ERROR AND SENSITIVTY ANALYSIS FOR SYSTEMS OF LINEAR EQUATIONS

- Read parts of sections 2.6 and 3.5.3
- Conditioning of linear systems.
- Estimating errors for solutions of linear systems
- Backward error analysis
- Relative element-wise error analysis

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.

- Let E, be an $n \times n$ matrix and e be an n-vector.
- "Perturb" A into $A(\epsilon) = A + \epsilon E$ and b into $b + \epsilon e$.
- Note: $A + \epsilon E$ is nonsingular for ϵ small enough.
- Why?
- The solution $x(\epsilon)$ of the perturbed system is s.t.

$$(A + \epsilon E)x(\epsilon) = b + \epsilon e.$$

- Let $\delta(\epsilon) = x(\epsilon) - x$. Then,

$$(A + \epsilon E)\delta(\epsilon) = (b + \epsilon e) - (A + \epsilon E)x = \epsilon \ (e - Ex)$$
$$\delta(\epsilon) = \epsilon \ (A + \epsilon E)^{-1}(e - Ex).$$

- $x(\epsilon)$ is differentiable at $\epsilon = 0$ and its derivative is

$$x'(0) = \lim_{\epsilon \to 0} \frac{\delta(\epsilon)}{\epsilon} = A^{-1}(e - Ex).$$

- A small variation $[\epsilon E, \epsilon e]$ will cause the solution to vary by roughly $\epsilon x'(0) = \epsilon A^{-1}(e - Ex)$.
- The relative variation is such that

$$\frac{\|x(\epsilon)-x\|}{\|x\|} \le \epsilon \|A^{-1}\| \left(\frac{\|e\|}{\|x\|} + \|E\| \right) + O(\epsilon^2).$$

- Since $\|b\| \le \|A\|\|x\|$:

$$\frac{\|x(\epsilon)-x\|}{\|x\|} \le \epsilon \|A\|\|A^{-1}\| \left(\frac{\|e\|}{\|b\|} + \frac{\|E\|}{\|A\|} \right) + O(\epsilon^2)$$

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

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

- Note: $\kappa_2(A) = \sigma_{max}(A)/\sigma_{min}(A)$ = ratio of largest to smallest singular values of A. Allows to define $\kappa_2(A)$ when 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 α, 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 ∞-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 α – but all the eigenvalues of A_n are equal to one.

Rigorous norm-based error bounds

➤ First need to show that $A + E$ is nonsingular if A is nonsingular and E is small. Begin with simple case:

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

c) From which we get:

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

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

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

e) Finally:

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

➤ Can generalize result:

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

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

Proof is based on relation $A + E = A(I + A^{-1}E)$ and use of previous lemma.

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

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

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

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

Taking norms $\to$ $\|y-x\| \le \|(A+E)^{-1}\| \, [\|e\| + \|E\|\|x\|]$
Dividing by $\|x\|$ and using result of lemma

$$\frac{\|y-x\|}{\|x\|} \le \|(A+E)^{-1}\| \, [\|e\|/\|x\| + \|E\|]$$
$$\le \frac{\|A^{-1}\|}{1-\|A^{-1}\|\|E\|} [\|e\|/\|x\| + \|E\|]$$
$$\le \frac{\|A^{-1}\|\|A\|}{1-\|A^{-1}\|\|E\|} \left[\frac{\|e\|}{\|A\|\|x\|} + \frac{\|E\|}{\|A\|}\right]$$

Result follows by using inequality $\|A\|\|x\| \ge \|b\|$.... **QED**

Simplification when $e = 0$:

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

Simplification when $E = 0$:

$$\frac{\|x-y\|}{\|x\|} \le \|A^{-1}\| \, \|A\| \frac{\|e\|}{\|b\|}$$

➤ Slightly less general form: Assume that $\|E\|/\|A\| \le \delta$ and $\|e\|/\|b\| \le \delta$ and $\delta\kappa(A) < 1$ then

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

Another common form:

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

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

Normwise backward error

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

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

For a given y and given perturbation directions E, e, we define the Normwise backward error:

$$\eta_{E,e}(y) = \min\{\epsilon \mid (A+\Delta A)y = b + \Delta b;$$
$$\text{for all } \Delta A, \Delta b \quad \text{satisfying:} \quad \|\Delta\ A\| \le \epsilon\|E\|;$$
$$\text{and} \quad \|\Delta\ b\| \le \epsilon\|e\|\}$$

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

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

➤ y is given (a computed solution). E and e to be selected (most likely 'directions of perturbation for A and b').

➤ Typical choice: $E = A$, $e = b$

✍ Explain why this is not unreasonable

Let $r = b - Ay$. Then we have:

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

Normwise backward error is for case $E = A, e = b$:

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

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 6×6 Vandermonde system $Ax = b$ where $a_{ij} = j^{2(i-1)}$, $b = A * [1, 1, \cdots, 1]^T$. We perturb A by E, with $|E| \le 10^{-10}|A|$ 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.

Proof of Theorem 3

Let $D \equiv \|E\|\|y\| + \|e\|$ and $\eta \equiv \eta_{E,e}(y)$. The theorem states that $\eta = \|r\|/D$. Proof in 2 steps.

First: Any $\delta A, \delta b$ pair satisfying (1) is such that $\epsilon \ge \|r\|/D$. Indeed from (1) we have (recall that $r = b - Ay$)

$$Ay + \Delta Ay = b + \Delta b \to r = \Delta Ay - \Delta b \to$$

$$\|r\| \le \|\Delta A\|\|y\| + \|\Delta b\| \le \epsilon(\|E\|\|y\| + \|e\|) \to \epsilon \ge \frac{\|r\|}{D}$$

Second: We need to show an instance where the minimum value of $\|r\|/D$ is reached. Take the pair $\Delta A, \Delta b$:

$$\Delta A = \alpha r z^T; \quad \Delta b = \beta r \quad \text{with } \alpha = \frac{\|E\|\|y\|}{D}; \quad \beta = \frac{\|e\|}{D}$$

The vector z depends on the norm used - for the 2-norm: $z = y/\|y\|^2$. Here: Proof only for 2-norm

a) We need to verify that first part of (1) is satisfied:

$$\begin{aligned}(A + \Delta A)y &= Ay + \alpha r \frac{y^T}{\|y\|^2} y = b - r + \alpha r \\ &= b - (1 - \alpha)r = b - \left(1 - \frac{\|E\|\|y\|}{\|E\|\|y\| + \|e\|}\right) r \\ &= b - \frac{\|e\|}{D} r = b + \beta r \quad \to \\ (A + \Delta A)y &= b + \Delta b \quad \leftarrow \text{The desired result}\end{aligned}$$

b) Finally: Must now verify that $\|\Delta A\| = \eta\|E\|$ and $\|\Delta b\| = \eta\|e\|$. Exercise: Show that $\|uv^T\|_2 = \|u\|_2\|v\|_2$

$$\|\Delta A\| = \frac{|\alpha|}{\|y\|^2}\|ry^T\| = \frac{\|E\|\|y\|}{D}\frac{\|r\|\|y\|}{\|y\|^2} = \eta\|E\|$$

$$\|\Delta b\| = |\beta|\|r\| = \frac{\|e\|}{D}\|r\| = \eta\|e\| \qquad QED$$

Componentwise backward error

A few more definitions on norms...

- A norm is absolute $\||x|\| = \|x\|$ for all x. (satisfied by all p-norms).
- A norm is monotone if $|x| \le |y| \to \|x\| \le \|y\|$.
- It can be shown that these two properties are equivalent.

Show: a function which satisfies the first 2 requirements of vector norms (1. $\phi(x) \ge 0$ (==0, iff $x = 0$) and 2. $\phi(\lambda x) = |\lambda|\phi(x)$) satisfies the triangle inequality iff its unit ball is convex.

(Continued) Use the above to construct a norm in $\mathbb{R}^2$ that is *not* absolute.

✍ Define absolute *matrix* norms in same way. Which of the norms $\|A\|_1, \|A\|_\infty, \|A_2\|$, and $\|A\|_F$ are absolute?

✍ Recall that for any matrix $fl(A) = A + E$ with $|E| \le \underline{\mathbf{u}}\,|A|$. For an absolute matrix norm

$$\frac{\|E\|}{\|A\|} \le \underline{\mathbf{u}}$$

What does this imply?

- Component-wise analysis requires that we use norms that are *absolute*
- We will restrict analysis to $\|.\|_\infty$
- See sec. 2.6.5 of text.

- Analogue of theorem 2 for case $E = |A|, e = |b|$:

THEOREM 4 Let $Ax = b$ and $(A + \Delta A)y = b + \Delta b$ where $|\Delta A| \le \epsilon|A|$ and $|\Delta b| \le \epsilon|b|$. Assume that $\epsilon\kappa_\infty(A) = r < 1$. Then $A + \Delta A$ is nonsingular and

$$\frac{\|x - y\|_\infty}{\|x\|_\infty} \le \frac{2\epsilon}{1 - r}\||A^{-1}|\ |A|\|_\infty$$

- Componentwise relative condition number :

$$\kappa_\infty^C(A) \equiv \|\ |A^{-1}|\ |A|\ \|_\infty$$

✍ Redo example seen after Theorem 3, (6×6 Vandermonde system) using componentwise analysis.

Componentwise backward error for $y \equiv$ is the smallest ϵ for which

$$(2) \begin{cases} (A + \Delta A)y = & b + \Delta b; \\ |\Delta\ A| \le \epsilon E; & |\Delta b| \le \epsilon e \end{cases}$$

Denoted by $\omega_{E,e}(y)$.

THEOREM 5 [Oettli-Prager] Let $r = b - Ay$ (residual). Then

$$\omega_{E,e}(y) = \max_i \frac{|r_i|}{(E|y| + e)_i}.$$

Zero denominator case: $0/0 \equiv 0$ and nonzero/ $0 \equiv \infty$

Example of ill-conditioning: The Hilbert Matrix

- Notorious example of ill conditioning.

$$H_n = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{n+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & & \cdots & \frac{1}{2n-1} \end{pmatrix} \quad \text{i.e.,} \quad h_{ij} = \frac{1}{i + j - 1}$$

- For $n = 5$ $\kappa_2(H_n) = 4.766.. \times 10^5$.
- Let $b_n = H_n(1, 1, \ldots, 1)^T$.
- Solution of $H_n x = b$ is $(1, 1, \ldots, 1)^T$.
- Let $n = 5$ and perturb $h_{5,1} = 0.2$ into 0.20001.
- New solution: $\boxed{(0.9937, 1.1252, 0.4365, 1.865, 0.5618)^T}$

Estimating condition numbers.

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

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

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

$$\|x\| = \|A^{-1}Ax\| \le \|A^{-1}\|\ \|Ax\| = \|A^{-1}\|\|(A-B)x\|$$
$$\le \|A^{-1}\|\ \|A-B\|\|x\|$$

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

Example:

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

Then $\frac{1}{\kappa_1(A)} \le \frac{0.01}{2}$ ➤ $\kappa_1(A) \ge 200$.

➤ It can be shown that (Kahan)

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

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

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

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

Proof of inequality.

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

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

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

$$\|b\| = \|Ax\| \le \|A\|\ \|x\| \quad \rightarrow \quad \|x\| \ge \frac{\|b\|}{\|A\|}$$

Therefore,

$$\frac{\|x-\tilde{x}\|}{\|x\|} \le \frac{\|A^{-1}\|\ \|r\|}{\|b\|/\|A\|} = \kappa(A)\frac{\|r\|}{\|b\|} \qquad \blacksquare$$

✍ Show that

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

THEOREM 6 Let A be a nonsingular matrix and $\tilde{x}$ an approximate solution to $Ax = b$. Then for any norm $\|.\|$,

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

In addition, we have the relation

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

in which $\kappa(A)$ is the condition number of A associated with the norm $\|.\|$.

Iterative refinement

- Define residual vector:

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

- We have seen that: $x - \tilde{x} = A^{-1}r$, i.e., we have

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

- Idea: Compute r accurately (double precision) then solve

$$A\delta = r$$

... and correct $\tilde{x}$ by

$$\tilde{x} := \tilde{x} + \delta$$

... repeat if needed.

- Read Section 3.5.3 for details