AMS526: Numerical Analysis I
(Numerical Linear Algebra)
Lecture 19: Sensitivity of Eigenvalues;
Arnoldi Iterations

Xiangmin Jiao
Stony Brook University
Outline

1 Sensitivity of Eigenvalues

2 Krylov Subspace and Arnoldi Iterations
Residual of Eigenvalue Problems

- We consider how eigenvalues and eigenvectors of $A + \delta A$ differ from those of $A$
- A small residual $r = Ax - \lambda x$ guarantees small perturbation to $A$ in backward analysis

**Theorem**

Given $A \in \mathbb{C}^{n \times n}$, let $x$ be an approximate eigenvector of $A$ with $\|x\|_2 = 1$, $\lambda$ an associated approximate eigenvalue, and $r = Ax - \lambda x$ the residual. Then $\lambda$ and $x$ are an exact eigenpair of some perturbed matrix $A + \delta A$, where $\|\delta A\|_2 = \|r\|_2$.

**Proof.**

Let $\delta A = -rx^*$, which satisfies $\|\delta A\|_2 = \|r\|_2\|x\|_2 = \|r\|_2$. Then

$$(A + \delta A)x = Ax - rx^*x = Ax - r = \lambda x.$$
Sensitivity of Eigenvalues

- Condition number of matrix $X$ determines sensitivity of eigenvalues

**Theorem**

Let $A \in \mathbb{C}^{n \times n}$ be a nondefective matrix, and suppose $A = X\Lambda X^{-1}$, where $X$ is nonsingular and $\Lambda$ is diagonal. Let $\delta A \in \mathbb{C}^{n \times n}$ be some perturbation of $A$, and let $\mu$ be an eigenvalue of $A + \delta A$. Then $A$ has an eigenvalue $\lambda$ such that

$$ |\mu - \lambda| \leq \kappa_p(X) \|\delta A\|_p $$

for $1 \leq p \leq \infty$.

- $\kappa_p(X)$ measures how far eigenvectors are from linear dependence
- For normal matrices, condition number $\kappa_2(X) = 1$ and $\kappa_p(X) = O(1)$, so eigenvalues of normal matrices are always well-conditioned
Sensitivity of Eigenvalues

Proof.

Let \( \delta \Lambda = X^{-1} (\delta A) X \). Then

\[
\| \delta \Lambda \|_p \leq \| X^{-1} \|_p \| \delta A \|_p \| X \|_p = \kappa_p(X) \| \delta A \|_p.
\]

Let \( y \) be an eigenvector of \( \Lambda + \delta \Lambda \) associated with \( \mu \). Suppose \( \mu \) is not an eigenvalue of \( A \), so \( \mu I - \Lambda \) is nonsingular.

\[
(\Lambda + \delta \Lambda)y = \mu y \Rightarrow (\mu I - \Lambda)y = (\delta \Lambda) y \Rightarrow y = (\mu I - \Lambda)^{-1} (\delta \Lambda) y.
\]

Thus

\[
\left\| (\mu I - \Lambda)^{-1} \right\|_p^{-1} \leq \| \delta \Lambda \|_p.
\]

\[
\| (\mu I - \Lambda)^{-1} \|_p = |\mu - \lambda|^{-1}, \text{ where } \lambda \text{ is the eigenvalue of } A \text{ closest to } \mu.
\]

Thus,

\[
|\mu - \lambda| \leq \| \delta \Lambda \|_p \leq \kappa_p(X) \| \delta A \|_p.
\]
Left and Right Eigenvectors

- To analyze sensitivity of individual eigenvalues, we need to define left and right eigenvectors
  - $Ax = \lambda x$ for nonzero $x$ then $x$ is right eigenvector associated with $\lambda$
  - $y^*A = \lambda y^*$ for nonzero $y$, then $y$ is left eigenvector associated with $\lambda$
- Left eigenvectors of $A$ are right eigenvectors of $A^*$

**Theorem**

Let $A \in \mathbb{C}^{n \times n}$ have distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ with associated linearly independent right eigenvectors $x_1, \ldots, x_n$ and left eigenvectors $y_1, \ldots, y_n$. Then $y_j^*x_i \neq 0$ if $i = j$ and $y_j^*x_i = 0$ if $i \neq j$.

**Proof.**

If $i \neq j$, $y_j^*Ax_i = \lambda_i y_j^*x_i$ and $y_j^*Ax_i = \lambda_j y_j^*x_i$. Since $\lambda_i \neq \lambda_j$, $y_j^*x_i = 0$. If $i = j$, since $\{x_i\}$ form a basis for $\mathbb{C}^n$, $y_i^*x_i = 0$ together with $y_i^*x_j = 0$ would imply that $y_i = 0$. This leads to a contradiction.
Sensitivity of Individual Eigenvalues

- We analyze sensitivity of individual eigenvalues that are distinct.

**Theorem**

Let $A \in \mathbb{C}^{n \times n}$ have $n$ distinct eigenvalues. Let $\lambda$ be an eigenvalue with associated right and left eigenvectors $x$ and $y$, respectively, normalized so that $\|x\|_2 = \|y\|_2 = 1$. Let $\delta A$ be a small perturbation satisfying $\|\delta A\|_2 = \epsilon$, and let $\lambda + \delta \lambda$ be the eigenvalue of $A + \delta A$ that approximates $\lambda$. Then

$$|\delta \lambda| \leq \frac{1}{|y^* x|} \epsilon + O(\epsilon^2).$$

- $\kappa = 1/|y^* x|$ is condition number for eigenvalue $\lambda$
- A simple eigenvalue is sensitive if its associated right and left eigenvectors are nearly orthogonal.
Sensitivity of Individual Eigenvalues

**Proof.**

We know that $|\delta \lambda| \leq \kappa_p(X)\epsilon = O(\epsilon)$. In addition, $\delta x = O(\epsilon)$ when $\lambda$ is a simple eigenvalue (proof omitted). Because

$$(A + \delta A)(x + \delta x) = (\lambda + \delta \lambda)(x + \delta x),$$

thus

$$(\delta A)x + A(\delta x) + O(\epsilon^2) = (\delta \lambda)x + \lambda(\delta x) + O(\epsilon^2).$$

Left multiplying by $y^*$ and using equation $y^*A = \lambda y^*$, we obtain

$$y^*(\delta A)x + O(\epsilon^2) = (\delta \lambda)y^*x + O(\epsilon^2)$$

and hence

$$\delta \lambda = \frac{y^*(\delta A)x}{y^*x} + O(\epsilon^2).$$

Since $|y^*(\delta A)x| \leq \|y\|_2 \|(\delta A)\|_2 \|x\|_2 = \epsilon$, $|\delta \lambda| \leq \frac{1}{|y^*x|}\epsilon + O(\epsilon^2)$.
Sensitivity of Multiple Eigenvalues and Eigenvectors

- Sensitivity of multiple eigenvalues is more complicated
  - For multiple eigenvalues, left and right eigenvectors can be orthogonal, hence very ill-conditioned
  - In general, multiple or close eigenvalues can be poorly conditioned, especially if matrix is defective

- Condition numbers of eigenvectors are also difficult to analyze
  - If matrix has well-conditioned and well-separated eigenvalues, then eigenvectors are well-conditioned
  - If eigenvalues are ill-conditioned or closely clustered, then eigenvectors may be poorly conditioned
Outline

1. Sensitivity of Eigenvalues

2. Krylov Subspace and Arnoldi Iterations
Krylov Subspace Methods

- Given $A$ and $b$, Krylov subspace

$$\{b, Ab, A^2b, \ldots, A^{k-1}b\}$$

<table>
<thead>
<tr>
<th></th>
<th>linear systems</th>
<th>eigenvalue problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hermitian</td>
<td>CG</td>
<td>Lanczos</td>
</tr>
<tr>
<td>Nonhermitian</td>
<td>GMRES, BiCG, etc.</td>
<td>Arnoldi</td>
</tr>
</tbody>
</table>

- CG, GMRES etc. are Krylov subspace methods for solving sparse linear systems (later)

- Lanczos and Arnoldi iterations are Krylov subspace methods for reduction to Hessenberg form
Review: Reduction to Hessenberg Form

- **General A**: First convert to *upper-Hessenberg* form, then to upper triangular

  \[
  \begin{bmatrix}
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \end{bmatrix}
  \xrightarrow{\text{Phase 1}}
  \begin{bmatrix}
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \end{bmatrix}
  \xrightarrow{\text{Phase 2}}
  \begin{bmatrix}
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \end{bmatrix}
  \]

- **Hermitian A**: First convert to *tridiagonal* form, then to diagonal

  \[
  \begin{bmatrix}
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \end{bmatrix}
  \xrightarrow{\text{Phase 1}}
  \begin{bmatrix}
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \end{bmatrix}
  \xrightarrow{\text{Phase 2}}
  \begin{bmatrix}
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \times & \times & \times & \times & \times \\
  \end{bmatrix}
  \]

- In general, phase 1 is direct and requires $O(m^3)$ flops
Arnoldi Iteration

- The Arnoldi iteration reduces a general, nonsymmetric matrix $A$ to Hessenberg form by similarity transformation $A = QHQ^*$
- It is analogous to Gram-Schmidt-style iteration instead of Householder reflections
- Let $Q_k = [q_1 \mid q_2 \mid \cdot \mid q_k]$ be $m \times k$ matrix with first $k$ columns of $Q$ and $\tilde{H}_k$ be $(k + 1) \times k$ upper-left section of $H$, i.e., $\tilde{H}_k = H_{1:k+1,1:k}$
- Consider first $k$ columns of $AQ = QH$, or $AQ_k = Q_{1:k} = Q_{k+1} \tilde{H}_k$

$$
\begin{bmatrix}
A \\
\\
A
\end{bmatrix}
\begin{bmatrix}
q_1 \\
\cdots \\
q_k
\end{bmatrix}
= 
\begin{bmatrix}
q_1 \\
\cdots \\
q_{k+1}
\end{bmatrix}
\begin{bmatrix}
h_{11} & \cdots & h_{1k} \\
h_{21} & \ddots & \vdots \\
h_{k+1,1} & \cdots & h_{k+1,k}
\end{bmatrix}
\begin{bmatrix}
Q_k \\
Q_{k+1}
\end{bmatrix}
\begin{bmatrix}
\tilde{H}_k
\end{bmatrix}
$$

- Question: How do we choose $q_1$?
Arnoldi Algorithm

- Start with a random \( q_1 \), then determine \( q_2 \) and \( \tilde{H}_1 \), and so on.
- The \( k \)th columns of \( AQ_k = Q_{k+1}\tilde{H}_k \) can be written as

\[
Aq_k = h_{1k}q_1 + \cdots + h_{kk}q_k + h_{k+1,k}q_{k+1}
\]

where \( h_{ik} = q_i^* Aq_k \).

Algorithm: Arnoldi Iteration

- given random nonzero \( b \), let \( q_1 = b/\|b\| \)
- for \( k = 1, 2, 3, \ldots \)
  - \( v = Aq_k \)
  - for \( j = 1 \) to \( k \)
    - \( h_{jk} = q_j^* v \)
    - \( v = v - h_{jk}q_j \)
    - \( h_{k+1,k} = \|v\| \)
    - \( q_{k+1} = v/h_{k+1,k} \)

- Question: What if \( q_1 \) happens to be an eigenvector?
QR Factorization of Krylov Matrix

- The vector $q_j$ from Arnoldi are orthonormal bases of successive Krylov subspaces

\[ K_k = \langle b, Ab, \ldots, A^{k-1}b \rangle = \langle q_1, q_2, \ldots, q_k \rangle \subseteq \mathbb{C}^m \]

- $Q_k$ is reduced QR factorization $K_k = Q_k R_k$ of Krylov matrix

\[
K_k = \begin{bmatrix}
    b & Ab & \cdots & A^{k-1}b \\
\end{bmatrix}
\]

- The projection of $A$ onto this space gives $k \times k$ Hessenberg matrix

\[ H_k = Q_k^* A Q_k = \tilde{H}_{1:k,1:k} \]

- Eigenvalues of $H_k$ (known as Ritz values) produce good approximations of those of $A$