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.
We perturb $A$ into $A + E$ and $b$ into $b + e_b$. Can we bound the perturbation to the solution?

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

\[
\| (I - E)^{-1} \| = \left\| \lim_{k \to \infty} \sum_{i=0}^{k} E^i \right\| = \lim_{k \to \infty} \left\| \sum_{i=0}^{k} E^i \right\| \\
\leq \lim_{k \to \infty} \sum_{i=0}^{k} \| E^i \| \leq \lim_{k \to \infty} \sum_{i=0}^{k} \| E \|^i \\
\leq \frac{1}{1 - \| E \|}
\]
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.

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 → \[\|y - x\| \leq \|(A + E)^{-1}\| \left(\|e_b\| + \|E\|\|x\|\right)\]

Dividing by \(\|x\|\) and using result of lemma

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

\[ \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\||E\|} \left(\|e_b\|/\|x\| + \|E\|\right) \]

\[ \leq \frac{\|A^{-1}\||A\|}{1 - \|A^{-1}\||E\|} \left(\|e_b\|/\|x\| + \|E\|\right) \]

Result follows by using inequality \(\|A\|\|x\| \geq \|b\|\).... QED
The quantity \( \kappa(A) = \| A \| \| A^{-1} \| \) is called the condition number of the linear system with respect to the norm \( \| \cdot \| \). When using the \( p \)-nормs we write:

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

Note: \( \kappa_2(A) = \sigma_{\text{max}}(A)/\sigma_{\text{min}}(A) \) = 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.
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.
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
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 $\left( A + \frac{ry^T}{y^Ty} \right) y = b$
- The relative perturbation to the matrix is
  \[
  \frac{\|ry^T\|}{\|y^Ty\|_2} / \|A\|_2 = \frac{\|r\|_2}{\|A\| \|y\|_2}
  \]
For a given $y$ and given perturbation directions $E$, $e_b$, we define the Normwise backward error:

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

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

$$(1) \ \begin{cases} (A + \Delta A)y = b + \Delta b; \\
\|\Delta A\| \leq \epsilon \|E\|; \quad \|\Delta b\| \leq \epsilon \|e_b\| 
\end{cases}$$


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

- Typical choice: $E = A$, $e_b = b$

_canvas

- Explain why this is not unreasonable

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

**THEOREM 3:**

$$
\eta_{E,e_b}(y) = \frac{\|r\|}{\|E\|\|y\| + \|e_b\|}
$$

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

$$
\eta_{A,b}(y) = \frac{\|r\|}{\|A\|\|y\| + \|b\|}
$$
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 $6 \times 6$ Vandermonde system $Ax = b$ where $a_{ij} = j^{2(i-1)}$, $b = A \ast [1, 1, \cdots, 1]^T$. 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.
Proof of Theorem 3

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

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

\[
Ay + \Delta Ay = b + \Delta b \rightarrow r = \Delta Ay - \Delta b \rightarrow
\]

\[
\|r\| \leq \|\Delta A\|\|y\| + \|\Delta b\| \leq \epsilon(\|E\|\|y\| + \|e_b\|) \rightarrow \epsilon \geq \frac{\|r\|}{D}
\]

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

\[
\Delta A = \alpha rz^T; \quad \Delta b = \beta r \quad \text{with } \alpha = \frac{\|E\|\|y\|}{D}; \quad \beta = \frac{\|e_b\|}{D}
\]
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:

$$(A + \Delta A)y = Ay + \alpha r \frac{y^T}{\|y\|^2}y = b - r + \alpha r$$

$$= b - (1 - \alpha)r = b - \left(1 - \frac{\|E\|\|y\|}{\|E\|\|y\| + \|e_b\|}\right)r$$

$$= b - \frac{\|e_b\|}{D}r = b + \beta r \quad \rightarrow$$

$$(A + \Delta A)y = b + \Delta b \quad \leftarrow \text{The desired result}$$
Finally:  

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

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

\[
\| \Delta b \| = |\beta| \| r \| = \frac{\| e_b \|}{D} \| r \| = \eta \| e_b \| \quad QED
\]
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 $[\cdots \pm 1 \cdots]$ 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 \( Bx = 0 \).

\[
\|x\| = \|A^{-1}Ax\| \leq \|A^{-1}\| \|Ax\| = \|A^{-1}\| \|(A - B)x\|
\]

\[
\leq \|A^{-1}\| \|A - B\| \|x\|
\]

Divide both sides by \( \|x\| \times \kappa(A) = \|x\| \|A\| \|A^{-1}\| \Rightarrow \) 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 \( \frac{1}{\kappa_1(A)} \leq \frac{0.01}{2} \quad \Rightarrow \quad \kappa_1(A) \geq \frac{2}{0.01} = 200. \)

\[\Rightarrow \quad \text{It can be shown that (Kahan)} \]

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

5-23
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 \( \|r\| \)?

\( \Rightarrow \) One option is to use the inequality

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

\( \Rightarrow \) 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,

\[
\frac{\|x - \tilde{x}\|}{\|x\|} \leq \frac{\|A^{-1}\| \|r\|}{\|b\|/\|A\|} = \kappa(A) \frac{\|r\|}{\|b\|}
\]

\[\Box\]

Show that

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