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 11, 2018

Conditioning of Linear Least Squares

- Consider fitting a line to a collection of points, then perturbing the points:
  - *If our line closely fits all of the points, a small perturbation to the points will not change the ideal fit line (least squares solution) much. Note that, if a least squares solution has a very small residual, any other solution with a residual close to as small, should be close to parallel to this solution.*
  - *When the points are distributed erratically and do not admit a reasonable linear fit, then the least squares solution has a large residual, and totally different lines may exist with a residual nearly as small. For example, if the points are in a ball around the origin, any linear fit has the same residual. A tiny perturbation could then perturb the least squares solution to be perpendicular to the original.*
- LLS is ill-posed for any $\boldsymbol{A}$, unless we consider solving for a particular $\boldsymbol{b}$
  - *If $\boldsymbol{b}$ is outside the span $\boldsymbol{A}$ then any perturbation to $\boldsymbol{A}$ or $\boldsymbol{b}$ can completely defines the new solution. Similarly, if most of $\boldsymbol{b}$ is outside the span of $\boldsymbol{A}$, a perturbation can cause the solution to fluctuate wildly.*
  - *On other hand, if for a particular $\boldsymbol{b}$ we can find a solution with (near-)zero residual, a small relative perturbation to $\boldsymbol{b}$ or $\boldsymbol{A}$ will have an effect similar to that of a linear system perturbation (growth bounded by $\kappa(\boldsymbol{A}) = \sigma_{max}/\sigma_{min}$).*

Stability of Normal Equations

- Normal equations solve $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{b}$:

  Consider perturbation $\boldsymbol{\delta b}$ to $\boldsymbol{b}$, in the worst case $\boldsymbol{A}\boldsymbol{b}$ is perturbed by $\sigma_{max}\boldsymbol{\delta b}$. Further $\kappa(\boldsymbol{A}^T\boldsymbol{A}) = \kappa(\boldsymbol{A})^2$ so the relative error can be amplified by the square of the expected condition number, given a small residual in the true solution.
- We can also obtain a QR factorization by means of Cholesky of $\boldsymbol{A}^T\boldsymbol{A}$ (known as Cholesky-QR):
  - *The QR factorization $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$ so long as $\boldsymbol{R}$ has a positive diagonal, so we have*

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

    and thus can obtain $\boldsymbol{R}$ by Cholesky $\boldsymbol{A}^T\boldsymbol{A}$, which is symmetric positive definite.
  - *However, if condition number of $\boldsymbol{A}^T\boldsymbol{A}$ is high, we can lose positive-definiteness due to roundoff error and Cholesky factorization can break down.*
  - *Moreover, if we obtain $\hat{\boldsymbol{Q}}$ via $\boldsymbol{A}\boldsymbol{R}^{-1}$, round-off error can result in a loss of orthogonality in the columns of $\hat{\boldsymbol{Q}}$.*
  - *So long as $\kappa(\boldsymbol{A}) < \sqrt{1/\epsilon_{mach}}$, a second iteration of Cholesky-QR on the computed reduced $\hat{\boldsymbol{Q}} = \boldsymbol{Q}\hat{\boldsymbol{R}}$ obtains an orthogonal $\boldsymbol{Q}$ and a correction $\hat{\boldsymbol{R}}$ to $\boldsymbol{R}$.*