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 T:

Rayleigh Quotient

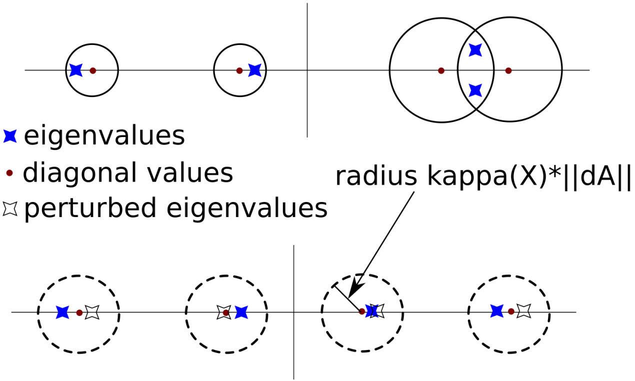
- For any vector $\boldsymbol{x}$, the *Rayleigh quotient* provides an estimate for some eigenvalue of $\boldsymbol{A}$:

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 A to run power iteration on 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

Demo: *Orthogonal Iteration*
Activity: *Orthogonal Iteration*

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

- Orthogonal iteration performs QR at each step to ensure stability

QR Iteration

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

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

QR Iteration with Shift

- QR iteration can be accelerated using shifting:

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

Eigenvalue Algorithms Review

Extremal (min/max) eigenvalue – power iteration
inverse iteration
Rayleigh-Quotient iter.

Full EVD – Orthogonal Iteration
QR Iteration (shifting)

reduction
- upper-Hessenberg
- tridiagonal for symmetric

QR Iteration Complexity

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

$A_i = Q_i R_i$

$A_{i+1} = R_i Q_i$

$A_0 = Q^T A Q$ → can obtain by Householder-like procedure

→ upper-Hessenberg

→ tridiagonal

---

if A_0 is upper-Hessenberg, so Q_i is upper-Hessenberg
$R_i Q_i$ is also

n-1 Givens rotations to compute
$O(n^2)$ cost
each rotation acts on two rows

.. A_0 is tridiagonal
$R_i Q_i$ is also, each Givens rotation has cost $O(1)$

Solving Tridiagonal Symmetric Eigenproblems

A variety of methods exists for the tridiagonal eigenproblem:

- QR iteration $O(n^2)$ for eigenvalues, $O(n^3)$ eigenvectors — bidiagonal singular value problem
- Divide and conquer

$$T = \begin{bmatrix} T_1 & r \\ r & T_2 \end{bmatrix} = \begin{bmatrix} T_1 & r e_{n/2} e_1^T \\ r e_1 e_{n/2}^T & T_2 \end{bmatrix} = \begin{bmatrix} \hat{T}_1 & \\ & \hat{T}_2 \end{bmatrix} + r \begin{bmatrix} e_{n/2} \\ e_1 \end{bmatrix} \begin{bmatrix} e_1^T & e_{n/2}^T \end{bmatrix}$$

$$r = t_{n/2+1,\, n/2} = t_{n/2,\, n/2+1}$$

$$T = \begin{bmatrix} \underbrace{Q_1 D_1 Q_1^T}_{T_1} & \\ & \underbrace{Q_2 D_2 Q_2^T}_{T_2 \text{ (recursion)}} \end{bmatrix} + r \, [\;] = \begin{bmatrix} Q_1 & \\ & Q_2 \end{bmatrix} \left(\overbrace{\begin{bmatrix} D_1 & \\ & D_2 \end{bmatrix}}^{D} + r \overbrace{\begin{bmatrix} Q_1^T e_{n/2} \\ Q_2^T e_1 \end{bmatrix}}^{u} \cdot \overbrace{\begin{bmatrix} Q_1^T e_{n/2} \\ Q_2^T e_1 \end{bmatrix}^T}^{u^T} \right) \cdot \begin{bmatrix} Q_1^T & \\ & Q_2^T \end{bmatrix}$$

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

$$A = D + \alpha u u^T.$$

$$0 = \underbrace{\det(A - \lambda I)}_{F(\lambda)}$$

(λ ↑ eigenvalues)

looking for $F(\lambda) = 0$

$$\det(A - \lambda I) = 1 + \alpha u^T (D - \lambda I)^{-1} u$$

Introduction to Krylov Subspace Methods

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

$$\boldsymbol{K}_k = [\boldsymbol{x_0} \quad \boldsymbol{A}\boldsymbol{x_0} \quad \cdots \quad \boldsymbol{A}^{k-1}\boldsymbol{x_0}]$$

mat-vec

$A^2 x_0$

starting guess

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

$K_n = [k_n^{(0)} \cdots k_n^{(n-1)}]$

$C = K_n^{-1} A K_n = K_n^{-1}[Ax_0 \; A^2x_0 \cdots A^k x_0] = K_n^{-1}[k_n^{(1)}, k_n^{(2)}, \cdots k_n^{(n-1)}, Ak_n^{(n-1)}]$

$$= \begin{bmatrix} 0 & & & \\ 1 & 0 & & Ak_n^{(n-1)} \\ & 1 & \ddots & \\ & & 1 & \end{bmatrix}$$

Krylov Subspaces

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

  QR factorization of K_k — 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:

  power iteration computes $K_n^{(i)}$ at ith step

Krylov Subspace Methods

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

$$Q_k X = A Q_k$$

$$\underbrace{Q_k^T Q_k}_{I} X = \underbrace{Q_k^T A Q_k}_{H_k}$$

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

triangular, triangular

$$C = K_n^{-1} A K_n = C = R_n^{-1} \underbrace{Q_n^T A Q_n}_{H_n} R_n$$

$$H_n = \begin{bmatrix} H_k & | & | & | & | \end{bmatrix}$$

$$H_k = R_n C R_n^{-1}$$

Rayleigh-Ritz Procedure

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

$H_k = XDX^{-1}$ | if A is symmetric

Ritz vectors $= Q_k X$ | $H_k = Q_k^T A Q_k$ is tridiagonal

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

If A is SPD, then $\lambda_{max}(A) = \max_{x \neq 0} \frac{x^T A x}{x^T x}$

$$\max_{x \in span(Q_k)} \frac{x^T A x}{x^T x} = \max_{y \in \mathbb{R}^k} \frac{y^T \overbrace{Q_k^T A Q_k}^{H_k} y}{y^T \underbrace{Q_k^T Q_k}_{I} y} = \max_{y \in \mathbb{R}^k} \frac{y^T H_k y}{y^T y} = \lambda_{max}(H_k)$$

$x = Q_k y$

Arnoldi Iteration

- Arnoldi iteration computes $\boldsymbol{H} = \boldsymbol{H}_n$ directly using the recurrence $\boldsymbol{q}_i^T \boldsymbol{A} \boldsymbol{q}_j = h_{ij}$, where $\boldsymbol{q}_l$ is the lth column of $\boldsymbol{Q}_n$:

$$A = Q_n H_n Q_n^T$$

$$q_i^T A q_j = \underbrace{q_i^T Q_n}_{e_i^T} H_n \underbrace{Q_n^T q_j}_{e_j} = h_{ij}$$

Arnoldi

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

if $H_n = T_n$ is tridiagonal

$$h_{ij} = \begin{cases} |i-j|>1 : & 0 \\ |i-j|\le 1 : & q_i^T A q_j \end{cases}$$

Lanczos

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:

- Alternative canonical forms and methods exist that are specialized to the generalized eigenproblem.