Least squares problems and QR factorization

We see a different algorithm to solve least-squares problems using the QR factorization of A.

Start from

$$A = QR, \quad Q = \begin{bmatrix} Q_0 & Q_c \end{bmatrix}, \quad R = \begin{bmatrix} R_0 \\ 0 \end{bmatrix}.$$

Since orthogonal matrices preserve the 2-norm,

$$\begin{aligned}
\|A\mathbf{x} - \mathbf{y}\| &= \|Q^T(A\mathbf{x} - \mathbf{y})\| = \|Q^T QR\mathbf{x} - Q^T\mathbf{y}\| \\
&= \|R\mathbf{x} - Q^T\mathbf{y}\| = \left\| \begin{bmatrix} R_0 \\ 0 \end{bmatrix} \mathbf{x} - \begin{bmatrix} Q_0^T \\ Q_c^T \end{bmatrix} \mathbf{y} \right\| \\
&= \left\| \begin{bmatrix} R_0\mathbf{x} - Q_0^T\mathbf{y} \\ Q_c^T\mathbf{y} \end{bmatrix} \right\|.
\end{aligned}$$

Solving least squares with QR

$$\|A\mathbf{x} - \mathbf{y}\| = \left\| \begin{bmatrix} R_0\mathbf{x} - Q_0^T\mathbf{y} \\ Q_c^T\mathbf{y} \end{bmatrix} \right\|$$

How can we minimize the norm of this vector?

Bottom block same value, regardless of $\mathbf{x}$. The squares of those entries will always be in the sum.

Top block We can choose $\mathbf{x}$ to make its entries smaller. Can we get $R_0\mathbf{x} - Q_0^T\mathbf{y} = 0$? Yes, if R_0 invertible.

When is R_0 invertible?

Related to a result we have seen earlier. If $A = QR$, with Q orthogonal, then

$$A^T A = (QR)^T (QR) = R^T \underbrace{Q^T Q}_{=I} R = R^T R = \begin{bmatrix} R_0^T & 0 \end{bmatrix} \begin{bmatrix} R_0 \\ 0 \end{bmatrix} = R_0^T R_0.$$

A has full column rank $\iff$ $A^T A$ is posdef $\iff$ $A^T A = R_0^T R_0$ is invertible $\iff$ R_0 is inverbile.

(Note for your future self revising: $R_0^T R_0$ is the Cholesky factorization of $A^T A$, which we shall see later in the course.)

Algorithm

We have proved the following.

Lemma

If $A = QR = \begin{bmatrix} Q_0 & Q_c \end{bmatrix} \begin{bmatrix} R_0 \\ 0 \end{bmatrix}$ (and has full column rank), then the solution of $\min \|A\mathbf{x} - \mathbf{y}\|$ is given by $\mathbf{x} = R_0^{-1}(Q_0^T)\mathbf{y}$.

The thin QR factorization $A = Q_0 R_0$ contains all we need here.

Corollary formula for the pseudoinverse $A^+ = R_0^{-1} Q_0^T$.

Cost:

1. Thin QR: $O(mn^2)$
2. Multiplication $\mathbf{c} = (Q_0^T)\mathbf{b}$: $O(mn)$.
3. Triangular system solution $R_0\mathbf{x} = \mathbf{c}$: $O(n^2)$ with back-substitution.

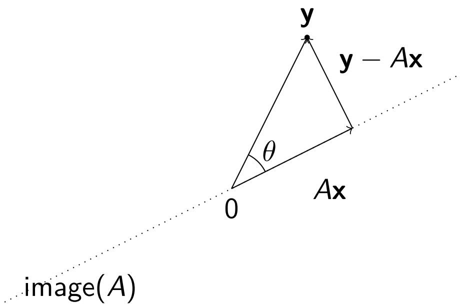
The dominant part is the thin QR computation.

The geometric picture

$\mathbf{y}$ is split into two orthogonal components: $A\mathbf{x}$ and $\mathbf{y} - A\mathbf{x}$.

- $A\mathbf{x} = Q_0 R_1 R_0^{-1} Q_0^T \mathbf{y} = Q_0 Q_0^T \mathbf{y}$ gives the projection of $\mathbf{y}$ onto $\operatorname{Im} A$. It has length $\|Q_0^T \mathbf{y}\|$.
- The residual $\|A\mathbf{x} - \mathbf{y}\|$ (i.e., the optimum value) is $\|Q_c^T \mathbf{y}\|$.

The vectors $\mathbf{y}$, $A\mathbf{x}$ and $A\mathbf{x} - \mathbf{y}$ form a right triangle with side lengths $\|\mathbf{y}\|, \|Q_0^T \mathbf{y}\|, \|Q_c^T \mathbf{y}\|$.



Exercises

1. Is it true that every symmetric, positive definite matrix $Q \in \mathbb{R}^{m \times m}$ can be written as $Q = A^T A$ for some $A \in \mathbb{R}^{m \times m}$? Hint: start from eigendecomposition $Q = U \Lambda U^T$, and consider $\begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_m} \end{bmatrix}$.

Book references: Demmel, 3.2.2; Trefethen-Bau, Lecture 11.