CS 450: Numerical Anlaysis

Lecture 10
Chapter 4 – Eigenvalue Problems
Theory of Eigenvalue Solvers

Edgar Solomonik

Department of Computer Science
University of Illinois at Urbana-Champaign

February 26, 2018

Perturbation Analysis of Eigenvalue Problems

- Suppose we seek eigenvalues $\boldsymbol{D} = \boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}$, but find those of a slightly perturbed matrix $\boldsymbol{D} + \boldsymbol{\delta D} = \hat{\boldsymbol{X}}^{-1}(\boldsymbol{A} + \boldsymbol{\delta A})\hat{\boldsymbol{X}}$:
  Note that the eigenvalues of $\boldsymbol{X}^{-1}(\boldsymbol{A} + \boldsymbol{\delta A})\boldsymbol{X} = \boldsymbol{D} + \boldsymbol{X}^{-1}\boldsymbol{\delta A}\boldsymbol{X}$ are also $\boldsymbol{D} + \boldsymbol{\delta D}$. So if we have perturbation to the matrix $||\boldsymbol{\delta A}||_F$, its effect on the eigenvalues corresponds to a (non-diagonal/arbitrary) perturbation of a diagonal matrix of eigenvalues, with norm
  $$||\boldsymbol{X}^{-1}||_2||\boldsymbol{\delta A}||_F||\boldsymbol{X}||_2 = \kappa(\boldsymbol{X})||\boldsymbol{\delta A}||_F.$$

- Thus it generally suffices to consider the effect of a (possibly larger) perturbation $\boldsymbol{\delta}\hat{\boldsymbol{A}}$ to a diagonal matrix: *Given a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, let $r_i = \sum_{j\neq i}|a_{ij}|$. We define the Gershgorin disks as*
  $$D_i = \{z \in \mathbb{C} : |z - a_{ii}| \leq r_i\}.$$
  The eigenvalues $\lambda_1, \dots, \lambda_n$ of any matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ are contained in the union of the Gershgorin disks, $\forall i \in \{1, \dots, n\}, \lambda_i \in \bigcup_{j=1}^{n} D_j$.

Conditioning of Particular Eigenpairs

- Consider the effect of a matrix perturbation on an eigenvalue λ associated with a right eigenvector $\boldsymbol{x}$ and a left eigenvector $\boldsymbol{y}^H$, $\lambda = \boldsymbol{y}^H \boldsymbol{A}\boldsymbol{x}/\boldsymbol{y}^H\boldsymbol{x}$

  For a sufficiently small perturbation $\boldsymbol{\delta A}$, the change the eigenvalue λ is perturbed to an eigenvalue $\hat{\lambda}$ of $\hat{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{\delta A}$,

$$\hat{\lambda} - \lambda = \boldsymbol{y}^H \boldsymbol{\delta A}\boldsymbol{x}/\boldsymbol{y}^H\boldsymbol{x} \leq \frac{||\boldsymbol{\delta A}||}{|\boldsymbol{y}^H\boldsymbol{x}|}$$

- Compare the accuracy of a Rayleigh quotient eigenvalue approximation given a normalized perturbed eigenvector (e.g. iterative guess) $\hat{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{\delta x}$, and an estimate that used both eigenvectors (also $\hat{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{\delta y}$),

$$|\hat{\lambda}_{xAx} - \lambda| \approx \boldsymbol{\delta x}^H\boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}\boldsymbol{A}\boldsymbol{\delta x} \leq \lambda||\boldsymbol{\delta x}|| + \Big(\lambda(\boldsymbol{y}^H\boldsymbol{x}) + (1 - \boldsymbol{y}^H\boldsymbol{x})||\boldsymbol{A}||\Big)||\boldsymbol{\delta x}||$$

$$|\hat{\lambda}_{yAx} - \lambda| \approx \boldsymbol{\delta y}^H\boldsymbol{A}\boldsymbol{x} + \boldsymbol{y}^H\boldsymbol{A}\boldsymbol{\delta x} \leq \lambda(||\boldsymbol{\delta x}|| + ||\boldsymbol{\delta y}||)$$

Orthogonal Iteration via QR Iteration

- In orthogonal iteration $\hat{\boldsymbol{Q}}_{i+1}\hat{\boldsymbol{R}}_{i+1} = \boldsymbol{A}\hat{\boldsymbol{Q}}_i$, QR iteration computes $\boldsymbol{A}_{i+1} = \boldsymbol{R}_i\boldsymbol{Q}_i = \hat{\boldsymbol{Q}}_{i+1}^T\boldsymbol{A}\hat{\boldsymbol{Q}}_{i+1}$ at iteration i:
  - *Using induction, assume* $\boldsymbol{A}_i = \hat{\boldsymbol{Q}}_i^T\boldsymbol{A}\hat{\boldsymbol{Q}}_i$
  - *QR iteration finds* $\boldsymbol{Q}_i\boldsymbol{R}_i = \boldsymbol{A}_i$
  - *Orthogonal iteration computes*

$$\hat{\boldsymbol{Q}}_{i+1}\hat{\boldsymbol{R}}_{i+1} = \boldsymbol{A}\hat{\boldsymbol{Q}}_i = \hat{\boldsymbol{Q}}_i\boldsymbol{A}_i = \hat{\boldsymbol{Q}}_i\boldsymbol{Q}_i\boldsymbol{R}_i$$

    and so then we have $\boldsymbol{A}_{i+1} = \hat{\boldsymbol{Q}}_{i+1}^T\boldsymbol{A}\hat{\boldsymbol{Q}}_{i+1} = \hat{\boldsymbol{R}}_{i+1}\hat{\boldsymbol{Q}}_i^T\hat{\boldsymbol{Q}}_{i+1} = \boldsymbol{R}_i\boldsymbol{Q}_i$

QR Iteration with Shift

- Describe QR iteration with shifting

$$\begin{aligned} \boldsymbol{Q}_i \boldsymbol{R}_i &= \boldsymbol{A}_i - \sigma_i \boldsymbol{I} \\ \boldsymbol{A}_{i+1} &= \boldsymbol{R}_i \boldsymbol{Q}_i + \sigma_i \boldsymbol{I} \end{aligned}$$

note that $\boldsymbol{A}_{i+1}$ is similar to $\boldsymbol{A}_i$

- The shift is typically selected to accelerate convergence with respect to a particular eigenvalue:

  We can select the shift as the bottom right element of $\boldsymbol{A}_i$, which would be the smallest eigenvalue if $\boldsymbol{A}_i$ is triangular (we have converged). Such shifting should accelerate convergence of the last column of $\boldsymbol{A}_i$, once finished we should operate only on the first $n-1$, and so on.

Hessenberg and Tridiagonal Form

- Describe reduction to Hessenberg form

  QR provides us with a way to reduce a matrix to triangular form by an orthogonal transformation. For the eigenvalue problem, we wish to perform similarity transformations, which need to be applied from both sides. Note that if simply compute the QR factorization of the first panel of $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2]$*,* $\boldsymbol{A}_1 = \boldsymbol{Q}\boldsymbol{R}$*, then* $\boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{Q}$ *is still generally dense* $\boldsymbol{Q}^T\boldsymbol{A}$ *has at least as many zeros as in* $\boldsymbol{R}$*. However, we can use QR factorization to introduce zeros by similarity transformation via*

$$\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{Q}\boldsymbol{R} & \boldsymbol{A}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \\ & \boldsymbol{Q} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12}\boldsymbol{Q} \\ \boldsymbol{R} & \boldsymbol{Q}^T\boldsymbol{A}_{22}\boldsymbol{Q} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \\ & \boldsymbol{Q}^T \end{bmatrix}$$

  if we pick $\boldsymbol{A}_{11}$ *to be* 1×1 *then doing this for each panel will yield an upper-Hessenberg result.*

- Describe reduction to tridiagonal form in symmetric case

  In the symmetric case $\boldsymbol{A}_{12}\boldsymbol{Q} = \boldsymbol{A}_{21}^T\boldsymbol{Q} = (\boldsymbol{Q}^T\boldsymbol{A}_{21})^T = \boldsymbol{R}^T$*, so we reduce to a banded (tridiagonal if* $\boldsymbol{A}_{11}$ *is* 1×1*) matrix.*

QR Iteration Complexity

- Compare complexity of QR iteration for various matrices

  Reduction to upper-Hessenberg or tridiagonal in the symmetric case, cost $O(n^3)$ *operations and can be done in a similar style to Householder QR.* *Given an upper-Hessenberg matrix,*

  - *reduction to upper-triangular requires* $n-1$ *Givens rotations,*
  - *computation of* $\boldsymbol{RQ}$ *can be done by application of the* $n-1$ *Givens rotations to* $\boldsymbol{R}$ *from the right*

  both steps cost $O(n^2)$ *computation, for* $O(n^3)$ *overall assuming QR iteration converges in* $O(n)$ *steps.*

  Given a tridiagonal matrix the same two general steps are required, but now each step costs $O(n)$*, so overall the eigenvalues and eigenvectors of a tridiagonal matrix can be computed with* $O(n^2)$ *work.*