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

right-eigenpair if $Ax = \lambda x$ | left-eigenpair $x^T A = \lambda x^T$, $A^T x = \lambda x$ | eigenvector x, αx is also: $\alpha A x = \alpha \lambda x$

if symmetric

1D eigenspace $\in \mathbb{C}^n$

$Ay = \lambda y$ then $A(x+y) = \lambda(x+y)$

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

if $a + bi$ → eigenvalue, also $a - bi$ → eigenvalue | up to n diff. eigenvalues, sum of multiplicities is $\leq n$

Eigenvalue Decomposition

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

$A = XDX^{-1}$ → X^{-1} is left eigenvectors (rows)

X: right eigenvectors

$AX = XD = [d_{11}x_1, d_{22}x_2 \dots d_{nn}x_n]$

if not diagonalizable, it is defective

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

$B = XDX^{-1}$

$A = ZXDX^{-1}Z^{-1}$ (ZX: eigenvectors)

A and B are orthogonally similar if $\exists X$, s.t. $X^{-1} = X^T$, and $XAX^T = B$

If A is normal, so $A^H A = A A^H$
if real $A^T A = A A^T$ } symmetric

then $A = X D X^T = X D X^{-1}$
orthogonally similar

$A = U \Sigma V^T$

$A A^T = U \Sigma^2 U^T$ } eigenvalue decompositions
$A^T A = V \Sigma^2 V^T$

if A is normal, then $U = V$, so $A = U \Sigma U^T$

property	similarity	matrix
symmetric	orthogonally	real diagonal
normal (real)	orthogonally	real diagonal
Hermitian	unitarily	real diagonal
$A = A^H$	$X = X^H$	

Rayleigh Quotient and Power Iteration

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

$$\rho_A(x) = \frac{x^T A x}{x^T x}$$

if x is eigenvector $\rho_A(x) = \lambda$

is not eigenvector $\rho_A(x) = \arg\min_{\alpha} \|x\alpha - y\|_2$, $y = Ax$

normal eqns are $x^T x \alpha = x^T y$

$$\alpha = \frac{x^T y}{x^T x} = \frac{x^T A x}{x^T x}$$

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

pick a random $x^{(0)}$ and iterate | linear convergence at rate λ_1/λ_2

$$x^{(i)} = A x^{(i-1)} = A(\alpha_1 y_1 + \alpha_2 y_2 + \dots) = \alpha_1 \lambda_1 y_1 + \alpha_2 \lambda_2 y_2 + \dots$$

if A is symmetric, eigvals $\lambda_1 > \lambda_2 > \dots$

eigvectors $y_1, y_2, \dots$

Inverse Iteration and Rayleigh Quotient Iteration

converge to eigval closest to σ

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

$A^{-1} = XD^{-1}X^{-1}$ $\quad \lambda_{max}(A^{-1}) = 1/\lambda_{min}(A)$

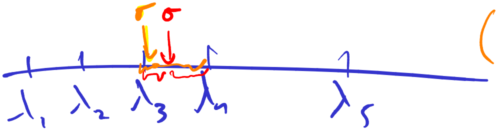
shifted inverse iteration
decompose $A - \sigma I$

$A = XDX^{-1}$

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

solve linear system at each step | given any factorization of A, solve costs $O(n^2)$

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



$(A - \rho_A(x_i) I) x_{i+1} = x_i$

solve

iterations cost $O(n^3)$, cubic convergence

$\lambda_{max}((A-\sigma I)^{-1}) = \frac{1}{\lambda_3 - \sigma}$

$\lambda_{2nd\text{-}max}((A-\sigma I)^{-1}) = \frac{1}{\lambda_4 - \sigma}$

conv rate $\frac{\lambda_{max}(\ldots)}{\lambda_{2nd\text{-}max}(\ldots)} = \frac{|\lambda_4 - \sigma|}{|\lambda_3 - \sigma|}$ (large / small)

Perturbation Analysis of Eigenvalue Problems

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

$\hat{X} = X + \delta X$

$B = \hat{X}^{-1} A \hat{X}$

$B = \tilde{D} + \delta B$

diagonal

$\delta D = X^{-1} \delta S X$

$\|\delta D\| \le \|X^{-1}\| \|\delta S\| \|X\| = \kappa(X) \|\delta S\|$

general perturbation, moreover eig$(B) = D$

- Gershgorin's theorem allows us to bound the effect of the perturbation on the eigenvalues of a (diagonal) matrix:

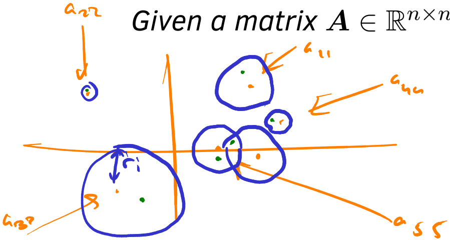
Given a matrix $A \in \mathbb{R}^{n \times n}$, *let* $r_i = \sum_{j \neq i} |a_{ij}|$, *define the Gershgorin disks as*

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

sum of absolute values of ith row of A

eigenvalue

diagonal entries



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

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

Canonical Forms

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

Computing Eigenvectors of Matrices in Schur Form

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

- Its easy to obtain eigenvectors of triangular matrix T:

Matrix Reductions

- Any matrix is orthogonally similar to an *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 is the starting point for computing many eigenvectors:

- Orthogonal iteration performs QR at each step to ensure stability

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:

QR Iteration with Shift

- Describe QR iteration with 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

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

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

- $\boldsymbol{H}_k$ 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

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

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

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