



C S C I 5304

Fall 2021

COMPUTATIONAL ASPECTS OF MATRIX THEORY

Class time : MW 4:00 – 5:15 pm
Room : Keller 3-230 or Online
Instructor : Daniel Boley

Lecture notes: http://www-users.cselabs.umn.edu/classes/Fall-2021/csci5304/

August 27, 2021

# ERROR AND SENSITIVITY ANALYSIS FOR SYSTEMS OF LINEAR EQUATIONS

---

- Conditioning of linear systems.
- Estimating errors for solutions of linear systems
- (Normwise) Backward error analysis
- Estimating condition numbers ..

## *Perturbation analysis for linear systems ( $Ax = b$ )*

Question addressed by perturbation analysis: determine the variation of the solution  $x$  when the data, namely  $A$  and  $b$ , undergoes small variations. Problem is **Ill-conditioned** if small variations in data cause very large variation in the solution.

### *Setting:*

➤ We perturb  $A$  into  $A + E$  and  $b$  into  $b + e_b$ . Can we bound the resulting change (perturbation) to the solution?

*Preparation:* We begin with a lemma for a simple case

## *Rigorous norm-based error bounds*

**LEMMA:** If  $\|E\| < 1$  then  $I - E$  is nonsingular and

$$\|(I - E)^{-1}\| \leq \frac{1}{1 - \|E\|}$$

*Proof* is based on following 5 steps

a) Show: If  $\|E\| < 1$  then  $I - E$  is nonsingular

b) Show:  $(I - E)(I + E + E^2 + \dots + E^k) = I - E^{k+1}$ .

c) From which we get:

$$(I - E)^{-1} = \sum_{i=0}^k E^i + (I - E)^{-1} E^{k+1} \rightarrow$$

d)  $(I - E)^{-1} = \lim_{k \rightarrow \infty} \sum_{i=0}^k E^i$ . We write this as

$$(I - E)^{-1} = \sum_{i=0}^{\infty} E^i$$

e) Finally:

$$\begin{aligned} \|(I - E)^{-1}\| &= \left\| \lim_{k \rightarrow \infty} \sum_{i=0}^k E^i \right\| = \lim_{k \rightarrow \infty} \left\| \sum_{i=0}^k E^i \right\| \\ &\leq \lim_{k \rightarrow \infty} \sum_{i=0}^k \|E^i\| \leq \lim_{k \rightarrow \infty} \sum_{i=0}^k \|E\|^i \\ &\leq \frac{1}{1 - \|E\|} \end{aligned}$$

➤ Can generalize result:

**LEMMA:** If  $\mathbf{A}$  is nonsingular and  $\|\mathbf{A}^{-1}\| \|\mathbf{E}\| < 1$  then  $\mathbf{A} + \mathbf{E}$  is non-singular and

$$\|(\mathbf{A} + \mathbf{E})^{-1}\| \leq \frac{\|\mathbf{A}^{-1}\|}{1 - \|\mathbf{A}^{-1}\| \|\mathbf{E}\|}$$

- Proof is based on relation  $A + E = A(I + A^{-1}E)$  and use of previous lemma.
- Now we can prove the main theorem:

**THEOREM 1:** Assume that  $(A + E)y = b + e_b$  and  $Ax = b$  and that  $\|A^{-1}\| \|E\| < 1$ . Then  $A + E$  is nonsingular and

$$\frac{\|x - y\|}{\|x\|} \leq \frac{\|A^{-1}\| \|A\|}{1 - \|A^{-1}\| \|E\|} \left( \frac{\|E\|}{\|A\|} + \frac{\|e_b\|}{\|b\|} \right)$$



**Proof:** From  $(A + E)y = b + e_b$  and  $Ax = b$  we get  $(A + E)(y - x) = e_b - Ex$ . Hence:

$$y - x = (A + E)^{-1}(e_b - Ex)$$

Taking norms  $\rightarrow \|y - x\| \leq \|(A + E)^{-1}\| [\|e_b\| + \|E\|\|x\|]$   
 Dividing by  $\|x\|$  and using result of lemma

$$\begin{aligned} \frac{\|y - x\|}{\|x\|} &\leq \|(A + E)^{-1}\| [\|e_b\|/\|x\| + \|E\|] \\ &\leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\|E\|} [\|e_b\|/\|x\| + \|E\|] \\ &\leq \frac{\|A^{-1}\|\|A\|}{1 - \|A^{-1}\|\|E\|} \left[ \frac{\|e_b\|}{\|A\|\|x\|} + \frac{\|E\|}{\|A\|} \right] \end{aligned}$$

Result follows by using inequality  $\|A\|\|x\| \geq \|b\| \dots$

**QED**

The quantity  $\kappa(\mathbf{A}) = \|\mathbf{A}\| \|\mathbf{A}^{-1}\|$  is called the **condition number** of the linear system with respect to the norm  $\|\cdot\|$ . When using the  $p$ -norms we write:

$$\kappa_p(\mathbf{A}) = \|\mathbf{A}\|_p \|\mathbf{A}^{-1}\|_p$$

- Note:  $\kappa_2(\mathbf{A}) = \sigma_{max}(\mathbf{A}) / \sigma_{min}(\mathbf{A})$  = ratio of largest to smallest singular values of  $\mathbf{A}$ . Allows to define  $\kappa_2(\mathbf{A})$  when  $\mathbf{A}$  is not square.
- Determinant *\*is not\** a good indication of sensitivity
- Small eigenvalues *\*do not\** always give a good indication of poor conditioning.

**Example:** Consider, for a large  $\alpha$ , the  $n \times n$  matrix

$$A = I + \alpha e_1 e_n^T$$





➤ Inverse of  $A$  is :  $A^{-1} = I - \alpha e_1 e_n^T$  ➤ For the  $\infty$ -norm we have

$$\|A\|_\infty = \|A^{-1}\|_\infty = 1 + |\alpha|$$

so that

$$\kappa_\infty(A) = (1 + |\alpha|)^2.$$

➤ Can give a very large condition number for a large  $\alpha$  – but all the eigenvalues of  $A$  are equal to one.

-  1 Show that  $\kappa(I) = 1$  ;
-  2 Show that  $\kappa(A) \geq 1$  ;
-  3 Show that  $\kappa(A) = \kappa(A^{-1})$
-  4 Show that for  $\alpha \neq 0$ , we have  $\kappa(\alpha A) = \kappa(A)$

Simplification when  $e_b = 0$  :

$$\frac{\|x - y\|}{\|x\|} \leq \frac{\|A^{-1}\| \|E\|}{1 - \|A^{-1}\| \|E\|}$$

Simplification when  $E = 0$  :

$$\frac{\|x - y\|}{\|x\|} \leq \|A^{-1}\| \|A\| \frac{\|e_b\|}{\|b\|}$$

► Slightly less general form: Assume that  $\|E\|/\|A\| \leq \delta$  and  $\|e_b\|/\|b\| \leq \delta$  and  $\delta\kappa(A) < 1$  then

$$\frac{\|x - y\|}{\|x\|} \leq \frac{2\delta\kappa(A)}{1 - \delta\kappa(A)}$$

 Show the above result

*Another common form:*

**THEOREM 2:** Let  $(A + \Delta A)y = b + \Delta b$  and  $Ax = b$  where  $\|\Delta A\| \leq \epsilon \|E\|$ ,  $\|\Delta b\| \leq \epsilon \|e_b\|$ , and assume that  $\epsilon \|A^{-1}\| \|E\| < 1$ . Then

$$\frac{\|x - y\|}{\|x\|} \leq \frac{\epsilon \|A^{-1}\| \|A\|}{1 - \epsilon \|A^{-1}\| \|E\|} \left( \frac{\|e_b\|}{\|b\|} + \frac{\|E\|}{\|A\|} \right)$$

➤ Results to be seen later are of this type.

## *Normwise backward error*

➤ We solve  $Ax = b$  and find an approximate solution  $y$

*Question:* Find smallest perturbation to apply to  $A, b$  so that \*exact\* solution of perturbed system is  $y$

## *Normwise backward error in just $A$ or $b$*

Suppose we model entire perturbation in RHS  $b$ .

- Let  $r = b - Ay$  be the residual.  
Then  $y$  satisfies  $Ay = b + \Delta b$  with  $\Delta b = -r$  exactly.
- The relative perturbation to the RHS is  $\frac{\|r\|}{\|b\|}$ .

Suppose we model entire perturbation in matrix  $A$ .

- Then  $y$  satisfies  $\left(A + \frac{ry^T}{y^T y}\right) y = b$
- The relative perturbation to the matrix is

$$\left\| \frac{ry^T}{y^T y} \right\|_2 / \|A\|_2 = \frac{\|r\|_2}{\|A\| \|y\|_2}$$



## *Normwise backward error in both $A$ & $b$*

For a given  $\mathbf{y}$  and given perturbation directions  $\mathbf{E}$ ,  $\mathbf{e}_b$ , we define the **Normwise backward error**:

$$\eta_{\mathbf{E}, \mathbf{e}_b}(\mathbf{y}) = \min\{\epsilon \mid (\mathbf{A} + \Delta\mathbf{A})\mathbf{y} = \mathbf{b} + \Delta\mathbf{b};$$

where  $\Delta\mathbf{A}, \Delta\mathbf{b}$  satisfy:  $\|\Delta\mathbf{A}\| \leq \epsilon\|\mathbf{E}\|;$   
and  $\|\Delta\mathbf{b}\| \leq \epsilon\|\mathbf{e}_b\|\}$

In other words  $\eta_{\mathbf{E}, \mathbf{e}_b}(\mathbf{y})$  is the smallest  $\epsilon$  for which

$$(1) \begin{cases} (\mathbf{A} + \Delta\mathbf{A})\mathbf{y} = & \mathbf{b} + \Delta\mathbf{b}; \\ \|\Delta\mathbf{A}\| \leq \epsilon\|\mathbf{E}\|; & \|\Delta\mathbf{b}\| \leq \epsilon\|\mathbf{e}_b\| \end{cases}$$

- $\mathbf{y}$  is given (a computed solution).  $\mathbf{E}$  and  $\mathbf{e}_b$  to be selected (most likely 'directions of perturbation for  $\mathbf{A}$  and  $\mathbf{b}$ ').
- Typical choice:  $\mathbf{E} = \mathbf{A}$ ,  $\mathbf{e}_b = \mathbf{b}$

 Explain why this is not unreasonable

Let  $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{y}$ . Then we have:


**THEOREM 3:** 
$$\eta_{\mathbf{E}, \mathbf{e}_b}(\mathbf{y}) = \frac{\|\mathbf{r}\|}{\|\mathbf{E}\|\|\mathbf{y}\| + \|\mathbf{e}_b\|}$$

Normwise backward error is for case  $\mathbf{E} = \mathbf{A}$ ,  $\mathbf{e}_b = \mathbf{b}$ :

$$\eta_{\mathbf{A}, \mathbf{b}}(\mathbf{y}) = \frac{\|\mathbf{r}\|}{\|\mathbf{A}\|\|\mathbf{y}\| + \|\mathbf{b}\|}$$



Show how this can be used in practice as a means to stop some iterative method which computes a sequence of approximate solutions to  $Ax = b$ .

 Consider the  $6 \times 6$  Vandermonde system  $Ax = b$  where  $a_{ij} = j^{2(i-1)}$ ,  $b = A * [1, 1, \dots, 1]^T$ . We perturb  $A$  by  $E$ , with  $|E| \leq 10^{-10}|A|$  and  $b$  similarly and solve the system. Evaluate the backward error for this case. Evaluate the forward bound provided by Theorem 2. Comment on the results.

## *Estimating condition numbers.*

- Often we just want to get a lower bound for condition number [it is 'worse than ...']
- We want to estimate  $\|A\| \|A^{-1}\|$ .
- The norm  $\|A\|$  is usually easy to compute but  $\|A^{-1}\|$  is not.
- We want: Avoid the expense of computing  $A^{-1}$  explicitly.

### *Idea:*

- Select a vector  $v$  so that  $\|v\| = 1$  but  $\|Av\| = \tau$  is small.
- Then:  $\|A^{-1}\| \geq 1/\tau$  (show why) and:

$$\kappa(A) \geq \frac{\|A\|}{\tau}$$

- Condition number worse than  $\|A\|/\tau$ .
- Typical choice for  $v$ : choose  $[\dots \pm 1 \dots]$  with signs chosen on the fly during back-substitution to maximize the next entry in the solution, based on the upper triangular factor from Gaussian Elimination.
- Similar techniques used to estimate condition numbers of large matrices in matlab.

## Condition numbers and near-singularity

- $1/\kappa \approx$  relative distance to nearest singular matrix.

Let  $A, B$  be two  $n \times n$  matrices with  $A$  nonsingular and  $B$  singular. Then

$$\frac{1}{\kappa(A)} \leq \frac{\|A - B\|}{\|A\|}$$

Proof:  $B$  singular  $\rightarrow \exists x \neq 0$  such that  $Bx = 0$ .

$$\begin{aligned}\|x\| &= \|A^{-1}Ax\| \leq \|A^{-1}\| \|Ax\| = \|A^{-1}\| \|(A - B)x\| \\ &\leq \|A^{-1}\| \|A - B\| \|x\|\end{aligned}$$

Divide both sides by  $\|x\| \times \kappa(A) = \|x\| \|A\| \|A^{-1}\|$  ➤ result.  
QED.

**Example:**

$$\text{let } A = \begin{pmatrix} 1 & 1 \\ 1 & 0.99 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$\text{Then } \frac{1}{\kappa_1(A)} \leq \frac{0.01}{2} \Rightarrow \kappa_1(A) \geq \frac{2}{0.01} = 200.$$

➤ It can be shown that (Kahan)

$$\frac{1}{\kappa(A)} = \min_B \left\{ \frac{\|A - B\|}{\|A\|} \mid \det(B) = 0 \right\}$$



## *Estimating errors from residual norms*

Let  $\tilde{x}$  an approximate solution to system  $Ax = b$  (e.g., computed from an iterative process). We can compute the residual norm:

$$\|r\| = \|b - A\tilde{x}\|$$

Question: How to estimate the error  $\|x - \tilde{x}\|$  from  $\|r\|$ ?

- One option is to use the inequality

$$\frac{\|x - \tilde{x}\|}{\|x\|} \leq \kappa(A) \frac{\|r\|}{\|b\|}.$$

- We must have an estimate of  $\kappa(A)$ .

### *Proof of inequality.*

First, note that  $A(x - \tilde{x}) = b - A\tilde{x} = r$ . So:

$$\|x - \tilde{x}\| = \|A^{-1}r\| \leq \|A^{-1}\| \|r\|$$

Also note that from the relation  $b = Ax$ , we get

$$\|b\| = \|Ax\| \leq \|A\| \|x\| \quad \rightarrow \quad \|x\| \geq \frac{\|b\|}{\|A\|}$$

Therefore,

$$\frac{\|x - \tilde{x}\|}{\|x\|} \leq \frac{\|A^{-1}\| \|r\|}{\|b\|/\|A\|} = \kappa(A) \frac{\|r\|}{\|b\|} \quad \square$$

 Show that

$$\frac{\|x - \tilde{x}\|}{\|x\|} \geq \frac{1}{\kappa(A)} \frac{\|r\|}{\|b\|}.$$