CS 450: Numerical Analysis
Eigenvalue Problems

University of Illinois at Urbana-Champaign

¹These 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 $A$ has eigenvector-eigenvalue pair (eigenpair) $(\lambda, x)$ if
  - right-eigenpair if $A x = \lambda x$
  - left-eigenpair if $x^T A = \lambda x^T$
  - if symmetric
    - if $A$ is a symmetric matrix
      - $A^T A = \lambda x x^T$

- Each $n \times n$ matrix has up to $n$ eigenvalues, which are either real or complex
  - if $a + bi$ is an eigenvalue
    - also $a - bi$ is an eigenvalue
    - sum of multiplicities is $\leq n$
Eigenvalue Decomposition

- If a matrix $A$ is diagonalizable, it has an eigenvalue decomposition $A = XDX^{-1}$ where $X^{-1}$ is left eigenvectors (rows).

$AX = XD = \begin{bmatrix} d_1x_1, & d_2x_2, & \ldots & d_nx_n \end{bmatrix}$

If not diagonalizable, it is defective.

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

$B = XDX^{-1}$

$A = ZXDX^{-1}Z^{-1}$

A and $B$ are orthogonally similar if $\exists X$, s.t. $X = X^T$, and $XAX^T = B$.
If $A$ is normal, so $A^H A = A A^H$ is symmetric

if real $A^T A = A A^T$

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

orthogonally similar

$A = U S V^T$

$A A^T = U S^2 U^T$ eigenvalue decompositions

$A^T A = V S^2 V^T$

if $A$ is normal, then $U = V$, so $A = U S U^T$
property
symmetric
normal (real)
Hermition
A = A^H

similarity
orthogonally
orthogonally
written by
X = X^H

matrix
real diagonal
real diagonal
real diagonal
Rayleigh Quotient and Power Iteration

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

$$\rho_A(x) = \frac{x^T A x}{x^T x} \quad \text{if } x \text{ is eigenvector} \quad \rho_A(x) = \lambda$$

$$\rho_A(x) = \arg\min_{\lambda} \|x - \lambda x\|_2 \quad \text{if } x \text{ is not eigenvector}$$

Normal eqns are $x^T x = x^T y$, $\lambda = \frac{x^T A y}{x^T x}$, $A x$

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

Start with a random $x^{(0)}$ and iterate linearly towards $\lambda_1/\lambda_2$

$$x^{(i+1)} = A x^{(i)} = A (x_1 y_1 + x_2 y_2 + \ldots) = x_1 y_1 + x_2 \lambda_2 y_2 + \ldots$$

If $A$ is symmetric, eigenvalues $\lambda_1 > \lambda_2 > \ldots$

Eigenvectors $y_1, y_2, \ldots$. 
Inverse Iteration and Rayleigh Quotient Iteration

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

$$A^{-1} = XD^{-1}X^{-1}$$  \[ \lambda_{\max}(A^{-1}) = \frac{1}{\lambda_{\min}(A)} \]

$A = XDX^{-1}$

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

solve linear system at each step (solve costs $O(n^3)$)

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

$$\left(A - \sigma A(x_i)I\right)x_{i+1} = x_i$$

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

conv rate \[ \frac{\lambda_{\max}(\cdot)}{\lambda_{\min}(\cdot)} \]

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

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

$\beta = \hat{D} + \delta \hat{B}$

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 

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

Consider the effect of a matrix perturbation on an eigenvalue $\lambda$ associated with a right eigenvector $x$ and a left eigenvector $y^H$, $\lambda = y^H Ax / y^H x$

A more accurate eigenvalue approximation than Rayleigh quotient for a normalized perturbed eigenvector (e.g. iterative guess) $\hat{x} = x + \delta x$, can be obtain with an estimate of both eigenvectors (also $\hat{y} = y + \delta y$),
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:

- It's 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 $H$, i.e. $A = QHQ^T$:

- 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
In orthogonal iteration $\hat{Q}_{i+1}^T \hat{R}_{i+1} = A \hat{Q}_i$, QR iteration computes $A_{i+1} = R_i Q_i = \hat{Q}_{i+1}^T A \hat{Q}_{i+1}$ at iteration $i$: 

QR Iteration
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

\[ A = D + \alpha uu^T \]
Solving Tridiagonal Symmetric Eigenproblems (II)

- Jacobi iteration

- Bisection

- Relatively robust representation (RRR and MRRR)
Define $k$-dimensional Krylov subspace matrix

$$K_k = \begin{bmatrix} x_0 & Ax_0 & \cdots & A^{k-1}x_0 \end{bmatrix}$$

Show that $K_n^{-1}AK_n$ is a companion matrix $C$:
Krylov Subspaces

- Given $QR = K_k$, we obtain an orthonormal basis for the Krylov subspace,

\[ \mathcal{K}_k(A, x_0) = \text{span}(Q) = \{ \rho_A(A)x_0 : \deg(\rho_A) < k \} \]

- Consider whether $k - 1$ steps of power iteration starting from $x_0$ lead to an approximation in the Krylov subspace, also consider QR (subspace) iteration:
Given $QR = K_k$, we obtain an orthonormal basis for the Krylov subspace and $H_k = Q^T AQ$ which minimizes $\| AQ - QH \|_2$.

$H_k$ is Hessenberg, because the companion matrix $C_k$ is Hessenberg:
Rayleigh-Ritz Procedure

- The eigenvalues/eigenvectors of $H_k$ are the *Ritz values/vectors*:

- The Ritz vectors and values are the *ideal approximations* of the actual eigenvalues and eigenvectors based on only $H_k$ and $Q$: 
Arnoldi Iteration

- Arnoldi iteration computes $H$ directly using the recurrence $q_i^T A q_j = h_{ij}$.

- 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 $k$th 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 $H_k$, $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$ with respect to eigenvalues of $A$, $\lambda_1 \geq \cdots \geq \lambda_n$ 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 $A = XDX^{-1}$ how can we compute $A^k$?

- What about $e^A$, $\log(A)$? generally $f(A)$?
Consider solutions to an ordinary differential equation of the form \( \frac{dx}{dt}(t) = Ax(t) + f(t) \) with \( x(0) = x_0 \):

\[
x(t) = e^{tA}x_0 + \int_0^t e^{(t-\tau)A}f(\tau)d\tau
\]

Using \( A = XDX^{-1} \) permits us to compute the solution explicitly (Jordan form also suffices if \( A \) is defective):
Consider a more general linear differential equation of the form
\[ B \frac{dx}{dt}(t) = Ax(t) + f(t) \] with \( x(0) = x_0 \), which we can reduce to the usual form by premultiplying with \( B^{-1} \):

If we can find \( X \) such that \( A = XD_A X^{-1} \) and \( B = XD_B X^{-1} \) we could solve this equation while preserving symmetry of \( A \) and \( B \):
Generalized Eigenvalue Problem

- A generalized eigenvalue problem has the form $Ax = \lambda Bx$,

- When $A$ and $B$ are symmetric, if one is SPD, we can perform Cholesky on $B$, multiply $A$ by the inverted factors, and diagonalize it:
Canonical Forms Generalized Eigenvalue Problem

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

- For some unitary $P, Q$, $A = PT_AQ^H$ and $B = PT_BQ^H$ where $T_A$ and $T_B$ are triangular
In a polynomial eigenvalue problem, we seek solutions $\lambda, x$ to

$$\sum_{i=0}^{d} \lambda^i A_i x = 0$$

Assuming for simplicity that $A_d = I$, solutions are given by solving the matrix eigenvalue problem with the block-companion matrix

$$[-A_{d-1} \cdots -A_0]$$

$$[I \quad 0 \quad \cdots]$$

$$[\cdots \quad \cdots]$$