AMS526: Numerical Analysis I
(Numerical Linear Algebra)
Lecture 6: Vector and Matrix Norms;
Condition Numbers

Xiangmin Jiao

Stony Brook University
Outline

1. Vector Norms
2. Matrix Norms
3. Conditioning and Condition Numbers
Definition of Norms

- Norm captures “size” of vector or “distance” between vectors
- There are many different measures for “sizes” but a norm must satisfy some requirements:

- **(positive definiteness)**
  \[ \| x \| > 0 \] if \( x \neq 0 \), and \( \| 0 \| = 0 \)

- **(absolute homogeneity)**
  \[ \| \alpha x \| = |\alpha| \| x \| \]

- **(triangle inequality)**
  \[ \| x + y \| \leq \| x \| + \| y \| \]
Definition of Norms

- Norm captures “size” of vector or “distance” between vectors
- There are many different measures for “sizes” but a norm must satisfy some requirements:

**Definition**

A *norm* is a function $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}$ that assigns a real-valued length to each vector. It must satisfy the following conditions:

1. $\| x \| > 0$ if $x \neq 0$, and $\| 0 \| = 0$, (positive definiteness)
2. $\| \alpha x \| = |\alpha| \| x \|$. (absolute homogeneity)
3. $\| x + y \| \leq \| x \| + \| y \|$. (triangle inequality)
Definition of Norms

- Norm captures “size” of vector or “distance” between vectors
- There are many different measures for “sizes” but a norm must satisfy some requirements:

**Definition**

A norm is a function $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}$ that assigns a real-valued length to each vector. It must satisfy the following conditions:

1. $\| x \| > 0$ if $x \neq 0$, and $\| 0 \| = 0$, (positive definiteness)
2. $\| \alpha x \| = |\alpha| \| x \|$. (absolute homogeneity)
3. $\| x + y \| \leq \| x \| + \| y \|$. (triangle inequality)

- An example is Euclidean length (i.e., $\| x \| = \sqrt{\sum_{i=1}^{n} |x_i|^2}$)
$p$-norms

- Euclidean length is a special case of $p$-norms, for $1 \leq p \leq \infty$, defined as
  \[ \|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p} \]

- For $p$-norms, the triangle inequality is called Minkowski Inequality
$p$-norms

- Euclidean length is a special case of $p$-norms, for $1 \leq p \leq \infty$, defined as

$$\|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

- For $p$-norms, the triangle inequality is called Minkowski Inequality

- Euclidean norm is 2-norm $\|x\|_2$ (i.e., $p = 2$)
**$p$-norms**

- Euclidean length is a special case of $p$-norms, for $1 \leq p \leq \infty$, defined as

$$
\|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p}
$$

- For $p$-norms, the triangle inequality is called Minkowski Inequality

- Euclidean norm is 2-norm $\|x\|_2$ (i.e., $p = 2$)

- 1-norm: $\|x\|_1 = \sum_{i=1}^{n} |x_i|$
**p-norms**

- Euclidean length is a special case of $p$-norms, for $1 \leq p \leq \infty$, defined as
  \[
  \|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p}
  \]
- For $p$-norms, the triangle inequality is called Minkowski Inequality
- Euclidean norm is 2-norm $\|x\|_2$ (i.e., $p = 2$)
- 1-norm: $\|x\|_1 = \sum_{i=1}^{n} |x_i|$
- $\infty$-norm: $\|x\|_\infty$. What is its value?
\( p \)-norms

- Euclidean length is a special case of \( p \)-norms, for \( 1 \leq p \leq \infty \), defined as
  \[
  \|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p}
  \]

- For \( p \)-norms, the triangle inequality is called Minkowski Inequality

- Euclidean norm is 2-norm \( \|x\|_2 \) (i.e., \( p = 2 \))

- 1-norm: \( \|x\|_1 = \sum_{i=1}^{n} |x_i| \)

- \( \infty \)-norm: \( \|x\|_\infty \). What is its value?
  
  ▶ Answer: \( \|x\|_\infty = \max_{1 \leq i \leq n} |x_i| \)
$p$-norms

- Euclidean length is a special case of $p$-norms, for $1 \leq p \leq \infty$, defined as
  \[ \|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p} \]

- For $p$-norms, the triangle inequality is called Minkowski Inequality

- Euclidean norm is 2-norm $\|x\|_2$ (i.e., $p = 2$)

- 1-norm: $\|x\|_1 = \sum_{i=1}^{n} |x_i|$

- $\infty$-norm: $\|x\|_\infty$. What is its value?
  - Answer: $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$

- Why we require $p \geq 1$? What happens if $0 \leq p < 1$?
Cauchy-Schwarz and Hölder Inequalities

- Cauchy-Schwarz inequality

\[ |x^T y| \leq \|x\|_2 \|y\|_2 \]

Simple proof: \[ |x^T y| = \|x\|_2 \|y\|_2 |\cos \theta|, \text{ where } \theta \text{ is angle between } x \text{ and } y \]

- Hölder inequality: Let \( p \) and \( q \) satisfy \( \frac{1}{p} + \frac{1}{q} = 1 \) with \( 1 \leq p, q \leq \infty \), then

\[ |x^T y| \leq \|x\|_p \|y\|_q \]

- Cauchy-Schwarz inequality is a special case of Hölder inequality
Weighted $p$-norms

- A generalization of $p$-norm is *weighted $p$-norm*, which assigns different weights (priorities) to different components.
  - It is anisotropic instead of isotropic

- Algebraically, $\|x\|_W = \|Wx\|$, where $W$ is diagonal matrix with $i$th diagonal entry $w_i \neq 0$ being weight for $i$th component.

- In other words,
  \[
  \|x\|_W = \left( \sum_{i=1}^{n} |w_i x_i|^p \right)^{1/p}
  \]

- What happens if we allow $w_i = 0$?
A-norm of Vectors

- Can we further generalize it to allow $W$ being arbitrary matrix?
- No. But we can generalize $W$ to be an arbitrary nonsingular matrix.
- Given a positive definite matrix $A \in \mathbb{R}^{n \times n}$, the $A$-norm on $\mathbb{R}^n$ is
  \[ \|x\|_A = \sqrt{x^T A x} \]
- Note: Weighted $p$-norm with $W$ is $A$-norm with $A = W^2$.
- These conventions are somewhat inconsistent, but they are both commonly used in the literature.
Outline

1. Vector Norms

2. Matrix Norms

3. Conditioning and Condition Numbers
Frobenius Norm

- One can define a norm by viewing \( m \times n \) matrix as vectors in \( \mathbb{R}^{mn} \)
- One useful norm is Frobenius norm (a.k.a. Hilbert-Schmidt norm)
\[
\| A \|_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} |a_{ij}|^2} = \sqrt{\sum_{j=1}^{n} \| a_j \|_2^2}
\]
i.e., 2-norm of \( nm \)-vector
- Furthermore,
\[
\| A \|_F = \sqrt{\text{tr}(A^T A)}
\]
where \( \text{tr}(B) \) denotes trace of \( B \), the sum of its diagonal entries
Frobenius Norm

- One can define a norm by viewing $m \times n$ matrix as vectors in $\mathbb{R}^{mn}$
- One useful norm is Frobenius norm (a.k.a. Hilbert-Schmidt norm)

$$\|A\|_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} |a_{ij}|^2} = \sqrt{\sum_{j=1}^{n} \|a_j\|_2^2}$$

i.e., 2-norm of $nm$-vector

- Furthermore,

$$\|A\|_F = \sqrt{\text{tr}(A^T A)}$$

where $\text{tr}(B)$ denotes trace of $B$, the sum of its diagonal entries

- Note that for $A \in \mathbb{R}^{n \times \ell}$ and $B \in \mathbb{R}^{\ell \times m}$,

$$\|AB\|_F \leq \|A\|_F \|B\|_F$$

because

$$\|AB\|_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} |a_i^T b_j|^2 \leq \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\|a_i\|_2 \|b_j\|_2\right)^2 = \|A\|_F^2 \|B\|_F^2$$
General Definition of Matrix Norms

- However, viewing $m \times n$ matrix as vectors in $\mathbb{R}^{mn}$ is not always useful, because matrix operations do not behave this way.

- Similar to vector norms, *general matrix norms* has the following properties (for $A, B \in \mathbb{R}^{m \times n}$)

  1. $\|A\| \geq 0$, and $\|A\| = 0$ only if $A = 0$,
  2. $\|A + B\| \leq \|A\| + \|B\|$,
  3. $\|\alpha A\| = |\alpha|\|A\|$.

- In addition, a matrix norm for $A, B \in \mathbb{R}^{n \times n}$ typically satisfies

  $\|AB\| \leq \|A\|\|B\|$, \quad \text{(submultiplicativity)}

  which is a generalization of Cauchy-Schwarz inequality.

- Note that Watkins defines matrix norms only $n \times n$ matrices, but we define them for $m \times n$ matrices, which is in fact important.
Norms Induced by Vector Norms

- Matrix norms can be *induced* from vector norms, which can better capture behaviors of matrix-vector multiplications.

**Definition**

Given vector norms $\| \cdot \|_{(n)}$ and $\| \cdot \|_{(m)}$ on domain and range of $A \in \mathbb{R}^{m \times n}$, respectively, the induced matrix norm $\| A \|_{(m,n)}$ is the smallest number $C \in \mathbb{R}$ for which the following inequality holds for all $x \in \mathbb{R}^n$:

$$\| A x \|_{(m)} \leq C \| x \|_{(n)}.$$
Norms Induced by Vector Norms

- Matrix norms can be *induced* from vector norms, which can better capture behaviors of matrix-vector multiplications.

**Definition**

Given vector norms \( \| \cdot \|_{(n)} \) and \( \| \cdot \|_{(m)} \) on domain and range of \( A \in \mathbb{R}^{m \times n} \), respectively, the induced matrix norm \( \| A \|_{(m,n)} \) is the smallest number \( C \in \mathbb{R} \) for which the following inequality holds for all \( x \in \mathbb{R}^n \):

\[
\| Ax \|_{(m)} \leq C \| x \|_{(n)}.
\]

- In other words, it is supremum of \( \| Ax \|_{(m)}/\| x \|_{(n)} \) for all \( x \in \mathbb{R}^n \\backslash \{0\} \)
- Maximum factor by which \( A \) can “stretch” \( x \in \mathbb{R}^n \)

\[
\| A \|_{(m,n)} = \sup_{x \in \mathbb{R}^n, x \neq 0} \| Ax \|_{(m)}/\| x \|_{(n)} = \sup_{x \in \mathbb{R}^n, \| x \|_{(n)} = 1} \| Ax \|_{(m)}.
\]
Norms Induced by Vector Norms

- Matrix norms can be *induced* from vector norms, which can better capture behaviors of matrix-vector multiplications

**Definition**

Given vector norms $\| \cdot \|_n$ and $\| \cdot \|_m$ on domain and range of $A \in \mathbb{R}^{m \times n}$, respectively, the induced matrix norm $\| A \|_{(m,n)}$ is the smallest number $C \in \mathbb{R}$ for which the following inequality holds for all $x \in \mathbb{R}^n$:

$$\| Ax \|_m \leq C \| x \|_n.$$

- In other words, it is supremum of $\| Ax \|_m / \| x \|_n$ for all $x \in \mathbb{R}^n \setminus \{0\}$
- Maximum factor by which $A$ can “stretch” $x \in \mathbb{R}^n$

$$\| A \|_{(m,n)} = \sup_{x \in \mathbb{R}^n, x \neq 0} \| Ax \|_m / \| x \|_n = \sup_{x \in \mathbb{R}^n, \| x \|_n = 1} \| Ax \|_m.$$

- Is vector norm consistent with matrix norm of $m \times 1$-matrix?
1-norm

- By definition

\[ \|A\|_1 = \sup_{x \in \mathbb{R}^n, \|x\|_1 = 1} \|Ax\|_1 \]

What is it equal to?

- Maximum of 1-norm of column vectors of \( A \)
- Or maximum of column sum of absolute values of \( A \), "column-sum norm"

To show it, note that for \( x \in \mathbb{R}^n \) and \( \|x\|_1 = 1 \)

\[ \|Ax\|_1 \leq n \sum_{j=1}^{n} |x_j| \|a_j\|_1 \leq \max_{1 \leq j \leq n} \|a_j\|_1 \|x\|_1 \]

Let \( k = \arg \max_{1 \leq j \leq n} \|a_j\|_1 \), then

\[ \|Ae_k\|_1 = \|a_k\|_1 \]

so max \( 1 \leq j \leq n \) \( \|a_j\|_1 \) is tight upper bound.
1-norm

- By definition
  \[ \|A\|_1 = \sup_{x \in \mathbb{R}^n, \|x\|_1 = 1} \|Ax\|_1 \]

- What is it equal to?
1-norm

By definition

\[ \|A\|_1 = \sup_{x \in \mathbb{R}^n, \|x\|_1 = 1} \|Ax\|_1 \]

What is it equal to?

- Maximum of 1-norm of column vectors of \( A \)
- Or maximum of column sum of absolute values of \( A \), “column-sum norm”
1-norm

- By definition
  \[ \|A\|_1 = \sup_{x \in \mathbb{R}^n, \|x\|_1 = 1} \|Ax\|_1 \]

- What is it equal to?
  - Maximum of 1-norm of column vectors of \(A\)
  - Or maximum of column sum of absolute values of \(A\), “column-sum norm”

- To show it, note that for \(x \in \mathbb{R}^n\) and \(\|x\|_1 = 1\)

\[
\|Ax\|_1 = \left\| \sum_{j=1}^{n} x_j a_j \right\|_1 \leq \sum_{j=1}^{n} |x_j| \|a_j\|_1 \leq \max_{1 \leq j \leq n} \|a_j\|_1 \|x\|_1
\]

- Let \(k = \arg \max_{1 \leq j \leq n} \|a_j\|_1\), then \(\|Ae_k\|_1 = \|a_k\|_1\), so \(\max_{1 \leq j \leq n} \|a_j\|_1\) is tight upper bound
∞-norm

- By definition
  \[ \|A\|_\infty = \sup_{x \in \mathbb{R}^n, \|x\|_\infty = 1} \|Ax\|_\infty \]

- What is \( \|A\|_\infty \) equal to?

- Maximum of 1-norm of column vectors of \( A^T \)
- Or maximum of row sum of absolute values of \( A \), “row-sum norm”

To show it, note that for \( x \in \mathbb{R}^n \) and \( \|x\|_\infty = 1 \)

\[ \|Ax\|_\infty = \max_{1 \leq i \leq m} |a_{i,\cdot}^T x| \leq \max_{1 \leq i \leq m} \|a_{i,\cdot}^T\|_1 \|x\|_\infty \]

where \( a_{i,\cdot} \) denotes the \( i \)th row vector of \( A \) and \( \|a_{i,\cdot}^T\|_1 = \sum_{j=1}^n |a_{ij}| \) is a tight bound.

Which vector can we choose to reach the bound?
\( \infty \)-norm

- By definition

\[
\|A\|_\infty = \sup_{x \in \mathbb{R}^n, \|x\|_\infty = 1} \|Ax\|_\infty
\]

- What is \( \|A\|_\infty \) equal to?
  - Maximum of 1-norm of column vectors of \( A^T \)
  - Or maximum of row sum of absolute values of \( A \), “row-sum norm”

- To show it, note that for \( x \in \mathbb{R}^n \) and \( \|x\|_\infty = 1 \)

\[
\|Ax\|_\infty = \max_{1 \leq i \leq m} |a_{i,:}|x| \leq \max_{1 \leq i \leq m} \|a_{i,:}^T\|_1 \|x\|_\infty
\]

where \( a_{i,:} \) denotes \( i \)th row vector of \( A \) and \( \|a_{i,:}^T\|_1 = \sum_{j=1}^n |a_{ij}| \)

- Furthermore, \( \|a_{i,:}^T\|_1 \) is a tight bound.

- Which vector can we choose to reach the bound?
2-norm

- What is 2-norm of a matrix?

Answer: Its largest singular value, which we will explain in later lectures.

What is 2-norm of a diagonal matrix $D$?

Answer: $\|D\|_2 = \max_{n} \{ |d_{ii}| \}$

What is 2-norm of rank-one matrix $uv^T$? Hint: Use Cauchy-Schwarz inequality.

Answer: $\|uv^T\|_2 = \|u\|_2 \|v\|_2$. 
2-norm

- What is 2-norm of a matrix?
  Answer: Its largest singular value, which we will explain in later lectures

- What is 2-norm of a diagonal matrix $D$?
  Answer: $\|D\|_2 = \max_{i} \{ |d_{ii}| \}$

- What is 2-norm of rank-one matrix $uv^T$? Hint: Use Cauchy-Schwarz inequality.
  Answer: $\|uv^T\|_2 = \|u\|_2 \|v\|_2$
2-norm

- What is 2-norm of a matrix?
  Answer: Its largest singular value, which we will explain in later lectures

- What is 2-norm of a diagonal matrix $D$?
What is 2-norm of a matrix?
Answer: Its largest singular value, which we will explain in later lectures

What is 2-norm of a diagonal matrix $D$?
Answer: $\|D\|_2 = \max_{i=1}^{n} \{|d_{ii}|\}$
2-norm

- What is 2-norm of a matrix? Answer: Its largest singular value, which we will explain in later lectures.

- What is 2-norm of a diagonal matrix $D$? Answer: $\|D\|_2 = \max_{i=1}^{n} \{|d_{ii}|\}$

- What is 2-norm of rank-one matrix $uv^T$? Hint: Use Cauchy-Schwarz inequality.
What is 2-norm of a matrix?
Answer: Its largest singular value, which we will explain in later lectures.

What is 2-norm of a diagonal matrix $D$?
Answer: $\|D\|_2 = \max_{i=1}^{n}\{|d_{ii}|\}$

What is 2-norm of rank-one matrix $uv^T$? Hint: Use Cauchy-Schwarz inequality.
Answer: $\|uv^T\|_2 = \|u\|_2 \|v\|_2$. 
Bounding Matrix-Matrix Multiplication

Let $A$ be an $l \times m$ matrix and $B$ an $m \times n$ matrix, then

$$\|AB\|_{(l,n)} \leq \|A\|_{(l,m)} \|B\|_{(m,n)}$$
Bounding Matrix-Matrix Multiplication

Let $A$ be an $l \times m$ matrix and $B$ an $m \times n$ matrix, then

$$\|AB\|_{(l,n)} \leq \|A\|_{(l,m)} \|B\|_{(m,n)}$$

To show it, note for $x \in \mathbb{R}^n$

$$\|ABx\|_{(l)} \leq \|A\|_{(l,m)} \|Bx\|_{(m)} \leq \|A\|_{(l,m)} \|B\|_{(m,n)} \|x\|_{(n)},$$
Bounding Matrix-Matrix Multiplication

- Let $A$ be an $l \times m$ matrix and $B$ an $m \times n$ matrix, then

$$\|AB\|_{(l,n)} \leq \|A\|_{(l,m)} \|B\|_{(m,n)}$$

- To show it, note for $x \in \mathbb{R}^n$

$$\|ABx\|_{(l)} \leq \|A\|_{(l,m)} \|Bx\|_{(m)} \leq \|A\|_{(l,m)} \|B\|_{(m,n)} \|x\|_{(n)},$$

- In general, this inequality is not an equality
- In particular, $\|A^p\| \leq \|A\|^p$ but $\|A^p\| \neq \|A\|^p$ in general for $p \geq 2$
Invariance under Orthogonal Transformation

- Given matrix $Q \in \mathbb{R}^{\ell \times m}$ with $\ell \geq m$. If $Q^TQ = I$, then $Qx$ for $x \in \mathbb{R}^m$ corresponds to orthogonal transformation to coordinate system in $\mathbb{R}^\ell$
- If $Q \in \mathbb{R}^{m \times m}$, then $Q$ is said to be an orthogonal matrix

**Theorem**

For any $A \in \mathbb{R}^{m \times n}$ and $Q \in \mathbb{R}^{\ell \times m}$ with $Q^TQ = I$ and $\ell \geq m$, we have

$$\|QA\|_2 = \|A\|_2 \text{ and } \|QA\|_F = \|A\|_F.$$  

In other words, 2-norm and Frobenius norms are invariant under orthogonal transformation.
Invariance under Orthogonal Transformation

- Given matrix $Q \in \mathbb{R}^{\ell \times m}$ with $\ell \geq m$. If $Q^T Q = I$, then $Qx$ for $x \in \mathbb{R}^m$ corresponds to orthogonal transformation to coordinate system in $\mathbb{R}^\ell$.
- If $Q \in \mathbb{R}^{m \times m}$, then $Q$ is said to be an orthogonal matrix.

**Theorem**

For any $A \in \mathbb{R}^{m \times n}$ and $Q \in \mathbb{R}^{\ell \times m}$ with $Q^T Q = I$ and $\ell \geq m$, we have

$$\|QA\|_2 = \|A\|_2 \text{ and } \|QA\|_F = \|A\|_F.$$ 

In other words, 2-norm and Frobenius norms are invariant under orthogonal transformation.

Proof for 2-norm: $\|Qy\|_2 = \|y\|_2$ for $y \in \mathbb{R}^m$ and therefore $\|QAx\|_2 = \|Ax\|_2$ for $x \in \mathbb{R}^n$. It then follows from definition of 2-norm.
Invariance under Orthogonal Transformation

- Given matrix $Q \in \mathbb{R}^{\ell \times m}$ with $\ell \geq m$. If $Q^T Q = I$, then $Qx$ for $x \in \mathbb{R}^m$ corresponds to orthogonal transformation to coordinate system in $\mathbb{R}^\ell$.
- If $Q \in \mathbb{R}^{m \times m}$, then $Q$ is said to be an orthogonal matrix.

**Theorem**

For any $A \in \mathbb{R}^{m \times n}$ and $Q \in \mathbb{R}^{\ell \times m}$ with $Q^T Q = I$ and $\ell \geq m$, we have

$$\|QA\|_2 = \|A\|_2 \text{ and } \|QA\|_F = \|A\|_F.$$  

In other words, 2-norm and Frobenius norms are invariant under orthogonal transformation.

**Proof for 2-norm:** $\|Qy\|_2 = \|y\|_2$ for $y \in \mathbb{R}^m$ and therefore $\|QAx\|_2 = \|Ax\|_2$ for $x \in \mathbb{R}^n$. It then follows from definition of 2-norm.

**Proof for Frobenius norm:**

$$\|QA\|_F^2 = \text{tr} \left( (QA)^T QA \right) = \text{tr} \left( A^T Q^T QA \right) = \text{tr} (A^T A) = \|A\|_F^2.$$
Outline

1. Vector Norms
2. Matrix Norms
3. Conditioning and Condition Numbers
Overview of Error Analysis

- Error analysis is important subject of numerical analysis
- Given a problem $f$ and an algorithm $\tilde{f}$ with an input $x$, the absolute error is $\|\tilde{f}(x) - f(x)\|$ and relative error is $\|\tilde{f}(x) - f(x)\|/\|f(x)\|$.
- What are possible sources of errors?

▶ Round-off error (input, computation), truncation (approximation) error

We would like the solution to be accurate, i.e., with small errors. The error depends on the property (conditioning) of the problem and the property (stability) of the algorithm.
Overview of Error Analysis

- Error analysis is an important subject of numerical analysis.
- Given a problem \( f \) and an algorithm \( \tilde{f} \) with an input \( x \), the absolute error is \( \| \tilde{f}(x) - f(x) \| \) and relative error is \( \| \tilde{f}(x) - f(x) \| / \| f(x) \| \).
- What are possible sources of errors?
  - Round-off error (input, computation), truncation (approximation) error
- We would like the solution to be accurate, i.e., with small errors.
- The error depends on property (conditioning) of the problem, property (stability) of the algorithm.
Absolute Condition Number

- Condition number is a measure of sensitivity of a problem
- Absolute condition number of a problem \( f \) at \( x \) is

\[
\hat{\kappa} = \lim_{\varepsilon \to 0} \sup_{\|\delta x\| \leq \varepsilon} \frac{\|\delta f\|}{\|\delta x\|}
\]

where \( \delta f = f(x + \delta x) - f(x) \)

- Less formally, \( \hat{\kappa} = \sup_{\delta x} \frac{\|\delta f\|}{\|\delta x\|} \) for infinitesimally small \( \delta x \)
- If \( f \) is differentiable, then

\[
\hat{\kappa} = \|J(x)\|
\]

where \( J \) is the Jacobian of \( f \) at \( x \), with \( J_{ij} = \partial f_i / \partial x_j \), and the matrix norm is induced by vector norms on \( \partial f \) and \( \partial x \)

- Question: What is absolute condition number of \( f(x) = \alpha x \)?
- Question: Is absolute condition number scale invariant?
Relative Condition Number

- **Relative condition number** of $f$ at $x$ is

\[
\kappa = \lim_{\varepsilon \to 0} \sup_{\|\delta x\| \leq \varepsilon} \frac{\|\delta f\|/\|f(x)\|}{\|\delta x\|/\|x\|}
\]

- Less formally, $\kappa = \sup_{\delta x} \frac{\|\delta f\|/\|\delta x\|}{\|f(x)\|/\|x\|}$ for infinitesimally small $\delta x$

- Different types of norms lead to different condition numbers

- If $f$ is differentiable, then $\kappa = \frac{\|J(x)\|}{\|f(x)\|/\|x\|}$

- Question: What is relative condition number of $f(x) = \alpha x$?

- Question: Is relative condition number scale invariant?

- In numerical analysis, we in general use relative condition number

- A problem is **well-conditioned** (or **ill-conditioned**) if $\kappa$ is small (or large)
Examples

- Example: Function \( f(x) = \sqrt{x} \)
Examples

- **Example: Function** $f(x) = \sqrt{x}$
  - Absolute condition number of $f$ at $x$ is $\hat{\kappa} = \|J\| = 1/(2\sqrt{x})$
    - Note: We are talking about the condition number of the problem for a given $x$
  - Relative condition number $\kappa = \frac{\|J\|}{\|f(x)\|/\|x\|} = \frac{1/(2\sqrt{x})}{\sqrt{x}/x} = 1/2$

- **Example: Function** $f(x) = x_1 - x_2$, where $x = (x_1, x_2)^T$
Examples

- **Example: Function** $f(x) = \sqrt{x}$
  - Absolute condition number of $f$ at $x$ is $\hat{\kappa} = \|J\| = 1/(2\sqrt{x})$
    - Note: We are talking about the condition number of the problem for a given $x$
  - Relative condition number $\kappa = \frac{\|J\|}{\|f(x)\|/\|x\|} = \frac{1/(2\sqrt{x})}{\sqrt{x}/x} = 1/2$

- **Example: Function** $f(x) = x_1 - x_2$, where $x = (x_1, x_2)^T$
  - Absolute condition number of $f$ at $x$ in $\infty$-norm is $\hat{\kappa} = \|J\|_\infty = \|(1, -1)\|_\infty = 2$
  - Relative condition number $\kappa = \frac{\|J\|_\infty}{\|f(x)\|_\infty / \|x\|_\infty} = \frac{2}{|x_1 - x_2| / \max\{|x_1|, |x_2|\}}$
    - $\kappa$ is arbitrarily large ($f$ is ill-conditioned) if $x_1 \approx x_2$ (hazard of cancellation error)

- **Note:** From now on, we will talk about only relative condition number