CS 450: Numerical Anlaysis

Lecture 6

Chapter 3 – Linear Least Squares

QR Factorization

Edgar Solomonik

Department of Computer Science
University of Illinois at Urbana-Champaign

February 2, 2018

Linear Least Squares

$x^* = \operatorname{argmin}_x f(x) \quad \Big| \quad \min_x f(x) = f(x^*)$

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

$x^* = \operatorname{argmin}_x ||Ax-b||_2^2 = \operatorname{argmin}_x \langle Ax-b, Ax-b\rangle$

- Given the reduced SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ we have $\boldsymbol{x}^\star = \underbrace{\boxed{\boldsymbol{V}\Sigma^{-1}\boldsymbol{U}^T}}_{A^+}\boldsymbol{b}$:

$$x^* = \operatorname{argmin}_x (Ax-b)^T(Ax-b) = (U\Sigma V^T x - b)^T\underbrace{(U\Sigma V^T x - b)}_{UU^T}$$

$$= (\cancel{U^T U}\Sigma V^T x - U^T b)^T(\cancel{U^T U}\Sigma V^T x - U^T b)$$

$$= (\Sigma V^T x - U^T b)^T(\Sigma V^T x - U^T b) = (V^T x - \Sigma^{-1}U^T b)$$

VV

$(V^T x - \Sigma^{-1}U^T b)$

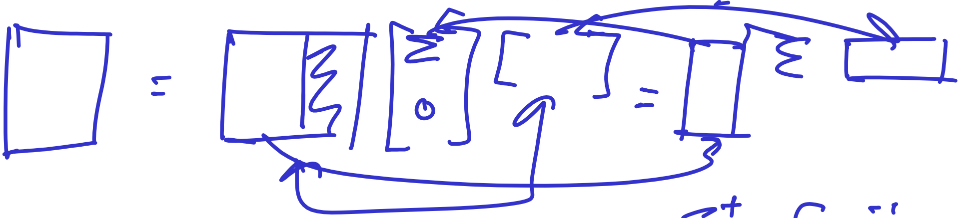
$\boxed{x = V\Sigma^{-1}U^T b}$

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 reduced SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ we have $\boxed{\boldsymbol{x}^\star = \boldsymbol{V}\Sigma^{-1}\boldsymbol{U}^T\boldsymbol{b}}$:

Σ^+



square

$$\Sigma^+ = \begin{bmatrix} \Sigma^{-1} & 0 \end{bmatrix}$$

Normal Equations

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

$$(U\Sigma V^T)^T (U\Sigma V^T)\, x^* = (U\Sigma V^T)^T b$$

$$\cancel{V}\cancel{\Sigma}\cancel{U^T}\, U \Sigma V^T x^* = \cancel{V}\cancel{\Sigma}\cancel{U^T}\, b \Rightarrow x^* = V\Sigma^{-1}U^T b$$

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

$B = \underbrace{A^T A}$ - symmetric

- positive definite, eigenvalues of B are squares of the singular values of A

$Bx = y$ can be solved by $B = LL^T$

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 $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$

$A = QR$ ← upper-triangular

↑ orthogonal

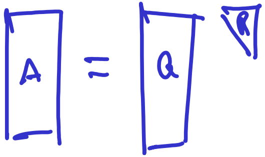
$Ax = b$

$QRx = b$

$x = R^{-1}Q^T b$

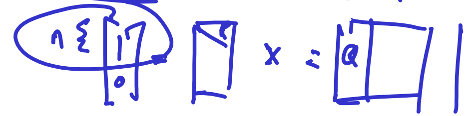
$Rx = Q^T b$

- 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



$Ax \cong b$

$Rx = Q^T b$



Gram-Schmidt Orthogonalization

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

A → Q ← orthonormal cols.

let a_i be ith column of A

span(A) = span(Q)

q_i B q_1

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

$q_1 = \frac{a_1}{\|a_1\|}$, $q_2 = a_2 - a_1 \frac{\langle a_1, a_2 \rangle}{\langle a_1, a_1 \rangle}$

normalize

$q_3 = a_3 - a_1 \frac{\langle a_1, a_3 \rangle}{\langle a_1, a_1 \rangle} - a_2 \frac{\langle \ \rangle}{\langle \ \rangle}$

$A = Q \nabla$

$a_1 = q_1 \|a_1\|$

$a_2 = q_1 \cdots + q_2 \cdots$

$a_3 = q_1 \qquad q_2 \qquad q_3$

Householder QR Factorization

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

$$Q \begin{bmatrix} z_1 \\ \vdots \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} ||z|| \\ 0 \\ \vdots \\ 0 \end{bmatrix} = Qz = ||z||_2 e_1$$

$$z = Q\, ||z||_2 e_1$$

$$QR = Q_1 \ldots Q_n R$$

Computing Householder Transformations

- To find a Householder transformation that annihilates a given vector $\boldsymbol{z}$, compute $\boldsymbol{u} = \frac{\boldsymbol{z} \pm ||\boldsymbol{z}||_2 \boldsymbol{e}_1}{||\boldsymbol{z} \pm ||\boldsymbol{z}||_2 \boldsymbol{e}_1||_2}$

- 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

Applying Householder Transformations

- The product $\boldsymbol{Q}\boldsymbol{w}$ 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 $\boldsymbol{u}$)