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

matrix	*similarity*	*reduced form*
SPD		
real symmetric		
Hermitian		
normal		
real		
diagonalizable		
arbitrary		

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*:

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 $\boldsymbol{T}$:

Rayleigh Quotient

- For any vector $\boldsymbol{x}$, the *Rayleigh quotient* provides an estimate (lower-bound) on some eigenvalue λ of $\boldsymbol{A}$:

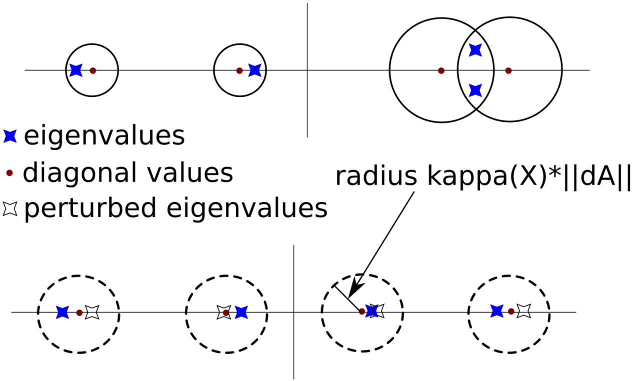
Perturbation Analysis of Eigenvalue Problems

- Suppose we seek eigenvalues $\boldsymbol{D} = \boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}$, but find those of a slightly perturbed matrix $\boldsymbol{D} + \boldsymbol{\delta D} = \hat{\boldsymbol{X}}^{-1}(\boldsymbol{A} + \boldsymbol{\delta A})\hat{\boldsymbol{X}}$:

- 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 obtain with an estimate of both eigenvectors (also $\hat{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{\delta y}$),

Power Iteration

- *Power iteration* can be used to compute the largest eigenvalue of a real symmetric matrix A:

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

Inverse and Rayleigh Quotient Iteration

Activity: *Inverse Iteration with a Shift*
Activity: *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

Review: eigenvalue problems

- sensitivity/conditioning

  $\kappa(X)$ ← eigenvector matrix

  (λ, x, y), $\langle x, y \rangle$

  (x ↑ left, y ↑ right; $\langle x, y \rangle$ ↑ angle between eigenvectors)

- canonical forms/similarity

  Schur decomposition

  $$A = Q^H U Q$$

  ▽ ← diagonal contains eigenvalues

- iterative algorithms for a particular eigenvalue
  - power iteration $O(n^2)$/iteration
  - inverse iteration $O(n^3)$ start-up, $O(n^2)$/iteration
  - Rayleigh-quotient iteration, $O(n^3)$/iteration

Deflation

- Power, inverse, and Rayleigh-quotient iteration compute a single eigenpair, to obtain further eigenpairs, can perform *deflation* (λ_1, x_1)

$B = A - \lambda_1 x_1 v^H$

choose $v = x$

$B = A - \lambda_1 x_1 x_1^H$

$\text{eigs}(A) = \lambda_1, \lambda_2, \dots \lambda_n$

$\text{eigs}(B) = 0, \lambda_2, \dots \lambda_n$

if A is symmetric, $A = \sum_{i=1}^{n} \lambda_i x_i x_i^H$, so $B = \sum_{i=2}^{n} \lambda_i x_i x_i^H$

eigenvectors $x_2 \dots x_n$ would be preserved

A is nonsymmetric, $A = \underbrace{[\,]}_{Q} \underbrace{\begin{bmatrix} \lambda_1 & \cdots \\ & \ddots & \vdots \\ & & \lambda_n \end{bmatrix}}_{T} \underbrace{[\quad]}_{Q^H}$ ← Schur decomposition

$q_1 = Qe_1$ is an eigenvector of A, moreover $q_1 = x_1$

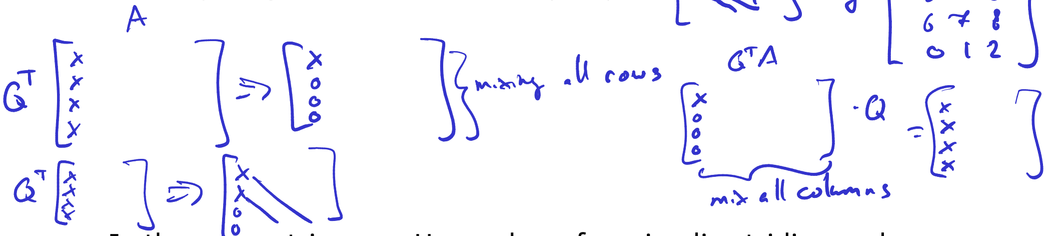
$\underbrace{Q \overbrace{T \underbrace{Q^H Q e_1}_{e_1}}^{\lambda e_1}}_{\lambda_1 Q e_1}$

$B = A - \lambda_1 q_1 q_1^H = QTQ^H - \lambda_1 q_1 q_1^H = Q(T - \lambda_1 \underbrace{Q^H q_1}_{e_1} \underbrace{q_1^H Q}_{e_1^T}) Q^H$

$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \ddots & \\ 0 & & & 0 \end{bmatrix}$

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. $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{H}\boldsymbol{Q}^T$:

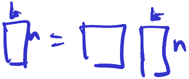


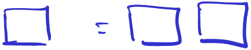
- In the symmetric case, Hessenberg form implies tridiagonal:

Simultaneous and Orthogonal Iteration

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

$X^{(i+1)} = AX^{(i)}$ or





span($X^{(i)}$) will converge to the span of k largest eigenvectors of A

- Orthogonal iteration performs QR at each step to ensure stability

$X^{(i+1)} = AQ^{(i)}$

$Q^{(i+1)}R^{(i+1)} = X^{(i+1)}$

$\lim_{i\to\infty}$ span($Q^{(i)}$) is span of k largest eigenvectors

if we perform orth. iteration of $A^T A$

$X^{(i+1)} = A^T(AQ^{(i)})$

QR Iteration

- QR iteration reformulates orthogonal iteration for $n = k$ to reduce cost/step,

$A_0 = A$

$Q_i R_i = A_i$

$A_{i+1} = R_i Q_i$

Orthogonal iteration

$\hat{X}_i := A \hat{Q}_i$

$[\hat{Q}_{i+1}, \hat{R}_{i+1}] = QR(\hat{X}_i)$

- Using induction, we assume $\boldsymbol{A}_i = \hat{\boldsymbol{Q}}_i^T \boldsymbol{A} \hat{\boldsymbol{Q}}_i$ and show that QR iteration obtains $\boldsymbol{A}_{i+1} = \hat{\boldsymbol{Q}}_{i+1}^T \boldsymbol{A} \hat{\boldsymbol{Q}}_{i+1}$

$\hat{Q}_i$ is Q computed by orthogonal iteration

$\hat{Q}_{i+1} = Q_i \hat{Q}_i$ (QR iteration)

$A_{i+1} = R_i Q_i = Q_i^T A_i Q_i = \underbrace{Q_i^T \hat{Q}_i^T}_{\hat{Q}_{i+1}^T} A \underbrace{\hat{Q}_i Q_i}_{\hat{Q}_{i+1}}$

QR Iteration with Shift

- QR iteration can be accelerated using shifting:

$$Q_i R_i = A_i - \sigma I$$

shift

$$A_{i+1} = R_i Q_i + \sigma I$$

A_{i+1} is similar to A_i

$R_i Q_i = Q_i^T (A_i - \sigma I) Q_i$

$Q_i (A_{i+1} - \sigma I) Q_i^T = Q_i R_i$

$R_i Q_i$ is similar to $Q_i R_i$

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

$\sigma = a_{nn}^{(i)}$ so last entry of A

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:

- QR iteration
- Divide and conquer

Solving the Secular Equation for Divide and Conquer

To solve the eigenproblem at each step, the divide and conquer method needs to diagonalize a rank-1 perturbation of a diagonal matrix

$$\boldsymbol{A} = \boldsymbol{D} + \alpha \boldsymbol{u}\boldsymbol{u}^T.$$

Introduction to Krylov Subspace Methods

- Define k-dimensional Krylov subspace matrix

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

- Show that $\boldsymbol{K}_n^{-1}\boldsymbol{A}\boldsymbol{K}_n$ is a companion matrix $\boldsymbol{C}$:

Krylov Subspaces

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

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

- Consider whether $k - 1$ steps of power iteration starting from $\boldsymbol{x_0}$ lead to an approximation in the Krylov subspace, also consider QR (subspace) iteration:

Krylov Subspace Methods

- Given $\boldsymbol{Q}\boldsymbol{R} = \boldsymbol{K}_k$, we obtain an orthonormal basis for the Krylov subspace and $\boldsymbol{H}_k = \boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{Q}$ which minimizes $||\boldsymbol{A}\boldsymbol{Q} - \boldsymbol{Q}\boldsymbol{H}||_2$:

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

Rayleigh-Ritz Procedure

Demo: *Arnoldi vs Power Iteration*

- 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 $\boldsymbol{Q}$:

Arnoldi Iteration

- Arnoldi iteration computes $\boldsymbol{H}$ directly using the recurrence $\boldsymbol{q}_i^T \boldsymbol{A} \boldsymbol{q}_j = h_{ij}$:

- After each matrix-vector product, orthogonalization is done with respect to each previous vector:

Lanczos Iteration

- Lanczos iteration provides a method to reduce a symmetric matrix to 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:
- Alternative canonical forms and methods exist that are specialized to the generalized eigenproblem.

Nonlinear Eigenvalue Problem

- In a polynomial eigenvalue problem, we seek solutions $\lambda, \boldsymbol{x}$ to

$$\sum_{i=0}^{d} \lambda^i \boldsymbol{A}_i \boldsymbol{x} = \boldsymbol{0}$$

- Assuming for simplicity that $\boldsymbol{A}_d = \boldsymbol{I}$, solutions are given by solving the matrix eigenvalue problem with the block-companion matrix

$$\begin{bmatrix} -\boldsymbol{A}_{d-1} & \cdots & -\boldsymbol{A}_0 \\ \boldsymbol{I} & \boldsymbol{0} & \cdots \\ & \ddots & \ddots \end{bmatrix}$$