Quick Review: Least Squares Problem

- For matrix A (mxn), and vector b (mx1), find vector x (nx1) that solves

$$\min_x \|b - Ax\|^2{}_2$$

- We determined x s.t. $A^TAx = A^Tb$ is a solution to the least squares problem when A^TA is positive definite.
- $A^TAx = A^Tb$ are called the Normal Equations.

Techniques for solving Normal Equations

1. GEPP on A^TA produces L, U, P s.t. $PA^TA=LU$, leads to solving two triangular systems
2. Cholesky decomposition on $A^TA = LL^T$, leads to solving two triangular systems. Decomposition takes about ½ steps of GEPP.
3. QR factorization on A, where Q is orthogonal, R upper triangular (with positive diagonals), leads to $A^TA=(QR)^TQR = R^TQ^TQR = R^TR$, and

   $A^Tb = (QR)^Tb = R^TQ^Tb$. So, we solve $Rx=Q^Tb$.

More on Cholesky

- Note: Matlab's `chol` function produces upper triangular R, such that $M = R^T R$ (rather than L such that $M = LL^T$), for a positive definite M.
- `chol(A, 'lower')` produces L instead.
- Recall: must calculate $A^T A$ product before factoring to solve least squares.
- Exercise: Multiplying an nxm matrix by an mxn matrix requires approximately $2mn^2$ flops

Example: Cholesky

- For our problem $A^TA = \begin{bmatrix} 19 & 4 & 1 \\ 4 & 11 & 7 \\ 1 & 7 & 10 \end{bmatrix}$

- We want to find L, such that

$$\begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix} = \begin{bmatrix} 19 & 4 & 1 \\ 4 & 11 & 7 \\ 1 & 7 & 10 \end{bmatrix}$$

Example: Cholesky (continued)

- This leads to

$$L = \begin{bmatrix} 4.3589 & 0 & 0 \\ 0.9177 & 3.1871 & 0 \\ 0.2294 & 2.1303 & 2.3258 \end{bmatrix}$$

- Solving L y = $A^T b$ (forward substitution)

→ y = $[5.7354, 5.2514, 2.7936]^T$

- Then solving $L^T x = y$ (backward substitution)

→ x = $[1.0747, 0.8448, 1.2011]^T$

Back to QR: How to find QR?

Assuming A is full column rank -

- Q: mxn – orthogonal matrix ($Q^TQ = I$)
- R: nxn – upper triangular ($r_{ii}>0$)

$$
\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix}
$$

This gives us …

$a_1 = q_1 r_{11}$

$a_2 = q_1 r_{12} + q_2 r_{22}$

$a_3 = q_1 r_{13} + q_2 r_{23} + q_3 r_{33}$

.

.

.

$a_n = q_1 r_{1n} + q_2 r_{2n} + q_3 r_{3n} + \ldots + q_n r_{nn}$

where $q_k^T q_k = 1$, $q_k^T q_j = 0$, $r_{ii} > 0$.

Note: a_k and q_k are mx1 vectors, r_{ij} are scalars.

- Step 1: Determine q_1 and first row of R

- $a_1 = q_1 r_{11}$

→Set $r_{11} = \|a_1\|$ and $q_1 = a_1/r_{11}$

- For k=2:n, $a_k = \sum_{j=1}^{k} r_{jk} q_j$
- Therefore, $q_1{}^T a_k = \sum_{j=1}^{k} r_{jk} q_1{}^T q_j = r_{1k}$

- Step 2: Determine q_2 and second row of R

- $a_2 = q_1 r_{12} + q_2 r_{22} \rightarrow q_2 r_{22} = a_2 - q_1 r_{12} = \widehat{a_2}$

$\rightarrow$Set $r_{22} = \|\widehat{a_2}\|$ and $q_2 = \widehat{a_2}/r_{22}$

- For k=3:n, $a_k = \sum_{j=1}^{k} r_{jk} q_j$
- Therefore, $q_2{}^T a_k = \sum_{j=1}^{k} r_{jk} q_2{}^T q_j = r_{2k}$

- Step p: Determine q_p and p^{th} row of R (p=3,…,n)

- $a_p = \sum_{j=1}^{p} r_{jp} q_j = (\sum_{j=1}^{p-1} r_{jp} q_j) + r_{pp} q_p$

→ $r_{pp} q_p = a_p - \sum_{j=1}^{p-1} r_{jp} q_j = \widehat{a_p}$

→Set $r_{pp} = \left\|\widehat{a_p}\right\|$ and $q_p = \widehat{a_p}/r_{pp}$

- For k=p+1:n, $a_k = \sum_{j=1}^{k} r_{jk} q_j$
- Therefore, ${q_p}^T a_k = \sum_{j=1}^{k} r_{jk} {q_p}^T q_j = r_{pk}$

QR example

For $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 2 \\ 3 & 2 & 0 \\ 2 & -2 & -1 \\ -1 & 1 & 2 \end{bmatrix}$, find Q and R s.t. A = QR.

Solve

$$[a_1 \quad a_2 \quad a_3] = [q_1 \quad q_2 \quad q_3] \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix}$$

where $q_1^T q_1 = q_2^T q_2 = q_3^T q_3 = 1$, $q_1^T q_2 = q_1^T q_3 = q_2^T q_3 = 0$

So, we have:

- $a_1 = q_1 r_{11}$
- $a_2 = q_1 r_{12} + q_2 r_{22}$
- $a_3 = q_1 r_{13} + q_2 r_{23} + q_3 r_{33}$

Step1: set $r_{11} = \|a_1\| = 4.36$

$q_1 = a_1 / r_{11} = [0.23, 0.46. 0.69, 0.46, -0.23]^T$

$r_{12} = q_1^T a_2 = 0.92$

$r_{13} = q_1^T a_3 = 0.23$

So, we have:

- $a_2 = q_1 r_{12} + q_2 r_{22}$
- $a_3 = q_1 r_{13} + q_2 r_{23} + q_3 r_{33}$

Step2: $q_2 r_{22} = a_2 - q_1 r_{12}$

$= \widehat{a_2} = [.79,.58,1.37,-2.42,1.21]^T$

Set $r_{22} = \|\widehat{a_2}\| = 3.1871$

$q_2 = \widehat{a_2} / r_{22} = [0.25, 0.18, 0.43, -0.76, 0.38]^T$

$r_{23} = q_2^T a_3 = 2.13$

So, we have:

- $a_3 = q_1r_{13} + q_2r_{23} + q_3r_{33}$

Step3: $q_3r_{33} = a_3 - q_1r_{13} - q_2r_{23}$

$= \widehat{a_3} = [1.47, 2.28, 0.76, -2.72, 2.86]^T$

Set $r_{33} = \|\widehat{a_3}\| = 2.33$

$q_3 = \widehat{a_3} / r_{33} = [0.18, 0.65, -0.46, 0.22, 0.53]^T$

Summarizing our example

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 2 \\ 3 & 2 & 0 \\ 2 & -2 & -1 \\ -1 & 1 & 2 \end{bmatrix} = \mathrm{QR}$$

$$= \begin{bmatrix} 0.23 & 0.35 & 0.18 \\ 0.46 & 0.18 & 0.65 \\ 0.69 & 0.43 & -0.46 \\ 0.46 & -0.76 & 0.22 \\ -0.23 & 0.38 & 0.53 \end{bmatrix} \begin{bmatrix} 4.36 & 0.92 & 0.23 \\ 0 & 3.19 & 2.13 \\ 0 & 0 & 2.33 \end{bmatrix}$$

Solving the Normal Equations

$$Rx = Q^T b$$

$$\begin{bmatrix} 4.36 & 0.92 & 0.23 \\ 0 & 3.19 & 2.13 \\ 0 & 0 & 2.33 \end{bmatrix} x = \begin{bmatrix} 5.74 \\ 5.25 \\ 2.79 \end{bmatrix}$$

Using backward substitution, we get

$$x = \begin{bmatrix} 1.07 \\ 0.84 \\ 1.20 \end{bmatrix}$$

How many steps to find Q and R?

- For p=1:n,
  - Calculate $\widehat{a_p}$, r_{pp} , q_p
  - For k=p+1:n, calculate r_{pk}
- This requires approximately $2mn^2$ flops

Comparing performance: Stability

- Consider the overdetermined system

$$\begin{bmatrix} 1 & -1 \\ 0 & 10^{-5} \\ 0 & 0 \end{bmatrix} x = \begin{bmatrix} 0 \\ 10^{-5} \\ 1 \end{bmatrix}$$

- Normal equations reduce to

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 + 10^{-5} \end{bmatrix} x = \begin{bmatrix} 0 \\ 10^{-10} \end{bmatrix}$$

- Exact solution is: x_1=x_2=1
- Consider calculations with 8 significant digits (allowing only exponent 0)

Cholesky

- Compute $A^TA \rightarrow \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$
- A^TA is singular → Cannot calculate L
- No solution by Cholesky

QR

- Factor A = QR = $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 10^{-5} \end{bmatrix}$
- $Q^T b = \begin{bmatrix} 0 \\ 10^{-5} \end{bmatrix}$

→ $x_1 = x_2 = 1$

QR is the more stable algorithm in this case (and in general)

Comparing performance: Efficiency

- Cholesky requires about $mn^2 + n^3/3$ flops
- QR requires about $2mn^2$ flops
- Cholesky is faster (if it applies)

Note: if A is large and sparse

- A^TA is sparse as well, and Cholesky can work very fast
- QR is not faster for sparse matrices

Overall – tend to use QR (unless large and sparse)