# A NOTE ON THE PSEUDOINVERSE

DAVID CLARK

The singular value decomposition (SVD) for real matrices has a nice application: the construction of the *pseudoinverse* of a matrix, which (i) gives us a general tool for solving a linear system $Ax = b$, and (ii) admits an elegant geometric interpretation. Here we'll briefly prove some elementary properties of the pseudoinverse. First, let's restate the theorem in which the construction is rooted.

**Theorem 1** (Singular Value Decomposition)**.** *Let $A$ be any real $m$ by $n$ matrix with rank $r$. Then we can write $A = U\Sigma V^T$, where $U$ and $V$ are orthogonal, and $\Sigma$ is $m$ by $n$ diagonal, with $r$ nonzero entries given by the positive square roots of eigenvalues of $A^TA$ and $AA^T$.*

In fact, the matrices $U$ and $V$ have columns the eigenvectors of $AA^T$ and $A^TA$, respectively. Note that it's standard to write the diagonal entries of $\Sigma$, often called the *singular values of* $A$, in descending order.

Now we define the pseudoinverse.

**Definition 0.1.** *The pseudoinverse of a matrix $A = U\Sigma V^T$, denoted $A^+$, is given by*

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

*where $\Sigma^+$ is obtained by transposing $\Sigma$ and inverting all nonzero entries.*

Its easy to verify that the pseudoinverse of an invertible matrix is its honest inverse: $A^+A = AA^+ = I$. For noninvertible or nonsquare $A$, we may have either or neither of these equalities, depending on $A$'s rank. So given a linear system corresponding to one of the latter cases, what exactly does the pseudoinverse tell us?

One answer comes from the theory of least squares solutions, which is based in the following statement.

**Claim 1.** *Let $Ax = b$ be any linear system. Then the system $A^TAx = A^Tb$ is consistent, i.e., $A^Tb \in C(A^TA)$, the column space of $A^TA$.*

*Proof.* Note first that, for any $A$, $N(A) = N(A^TA)$, i.e., the nullspaces of $A$ and $A^TA$ coincide. One inclusion is clear. On the other hand, if $A^TAx = 0$ then $Ax$ is both in the nullspace of $A^T$ and the column space of $A$. By orthogonality then, we know that $Ax = 0$.

Now, to show $A^Tb \in C(A^TA)$ it's sufficient to show that $A^Tb \perp N((A^TA)^T) = N(A^TA)$. If $y \in N(A^TA)$, we are guaranteed that $Ay = 0$ by the previous observation. Then $y \cdot A^Tb = y^TA^Tb = (Ay)^Tb = 0$. □

Even if our original system $Ax = b$ were not consistent, we know this new one is. In fact, any solution to this new system gives a *least squares solution* to $Ax = b$,

---

*Date*: December 28, 2007.

i.e., a vector $\hat{x}$ as close to $C(A)$, in a least squares sense, as possible. It's easy to see why this is true. Any solution to $A^TAx = A^Tb$ satisfies $A^T(A\hat{x} - b) = 0$; in other words, the error vector $A\hat{x} - b$ is perpendicular to the column space of $A$, guaranteeing its minimal distance. (Of course, if $Ax = b$ is consistent, this error vector is zero, and $\hat{x}$ is an honest solution.)

One might hope that the pseudoinverse helps in solving such inconsistent systems. This is indeed the case; in particular, given any system $Ax = b$, the vector $x^+ := A^+b$ is a least squares solution.

**Claim 2.** *Let $A = U\Sigma V^T$, and $x^+ = A^+b = V\Sigma^+U^Tb$. Then $A^TAx^+ = A^Tb$.*

*Proof.* It's sufficient to show that $A^T(Ax^+ - b) = 0$. We have

$$\begin{aligned} Ax^+ - b &= (U\Sigma V^T)V\Sigma^+U^Tb - b \\ &= (U\Sigma\Sigma^+U^T - I)b \\ &= (U(\Sigma\Sigma^+ - I)U^Tb. \end{aligned}$$

Thus,

$$\begin{aligned} A^T(Ax^+ - b) &= V\Sigma^TU^TU(\Sigma\Sigma^+ - I)U^Tb \\ &= V\Sigma^T(\Sigma\Sigma^+ - I)U^Tb. \end{aligned}$$

Observe now that, in block form, we have

$$\Sigma^T = \begin{pmatrix} \Sigma_r^T & 0_{r,m-r} \\ 0_{n-r,r} & 0_{n-r,m-r} \end{pmatrix},$$

where $\Sigma_r$ is the $r$ by $r$ submatrix of nonzero diagonal entries in $\Sigma$, and

$$\Sigma\Sigma^+ - I = \begin{pmatrix} 0_r & 0_{r,m-r} \\ 0_{m-r,r} & -I_{m-r} \end{pmatrix}.$$

Multiplying these matrices gives a big zero, of course, and our result. □

Thus, the pseudoinverse always gives us a least squares solution to $Ax = b$. For systems in which $A$ has linearly independent columns, this is the only least squares solution, also given by $\hat{x} = (A^TA)^{-1}A^Tb$. However, consider the worst possible case: an inconsistent $Ax = b$ in which $A$ has dependent columns. Here, there is not a unique least squares solution, but in fact infinitely many: just consider the set of solutions to $A^TAx = A^Tb$, where we now have $A^TA$ nonsingular. The pseudoinverse gives us exactly one such solution $x^+$; is there something special about this vector? No diggity.

**Claim 3.** *The vector $x^+$ is the* shortest *least squares solution to $Ax = b$, i.e.,*

$$\|x^+\| = \min\{\|x\| : \|Ax - b\| \text{ is minimal}\}$$

*Proof.* In the following equalities, we will use the fact that $U$ and $V$ are orthogonal, and thus length-preserving, several times.

$$\begin{aligned}
\min\{\|x\| : \|Ax - b\| \text{ is minimal}\} &= \min\{\|x\| : \|U\Sigma V^T x - b\| \text{ is minimal}\} \\
&= \min\{\|x\| : \|\Sigma V^T x - U^T b\| \text{ is minimal}\} \\
&= \min\{\|V^T x\| : \|\Sigma V^T x - U^T b\| \text{ is minimal}\} \\
&= \min\{\|y\| : \|\Sigma y - U^T b\| \text{ is minimal}\}
\end{aligned}$$

Here we've made the substitution $y := V^T x$ to get the last line.

Let's now consider the system $\Sigma y = U^T b$, which we'll write using blocks as

$$\begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} y = U^T b.$$

We want to find a least squares solution of minimal length. Since $\Sigma$ is diagonal, however, there is an obvious candidate:

$$y^+ := \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^T b = \Sigma^+ U^T b.$$

A few moments reflection should confirm this. Thus, $x^+ := V y^+ = V\Sigma^+ U^T b$ will attain the minimum length described above.

□

This is pretty remarkable: the pseudoinverse gives us, in a natural geometric sense, the *best possible* solution to any linear system $Ax = b$. It's worth mentioning another characterization of this special solution: the vector $x^+$ is always in the row space of $A$.

**Claim 4.** *$A^+ b \in C(A^T)$ for any $b$.*

*Proof.* Using orthogonal compliments, we can write $A^+ b$ as a unique linear combination

$$x^+ = A^+ b = x_r + x_n,$$

where $x_r \in C(A^T)$ and $x_n \in N(A)$. We know that this vector is the minimal length least squares solution to $Ax = b$, i.e., $\|x_r + x_n\|$ is minimal among $x$ such that $A^T A x = A^T b$. Observe, however, that this vector is still a least squares solution even with the nullspace component discarded, since $A^T A x_r = A^T A(x_r + x_n) = A^T b$. Further, we note that

$$\begin{aligned}
\|x_r\|^2 &= x_r \cdot x_r \\
&= x_r \cdot x_r + 2x_r \cdot x_n \\
&\leq x_r \cdot x_r + 2x_r \cdot x_n + x_n \cdot x_n \\
&= \|x_r + x_n\|^2,
\end{aligned}$$

using, of course, the fact that $x_r$ and $x_n$ are orthogonal. Thus $x^+$ achieves minimal length when $x_n = 0$, giving the result.

□

So we've shown that $A^+$ maps to $C(A^T)$. But since these matrices clearly have the same rank, we get that

$$C(A^+) = C(A^T).$$

Do we have a similar correspondence for the nullspaces? Fo' sho'.

**Claim 5.** *$N(A^+) = N(A^T)$.*

*Proof.* To show containment, let $b \in N(A^T)$, so $V\Sigma^T U^T b = 0$, and clearly also $\Sigma^T U^T b = 0$. But since $\Sigma^T$ is diagonal, this implies that $\Sigma^+ U^T b = 0$. Thus $A^+ b = V\Sigma^T U^T b = 0$. Our statement then follows from the Rank-Nullity Theorem, as $A^+$ and $A^T$ have the same rank and thus the same nullspace dimension. □

In particular, this tells us that $\ker(A^+|_{C(A)}) = 0$, since $C(A)$ is the orthogonal complement of $N(A^T)$. And because $A^+$ maps $C(A)$ to $C(A^T)$, and $\dim C(A) = \dim C(A^T)$, we see that this map is actually an isomorphism (that is, it is one-to-one and onto, or equivalently, invertible).

So we've shown a lot about our matrix $A^+$: how it helps us solve the system $Ax = b$, and many ways in which it relates to the original matrix $A$. But what really justifies calling this thing *the pseudoinverse*? You've probably already guessed.

**Claim 6.** *Restricted to $C(A^T)$ and $C(A)$, respectively, $A$ and $A^+$ are mutual inverses.*

*Proof.* Let $x \in C(A^T)$, so $x = A^T y$ for some $y$. Then

$$\begin{aligned} A^+Ax - x &= A^+AA^Ty - A^Ty \\ &= V\Sigma^+\Sigma\Sigma^T U^T y - V\Sigma^T U^T y \\ &= V(\Sigma^+\Sigma - I)\Sigma^T U^T y \\ &= 0, \end{aligned}$$

where the last line follows from a similar computation to the one used in the proof of Claim 2. This establishes that $A^+Ax = x$. The argument is essentially the same to show $AA^+b = b$ for all $b \in C(A)$.

□

Thus $A^+$ inverts $A$ exactly where it can: on its column space.