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

eigs A^k if D are eigs of A? D^k

$$A = XDX^{-1}, \quad A^2 = XD\underbrace{X^{-1}X}_{I}DX^{-1} = XD^2X^{-1}$$

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

if A is real, then if $a+bi$ is an eig. of A, then $a-bi$ is also

Eigenvalue Decomposition

- If a matrix A is diagonalizable, it has an *eigenvalue decomposition*

$$A = XDX^{-1} \qquad X = [x_1, \ldots, x_n]$$

$$Ax_i = XDX^{-1}x_i = XDe_i = X\lambda_i e_i = \lambda_i x_i$$

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

invertible, so Z^{-1} exists

unitary: $Z^H = Z^{-1}$

orthogonal: $Z^T = Z^{-1}$

Similarity of Matrices

matrix	similarity	reduced form
real SPD	orthogonal	real positive diagonal
real symmetric	orthogonal	tridiagonal diagonal } real
$A = A^H$ Hermitian	unitary	real diagonal
$AA^H = A^H A$ normal	unitary	diagonal
real	orthogonally	Hessenberg (real) 
diagonalizable	invertible	diagonal
arbitrary	unitarily invertible	triangular (Schur) bidiagonal (Jordan)

Canonical Forms

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

$$A = Z \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{bmatrix} Z^{-1}, \quad \text{each } J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}$$

unique modulo permutation of J_i's

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

$$A = Q \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}}_{T} Q^H \qquad \det(T - t_{ii} I) = 0$$

eigenvalue of T

columns of Q - Schur vectors

t_{ii}

Computing Eigenvectors of Matrices in Schur Form

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

A is similar to B $\Rightarrow$ $A = ZBZ^{-1}$

$B = XDX^{-1}$, $A = ZXDX^{-1}Z^{-1} = YDY^{-1}$

- Its easy to obtain eigenvectors of triangular matrix $\boldsymbol{T}$:

$Te_1 = \lambda_1 e_1$ ($\lambda_1 = t_{11}$), e_1 is an eigenvector

$$(T - \lambda_i I)x = 0$$

$$\begin{bmatrix} U_{11} & u & T_{13} \\ & 0 & v^T \\ & & U_{22} \end{bmatrix} \begin{bmatrix} U_{11}^{-1}u \\ -1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

need to solve $U_{11} v = -u \Rightarrow \begin{bmatrix} v \\ -1 \\ 0 \\ \vdots \end{bmatrix} = \begin{bmatrix} U_{11}U_{11}^{-1}u - u \\ 0 \end{bmatrix}$

Rayleigh Quotient

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

$$\rho_A(x) = \frac{x^H A x}{x^H x}, \quad \text{sol'n to} \quad x\alpha \cong Ax$$

?

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 $\boldsymbol{D} + \boldsymbol{\delta D} = \hat{\boldsymbol{X}}^{-1}(\boldsymbol{A} + \boldsymbol{\delta A})\hat{\boldsymbol{X}}$:

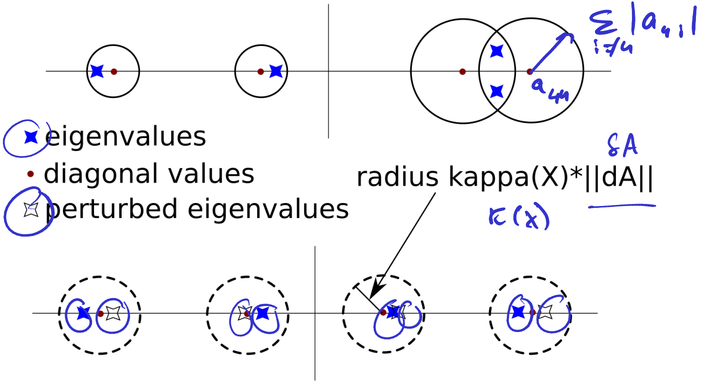
$$B = X^{-1}(A + \delta A)X = D + \underbrace{X^{-1}\delta A X}_{\delta B}$$

$$\|\delta B\| \le \|X^{-1}\|\,\|\delta A\|\,\|X\| = \kappa(X)\|\delta A\|$$

- 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*

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

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

$$\frac{y^H A x}{y^H x}$$

$$y^H(Ax) = y^H \lambda x$$

$$(y^H A)x = y^H \lambda x$$

$$\left| \frac{y^H(A+\delta A)x}{y^H x} - \lambda \right| = \frac{|y^H \delta A x|}{|y^H x|} \le \frac{\|\delta A\| \ldots}{|y^H x|}$$

- 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 $\hat{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{\delta y}$),

$$\underbrace{\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}}_{X} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \underbrace{\begin{bmatrix} y_1^H \\ \vdots \\ y_n^H \end{bmatrix}}_{X^{-1}}$$

$$\left| \frac{(y+\delta y)^H A (x+\delta x)}{y^H x} - \lambda \right| = |\lambda| \frac{\|\delta y\| + \|\delta x\| \ldots}{|y^H x|}$$

[illegible]

Power Iteration

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

$$x^{(i+1)} = A x^{(i)}$$

$$x^{(i+1)} = x^{(i+1)} / \| x^{(i+1)} \|$$

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

$$A = U D V^H \Rightarrow A^k = U D^k V^H$$

$$x^{(k)} = A^k x^{(0)} = U D^k V^H x^{(0)} = \sum_{i=1}^{n} u_i \lambda_i^k \underbrace{v_i^H x^{(0)}}_{\alpha^{(i,k)}} \quad \langle v_i, x^{(0)} \rangle$$

$$\alpha^{(i,k)} = \lambda_i^k \alpha^{(i,0)}$$

$$\left| \frac{\alpha^{(i,k)}}{\alpha^{(j,k)}} \right| = \left| \frac{\lambda_i}{\lambda_j} \right|^k \cdot C$$

Inverse Iteration 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

(λ_1, x_1)

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

$$B = A - \lambda_1 x_1 v^H = \left(\sum_{i=1}^{n} \lambda_i x_i y_i^H\right) - \lambda_1 x_1 v^H$$

for symmetric, pick $v^H = x_1^H$, preserves eigenvectors

for nonsymmetric, $v^H = x_1^H$, preserve Schur vectors

Schur decomp. B $A = QTQ^H$ $\quad [\lambda_1 \ \ddots \ \lambda_n]$

$$= A - \lambda_1 x x^H = Q(T - Q^H \lambda_1 x x^H Q)Q^H$$

$\lambda_1 e_1 e_1^T$

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

QR Iteration

- QR iteration reformulates orthogonal iteration to reduce cost/step to $O(nk^2)$
- 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:

QR Iteration Complexity

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

Solving Tridiagonal Symmetric Eigenproblems

A rich variety of methods exists for the tridiagonal eigenproblem:

- QR iteration
- Divide and conquer

Solving the Secular Equation

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

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

Solving Tridiagonal Symmetric Eigenproblems (II)

- Jacobi iteration
- Bisection
- Relatively robust representation (RRR and MRRR)

Introduction to Krylov Subspace Methods

- Define k-dimensional Krylov subspace matrix

$$\boldsymbol{K}_k = \begin{bmatrix} \boldsymbol{x}_{\mathbf{0}} & \boldsymbol{A}\boldsymbol{x}_{\mathbf{0}} & \cdots & \boldsymbol{A}^{k-1}\boldsymbol{x}_{\mathbf{0}} \end{bmatrix}$$

- 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

- 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:

Convergence of Lanczos Iteration

- Cauchy interlacing theorem: eigenvalues of $\boldsymbol{H}_k$, $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$ with respect to eigenvalues of $\boldsymbol{A}$, $\lambda_1 \geq \cdots \geq \lambda_n$ satisfy

$$\lambda_i \leq \tilde{\lambda}_i \leq \lambda_{n-k+i}$$

- Convergence to extremal eigenvalues is generally fastest

Applications of Eigenvalue Problems: Matrix Functions

- Given $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{-1}$ how can we compute $\boldsymbol{A}^k$?

- What about $e^{\boldsymbol{A}}$? $\log(\boldsymbol{A})$? generally $f(\boldsymbol{A})$?

Applications of Eigenvalue Problems: Differential Equations

- Consider solutions to an ordinary differential equation of the form $\frac{d\boldsymbol{x}}{dt}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{f}(t)$ with $\boldsymbol{x}(0) = \boldsymbol{x_0}$:

$$\boldsymbol{x}(t) = e^{t\boldsymbol{A}}\boldsymbol{x}_0 + \int_0^t e^{(t-\tau)\boldsymbol{A}}\boldsymbol{f}(\tau)d\tau$$

- Using $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{-1}$ permits us to compute the solution explicitly (Jordan form also suffices if $\boldsymbol{A}$ is defective):

Differential Equations using the Generalized Eigenvalue Problem

- Consider a more general linear differential equation of the form $\boldsymbol{B}\frac{d\boldsymbol{x}}{dt}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{f}(t)$ with $\boldsymbol{x}(0) = \boldsymbol{x_0}$, which we can reduce to the usual form by premultiplying with $\boldsymbol{B}^{-1}$:

- If we can find $\boldsymbol{X}$ such that $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{D_A}\boldsymbol{X}^{-1}$ and $\boldsymbol{B} = \boldsymbol{X}\boldsymbol{D_B}\boldsymbol{X}^{-1}$ we could solve this equation while preserving symmetry of $\boldsymbol{A}$ and $\boldsymbol{B}$:

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

Canonical Forms Generalized Eigenvalue Problem

- For nonsingular $\boldsymbol{U}, \boldsymbol{V}$, $\boldsymbol{A} - \lambda \boldsymbol{B} = \boldsymbol{U}(\boldsymbol{J} - \lambda \boldsymbol{I})\boldsymbol{V}^T$ where $\boldsymbol{J}$ is in Jordan form

- For some unitary $\boldsymbol{P}, \boldsymbol{Q}$, $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{T}_{\boldsymbol{A}}\boldsymbol{Q}^H$ and $\boldsymbol{B} = \boldsymbol{P}\boldsymbol{T}_{\boldsymbol{B}}\boldsymbol{Q}^H$ where $\boldsymbol{T}_{\boldsymbol{A}}$ and $\boldsymbol{T}_{\boldsymbol{B}}$ are triangular

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

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