Least-squares via QR factorization

Stephen Boyd and Sanjay Lall

EE263
Stanford University

Least-squares via QR factorization

- $A \in \mathbb{R}^{m \times n}$ skinny, full rank
- factor as $A = QR$ with $Q^T Q = I_n$, $R \in \mathbb{R}^{n \times n}$ upper triangular, invertible
- pseudo-inverse is
$$A^\dagger = (A^T A)^{-1} A^T = (R^T Q^T Q R)^{-1} R^T Q^T = R^{-1} Q^T$$
so $x_{\mathrm{ls}} = R^{-1} Q^T y$
- projection on **range**(A) given by matrix
$$A(A^T A)^{-1} A^T = A R^{-1} Q^T = Q Q^T$$

Least-squares via full QR factorization

- full QR factorization:

$$A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

$\begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \in \mathbb{R}^{m \times m}$ orthogonal, $R_1 \in \mathbb{R}^{n \times n}$ upper triangular, invertible

- multiplication by orthogonal matrix doesn't change norm, so

$$\begin{aligned}
\|Ax - y\|^2 &= \left\| \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} x - y \right\|^2 \\
&= \left\| \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}^\mathsf{T} \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} x - \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}^\mathsf{T} y \right\|^2 \\
&= \left\| \begin{bmatrix} R_1 x - Q_1^\mathsf{T} y \\ -Q_2^\mathsf{T} y \end{bmatrix} \right\|^2 \\
&= \|R_1 x - Q_1^\mathsf{T} y\|^2 + \|Q_2^\mathsf{T} y\|^2
\end{aligned}$$

Least-squares via full QR factorization

so for any y,

$$\|Ax - y\|^2 = \|R_1 x - Q_1^T y\|^2 + \|Q_2^T y\|^2$$

- this is evidently minimized by choice $x_{\mathrm{ls}} = R_1^{-1} Q_1^T y$ (which makes first term zero)
- residual with optimal x is
$$Ax_{\mathrm{ls}} - y = -Q_2 Q_2^T y$$
- $Q_1 Q_1^T$ gives projection onto **range**(A)
- $Q_2 Q_2^T$ gives projection onto **range**$(A)^\perp$

Growing sets of regressors

consider *family* of least-squares problems

$$\text{minimize} \quad \left\| \sum_{i=1}^{p} x_i a_i - y \right\|$$

for $p = 1, \ldots, n$

($a_1, \ldots, a_p$ are called *regressors*)

- approximate y by linear combination of $a_1, \ldots, a_p$
- project y onto $\textbf{span}\{a_1, \ldots, a_p\}$
- regress y on $a_1, \ldots, a_p$
- as p increases, get better fit, so optimal residual decreases

Growing sets of regressors

solution for each $p \leq n$ is given by

$$x_{\mathrm{ls}}^{(p)} = (A_p^\mathsf{T} A_p)^{-1} A_p^\mathsf{T} y = R_p^{-1} Q_p^\mathsf{T} y$$

where

- $A_p = [a_1 \cdots a_p] \in \mathbb{R}^{m \times p}$ is the first p columns of A
- $A_p = Q_p R_p$ is the QR factorization of A_p
- $R_p \in \mathbb{R}^{p \times p}$ is the leading $p \times p$ submatrix of R
- $Q_p = [q_1 \cdots q_p]$ is the first p columns of Q

Least-norm solution via QR factorization

find QR factorization of A^T, $i.e.$, $A^\mathsf{T} = QR$, with

- $Q \in \mathbb{R}^{n\times m}$, $Q^\mathsf{T} Q = I$
- $R \in \mathbb{R}^{m\times m}$ upper triangular, nonsingular

then

- $x_{\mathrm{ln}} = A^\mathsf{T}(AA^\mathsf{T})^{-1} y = QR^{-T} y$
- $||x_{\mathrm{ln}}|| = ||R^{-T} y||$