end{bmatrix}$, $V = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$.
Hence, $A^+ = V\Sigma^+ U^T$ where $\Sigma^+ = \begin{bmatrix} 1/\sqrt{10} & 0 \\ 0 & 0 \end{bmatrix}$.

Multiplying these matrices (which may not be necessary or helpful for applications), $A^+ = \frac{1}{10}\begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}$.

Note. Since A does not have full column rank, $A^+ = (A^T A)^{-1} A^T$ cannot be used. That's because $A^T A$ is not invertible.

Comment. Here, $A^+ A = \boldsymbol{v}_1 \boldsymbol{v}_1^T = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $AA^+ = \boldsymbol{u}_1 \boldsymbol{u}_1^T = \frac{1}{5}\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$ are not visually like the identity. However, note that these are the (orthogonal) projections onto $\boldsymbol{v}_1$ and $\boldsymbol{u}_1$ respectively (in particular, the eigenvalues are $1, 0$).

Armin Straub
straub@southalabama.edu