Numerical Matrix Analysis
Lecture Notes #9 — Conditioning and Stability
Conditioning and Condition Numbers

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 2016
Outline

1. Looking Back...
   - Things We Ignored, or Waved Our Hands At
   - Conditioning
   - Conditioning of a Problem — “The Intrinsic Difficulty”

2. Conditioning Examples
   - Cancellation; Polynomial Roots; Matrix Eigenvalues
   - Conditioning of Fundamental Linear Algebra Operations
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.
For us a (linear algebra) problem is a function

\[ f : X \rightarrow Y \]

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

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

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

Clearly, we must quantify what “small” and “large” mean...
Absolute Condition Number

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

$$
\delta f \overset{\text{def}}{=} f(\bar{x} + \delta \bar{x}) - f(\bar{x})
$$

be the corresponding change in $f$.

The **absolute condition number** $\hat{\kappa}(\bar{x})$ of the problem $f$ at $\bar{x}$ is defined as

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

**Think:** the supremum over small perturbations.
Absolute Condition Number

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

$$\delta f \overset{\text{def}}{=} f(\bar{x} + \delta \bar{x}) - f(\bar{x})$$

be the corresponding change in $f$.

The absolute condition number $\hat{\kappa}(\bar{x})$ of the problem $f$ at $\bar{x}$ is defined as

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

Think: the supremum over small perturbations.

For notational convenience we usually drop the limit, and write

$$\hat{\kappa}(\bar{x}) = \sup_{\delta \bar{x}} \frac{\|\delta f\|}{\|\delta \bar{x}\|}$$

with the understanding that $\delta \bar{x}$ and $\delta f$ are infinitesimal.
Looking Back...
Conditioning Examples
Things We Ignored, or Waved Our Hands At
Conditioning of a Problem — “The Intrinsic Difficulty”

Differentiability and the Absolute Condition Number

If \( f \) is differentiable, we can define the **Jacobian**

\[
J(\bar{x}) = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\
\frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n}
\end{bmatrix}.
\]

Now, we have

\[
\delta f = J(x) \delta \bar{x}, \quad \|\delta \bar{x}\| \rightarrow 0,
\]

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

\[
\hat{\kappa} = \|J(\bar{x})\|.
\]
The relative condition number $\kappa(\bar{x})$ of the problem is defined as

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

or, compactly,

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

If/When $f$ is differentiable we get

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

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

Another reason is that even if the absolute condition number is small, the relative condition number can still be large, if \( \frac{\|f(\bar{x})\|}{\|\bar{x}\|} \) is small. Here, a small absolute perturbation of \( f(\bar{x}) \) may make the result \( f(\bar{x} + \delta \bar{x}) \) almost completely independent of \( f(\bar{x}) \), *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 \( \kappa \sim 10^6, 10^{16} \) then it is “large,” and the problem is ill-conditioned.
Example: Quantifying “Cancellation Error”

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

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

The Jacobian of $f$ is

$$J(\bar{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 1 & -1 \end{bmatrix}.$$
We consider the problem $f : \mathbb{C}^2 \to \mathbb{C}$ defined by

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

The Jacobian of $f$ is

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

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

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

Now, if $x_1 \approx x_2$, the problem is clearly ill-conditioned; otherwise it is well-conditioned.
Example: Quantifying “Cancellation Error”

\[ \log_{10}(\kappa) \text{ for Subtraction} \]
Looking Back...
Conditioning Examples

Conditioning of Fundamental Linear Algebra Operations

Wilkinson’s Classic Example

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

\[
x^2 - 2x + 1 = (x - 1)^2
\]
\[
x^2 - 2x + 0.9999 = (x - 0.99)(x - 1.01)
\]
\[
x^2 - 2x + 0.999999 = (x - 0.999)(x - 1.001)
\]

Here, a perturbation of $10^{-4}$ ($10^{-6}$) in one coefficient moved both roots $10^{-2}$ ($10^{-3}$).
Wilkinson’s Classic Example

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

\[ x^2 - 2x + 1 = (x - 1)^2 \]
\[ x^2 - 2x + 0.9999 = (x - 0.99)(x - 1.01) \]
\[ x^2 - 2x + 0.999999 = (x - 0.999)(x - 1.001) \]

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

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

\[
\lim_{\delta(\text{coeff}) \to 0} \frac{d}{d[\delta(\text{coeff})]} \sqrt{\delta(\text{coeff})} = \lim_{\delta(\text{coeff}) \to 0} \frac{1}{2\sqrt{\delta(\text{coeff})}} \rightarrow \infty
\]

the condition number $\kappa = \infty$. 

Peter Blomgren, (blomgren.peter@gmail.com)
Even when the roots are single (distinct), polynomial root-finding is ill-conditioned:

If the $i$th coefficient $a_i$ of a polynomial $p(x)$ is perturbed by $\delta a_i$, the perturbation of the $j$th 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 $\kappa_{ji}$ is the condition number of $x_j$ with respect to perturbation of the coefficient $a_i$.

This number can be very large.
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 \( \mathcal{N}(0,1) \) distribution (mean zero, variance 1).
Wilkinson’s Classic Example
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 basis of expansion $\{x^k\}_{k=0,1,\ldots}$.

Using, e.g., the 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.

Consider the two matrices

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

\[
\lambda = \{1, 1\}, \quad \vec{u}_1 = \begin{bmatrix}
1 \\
0
\end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix}
-1 \\
0
\end{bmatrix}
\]
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

\[
\lambda = \{1, 1\}, \quad \tilde{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \tilde{u}_2 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}
\]

The eigenvalues and eigenvectors of \( A_2 \) are

\[
\lambda = \{2, 0\}, \quad \tilde{u}_1 = \begin{bmatrix} 0.999999500000037 \\ 0.00099999950000 \end{bmatrix}, \quad \tilde{u}_2 = \begin{bmatrix} -0.999999500000037 \\ 0.00099999950000 \end{bmatrix}
\]

Clearly, this problem is quite ill-conditioned. If \( A \) is symmetric, then the eigenvalues are well-conditioned, with \( \kappa = \|A\|_2/|\lambda| \).
Next, we take a closer look at the conditioning of the basic building blocks of linear algebra

- Matrix-Vector multiplications, \( \bar{y} := A\bar{x} \)
- A Matrix, \( A \in \mathbb{C}^{m \times n} \)
- A system of equations, \( A\bar{x} = \bar{b} \)
Conditioning of Matrix-Vector Multiplication

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

From the definition of the condition number we have

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

that is

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

If $A$ is square and non-singular, we can use the fact\(^\dagger\) that $\frac{\|\bar{x}\|}{\|A\bar{x}\|} \leq \|A^{-1}\|$ and get a bound independent of $\bar{x}$

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

\(^\dagger\) $\|\bar{x}\| = \|A^{-1}A\bar{x}\| \leq \|A^{-1}\| \|A\bar{x}\|$.
Theorem

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

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

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

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

with respect to perturbations in $\bar{b}$. If $\|\cdot\| = \|\cdot\|_2$ equalities hold if $\bar{x}$ is a multiple of a right singular vector of $A$ corresponding to the minimal singular value $\sigma_m$, and if $\bar{b}$ is a multiple of a left singular vector of $A$ corresponding to the maximal singular value $\sigma_1$. 
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

$$\kappa \leq \|A\|_2 \|(A^* A)^{-1} A^*\|_2 \leq \|A\|_2 \|(A^* A)^{-1}\|_2 \|A^*\|_2$$

$$= \frac{1}{\sigma_n^2} \sigma^2 \frac{\sigma_1}{\sigma_n} = \left[\frac{\sigma_1}{\sigma_n}\right]^2$$
Conditioning of Matrix-Vector Multiplication

<table>
<thead>
<tr>
<th>Approach</th>
<th>Pseudo-inverse</th>
<th>Conditioning(^\dagger)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal Equations</td>
<td>(A^\dagger = (A^* A)^{-1} A^*)</td>
<td>(\kappa(\bar{x}</td>
</tr>
<tr>
<td>QR-Factorization</td>
<td>(A^\dagger = R^{-1} Q^*)</td>
<td>(\kappa(\bar{x}</td>
</tr>
<tr>
<td>SVD</td>
<td>(A^\dagger = V \Sigma U^*)</td>
<td>(\kappa(\bar{x}</td>
</tr>
</tbody>
</table>

\(^\dagger\) Conditioning of solving for \(\bar{x} = A^\dagger \tilde{b}\), wrt. \(\delta \tilde{b}\).
The product $\|A\| \|A^{-1}\|$ is ubiquitous in numerical analysis, and has its own name — the **condition number** of the matrix $A$;

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

relative to the norm $\| \cdot \|$.

In this instance, the condition number is attached to the matrix $A$, not (as earlier) to a problem.
The product $\|A\| \|A^{-1}\|$ is ubiquitous in numerical analysis, and has its own name — the **condition number** of the matrix $A$;

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

Peter Blomgren, (blomgren.peter@gmail.com)
The Condition Number of a Matrix: Comments

Geometrically $\kappa(A)$ is the eccentricity of the hyper-ellipse $A S^{n-1}$ — 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

$$\kappa(A) = \frac{\sigma_1}{\sigma_n}.$$
We have considered $A\bar{x} = \bar{b}$ where $A$ was fixed, and we perturbed either $\bar{x}$ or $\bar{b}$ and looked at the effect on $\bar{b}$ or $\bar{x}$.

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

$$(A + \delta A)(\bar{x} + \delta \bar{x}) = \bar{b}$$

$A\bar{x} + \delta A\bar{x} + A\delta \bar{x} + \delta A\delta \bar{x} = \bar{b}$ expanded

$\delta A\bar{x} + A\delta \bar{x} + \delta A\delta \bar{x} = 0$ used $A\bar{x} = \bar{b}$

$\delta A\bar{x} + A\delta \bar{x} = 0$ dropped doubly infinitesimal term

Now,

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

$$
\|\delta \bar{x}\| \leq \|A^{-1}\| \|\delta A\| \|\bar{x}\|
$$

$$
\frac{\|\delta \bar{x}\|}{\|\bar{x}\|} \leq \|A^{-1}\| \|\delta A\|
$$

$$
\frac{\|\delta \bar{x}\|}{\|\bar{x}\|} \left/ \frac{\|\delta A\|}{\|A\|} \right. \leq \|A^{-1}\| \|A\|
$$

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

Peter Blomgren, 〈blomgren.peter@gmail.com〉

Conditioning and Condition Numbers — (25/26)
Trefethen-\&-Bau-12.3(a,b,c)

Note for HW\#4: Please, don’t print any $80 \times 80$-matrices!