Numerical Linear Algebra

Lecture Notes 3:Norms of Vectors and Matrix

Professor: Zhihua Zhang

1 Vector Norms

Definition 1.1 (Vector Norm)**.** *A vector norm* $||\mathbf{x}||$ *is any mapping from* $\mathbb{R}^n$ *to* $\mathbb{R}$ *with the following four properties.*

- $||\mathbf{x}|| \geq 0$ *for all* $\mathbf{x} \in \mathbb{R}^n$
- $||\mathbf{x}|| = 0$ *iff* $\mathbf{x} = 0$
- $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
- $||\alpha\mathbf{x}|| = |\alpha|||\mathbf{x}||$, $\mathbf{x} \in \mathbb{R}^n$

Theorem 1.1. $||\cdot||$ *is convex.*

Proof. For any $\alpha \in [0, 1]$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$||\alpha\mathbf{x} + (1-\alpha)\mathbf{y}|| \leq ||\alpha\mathbf{x}|| + ||(1-\alpha)\mathbf{y}|| = \alpha||\mathbf{x}|| + (1-\alpha)||\mathbf{y}||$$

□

Definition 1.2 (p-norm)**.** *Let* $p \leq 1$ *be a real number.*

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

Example 1.1.

$$||\mathbf{x}||_1 = \sum_{i=1}^{n} |x_i|$$

$$||\mathbf{x}||_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}$$

$$||\mathbf{x}||_\infty = \max_i(|x_i|)$$

Definition 1.3 (0-norm)**.**

$$||\mathbf{x}||_0 = \#(i|x_i \neq 0)$$

that is a total number of non-zero elements in a vector.

Remark: Strictly speaking, $||\cdot||_0$ is not a norm as it doesn't satisfy the property 4 of the vector norm definition.

1.1 Inner Products and Norms

Theorem 1.2 (Cauchy-Schwarz Inequality)**.** *For all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*,*

$$|\mathbf{x}^T \mathbf{y}| = |<\mathbf{x}, \mathbf{y}>| \leq \sqrt{<\mathbf{x}, \mathbf{x}>}\sqrt{<\mathbf{y}, \mathbf{y}>}$$

Theorem 1.3 (Hölder's Inequality)**.** *For all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*,* $p, q \in \mathbb{R}$*, and* $\frac{1}{p} + \frac{1}{q} = 1$*, then*

$$|\mathbf{x}^T \mathbf{y}| \leq ||\mathbf{x}||_p ||\mathbf{y}||_q$$

Theorem 1.4. *For all* $\mathbf{x} \in \mathbb{R}^n$*,*

$$\begin{aligned} ||\mathbf{x}||_2 &\leq ||\mathbf{x}||_1 \leq \sqrt{n}||\mathbf{x}||_2 \\ ||\mathbf{x}||_\infty &\leq ||\mathbf{x}||_2 \leq \sqrt{n}||\mathbf{x}||_\infty \\ ||\mathbf{x}||_\infty &\leq ||\mathbf{x}||_1 \leq n||\mathbf{x}||_\infty \end{aligned}$$

Theorem 1.5. *Let* $||\cdot||_\alpha$ *and* $||\cdot||_\beta$ *be two norms on* $\mathbb{R}^n$*. There are two constants* $c_1, c_2 \geq 0$*, such that for all* $\mathbf{x} \in \mathbb{R}^n$*,*

$$c_1||\mathbf{x}||_\alpha \leq ||\mathbf{x}||_\beta \leq c_2||\mathbf{x}||_\alpha$$

1.2 Convergence

A sequence $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots$ where each $\mathbf{x}^{(i)} \in \mathbb{R}^n$ is said to converge to $\mathbf{x} \in \mathbb{R}^n$ if

$$\lim_{k \to \infty} ||\mathbf{x}^{(k)} - \mathbf{x}|| = 0$$

A sequence $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots$ is called Cauchy if for all $\epsilon > 0$ there exists a positive integer $N(\epsilon)$ such that

$$n, m \geq N(\epsilon) \implies ||\mathbf{x}^{(n)} - \mathbf{x}^{(m)}|| \leq \epsilon$$

1.3 Dual Norms

Definition 1.4. *Let* $||\cdot||$ *be a norm on* $\mathbb{R}^n$*, the function*

$$||\mathbf{y}||_D = \max_{||\mathbf{x}||=1} |\mathbf{y}^T \mathbf{x}| = \max_{||\mathbf{x}||=1} \mathbf{y}^T \mathbf{x}$$

is the dual norm of $||\cdot||$*.*

Theorem 1.6. *The dual norm is a norm.*

Proof. (1) and (4) is obvious.

(2) If $\mathbf{y} \neq 0$, then $||\mathbf{y}||_D = \max_{||\mathbf{x}||=1} |\mathbf{y}^T \mathbf{x}| \geq \left|\mathbf{y}^T \frac{\mathbf{y}}{||\mathbf{y}||}\right| = \frac{||\mathbf{y}||_2^2}{||\mathbf{y}||} \geq 0$.

(3)

$$\begin{aligned} ||\mathbf{y} + \mathbf{z}||_D &= \max_{||\mathbf{x}||=1} |(\mathbf{y} + \mathbf{z})^T \mathbf{x}| \\ &\leq \max_{||\mathbf{x}||=1} (|\mathbf{y}^T \mathbf{x}| + |\mathbf{z}^T \mathbf{x}|) \\ &\leq \max_{||\mathbf{x}||=1} |\mathbf{y}^T \mathbf{x}| + \max_{||\mathbf{x}||=1} |\mathbf{z}^T \mathbf{x}| \\ &= ||\mathbf{y}||_D + ||\mathbf{z}||_D \end{aligned}$$

□

Lemma 1.1. *Let $||\cdot||$ be a norm and $||\cdot||_D$ be its dual norm, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,*

$$||\mathbf{y}^T\mathbf{x}||_D \le ||\mathbf{x}|| \cdot ||\mathbf{y}||_D$$
$$||\mathbf{y}^T\mathbf{x}||_D \le ||\mathbf{x}||_D \cdot ||\mathbf{y}||$$

Proof.

$$||\mathbf{y}||_D = \max_{||\mathbf{z}||=1} |\mathbf{y}^T\mathbf{z}| \ge \left|\mathbf{y}^T \frac{\mathbf{x}}{||\mathbf{x}||}\right| = \frac{|\mathbf{y}^T\mathbf{x}|}{||\mathbf{x}||}$$

□

Example 1.2.

$$|\mathbf{y}^T\mathbf{x}| = |\sum_{i=1}^{n} y_i x_i| \le \sum_{i=1}^{n} |y_i| \cdot |x_i| \le \max_i |y_i| \sum_{i=1}^{n} |x_i| = ||\mathbf{y}||_\infty ||\mathbf{x}||_1$$

Example 1.3.

$$|\mathbf{x}^T\mathbf{y}| \le ||\mathbf{x}||_2 \cdot ||\mathbf{y}||_2$$

Theorem 1.7. *Let $||\cdot||$ be a vector norm on $\mathbb{R}^n$, and $||\cdot||_D$ be its dual norm, and $c > 0$ be given, then $||\mathbf{x}|| = c||\mathbf{x}||_D$ for all $\mathbf{x} \in \mathbb{R}^n$ iff $||\cdot|| = \sqrt{c}||\cdot||_2$. In particular, $||\mathbf{x}|| = ||\mathbf{x}||_D$ iff $||\cdot||$ is the l-2 norm $||\cdot||_2$.*

Proof. (a) If $||\cdot|| = \sqrt{c}||\cdot||_2$ and $\mathbf{x} \in \mathbb{R}^n$, then

$$||\mathbf{x}||_D = \max_{||\mathbf{y}||=1} |\mathbf{x}^T\mathbf{y}| = \max_{||\mathbf{y}||=1} \left|\mathbf{x}^T \frac{\sqrt{c}\mathbf{y}}{\sqrt{c}}\right| = \max_{||\mathbf{z}||_2=1} \frac{|\mathbf{x}^T\mathbf{z}|}{\sqrt{c}} = \frac{1}{\sqrt{c}}||\mathbf{x}||_{2.D} = \frac{1}{\sqrt{c}}||\mathbf{x}||_2 = \frac{1}{c}||\mathbf{x}||$$

(b) If $||\cdot|| = c||\cdot||_D$, then

$$||\mathbf{x}||_2^2 = \mathbf{x}^T\mathbf{x} \le ||\mathbf{x}|| \cdot ||\mathbf{x}||_D = \frac{1}{c}||\mathbf{x}||^2$$

So $||\mathbf{x}|| \ge \sqrt{c}||\mathbf{x}||_2$.

$$\begin{aligned}
\frac{1}{c}||\mathbf{x}|| = ||\mathbf{x}||_D &= \max_{||\mathbf{y}||=1} |\mathbf{x}^T\mathbf{y}| \\
&= \max_{\mathbf{y}\ne 0} \left|\mathbf{x}^T \frac{\mathbf{y}}{||\mathbf{y}||}\right| \\
&= \max_{\mathbf{y}\ne 0} \left|\mathbf{x}^T\mathbf{y} \frac{||\mathbf{y}||_2}{||\mathbf{y}||_2} \frac{1}{||\mathbf{y}||}\right| \\
&= \max_{\mathbf{y}\ne 0} \left|\mathbf{x}^T \frac{\mathbf{y}}{||\mathbf{y}||_2}\right| \frac{||\mathbf{y}||_2}{||\mathbf{y}||} \\
&\le \frac{1}{\sqrt{c}} \max_{\mathbf{y}\ne 0} \left|\mathbf{x}^T \frac{\mathbf{y}}{||\mathbf{y}||_2}\right| \\
&= \frac{1}{\sqrt{c}}||\mathbf{x}||_2
\end{aligned}$$

So $||\mathbf{x}|| = \sqrt{c}||\mathbf{x}||_2$ because $||\mathbf{x}|| \ge \sqrt{c}||\mathbf{x}||_2$ and $||\mathbf{x}|| \le \sqrt{c}||\mathbf{x}||_2$.

□

2 Matrix Norms

Definition 2.1. *$f : \mathbb{R}^{m\times n} \longrightarrow \mathbb{R}$ is a matrix norm if the following properties hold,*

- $f(\mathbf{A}) \geq 0$, *for all* $\mathbf{A} \in \mathbb{R}^{m\times n}$
- $f(\mathbf{A}) = 0$ *iff* $\mathbf{A} = 0$
- $f(\mathbf{A}+\mathbf{B}) \leq f(\mathbf{A}) + f(\mathbf{B})$, *for all* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$
- $f(\alpha\mathbf{A}) = |\alpha| f(\mathbf{A})$, *for all* $\alpha \in \mathbb{R}$, $\mathbf{A} \in \mathbb{R}^{m\times n}$

Definition 2.2 (F-norm)**.**

$$||\mathbf{A}||_F = \left(\sum_{i,j} A_{ij}^2\right)^{\frac{1}{2}} = tr(\mathbf{A}\mathbf{A}^T)^{\frac{1}{2}}$$

Definition 2.3. *Let $||\cdot||$ be a matrix norm, $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times p}$. we say $||\cdot||$ be consistent if $||\mathbf{AB}|| \leq ||\mathbf{A}|| \cdot ||\mathbf{B}||$.*

Remark: Not all matrix norms are consistent.

Definition 2.4. *Let $\mathbf{A} \in \mathbb{R}^{m\times n}$, $||\cdot||$ be a vector norm on $\mathbb{R}^n$, then*

$$|||\mathbf{A}||| = \sup_{\mathbf{x}\neq 0} \frac{||\mathbf{Ax}||}{||\mathbf{x}||} = \max_{||\mathbf{x}||=1} ||\mathbf{Ax}||$$

is called an operator norm or induced norm.

Theorem 2.1. *An operator norm is a consistent matrix norm.*

Proof. (1) If $\mathbf{A} \neq 0$, there exists some $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that $\mathbf{A}\hat{\mathbf{x}} \neq 0$. So we have $||\hat{\mathbf{x}}|| > 0$, $||\mathbf{A}\hat{\mathbf{x}}|| > 0$,

$$|||\mathbf{A}||| = \max_{\mathbf{x}\neq 0} \frac{||\mathbf{Ax}||}{||\mathbf{x}||} \geq \frac{||\mathbf{A}\hat{\mathbf{x}}||}{||\hat{\mathbf{x}}||} > 0$$

(2)

$$|||\alpha\mathbf{A}||| = \max_{\mathbf{x}\neq 0} \frac{||\alpha\mathbf{Ax}||}{||\mathbf{x}||} = |\alpha| \cdot |||\mathbf{A}|||$$

(3)

$$|||\mathbf{A}+\mathbf{B}||| = \max_{\mathbf{x}\neq 0} \frac{||(\mathbf{A}+\mathbf{B})\mathbf{x}||}{||\mathbf{x}||} \leq \max_{\mathbf{x}\neq 0} \frac{||\mathbf{Ax}|| + ||\mathbf{Bx}||}{||\mathbf{x}||} \leq \max_{\mathbf{x}\neq 0} \frac{||\mathbf{Ax}||}{||\mathbf{x}||} + \max_{\mathbf{x}\neq 0} \frac{||\mathbf{Bx}||}{||\mathbf{x}||} = |||\mathbf{A}||| + |||\mathbf{B}|||$$

(4)

$$||\mathbf{ABx}|| \leq |||\mathbf{A}||| \cdot ||\mathbf{Bx}|| \leq |||\mathbf{A}||| \cdot |||\mathbf{B}||| \cdot ||\mathbf{x}||$$

which means that

$$\frac{||\mathbf{ABx}||}{||\mathbf{x}||} \leq |||\mathbf{A}||| \cdot |||\mathbf{B}|||$$

is true for all $\mathbf{x} \neq 0$, so

$$|||\mathbf{AB}||| \leq |||\mathbf{A}||| \cdot |||\mathbf{B}|||$$

$\square$

Remark: Not all consistent matrix norms are operator norms.

Definition 2.5. *For $p \geq 1$,*

$$|||\mathbf{A}|||_p = \max_{\mathbf{x} \neq} \frac{||\mathbf{A}\mathbf{x}||_p}{||\mathbf{x}||_p}$$

Theorem 2.2.

$$||\mathbf{QAZ}||_F = ||\mathbf{A}||_F$$

and

$$||\mathbf{QAZ}||_2 = ||\mathbf{A}||_2$$

for $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}_m$ and $\mathbf{Z}\mathbf{Z}^T = \mathbf{Z}^T\mathbf{Z} = \mathbf{I}_n$.

Proof.

$$||\mathbf{QAZ}||_F = tr(\mathbf{QAZZ}^T\mathbf{A}^T\mathbf{Q}^T)^{\frac{1}{2}} = tr(\mathbf{QAA}^T\mathbf{Q}^T)^{\frac{1}{2}} = tr(\mathbf{AA}^T)^{\frac{1}{2}} = ||\mathbf{A}||_F$$

$$||\mathbf{QAZ}||_2 = \max_{\mathbf{x} \neq 0} \frac{||\mathbf{QAZx}||_2}{||\mathbf{x}||_2} = \max_{\mathbf{y} \neq 0} \frac{||\mathbf{QAy}||_2}{||\mathbf{y}||_2} = \max_{\mathbf{y} \neq 0} \frac{||\mathbf{Ay}||_2}{||\mathbf{y}||_2} = ||\mathbf{A}||_2$$

$\square$

Theorem 2.3.

$$|||\mathbf{A}|||_\infty = \max_{\mathbf{x} \neq 0} \frac{||\mathbf{Ax}||_\infty}{||\mathbf{x}||_\infty} = \max_i \sum_j |A_{ij}|$$

$$|||\mathbf{A}|||_1 = \max_{\mathbf{x} \neq 0} \frac{||\mathbf{Ax}||_1}{||\mathbf{x}||_1} = |||\mathbf{A}^T|||_\infty = \max_j \sum_i |A_{ij}|$$

Theorem 2.4.

$$|||\mathbf{A}|||_2 = \max_{\mathbf{x} \neq 0} \frac{||\mathbf{Ax}||_2}{||\mathbf{x}||_2} = \sqrt{\lambda_{max}(\mathbf{A}^T\mathbf{A})}$$

$$|||\mathbf{A}|||_2 = |||\mathbf{A}^T|||_2$$