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 $\boldsymbol{\Sigma}$:

Demo: *Polynomial fitting via the normal equations*

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

Demo: *Normal equations vs Pseudoinverse*
Demo: *Issues with the 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 $\boldsymbol{A} = \boldsymbol{QR}$
- 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

Demo: *Gram-Schmidt–The Movie*
Demo: *Gram-Schmidt and Modified Gram-Schmidt*

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

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

Demo: *3x3 Householder demo*

Householder QR Factorization

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

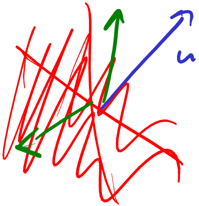
- Imposing this form on Q leaves exactly two choices for u given z,

$$\boldsymbol{u} = \frac{\boldsymbol{z} \pm ||\boldsymbol{z}||_2 \boldsymbol{e}_1}{||\boldsymbol{z} \pm ||\boldsymbol{z}||_2 \boldsymbol{e}_1||_2}$$

Applying Householder Transformations

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

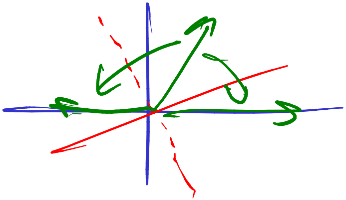
$I - 2uu^T$



- 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)

$$Qw = \pm e_1 \|w\|$$

$$\begin{bmatrix} \pm\|w\| \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



$Q = \text{House}(v)$

$Qv = \pm e_1 \|v\|$

$Q = I - 2uu^T$ $\quad U(n)$

$Qw = \underbrace{w - 2u\langle u, w\rangle}_{O(n)}$

$[Q, R] = QR(A) \quad A \in m \times n$

$Q = I$

For $j = 1$ to $\#\text{cols}(A)$ $\quad n$

$\quad R^{(n-j)\times(n-j)}$ $\quad Q_j = \text{House}(A[j:, j])$

$\quad A[j:, j:] \leftarrow Q_j^T A[j:, j:] \Leftarrow O(mn)$

$\quad \underbrace{m}_{}, \; n-j$

$\quad Q \leftarrow Q \begin{bmatrix} I & \\ & Q_j \end{bmatrix}$

$R = A$

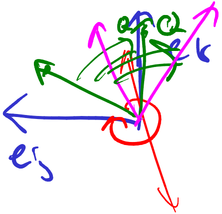
$\begin{bmatrix} x & x & x \\ x & x & x \\ x & x & x \end{bmatrix} \Rightarrow \begin{bmatrix} x & x & x \\ 0 & x & x \\ 0 & 0 & x \end{bmatrix}$

$\Downarrow$ overall (across j)

$O(mn^2)$

Givens Rotations

- Householder reflectors reflect vectors, Givens rotations rotate them



$$G^{-1} = G^T$$

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

$$ca + sb = \sqrt{a^2+b^2}$$

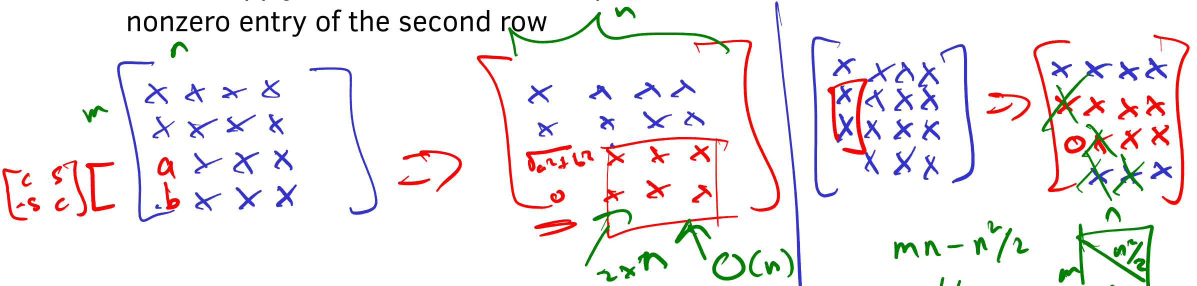
$$-sa + cb = 0 \Rightarrow c = \frac{sa}{b} = \frac{a}{\sqrt{a^2+b^2}}$$

$$G\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sqrt{a^2+b^2} \\ 0 \end{bmatrix}$$

$$\frac{sa^2}{b} + sb = \sqrt{a^2+b^2} \Rightarrow \underbrace{sa^2 + sb^2}_{s(a^2+b^2)} = b\sqrt{a^2+b^2} \Rightarrow s = \frac{b}{\sqrt{a^2+b^2}}$$

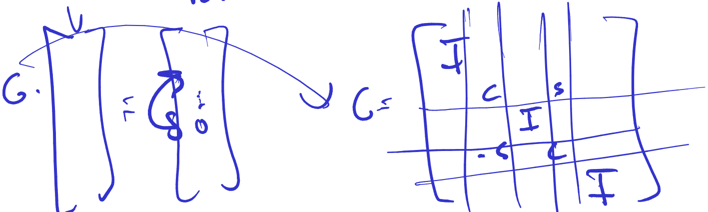
QR via Givens Rotations

- 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

for n×n matrix



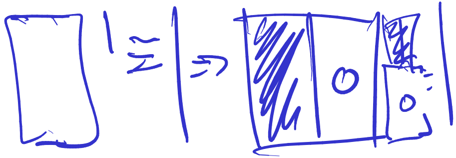
mn · O(n) ⇒ O(mn²)

Rank-Deficient Least Squares

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

$$Ax = b$$

$$Ax \cong b$$



- Rank-deficient least squares problems seek a minimizer $\boldsymbol{x}$ of $||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_2$ of minimal norm $||\boldsymbol{x}||_2$

$$x^* = \underset{x}{\operatorname{argmin}} \, ||x||_2 \quad \text{s.t. } Ax - b \perp \operatorname{span}(A)$$

$$\text{i.e., } ||Ax^* - b||_2 \leq ||Ax - b||_2 \quad \forall x$$

Truncated SVD

Demo: Image compression

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

$A \in \mathbb{R}^{m \times n}$ is low rank if $\operatorname{rank}(A) < \min(m,n)$ $\Rightarrow A = \underset{m \times r}{[\;]}\,\underset{r \times n}{[\;]}$

$$[\;] = [U][\Sigma][V^T] = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

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

$$\min_{\operatorname{rank}(\tilde{A}) = r} \|A - \tilde{A}\|_{2 \text{ or } F}$$

minimized $\tilde{A} = \sum_{i=1}^{r} \sigma_i u_i v_i^T$

$$\underbrace{\sigma_1 \geq \sigma_2 \cdots \geq \sigma_r \cdots \geq \sigma_n}_{\text{singular values of } A}$$

$$A^{+} = \sum_{i=1}^{r} \underline{\frac{1}{\sigma_i}} v_i u_i^T$$

$$A = \begin{bmatrix} \\ u \\ \\ \end{bmatrix} \begin{bmatrix} S & 0 \end{bmatrix} \begin{bmatrix} V^T \end{bmatrix}$$

$$\min \|Ax - b\|_4 \quad \|x\|$$

$$x^* = A^{+} b$$

$$Ax = b$$

$$S V_1 x = u_1^T b$$

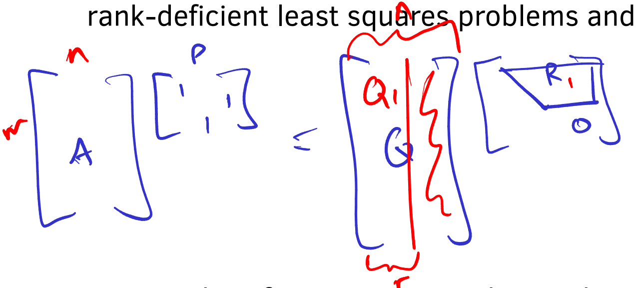
$$\begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} S & 0 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} b \\ \\ \end{bmatrix}$$

$V_2 x = 0$ to get

$\min \|x\|_2$

QR with Column Pivoting

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



$AP = Q_1 R_1$

$A = Q_1 R_1$

$R_1 x \cong Q_1^T b$ with $\|x\|_2$

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

target norm

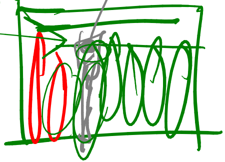
$[Q, R, P] \leftarrow QRCP(A) \implies AP = QR$

$\mu_j = \|A[:, j]\|_2^2$

rank(A)

for $j = 1$ to # cols (A)

$\|A[j:n, k]\|_2^2$



$j^* = \underset{k \geq j}{\operatorname{argmin}}\ \mu_k$

$P[j^*, j] = 1$

$A[:, j] \leftrightarrow A[:, j^*]$

$Q = \text{House}(A[j:n, j])$

$A[j:n, j:n] \leftarrow Q^T A[j:n, j:n]$

$\mu[j+1:n] \leftarrow \mu[j+1:n] - A[j, j+1:n]^2$

Eigenvalues and Eigenvectors

$$Av = \lambda v$$

$\mathbb{C}$, e.g. 0, 2.3, $1-i$, $1+i$

eigenvalues

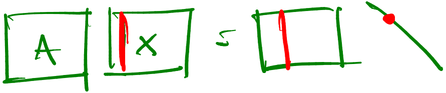
similarity transformation

$$B = ZAZ^{-1}$$

$$A = XDX^{-1}$$

eigenvectors

$$AX = XD \implies$$



$A = USU^T$

$A^TA = US\cancel{U^TU}SU^T$

$$A^TA = V S^2 V^{-1} \implies V^{-1} = V^T$$

singular values