Condition Numbers and Perturbations

When we solve a linear system or a linear least squares problem, we determine the solution $X$ as a function of the data $(A, b)$.

i.e.

\[ X = f(A, b) \]

For a linear system $Ax = b$

\[ f(A, b) = A^{-1}b \]

For a least squares problem

\[ f(A, b) = (A^T A)^{-1} A^T b \]

If, for some reason, we actually have

\[ \hat{X} = f(\hat{A}, \hat{b}) \]

for some perturbed data

\[ \hat{A} = A + E, \quad \hat{b} = b + d \]
We are very interested in how well \( \hat{x} \) approximates \( x \). How close is \( \hat{x} \) to \( x \)?

Our tool for answering this question is the norm. (Recall: any norm will do. Some are more convenient than others.)

We want to know

\[ \| \hat{x} - x \| \] (Absolute error)

However, this measure may be misleading. Let's think of \( \hat{x} \) and \( x \) a scalars for the moment.

\( \hat{x} = (1.235) \times 10^5 \), \( x = (1.234) \times 10^5 \)

\[ \| \hat{x} - x \| = (0.001) \times 10^5 = 1 \]

Which doesn't look very small. However, we see that the three leading digits of \( \hat{x} \) and \( x \) agree and this would be true if we replaced \( 10^5 \) by \( 10^p \) for any integer \( p \).

Is there any way to discover this?

Consider

\[ \frac{\| \hat{x} - x \|}{\| x \|} \] (Relative error)
We can compute this and estimate
\[
\frac{\|x_1 - x_1\|}{\|x_1\|} = \frac{(0.0001) \times 10^5}{0.1234 \times 10^5} = 8 \times 10^{-4}
\]

This tells us that at least 3 leading digits agree. We also see this works for any scaling \(10^p\) in place of \(10^5\).

\[
\frac{\|x_2 - x_2\|}{\|x_2\|} = \frac{(0.0001) \times 10^p}{0.1234 \times 10^5} = \frac{10^{-4}}{1.234} \approx 8 \times 10^{-4}
\]

The bound \(8 \times 10^{-4}\) says \(x_1\) and \(x_2\) differ in the 4-th place beyond the decimal and the leading 3 digits agree regardless of the scaling \(10^p\).

Obviously, relative error is what we want.

\textbf{Caution:} For vectors \(x, \hat{x}\) this measure only gives information about the largest (in magnitude) components.

Example: \(x = \begin{pmatrix} 123.5 \times 10^5 \\ 2222 \end{pmatrix}, \hat{x} = \begin{pmatrix} 1234 \times 10^5 \\ 1111 \end{pmatrix}\)

\[
\frac{\|x - \hat{x}\|_\infty}{\|x\|_\infty} = \frac{(0.0001) \times 10^5}{1234 \times 10^5} = 8 \times 10^{-4}
\]

as before. But, the second component
has no leading digits in agreement with the second component of x.

For linear systems, and least squares problems, this is what we would like, but it is really a general problem, not just for computation, but for any situation where there is uncertainty in the data.

Eventually, we shall be primarily concerned with perturbations due to round off error. However, the mathematics of perturbation theory is not concerned about the source of the perturbations. Don’t confuse the two notions. Round off error is just one possible source of perturbation.

In general, we might consider the problem of evaluating a vector valued function

\[ f(a) \]

In linear systems, think of a

as a vector representing all of the parameters \((A,b)\).
Given a perturbed \( \hat{a} = a + \Delta a \), we want to get an upper bound for

\[
\frac{\| f(\hat{a}) - f(a) \|}{\| f(a) \|} = \frac{\| \hat{a} - a \|}{\| a \|}
\]

in terms of the relative perturbation to the data

\[
\frac{\| \hat{a} - a \|}{\| a \|} = \frac{\| \Delta a \|}{\| a \|}
\]

If we put \( \hat{x} = f(\hat{a}) \), \( x = f(a) \), we get

\[
\frac{\| \hat{x} - x \|}{\| x \|} = \frac{\| f(\hat{a}) - f(a) \|}{\| f(a) \|}
\]

We want an expression of the form

\[
\frac{\| f(\hat{a}) - f(a) \|}{\| f(a) \|} \leq K(a) \frac{\| \hat{a} - a \|}{\| a \|}
\]

That is valid for all \( \hat{a} \) that are reasonably close to \( a \).
If \( K(a) \frac{\| \hat{a} - a \|}{\| a \|} = c \| a \|^p \)

where \( c \approx 10 \) then the leading \( p-1 \) digits of the largest components of \( \hat{x} = f(\hat{a}) \) will agree with those of \( x = f(a) \).

\( K(a) \) is the condition number with respect to evaluating \( f \) at \( a \).

How do we find \( K(a) \) for a given function \( f \)?

First consider a scalar function of a scalar argument \( a \).

Example: \( f(a) = \tan(a) \), \( -\frac{\pi}{2} < a < \frac{\pi}{2} \)

If \( f \) is twice differentiable at \( a \), then

\[
f(\hat{a}) = f(a) + f'(a)(\hat{a} - a) + O(\| \hat{a} - a \|^2)
\]

\[
\frac{\| f(\hat{a}) - f(a) \|}{\| f(a) \|} = \left( \frac{\| f'(a) \| \| a \|}{\| f(a) \|} \right) \frac{\| \hat{a} - a \|}{\| a \|} + O(\| \hat{a} - a \|)
\]

\[
\geq K(a) \frac{\| \hat{a} - a \|}{\| a \|}
\]
When $f$ is a vector-valued function of a vector argument $a$, we have

$$f(\tilde{a}) = f(a) + J(a)(\tilde{a} - a) + o(\|\tilde{a} - a\|^2)$$

as before, with $J(a)$ the Jacobian of $f$ evaluated at $a$.

$$J(a)_{ij} = \frac{\partial f_i}{\partial a_j}(a)$$

and

$$k(a) = \frac{\|J(a)\|_F}{\|f(a)\|}$$

where $\|\cdot\|_F$ is the matrix norm induced by the vector norm $\|\cdot\|$.

A more general definition of $k(a)$ is given in TrB p. 90.

Note: They talk about $f(x)$ and I always find this confusing when we start talking about $x = f(A, b)$. 
Condition of Linear Systems

To understand the condition number associated with a linear system $Ax = b$ with $A$ square and nonsingular, we could work with the general definition we just gave. This is done in T&B. We shall take a more traditional approach here. We shall let $\| \cdot \|$ stand for any of the vector, and corresponding induced $p$-norms.

Perturbations in RHS

Lemma: Let $Ax = b$, $A\hat{x} = \hat{b}$

Then

\[
\frac{\|\hat{x} - x\|}{\|x\|} \leq \kappa(A) \frac{\|\hat{b} - b\|}{\|b\|}
\]

where $\kappa(A) = \|A\|\|A^{-1}\|$.

Proof: $\hat{x} - x = A^{-1}(\hat{b} - b) = A^{-1}(\hat{b} - b)$

Thus $\|\hat{x} - x\| \leq \|A^{-1}\| \|\hat{b} - b\| (1.1)$.

We also have $\|b\| = \|Ax\| \leq \|A\| \|x\|$

so $\frac{1}{\|x\|} \leq \frac{\|A\|}{\|b\|} (1.2)$

Combining (1.1), 2

\[
\frac{1}{\|x\|} \|\hat{x} - x\| \leq \frac{\|A\|\|\hat{b} - b\|}{\|b\|} \leq \frac{\kappa(A)\|\hat{b} - b\|}{\|b\|}
\]
Let \( f(A, b) = A^{-1}b \)

In the context of what we had before,

\[ x = f(A, b), \quad \hat{x} = f(A, \hat{b}) \]

The perturbations are only in the RHS \( b \)
We see that the condition number only depends on \( A \).

**Lemma 2.** For any given \( \hat{x} \), form

The linear system residual \( b - Ax \)
Let \( Ax = \hat{b} \). Then

\[ \frac{\| \hat{x} - x \|}{\| x \|} \leq \kappa(A) \frac{\| b - Ax \|}{\| b \|} \]

**Proof:** Put \( \hat{b} = Ax \) in the previous lemma. \( \Box \)

**Interpretation:** If \( \hat{x} \) is any approximation to the solution \( x \) to \( Ax = b \), you can tell how well \( \hat{x} \) approximates \( x \) in a relative sense. If you know \( \kappa(A) \) and \( b - Ax \)

<table>
<thead>
<tr>
<th>Condition Number</th>
<th>Linear System Residual</th>
</tr>
</thead>
</table>


Very Important: A small linear system residual does not assure an accurate answer \( \hat{x} \). You must know \( \kappa(A) \), the condition number.

If (as is often the case):

\[
\frac{\|b - Ax\|}{\|b\|} \approx 10^{-16}
\]

and

\[
\kappa(A) = 10^d
\]

Then

\[
\frac{\|\hat{x} - x\|}{\|x\|} \leq 10^d (10^{-16}) = 10^{-(16-d)}
\]

Remember what we found before:

Components of largest magnitude will agree up to the \( p \)-th place after the decimal (assuming normalization).

where \( p = 16 - d \).
Perturbations to $A$

What if we have the same RHS but allows perturbations to $A$.

Suppose \[(A + E)\hat{x} = b, \quad Ax = b\]

We could use Lemma 1 again.

Let \[\hat{b} = b - EX\hat{x}\]. From Lemma 1

\[
\frac{\|\hat{x} - x\|}{\|x\|} \leq K(A) \frac{\|\hat{b} - b\|}{\|b\|} = K(A) \frac{\|E\hat{x}\|}{\|x\|}
\]

This result is true but we have no idea of how large $\|E\|_2$ might be. Note: this is different than the situation in Lemma 2. Here we were given $\hat{x}$. Here $\hat{x}$ has been computed and depends on $A$ and the perturbation $E$.

We can do much better with this...
Note

\[(A+E)\hat{x} = b \Rightarrow (I + A^{-1}E)\hat{x} = A^{-1}b = x\]

If \( I + A^{-1}E \) is invertible, then

\[\hat{x} = (I + A^{-1}E)^{-1}x\]

To work with this we need to know something about inverses of perturbations to the identity \( I + F \).

**Lemma 3.** If \( \|F\| < 1 \), then

1. \( I + F \) is nonsingular

2. \( \| (I + F)^{-1} \| \leq \frac{1}{1 - \|F\|} \).

**Proof:** First we show \( I + F \) is nonsingular by proving \( x \neq 0 \Rightarrow (I + F)x \neq 0 \).

If \( x \neq 0 \), then

\[
\| (I + F)x \| \geq \|x\| - \|Fx\| \quad \text{(triangle )}
\]

\[
\geq \|x\| - \|F\|\|x\| \quad \text{(\( \|F\| \leq \|x\|\))}
\]

\[
= (1 - \|F\|)\|x\| > 0,
\]
Since $I + F$ is non-singular, we have

\[ 1 = \| I \| = \| (I + F)^{-1} (I + F) \| = \| (I + F)^{-1} + (I + F)^{-1} F \| \geq \| (I + F)^{-1} \| - \| (I + F)^{-1} F \| \geq \| (I + F)^{-1} \| - \| (I + F)^{-1} I \| \| F \| = \| (I + F)^{-1} \| (1 - \| F \|) \]

Dividing on both sides gives

\[ \| (I + F)^{-1} \| \leq \frac{1}{1 - \| F \|} \]

We have specified $\| \cdot \|$ to be any p-norm.

The properties we used are

1) $\| M x \| \leq \| M \| \cdot \| x \|$

2) $\| A B U \| \leq \| A \| \cdot \| B \| \cdot \| U \|$

3) $\| I \| = 1$

Thus, we could easily be more general about the norms we are using.

Any (consistent) set of matrix and vector norms satisfying 1, 2, 3 would work.
Let's return to the problem at hand.

\textbf{Lemma 4} Suppose \( A \) nonsingular with
\[
(A + E)x = b, \quad Ax = b
\]

If \( \|A^{-1}\|\|E\| < 1 \), then

\[
\frac{\|\hat{x} - x\|}{\|x\|} \leq \frac{\|A^{-1}E\|}{1 - \|A^{-1}E\|} = \frac{\kappa(A) \|E\|}{\|A\|} \frac{\|E\|}{\|A\|}
\]

\textbf{Proof} \( \|A^{-1}E\| \leq \|A^{-1}\|\|E\| < 1 \), so \( I + A^{-1}E \) is nonsingular, and we have already seen that

\[
\hat{x} = (I + A^{-1}E)^{-1}x
\]

\[
\hat{x} - x = [(I + A^{-1}E)^{-1}I]x
\]

\[
= [I - (I + A^{-1}E)](I + A^{-1}E)^{-1}x
\]

\[
= -A^{-1}E(I + A^{-1}E)^{-1}x
\]

Thus

\[
\|\hat{x} - x\| \leq \|A^{-1}E\|\|I + A^{-1}E\|\|x\|
\]

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

From Lemma 3.
Thus

\[
\frac{\|\hat{x} - x\|}{\|x\|} \leq \frac{\|Ae\|}{\|x\|} = \frac{K(A)\|e\|}{\|x\|} = \frac{K(A)\|e\|}{\|x\| (1 - K(A)\|e\|/\|x\|)}
\]

Since \(\|A^{-1}e\| \leq \|A^{-1}\|\|e\| = (\|AA^{-1}\| - 1)/\|x\|\),

\[\|e\| \leq 10^{-16}\|x\|\]

Note: If \(\|e\| \leq 10^{-16}\|x\|\) (perturbations at roundoff level)

and \(K(A) = 10^d\), this says

1) \(\|A^{-1}\| \leq 10^{-16}\) \(K(A) = 10^{-(16-d)}\)

Our result is valid if \(d < 16\)

2) \(\frac{\|\hat{x} - x\|}{\|x\|} \leq \frac{10^{-16}}{1 - 10^{-16-d}} \approx 10^{-(16-d)}\)

Thus, when \(e\) is roughly the order of roundoff in (e.g. 16-bit arithmetic
double precision on most machines today)

Then there are roughly \(16-d\) digits of accuracy in \(\hat{x}\) as an
approximation to $x$, $K(A) = 10^d$

means you lose $d$ digits of accuracy due to perturbations in $A$.

**Surprise**: The result is good for any.

RHS $b$ !

It is not much more difficult to consider perturbations to $A$ and $b$
at the same time. We'll post some

this till later though.
Condition of the Linear Least Squares

Problem \( \min \| b - Ax \|_2 \).

A Partial Result: With very little extra work, we can gain a partial understanding of the perturbation theory for LS. We only consider the full rank case. Mathematically, it is OK to write the solution as

\[
x = (A^TA)^{-1} A^Tb = A^E b
\]

Lemma 5: Let \( A \in \mathbb{R}^{m \times n}, \ m \geq n \). Suppose \( \hat{x} \) solves \( \min \| b - A\hat{x} \|_2 \), \( x \) solves \( \min \| b - Ax \|_2 \).

Then

\[
\frac{\| \hat{x} - x \|_2}{\| x \|_2} \leq K(A) \frac{\| b - b \|_2}{\| Pb \|_2} = \frac{K(A)}{\cos(\theta)} \frac{\| b \|_2}{\| b \|_2}
\]

where \( K(A) = \| A^E \|_2 / \| A \|_2 \).

\( P \) is an orthogonal projector onto \( \text{Range}(A) \).

\( \cos(\theta) = \| Pb \|_2 / \| b \|_2 \) cosine of angle \( \theta \) that \( b \) makes with \( \text{Range}(A) \).
Proof: Just as with the linear system, we have
\[ \hat{x} - \hat{x} = A^T \hat{b} - A^T b = A^T (\hat{b} - b) \]

Thus
\[ ||\hat{x} - \hat{x}||_2 \leq ||A^T||_2 ||\hat{b} - b||_2 \]

From our discussion of LS, we know \( \hat{x} \) solves \( \min ||b - A\hat{x}||_2 \) if \( L \)

\[ A\hat{x} = \hat{b} \]

where \( \hat{b} \) is the orthogonal projection of \( b \)
onto \( \text{Range}(A) \)

Thus
\[ ||\hat{b}||_2 = ||A\hat{x}||_2 \leq ||A||_2 ||\hat{x}||_2 \]

so
\[ \frac{1}{||\hat{A}||} \leq \frac{||\hat{b}||_2}{||\hat{b}||_2} \]

and we have
\[ \frac{||\hat{x} - \hat{x}||_2}{||A||_2 ||\hat{x}||_2} \leq \left( \frac{||A^T||_2 ||A||_2}{||\hat{b}||_2} \right) \frac{||\hat{b} - b||_2}{||\hat{b}||_2} \]

\[ = \frac{\cos(\theta)}{\cos(\theta)} \frac{||\hat{b} - b||_2}{||\hat{b}||_2} \]

where \( \cos(\theta) = \frac{||\hat{b}||_2}{||\hat{b}||_2} \)
Interpretation:

Note that our condition number \( \frac{K(A)}{\cos(\theta)} \)
now depends on the R& S b
relative to Range(A).

First note when A is square and non-sing.

Then
\[
A^T = (A^TA)^{-1} A^T = A^T A^{-1} A^T = A^{-1}, \quad \text{and} \quad b \in \text{Range}(A)
\]

Also note when b \in \text{Range}(A) then P b = b, \quad \cos(\theta) = 1
(even for m \geq n)

So the result for linear systems is recurrent
in both cases.

However, when m > n and b \not\in \text{Range}(A)
then \cos(\theta) comes into play and we see the problem becomes
more ill-conditioned as b
becomes more nearly orthogonal to Range(A).

This phenomenon can be appreciated
geometrically.
All of the vectors \( b_1, b_2, b_3, b_4 \ldots \) have the same projection \( Ax \) onto \( \text{Range}(A) \). But, as the angle approaches \( \frac{\pi}{2} \), the vector \( b \) is becoming arbitrarily large.

Even though \( \frac{\|b - b\|}{\|b\|} < \varepsilon \) might hold

\[
\cos(\theta) = \frac{\|Ax\|}{\|b\|} = \frac{\|Pb\|}{\|b\|} \quad \text{becomes}
\]

arbitrarily small. The projection of the possible perturbations becomes larger and larger with respect to the fixed \( \|Ax\| \).
Singular Value and the 2-Norm
condition number.

Our definition of condition number depends on choice of norm.

If we have $A = QR$ (short form)

Then $A^T = R^*Q^*$ (With $R$, this could be computed by back-substitution)

$$K(A) = \| A^T \| \| A \|$$

This is something we could actually compute for

$\| \cdot \|_1 = \| \cdot \|_1$ or $\| \cdot \|_1 = \| \cdot \|_2$

But it is very expensive. For

$\| \cdot \|_1 = \| \cdot \|_2$

we need the SVD.
From the short form of SVD

\[ A = USV^* \]
\[ U^*U = V^*V = I_n \]
\[ S = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \]
\[ \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0 \]

we have

\[ A^T = VS^{-1}U^* \]

and we can see that

\[ \|A^T\|_2 = \frac{1}{\sigma_n} \]

(\(\frac{1}{\sigma_n}\) is the largest singular value of \(A^T\))

Thus

\[ \|AA^T\|_2 = \frac{\sigma_1}{\sigma_n} \]

and of course this reduces to

\[ \|AA^T\|_2 = \frac{\sigma_1}{\sigma_n} \text{ when } A \text{ is square.} \]

The 2-norm condition number

\[ \chi_2(A) = \frac{\sigma_1}{\sigma_n} \]

The ratio of largest to smallest singular values
Condition of the Linear Least Squares Problem

We shall complete the story by developing bounds for perturbations in $A$. To do this, we shall need the following results.

**Lemma 6** If $A = QR$ (short form)

Then $K_2(A) = K_2(R)$

**Lemma 7** $K_2(A)^2 = \| (A^T A)^{-1} \|_2 \| A \|_2^2$

**Lemma 8** Suppose $(I + F)Z = g$.

If $\| E \|_2 \leq \frac{1}{10}$, then $\| Z \|_2 \leq C_0 \| g \|_2$, $C_0 = \frac{10}{9}$

**Lemma 9** If $R_1 = R + E$, $\| R_1^{-1} E \|_2 \leq \frac{1}{10}$

Then $\| R_1^{-1} \| \leq C_0 \| R^{-1} \|$

**Lemma 10** $\| (R^T R)^{-1} E E^T \|_2 \leq (K_2(R))^2 \left( \frac{\| E \|_2}{\| R \|_2} \right)^2$

**Proof:** 6, 7 follow from SVD, 8, 9 from Lemma 3

10 follows from 7
Theorem. Let $A \in \mathbb{R}^{m \times n}$, $m \leq n$ be full rank.

Suppose $(A+E)\hat{x} = b$, $Ax = x$.

If $K_e(A)\frac{\|E\|_2}{\|A\|_2} \leq \frac{1}{10}$, then

\[
\frac{\|\hat{x} - x\|_2}{\|x\|_2} \leq \frac{\|E\|_2}{\|A\|_2} \left[ K_e(A) + K_e^2(A) \tan \theta \right]
\]

where $c \leq 3$, $\theta = \angle (A,b)$.

Proof: Let $A = QR$ (SVD form).

Put $E = QE_0 + E_1$, $E_0 = QT E$, $E_1 = (I - QT)E$.

Let $R_1 = R + E_0$. Then $A + E = QR_1 + E_1$.

Let $F = (R^T R_1)^{-1} E_1 E_1$, $G = R^{-1} E_0$.

From Lemmas 6-10

\[
\|F\|_2 \leq \frac{1}{c} K(A)^2 \left( \frac{\|E\|_2}{\|A\|_2} \right)^2, \quad \|G\|_2 \leq K(A) \frac{\|E\|_2}{\|A\|_2}.
\]

Let $b_1 = QT b$, $b_2 = (I - QT)b$. 

Now,
\[
(A+E^T(A+E))\hat{x} = (A+E)^Tb
\]
\[
(R^T R + E^T E)\hat{x} = R^T Q^Tb + E^Tb
\]
Then
\[
(I+E)\hat{x} = R_1^{-1} Q^Tb + (R_2 R_1^{-1} E^T)b_2
\]
\[
= (I+G)^{-1} R_1^{-1} Q^Tb + g_2
\]
Thus
\[
(I+F)(\hat{x} - x) = [(I+G)^{-1} - I]x + g_2 - Fx
\]
\[
= -G(I+G)^{-1}x + g_2 - Fx
\]
\[
= g_1 + g_2 + g_3
\]

Lemma’s 5+10 ⇒
\[
\frac{\|x - x_1\|_2}{\|x\|_2} \leq C_0 \left[ \frac{\|z_1\|_2}{\|x\|_2} + \frac{\|z_2\|_2}{\|x\|_2} + \frac{\|z_3\|_2}{\|x\|_2} \right]
\]

Now,
1. \[
\frac{\|z_1\|_2}{\|x\|_2} \leq C_0 \|g_1\|_2 \leq \epsilon K(R) \frac{\|E\|}{\|A\|} = C_0 K(A) \frac{\|E\|}{\|A\|}
\]
2. \[
\frac{\|z_2\|_2}{\|x\|_2} \leq \frac{\|R_1^{-1}E\|_2 \|b_2\|_2}{\|x\|_2} \leq C_0^2 K(A) \frac{\|E\|}{\|A\|} \frac{\|b_2\|_2}{\|x\|_2}
\]
3. \[
\frac{\|z_3\|_2}{\|x\|_2} = \frac{\|Fx\|_2}{\|x\|_2} \leq C_0 \left( K(A) \frac{\|E\|}{\|A\|} \right)^2
\]

The result follows from (11), (12).
The important terms in this derivation are the expressions for vectors $g_1$, $g_2$:

$$g_1 = [(I+G)^{-1} - I]x = -G(I+G)^{-1}x$$

where $G = R^TE_0 = R^TE$ is analogous to the result for linear systems.

The term

$$g_2 = (R^TR)^{-1}e^Tb = (R^TR)^{-1}e^T(I-QQ^T)b = (R^TR)^{-1}e^Tb_2$$

arises from the relation between $b$ and range ($A$). Ultimately we end up with

$$\frac{\|g_2\|^2}{\|b\|^2} \leq C_a \frac{x^T(A)^2(x^T(A)^2 + (\frac{\|e\|^2}{\|A\|^2}) \tan (\theta))}{\|e\|^2}$$

This term vanishes when $b \in \text{range}(A)$ (i.e., when $\theta = 0$) and gives the linear system result when $A$ is non-

The rest of the discussion just involves the detail needed to relate everything.
to the original quantities $x, A, E, GR, b$.

The full Monty

The results for perturbations to $b$ only and to $A$ only can be combined, either directly or by use of the triangle inequality together with lemmas 6-10, to give

Theorem: Let $A$ be full rank and let

$$
\varepsilon = \max \left[ \frac{11111}{11111111}, \frac{11111111}{11111111} \right], \quad \frac{11111111}{11111111} < 10
$$

Let

$$(A+E)^{\ast}x = b+f, \quad A^{\ast}x = b.$$

Then

$$
\frac{\| \hat{x} - x \|_2}{\| x \|_2} \leq \varepsilon \left[ \frac{K(A)}{\cos(\theta)} + K_2(A)^2 \tan \theta \right]
$$

where $C$ is a modest constant ($C=O(1)$) and

$$
\theta = \angle(A, b).
$$
Sensitivity of the Full Rank L.S. Problem

Lemma: \( \| A \|_2 (A^T A)^{-1} A^T \|_2 = \frac{\sigma_1}{\sigma_n} = \kappa_2(A) \)

\( \| A \|_2 (A^T A)^{-1} \|_2 = \| (A^T A) \|^{-1/2} \| (A^T A)^{-1/2} \| = \left( \frac{\sigma_1}{\sigma_n} \right)^2 = \kappa^2_2(A) \)

Proof: Exercise

Theorem: Suppose \( x, r \) and \( \tilde{x}, \tilde{r} \) satisfy

\[ A x = b \quad r = b - A x \]

\[ A \tilde{x} = \tilde{b} \quad \tilde{r} = \tilde{b} - A \tilde{x} \]

where \( A, \tilde{A} \in \mathbb{R}^{m \times n}, \ m \geq n, \ 0 \neq b, \ \tilde{b} \in \mathbb{R}^m \).

Assume \( \varepsilon = \max \left\{ \frac{\| \tilde{A} - A \|_2}{\| A \|_2}, \frac{\| \tilde{b} - b \|_2}{\| b \|_2} \right\} < \frac{\sigma_1}{\sigma_n} \)

and \( 1 \neq \sin(\theta) = \frac{\| b - A x \|}{\| b \|} \quad \cos(\theta) = \frac{\| A x \|}{\| b \|} \)

Then

\[ \frac{\| \tilde{x} - x \|_2}{\| x \|_2} \leq \varepsilon \left( \frac{2 \kappa_2(A)}{\cos(\theta)} + \frac{\tan(\theta) \kappa_2(A)^2}{\cos(\theta)} + O(\varepsilon^2) \right) \]
Proof: Put \( E = (\hat{A} - A) \hat{\varepsilon} \), \( f = (b^T - b) \hat{\varepsilon} \).

The hypothesis implies \( \|\hat{A} - A\| < \varepsilon \) and thus \( \text{rank}(A + tE) = n \) for \( t \in [0, \varepsilon] \).

Also, the solution \( x(t) \) to

\[
(A + tE)^T(A + tE)x(t) = (A + tE)^T(b + tf)
\]

is continuously differentiable for all \( t \in [0, \varepsilon] \).

Hence,

\[
x(0) = x(0) + t \dot{x}(0) + O(t^2),
\]

and

\[
\dot{x} = x + \varepsilon \dot{x}(0) + O(\varepsilon^2), \quad x \neq 0.
\]

Thus

\[
\|\hat{x} - x\| \leq \varepsilon \frac{\|x(0)\|}{\|x\|} + O(\varepsilon^2),
\]

A bound on \( \|x(0)\| \) may be obtained by differentiating \((x)\) and setting \( t = 0 \).

Differentiation gives

\[
[(A + tE)^T E + E^T (A + tE)]x(t) + (A + tE)^T (A + tE)x(t) =
\]

\[
E^T (b + tf) + (A + tE)^T f.
\]

Setting \( t = 0 \) gives

\[
(A^T E + E^T A)x + A^T A x(0) = E^T b + A^T f.
\]
Thus
\[ x(0) = (A^T A)^{-1} E^T (b - Ax) + A^T (f - Ex) \]
\[ = (A^T A)^{-1} A^T (f - Ex) + (A^T A)^{-1} E^T r \]
The definition of \( E \) implies \( \| E \| \leq \| b \| \) and \( \| E \| \leq \| b \| \).
Thus
\[ \frac{\| y(0) \|}{\| x \|} \leq \| (A^T A)^{-1} A^T \| \| A \| \left( \frac{\| b \|}{\| b \| \| x \|} + 1 \right) \]
\[ + \| (A^T A)^{-1} \| \| A \| \| z \| \leq \frac{\| z \|}{\| A^T A \|} \leq \kappa_2(A) \left( \frac{2}{\cos(\theta)} \right) + \kappa_2(A)^2 \frac{\sin(\theta)}{\cos(\theta)} \]
Since \( \| A \| \| x \| \geq \| Ax \| \) and \( \frac{1}{\cos(\theta)} > 1 \),
This establishes the desired bound.

Ref: See Thm 5.3.1, p242, Golub & VanLoan, 3rd ed.