Numerical Matrix Analysis

Notes #9 — Conditioning and Stability
Conditioning and Condition Numbers

Peter Blomgren
⟨blomgren@sdsu.edu⟩

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 2024
(Revised: February 21, 2024)

Outline

Student Learning Targets, and Objectives

Target Condition Numbers
- **Objective** Know the definition of, and be able to computer, Absolute and Relative Condition Numbers
- **Objective** Condition number is the "special" case of continuity

Target Key Conditioning Examples
- **Objective** Cancellation Error
- **Objective** Wilkinson's Classic Example — Ill-conditioning of root finding
- **Objective** Eigenvalues and Eigenvectors

What We Swept Under The Rug...

So far we have not discussed stability in a systematic way...

... unless vigorous hand-waving and "proof by picture" qualifies as systematic.

We now turn our attention to the issue of **stability**, and look at several things:

- Conditioning — sensitivity to perturbations.
- Finite precision floating point arithmetic — representation errors.
- Stability of Algorithms.

Conditioning : Perturbation Behavior of the Mathematical Problem.
Stability : Perturbation Behavior of an Algorithm.

The Condition of a Problem

For us a (linear algebra) problem is a function

$$f : X \to Y$$

where $X \subseteq \mathbb{C}^n$, and $Y \subseteq \mathbb{C}^m$. Given some input $\vec{x} \in X$, we produce an answer $\vec{y} \in Y$.

A **well-conditioned problem** has the property that small perturbations (changes) in $\vec{x}$ leads to small changes in $\vec{y} = f(\vec{x})$.

An **ill-conditioned problem** has the property that small perturbations (changes) in $\vec{x}$ leads to large changes in $\vec{y} = f(\vec{x})$.

Clearly, we must quantify what "small" and "large" mean...

Absolute Condition Number

Let $\delta\vec{x}$ denote a small perturbation of $\vec{x}$, and let

$$\delta f \stackrel{\text{def}}{=} f(\vec{x}+\delta\vec{x}) - f(\vec{x})$$

be the corresponding change in f.

The **Absolute Condition Number** $\widehat{\kappa}(\vec{x})$ of the problem f **at** $\vec{x}$ is defined:

$$\widehat{\kappa}(\vec{x}) = \lim_{\Delta\to 0} \sup_{\|\delta\vec{x}\|\leq\Delta} \frac{\|\delta f\|}{\|\delta\vec{x}\|}$$

Think: the supremum over small perturbations.

For **notational convenience** we usually drop the limit, and write

$$\widehat{\kappa}(\vec{x}) = \sup_{\delta\vec{x}} \frac{\|\delta f\|}{\|\delta\vec{x}\|}$$

with the understanding that $\delta\vec{x}$ and δf are infinitesimal.

Differentiability and the Absolute Condition Number

If $f : \mathbb{C}^n \mapsto \mathbb{C}^m$ is differentiable, we can define the **Jacobian**

$$J(\vec{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}.$$

Now, we have

$$\delta f = J(\vec{x})\delta\vec{x}, \quad \|\delta\vec{x}\| \to 0,$$

and we can express the condition number in terms of the Jacobian

$$\widehat{\kappa}(\vec{x}) = \|J(\vec{x})\|.$$

Relative Condition Number

The **Relative Condition Number** $\kappa(\vec{x})$ of the problem is defined as

$$\kappa(\vec{x}) = \lim_{\Delta\to 0} \sup_{\|\delta\vec{x}\|\leq\Delta} \left[\frac{\|\delta f\|}{\|f(\vec{x})\|} \Big/ \frac{\|\delta\vec{x}\|}{\|\vec{x}\|} \right],$$

or, compactly,

$$\kappa(\vec{x}) = \sup_{\delta\vec{x}} \left[\frac{\|\delta f\|}{\|f(\vec{x})\|} \Big/ \frac{\|\delta\vec{x}\|}{\|\vec{x}\|} \right].$$

If/When f is differentiable we get

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

Absolute vs. Relative Condition Numbers

The **relative condition number** tends to be the more useful description of error propagation in numerical analysis.

Part of the reason is that errors introduced due to floating point arithmetic during computations are relative to the size of the computed quantities.

Another reason is that even if the absolute condition number is small, the relative condition number can still be large, if $\frac{\|f(\vec{x})\|}{\|\vec{x}\|}$ is small. Here, a small absolute perturbation of $f(\vec{x})$ may make the result $f(\vec{x}+\delta\vec{x})$ almost completely independent of $f(\vec{x})$, *i.e.* completely dominated by the perturbation.

Rules of Thumb: If $\kappa \sim 1, 10, 10^2$ then it is "small," and the problem is well-conditioned; if $\kappa \sim 10^{10} \to 10^{16}$ then it is "large," and the problem is ill-conditioned.

Example: Quantifying "Cancellation Error" 1 of 2

We consider the problem $f : \mathbb{C}^2 \to \mathbb{C}$ defined by

$$f(\vec{x}) = x_1 - x_2, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{C}^2$$

The Jacobian of f is

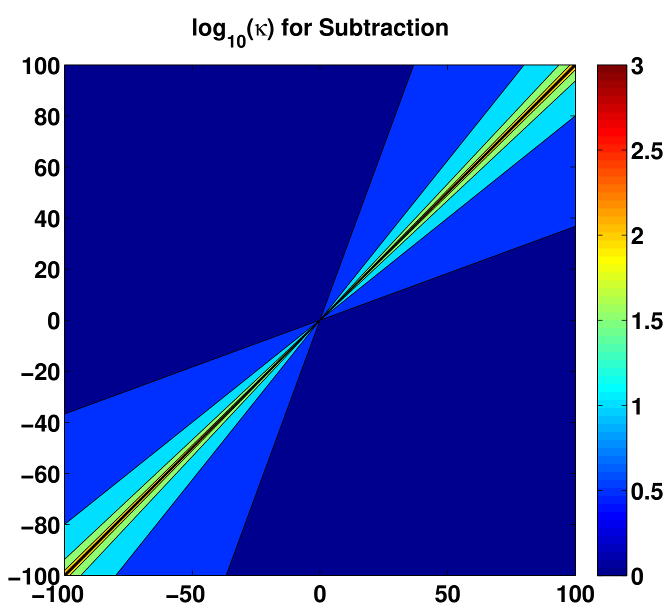
$$J(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 1 & -1 \end{bmatrix}.$$

If we use the ∞-norm, we have $\|J(\vec{x})\|_\infty = 2$, and

$$\kappa(\vec{x}) = \frac{\|J(\vec{x})\|_\infty}{\|f(\vec{x})\|/\|\vec{x}\|_\infty} = \frac{2\,\max\{|x_1|,|x_2|\}}{|x_1 - x_2|}.$$

Now, if $x_1 \approx x_2$, the problem is clearly ill-conditioned; otherwise it is well-conditioned.

Example: Quantifying "Cancellation Error" 2 of 2



Wilkinson's Classic Example : Root Finding 1 of 4

Finding the roots of a polynomial given the polynomial coefficients is a classic example of an ill-conditioned problem, *e.g.*

$$\begin{aligned} x^2 - 2x + 1 &= (x-1)^2 \\ x^2 - 2x + 0.9999 &= (x-0.99)(x-1.01) \\ x^2 - 2x + 0.999999 &= (x-0.999)(x-1.001) \end{aligned}$$

Here, a perturbation of 10^{-4} (10^{-6}) in one coefficient moved both roots 10^{-2} (10^{-3}).

The roots change $\propto \sqrt{\delta(\text{coeff})}$, and since

$$\lim_{\delta(\text{coeff})\searrow 0} \frac{\delta(\text{roots})}{\delta(\text{coeff})} = \lim_{\delta(\text{coeff})\searrow 0} \frac{\sqrt{\delta(\text{coeff})}}{\delta(\text{coeff})} = \lim_{\delta(\text{coeff})\searrow 0} \frac{1}{\sqrt{\delta(\text{coeff})}} \to \infty$$

the condition number is: $\kappa = \infty$.

Wilkinson's Classic Example

Even when the roots are single (distinct), polynomial root-finding is ill-conditioned:

If the ith coefficient a_i of a polynomial $p(x)$ is perturbed by δa_i, the perturbation of the jth root x_j is

$$\delta x_j = -\frac{(\delta a_i) x_j^i}{p'(x_j)}, \quad \text{and} \quad \kappa_{ji} = \frac{|a_i x_j^{i-1}|}{|p'(x_j)|}$$

where κ_{ji} is the condition number of x_j with respect to perturbation of the coefficient a_i.

This number can be very large.

Wilkinson's Classic Example : Root Finding

The classic example is the roots of Wilkinson's polynomial

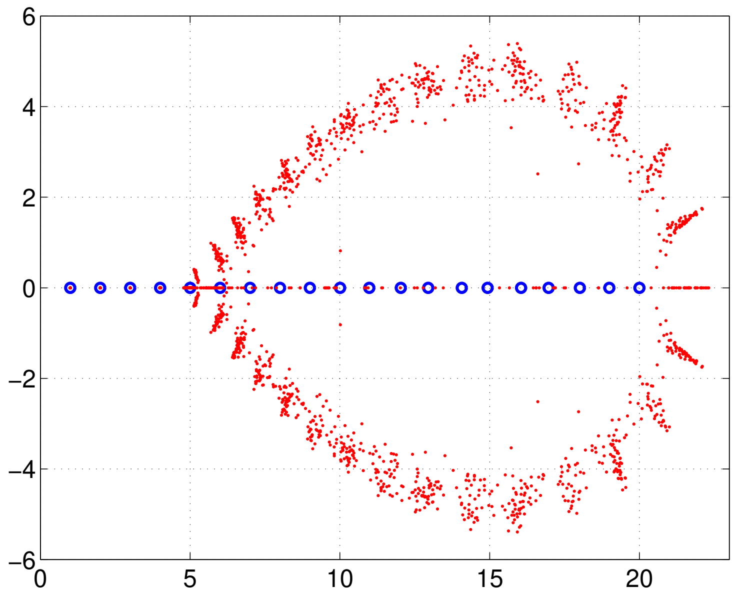
$$p(x) = \prod_{i=1}^{20} (x - i) = a_0 + a_1 x + \cdots + a_{19} x^{19} + x^{20}$$

with the unperturbed roots $\{1, 2, \ldots, 19, 20\}$.

It turns out that the most sensitive root is $r_{15} = 15$, and it is most sensitive to perturbations in $a_{15} \approx -1.67 \times 10^9$, with $\kappa_{15,15} \approx 4.6602 \times 10^{12}$.

The **figure** on the next slide shows the distribution of roots of 100 randomly perturbed Wilkinson polynomials. The coefficients have been perturbed $\tilde{a}_i = (1 + 10^{-10} r_i) a_i$, where r_i is drawn from the $N(0,1)$ distribution (mean zero, variance 1).

Wilkinson's Classic Example : Root Finding



Polynomial Roots: Comments

It turns out that polynomial rootfinding does not have to be as ill-conditioned as we have described. The ill-conditioning as described is largely associated with the **unfortunate choice of basis**: $\{ x^k \}_{k=0,1,\ldots}$ (the standard polynomial basis).

Using, *e.g.* the **orthonormal** Chebyshev polynomial basis $\{ T_n(x) \}_{n=0,1,\ldots}$ can improve the conditioning significantly.

[L.N. Trefethen 2012], *Six Myths of Polynomial Interpolation and Quadrature*, Mathematics Today 47, no. 4, pp. 184–188.

[L.N. Trefethen 2013], *Approximation Theory and Approximation Practice*, Society for Industrial and Applied Mathematics.

Eigenvalues of a Non-Symmetric Matrix

Consider the two matrices

$$A_1 = \begin{bmatrix} 1 & 1000 \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad A_2 = \begin{bmatrix} 1 & 1000 \\ 10^{-3} & 1 \end{bmatrix}$$

The eigenvalues and eigenvectors of A_1 are

$$\lambda(A_1) = \{1, 1\}, \quad \vec{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

The eigenvalues and eigenvectors of A_2 are

$$\lambda(A_2) = \{2, 0\}, \quad \vec{u}_1 = \begin{bmatrix} 0.99999950000037 \\ 0.00099999950000 \end{bmatrix}, \; \vec{u}_2 = \begin{bmatrix} -0.99999950000037 \\ 0.00099999950000 \end{bmatrix}$$

Clearly, this problem is quite ill-conditioned. When A is **symmetric**, then the eigenvalues are better conditioned, with $\kappa(\lambda_i) = \|A\|_2/|\lambda_i|$.

Conditioning of Fundamental Linear Algebra Operations

Next, we take a closer look at the conditioning of the basic building blocks of linear algebra, and note that these are at least in some sense different things...

- **[An operation/function]**
  — Matrix-Vector multiplication, $\vec{y} \leftarrow A\vec{x}$
- **[A mathematical "object"]**
  — The Matrix, $A \in \mathbb{C}^{m \times n}$
- **[A mathematical "problem"]**
  — The System of Equations, $A\vec{x} = \vec{b}$

Conditioning of Matrix-Vector Multiplication 1 of 4

Let $A \in \mathbb{C}^{m \times n}$, $\vec{x} \in \mathbb{C}^n$, and consider the product $A\vec{x}$. For now, consider perturbations $\delta\vec{x}$ only.

From the definition of the condition number we have

$$\kappa(\vec{x}) = \sup_{\delta\vec{x}} \left[\frac{\|A(\vec{x}+\delta\vec{x}) - A\vec{x}\|}{\|A\vec{x}\|} \Big/ \frac{\|\delta\vec{x}\|}{\|\vec{x}\|} \right] = \sup_{\delta\vec{x}} \left[\frac{\|A\delta\vec{x}\|}{\|\delta\vec{x}\|} \Big/ \frac{\|A\vec{x}\|}{\|\vec{x}\|} \right]$$

that is

$$\kappa(\vec{x}) = \|A\| \frac{\|\vec{x}\|}{\|A\vec{x}\|}$$

If A is square and non-singular, we can use the fact[‡] that $\frac{\|\vec{x}\|}{\|A\vec{x}\|} \leq \|A^{-1}\|$ and get a bound independent of $\vec{x}$

$$\kappa(\vec{x}) \leq \|A\| \, \|A^{-1}\|$$

[‡] $\|\vec{x}\| = \|A^{-1}A\vec{x}\| \leq \|A^{-1}\| \, \|A\vec{x}\|$.

Conditioning of Matrix-Vector Multiplication 2 of 4

Theorem (Conditioning of Matrix-Vector Multiplication)

Let $A \in \mathbb{C}^{m \times m}$ be non-singular and consider the equation $A\vec{x} = \vec{b}$. The problem of computing $\vec{b}$, given $\vec{x}$, has condition number

$$\kappa = \|A\| \frac{\|\vec{x}\|}{\|\vec{b}\|} \leq \|A\| \, \|A^{-1}\|$$

with respect to perturbations in $\vec{x}$. The problem of computing $\vec{x}$, given $\vec{b}$ ($A^{-1}\vec{b} = \vec{x}$), has condition number

$$\kappa = \|A^{-1}\| \frac{\|\vec{b}\|}{\|\vec{x}\|} \leq \|A^{-1}\| \, \|A\|$$

with respect to perturbations in $\vec{b}$. If $\|\cdot\| = \|\cdot\|_2$ equalities hold if $\vec{x}$ is a multiple of a right singular vector of A corresponding to the minimal singular value σ_m, and if $\vec{b}$ is a multiple of a left singular vector of A corresponding to the maximal singular value σ_1.

Conditioning of Matrix-Vector Multiplication

Note: If $A \in \mathbb{C}^{m \times n}$ is a full-rank ($m \geq n$) non-square matrix, then the previously stated results hold with A^{-1} replaced by the pseudo-inverse, *e.g.*

$$A^{\dagger} = (A^* A)^{-1} A^*$$

I.e.

$$\kappa \leq \|A\| \, \|A^{\dagger}\|$$

With this particular pseudo-inverse we have

$$\begin{aligned} \kappa &\leq \|A\|_2 \, \|(A^* A)^{-1} A^*\|_2 \leq \|A\|_2 \, \|(A^* A)^{-1}\|_2 \, \|A^*\|_2 \\ &= \sigma_1 \frac{1}{\sigma_n^2} \sigma_1 = \left[\frac{\sigma_1}{\sigma_n}\right]^2 \end{aligned}$$

Conditioning of Matrix-Vector Multiplication

Approach	Pseudo-inverse	Conditioning‡
Normal Equations	$A^{\dagger} = (A^* A)^{-1} A^*$	$\kappa(\vec{x} \vert \delta\vec{b}) \leq (\sigma_1/\sigma_n)^2$
QR-Factorization	$A^{\dagger} = R^{-1} Q^*$	$\kappa(\vec{x} \vert \delta\vec{b}) \leq (\sigma_1/\sigma_n)$
SVD	$A^{\dagger} = V \Sigma U^*$	$\kappa(\vec{x} \vert \delta\vec{b}) \leq (\sigma_1/\sigma_n)$

This shows that the solution strategy involving the normal equations is subject to a larger condition number than the *QR* or SVD approaches.

‡ Conditioning of solving for $\vec{x} = A^{\dagger}\vec{b}$, wrt. $\delta\vec{b}$.

The Condition Number of a Matrix

The product $\|A\| \, \|A^{-1}\|$ is ubiquitous in numerical analysis, and has its own name — the **condition number** of the matrix A;

$$\kappa(A) = \|A\| \, \|A^{-1}\|, \quad \text{relative to the norm } \|\cdot\|.$$

In this instance, the condition number is attached to the matrix A, not (as earlier) to a problem.

If $\kappa(A)$ is small the matrix is well-conditioned, otherwise ill-conditioned.

If $\|\cdot\| = \|\cdot\|_2$, then

$$\|A\| = \sigma_1, \quad \|A^{-1}\| = 1/\sigma_m, \quad \text{thus} \quad \kappa(A) = \frac{\sigma_1}{\sigma_m}.$$

When A is singular, $\kappa(A) = \infty$.

The Condition Number of a Matrix: Comments

Geometrically $\kappa(A)$ is the eccentricity of the hyper-ellipse $A\mathbb{S}^{n-1}$ — the ratio of the major and minor semi-axes.

In many problems this ratio is referred to as the **separation of scales**.

In Ordinary Differential Equations, the term **stiffness** is used.

Since $1 \leq \kappa(A) \leq \infty$, it is sometimes more convenient to compute the reciprocal condition number $1/\kappa(A)$. If $1/\kappa(A) \sim 10^{-d}$ then application of A (or A^{-1}) to a vector will roughly result in a loss of d significant digits of accuracy.

For non-square $A \in \mathbb{C}^{m \times n}$ $(m \geq n)$ of full rank, the most useful generalization of the condition number is

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

Condition of a System of Equations — 1 of 2

We have considered $A\vec{x} = \vec{b}$ where A was fixed, and we perturbed either $\vec{x}$ or $\vec{b}$ and looked at the effect on $\vec{b}$ or $\vec{x}$.

Now, let's perturb $A \mapsto A + \delta A$, while holding $\vec{b}$ fixed, and study the effect on $\vec{x}$, we must have

$(A+\delta A)(\vec{x}+\delta\vec{x}) = \vec{b}$	
$A\vec{x} + \delta A\vec{x} + A\delta\vec{x} + \delta A\delta\vec{x} = \vec{b}$	expanded
$\delta A\vec{x} + A\delta\vec{x} + \delta A\delta\vec{x} = 0$	used $A\vec{x} = \vec{b}$
$\delta A\vec{x} + A\delta\vec{x} = 0$	dropped doubly infinitesimal term

Now,

$$\delta\vec{x} = -A^{-1}(\delta A\vec{x}).$$

Condition of a System of Equations — 2 of 2

From $\delta\vec{x} = -A^{-1}(\delta A\vec{x})$ we get

$$\begin{aligned}
\|\delta\vec{x}\| &\leq \|A^{-1}\|\,\|\delta A\|\,\|\vec{x}\| \\
\frac{\|\delta\vec{x}\|}{\|\vec{x}\|} &\leq \|A^{-1}\|\,\|\delta A\|\,\underbrace{\frac{\|A\|}{\|A\|}}_{1} \\
\frac{\|\delta\vec{x}\|}{\|\vec{x}\|}\Big/\frac{\|\delta A\|}{\|A\|} &\leq \|A^{-1}\|\,\|A\|
\end{aligned}$$

Hence, the condition number of the problem of computing $\vec{x} = A^{-1}\vec{b}$, with respect to perturbations in A, is bounded by $\kappa(A)$.

From the earlier discussion, we know that the condition number of the problem of computing $\vec{x} = A^{-1}\vec{b}$, with respect to perturbations in $\vec{b}$, is bounded by $\kappa(A)$.

Homework #5 — Due Date in Canvas/Gradescope

Homework #5: Trefethen-&-Bau-12.3(a,b,c)

Note for HW#4: Please, don't print any (80×80)-matrices!

Homework AI-Policy Spring 2024

AI-era Policies — SPRING 2024

AI-3 Documented: *Students can use AI in any manner for this assessment or deliverable, but they must provide appropriate documentation for all AI use.*

This applies to ALL MATH-543 WORK during the SPRING 2024 semester.

The goal is to leverage existing tools and resources to generate HIGH QUALITY SOLUTIONS to all assessments.

You MUST document what tools you use and HOW they were used (including prompts); AND how results were VALIDATED.

BE PREPARED to DISCUSS homework solutions and AI-strategies. **Participation in the in-class discussions will be an essential component of the grade for each assessment.**