Conditioning of Linear Least Squares

Consider fitting a line to a collection of points, then perturbing the points:

- Linear least squares is ill-posed for any $A$, unless we consider solving for a particular $b$

If residual is small

$$\kappa(A) = \frac{\sigma_{\text{max}}}{\sigma_{\text{min}}}$$

is condition number of $A$

$\kappa(A) \neq \|A\| \cdot \|A^{-1}\|$ if $A$ is rectangular
Stability of Normal Equations

- Normal equations solve $A^T A x = A b$:

$$\kappa(A^T A) = \kappa(A)^2$$

$$S x = y$$

- We can often improve the solution to the normal equations by performing them again on the reduced $Q$:

$$\text{Cholesky } L L^T \Rightarrow A^T A = L L^T$$

$$A = QR \Rightarrow A^T A = R^T R$$

$$\Delta \Delta \Delta$$

$$\text{Cholesky } Q R$$

$$\text{so } A = \hat{Q} \hat{R} = Q \bar{R}$$

$$\hat{Q} = Q \bar{R}$$
Gram-Schmidt Orthogonalization

- Classical and Modified Gram-Schmidt process for QR:
  The $i$th column of $Q$ is $q_i = b_i/||b_i||_2$ where for CGS,
  $$b_i = a_i - \sum_{j=1}^{i-1} \langle a_j, a_i \rangle a_j,$$
  while for MGS
  $$b_i = a_i - \sum_{j=1}^{i-1} \langle q_j, a_i \rangle a_j.$$

- The cost of Gram-Schmidt is $2mn^2$ to leading order if $A \in \mathbb{R}^{m \times n}$:
  $$\text{inner product } \langle q_j, a_i \rangle \quad 2m \quad \text{adds and mults} \quad \sum_{i=1}^{n-1} = 2mn^2$$
Error in MGS Orthogonalization

- MGS can be expressed in terms of projection matrices \( P_i = I - q_i q_i^T \)

\[ P_i a_j = a_j - q_i q_i^T a_j = a_j - q_i (q_i^T a_j) \]

- The error in \( q_n \) due to a perturbation in \( a_n \) is amplified by \( \prod_{i=1}^{n-1} \kappa(P_i) \)

loss of orthogonality due to error in updating (applying tech \( \Pi_i \))
Householder QR: Eliminating Error in MGS

- Householder QR eliminates error amplification by using orthogonal Householder matrices (reflectors) rather than projection matrices:

\[ Q_i = I - 2u_iu_i^T \]

- The cost of using Householder QR to solve a least squares problem is \( 2mn^2 - 2n^3/3 \) to leading order.
Givens Rotations

- *Givens rotations* eliminate one element at a time $G = \begin{bmatrix} c & s \\ -s & c \end{bmatrix}$

  $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 11 \\ 21 \end{bmatrix}$

- Givens rotations can be advantageous when working with *sparse* matrices
Solving Rank-Deficient Least Squares Problems

- The **pseudoinverse** is defined by \( A^\dagger = V \Sigma^\dagger U^T \) where \( \Sigma^\dagger \) inverts only nonzero elements in \( \Sigma \), it satisfies \( AA^\dagger A = A \)

\[
A^\dagger = (A^T A)^{-1} A = \begin{bmatrix} \vdots \end{bmatrix} \begin{bmatrix} \vdots \end{bmatrix} \begin{bmatrix} \vdots \end{bmatrix}
\]

\( A \) is full rank \( A \) arbitrary \( \Rightarrow \) call

- Given a least squares problem \( Ax \approx b \), where \( A \) is rank-deficient, we can solve it via the pseudoinverse

\[
x = A^\dagger b
\]
An effective way to solve rank-deficient least squares problems without the SVD, is using QR with column pivoting \( AP = QR \).

The rank-deficient least squares problem \( Ax \approx b \) can be solved by QR with column pivoting.
Aggregation of Transformations

- Householder transformations can be aggregated in the form $I - YT YT^T$ where $Y$ is lower-trapezoidal and $T$ is upper-triangular.

- Given an arbitrary orthogonal matrix $Q$, we can compute $Y$ via LU factorization of $I - Q$. 