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 7, 2018

Linear Least Squares

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

  Since $m \geq n$, the minimizer generally does not attain a zero residual $\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$. We can rewrite the optimization problem constraint via

$$\boldsymbol{x}^\star = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{argmin}} ||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_2^2 = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{argmin}} \left[(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})\right]$$

- Given the SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ we have $\boldsymbol{x}^\star = \boldsymbol{V}\boldsymbol{\Sigma}^\dagger\boldsymbol{U}^T\boldsymbol{b}$:

$$\begin{aligned}
0 &= (\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x}^\star - \boldsymbol{b})^T\boldsymbol{U}\boldsymbol{U}^T(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x}^\star - \boldsymbol{b}) - \min_{\boldsymbol{x}\in\mathbb{R}^n} ||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_2 \\
&= (\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x}^\star - \boldsymbol{U}^T\boldsymbol{b})^T(\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x}^\star - \boldsymbol{U}^T\boldsymbol{b}) - \min_{\boldsymbol{x}\in\mathbb{R}^n} ||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_2 \\
&= (\boldsymbol{V}^T\boldsymbol{x}^\star - \boldsymbol{\Sigma}^\dagger\boldsymbol{U}^T\boldsymbol{b})^T\boldsymbol{\Sigma}^2(\boldsymbol{V}^T\boldsymbol{x}^\star - \boldsymbol{\Sigma}^\dagger\boldsymbol{U}^T\boldsymbol{b}) \\
&= (\boldsymbol{x}^\star - \boldsymbol{V}\boldsymbol{\Sigma}^\dagger\boldsymbol{U}^T\boldsymbol{b})^T\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^T(\boldsymbol{x}^\star - \boldsymbol{V}\boldsymbol{\Sigma}^\dagger\boldsymbol{U}^T\boldsymbol{b})
\end{aligned}$$

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

Normal Equations

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

$$\begin{aligned}
(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^T\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x} &= (\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^T\boldsymbol{b} \\
\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x} &= \boldsymbol{\Sigma}^T\boldsymbol{U}^T\boldsymbol{b} \\
\boldsymbol{V}^T\boldsymbol{x} &= (\boldsymbol{\Sigma}^T\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}^T\boldsymbol{U}^T\boldsymbol{b} = \boldsymbol{\Sigma}^\dagger\boldsymbol{U}^T\boldsymbol{b} \\
\boldsymbol{x} &= \boldsymbol{V}\boldsymbol{\Sigma}^\dagger\boldsymbol{U}^T\boldsymbol{b} = \boldsymbol{x}^\star
\end{aligned}$$

- However, solving the normal equations is a more ill-conditioned problem then the original least squares algorithm
  Generally we have $\kappa(\boldsymbol{A}^T\boldsymbol{A}) = \kappa(\boldsymbol{A})^2$ (the singular values of $\boldsymbol{A}^T\boldsymbol{A}$ are the squares of those in $\boldsymbol{A}$). Consequently, solving the least squares problem via the normal equations may be unstable because it involves solving a problem that has worse conditioning than the initial least squares problem.

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

  Existence and uniqueness shown constructively by Gram-Schmidt orthogonalization process.
  We have $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{R}^T\boldsymbol{R}$, so the solution to the normal equations (which is also the minimizer $\boldsymbol{x}^\star$ satisfies $\boldsymbol{R}^T\boldsymbol{R}\boldsymbol{x}^\star = \boldsymbol{R}^T\boldsymbol{Q}^T\boldsymbol{b}$. Furthermore, it suffices to solve $\boldsymbol{R}\boldsymbol{x}^\star = \boldsymbol{Q}^T\boldsymbol{b}$, which can be done by backward substitution after transforming $\boldsymbol{b}$.

- 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

  A full QR factorization gives $\boldsymbol{Q} \in \mathbb{R}^{m\times m}$ and $\boldsymbol{R} \in \mathbb{R}^{m\times n}$, but since $\boldsymbol{R}$ is upper triangular, the latter $m-n$ columns of $\boldsymbol{Q}$ are only constrained so as to keep $\boldsymbol{Q}$ orthogonal. The reduced QR factorization is given by taking the first n columns $\boldsymbol{Q}$ and $\hat{\boldsymbol{Q}}$ the upper-triangular block of $\boldsymbol{R}$, $\hat{\boldsymbol{R}}$.

Gram-Schmidt Orthogonalization

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

  The Gram-Schmidt process orthogonalizes a rectangular matrix, i.e. it finds a set of orthonormal vectors with the same span as the columns of the given matrix. If $\boldsymbol{a}_i$ is the ith column of the input matrix, the ith orthonormal vector (ith column of $\boldsymbol{Q}$) is

$$\boldsymbol{q}_i = \boldsymbol{b}_i / ||\boldsymbol{b}_i||_2, \quad \boldsymbol{b}_i = \boldsymbol{a}_i - \sum_{j=1}^{i-1} \langle \boldsymbol{q}_j, \boldsymbol{a}_i \rangle \boldsymbol{q}_j.$$

- **Modified Gram-Schmidt process for QR:** *Better numerical stability is achieved by orthogonalizing each vector with respect to each previous vector in sequence (modifying the vector prior to orthogonalizing to the next vector),*

$$\boldsymbol{q}_i = \boldsymbol{b}_i / ||\boldsymbol{b}_i||_2, \quad \boldsymbol{b}_i = \boldsymbol{b}_i^{(i-1)}, \quad \boldsymbol{b}_i^{(j)} = \boldsymbol{b}_i^{(j-1)} - \langle \boldsymbol{q}_j, \boldsymbol{b}_i^{(j-1)} \rangle \boldsymbol{q}_j, \quad \boldsymbol{b}_i^{(0)} = \boldsymbol{a}_i.$$

Householder QR Factorization

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

  Householder QR achieves unconditional stability, by applying only orthogonal transformations to reduce the matrix to upper-triangular form. Householder transformations (reflectors) are orthogonal matrices, that reduce a vector to a multiple of the first elementary vector, $\alpha e_1 = Qz$ Because multiplying a vector by an orthogonal matrix preserves its norm, we must have that $|\alpha| = ||z||_2$. As we will see, this transformation can be achieved by a rank-1 perturbation of identify of the form $Q = I - 2uu^T$ where u is a normalized vector. Householder matrices are both symmetric and orthogonal implying that $Q = Q^{-1}$. Imposing this form on Q leaves exactly two choices for u given z.