CS 450: Numerical Anlaysis

Lecture 7

Chapter 3 – Linear Least Squares

Stability and Efficiency of Linear Least Squares Algorithms

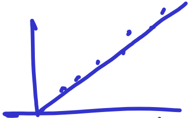
Edgar Solomonik

Department of Computer Science
University of Illinois at Urbana-Champaign

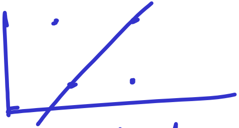
February 7, 2018

Conditioning of Linear Least Squares

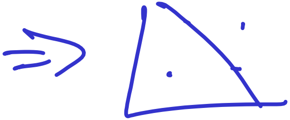
- Consider fitting a line to a collection of points, then perturbing the points:



residual small



residual large



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

if residual is small

$\kappa(A)$ condition number of A

is condition of LLS

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

$\kappa(A) \neq \|A\| \|A^{-1}\|$ if A is rectangular

Stability of Normal Equations

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

y

$$\kappa(\underbrace{A^T A}_{S}) = \kappa(A)^2$$

$$S x = y$$

why is $\dfrac{\|Ax\|}{\|b\|} \leq 1$

$$R x = Q^T b$$

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

Cholesky-QR

$$\mathrm{chol}(A^T A) = L L^T \Rightarrow A^T A = L L^T$$

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

Cholesky-QR2

so $A \approx \hat{Q}\hat{R} = Q \underbrace{\bar{R}\hat{R}}_{R}$

$$\hat{Q} = Q\bar{R}$$

Gram-Schmidt Orthogonalization

- Classical and Modified Gram-Schmidt process for QR:
  The ith column of $\boldsymbol{Q}$ is $\boldsymbol{q}_i = \boldsymbol{b}_i/||\boldsymbol{b}_i||_2$ where for CGS,

$$\boldsymbol{b}_i = \boldsymbol{a}_i - \sum_{j=1}^{i-1} \frac{\langle \boldsymbol{a}_j, \boldsymbol{a}_i \rangle}{\langle \boldsymbol{a}_j, \boldsymbol{a}_j \rangle} \boldsymbol{a}_j,$$

  while for MGS

$$\boldsymbol{b}_i = \boldsymbol{a}_i - \sum_{j=1}^{i-1} \langle \boldsymbol{q}_j, \overset{2m}{\boldsymbol{a}_i} \rangle \underset{2m}{\boldsymbol{a}_j}.$$

$\text{span}([a_1, \dots a_{i-1}])$

$= \text{span}([q_1 \dots q_{i-1}])$

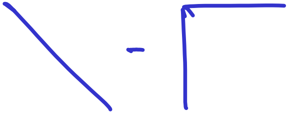
- The cost of Gram-Schmidt is $2mn^2$ to leading order if $\boldsymbol{A} \in \mathbb{R}^{m \times n}$: $\sum_{i=1}^{n} (n-i)4m = 2mn^2$

inner product $\langle q_j, a_i \rangle$ $2m$

adds and mults

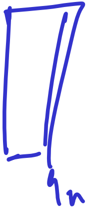
Error in MGS Orthogonalization

- MGS can be expressed in terms of projection matrices $\boldsymbol{P}_i = \boldsymbol{I} - \boldsymbol{q}_i \boldsymbol{q}_i^T$



$P_i a_j = a_j - q_i q_i^T a_j = a_j - q_i \langle q_i, a_j \rangle$

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



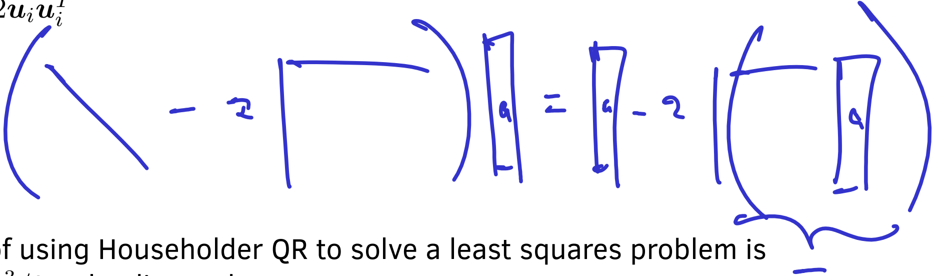
loss of orthogonality due to error
in updating (applying each P_i)

Householder QR: Eliminating Error in MGS

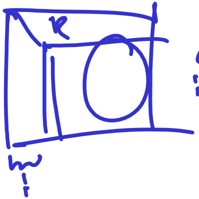
- Householder QR eliminates error amplification by using orthogonal Householder matrices (reflectors) rather than projection matrices $\boldsymbol{Q}_i = \boldsymbol{I} - 2\boldsymbol{u}_i\boldsymbol{u}_i^T$

$v = a \pm \|a\| e$

$a[0] > 0$ then $+$



- The cost of using Householder QR to solve a least squares problem is $2mn^2 - 2n^3/3$ to leading order



$\sum_{i=1}^{n} 2(m-i)(n-i) = 2mn^2 - 2n^3/3$

$O(mn^2)$

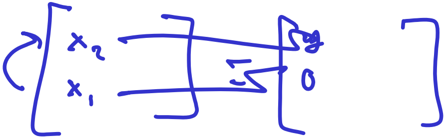
$H_2 H_1 R = A$

$Ax \cong b$

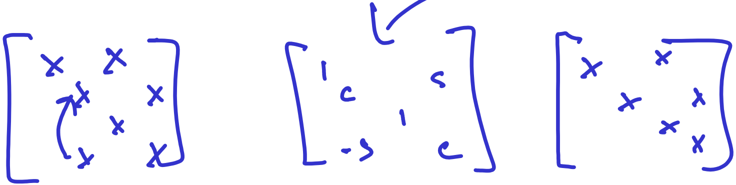
$Rx \cong H_1 H_2 b$

Givens Rotations

- *Givens rotations* eliminate one element at a time $G = \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \|z\| \\ 0 \end{bmatrix}$



- Givens rotations can be advantageous when working with *sparse* matrices



Solving Rank-Deficient Least Squares Problems

- The *pseudoinverse* is defined by $\boxed{A^{\dagger} = V\Sigma^{\dagger}U^{T}}$ where $\Sigma^{\dagger}$ inverts only nonzero elements in Σ, it satisfies $AA^{\dagger}A = A$

$$\underbrace{A^{+} = (A^{T}A)^{-1}A}_{A \text{ is full rank}} \quad = \quad \underbrace{[\ ]\,[\ddots]\,[\ ]}_{A \text{ arbitrary}}$$

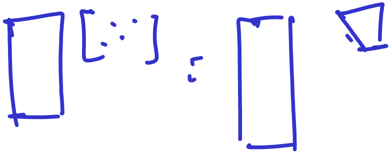
reciprocate diagonal entries

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

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

QR with Column Pivoting

- An effective way to solve rank-deficient least squares problems without the SVD, is using QR with column pivoting $\boldsymbol{AP} = \boldsymbol{QR}$



- The rank-deficient least squares problem $\boldsymbol{Ax} \cong \boldsymbol{b}$ can be solved by QR with column pivoting

Aggregation of Transformations

- Householder transformations can be *aggregated* in the form $\boldsymbol{I} - \boldsymbol{Y}\boldsymbol{T}\boldsymbol{Y}^T$ where $\boldsymbol{Y}$ is lower-trapezoidal and $\boldsymbol{T}$ is upper-triangular

- Given an arbitrary orthogonal matrix $\boldsymbol{Q}$, we can compute $\boldsymbol{Y}$ via LU factorization of $\boldsymbol{I} - \boldsymbol{Q}$