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 \( A \) has eigenvector-eigenvalue pair (eigenpair) \((\lambda, x)\) if

Each \( n \times n \) matrix has up to \( n \) eigenvalues, which are either real or complex.
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

<table>
<thead>
<tr>
<th>matrix</th>
<th>similarity</th>
<th>reduced form</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>real symmetric</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hermitian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>normal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>real</td>
<td></td>
<td></td>
</tr>
<tr>
<td>diagonalizable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>arbitrary</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Canonical Forms

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

- Any diagonalizable matrix is *unitarily similar* to a triangular matrix, giving its *Schur form*:
Given the eigenvectors of one matrix, we seek those of a similar matrix:

Its easy to obtain eigenvectors of triangular matrix $T$: 

Activity: Calculating Eigenpairs of a Triangular Matrix
Rayleigh Quotient

- For any vector $x$, the *Rayleigh quotient* provides an estimate for some eigenvalue of $A$: 
Suppose we seek eigenvalues \( D = X^{-1}AX \), but find those of a slightly perturbed matrix \( D + \delta D = \hat{X}^{-1}(A + \delta A)\hat{X} \):

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 $y$, $\lambda = y^H Ax / y^H x$

A more accurate eigenvalue approximation than Rayleigh quotient for a normalized perturbed eigenvector (e.g., iterative guess) $\hat{x} = x + \delta x$, can be 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 $A$:

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

\[ X^{(k+1)} = A X^{(k)} \]

leading to eigenvalues of \( A \)

\[ \text{span}(X^{(k)}) = \text{span}(\lambda_1, \ldots, \lambda_k) \]

\[ \lim_{k \to \infty} \]

- Orthogonal iteration performs QR at each step to ensure stability

\[ O(m^2) \]

\[ X^{(1)} = Q R \]

\[ \text{span}(X^{(1)}) = \text{span}(Q^{(1)}) \]

\[ O(n^2) \]

\[ X^{(i+1)} = A Q^{(i)} \]

Ortho-iter (\( A^T A \)) = "Randomized SVD"

- eigenvalues of \( A^T A \) are singular vectors of \( A \)
QR Iteration

- QR iteration reformulates orthogonal iteration for \( n = k \) to reduce cost/step,

\[
\hat{Q}_{i+1} R_{i+1} = A \hat{Q}_i
\]

Using induction, we assume \( A_i = \hat{Q}_i^T A \hat{Q}_i \) and show that QR iteration obtains \( A_{i+1} = \hat{Q}_{i+1}^T A \hat{Q}_{i+1} \).
An iteration converges to Schur form

\[ A = Q^T \Delta Q \]

\[ \hat{Q}_i = Q \]

\[ A \hat{Q}_i = Q^T \Delta \]

\[ \hat{Q}_i \in \mathbb{R} \]
QR Iteration with Shift

- QR iteration can be accelerated using shifting:

$$Q_i R_i = A_i - \sigma_i I$$

$$A_{i+1} = R_i Q_i + \sigma_i I$$

- The shift is typically selected to accelerate convergence with respect to a particular eigenvalue:

The shift is also performing a QZ iteration on $$A_i$$ at each step.
QR Iteration Complexity

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

$$O(n^3) \quad \text{of symmetric} \quad O(n)$$

$$\text{QR}(A) = \begin{bmatrix} \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \end{bmatrix}$$

QR iteration cost/iteration is $$O(n^2)$$ for nonsym
$$O(n)$$ for sym
Solving Tridiagonal Symmetric Eigenproblems

A variety of methods exists for the tridiagonal eigenproblem:

- **QR iteration**
  - \( O(n^3) \) QR factorizations / eigenvalue => \( O(n^3) \) cost
  - \( O(n^3) \) eigenvectors

- **Divide and conquer**
  - Nonlinear equation to combine eigenvalue decompositions of \( T_1 \) and \( T_2 \)
  - MRRR engines in \( O(n^2) \) cost
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

\[ A = D + \alpha uu^T. \]
Krylov subspace methods work with information contained in the $n \times k$ matrix

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

(Assume $K_n$ is invertible)

The matrix $K_n^{-1}AK_n$ is a companion matrix $C$:

$$K_n^{(m)} = A^{i-1}x_0$$

$$AK_n = \begin{bmatrix} A^{(m)} & \cdots & A^{(n)} \end{bmatrix} = \begin{bmatrix} k_n^{(m)} & \cdots & k_n^{(n)} \end{bmatrix}$$

$$K_n^{-1}AK_n = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

$\rightarrow$ companion matrix
Krylov Subspaces

Given $Q_k R_k = K_k$, we obtain an orthonormal basis for the Krylov subspace,

$$K_k(A, x_0) = \text{span}(Q_k) = \{p(A)x_0 : \deg(p) < k\},$$

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 $H_k = Q_k^T AQ_k$ minimizes $\|AQ_k - Q_k H_k\|_2$:

  $Q_k x \approx AQ_k$

  $Q_k^+ Q_k x = Q_k^+ AQ_k$

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

  $Q_k^T R_n = C_n$

  $K_n^{-1} A L_n = C_n$

  $R_n^T Q_n^+ A Q_n R_n = C_n$

  $H_n = Q_n^+ A Q_n = R_n C_n R_n^{-1}$
Rayleigh-Ritz Procedure

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

$$H_k x = \lambda x$$

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

$$\max_{x \in \text{span}(Q_k)} \frac{x^T A x}{x^T x} = \max_{y \in Q_k} y^T A Q_k y$$

**Demo:** Arnoldi vs Power Iteration

**Activity:** Computing the Maximum Ritz Value
Arnoldi Iteration

- Arnoldi iteration computes $H = H_n$ directly using the recurrence $q_i^T A q_j = h_{ij}$, where $q_l$ is the $l$th column of $Q_n$:

- After each matrix-vector product, orthogonalization is done with respect to each previous vector:

\[
q_{i+1} = (u_j - \sum_{s < i} h_{is} q_s) \text{ normalized}
\]

\[
u_j = A q_j \quad \text{and} \quad h_{ij} = q_i^T u_j \quad \text{for each } i < j
\]
Lanczos Iteration

- Lanczos iteration provides a method to reduce a symmetric matrix to a tridiagonal matrix:

\[ A_k = Q_k^T A Q_k \]

is upper Hessenberg

is symmetric if \( A \) is symmetric

\( \Rightarrow \) tridiagonal

- After each matrix-vector product, it suffices to orthogonalize with respect to two previous vectors:

\[ v_k = u_k - v_{k-1} \]

need to compute only \( u_{k-1, k} = h_{k, k+1} - q_k^T \tilde{u}_{k-1} \)
Cost Krylov Subspace Methods

The cost of matrix-vector multiplication when the matrix has $m$ nonzeros:

\[ V_i = \sum a_{ij} u_j \]

\[ V_i = 0 \text{ for each nonzero } a_{ij} \text{ in } A \]

\[ V_i = a_{ij} u_j \]

The cost of orthogonalization at the $k$th iteration of a Krylov subspace method is

\[ \text{per iteration cost: } \begin{cases} O(nk) & \text{for Arnoldi} \\ O(n^2) & \text{Lanczos} \end{cases} \]

\[ \text{over all iterations: } \begin{cases} O(nk^2) & \text{for Arnoldi} \\ O(n^3) & \text{for Lanczos} \end{cases} \]

where $n,k \leq m$.
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 $Ax = \lambda Bx$,

- When $A$ and $B$ are symmetric and $B$ is SPD, we can perform Cholesky on $B$, multiply $A$ by the inverted factors, and diagonalize it:

- Alternative canonical forms and methods exist that are specialized to the generalized eigenproblem.