CS 450: Numerical Analysis

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

\[ Ax = \lambda x \]

- $\lambda x$ is also an eigenvector

- $\lambda, x \in \mathbb{R}^n, x \neq 0$, $Ax = \lambda x$

- Eigenspace, $\dim(x) \leq$ multiplicity of $\lambda$

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

- If we have $\lambda = a + bi$

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

- For diagonalizable matrices, $\lambda_1, \ldots, \lambda_k, x_1, \ldots, x_k$

\[ \sum_{i=1}^{k} \dim(x_i) = \dim(A) \frac{n}{n} \]
Eigenvalue Decomposition

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

\[ A = X \Lambda X^{-1} \]

- $A$ and $B$ are similar, if there exist $Z$ such that $A = Z B Z^{-1}$
### Similarity of Matrices

<table>
<thead>
<tr>
<th>A matrix</th>
<th>B = ZAZ&lt;sup&gt;-1&lt;/sup&gt;</th>
<th>reduced form</th>
</tr>
</thead>
<tbody>
<tr>
<td>real symmetric</td>
<td>orthogonal</td>
<td>diagonal, positive</td>
</tr>
<tr>
<td>SPD</td>
<td>orthogonal</td>
<td>real tridiagonal</td>
</tr>
<tr>
<td>Hermitian</td>
<td>unitary</td>
<td>real diagonal</td>
</tr>
<tr>
<td>normal</td>
<td>unitary</td>
<td>real Hessenberg</td>
</tr>
<tr>
<td>real</td>
<td>orthogonal</td>
<td>diagonal</td>
</tr>
<tr>
<td>diagonalizable</td>
<td>invertible</td>
<td>diagonal</td>
</tr>
<tr>
<td>arbitrary</td>
<td>unitary, invertible</td>
<td>triangular (Schur form)</td>
</tr>
</tbody>
</table>

- **real symmetric**: \( \lambda > 0 \)
- **SPD**: symmetric positive definite
- **Hermitian**: \( A^H = A \)
- **normal**: \( A^H A = A A^H \)
Canonical Forms

- Any matrix is \textit{similar} to a bidiagonal matrix, giving its \textit{Jordan form}:

\[
A = X \begin{bmatrix}
J_1 & & \\
& \ddots & \\
& & J_k
\end{bmatrix} X^{-1}
J_k = \begin{bmatrix}
\lambda_1 & & \\
& \ddots & \\
& & \lambda_k
\end{bmatrix}
\]

- Any diagonalizable matrix is \textit{unitarily similar} to a triangular matrix, giving its \textit{Schur form}:

\[
A = QTQ^H
\]

where \(T\) is a triangular matrix and \(Q\) is a unitary matrix.
Eigenvectors from Schur Form

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

\[ T_x = \lambda x \]
\[ A = Q \varnothing Q^T \]
\[ T = X \varnothing X^{-1} \]
\[ \text{eigvec}(A) = QX \]

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

\[ 0 = (AX - \lambda X) = (A - \lambda I)X \]
Rayleigh Quotient

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

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

\[ \rho_A(x) = \lambda \quad \text{for some } \lambda \in \sigma(A) \]

\[ x^T A x = \lambda x^T x \]

The **field of values** $\sigma(A)$ is given by:

\[ \sigma(A) = \{ \rho_A(x) : x \in \mathbb{R}^n \} \]

\[ \rho_A(x) \leq \sigma_{\max}(A) \]

$y = Ax$, $\rho_A(x)$ is a good estimate of $\lambda$ even associated with a vector near $x$, since it is the solution $x^2 = y$.

\[ \begin{bmatrix} x^2 \end{bmatrix} = \begin{bmatrix} y \end{bmatrix} \]

\[ \begin{bmatrix} x^T x & 0 \\ 0 & x^T x \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix} = \begin{bmatrix} x^Ty \\ 1 \end{bmatrix} \]

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

\[
X^{-1}(A + \delta A)X = D + X^{-1}\delta A X = \begin{bmatrix}
\vdots & \vdots \\
\vdots & \vdots \\
\end{bmatrix}
\]

\[
\|X^{-1}\delta A X\| \leq \|X^{-1}\| \|\delta A\| \|X\| = \kappa(X) \|\delta A\|
\]

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

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

\[
D_i = \{ z \in \mathbb{C} : |z - a_{ii}| \leq r_i \}.
\]
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 $y$, $\lambda = y^H A x / y^H x$

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*
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 for $n = k$ to reduce cost/step,

- 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 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

\[ A = D + \alpha uu^T. \]
Krylov subspace methods work with information contained in the $n \times k$ matrix $K_k = [x_0 \ Ax_0 \ \cdots \ A^{k-1}x_0]$

The matrix $K_n^{-1}AK_n$ is a companion matrix $C$: 
Given $Q_k R_k = K_k$, we obtain an orthonormal basis for the Krylov subspace,

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

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

\[ H_k = Q_k^T A Q_k \] minimizes \[ \| A Q_k - Q_k H_k \|_2 : \]

\[ H_k \] is Hessenberg, because the companion matrix \[ C_k \] is Hessenberg:
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 $Q_k$:
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:
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 $k$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:
Generalized Eigenvalue Problem

- A generalized eigenvalue problem has the form \( Ax = \lambda B x \),

- 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.