CS 450: Numerical Analysis\(^1\)

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 $\mathbf{A}$ has eigenvector-eigenvalue pair (eigenpair) $(\lambda, \mathbf{x})$ if

$$e^{\lambda t} \mathbf{A}^t \text{ if } D \text{ are } e^{\lambda t} \mathbf{A}^t \text{ of } \mathbf{A} ? D^t$$

$$\mathbf{A} = \mathbf{X} \mathbf{D} \mathbf{X}^{-1}, \quad \mathbf{A}^2 = \mathbf{X} \mathbf{D} \mathbf{X}^{-1} \mathbf{D} \mathbf{X}^{-1} = \mathbf{X} \mathbf{D}^2 \mathbf{X}^{-1}$$

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

  if $\mathbf{A}$ is real, then if $a + bi$ is an eig. of $\mathbf{A}$, then $a - bi$ is also
Eigenvalue Decomposition

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

$$A = XDX^{-1} \quad X = \begin{bmatrix} x_1, \ldots, x_n \end{bmatrix}$$

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

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

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

$Z^T = Z^{-1}$
### Similarity of Matrices

<table>
<thead>
<tr>
<th>matrix</th>
<th>similarity</th>
<th>reduced form</th>
</tr>
</thead>
<tbody>
<tr>
<td>real</td>
<td>SPD</td>
<td>real positive diagonal</td>
</tr>
<tr>
<td>real symmetric</td>
<td>orthogonal</td>
<td>tridiagonal $\mathbb{R}$</td>
</tr>
<tr>
<td>Hermitian</td>
<td>unitary</td>
<td>real diagonal</td>
</tr>
<tr>
<td>normal</td>
<td>unitary</td>
<td>diagonal</td>
</tr>
<tr>
<td>real</td>
<td>orthogonally</td>
<td>Hessenberg (real)</td>
</tr>
<tr>
<td>diagonalizable</td>
<td>invertible</td>
<td>diagonal</td>
</tr>
<tr>
<td>arbitrary</td>
<td>unitarily</td>
<td>triangular (Schur)</td>
</tr>
<tr>
<td></td>
<td>invertible</td>
<td>bidiagonal (Jordan)</td>
</tr>
</tbody>
</table>
Canonical Forms

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

\[
A = Z \begin{bmatrix}
J_1 & & & \\
& J_2 & & \\
& & \ddots & \\
& & & J_n
\end{bmatrix} Z^\dagger, \text{ each } J_i = \begin{bmatrix}
\lambda_i & 1 & & \\
0 & \lambda_i & & \\
& 0 & \ddots & \\
& & 0 & \lambda_i
\end{bmatrix}
\]

unique modulo permutation of \( J_i \)'s

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

\[
A = Q \begin{bmatrix}
\lambda_1 & & & \\
& \ddots & & \\
& & \ddots & \\
& & & \lambda_n
\end{bmatrix} Q^H, \quad \det(T - \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}) = 0
\]

column of \( Q \)-Schur vectors

\[
T_{ii} -> \theta_i
\]
Computing Eigenvectors of Matrices in Schur Form

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

\[
A \text{ is similar to } B \Rightarrow A = Z \Sigma Z^{-1}
\]

\[
B = XDX^{-1}, \quad A = \begin{bmatrix} 2 & X \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} X^{-1} \\ 2 \end{bmatrix} = YDY^{-1}
\]

- It's easy to obtain eigenvectors of triangular matrix \( T \):

\[
T e_1 = \lambda_1 e_1
\]

\[
\begin{bmatrix} 1 & \vdots \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \\ 0 & \cdots & 0 \\ \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ \end{bmatrix}
\]

\[
(T - \lambda I) \begin{bmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{n-1,n} \\ u_{n,n} \\ \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 0 \\ \end{bmatrix}
\]

\[
\begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n1} & u_{n2} & \cdots & u_{nn} \\ \end{bmatrix} \begin{bmatrix} u_{11}^{-1} \\ u_{21}^{-1} \\ \vdots \\ u_{n1}^{-1} \\ \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ \end{bmatrix}
\]

\[
u_{11} u_{11} = -u_{1n} \Rightarrow \begin{bmatrix} u_{11} \\ \vdots \\ \end{bmatrix} = \begin{bmatrix} u_{11}^{-1} \\ \vdots \\ \end{bmatrix}
\]
Rayleigh Quotient

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

$$\rho_A(x) = \frac{\langle x, Ax \rangle}{\|x\|^2}$$

solving to

$$x^* A x$$
Perturbation Analysis of Eigenvalue Problems

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

\[
B = \hat{X}^{-1}(A + \delta A)\hat{X} = D + \hat{X}^{-1}\delta AX
\]

\[
|\|B\|_2 \leq \|\hat{X}^{-1}\|_2\|\delta A\|_2\|\hat{X}\|_2 \leq \epsilon(A)\|\delta A\|_2
\]

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

\[
D_i = \{z \in \mathbb{C} : |z - a_{ii}| \leq r_i\}.
\]
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 $X^{-1}(A + \delta A)X$, which contain the eigenvalues $D$ of $A$ and the perturbed eigenvalues $D + \delta D$ of $A + \delta A$ provided that $||\delta A||$ is sufficiently small.
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 $y$, $\lambda = y^H A x / y^H x$

\[
\frac{y^H A x}{y^H x} = y^H \lambda x
\]

\[
(y^H A x) x = y^H \lambda x
\]

\[
\begin{vmatrix}
\frac{y^H A x}{y^H x}
\end{vmatrix} = 1
\]

\[
\left| \frac{y^H (A + \delta A) x}{y^H x} - \lambda \right| = \frac{||\delta A||}{||y^H x||} \leq \frac{||\delta A||}{||y^H x||}
\]

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

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

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

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

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

$$A = U \Lambda V^H \quad \Rightarrow \quad A^k = U \Lambda^k V^H$$

$$x^{(k)} = A^k x^{(0)} = \sum_{i=1}^n u_i \lambda_i^k v_i^H x^{(0)}$$

$$\alpha^{(i,k)} = \frac{\langle v_i, x^{(0)} \rangle}{\langle \alpha^{(i,k)} \rangle = \frac{A x^{(0)}}{\| x^{(0)} \|}}$$
Inverse Iteration and 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.

\[ \beta = A - \lambda_i x, v^H = \left( \sum_{i=1}^{\infty} \lambda_i x_i v_i^H \right) - \lambda_i x_i v_i^H \]

- For symmetric, \( v^H = x_i^H \), preserves eigenvectors.
- For nonsymmetric, \( v = x_i^H \), preserves Schur vectors.

Schur decom. \( A = QTQ^H \)

\[ \begin{bmatrix} \lambda_i & * \\ \ast & \ddots \end{bmatrix} = A - \lambda_i x_i x_i^H = Q C (T - \lambda_i x_i x_i^H)^{-1} Q^H \]
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 $H$, i.e. $A = QHQ^T$:

- 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 $A_i = \hat{Q}_i^T A \hat{Q}_i$ and show that QR iteration obtains $A_{i+1} = \hat{Q}_{i+1}^T A \hat{Q}_{i+1}$
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

$$A = D + \alpha uu^T$$
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 \textit{Ritz values/vectors}:

- The Ritz vectors and values are the \textit{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 $f(A)$?
Applications of Eigenvalue Problems: Differential Equations

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 \( B \):
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
Nonlinear Eigenvalue Problem

- 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

$$
\begin{bmatrix}
-A_{d-1} & \cdots & -A_0 \\
I & 0 & \cdots \\
& \ddots & \ddots
\end{bmatrix}
$$