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 $\boldsymbol{x}^\star = \underbrace{\boldsymbol{V}\boldsymbol{\Sigma}^\dagger\boldsymbol{U}^T}_{\boldsymbol{A}^\dagger}\boldsymbol{b}$, where $\boldsymbol{\Sigma}^\dagger$ contains the reciprocal of all nonzeros in $\boldsymbol{\Sigma}$:

Data Fitting via Linear Least Squares

- Given a set of m points with coordinates $\boldsymbol{x}$ and $\boldsymbol{y}$, seek an $n-1$ degree polynomial p so that $p(x_i) \approx y_i$ by minimizing

- we can write this objective as a linear least squares problem

Conditioning of Linear Least Squares

- Consider a perturbation $\boldsymbol{\delta b}$ to the right-hand-side $\boldsymbol{b}$

- The amplification in relative perturbation magnitude (from $\boldsymbol{b}$ to $\boldsymbol{x}$) depends on how much of $\boldsymbol{b}$ is spanned by the columns of $\boldsymbol{A}$,

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:**

Householder QR Factorization

***Demo:** 3x3 Householder demo*

- **A Householder transformation $Q = I - 2uu^T$ is an orthogonal matrix defined to annihilate entries of a given vector z, so $Qz = \pm||z||_2 e_1$:**

- 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 $\boldsymbol{u}$)

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

Rank-Deficient Least Squares

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

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

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 $\epsilon_{\text{mach}}\sigma_{\text{max}}$
- 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