Least Squares Solutions

Suppose that a linear system $Ax = b$ is inconsistent. This is often the case when the number of equations exceeds the number of unknowns (an overdetermined linear system). If a tall matrix A and a vector b are randomly chosen, then $Ax = b$ has no solution with probability 1.

In geometric terms, inconsistency means that b is not in the image of A. If so, it may still be reasonable to look for x such that $y = Ax$ is as close to b as possible, i.e.,

$$\|Ax - b\| \quad \text{is a minimum.}$$

In other words, we are interested in a vector x^* such that

$$Ax^* = \operatorname{proj}_{\operatorname{im} A} b.$$

Any such vector x^* is called a *least squares solution* to $Ax = b$, as it minimizes the sum of squares

$$\|Ax - b\|^2 = \sum_k ((Ax)_k - b_k)^2.$$

For a consistent linear system, there is no difference between a least squares solution and a regular solution.

Consider the following derivation:

$$\begin{aligned} &Ax^* = \operatorname{proj}_{\operatorname{im} A} b \\ &b - Ax^* \perp \operatorname{im} A \qquad (b - Ax^* \text{ is normal to } \operatorname{im} A) \\ &b - Ax^* \text{ is in } \ker A^\top \\ &A^\top (b - Ax^*) = 0 \\ &A^\top A x^* = A^\top b \qquad (\textit{normal equation}). \end{aligned}$$

Note that $A^\top A$ is a symmetric square matrix. If $A^\top A$ is invertible, and this is the case whenever A has trivial kernel, then the least squares solution is unique:

$$x^* = (A^\top A)^{-1} A^\top b.$$

Moreover,

$$Ax^* = A(A^\top A)^{-1} A^\top b,$$

so $A(A^\top A)^{-1} A^\top$ is the standard matrix of the orthogonal projection onto the image of A. If $A^\top A$ is not invertible, there are infinitely many least squares solutions. They all yield the same Ax^*.

Here are some supporting propositions and examples.

Proposition. $Ax \cdot y = x \cdot A^\top y$
Proof. Exercise.

Proposition. $(\operatorname{im} A)^\perp = \ker A^\top$
Proof. Exercise.

Proposition. $\ker A = \ker A^\top A$
Proof. If $Ax = 0$, then $A^\top Ax = 0$. If $A^\top Ax = 0$, $\|Ax\|^2 = (Ax)^\top Ax = x^\top A^\top Ax = 0$. □

Proposition. $\operatorname{im} A^\top = \operatorname{im} A^\top A$
Proof. $\operatorname{im} A^\top = (\ker A)^\perp = (\ker A^\top A)^\perp = ((\operatorname{im} A^\top A)^\perp)^\perp = \operatorname{im} A^\top A$ □

Example. The linear system $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is inconsistent.

The vector $b = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is not on the line $\operatorname{im} A = \operatorname{span}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

The associated normal equation is $\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$.

The matrix $A^\top A = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$ is not invertible.

The least squares solutions are $x^* = \begin{pmatrix} 1 \\ .5 \end{pmatrix} + \operatorname{span}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

The orthogonal projection of b onto $\operatorname{im} A$ is $Ax^* = \begin{pmatrix} 1.5 \\ 1.5 \end{pmatrix}$.

Example. The linear system $\begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ is inconsistent.

The vector $b = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ is not in the plane $\operatorname{im} A = \operatorname{span}\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right)$.

The associated normal equation is $\begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

The matrix $A^\top A = \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix}$ is invertible, $(A^\top A)^{-1} = \frac{1}{2}\begin{pmatrix} 2 & -2 \\ -2 & 3 \end{pmatrix}$.

The least squares solution is unique, $x^* = \begin{pmatrix} 1 \\ -.5 \end{pmatrix}$.

The orthogonal projection of b onto $\operatorname{im} A$ is $Ax^* = \begin{pmatrix} .5 \\ .5 \\ 1 \end{pmatrix}$.

The matrix of the orthogonal projection onto $\operatorname{im} A$ is $A(A^\top A)^{-1}A^\top = \begin{pmatrix} .5 & .5 & 0 \\ .5 & .5 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.