

Norms, Convergence, and Derivative

Dear all, nice to meet you! This is my first RC, and maybe many of you still don't know me. I'm Yang Mo, a sophomore student majoring in ECE. In my view, VV285 is one of the most useful and interesting course among all the courses in JI. Hope you can enjoy this course!

Also, I believe different TAs will hold different styles of RC. Some may focus on advanced techniques, some on basic concepts, some on proof, etc. For me, I hope you can "visualize" all the calculations on high dimensions. So my RC will more focus on helping you to understand what happened in the high dimension space. The exercise I choose might be easy, but they can help you understand what happened.

And I know some of you are very good at math, much better than me. So you can simply do something more interesting rather than reading the slide. If you want to learn all the things in a VV186 way, it is recommended to read textbooks from the math department or look at the RC slides provided by Pingbang Hu, a previous TA who works on TCS at UIUC:

<https://github.com/sleepymalc/VV285-2021SU>

If you think you didn't have enough exercise or you think my exercises are too obvious, you can look at the RC slides from another previous TA, Yuxiang Chen:

[AlaricRFF/AlaricRFF-VV285_SU2022_Yuxiang-Chen.github.com](https://github.com/AlaricRFF/AlaricRFF-VV285_SU2022_Yuxiang-Chen.github.com)

1 Norms

The most common norm that we're familiar with is on \mathbb{R}^n

$$\|(x_i)\|_2 = \sqrt{\sum_i x_i^2} \quad (1)$$

$$\|(x_i)\|_\infty = \max\{|x_i| : 1 \leq i \leq n\} \quad (2)$$

In VV186, you are required to study some norms in Assignment 7. Do you still remember them?

Norm : 3 things that you should notice Suppose $\|\cdot\|$ is a norm

① $\|x\| \geq 0$ for $\forall x$. $\|x\|=0$ iff $x=0$

② $\|k \cdot x\| = |k| \cdot \|x\|$ for $\forall x, \forall k \in \mathbb{R}$

③ $\|x+y\| \leq \|x\| + \|y\|$

Exercise 7.4

In this exercise, $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ denote elements of \mathbb{R}^n .

- i) Use (1) of Exercise 7.3 to prove Hölders inequality:

$$\sum_{j=1}^n |x_j y_j| \leq \left(\sum_{j=1}^n |x_j|^p \right)^{1/p} \left(\sum_{j=1}^n |y_j|^q \right)^{1/q} \quad \text{for } p \in \mathbb{N} \setminus \{0\}, \quad \frac{1}{p} + \frac{1}{q} = 1$$

Hint: first deduce from Jensen's inequality that $a \cdot b \leq a^p/p + b^q/q$ for $a, b > 0$ and $p^{-1} + q^{-1} = 1$. Then let $\tilde{x} = x/\|x\|_p$, $\tilde{y} = y/\|y\|_q$ and show that $\sum |\tilde{x}_j \tilde{y}_j| \leq 1$.

(2 Marks)

- ii) Prove Minkowski's Inequality:

$$\left(\sum_{j=1}^n |x_j + y_j|^p \right)^{1/p} \leq \left(\sum_{j=1}^n |x_j|^p \right)^{1/p} + \left(\sum_{j=1}^n |y_j|^p \right)^{1/p} \quad p \in \mathbb{N} \setminus \{0\}$$

Hint: use $|x_j + y_j|^p \leq (|x_j| + |y_j|)|x_j + y_j|^{p-1}$. Then apply Hölder's inequality. It will be useful to prove that $p^{-1} + q^{-1} = 1$ implies $(p-1)q = p$.

(2 Marks)

- iii) Deduce that

$$\|x\|_p := \left(\sum_{j=1}^n |x_j|^p \right)^{\frac{1}{p}},$$

defines a norm on \mathbb{R}^n for all $p \in \mathbb{N} \setminus \{0\}$.

(2 Marks)

- iv) Show that $\|x\|_p \geq \|x\|_q$ for $p < q$, $p, q \in \mathbb{N} \setminus \{0\}$ and $x \neq 0$. *Hint:* Set $\xi := \|x\|_p$ so $|x_j|/\xi \leq 1$ for $j = 1, \dots, n$. Deduce $(|x_j|/\xi)^q \leq (|x_j|/\xi)^p$ for $p < q$ and hence

$$\sum_{j=1}^n \left(\frac{|x_j|}{\xi} \right)^q \leq \sum_{j=1}^n \left(\frac{|x_j|}{\xi} \right)^p = 1.$$

(2 Marks)

- v) Verify that $\|x\|_\infty := \max_{1 \leq j \leq n} |x_j|$ defines a norm on \mathbb{R}^n and prove that $\|x\|_\infty = \lim_{p \rightarrow \infty} \|x\|_p$ for any $x \in \mathbb{R}^n$.

Hint: Use that $\lim_{m \rightarrow \infty} \sqrt[m]{y} = 1$ for fixed $y > 0$. (See Exercise 7.3.)

(2 Marks)

I've put all the things about this exercise in the appendix and we can discuss about it.

2 Open and Close Set

Open Set

For every x in A , there exists $\varepsilon > 0$ such that $B(x, \varepsilon) \subseteq A$, where $B(x, \varepsilon)$ denotes the open ball of radius ε centered at x .

Closed Set

Formally, a set A is said to be closed if it contains all of its accumulation points.

2.1 Excercise

Prove: A bounded infinite sequence of points on \mathbb{R}^2 always has a convergent subsequence.

Remember in VV186 : bounded sequence must posses a converge subsequence.

For $(a_1, b_1), (a_2, b_2), \dots, (a_n, b_n), \dots$

\exists a subsequence a_{k_i} a_{k_i} converges to X .

For $b_{k_1}, b_{k_2}, \dots, b_{k_i}, \dots$

There exist a subsequence converges.

$b_{k_p m_1}, b_{k_p m_2}, \dots$ converges to γ
 $a_{k_p m_1}, b_{k_p m_2}, \dots$ converges to (x, γ) .

Suppose X is a complete normed space and S_1, S_2, \dots, S_n is a finite collection of subsets of X . Judge T or F :

T • If S_i 's are open, then $\cup_i S_i$ is open.

T • If S_i 's are open, then $\cap_i S_i$ is open.

T • If S_i 's are closed, then $\cup_i S_i$ is closed.

T • If S_i 's are closed, then $\cap_i S_i$ is closed.

F • If Y is a normed space and $f : X \rightarrow Y$ is continuous, then $f(O)$ is open whenever $O \subseteq X$ is open.

T • Suppose $K \subseteq X$ is compact and X is a normed space. Any continuous functions $f : K \rightarrow X$ is uniformly continuous.

① Suppose $a \in \cup_i S_i \quad \exists j \in S_j \Rightarrow \{x | \|x-a\| < \varepsilon\} \subseteq S_j$

② Suppose $a \in \cap_i S_i \quad \forall j \in S_j$

$$(a \in S_1) \wedge (a \in S_2) \wedge \dots \wedge (a \in S_n)$$

$$\Downarrow \quad \Downarrow \quad \Downarrow$$

$$B_{\varepsilon_1}(a) \in S_1 \quad B_{\varepsilon_2}(a) \in S_2 \quad B_{\varepsilon_n}(a) \in S_n$$

$$\text{let } \varepsilon = \frac{1}{2} \min \{\varepsilon_1, \dots, \varepsilon_n\}$$

$$\Rightarrow B_\varepsilon(a) \subseteq \cap_i S_i$$

③ let $O_i^c = S_i$ O_i is open $\cup_i S_i = (S_1 \cup S_2 \cup \dots \cup S_n) = (O_1^c \cup \dots \cup O_n^c) = (\underbrace{O_1 \cap \dots \cap O_n}_c)^c$ (Similar to De Morgan's Law) Open

④ $\cap_i S_i = (\cup_i O_i)^c$ is close

⑤ $X = (0, 2\pi) \quad f(x) = \sin x \quad Y = [-1, 1]$

⑥ K compact. $\Rightarrow K$ closed and bounded Suppose + not uniformly continuous,
 $\forall \varepsilon, \exists x, y |x-y| < \delta \quad |f(x)-f(y)| > \varepsilon$ not continuous.

Let X be a complete normed vector space. Then the set:

$$GL(X) := \{L \in \mathcal{L}(X, X) : L^{-1} \text{ exists}\} \quad (3)$$

Show that this set is open. Further more, show that if $\|L\| \leq 1$, then $1-L \in GL(X)$ and

$$(1-L)^{-1} = \sum_{n=0}^{\infty} L^n \quad (4)$$

where $L^0 := 1$ which is the identity.

⑦ $\sum_{n=0}^{\infty} L^n$ converge.

$$\sum_{n=0}^{\infty} \|L^n\| \leq \sum_{n=0}^{\infty} \|L\|^n \quad (\|L_1 \cdot L_2\| < \|L_1\| \cdot \|L_2\|)$$

$$= \frac{1}{1-\|L\|} \quad \text{is bounded.}$$

$$\text{For } \sum_{n=0}^{\infty} L^n, \quad 1 - \sum_{n=0}^{\infty} L^{n+1} = (1-L)(1+L+\dots+L^n)$$

$$\|(1-L) - 1\| = \|L^{n+1}\| \leq \|L\|^{n+1} \xrightarrow{n \rightarrow \infty} 0 \quad \text{So } (1-L)^{n+1} \xrightarrow{n \rightarrow \infty} 1$$

$$\Rightarrow 1 \leftarrow \sum_{n=0}^{\infty} (1-L) \left(\sum_{n=0}^{\infty} L^n \right) \Rightarrow (1-L)^{-1} = \sum_{n=0}^{\infty} L^n \quad (\|L\| < 1)$$

② $GL(x)$ is open
(so that)

Want to prove: Suppose $L_0 \in GL(x) \exists \delta > 0$ s.t. $B_\delta(L_0) \subset GL(x)$

$$B_\delta(L_0) = \{L \in \mathcal{L}(x, x) : \|L - L_0\| < \delta\}$$

Fixing L_0 . Set $\delta \in (0, \frac{1}{\|L_0^{-1}\|})$

$$\|L L_0^{-1} - 1\| = \|L L_0^{-1} - L_0 L_0^{-1}\| = \|L - L_0\| \cdot \|L_0^{-1}\| < \delta \cdot \|L_0^{-1}\| < 1$$

As $\|1 - L L_0^{-1}\| < 1$, $1 - (L L_0^{-1} - 1)$ has an inverse

\Rightarrow So $\overset{\sim}{L L_0^{-1}}$ has an inverse
invertible $\Rightarrow L$ is invertible.

$$\Rightarrow B_\delta(L_0) \subseteq GL(x).$$

3 Derivatives

Suppose $A \in Mat(n \times n : F)$, calculate their first derivatives:

- $f(x) = \langle x, Ax \rangle$
- $g(A) = \text{tr}(A^2)$

$$\begin{aligned} ① \quad f(x+h) &= \langle x+h, A(x+h) \rangle = \langle x+h, Ax+Ah \rangle = \langle x+h, Ax \rangle + \langle x+h, Ah \rangle \\ &= \langle x, Ax \rangle + \langle h, Ax \rangle + \langle x, Ah \rangle + \langle h, Ah \rangle \\ &= f(x) + \langle h, Ax \rangle + \langle x, Ah \rangle + o(h) \end{aligned}$$

$$\text{So } Df|_x h = \langle h, Ax \rangle + \langle x, Ah \rangle$$

$$\begin{aligned} ② \quad g(x+h) &= \text{tr}(A+h)^2 = \text{tr}[(A+h)(A+h)] = \text{tr} A^2 + \text{tr}(Ah) + \text{tr}(hA) + \text{tr} h^2 \\ &= g(A) + \text{tr}(Ah) + \text{tr}(hA) + \text{tr} h^2 \\ &\quad Df_A h = 2 \text{tr}(Ah). \end{aligned}$$

4 Appendix: The Exercise in VV186

4.1 1. About what is required to be a normed vector space

A vector space, or a linear space, is a set of objects which can be added and multiplied by scalars. The operation of addition and multiplication should follow a set of specific rules. For more detailed information, you can visit https://en.wikipedia.org/wiki/Vector_space.

For normed vector space, some additional requirements should be added. For a space, of course, we want to measure the distance between two elements. But, how can we measure? And is there any requirement for our method to measure?

In a normal way, we can define a distance for a 3-dimension space like that the distance between (a,b) and (c,d) is

$$\sqrt{(a-c)^2 + (b-d)^2} \quad (5)$$

However, this can be easily changed if I want to invent a new way to measure distance. Using the same example, I can also set my distance like this:

$$\sqrt[3]{(a-b)^3 + (c-d)^3} \quad (6)$$

No one can say this is problematic. So, what is distance, or what should it be like?

Well, mathematicians say that you can define a "distance" as long as it meets 3 requirements: (We denote distance as $\| \cdot \|$, and said V and U to be elements of the vector space)

1. $\| V \| \geq 0$ and $\| V \| = 0$ if and only if $V = 0$;
2. For all the scalars a , $\| a \cdot V \| = |a| \cdot \| V \|$
3. $\| U + V \| \leq \| U \| + \| V \|$

We can discover that the third one is the hardest to be satisfied. However, there's an interesting discovery: For all p belonging to natural numbers, these distances always meets the requirements: (U and V can be respectively written as (u_1, u_2, \dots, u_n) and (v_1, v_2, \dots, v_n))

$$\sqrt[p]{(u_1 - v_1)^p + (u_2 - v_2)^p + \dots + (u_n - v_n)^p} \quad (7)$$

This is even true when p goes to infinity. What happens and why?

4.2 2. Start from Jensen Inequality

When a function is convex, we can apply Jensen's Inequality to it. It estimates the properties of convex functions by assuming they behave linearly.

Jensen's Inequality tells us that, for a convex function $f(x)$ (whose domain is $[a,b]$), and $x_1, x_2, \dots, x_n \in [a,b]$, we can know that

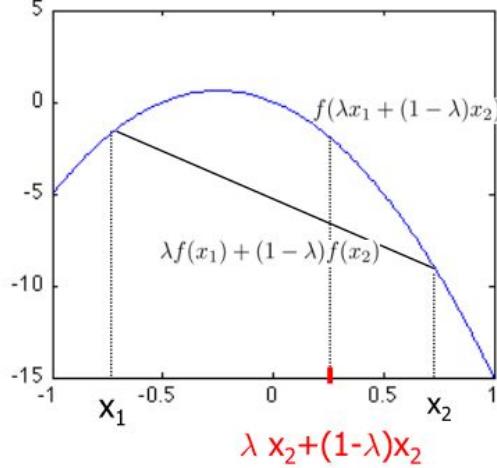
$$\frac{f(x_1) + f(x_2) + \dots + f(x_n)}{n} \geq f\left(\frac{x_1 + x_2 + \dots + x_n}{n}\right) \quad (8)$$

Similarly, you can get the reverse result if $f(x)$ is a concave function.

The detailed information can be viewed at https://en.wikipedia.org/wiki/Jensen_inequality.

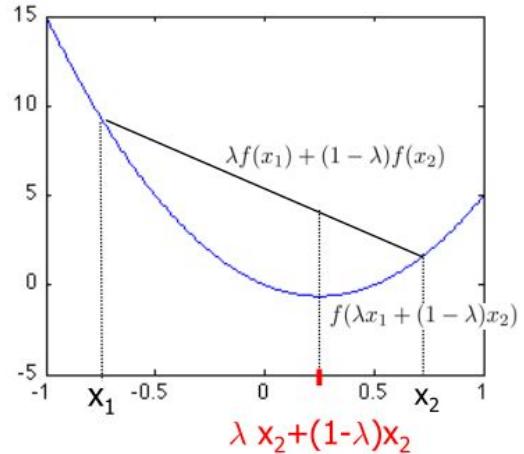
f is concave if and only

$$\forall x_1, x_2, \forall \lambda \in (0, 1), \\ f(\lambda x_1 + (1 - \lambda)x_2) \geq \lambda f(x_1) + (1 - \lambda)f(x_2)$$



f is convex if and only

$$\forall x_1, x_2, \forall \lambda \in (0, 1), \\ f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$$



4.3 3. From Jensen's inequality to Hölder's inequality

First, we need to know what's Hölder's inequality. Hölder's inequality can be written as:

$$\sum_{i=1}^n |x_i y_i| \leq \left(\sum_{j=1}^n |x_j|^p \right)^{\frac{1}{p}} \left(\sum_{j=1}^n |y_j|^q \right)^{\frac{1}{q}} \quad (9)$$

p and q should satisfy that $p \in \mathbb{N}$ and $1/p + 1/q = 1$.

4.3.1 Proof:

let

$$\tilde{x}_i = \frac{|x_i|}{\sqrt[p]{\sum_{j=1}^n |x_j|^p}}, \quad \tilde{y}_i = \frac{|y_i|}{\sqrt[q]{\sum_{j=1}^n |y_j|^q}} \quad (10)$$

We can discover for Jensen's inequality that

$$\prod_{i=1}^n a_i^{\lambda_i} \leq \sum_{i=1}^n \lambda_i a_i \quad (\sum \lambda_i = 1) \quad (11)$$

(which is easy to be proved by applying a log function on the RHS. Note that log function is concave and you can apply Jensen's inequality to it)

According to the upward formula, we can deduce that

$$\tilde{x}_i \cdot \tilde{y}_i \leq \frac{\tilde{x}_i^p}{p} + \frac{\tilde{y}_i^q}{q} \quad (\text{As } \frac{1}{p} + \frac{1}{q} = 1) \quad (12)$$

This can be transferred as

$$\frac{|x_i||y_i|}{\sqrt[p]{\sum_{j=1}^n |x_j|^p} \sqrt[q]{\sum_{j=1}^n |y_j|^q}} \leq \frac{|x_i|^p}{p \sum_{j=1}^n |x_j|^p} + \frac{|y_i|^q}{p \sum_{j=1}^n |y_j|^q} \quad (13)$$

Add all the different is up

$$\frac{\sum_{j=1}^n |x_j||y_j|}{\sqrt[p]{\sum_{j=1}^n |x_j|^p} \sqrt[q]{\sum_{j=1}^n |y_j|^q}} \leq \frac{\sum_{i=1}^n |x_i|^p}{p \sum_{j=1}^n |x_j|^p} + \frac{\sum_{i=1}^n |y_i|^q}{p \sum_{j=1}^n |y_j|^q} \quad (14)$$

As

$$\frac{\sum_{i=1}^n |x_i|^p}{p \sum_{j=1}^n |x_j|^p} + \frac{\sum_{i=1}^n |y_i|^q}{p \sum_{j=1}^n |y_j|^q} = \frac{1}{p} + \frac{1}{q} = 1 \quad (15)$$

It can be shorted as

$$\frac{\sum_{j=1}^n |x_j| |y_j|}{\sqrt[p]{\sum_{j=1}^n |x_j|^p} \sqrt[q]{\sum_{j=1}^n |y_j|^q}} \leq 1 \quad (16)$$

So we can move the denominator to the RHS. This is the same as

$$\sum_{j=1}^n |x_j y_j| \leq \left(\sum_{j=1}^n |x_j|^p \right)^{\frac{1}{p}} \left(\sum_{j=1}^n |y_j|^q \right)^{\frac{1}{q}} \quad (17)$$

Now we've got Hölder's inequality!

4.4 4. From Hölder's inequality to Minkowski' s inequality

Let's move on from Hölder's inequality to Minkowski' s Inequality.

Also, Minkowski' s Inequality can be written like this:

$$\left(\sum_{i=1}^n |x_i + y_i|^p \right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |y_i|^p \right)^{\frac{1}{p}} \quad (18)$$

4.4.1 Proof

The LHS can be written as

$$\left(\sum_{i=1}^n |x_i + y_i|^{p-1} \cdot |x_i| + \sum_{i=1}^n |x_i + y_i|^{p-1} \cdot |y_i| \right)^{\frac{1}{p}} \quad (19)$$

Applying Hölder's inequality, we can know that

$$\sum_{i=1}^n |x_i + y_i|^{p-1} \cdot |x_i| \leq \left(\sum_{i=1}^n |x_i + y_i|^p \right)^{\frac{p-1}{p}} \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \quad (\text{As } \frac{1}{p} + \frac{p-1}{p} = 1) \quad (20)$$

Similarly, we can have

$$\sum_{i=1}^n |x_i + y_i|^{p-1} \cdot |y_i| \leq \left(\sum_{i=1}^n |x_i + y_i|^p \right)^{\frac{p-1}{p}} \left(\sum_{i=1}^n |y_i|^p \right)^{\frac{1}{p}} \quad (21)$$

Add the two inequalities together, and we can get

$$\sum_{i=1}^n |x_i + y_i|^{p-1} \cdot |x_i| + \sum_{i=1}^n |x_i + y_i|^{p-1} \cdot |y_i| \leq \left(\sum_{i=1}^n |x_i + y_i|^p \right)^{\frac{p-1}{p}} \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |x_i + y_i|^p \right)^{\frac{p-1}{p}} \left(\sum_{i=1}^n |y_i|^p \right)^{\frac{1}{p}} \quad (22)$$

We can know that

$$\sum_{i=1}^n |x_i + y_i|^p \leq \left(\sum_{i=1}^n |x_i + y_i|^p \right)^{\frac{p-1}{p}} \left(\left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |y_i|^p \right)^{\frac{1}{p}} \right) \quad (23)$$

Finally, we can move the extra part on the RHS to the LHS, and we can get

$$\left(\sum_{i=1}^n |x_i + y_i|^p \right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |y_i|^p \right)^{\frac{1}{p}} \quad (24)$$

Now we've got Minkowski' s inequality!

4.5 5. What these inequalities mean when combined with normed vector space

Brilliant readers may have noticed that all these inequalities can be used to deal with vectors. We can rewrite the inequality in a vector-like way by assuming vectors X and Y to be like

$$X = (x_1, x_2, \dots, x_n), \quad Y = (y_1, y_2, \dots, y_n) \quad (25)$$

So we discover that we can define a new way to measure the distance like

$$\|X\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \quad (26)$$

Note that $\|X\|_p$ is only a way to note this is a different way of calculating distance instead of the traditional $|X|$.

According to Minkowski's inequality

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p \quad (27)$$

We can discover that this definition of distance works!

4.6 6. When p goes to infinity...

Well, in many cases, things get much more different when something goes to infinity. However, this definition still works when p goes to infinity!

We can prove that when p goes to infinity

$$\|X\|_\infty = \max\{x_1, x_2, \dots, x_n\} \quad (28)$$

This can be done quite easily using the X tilde.

4.6.1 Proof

We can put the distance into the limit

$$\lim_{p \rightarrow \infty} \|X\|_p = \lim_{p \rightarrow \infty} \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \quad (29)$$

Then we introduce X tilde

$$\tilde{x}_i = \frac{|x_i|}{\sqrt[p]{\sum_{j=1}^n |x_j|^p}} \quad (30)$$

We can know that

$$\lim_{p \rightarrow \infty} \|\tilde{X}\|_p = \lim_{p \rightarrow \infty} \left(\sum_{i=1}^n |\tilde{x}_i|^p \right)^{\frac{1}{p}} \quad (31)$$

And we know that, when

$$x_i < \max\{x_1, x_2, \dots, x_n\} \quad (32)$$

then

$$\tilde{x}_i < 1 \quad (33)$$

So

$$\lim_{p \rightarrow \infty} (|\tilde{x}_i|^p) = 0 \quad (34)$$

As

$$\lim_{p \rightarrow \infty} \left(\sum_{i=1}^n |\tilde{x}_i|^p \right)^{\frac{1}{p}} = \lim_{p \rightarrow \infty} \left(\sum_{i=1}^n \lim_{p \rightarrow \infty} |\tilde{x}_i|^p \right)^{\frac{1}{p}} \quad (35)$$

And as

$$\sum_{i=1}^n \lim_{p \rightarrow \infty} |\tilde{x}_i|^p \leq n \quad (\text{As } |\tilde{x}_i| \leq 1) \quad (36)$$

And we can know a fact that

$$\lim_{m \rightarrow \infty} \sqrt[m]{y} = 1 \text{ when } y > 0 \quad (37)$$

As

$$0 \leq \sum_{i=1}^n \lim_{p \rightarrow \infty} |\tilde{x}_i|^p \leq n \quad (38)$$

Then applying the fact

$$\lim_{p \rightarrow \infty} \left(\sum_{i=1}^n |\tilde{x}_i|^p \right)^{\frac{1}{p}} = \lim_{p \rightarrow \infty} \left(\sum_{i=1}^n \lim_{p \rightarrow \infty} |\tilde{x}_i|^p \right)^{\frac{1}{p}} = 1 \quad (39)$$

So we can change tilde back to the normal form and finally get

$$\| X \|_{\infty} = \max\{x_1, x_2, \dots, x_n\} \quad (40)$$

5 References

- VV285, slide. Horst Hohberger
- RC, slides. TA-Pingbang Hu
- RC, slides. TA-Chen Yuxiang

For further questions, you can contact me on wechat.



Fragments

芬兰



扫一扫上面的二维码图案，加我为朋友。

