Linear Least Squares Computations

Assuming that $A \in \mathbb{R}^{m \times n}$ has linearly independent columns, the problem

$$
\arg \min_{x} \|Ax - b\|_2
$$

(1)

has a unique solution, say $x_{LS}$, which is also the unique solution to the normal equations

$$
A^T Ax = A^T b.
$$

(2)

This suggests the normal equations approach to computing $x_{LS}$:

1. Form $C = A^T A$, and $w = A^T b$.
2. Compute the Cholesky factorization $C = LL^T$.
3. Solve $Ly = w$ by forsub and then $L^T x = y$ by backsub.

This algorithm requires $mn^2 + \frac{1}{3}n^3 + O(mn)$ flops (taking advantage of the symmetry of $C$). It is an important method because it is fast and doesn’t use very much memory. $Cx = w$ can be viewed as a compressed form of arg min$_x \|Ax - b\|_2$.

We have other methods that, while more costly, are more robust in the face of rounding errors. The other methods arrive at $x_{LS}$ by a different route. Recall that the normal equations were a result of requiring that $b - Ax$ be orthogonal (normal) to the subspace $S = \text{ColSp}(A)$. That is another way of saying that $Ax$ is the orthogonal projection of $b$ onto $S$. The solution to the normal equations is therefore the solution to the rectangular, but consistent, linear system

$$
Ax = Pb,
$$

(3)

where $P$ is the orthogonal projector onto $S = \text{ColSp}(A)$.

Now if $Z$ is any matrix whose columns form a basis for $S$, then the orthogonal projector onto $S$ is $P = Z(Z^T Z)^{-1}Z^T$, and (3) becomes

$$
Ax = Z(Z^T Z)^{-1}Z^T b.
$$

(4)

This family of equations, parametrized by $Z$, explains most methods. For example, the normal equations gives $x = (A^T A)^{-1}A^T b$, and premultiplying by $A$ gives $Ax = A(A^T A)^{-1}A^T b$, which is (4), with $Z = A$.

The most often used LS methods compute a matrix $Z = Q$ whose columns form an orthonormal basis for $S$. Since $Q$ and $A$ have the same column space, each column of $A$ is a linear combination of the columns of $Q$. That is $A = QR$, where $R \in \mathbb{R}^{n \times n}$, and (3) becomes $QRx = QQ^T b$. Premultiplying by $Q^T$ gives the (smaller and nonsingular) $n \times n$ system

$$
Rx = Q^T b.
$$