Normal Equations

If b is not in the column space of A, then $Ax = b$ has no solution; the system is *inconsistent*. This is typical if A is $m \times n$ with $m > n$, which we will assume here. Let us also assume that A has full rank. Since $Ax = b$ has no solution, one may reasonably be interested in finding a vector x which minimizes the difference between b and Ax:

$$\min_x \|Ax - b\|. \tag{1}$$

Equivalently: find a vector y in the column space of A which is closest to b (then x is the unique solution of the *consistent* system $Ax = y$). There are many norms that we might use in (1), but if we use the norm induced by the dot product, then (1) is called the *discrete linear least squares problem*:

$$\min_x \|Ax - b\|_2. \tag{2}$$

Now suppose that we want to find an element of $S \leq \mathbb{R}^n$ that is closest to some vector b (which is typically not in S). Our intuition says that we "drop a perpendicular" from b to S, and that is exactly right: Let $y \in S$ be such that $b - y \perp S$, and consider any vector $w = y + z \in S$.

$$\begin{aligned} \|b - w\|_2^2 &= (b - w)^T(b - w) \\ &= (b - y)^T(b - y) + z^T z \end{aligned}$$

which is (uniquely) minimized by $z = 0$ ("Calculus? We don't need no stinking calculus!"). Therefore a vector $y \in S$ which minimizes $\|b - y\|_2$ must satisfy $b - y \perp S$. In the language of orthogonal projections: *if $b = b_S + b_{S^\perp}$ is the direct sum representation of b in $\mathbb{R}^n = S \oplus S^\perp$, then $\|b - y\|_2$ is minimized at $y = b_S$.*

Now we can apply this to (2) by letting $S = \text{ColSp(A)}$. That is, we want $y = Ax$, and therefore $b - Ax \perp \text{ColSp(A)}$. Clearly this requires $(b - Ax)^T A = 0$ (right?). Transposing this equation gives

$$A^T A x = A^T b, \tag{3}$$

and this system of equations is called the *normal equations* for (2).

Since the columns of A are linearly independent, $Az = 0 \Leftrightarrow z = 0$, and thus $A^T A$ is nonsingular. Therefore the normal equations, and hence the least squares problem, has a unique solution. In the language of projections: the (unique) orthogonal projector onto the $\text{ColSp}(A)$ is $P = A(A^T A)^{-1} A^T$, giving $y = Pb = A(A^T A)^{-1} A^T b$, and from $y = Ax$, we take $x = (A^T A)^{-1} A^T b$. This is just the normal equations *solved*.

Notice that if Q is any matrix whose columns form a basis for ColSp(A), then we want $(b - Ax)^T Q = 0$ (right?), so a more general normal equation is $Q^T A x = Q^T b$. Thus, while providing *a* route to *the* LS solution, the normal equations (3) are *not the only route to its computation...*