AMS526: Numerical Analysis I (Numerical Linear Algebra)

Lecture 13: Conditioning of Least Squares Problems; Stability of Least Squares Algorithms

Xiangmin Jiao

SUNY Stony Brook

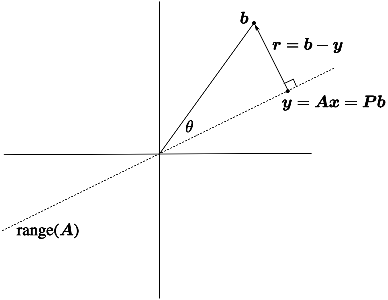
October 23, 2008

Four Conditioning Problems

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

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

- Its solution is $\boldsymbol{x} = \boldsymbol{A}^+\boldsymbol{b}$. Another quantity is $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} = \boldsymbol{P}\boldsymbol{b}$, where $\boldsymbol{P} = \boldsymbol{A}\boldsymbol{A}^+$

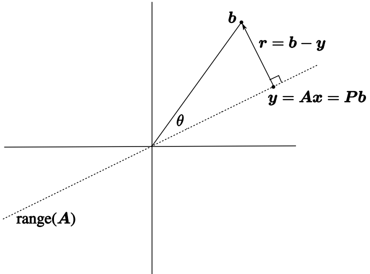


- Consider $\boldsymbol{A}$ and $\boldsymbol{b}$ as input data, and $\boldsymbol{x}$ and $\boldsymbol{y}$ as output. We then have four conditioning problems:

Input \ Output	$\boldsymbol{y}$	$\boldsymbol{x}$
$\boldsymbol{b}$	$\kappa_{\boldsymbol{b}\to\boldsymbol{y}}$	$\kappa_{\boldsymbol{b}\to\boldsymbol{x}}$
$\boldsymbol{A}$	$\kappa_{\boldsymbol{A}\to\boldsymbol{y}}$	$\kappa_{\boldsymbol{A}\to\boldsymbol{x}}$

Some Prerequisites

- These conditioning problems are important and subtle.
- We focus on the second column, namely $\kappa_{\boldsymbol{b}\to\boldsymbol{x}}$ and $\kappa_{\boldsymbol{A}\to\boldsymbol{x}}$
  - Understanding $\kappa_{\boldsymbol{b}\to\boldsymbol{y}}$ and $\kappa_{\boldsymbol{A}\to\boldsymbol{y}}$ is prerequisite
  - Note: Linear system $\boldsymbol{Ax} = \boldsymbol{b}$ for nonsingular $\boldsymbol{A} \in \mathbb{C}^{m\times m}$ is a special case of least squares problems, where $\boldsymbol{y} = \boldsymbol{b}$.
- We will focus on intuitive understanding instead of rigorous proof
- Three quantities (All norms are 2-norms)
  - Condition number of $\boldsymbol{A}$: $\kappa(\boldsymbol{A}) = \|\boldsymbol{A}\|\|\boldsymbol{A}^+\| = \sigma_1/\sigma_n$
  - Angle between $\boldsymbol{b}$ and $\boldsymbol{y}$: $\theta = \arccos \frac{\|\boldsymbol{y}\|}{\|\boldsymbol{b}\|}$
  - Orientation of $\boldsymbol{y}$ with range($\boldsymbol{A}$): $\eta = \frac{\|\boldsymbol{A}\|\|\boldsymbol{x}\|}{\|\boldsymbol{y}\|}$ (how far $\boldsymbol{y}$ is from $\boldsymbol{u}_1$, left singular vector of $\boldsymbol{A}$ corresponding to σ_1?)



Sensitivity of $\boldsymbol{y}$ to Perturbations in $\boldsymbol{b}$

- Intuition: The larger θ is, the more sensitive $\boldsymbol{y}$ is in terms of relative error
- Analysis: $\boldsymbol{y} = \boldsymbol{P}\boldsymbol{b}$, so

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

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

Input \ Output	$\boldsymbol{y}$	$\boldsymbol{x}$
$\boldsymbol{b}$	$\frac{1}{\cos\theta}$	
$\boldsymbol{A}$		

- Question: When the maximum is attained for perturbation $\delta\boldsymbol{b}$?

Sensitivity of $\boldsymbol{y}$ to Perturbations in $\boldsymbol{b}$

- Intuition: The larger θ is, the more sensitive $\boldsymbol{y}$ is in terms of relative error
- Analysis: $\boldsymbol{y} = \boldsymbol{P}\boldsymbol{b}$, so

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

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

Input \ Output	$\boldsymbol{y}$	$\boldsymbol{x}$
$\boldsymbol{b}$	$\frac{1}{\cos\theta}$	
$\boldsymbol{A}$		

- Question: When the maximum is attained for perturbation $\delta\boldsymbol{b}$?
- Answer: When $\delta\boldsymbol{b}$ is in range($\boldsymbol{A}$)

Sensitivity of $\boldsymbol{x}$ to Perturbations in $\boldsymbol{b}$

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

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

where $\eta = \|\boldsymbol{A}\|\|\boldsymbol{x}\|/\|\boldsymbol{y}\|$

Input \ Output	$\boldsymbol{y}$	$\boldsymbol{x}$
$\boldsymbol{b}$	$\frac{1}{\cos\theta}$	$\frac{\kappa(\boldsymbol{A})}{\eta\cos\theta}$
$\boldsymbol{A}$		

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

Sensitivity of $\boldsymbol{x}$ to Perturbations in $\boldsymbol{b}$

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

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

where $\eta = \|\boldsymbol{A}\|\|\boldsymbol{x}\|/\|\boldsymbol{y}\|$

Input \ Output	$\boldsymbol{y}$	$\boldsymbol{x}$
$\boldsymbol{b}$	$\frac{1}{\cos\theta}$	$\frac{\kappa(\boldsymbol{A})}{\eta\cos\theta}$
$\boldsymbol{A}$		

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

Sensitivity of $\boldsymbol{y}$ and $\boldsymbol{x}$ to Perturbations in $\boldsymbol{A}$

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

Input \ Output	$\boldsymbol{y}$	$\boldsymbol{x}$
$\boldsymbol{b}$	$\frac{1}{\cos\theta}$	$\frac{\kappa(\boldsymbol{A})}{\eta\cos\theta}$
$\boldsymbol{A}$	$\leq \frac{\kappa(\boldsymbol{A})}{\cos\theta}$	$\leq \kappa(\boldsymbol{A}) + \frac{\kappa(\boldsymbol{A})^2\tan\theta}{\eta}$

- For second row, bounds are not necessarily tight
- Assume $\cos\theta = O(1)$, $\kappa_{\boldsymbol{A}\to\boldsymbol{x}}$ can lie anywhere between $\kappa(\boldsymbol{A})$ and $O(\kappa(\boldsymbol{A})^2)$

Algorithms for Solving Least Squares Problems

There are many variants of algorithms for solving least squares problems

- Extremely unstable
  - Normal equations: solve $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b}$
- Unstable
  - Classical Gram-Schmidt
  - Modified Gram-Schmidt with explicit $\boldsymbol{Q}$
- Stable
  - Modified Gram-Schmidt with augmented system of equations with implicit $\boldsymbol{Q}$
  - Householder QR (with/without pivoting, explicit or implicit $\boldsymbol{Q}$)
  - Singular value decomposition
- Demo of different methods
- Note that in general, only SVD is robust for solving rank deficient least squares problems