CS 450: Numerical Analysis\footnote{These slides have been drafted by Edgar Solomonik as lecture templates and supplementary material for the book “Scientific Computing: An Introductory Survey” by Michael T. Heath (slides).}

Linear Least Squares

University of Illinois at Urbana-Champaign
Linear Least Squares

- Find $x^* = \text{argmin}_{x \in \mathbb{R}^n} ||Ax - b||_2$ where $A \in \mathbb{R}^{m \times n}$:

- Given the SVD $A = U \Sigma V^T$ we have $x^* = V \Sigma^\dagger U^T b$, where $\Sigma^\dagger$ contains the reciprocal of all nonzeros in $\Sigma$: 
Conditioning of Linear Least Squares

- Consider fitting a line to a collection of points, then perturbing the points:
  - LLS is ill-posed for any $A$, unless we consider solving for a particular $b$. 

**Demo:** Polynomial fitting via the normal equations
Normal Equations

- *Normal equations* are given by solving $A^T A x = A^T b$.

- However, solving the normal equations is a more ill-conditioned problem than the original least squares algorithm.
Solving the Normal Equations

- If $A$ is full-rank, then $A^T A$ is symmetric positive definite (SPD):

  - Since $A^T A$ is SPD we can use Cholesky factorization, to factorize it and solve linear systems:
QR Factorization

- If $A$ is full-rank there exists an orthogonal matrix $Q$ and a unique upper-triangular matrix $R$ with a positive diagonal such that $A = QR$

- A reduced QR factorization (unique part of general QR) is defined so that $Q \in \mathbb{R}^{m \times n}$ has orthonormal columns and $R$ is square and upper-triangular.

- We can solve the normal equations (and consequently the linear least squares problem) via reduced QR as follows.
Gram-Schmidt Orthogonalization

▶ Classical Gram-Schmidt process for QR:

▶ Modified Gram-Schmidt process for QR:
A Householder transformation $Q = I - 2uu^T$ is an orthogonal matrix defined to annihilate entries of a given vector $z$, so $\|z\|_2 Q e_1 = z$:

Imposing this form on $Q$ leaves exactly two choices for $u$ given $z$,

$$u = \frac{z \pm \|z\|_2 e_1}{\|z \pm \|z\|_2 e_1\|_2}$$
Applying Householder Transformations

- The product \( x = Qw \) can be computed using \( O(n) \) operations if \( Q \) is a Householder transformation.

- Householder transformations are also called *reflectors* because their application reflects a vector along a hyperplane (changes sign of component of \( w \) that is parallel to \( u \)).
Givens Rotations

- Householder reflectors reflect vectors, Givens rotations rotate them

- Givens rotations are defined by orthogonal matrices of the form \[
\begin{bmatrix}
c & s \\
-s & c
\end{bmatrix}
\]
QR via Givens Rotations

- We can apply a Givens rotation to a pair of matrix rows, to eliminate the first nonzero entry of the second row

- Thus, $n(n - 1)/2$ Givens rotations are needed for QR of a square matrix
Suppose we want to solve a linear system or least squares problem with a (nearly) rank deficient matrix $A$.

Rank-deficient least squares problems seek a minimizer $x$ of $\|Ax - b\|_2$ of minimal norm $\|x\|_2$.
Truncated SVD

- After floating-point rounding, rank-deficient matrices typically regain full-rank but have nonzero singular values on the order of $\epsilon_{\text{mach}} \sigma_{\text{max}}$

- By the *Eckart-Young-Mirsky theorem*, truncated SVD also provides the best low-rank approximation of a matrix (in 2-norm and Frobenius norm)
QR with Column Pivoting

- QR with column pivoting provides a way to approximately solve rank-deficient least squares problems and compute the truncated SVD.

- A pivoted QR factorization can be used to compute a rank-\(r\) approximation.