CS 450: Numerical Anlaysis[1]

Eigenvalue Problems

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).*

Eigenvalues and Eigenvectors

- A matrix $\boldsymbol{A}$ has eigenvector-eigenvalue pair (eigenpair) $(\lambda, \boldsymbol{x})$ if

- Each $n \times n$ matrix has up to n eigenvalues, which are either real or complex

Eigenvalue Decomposition

- If a matrix $\boldsymbol{A}$ is diagonalizable, it has an *eigenvalue decomposition*

- $\boldsymbol{A}$ and $\boldsymbol{B}$ are *similar*, if there exist $\boldsymbol{Z}$ such that $\boldsymbol{A} = \boldsymbol{Z}\boldsymbol{B}\boldsymbol{Z}^{-1}$

Similarity of Matrices

Invertible similarity transformations $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{-1}$

matrix ($\boldsymbol{A}$)	*reduced form* ($\boldsymbol{Y}$)
arbitrary	
diagonalizable	

Unitary similarity transformations $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{A}\boldsymbol{U}^{H}$

matrix ($\boldsymbol{A}$)	*reduced form* ($\boldsymbol{Y}$)
arbitrary	
normal	
Hermitian	

Orthogonal similarity transformations $\boldsymbol{Y} = \boldsymbol{Q}\boldsymbol{A}\boldsymbol{Q}^{T}$

matrix ($\boldsymbol{A}$)	*reduced form* ($\boldsymbol{Y}$)
real	
real symmetric	
SPD	

Canonical Forms

- Any matrix is *similar* to a bidiagonal matrix, giving its *Jordan form*:

- Any diagonalizable matrix is *unitarily similar* to a triangular matrix, giving its *Schur form*:

- Real matrices are *orthogonally similar* to a block-triangular real matrix with 1×1 or 2×2 blocks (real Schur form)

Eigenvectors from Schur Form

- Given the eigenvectors of one matrix, we seek those of a similar matrix:

- Its easy to obtain eigenvectors of triangular matrix T:

Rayleigh Quotient

- For any vector x, the *Rayleigh quotient* provides an estimate for some eigenvalue of A:

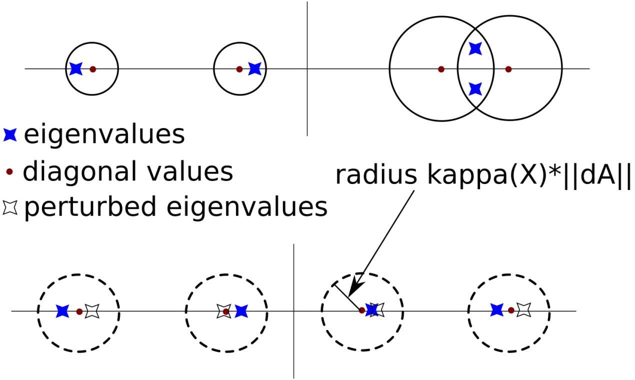
Perturbation Analysis of Eigenvalue Problems

- For non-defective $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{-1}$, the eigenvalues of $\boldsymbol{A} + \boldsymbol{\delta A} = \hat{\boldsymbol{X}}(\boldsymbol{D} + \boldsymbol{\delta D})\boldsymbol{X}^{-1}$ satisfy $\|\boldsymbol{\delta D}\| \leq \kappa(X)\|\boldsymbol{\delta A}\|$:

- Gershgorin's theorem allows us to bound the effect of the perturbation on the eigenvalues of a (diagonal) matrix:
  Given a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, *let* $r_i = \sum_{j\neq i} |a_{ij}|$, *define the Gershgorin disks as*

$$D_i = \{z \in \mathbb{C} : |z - a_{ii}| \leq r_i\}.$$

Gershgorin Theorem Perturbation Visualization



- Top corresponds to Gershgorin disks on complex plane of 4-by-4 real matrix.
- Bottom part corresponds to bounds on Gershgorin disks of $\boldsymbol{X}^{-1}(\boldsymbol{A}+\boldsymbol{\delta A})\boldsymbol{X}$, which contain the eigenvalues $\boldsymbol{D}$ of $\boldsymbol{A}$ and the perturbed eigenvalues $\boldsymbol{D}+\boldsymbol{\delta D}$ of $\boldsymbol{A}+\boldsymbol{\delta A}$ provided that $||\boldsymbol{\delta A}||$ is sufficiently small.

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 $\boldsymbol{y}$, $\lambda = \boldsymbol{y}^H\boldsymbol{A}\boldsymbol{x}/\boldsymbol{y}^H\boldsymbol{x}$

- A more accurate eigenvalue approximation than Rayleigh quotient for a normalized perturbed eigenvector (e.g., iterative guess) $\hat{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{\delta x}$, can be obtained with an estimate of both eigenvectors (also $\hat{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{\delta y}$),

Google's PageRank

A well-known application of eigenproblems is the problem of ranking n web-pages

Power Iteration

- *Power iteration* can be used to compute the largest eigenvalue of a real symmetric matrix $\boldsymbol{A}$:

- The error of power iteration decreases at each step by the ratio of the largest eigenvalues:

Rates of Convergence

- If the error at the kth step with respect to the desired solution is $\boldsymbol{e}_k$, rth order convergence implies that $\lim_{k\to\infty} \|\boldsymbol{e}_k\|/\|\boldsymbol{e}_{k-1}\|^r \leq C$

Inverse and Rayleigh Quotient Iteration

- *Inverse iteration* uses LU/QR/SVD of $\boldsymbol{A}$ to run power iteration on $\boldsymbol{A}^{-1}$

- *Rayleigh quotient iteration* provides rapid convergence to an eigenpair

Deflation

- Power, inverse, and Rayleigh-quotient iteration compute a single eigenpair, to obtain further eigenpairs, can perform *deflation*

Direct Matrix Reductions

- We can always compute an orthogonal similarity transformation to reduce a general matrix to *upper-Hessenberg* (upper-triangular plus the first subdiagonal) matrix $\boldsymbol{H}$, i.e. $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{H}\boldsymbol{Q}^T$:

- In the symmetric case, Hessenberg form implies tridiagonal:

Simultaneous and Orthogonal Iteration

- *Simultaneous iteration* provides the main idea for computing many eigenvectors at once:

- Orthogonal iteration performs QR at each step to ensure stability

Orthogonal Iteration Convergence

- If $\boldsymbol{A}$ has distinct eigenvalues and $\boldsymbol{R}_i$ has positive decreasing diagonal, the jth column of $\boldsymbol{Q}_i$ converges to the jth Schur vector of $\boldsymbol{A}$ linearly with rate $\max(|\lambda_{j+1}/\lambda_j|, |\lambda_j/\lambda_{j-1}|)$.

QR Iteration

- QR iteration reformulates orthogonal iteration for $n = k$ to reduce cost/step,
- If orthogonal iteration starts with $\hat{\boldsymbol{Q}}_1 = \boldsymbol{Q}_0$, then $\hat{\boldsymbol{Q}}_i = \prod_{j=0}^{i-1} \boldsymbol{Q}_j$,
- QR iteration converges to triangular $\boldsymbol{A}_i$ if the eigenvalues are distinct in modulus, and in general converges to block-triangular form with a block for each set of eigenvalues of equal modulus.

QR Iteration with Shift

- QR iteration can be accelerated using shifting:

- The shift is selected to accelerate convergence to an eigenvalue (pair):

QR Iteration Complexity

- QR iteration is accelerated by first reducing to upper-Hessenberg or tridiagonal form:

Solving Tridiagonal Symmetric Eigenproblems

A variety of methods exists for the tridiagonal eigenproblem:

Introduction to Krylov Subspace Methods

- *Krylov subspace methods* work with information contained in the $n \times k$ matrix

$$\boldsymbol{K}_k = \begin{bmatrix} \boldsymbol{x_0} & \boldsymbol{A}\boldsymbol{x_0} & \cdots & \boldsymbol{A}^{k-1}\boldsymbol{x_0} \end{bmatrix}$$

- Assuming $\boldsymbol{K}_n$ is invertible, the matrix $\boldsymbol{K}_n^{-1}\boldsymbol{A}\boldsymbol{K}_n$ is a *companion matrix* $\boldsymbol{C}$:

Krylov Subspaces

- Given $\boldsymbol{Q}_k\boldsymbol{R}_k = \boldsymbol{K}_k$, we obtain an orthonormal basis for the Krylov subspace,

$$\mathcal{K}_k(\boldsymbol{A}, \boldsymbol{x}_0) = span(\boldsymbol{Q}_k) = \{p(\boldsymbol{A})\boldsymbol{x}_0 : deg(p) < k\},$$

  where p is any polynomial of degree less than k.
- The Krylov subspace includes the $k-1$ approximate dominant eigenvectors generated by $k-1$ steps of power iteration:

Krylov Subspace Methods

- The $k \times k$ matrix $\boldsymbol{H}_k = \boldsymbol{Q}_k^T \boldsymbol{A} \boldsymbol{Q}_k$ minimizes $\|\boldsymbol{A}\boldsymbol{Q}_k - \boldsymbol{Q}_k \boldsymbol{H}_k\|_2$:

- $\boldsymbol{H}_k$ is Hessenberg, because the companion matrix $\boldsymbol{C}_k$ is Hessenberg:

Demo: Arnoldi vs Power Iteration

Rayleigh-Ritz Procedure

- The eigenvalues/eigenvectors of $\boldsymbol{H}_k$ are the *Ritz values/vectors*:

- The Ritz vectors and values are the *ideal approximations* of the actual eigenvalues and eigenvectors based on only $\boldsymbol{H}_k$ and $\boldsymbol{Q}_k$:

Arnoldi Iteration

- Arnoldi iteration computes the ith column of $\boldsymbol{H}_n$, $\boldsymbol{h}_i$ and the ith column of $\boldsymbol{Q}_n$ directly using the recurrence $\boldsymbol{A}\boldsymbol{q}_i = \boldsymbol{Q}_n\boldsymbol{h}_i = \sum_{j=1}^{i+1} h_{ji}\boldsymbol{q}_j$

Lanczos Iteration

- Lanczos iteration provides a method to reduce a symmetric matrix to a tridiagonal matrix:

- After each matrix-vector product, it suffices to orthogonalize with respect to two previous vectors:

Cost Krylov Subspace Methods

- The cost of matrix-vector multiplication when the matrix has m nonzeros

- The cost of orthogonalization at the kth iteration of a Krylov subspace method is

Restarting Krylov Subspace Methods

- In finite precision, Lanczos generally loses orthogonality, while orthogonalization in Arnoldi can become prohibitively expensive:

- Consequently, in practice, low-dimensional Krylov subspace methods are constructed repeatedly using carefully selected new starting vectors:

Generalized Eigenvalue Problem

- A generalized eigenvalue problem has the form $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{B}\boldsymbol{x}$,

- When $\boldsymbol{A}$ and $\boldsymbol{B}$ are symmetric and $\boldsymbol{B}$ is SPD, we can perform Cholesky on $\boldsymbol{B}$, multiply $\boldsymbol{A}$ by the inverted factors, and diagonalize it:

- Specialized canonical forms and methods exist for the generalized eigenproblem with fewer constraints on $\boldsymbol{B}$ and better cost/stability.