Numerical Matrix Analysis

Notes #14 — Conditioning and Stability: Least Squares Problems: Conditioning

Peter Blomgren
⟨blomgren.peter@gmail.com⟩

Department of Mathematics and Statistics
Dynamical Systems Group
Computational Sciences Research Center
San Diego State University
San Diego, CA 92182-7720

http://terminus.sdsu.edu/

Spring 2020

Outline

1. Recap
   - Backward Stability
2. Least Squares Problems
   - Introduction: Projection, Pseudo-Inverse
   - Conditioning
   - Dimensionless Parameters: $\kappa(A)$, θ, and η
3. Conditioning of LSQ Problems
   - Theorem
   - SVD Trickery
   - Proof

Recap: Last Time — Backward Stability of Back-Substitution

We looked at a backward stability proof in gory detail. — The technique is quite straight-forward, albeit somewhat tedious.

- We replace the floating point operators $\oplus$, $\ominus$, $\otimes$, and $\oslash$ with exact mathematical operations + relative error terms, *i.e.* $x \oplus y \rightsquigarrow (x+y)(1+\epsilon)$, where $|\epsilon| \leq \epsilon_{\text{mach}}$.
- Then we **interpret** the error as perturbations on the appropriate part of the problem formulation (so that that computed solution is the exact solution to a nearby problem).

Recap: Last Time

As we used the backward substitution algorithm for the detailed backward **stability** proof; we now turn to the least squares problems for a detailed discussion on **conditioning**...

...and we recall that `Accuracy(conditioning,stability)`, so these are all important pieces in the larger *"numerics jigsaw puzzle."*

Rewind (Computational Accuracy)

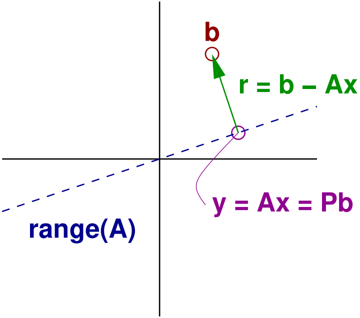
Suppose a backward stable algorithm is applied to solve a problem $f : X \mapsto Y$ with condition number κ in a floating point environment satisfying the floating point representation axiom, and the fundamental axiom of floating point arithmetic.

Then the relative errors satisfy

$$\frac{\|\tilde{f}(x) - f(x)\|}{\|f(x)\|} = \mathcal{O}(\kappa(x)\epsilon_{\text{mach}}).$$

Least Squares Problems...

Once again, we return to the least squares problem.



We measure everything in the two-norm, and let $\|\cdot\| = \|\cdot\|_2$; formally we are trying to solve

Given $A \in \mathbb{C}^{m\times n}$ of full rank, $m \geq n$, $\vec{b} \in \mathbb{C}^m$, find $\vec{x} \in \mathbb{C}^n$ such that $\|\vec{b} - A\vec{x}\|_2$ is minimized.

Least Squares Problems...

The conditioning of these problems depend on a combination of

(1) The conditioning of square systems of equations
(2) The geometry of orthogonal projections.

The topic is subtle, and has nontrivial implications for the **stability** (and ultimately, the **accuracy**) of least squares algorithms.

From our previous discussion of least squares problem we know

$$\begin{aligned} \vec{x} &= A^\dagger \vec{b}, \quad \text{where} \quad A^\dagger = (A^*A)^{-1}A^* \\ A\vec{x} &= \vec{y}, \quad \text{where} \quad \vec{y} = P\vec{b}, \quad P = AA^\dagger \end{aligned}$$

P is the **orthogonal projector** onto $\mathrm{range}(A)$, and $A^\dagger \in \mathbb{C}^{m\times m}$ is the **pseudo-inverse** of A. For this, theoretical infinite-precision, discussion the choice of implementation/expression for the pseudo-inverse does not matter.

Least Squares Problems... Conditioning

Conditioning is the measure of sensitivity of solutions to perturbations in the data.

Our data are

$$A \in \mathbb{C}^{m \times n}, \quad \text{and} \quad \vec{b} \in \mathbb{C}^{m},$$

and the solution is either the vector $\vec{x} \in \mathbb{C}^{n}$, or the vector $\vec{y} = P\vec{b}$ (depending on our point of view / application).

We end up with four combinations of input/output-perturbations:

$\downarrow$ **Input, Output** $\rightarrow$	$\vec{y}$	$\vec{x}$
$\vec{b}$	$\kappa(\vec{b} \rightarrow \vec{y})$	$\kappa(\vec{b} \rightarrow \vec{x})$
A	$\kappa(A \rightarrow \vec{y})$	$\kappa(A \rightarrow \vec{x})$

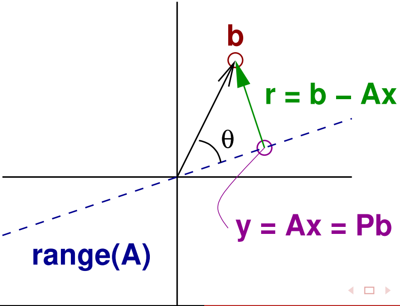
Three Dimensionless Parameters

We are going to express all the condition-numbers using three dimensionless parameters — $\kappa(A)$, θ, and η

$\kappa(\mathbf{A})$ is our old friend the condition number of the matrix A

$$\kappa(A) = \frac{\sigma_1}{\sigma_n}.$$

θ is the angle between $\vec{b}$ and $\vec{y} = A\vec{x} = P\vec{b}$,



Three Dimensionless Parameters

η is a measure of how much $\|\vec{y}\|$ falls short of its maximum value, given $\|A\|$ and $\|\vec{x}\|$: (or how misaligned $(\vec{y}, \vec{x})$ is with $(\vec{u}_1, \vec{v}_1)$. — Implications for "Model Quality")

$$\eta = \frac{\|A\| \, \|\vec{x}\|}{\|\vec{y}\|} = \frac{\|A\| \, \|\vec{x}\|}{\|A\vec{x}\|}.$$

These parameters lie in the ranges

$$\kappa(A) \in [1, \infty), \quad \theta \in \left[0, \frac{\pi}{2}\right], \quad \eta \in [\, 1, \kappa(A)\,),$$

and

$$\cos(\theta) = \frac{\|\vec{y}\|}{\|\vec{b}\|}, \quad \theta = \cos^{-1}\left(\frac{\|\vec{y}\|}{\|\vec{b}\|}\right).$$

Least Squares Problems... Conditioning Theorem

Theorem (Conditioning of the Least Squares Problems)

Let $\vec{b} \in \mathbb{C}^m$ and $A \in \mathbb{C}^{m \times n}$ of full rank be given.

The least squares problem, $\min_{\vec{x} \in \mathbb{C}^n} \|\vec{b} - A\vec{x}\|$ *has the following 2-norm relative condition numbers describing the sensitivities of $\vec{y}$ and $\vec{x}$ to perturbations in $\vec{b}$ and A:*

$\downarrow$ **Input, Output** $\rightarrow$	$\vec{y}$	$\vec{x}$
$\vec{b}$	$\dfrac{1}{\cos\theta}$	$\dfrac{\kappa(A)}{\eta\cos\theta}$
A	$\dfrac{\kappa(A)}{\cos\theta}$	$\kappa(A) + \dfrac{\kappa(A)^2\tan\theta}{\eta}$

The results in the first row are exact, being attained for certain perturbations $\delta\vec{b}$, and the results in the second row are upper bounds.

A Note on the Theorem $m = n$

In the special case $m = n$, the least squares problem reduces to a square non-singular problem, with $\theta = 0$, and the table looks like

$\downarrow$ **Input**, **Output** $\rightarrow$	$\vec{y}$	$\vec{x}$
$\vec{b}$	1	$\dfrac{\kappa(A)}{\eta}$
A	0	$\kappa(A)$

Since A is square + full rank, $\vec{y}$ is already in the range, so no projection is needed; hence the condition number is 0.

(Massively) Simplifying the Proof Using the SVD

We have shown (a long, long time ago) that every matrix has a singular value decomposition.

Let $U\Sigma V^* = A$ be the SVD of A. We can use U and V to get two convenient bases in which we prove the theorem. Since 2-norm perturbations are not changed by a unitary change of basis, the **perturbation behavior of A is the same as that of Σ.**

Without loss of generality we can assume that

$$A = \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} = \begin{bmatrix} A_1 \\ 0 \end{bmatrix} = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

Moving Along...

Now with

$$\vec{b} = \begin{bmatrix} \vec{b}_1 \\ \vec{b}_2 \end{bmatrix}, \quad \vec{b}_1 \in \mathbb{C}^n, \ \vec{b}_2 \in \mathbb{C}^{m-n}$$

the projection of $\vec{b}$ onto $\mathrm{range}(A)$ is trivial

$$\vec{y} = P\vec{b} = \begin{bmatrix} \vec{b}_1 \\ \vec{0} \end{bmatrix}$$

Now, $A\vec{x} = \vec{y}$ has the unique solution $\vec{x} = A_1^{-1}\vec{b}_1$.

We note that the orthogonal projector, and the pseudo-inverse of A take the forms

$$P = \begin{bmatrix} I_{n\times n} & 0 \\ 0 & 0 \end{bmatrix}, \qquad A^\dagger = \begin{bmatrix} A_1^{-1} & 0 \end{bmatrix}.$$

Part#1: Sensitivity of $\vec{y}$ to Perturbations in $\vec{b}$

$\tilde{\mathbf{y}} = \mathbf{P}\tilde{\mathbf{b}}$ is a linear differentiable map; and the Jacobian is P itself, with $\|P\| = 1$.

For a differentiable map $x \mapsto f(\vec{x})$ the condition number is

$$\kappa = \frac{\|J(\vec{x})\|}{\|f(\vec{x})\|/\|\vec{x}\|}.$$

Here we have,

$$\kappa(\vec{b} \to \vec{y}) = \frac{\|P\|}{\|\vec{y}\|/\|\vec{b}\|} = \frac{1}{\cos\theta} \quad \square.$$

Part#2: Sensitivity of $\vec{x}$ to Perturbations in $\vec{b}$

$\tilde{\mathbf{x}} = \mathbf{A}^\dagger \tilde{\mathbf{b}}$ is also linear, with Jacobian $J = A^\dagger$, so

$$\kappa(\vec{b} \to \vec{x}) = \frac{\|A^\dagger\|}{\|\vec{x}\|/\|\vec{b}\|} = \|A^\dagger\| \, \frac{\|\tilde{\mathbf{b}}\|}{\|\tilde{\mathbf{y}}\|} \, \frac{\|\tilde{\mathbf{y}}\|}{\|\tilde{\mathbf{x}}\|} = \|A^\dagger\| \, \frac{\mathbf{1}}{\cos\theta} \, \frac{\|\mathbf{A}\|}{\eta}$$

Finally, we recognize $\kappa(A) = \sigma_1 \cdot \dfrac{1}{\sigma_n} = \|A\| \, \|A^\dagger\|$ (in this case), and we have

$$\kappa(\vec{b} \to \vec{x}) = \frac{\kappa(A)}{\eta \cos\theta}. \qquad \square$$

That concludes the "easy" parts of the proof...

Part#3: Perturbations in A — A Help-Result

Perturbations in A affect the least squares problem in two ways

- The mapping of $\mathbb{C}^n$ onto $\mathrm{range}(A)$ is distorted.
- $\mathrm{range}(A)$ is also altered.

Perturbations in A affect the least squares problem in two ways

- The mapping of $\mathbb{C}^n$ onto $\mathrm{range}(A)$ is distorted.
- $\mathrm{range}(A)$ is also altered.

The changes in $\mathrm{range}(A)$ introduce a "tilt" of the space; and the question is what is the maximal tilt $\delta\alpha$ induced by a perturbation δA?

The image of the unit sphere in $\mathbb{R}^n$, $\mathbb{S}^{n-1}$ is $A\mathbb{S}^{n-1}$, a hyper-ellipse that "lies flat" in $\mathrm{range}(A)$.

Perturbations in A affect the least squares problem in two ways

- The mapping of $\mathbb{C}^n$ onto $\mathrm{range}(A)$ is distorted.
- $\mathrm{range}(A)$ is also altered.

The changes in $\mathrm{range}(A)$ introduce a "tilt" of the space; and the question is what is the maximal tilt $\delta\alpha$ induced by a perturbation δA?

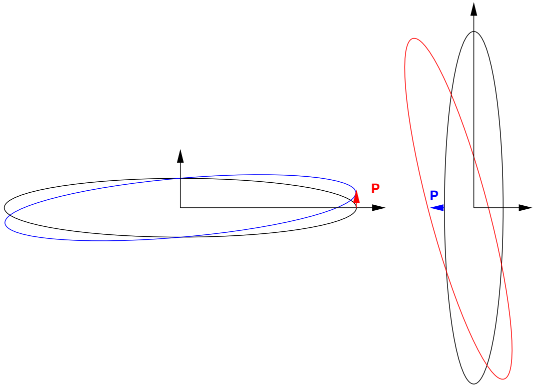
The image of the unit sphere in $\mathbb{R}^n$, $\mathbb{S}^{n-1}$ is $A\mathbb{S}^{n-1}$, a hyper-ellipse that "lies flat" in $\mathrm{range}(A)$.

We grab a point $\vec{p} = A\vec{v}$ on the hyper-ellipse (hence $\|\vec{v}\| = 1$, since $\vec{v} \in \mathbb{S}^{n-1}$); we introduce a perturbation $\delta\vec{p} \perp \mathrm{range}(A)$.

We can express this as a rank-1 matrix perturbation $\delta A = (\delta\vec{p})\vec{v}^* \Leftrightarrow (\delta A)\vec{v} = \delta\vec{p}$, and $\|\delta A\| = \|\delta\vec{p}\|$.

SAN DIEGO STATE UNIVERSITY

Now, clearly, if we want to maximize the tilt, we should grab the hyper-ellipse as close to the origin as possible



Hence, we let $\vec{p} = \sigma_n \vec{u}_n$ (the minor semi-axis in $A\mathbb{S}^{n-1}$.)

Now, since we have A in a convenient diagonal (Σ) form, $\vec{p}$ is the last column of A, $\vec{v}^* = (0, 0, \ldots, 0, 1)$, and δA is a perturbation below the diagonal in this (last) column.

$$\vec{p} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \sigma_n \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \delta\vec{p} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \delta p_{n+1} \\ \vdots \\ \delta p_m \end{bmatrix}, \quad \delta A = \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \\ 0 & 0 & \ldots & \delta A_{n+1,n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \delta A_{m,n} \end{bmatrix}$$

Now, since we have A in a convenient diagonal (Σ) form, $\vec{p}$ is the last column of A, $\vec{v}^* = (0, 0, \ldots, 0, 1)$, and δA is a perturbation below the diagonal in this (last) column.

$$\vec{p} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \sigma_n \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \delta\vec{p} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \delta p_{n+1} \\ \vdots \\ \delta p_m \end{bmatrix}, \quad \delta A = \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \\ 0 & 0 & \ldots & \delta A_{n+1,n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \delta A_{m,n} \end{bmatrix}$$

the tilting angle induced by this perturbation is

$$\tan(\delta\alpha) = \frac{\|\delta\vec{p}\|}{\sigma_n}.$$

We have

$$\tan(\delta\alpha) = \frac{\|\delta\vec{p}\|}{\sigma_n}.$$

Further,

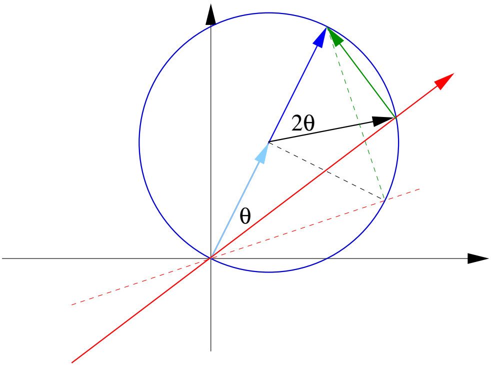
$$\|\delta\vec{p}\| = \|\delta A\|, \quad \delta\alpha \le \tan(\delta\alpha),$$

Hence,

$$\delta\alpha \le \frac{\|\delta A\|}{\sigma_n} = \frac{\|\delta A\|}{\|A\|}\kappa(A).$$

We are now ready to proceed with the proof...

Part#3: Sensitivity of $\vec{y}$ to Perturbations in A



Since $\vec{y}$ is the orthogonal projection of $\vec{b}$ onto $\mathrm{range}(A)$, it is determined by $\vec{b}$ and $\mathrm{range}(A)$ alone. Therefore we can study changes on $\vec{y}$ induced by tiltings $\delta\alpha$ of $\mathrm{range}(A)$.

No matter how we tilt $\mathrm{range}(A)$, $\vec{y} \in \mathrm{range}(A)$ must be orthogonal to $(\vec{b} - \vec{y}) \in \mathrm{range}(A)^{\perp}$. — As $\mathrm{range}(A)$ varies, the point $\vec{y}$ moves along a sphere of radius $\|\vec{b}\|/2$ centered at the point $\vec{b}/2$.

Part#3: Sensitivity of $\vec{y}$ to Perturbations in A

Tilting $\mathrm{range}(A)$ in the plane $\vec{0}$-$\vec{b}$-$\vec{y}$ by an angle $\delta\alpha$ changes the angle "2θ" at the central point $\vec{b}/2$ by $2\delta\alpha$.

The corresponding change $\delta\vec{y}$, is the base of an isosceles triangle with central angle $2\delta\alpha$, and edge length $\|\vec{b}\|/2$. Hence, $\|\delta\vec{y}\| = \|\vec{b}\|\, \sin(\delta\alpha)$

Tilting $\mathrm{range}(A)$ in any other plane results in a similar geometry in a different plane and perturbations smaller by a factor as small as $\sin\theta$.

For arbitrary perturbations we have

$$\|\delta\vec{y}\| \leq \|\vec{b}\|\, \sin(\delta\vec{a}) \leq \|\vec{b}\|\, \delta\alpha$$

Combining with previous results give us $\kappa(A \to \vec{y})$

$$\|\delta\vec{y}\| \leq \|\vec{b}\| \frac{\|\delta A\|}{\|A\|}\kappa(A) = \frac{\|\vec{y}\|}{\cos\theta}\,\frac{\|\delta A\|}{\|A\|}\kappa(A) \quad \Leftrightarrow \quad \frac{\|\delta\tilde{\mathbf{y}}\|}{\|\tilde{\mathbf{y}}\|} \Big/ \frac{\|\delta\mathbf{A}\|}{\|\mathbf{A}\|} \leq \frac{\kappa(\mathbf{A})}{\cos\theta}. \quad \square$$

SAN DIEGO STATE UNIVERSITY

Part#4: Sensitivity of $\vec{x}$ to Perturbations in A

We now analyze the most interesting relationship of the theorem; the sensitivity of the least squares solution to perturbations in A.

We write perturbations in two parts

$$\delta A = \begin{bmatrix} \delta A_1 \\ \delta A_2 \end{bmatrix}, \quad \delta A_1 \in \mathbb{C}^{n\times n}, \;\; \delta A_2 \in \mathbb{C}^{(m-n)\times n}$$

First, we look at the effects of $\delta \mathbf{A_1}$: these perturbations change the mapping of A in its range, **but does not change $\mathrm{range}(A)$ itself**, and hence not $\vec{y}$. We get

$$(A_1 + \delta A_1)\vec{x} = \vec{b}_1$$

The condition number for this operation is simply (as before)

$$\kappa(A_1 \to \vec{x}) = \frac{\|\delta \vec{x}\|}{\|\vec{x}\|} \Big/ \frac{\|\delta A_1\|}{\|A_1\|} \leq \kappa(A_1) = \kappa(A)$$

Part#4: Sensitivity of $\vec{x}$ to Perturbations in A

Next, we consider the effects of $\delta\mathbf{A_2}$. This perturbation tilts $\mathrm{range}(A)$ without changing the mapping of A within this space.

Part#4: Sensitivity of $\vec{x}$ to Perturbations in A

Next, we consider the effects of $\delta\mathbf{A_2}$. This perturbation tilts $\mathrm{range}(A)$ without changing the mapping of A within this space.

The point vector $\vec{b}_1$, and the point $\vec{y} = [\vec{b}_1^* \, \vec{0}^*]^*$ are perturbed, but A_1 is not. This corresponds to perturbing $\vec{b}_1$ in $\vec{x} = A_1^{-1}\vec{b}_1$, for which the condition number takes the form

$$\kappa = \frac{\|\delta\vec{x}\|}{\|\vec{x}\|} \Bigg/ \frac{\|\delta\vec{b}_1\|}{\|\vec{b}_1\|} \leq \frac{\kappa(A_1)}{\eta(A_1, \vec{x})} = \frac{\kappa(\mathbf{A})}{\eta}$$

Part#4: Sensitivity of $\vec{x}$ to Perturbations in A

Next, we consider the effects of $\delta\mathbf{A_2}$. This perturbation tilts $\mathrm{range}(A)$ without changing the mapping of A within this space.

The point vector $\vec{b}_1$, and the point $\vec{y} = [\vec{b}_1^* \, \vec{0}^*]^*$ are perturbed, but A_1 is not. This corresponds to perturbing $\vec{b}_1$ in $\vec{x} = A_1^{-1}\vec{b}_1$, for which the condition number takes the form

$$\kappa = \frac{\|\delta\vec{x}\|}{\|\vec{x}\|} \Bigg/ \frac{\|\delta\vec{b}_1\|}{\|\vec{b}_1\|} \leq \frac{\kappa(A_1)}{\eta(A_1,\vec{x})} = \frac{\kappa(\mathbf{A})}{\eta}$$

since...

$$\frac{\|\delta\vec{x}\|}{\|\vec{x}\|} \Bigg/ \frac{\|\delta\vec{b}_1\|}{\|\vec{b}_1\|} \leq \frac{\|J(\vec{x})\|}{\|\vec{x}\| \,/\, \|\vec{b}_1\|} = \frac{\|A_1^{-1}\|\,\|\vec{b}_1\|}{\|\vec{x}\|} = \frac{1}{\sigma_n}\frac{\|A_1\vec{x}\|}{\|\vec{x}\|} = \frac{\sigma_1}{\sigma_n}\frac{\|\mathbf{A_1}\tilde{\mathbf{x}}\|}{\|\mathbf{A_1}\|\,\|\tilde{\mathbf{x}}\|}$$

SAN DIEGO STATE UNIVERSITY

Part#4: Sensitivity of $\vec{x}$ to Perturbations in A

In order to close this argument out, we must relate $\delta\vec{b}_1$ to δA_2...

The vector $\vec{b}_1$ is $\vec{y}$ expressed in the coordinates of $\mathrm{range}(A)$. Therefore, the only changes in $\vec{y}$ that are realized as changes in $\vec{b}_1$ are those that are parallel to $\mathrm{range}(A)$.

Part#4: Sensitivity of $\vec{x}$ to Perturbations in A

If $\mathrm{range}(A)$ is tilted by $\delta\alpha$ in the $\vec{0}$-$\vec{b}$-$\vec{y}$ plane, the resulting perturbation $\delta\vec{y}$ is not parallel to $\mathrm{range}(A)$, but at an angle $\frac{\pi}{2} - \theta$, therefore

$$\|\delta\vec{b}_1\| = \|\delta\vec{y}\| \sin\theta \leq \|\vec{b}\|\, \delta\alpha \sin\theta.$$

Part#4: Sensitivity of $\vec{x}$ to Perturbations in A

If $\mathrm{range}(A)$ is tilted by $\delta\alpha$ in the $\vec{0}$-$\vec{b}$-$\vec{y}$ plane, the resulting perturbation $\delta\vec{y}$ is not parallel to $\mathrm{range}(A)$, but at an angle $\frac{\pi}{2} - \theta$, therefore

$$\|\delta\vec{b}_1\| = \|\delta\vec{y}\| \sin\theta \leq \|\vec{b}\|\, \delta\alpha \sin\theta.$$

If $\mathrm{range}(A)$ is tilted in a direction orthogonal to the $\vec{0}$-$\vec{b}$-$\vec{y}$ plane, $\delta\vec{y}$ is parallel to $\mathrm{range}(A)$, and we get $\|\delta\vec{y}\| \leq \|\vec{b}\|\, \delta\alpha \sin\theta$, and since $\|\delta\vec{b}_1\| \leq \|\delta\vec{y}\|$, we have

$$\|\delta\vec{b}_1\| \leq \|\vec{b}\|\, \delta\alpha \sin\theta, \quad \text{same argument as for } \kappa(A \to \vec{y}).$$

SAN DIEGO STATE UNIVERSITY

Part#4: Sensitivity of $\vec{x}$ to Perturbations in A

If $\mathrm{range}(A)$ is tilted by $\delta\alpha$ in the $\vec{0}$-$\vec{b}$-$\vec{y}$ plane, the resulting perturbation $\delta\vec{y}$ is not parallel to $\mathrm{range}(A)$, but at an angle $\frac{\pi}{2} - \theta$, therefore

$$\|\delta\vec{b}_1\| = \|\delta\vec{y}\| \sin\theta \leq \|\vec{b}\|\, \delta\alpha \sin\theta.$$

If $\mathrm{range}(A)$ is tilted in a direction orthogonal to the $\vec{0}$-$\vec{b}$-$\vec{y}$ plane, $\delta\vec{y}$ is parallel to $\mathrm{range}(A)$, and we get $\|\delta\vec{y}\| \leq \|\vec{b}\|\, \delta\alpha \sin\theta$, and since $\|\delta\vec{b}_1\| \leq \|\delta\vec{y}\|$, we have

$$\|\delta\vec{b}_1\| \leq \|\vec{b}\|\, \delta\alpha \sin\theta, \quad \text{same argument as for } \kappa(A \to \vec{y}).$$

We now have all the pieces to the puzzle... all we need is a bit of glue!

Part#4: Sensitivity of $\vec{x}$ to Perturbations in A

Since $\|\vec{b}_1\| = \|\vec{b}\| \cos\theta$ we can rewrite the previous inequality as

$$\frac{\|\delta\vec{b}_1\|}{\|\vec{b}_1\|} \leq \delta\alpha \tan\theta.$$

Part#4: Sensitivity of $\vec{x}$ to Perturbations in A

Since $\|\vec{b}_1\| = \|\vec{b}\| \cos\theta$ we can rewrite the previous inequality as

$$\frac{\|\delta\vec{b}_1\|}{\|\vec{b}_1\|} \leq \delta\alpha \tan\theta.$$

using the final result on slide 19 in the form

$$\frac{\delta\alpha\,\|A\|}{\|\delta A\|} \leq \kappa(A)$$

we have

$$\frac{\|\delta\vec{x}\|}{\|\vec{x}\|} \bigg/ \frac{\|\delta\mathbf{A_2}\|}{\|\mathbf{A}\|} = \frac{\|\delta\vec{b}_1\|}{\|\vec{b}_1\|}\,\frac{\kappa(A)}{\eta}\,\frac{\|\mathbf{A}\|}{\|\delta\mathbf{A_2}\|} \leq \frac{\tan\theta\,\kappa(A)}{\eta}\,\frac{\delta\alpha\,\|A\|}{\|\delta A\|} \leq \frac{\tan\theta\,\kappa(\mathbf{A})^2}{\eta}$$

Part#4: Sensitivity of $\vec{x}$ to Perturbations in A

Since $\|\vec{b}_1\| = \|\vec{b}\| \cos\theta$ we can rewrite the previous inequality as

$$\frac{\|\delta \vec{b}_1\|}{\|\vec{b}_1\|} \leq \delta\alpha \tan\theta.$$

using the final result on slide 19 in the form

$$\frac{\delta\alpha \, \|A\|}{\|\delta A\|} \leq \kappa(A)$$

we have

$$\frac{\|\delta\vec{x}\|}{\|\vec{x}\|} \Big/ \frac{\|\delta \mathbf{A_2}\|}{\|\mathbf{A}\|} = \frac{\|\delta\vec{b}_1\|}{\|\vec{b}_1\|} \frac{\kappa(A)}{\eta} \frac{\|\mathbf{A}\|}{\|\delta\mathbf{A_2}\|} \leq \frac{\tan\theta\,\kappa(A)}{\eta} \frac{\delta\alpha\,\|A\|}{\|\delta A\|} \leq \frac{\tan\theta\,\kappa(\mathbf{A})^2}{\eta}$$

Adding this to the contribution from δA_1 gives us

$$\kappa(A \to \vec{x}) = \kappa(A) + \frac{\tan\theta\,\kappa(A)^2}{\eta}. \quad \square$$

San Diego State University

One Final Comment

Clearly, finding the least squares solution $\vec{x}$ is a tough problem:

— The condition number contains the square of the condition number of the matrix A.

— Even for moderately ill-conditioned matrices, the least squares problem quickly becomes very ill-conditioned.

Next time we connect the conditioning results derived here with the stability (or lack thereof) of some numerical algorithms applied to the least squares problem.

Homework #6 — Due Friday April 10, 2020

TB-18.1:

PB-14.1: Consider the vector $\vec{x} \in \mathbb{R}^{101}$ consisting of equi-spaces points in the interval $[0, 1]$, e.g. **x = linspace(0,1,101)';** and let $A_k \in \mathbb{R}^{101\times(k+1)}$ be the matrix consisting of columns formed by (component-size powers $\{0, \ldots, k\}$ of the x-values (a Vandermonde Matrix). Let $c_k = \kappa(A_k)$ be the collection of condition numbers for these matrices.

- Plot $\vec{c}$ (use a log scale)
- We could use these matrices (A_k) to least-squares-fit polynomials (of matching degree k) to some data-set with 101 measurements. Is it necessarily better to have more model parameters (*i.e.* fitting a higher degree polynomial)? — Discuss.

https://en.wikipedia.org/wiki/Vandermonde_matrix