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

$$Ax = \lambda x$$

· αx is also an eigenvector

· λ, $\mathcal{X} \subseteq \mathbb{R}^n$, $x \in \mathcal{X}$, $Ax = \lambda x$

(↑ eigenspace), $\dim(\mathcal{X}) \leq$ multiplicity of λ

when '=', then the matrix is diagonalizable, otherwise, it is defective

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

even if the matrix is real

· if we have $\lambda = a + bi$

then we also have $\lambda = a - bi$

· for diagonalizable matrices, $\lambda_1 \dots \lambda_k$, $\mathcal{X}_1 \dots \mathcal{X}_k$

$$\sum_{i=1}^{k} \dim(\mathcal{X}_i) = \underbrace{\dim(A)}_{n}$$

Eigenvalue Decomposition

- If a matrix A is diagonalizable, it has an *eigenvalue decomposition*
- A and B are *similar*, if there exist Z such that $A = ZBZ^{-1}$

Similarity of Matrices

A matrix	$B = ZAZ^{-1}$ similarity ($\leftarrow Z$)	B reduced form
real symmetric positive definite SPD	orthogonal	diagonal, positive $D > 0$
real symmetric ($\lambda > 0$)	orthogonal	real tridiagonal, real diagonal
Hermitian ($A^H = A$)	unitary	real diagonal
normal ($A^H A = A A^H$)	unitary	diagonal
real	orthogonal	real Hessenberg
diagonalizable	invertible	diagonal
arbitrary	unitary invertible	triangular (Schur form) bidiagonal (Jordan normal form)

Canonical Forms

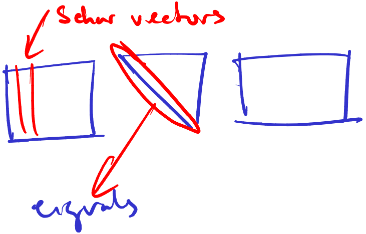
- Any matrix is *similar* to a bidiagonal matrix, giving its *Jordan form*:

$$A = X \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_k \end{bmatrix} X^{-1} \qquad J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}$$

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

$$A = Q T Q^H$$

(T: triangular; Q: unitary)



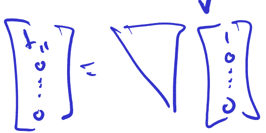
Eigenvectors from Schur Form

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

$Tx = \lambda x$ $\qquad A = QDQ^H$

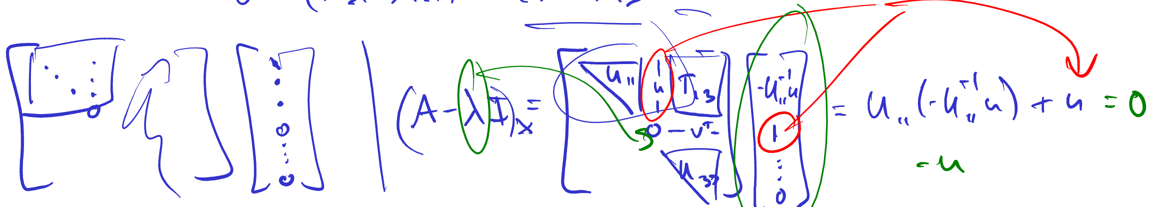
$T = XDX^{-1}$ $\qquad \text{eigvecs}(A) = QX$

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



$Ax = \lambda x$

$0 = (Ax - \lambda x) = (A - \lambda I)x$



Rayleigh Quotient

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

$\lambda = \|Av\| / \|v\|$ (a real scalar (positive))

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

$Av = \lambda v$ eigenpair (λ, v)

$\rho_A(v) = \lambda$, $\rho_A(\alpha v) = \lambda$

$\{\rho : \rho = \rho_A(x) \ \forall x \in \mathbb{R}^n\} \rightarrow$ field of values

$\leq |\sigma_{max}|$

$y = Ax$, $\rho_A(x)$ is a good estimate of the eigenvalue associated with a vector near x, since its the solution

$x \overset{?}{\alpha} \cong y$ $\quad [x][\overset{?}{\alpha}] \cong [y]$

normal eqs

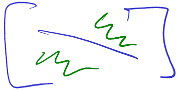
$x^T x \alpha = x^T y$ ($y = Ax$)

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

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

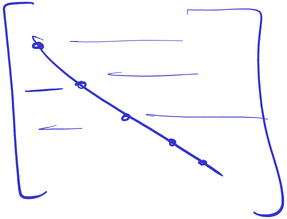
$$X^{-1}(A+\delta A)X = D + X^{-1}\delta A X =$$

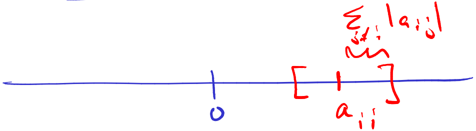


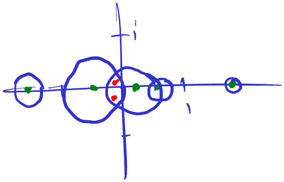
$$\|X^{-1}\delta A X\| \le \|X^{-1}\| \cdot \|\delta A\| \cdot \|X\| = \kappa(X)\|\delta A\|$$

- 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*

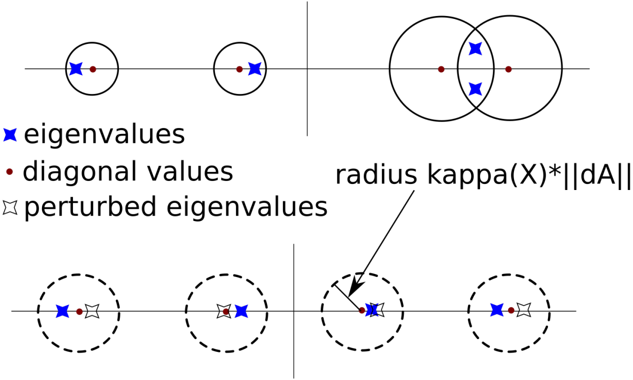
$$D_i = \{z \in \mathbb{C} : |z - a_{ii}| \le r_i\}.$$







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

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

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:

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

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

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:

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

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

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

- 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:

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

Demo: *Arnoldi Iteration*
Demo: *Arnoldi Iteration with Complex Eigenvalues*

- 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 $\boldsymbol{Q}_n$:

- 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 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.