Error and Sensitivty Analysis for Systems of Linear Equations Matrix Computations — CPSC 5006 E

Julien Dompierre

Department of Mathematics and Computer Science
Laurentian University

Sudbury, October 6, 2010

Outline

- Read sections 2.7, 3.3, 3.4, 3.5.
- Conditioning of linear systems.
- Estimating accuracy.
- Error analysis.

Perturbation Analysis (p. 80)

Consider a linear system $Ax = b$. The question addressed by perturbation analysis is to determine the variation of the solution x when the data, namely A and b, undergoes small variations. A problem is **ill-conditioned** if small variations in the data lead to very large variation in the solution.

Let E, be an $n \times n$ matrix and e be an n-vector.

"Perturb" A into $A(\varepsilon) = A + \varepsilon E$ and b into $b + \varepsilon e$.

Note: $A + \varepsilon E$ is non singular for ε small enough.
(Why?)

The solution $x(\varepsilon)$ of the perturbed system is such that

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

Perturbation Analysis (continued) (p. 81)

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

$$\begin{aligned} (A+\varepsilon E)\delta(\varepsilon) &= (b+\varepsilon e) - (A+\varepsilon E)x = \varepsilon\ (e - Ex) \\ \delta(\varepsilon) &= \varepsilon\ (A+\varepsilon E)^{-1}(e - Ex). \end{aligned}$$

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

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

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

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

Since $\|b\| \le \|A\|\|x\|$, we get

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

Condition Number (p. 81)

The quantity $\kappa(A) = \|A\| \, \|A^{-1}\|$ is called the **condition number** of the linear system with respect to the norm $\|\cdot\|$, with the convention that $\kappa(A) = \infty$ for singular A. When using the standard norms $\|\cdot\|_p$, $p = 1, \ldots, \infty$, we label $\kappa(A)$ with the same label as the associated norm. Thus,

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

Ill-Conditioned Matrices (p. 82)

The previous inequality can be written as

$$\frac{\|x(\varepsilon) - x\|}{\|x\|} \leq \kappa(A)(\rho_A + \rho_b) + O(\varepsilon^2)$$

where

$$\rho_A = \varepsilon \frac{\|E\|}{\|A\|} \qquad \rho_b = \varepsilon \frac{\|e\|}{\|b\|}$$

represent the relative error in A and b respectively. Thus the relative error in x can be $\kappa(A)$ times the relative error in A and b.

If $\kappa(A)$ is large, then A is said to be an **ill-conditioned** matrix. Matrix with small condition numbers are said to be **well-conditioned**.

$\kappa(A) \geq 1$ for all (submultiplicative and unitary) matrix norm.

Condition Number (p. 81)

The condition number $\kappa(\cdot)$ depends on the underlying norm and subscript used. In particular, with the 2-norm, we have

$$\kappa_2(A) = \|A\|_2 \, \|A^{-1}\|_2 = \frac{\sigma_1(A)}{\sigma_n(A)}$$

which is the ratio of largest to smallest singular values of A. This is also a measure of the elongation of the hyperellipsoid given by the set $\{Ax \text{ such that } \|x\|_2 = 1\}$.

$\kappa_2(A)$ is defined when A is not square.

$\kappa_2(Q) = 1$ if Q is an orthogonal matrix.

Any two condition numbers $\kappa_\alpha(\cdot)$ and $\kappa_\beta(\cdot)$ on $\mathbb{R}^{n\times n}$ are equivalent in that constants c_1 and c_2 can be found for which

$$c_1\kappa_\alpha(A) \leq \kappa_\beta(A) \leq c_2\kappa_\alpha(A)$$

for all A in $\mathbb{R}^{n\times n}$. For example, we have

$$\begin{aligned}
\frac{1}{n}\kappa_2(A) \leq \kappa_1(A) \leq n\kappa_2(A),\\
\frac{1}{n}\kappa_\infty(A) \leq \kappa_2(A) \leq n\kappa_\infty(A),\\
\frac{1}{n^2}\kappa_1(A) \leq \kappa_\infty(A) \leq n^2\kappa_1(A).
\end{aligned}$$

If a matrix is ill-conditioned in the α-norm, it is ill-conditioned in the β-norm modulo the constants c_1 and c_2 above.

Equivalence of Matrix Norms (p. 56)

These equivalences of condition numbers (previous slide) are a consequence of the equivalence of matrix norm.
The Frobenius and p-norms (especially $p = 1, 2, \infty$) satisfy certain inequalities that are frequently used in analysis of matrix computations. For $A \in \mathbb{R}^{m \times n}$ we have

$$\begin{aligned}
\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\,\|A\|_2 \\
\max_{i,j} |a_{ij}| \leq \|A\|_2 \leq \sqrt{mn}\,\max_{i,j} |a_{ij}| \\
\frac{1}{\sqrt{n}}\|A\|_\infty \leq \|A\|_2 \leq \sqrt{m}\,\|A\|_\infty \\
\frac{1}{\sqrt{m}}\|A\|_1 \leq \|A\|_2 \leq \sqrt{n}\,\|A\|_1
\end{aligned}$$

It is natural to consider how well determinant size measure ill-conditioning. If $\det(A) = 0$ is equivalent to singularity, then $\det(A) \approx 0$ equivalent to near singularity? No, determinant is **not** a good indication of sensitivity. Consider

$$B_n = \begin{bmatrix} 1 & -1 & -1 & \cdots & -1 \\ 0 & 1 & -1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & 1 & -1 \\ 0 & 0 & & \cdots & 1 \end{bmatrix}$$

The matrix B_n has determinant 1, but $\kappa_\infty(B_n) = n\,2^{n-1}$.

Determinants and Nearness of Singularity

On the other hand, a very well conditioned matrix can have very small determinant. For example

$$D_n = \mathrm{diag}(0.1, 0.1, \ldots, 0.1) = \begin{bmatrix} 0.1 & 0 & \cdots & 0 \\ 0 & 0.1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0.1 \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

The condition number $\kappa_p(D_n) = 1$ although $\det(N_n) = 10^{-n}$.

Eigenvalues and Condition Number

Small eigenvalues **do not** always give a good indication of poor conditioning. Consider matrices of the form

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

for large α. The inverse of A_n is

$$A_n^{-1} = I - \alpha e_1 e_n^T$$

and for the ∞-norm (maximum absolute value row sum) we have

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

so that

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

For a large α, this can give a very large condition number, whereas all the eigenvalues of A_n are equal to unity.

Lemma 2.3.3 (p. 58)

If $F \in \mathbb{R}^{n\times n}$ and $\|F\|_p < 1$, then $I - F$ is non singular and

$$(I - F)^{-1} = \sum_{k=0}^{\infty} F^k$$

with

$$\|(I - F)^{-1}\|_p \leq \frac{1}{1 - \|F\|_p}.$$

Rigorous Norm-Based Error Bounds (p. 83)

First need to show that $A + E$ is non singular if A is non singular and E is small:

LEMMA (p. 58): If A is non singular and $\|A^{-1}\| \; \|E\| < 1$ then $A + E$ is non-singular and

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

Theorem (1)

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

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

Proof

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

$$\begin{aligned} y - x &= A^{-1}[(e - Ex) - (E(y-x))] \to \\ \|y - x\| &\le \|A^{-1}\| \|(e - Ex) - (E(y-x))\| \\ &\le \|A^{-1}\| \left[\|e - Ex\| + \|E\| \|y - x\| \right] \end{aligned}$$

So $\|y - x\|(1 - \|A^{-1}\| \|E\|) \le \|A^{-1}\| [\|e\| + \|E\| \|x\|]$

....

Note: stated in a slightly weaker form in text. 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* $\|\Delta\ A\| \leq \varepsilon \|E\|$, $\|\Delta\ b\| \leq \varepsilon \|e\|$, *and assume that* $\varepsilon \|A^{-1}\| \|E\| < 1$. *Then*

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

Norm-wise Backward Error (p. 84)

Question: How much do we need to perturb data for an approximate solution y to be the exact solution of the perturbed system?

Norm-wise backward error for y is defined as smallest ε for which

$$(A + \Delta A)y = b + \Delta b; \quad \|\Delta\ A\| \leq \varepsilon \|E\|; \quad \|\Delta\ b\| \leq \varepsilon \|e\|$$

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

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

Typical choice: $E = A$, $e = b$

(Explain why this is not unreasonable)

Norm-wise Backward Error

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

Theorem (3)

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

Norm-wise 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 $|E| \leq 10^{-10}|A|$ 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.)

Component-wise 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 $|x| \leq |y| \rightarrow \|x\| \leq \|y\|$.

It can be shown that these two properties are equivalent.
... and some notation:

$$\begin{aligned} \omega_{E,e}(y) = \min\{\varepsilon \mid (A + \Delta A)y &= b + \Delta b, \\ |\Delta A| &\leq \varepsilon E, \quad |\Delta b| \leq \varepsilon e\} \end{aligned}$$

(where $E \geq 0, e \geq 0$) is the component-wise backward error.

Absolute Norm

Show: a function which satisfies the first two requirements of vector norms: (1) $\phi(x) \geq 0$ ($\phi(x) = 0$ 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 the same way. Which of the usual norms $\|A\|_1, \|A\|_\infty, \|A_2\|$, and $\|A\|_F$ are absolute?

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

$$\frac{\|E\|}{\|A\|} \leq \mathbf{u}$$

What does this imply?

Theorem of Oettli-Prager (p. 85)

Theorem (4 Oettli-Prager)

Let $r = b - Ay$ (residual). Then

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

zero denominator case:
$0/0 \to 0$
nonzero/ $0 \to \infty$

Analogue of theorem 2:

Theorem (5)

Let $Ax = b$ and $(A + \Delta A)y = b + \Delta b$ where $|\Delta A| \le \varepsilon E$ and $|\Delta b| \le \varepsilon e$. Assume that $\varepsilon \| |A^{-1}|E \| \le 1$, where $\| \cdot \|$ is an absolute norm. Then,

$$\frac{\|x - y\|}{\|x\|} \le \frac{\varepsilon}{1 - \varepsilon \| |A^{-1}|E \|} \frac{\| |A^{-1}|(E|x| + e) \|}{\|x\|}$$

In addition, equality achieved to order ε for infinity norm.

Implication

$$\limsup_{\varepsilon \to 0} \left\{ \frac{\|\Delta x\|_\infty}{\varepsilon \|x\|_\infty} : (A + \Delta A)(x + \Delta x) = b + \Delta b \right\}$$

is equal to

$$\text{cond}_{E,e}(A, x) \equiv \frac{\| |A^{-1}| (E|x| + e) \|_\infty}{\|x\|_\infty}$$

Condition number depends on x (i.e. on right-hand side b)
Special case $E = |A|, e = 0$ yields

$$\text{cond}(A, x) \equiv \frac{\| \ |A^{-1}| \ |A| \ |x| \ \|_\infty}{\|x\|_\infty}$$

Component-wise condition number :

$$\text{cond}(A) \equiv \| \ |A^{-1}| \ |A| \ \|_\infty$$

(Redo example seen after Theorem 3, (6×6 Vandermonde system) using component-wise analysis.)

Example of Ill-Conditioning: The Hilbert Matrix

Notorious example of ill conditioning.

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

For $n = 5$, $\kappa_2(H_n) = 4.766.. \times 10^5$.

Let $b_n = H_n \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$.

Solution of $H_n x = b$ is $\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$.

Let $n = 5$ and perturb $h_{5,1} = 0.2$ into 0.20001.

New solution: $\boxed{\begin{bmatrix} 0.9937 & 1.1252 & 0.4365 & 1.876 & 0.5618 \end{bmatrix}^T}$

Estimating Condition Numbers (p. 81)

Theorem

Let A, B be two $n \times n$ matrices with A non singular 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 $\|x\|$ $\kappa(A) = \|x\|\|A\|\;\|A^{-1}\|$ □

Example: Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 0.99 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Then

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

In fact it is true that

$$\frac{1}{\kappa_p(A)} = \min_{B \text{ s.t. } \det(B)=0} \frac{\|A-B\|_p}{\|A\|_p}$$

This result may be found in Kahan (1966) and shows that $\kappa_p(A)$ measures the relative p-norm distance from A to the set of singular matrices.

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,

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

(Show that

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

)

Theorem 6 (p. 81)

Theorem (6)

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

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

In addition, we have the relation

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

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

It can be shown that

$$\frac{1}{\kappa(A)} = \min_{B} \left\{ \frac{\|A - B\|}{\|A\|} \quad \text{such that } \det(B) = 0 \right\}$$

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.

Algorithm: Iterative Refinement

while *true*
  Compute $r = b - A\tilde{x}$
  Solve $A\delta = r$
  Compute $\tilde{x} := \tilde{x} + \delta$
  if $\|\delta\| \geq \varepsilon\|\tilde{x}\|$ **then** break
end

Why does this work? Model: each solution gets m digits at most because of the conditioning: For example 3 digits. At the first iteration, the error is roughly $\approx 0.001 \times \|b\|$.

Second iteration: error in δ is roughly $0.001 \times \|r\|$, but now $\|r\|$ is much smaller than $\|b\|$.
etc ...

Assume residual is computed exactly. Backward error analysis:

$$(A + F_k)\delta_k = r_k \quad \rightarrow \quad x_{k+1} = x_k + (A + F_k)^{-1} r_k$$

So: $r_{k+1} = b - Ax_{k+1} = \ldots = F_k(A + F_k)^{-1} r_k \rightarrow$

$$\|r_{k+1}\| \le \|F_k\| \|(A + F_k)^{-1}\| \|r_k\|$$

A previous result showed that if $\|F_k\| \; \|A^{-1}\| < 1$ then

$$\|F_k\| \|(A + F_k)^{-1}\| \le \frac{\|F_k\| \|A^{-1}\|}{1 - \|F_k\| \; \|A^{-1}\|}$$

So : process will converge if (suff. condition)

$$\|F_k\| \|A^{-1}\| \le \gamma < \frac{1}{2}$$

Iterative Refinement

Important: Iterative refinement won't work when the residual r consists only of noise. When $b - Ax$ is already very small ($\approx \varepsilon$) it is likely to be just noise, so not much can be done because

$$\delta = A^{-1}\text{noise}$$

Heuristic: If $\varepsilon = 10^{-d}$, and $\kappa_\infty(A) \approx 10^q$ then each iterative refinement step will gain about $d - q$ digits.

Iterative Refinement. A Scilab Experiment

```
n = 6
A = Hilbert( 6 )
b = A * ones(n,1)
inv(A)*b

B = A
B(6,1) = B(6,1) + 0.000001
x = inv(B) * b
x_exact = ones(n,1)
error = norm( x_exact - x, 2)

residue = b - A*x
correction = inv(B)*residue
x = x + correction
error = norm( x_exact - x, 2)

residue = b - A*x
correction = inv(B)*residue
x = x + correction
error = norm( x_exact - x, 2)
```

Repeat a couple of times...

Observation: We gain about 3 digits per iteration.

(Read Section 3.5.4 on estimating accuracy.)