CS 450: Numerical Anlaysis[1]

Linear Least Squares

University of Illinois at Urbana-Champaign

[1] *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

- Find $\boldsymbol{x}^\star = \operatorname{argmin}_{\boldsymbol{x} \in \mathbb{R}^n} ||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_2$ where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$:

- Given the SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ we have $\boldsymbol{x}^\star = \boldsymbol{V}\boldsymbol{\Sigma}^\dagger\boldsymbol{U}^T\boldsymbol{b}$, where $\boldsymbol{\Sigma}^\dagger$ contains the reciprocal of all nonzeros in $\boldsymbol{\Sigma}$:

Conditioning of Linear Least Squares

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

- LLS is ill-posed for any $\boldsymbol{A}$, unless we consider solving for a particular $\boldsymbol{b}$

Normal Equations

- *Normal equations* are given by solving $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b}$:

- However, solving the normal equations is a more ill-conditioned problem then the original least squares algorithm

Solving the Normal Equations

- If $\boldsymbol{A}$ is full-rank, then $\boldsymbol{A}^T\boldsymbol{A}$ is symmetric positive definite (SPD):
- Since $\boldsymbol{A}^T\boldsymbol{A}$ is SPD we can use Cholesky factorization, to factorize it and solve linear systems:

QR Factorization

- If $\boldsymbol{A}$ is full-rank there exists an orthogonal matrix $\boldsymbol{Q}$ and a unique upper-triangular matrix $\boldsymbol{R}$ with a positive diagonal such that $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$
- A reduced QR factorization (unique part of general QR) is defined so that $\boldsymbol{Q} \in \mathbb{R}^{m \times n}$ has orthonormal columns and $\boldsymbol{R}$ is square and upper-triangular
- We can solve the normal equations (and consequently the linear least squares problem) via reduced QR as follows

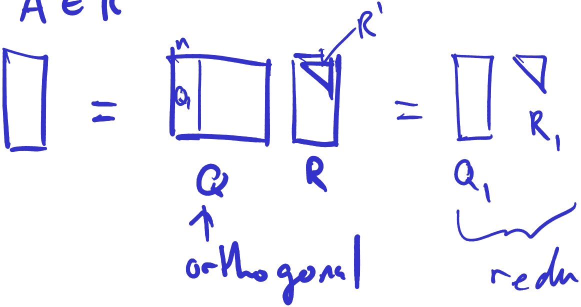
Gram-Schmidt Orthogonalization

- **Classical Gram-Schmidt process for QR:**

- **Modified Gram-Schmidt process for QR:**

Review: QR factorization

$A \in \mathbb{R}^{m \times n}$



$$Ax \cong b$$

$$QRx \cong b$$

$$Rx \cong Q^T b$$

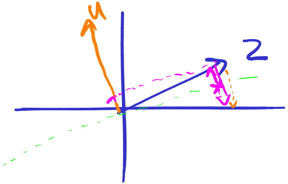
$$A^T A x = A^T b$$

$$\cancel{R^T} \underbrace{Q^T Q}_{I} R x = \cancel{R^T} Q^T b$$

$$\Longleftrightarrow \quad \underbrace{Rx = Q^T b}$$

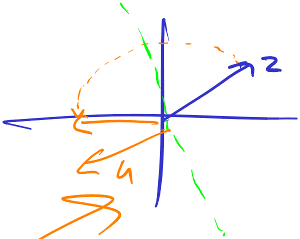
Householder QR Factorization

- **A Householder transformation $Q = I - 2uu^T$ is an orthogonal matrix defined to annihilate entries of a given vector z, so $||z||_2 Qe_1 = z$:**



$z - 2uu^T z$

$$\begin{bmatrix} ||z|| \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



$$\begin{bmatrix} -||z||_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- 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 $\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{w}$ can be computed using $O(n)$ operations if $\boldsymbol{Q}$ is a Householder transformation

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

$$Q^T = Q$$

$$Q_1 A \Longrightarrow \begin{bmatrix} \|a_1\| & u \\ 0 & B \end{bmatrix}$$

$$Q^T A = \text{closer to } R$$

$$\hat{Q}_2 B = \begin{bmatrix} \|b_1\| & v \\ 0 & C \end{bmatrix} \to \text{continue}$$

$$A = QR$$

Householder reflector

$$Q^T = Q_1 \cdot Q_2 \cdot \ldots \cdot Q_n I$$

$$\begin{bmatrix} 1 & \\ & \hat{Q}_2 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ & & Q_3 \end{bmatrix}$$

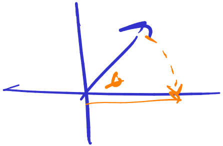
$$(I - 2uu^T) \underset{\uparrow\, n \times n}{S} \Rightarrow O(n^2)$$

$$O(n)$$

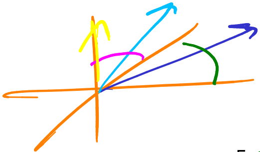
$$(I - 2uu^T) v = v - 2u \underline{\langle u, v \rangle}$$

Givens Rotations

- Householder reflectors reflect vectors, Givens rotations rotate them



orthogonal matrices are compositions of
- rotations
- reflections



- Givens rotations are defined by orthogonal matrices of the form $\begin{bmatrix} c & s \\ -s & c \end{bmatrix}$

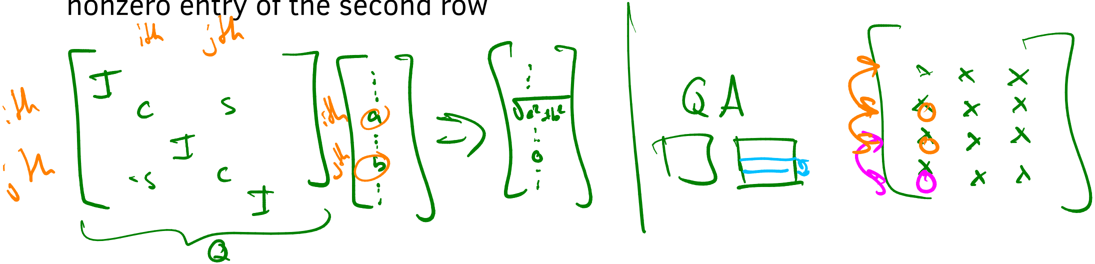
not symmetric

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sqrt{a^2+b^2} \\ 0 \end{bmatrix}$$

$$c = \frac{a}{\sqrt{a^2+b^2}} \qquad s = \frac{b}{\sqrt{a^2+b^2}}$$

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

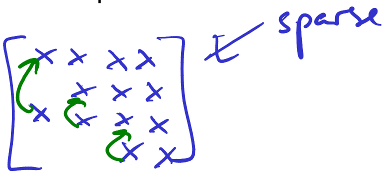


- Thus, $n(n-1)/2$ Givens rotations are needed for QR of a square matrix

$O(n)$ per Givens rotation

$O(n^3)$ in total

$\sim 2n^3$ in total vs $\frac{4}{3}n^3$ for Householder



Rank-Deficient Least Squares

- Suppose we want to solve a linear system or least squares problem with a (nearly) rank deficient matrix A

singular $\quad \exists\, x \neq 0$ s.t. $Ax = 0$, $\sigma_{min} = 0$

no unique soln to $Ax = b$ or $Ax \cong b$

if A is singular $fl(A)$ probably isn't, but $\sigma_{min} \approx \sigma_{max}\, \varepsilon_{mach}$

$\kappa(fl(A)) = 1/\varepsilon_{mach}$

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

$$x = A^{+} b \Rightarrow A^{+} = V \Sigma^{+} U^{T}, \quad \Sigma^{+} = \begin{bmatrix} \sigma_1^{+} & & \\ & \ddots & \\ & & \sigma_n^{+} \end{bmatrix} \quad \sigma_i^{+} = \begin{cases} 1/\sigma_i & \text{if } \sigma_i > 0 \\ 0 & \text{if } \sigma_i = 0 \end{cases}$$

for numerical rank $\sigma_i^{+} = 0$ if σ_i is very small

Truncated SVD

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

σ_{min}

(Tykhonov) alternative: regularization parameter

$$\min_x \|Ax-b\|_2 + \alpha\|x\|_2$$

$$\begin{bmatrix}A\\ \sqrt{\alpha}I\end{bmatrix}x \cong \begin{bmatrix}b\\ 0\end{bmatrix} \Rightarrow \min_x \left\|\begin{bmatrix}Ax-b\\ \sqrt{\alpha}x\end{bmatrix}\right\| =$$

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

$A \approx$ ⇐ low-rank approximation

rank-r truncated SVD is B, then $\|A-B\|_{2/F}$ is minimal

$$A = U \cdot \begin{bmatrix}\sigma_{max} & & & & \\ & \ddots & & & \\ & & 0 & & \\ & & & 0 & \\ & & & & 0\end{bmatrix} U^T = \quad \cdot \quad 0 \quad =$$

QR with Column Pivoting

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

$QR = AP$

↑ permutation matrices

$QR =$

$Ax \cong b$

$\underbrace{QRP^T x}_{y} \cong b$

$Ry = Q^T b \qquad x = Py$

$= AP$

- A pivoted QR factorization can be used to compute a rank-r approximation

taking first r columns of AP, and compute SVD

$O(nr^2)$

for each column (Householder reflector)

$\underbrace{\text{pivot}}_{P_1}$ column of largest norm to be leading column

$Q_1 A P_1 = \begin{bmatrix} \alpha & b \\ 0 & B \end{bmatrix} \rightarrow$ continue with B | Greedy algorithm