In general, for an overdetermined $m \times n$ system $Ax = b$, what Gauss and Legendre discovered is that there are solutions x minimizing

$$\|Ax - b\|_2^2$$

(where $\|u\|_2^2 = u_1^2 + \cdots + u_n^2$, the square of the Euclidean norm of the vector $u = (u_1, \ldots, u_n)$), and that these solutions are given by the square $n \times n$ system

$$A^\top A x = A^\top b,$$

called the *normal equations*. Furthermore, when the columns of A are linearly independent, it turns out that $A^\top A$ is invertible, and so x is unique and given by

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

Note that $A^\top A$ is a symmetric matrix, one of the nice features of the normal equations of a least squares problem. For instance, since the above problem in matrix form is represented as

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{pmatrix} \begin{pmatrix} d \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

the normal equations are

$$\begin{pmatrix} 3 & x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 & x_1^2 + x_2^2 + x_3^2 \end{pmatrix} \begin{pmatrix} d \\ c \end{pmatrix} = \begin{pmatrix} y_1 + y_2 + y_3 \\ x_1 y_1 + x_2 y_2 + x_3 y_3 \end{pmatrix}.$$

In fact, given any real $m \times n$ matrix A, there is always a unique x^+ of minimum norm that minimizes $\|Ax - b\|_2^2$, even when the columns of A are linearly dependent. How do we prove this, and how do we find x^+?

Theorem 23.1. *Every linear system $Ax = b$, where A is an $m \times n$ matrix, has a unique least squares solution x^+ of smallest norm.*

Proof. Geometry offers a nice proof of the existence and uniqueness of x^+. Indeed, we can interpret b as a point in the Euclidean (affine) space $\mathbb{R}^m$, and the image subspace of A (also called the column space of A) as a subspace U of $\mathbb{R}^m$ (passing through the origin). Then it is clear that

$$\inf_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 = \inf_{y \in U} \|y - b\|_2^2,$$

with $U = \operatorname{Im} A$, and we claim that x minimizes $\|Ax - b\|_2^2$ iff $Ax = p$, where p the orthogonal projection of b onto the subspace U.

Recall from Section 13.1 that the orthogonal projection $p_U \colon U \oplus U^\perp \to U$ is the linear map given by

$$p_U(u + v) = u,$$