AMS526: Numerical Analysis I
(Numerical Linear Algebra)
Lecture 11: Conditioning of Least Squares Problems;
Stability of Least Squares Algorithms

Xiangmin Jiao

SUNY Stony Brook
Outline

1. Conditioning of Least Squares Problems

2. Stability of Least Squares Algorithms
Four Conditioning Problems

- Least squares problem: Given $A \in \mathbb{C}^{m \times n}$ with full rank and $b \in \mathbb{C}^m$, 

  $$\min_{x \in \mathbb{C}^n} \| b - Ax \|$$

  Its solution is $x = A^+ b$. Another quantity is $y = Ax = Pb$, where $P = AA^+$

- Consider $A$ and $b$ as input data, and $x$ and $y$ as output. We then have four conditioning problems:

<table>
<thead>
<tr>
<th>Input \ Output</th>
<th>$y$</th>
<th>$x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>$\kappa b \rightarrow y$</td>
<td>$\kappa b \rightarrow x$</td>
</tr>
<tr>
<td>$A$</td>
<td>$\kappa A \rightarrow y$</td>
<td>$\kappa A \rightarrow x$</td>
</tr>
</tbody>
</table>

- These conditioning problems are important and subtle.
Some Prerequisites

- We focus on the second column, namely $\kappa b \rightarrow x$ and $\kappa A \rightarrow x$
- However, understanding $\kappa b \rightarrow y$ and $\kappa A \rightarrow y$ is prerequisite

- Three quantities: (All in 2-norms)
  - Condition number of $A$:
    \[ \kappa(A) = \|A\| \|A^+\| = \sigma_1/\sigma_n \]
  - Angle between $b$ and $y$:
    \[ \theta = \arccos \frac{\|y\|}{\|b\|}. \quad (0 \leq \theta \leq \pi/2) \]
  - Orientation of $y$ with range($A$):
    \[ \eta = \frac{\|A\| \|x\|}{\|y\|}. \quad (1 \leq \eta \leq \kappa(A)) \]
Sensitivity of $\mathbf{y}$ to Perturbations in $\mathbf{b}$

- Intuition: The larger $\theta$ is, the more sensitive $\mathbf{y}$ is in terms of relative error
- Analysis: $\mathbf{y} = \mathbf{Pb}$, so

$$
\kappa_{\mathbf{b} \rightarrow \mathbf{y}} = \frac{\| \mathbf{P} \|}{\| \mathbf{y} \|/\| \mathbf{b} \|} = \frac{\| \mathbf{b} \|}{\| \mathbf{y} \|} = \frac{1}{\cos \theta},
$$

where $\| \mathbf{P} \| = 1$

<table>
<thead>
<tr>
<th>Input \ Output</th>
<th>$\mathbf{y}$</th>
<th>$\mathbf{x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbf{b}$</td>
<td>$\frac{1}{\cos \theta}$</td>
<td></td>
</tr>
<tr>
<td>$\mathbf{A}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Question: When the maximum is attained for perturbation $\delta \mathbf{b}$?
Sensitivity of $y$ to Perturbations in $b$

- **Intuition:** The larger $\theta$ is, the more sensitive $y$ is in terms of relative error.
- **Analysis:** $y = Pb$, so

$$
\kappa_{b \rightarrow y} = \frac{\|P\|}{\|y\|/\|b\|} = \frac{\|b\|}{\|y\|} = \frac{1}{\cos \theta},
$$

where $\|P\| = 1$

<table>
<thead>
<tr>
<th>Input \ Output</th>
<th>$y$</th>
<th>$x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>$\frac{1}{\cos \theta}$</td>
<td></td>
</tr>
<tr>
<td>$A$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Question:** When the maximum is attained for perturbation $\delta b$?
- **Answer:** When $\delta b$ is in range($A$)
Sensitivity of $x$ to Perturbations in $b$

- **Intuition:** It depends on how sensitive $y$ is to $b$, and how $y$ lies within $\text{range}(A)$
- **Analysis:** $x = A^+ b$, so

$$
\kappa_{b \rightarrow x} = \frac{\|A^+\|}{\|x\|/\|b\|} = \|A^+\| \frac{\|b\|}{\|y\|} \frac{\|y\|}{\|x\|} = \|A^+\| \frac{1}{\cos \theta} \frac{\|A\|}{\eta} = \frac{\kappa(A)}{\eta \cos \theta},
$$

where $\eta = \|A\| \|x\|/\|y\|$

<table>
<thead>
<tr>
<th>Input \ Output</th>
<th>$y$</th>
<th>$x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>$\frac{1}{\cos \theta}$</td>
<td>$\frac{\kappa(A)}{\eta \cos \theta}$</td>
</tr>
<tr>
<td>$A$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sensitivity of $x$ to Perturbations in $b$

- Assume $\cos \theta = O(1)$, $\kappa_{b \rightarrow x} = \frac{\kappa(A)}{\eta \cos \theta}$ can lie anywhere between 1 and $O(\kappa(A))$!
- Question: When the maximum is attained for perturbation $\delta b$?
Sensitivity of $\mathbf{x}$ to Perturbations in $\mathbf{b}$

- Assume $\cos \theta = O(1)$, $\kappa_{b \rightarrow \mathbf{x}} = \frac{\kappa(A)}{\eta \cos \theta}$ can lie anywhere between 1 and $O(\kappa(A))$!
- Question: When the maximum is attained for perturbation $\delta \mathbf{b}$?
- Answer: When $\delta \mathbf{b}$ is in subspace spanned by left singular vectors corresponding to smallest singular values
- Question: What if $\mathbf{A}$ is a nonsingular matrix?
Sensitivity of \( x \) to Perturbations in \( b \)

- Assume \( \cos \theta = O(1) \), \( \kappa_{b\mapsto x} = \frac{\kappa(A)}{\eta \cos \theta} \) can lie anywhere between 1 and \( O(\kappa(A)) \)!

- Question: When the maximum is attained for perturbation \( \delta b \)?

- Answer: When \( \delta b \) is in subspace spanned by left singular vectors corresponding to smallest singular values

- Question: What if \( A \) is a nonsingular matrix?

- Answer: \( \kappa_{b\mapsto x} \) can lie anywhere between 1 and \( \kappa(A) \)!
Sensitivity of \( y \) and \( x \) to Perturbations in \( A \)

- The relationship are nonlinear, because \( \text{range}(A) \) changes due to \( \delta A \)
- Intuitions:
  - The larger \( \theta \) is, the more sensitive \( y \) is in terms of relative error.
  - Tilting of range(\( A \)) depends on \( \kappa(A) \).
  - For \( x \), it depends where \( y \) lies within range(\( A \))

<table>
<thead>
<tr>
<th>Input ( b )</th>
<th>Output ( \kappa(A) )</th>
<th>( \eta ) cos ( \theta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{1}{\cos \theta} )</td>
<td>( \frac{\kappa(A)}{\eta \cos \theta} )</td>
<td></td>
</tr>
</tbody>
</table>

For second row, bounds are not necessarily tight
- Assume \( \cos \theta = O(1) \), \( \kappa_{A \rightarrow x} \) can lie anywhere between \( \kappa(A) \) and \( O(\kappa(A)^2) \)
Condition Numbers of Linear Systems

- Linear system $Ax = b$ for nonsingular $A \in \mathbb{C}^{m \times m}$ is a special case of least squares problems, where $y = b$
- If $m = n$, then $\theta = 0$, so $\cos \theta = 1$ and $\tan \theta = 0$.

<table>
<thead>
<tr>
<th>Input \ Output</th>
<th>$y$</th>
<th>$x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>1</td>
<td>$\frac{\kappa(A)}{\eta}$</td>
</tr>
<tr>
<td>$A$</td>
<td>$\leq \kappa(A)$</td>
<td>$\leq \kappa(A)$</td>
</tr>
</tbody>
</table>
Outline

1. Conditioning of Least Squares Problems

2. Stability of Least Squares Algorithms
Algorithms for Solving Least Squares Problems

- There are many variants of algorithms for solving least squares problems
  - Householder QR (with/without pivoting, explicit or implicit $Q$): **Backward stable**
  - Classical Gram-Schmidt: **Unstable**
  - Modified Gram-Schmidt with explicit $Q$: **Unstable**
  - Modified Gram-Schmidt with augmented system of equations with implicit $Q$: **Backward stable**
  - Normal equations (solve $A^T Ax = A^T b$): **Very unstable**
  - Singular value decomposition: **Stable and most accurate**

- Note that in general, only SVD is robust for solving rank deficient least squares problems
Theorem

Let the full-rank least squares problem be solved using Householder triangularization on a computer satisfying the two axioms of floating point numbers. The algorithm is backward stable in the sense that the computed solution $\tilde{x}$ has the property

$$\|(A + \delta A)\tilde{x} - b\| = \min, \quad \frac{\|\delta A\|}{\|A\|} = O(\epsilon_{\text{machine}})$$

for some $\delta A \in \mathbb{C}^{m \times n}$.

- Backward stability of the algorithm is true whether $\hat{Q}^* b$ is computed via explicit formation of $\hat{Q}$ or computed implicitly.
- Backward stability also holds for Householder triangularization with arbitrary column pivoting $AP = \hat{Q}\hat{R}$.
Gram-Schmidt Orthogonalization

- Note that Gram-Schmidt orthogonalization in general is unstable, due to loss of orthogonality.
- However, Gram-Schmidt can be stabilized using an augmented system of equations:
  1. Compute QR factorization of augmented matrix: \([Q,R1]=\text{mgs}([A,b])\)
  2. Extract \(R\) and \(\hat{Q}^*b\) from \(R1\): \(R=R1(1:n,1:n); \ Qb=R1(1:n,n+1)\)
  3. Back solve: \(x=R\backslash Qb\)

**Theorem**

The solution of the full-rank least squares problem by Gram-Schmidt orthogonality is backward stable in the sense that the computed solution \(\tilde{x}\) has the property

\[
\|(A + \delta A)\tilde{x} - b\| = \min, \quad \frac{\|\delta A\|}{\|A\|} = O(\epsilon_{\text{machine}})
\]

for some \(\delta A \in \mathbb{C}^{m \times n}\), provided that \(\hat{Q}^* b\) is formed implicitly.
Other Methods

- The method of *normal equation* solves \( x = (A^T A)^{-1} A^T b \)

**Theorem**

*The solution of the full-rank least squares problem via normal equation is unstable. Stability can be achieved, however, by restriction to a class of problems in which \( \kappa(A) \) is uniformly bounded above or \( (\tan \theta / \eta) \) is uniformly bounded below.*

- Solution using SVD: \( A = \hat{U} \hat{\Sigma} \hat{V}^*, \ x = \hat{V} \hat{\Sigma}^{-1} \hat{U}^* b \)

**Theorem**

*The solution of the full-rank least squares problem by the SVD is backward stable.*