↙ it appears in various places and contexts

Fundamental Theorem of Linear Algebra

It states that the size and the rank of a matrix completely define the dimensions of its 4 fundamental subspaces.

$A \in \mathbb{M}_{m\times n}$

$$\begin{tikzcd} \operatorname{Ker}A \arrow[d, phantom, "\cap"] & & \operatorname{Im}A \arrow[d, phantom, "\cap"] \\ \mathbb{R}^n \arrow[rr, shift left, "A"] & & \mathbb{R}^m \arrow[ll, shift left, "A^T"] \\ \operatorname{coIm}A \arrow[u, phantom, "\cup"] & & \operatorname{coKer}A \arrow[u, phantom, "\cup"] \end{tikzcd}$$

[redacted]
- Fund. Th.
- $\langle\cdot,\cdot\rangle$
- $\|\cdot\|$
- CS ineq.
- Δ ineq.

• ○ ●

Th Let $A \in \mathbb{M}_{m\times n}$ and $\operatorname{rank}A = r$.

$\Rightarrow$ (1) $\dim \operatorname{Ker}A = n-r$

(2) $\dim \operatorname{coIm}A = r$

(3) $\dim \operatorname{Im}A = r$

(4) $\dim \operatorname{coKer}A = m-r$

In fact, this is only the first part of the theorem.

Corollary: $\boxed{\operatorname{rank}A = \operatorname{rank}A^T}$

Indeed, $\operatorname{rank}A^T \overset{(3)}{=} \dim \operatorname{Im}A^T \overset{def}{=} \dim \operatorname{coIm}A \overset{(2)}{=} \operatorname{rank}A$

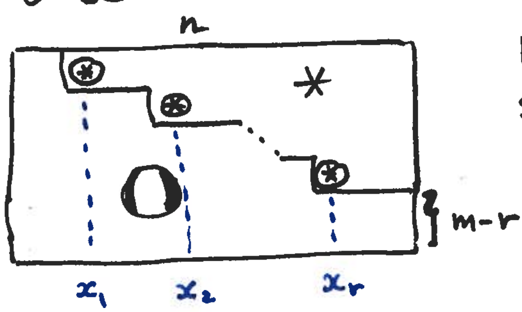
Remark: (2) & (3) give geometric interpretation of the rank. ↳ originally defined as the number of pivots.

(2): $\operatorname{rank}A$ = # lin. indep. rows of A.
(3): $\operatorname{rank}A$ = # lin. indp. columns of A] turns out these two num. are the same.

Proof

(1) Elementary row operations do not change the kernel:
Let $\bar{U}$ be the row echelon form of A:

[If $\tilde{A}$ is obtain from A by $a^i \rightsquigarrow a^i + \alpha a^j$ or $\begin{cases} a^i \rightsquigarrow a^j \\ a^j \rightsquigarrow a^i \end{cases}$ then $Ax=0 \Leftrightarrow \tilde{A}x=0$

$\bar{U} =$ 

$\operatorname{Ker}A = \operatorname{Ker}\bar{U}$
So it is enough to show that $\dim \operatorname{Ker}\bar{U} = n-r$.

Let us construct a basis in $\operatorname{Ker}\bar{U} = \{x : \bar{U}x = 0\}$
Let $x_1 \dots x_r$ denote the basic variables, corresponding to the pivots, and $x_{r+1} \dots x_n$ denote the remaining free variables.

As we know, the basic variables are expressed in terms of free variables:

$$\left(\begin{cases} x_1 = f_1(x_{r+1},\dots,x_n) \\ \vdots \\ x_r = f_r(x_{r+1},\dots,x_n) \end{cases} \quad \text{where } x_{r+1},\dots,x_n \text{ are arbitrary, and } f_1,\dots,f_r \text{ are linear functions}\right) \Leftrightarrow \bar{U}x = 0 \Leftrightarrow Ax = 0$$

Let basic free 

The claim is that $v_1 \dots v_{n-r}$ form a basis in $\operatorname{Ker}A = \ker\bar{U}$.

- $v_1 \dots v_{n-r}$ are lin. indep. (this is obvious)
- $\operatorname{span}(v_1 \dots v_{n-r}) = \operatorname{Ker}\bar{U}$. Let $y \in \operatorname{Ker}\bar{U}$.

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_r \\ y_{r+1} \\ \vdots \\ y_n \end{bmatrix} \begin{matrix} = f_1(y_{r+1}\cdots y_n) \\ \\ = f_r(y_{r+1}\cdots y_n) \\ \\ \\ \\ \end{matrix}$$

Then $y = \underbrace{y_{r+1}v_1 + \dots + y_n v_{n-r}}_{z}$

The last $(n-r)$ components of y and z are obviously the same.

Indeed: [$z_1 = y_{r+1}f_1(1,0,\dots 0) + \dots + y_n f_1(0,\dots 0,1) = f_1(y_{r+1},\dots,y_n) = y_1$
Similarly, $z_k = y_k \ \forall k$.

Therefore $\dim \operatorname{Ker}A = n-r$.

(2) Elementary row operations $\underbrace{\text{don't change}}_{\text{this is obvious, convince yourself.}}$ $\operatorname{coIm} A = \operatorname{span}\{a^1, \dots, a^m\}$.

Let us construct a basis in $\operatorname{coIm} A = \operatorname{coIm} \tilde{U}$.

Let $u^1, \dots, u^r$ be the first r pivot rows of $\tilde{U}$. They form a basis in $\operatorname{co Im} \tilde{U}$.

- $u^1 \dots u^r$ are linearly independent (this is obvious)
- $\operatorname{span}(u^1, \dots, u^r) = \operatorname{co Im} \tilde{U}$. (this is also obvious since the remaining rows of $\tilde{U}$ are all zero rows)

$\Rightarrow \dim \operatorname{co Im} A = r$.

(3) We claim that $\underbrace{a_1, \dots, a_r}_{\text{colums of } A \text{ that correspond to the pivots.}}$ form a basis of $\operatorname{Im} A = \operatorname{span}(a_1, \dots, a_n)$

- $\operatorname{span}(a_1, \dots, a_r) = \operatorname{Im} A$

  Let $b \in \operatorname{Im} A \Rightarrow \exists x : Ax = b \Rightarrow$ $b = A \cdot \begin{array}{l} \text{basic}\left\{\begin{matrix} \\ \\ \end{matrix}\right. \\ \text{free}\left\{\begin{matrix} \\ \\ \end{matrix}\right. \end{array} \begin{bmatrix} x_1 \\ \vdots \\ x_r \\ x_{r+1} \\ \vdots \\ x_n \end{bmatrix} = A \cdot \begin{bmatrix} g_1(x_{r+1} \dots x_n) \\ \vdots \\ g_r(x_{r+1} \dots x_n) \\ x_{r+1} \\ \vdots \\ x_n \end{bmatrix}$

  Free variables can take any values.

  Let us set $x_{r+1} = \dots = x_n = 0$.

$$\Rightarrow \quad b = A \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_r \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \tilde{x}_1 a_1 + \dots + \tilde{x}_r a_r \in \operatorname{span}(a_1 \dots a_r).$$

$\tilde{x}_i = g_i(0, \dots, 0)$

- $a_1, \dots, a_r$ are linearly independent.

  Suppose that $\alpha_1 a_1 + \dots + \alpha_r a_r = 0$ (we want to show that all $\alpha_i = 0$)

  $\Rightarrow A \cdot \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_r \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0$ But since all free variables are zeros, so are all basic variables:

  $\alpha_1 = f_1(0, \dots, 0) = 0$
  $\vdots$
  $\alpha_r = f_r(0, \dots 0) = 0$

  $\Rightarrow$ only trivial lin. combination of $a_1 \dots a_r$ is zero

  $\Rightarrow a_1 \dots a_r$ are lin. indep.

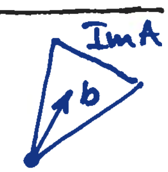
Thus, $\dim \operatorname{Im} A = r$.

(4) This is straightforward now.

$$\dim \underbrace{\operatorname{coKer} A} = \dim \underbrace{\operatorname{Ker} A^T} \overset{(1)}{=} m - \operatorname{rank}(A^T) \overset{\text{corollary of (2) and (3).}}{=} m - r.$$

$\square$

Let $A \in \mathbb{M}_{m \times n}$. Consider system $Ax = b$.



- $Ax = b$ is consistent (solution exists) $\Longleftrightarrow b \in \operatorname{Im} A = \operatorname{span}(a_1 \dots a_n) \Longleftrightarrow \operatorname{span}(a_1 \dots a_n, b) = \operatorname{span}(a_1 \dots a_n)$

  $\Longleftrightarrow \underbrace{\dim \operatorname{span}(a_1 \dots a_n, b)}_{\operatorname{rank}[A|b]} = \underbrace{\dim \operatorname{span}(a_1 \dots a_n)}_{\operatorname{rank} A}$

Suppose $Ax = b$ is consistent.

General solution: $x = x^* + \operatorname{Ker} A$ $\leftarrow$ superposition principle.

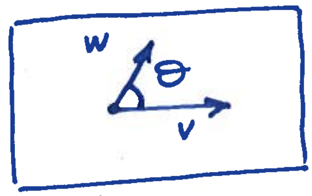
- Sol. unique $\Longleftrightarrow \operatorname{Ker} A = \{0\} \Longleftrightarrow \dim \operatorname{Ker} A = 0 \Longleftrightarrow n = \operatorname{rank} A$
- Inf. many $\Longleftrightarrow \dim \operatorname{Ker} A > 0 \Longleftrightarrow \operatorname{rank} A < n$

Let V be a vector space. So far we can add vectors in V and multiply them by scalars. But we don't know how to measure the length af a vector and how to compute the angle between two vectors. For example, if $V = C[a,b]$, the vector space of all continuous functions on $[a,b]$, what is the length of $f(x) = \sin x$? Or, what is the angle between $f(x) = \sin x$ and $g(x) = \cos x$?

this is what we need in applicat.

Recall that in $\mathbb{R}^n$, our prototypical example of a vector space, we can compute lengths and angles: if $v = (v_1, \dots v_n)^T \in \mathbb{R}^n$ and $w = (w_1, \dots w_n)^T \in \mathbb{R}^n$, then

- Dot product $\quad v \cdot w = v_1 w_1 + \dots + v_n w_n = \sum_{i=1}^{n} v_i w_i$
- Length $\left(\begin{matrix}\text{Euclidean}\\ \text{norm}\end{matrix}\right)$ $\quad \|v\| = \sqrt{v \cdot v} = \sqrt{\sum_{i=1}^{n} v_i^2}$
- Angle between v and w : $\quad \cos\theta = \dfrac{v \cdot w}{\|v\| \, \|w\|}$



$\mathbb{R}^n$

The concepts of inner product (generalization of the dot product) and norm (generalization of length) allows to compute lengths and angles in general vector spaces. They induce geometry on vector spaces.

Remark : Inner products and norms lie at the heart of linear algebra. They are very important for theoretical developments, practical applications and numerical algorithms.

Def An inner product on a vector space V is a function $\langle \cdot , \cdot \rangle : V \times V \to \mathbb{R}$ that satisfies the following properties :

1) Bilinear (linear in both arguments)

$$\langle \alpha v + \beta w , u \rangle = \alpha \langle v, u \rangle + \beta \langle w, u \rangle$$
$$\langle u , \alpha v + \beta w \rangle = \alpha \langle u, v \rangle + \beta \langle u, w \rangle$$

These properties are "extracted" from the properties of the dot product on $\mathbb{R}^n$.

2) Symmetric

$$\langle v, w \rangle = \langle w, v \rangle$$

3) Positive-definite

$$\langle v, v \rangle \geq 0$$
$$\langle v, v \rangle = 0 \iff v = 0$$

Def A vector space equipped with an inner product is called an inner product space.

Note : V can have many different inner products. E.g. on $\mathbb{R}^n$:
$$\langle v, w \rangle = \sum_{i=1}^{n} \alpha_i v_i w_i$$
where $\alpha_1, \dots \alpha_n \in \mathbb{R}_+$
($\alpha_i = 1 \Rightarrow$ dot product)

Def Given an inner product $\langle \cdot , \cdot \rangle$, the associated norm of $v \in V$ is $\quad \|v\| = \sqrt{\langle v, v \rangle}$

Example 1 Let $\alpha_1, \dots, \alpha_n > 0$, $V = \mathbb{R}^n$

$$\left.\begin{aligned} \langle v, w\rangle &= \sum_{i=1}^n \alpha_i v_i w_i \\ \|v\| &= \sqrt{\sum_{i=1}^n \alpha_i v_i^2} \end{aligned}\right\}$$ Weighted inner product and weighted norm on $\mathbb{R}^n$; α_i are weights

Often used in statistics and data fitting

$v = (v_1 \dots\dots\dots v_n)$ data vector
$\quad\ \ \alpha_1 \qquad\quad \alpha_n$

$\alpha_i \gg 1 \Rightarrow v_i$ is important
$\alpha_i \ll 1 \Rightarrow v_i$ is not important

We will meet this norm again when we will be discussing least square approx.

Example 2 Let $V = C[a,b]$

$$\left.\begin{aligned} \langle f, g\rangle &= \int_a^b f(x) g(x)\, dx \\ \|f\|_2 &= \sqrt{\int_a^b f(x)^2 dx} \end{aligned}\right\}$$ L^2 inner product and L^2 norm.

Why "L^2"? L stands for Lebesgue although he did not introduce it.

Because it is a special case of L^p norm

$$\|f\|_p = \left(\int_a^b |f(x)|^p dx\right)^{\frac{1}{p}}$$

For example, if $V = C[0, \pi]$, then

$$\langle \sin(x), \cos(x)\rangle = \int_0^\pi \sin(x)\cos(x)\, dx = \tfrac{1}{2}\sin^2 x \Big|_0^\pi = 0$$

$$\|1\|_2 = \left(\int_0^\pi 1\, dx\right)^{1/2} = \sqrt{\pi}.$$

Remark: The L^2 product and norm can be defined on more general (rather than continuous) functions, but for $C[a,b]$ it is easy to see that $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ are indeed well-defined. The most general spaces that admit L^2 products are Hilbert spaces $\longrightarrow$ more on this in ACM 1017

used in
function theory
Fourier analysis
Quantum mechanics.

Question Can we take take the space of all functions on $[a,b]$ instead of $C[a,b]$?

- No, $\langle\cdot,\cdot\rangle$ may not exist for non integrable functions

Can we take the space of integrable function?

- No, although $\langle\cdot,\cdot\rangle$ is well defined now, it does not define an inner product:

$$f(x) = \begin{cases} 1 & x = a \\ 0 & x \neq a \end{cases} \qquad f(x) \neq 0 \text{ but } \langle f, f\rangle = 0.$$

Let V be an inner product space (that is a vector space with $\langle\cdot,\cdot\rangle$). We can define the angle θ between two vectors $v, w \in V$ by analogy with $\mathbb{R}^n$:

$$\cos\theta = \frac{\langle v, w\rangle}{\|v\| \cdot \|w\|} \qquad (*)$$

In particular, v and w are orthogonal ($\perp$) if $\langle v, w\rangle = 0$

Hold on. In $\mathbb{R}^n$ we know that $\frac{v \cdot w}{\|v\| \cdot \|w\|} \in [-1, 1]$, but what about general inner products?

Is θ well defined by $(*)$? Yes, it is. Cauchy & Schwarz guarantee that.

Th (Cauchy-Schwarz) Let V be any vector space and $\langle\cdot,\cdot\rangle$ be any inner product on V.

Then $$\boxed{|\langle v,w\rangle| \le \|v\|\cdot\|w\|} \quad \forall v,w\in V$$

Moreover, $|\langle v,w\rangle| = \|v\|\cdot\|w\| \iff w=\alpha\cdot v$ $(w \parallel v)$

This inequality guarantees that the angle θ in (*) is well-def. But it has many much more important consequenses

Remark: Originally, it was proved for L^2 product. It was generalized to general inner products by Bunyakovsky.

Proof

The textbook gives a complitely correct, but a bit formal proof. Here is a more intuitive. [skipped the proof in class]

First, let us ged rid of the absolute value and rewrite the inequality in terms of $\langle\cdot,\cdot\rangle$ only.

$$\langle v,w\rangle^2 \le \|v\|^2\cdot\|w\|^2 = \langle v,v\rangle\cdot\langle w,w\rangle$$

Or, equivalently,

$$\langle v,v\rangle\cdot\langle w,w\rangle - \langle v,w\rangle\langle v,w\rangle \geqslant 0 \leftarrow \text{this is what we want to show.}$$

Second, let us assume $v\neq 0$ and $w\neq 0$, since otherwise the inequality (which becomes equality) is obvious.

Then we can divide both parts by $\langle w,w\rangle\neq 0$

$\langle w,w\rangle > 0$ (here we use positive-definiteness)

$$\langle v,v\rangle - \langle v,w\rangle\cdot\underbrace{\frac{\langle v,w\rangle}{\langle w,w\rangle}}_{\alpha} \geqslant 0$$

Using linearity:

$$\underline{\langle v, v-\alpha w\rangle} \geqslant 0 \leftarrow \text{we want to show this for any } v,w\neq 0\text{, where } \alpha=\frac{\langle v,w\rangle}{\langle w,w\rangle}$$

How can we possibly prove this? The only related property of inner products is $\langle u,u\rangle \geqslant 0$. Let us try to use it.

$$\cancel{0 \le \langle v-\alpha w\rangle} \quad 0 \le \langle v-\alpha w, v-\alpha w\rangle = \underline{\langle v, v-\alpha w\rangle} - \underbrace{\alpha\langle w,v\rangle}_{\frac{\langle v,w\rangle\langle w,v\rangle}{\langle w,w\rangle}} + \underbrace{\alpha^2\langle w,w\rangle}_{\frac{\langle v,w\rangle^2}{\langle w,w\rangle}}$$

"=" if and only if $v=\alpha w$

using symmetry $\langle v,w\rangle=\langle w,v\rangle$ $\longrightarrow$ $\frac{\langle v,w\rangle\langle w,v\rangle}{\langle w,w\rangle} = \frac{\langle v,w\rangle^2}{\langle w,w\rangle}$

Thus, $\langle v, v-\alpha w\rangle = \langle v-\alpha w, v-\alpha w\rangle \geqslant 0$

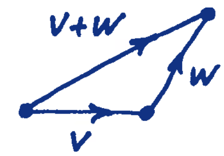
And we are done. $\square$

Remark: $|\langle v,w\rangle| \le \|v\|\cdot\|w\|$ is the most general form of the CS inequality. Writing it for specific inner products gives a variety of useful inequalities.

- Dot product : $\left|\sum_{i=1}^n x_i y_i\right| \le \sqrt{\sum_{i=1}^n x_i^2}\sqrt{\sum_{i=1}^n y_i^2}$
- L^2 inner product : $\left|\int_a^b f(x)g(x)\,dx\right| \le \left(\int_a^b f(x)^2 dx\right)^{1/2}\left(\int_a^b g(x)^2 dx\right)^{1/2}$

In $\mathbb{R}^n$, we have the triangle inequality : $\|v+w\| \leq \|v\| + \|w\|$

It turs out that this inequality holds in any inner space and it is a direct consequence of the CS inequality.



Th (Δ inequality) Let V be any vector space and $\langle \cdot, \cdot \rangle$ be any inner product on V. The associated norm $\|\cdot\|$ satisfies the triangle inequality :

$$\|v+w\| \leq \|v\| + \|w\| \qquad \forall v, w \in V$$

Proof: $\|v+w\|^2 = \langle v+w, v+w \rangle = \|v\|^2 + 2\langle v,w \rangle + \|w\|^2$

(we can drop $|\cdot|$ in CS) CS $\longrightarrow \leq \|v\|^2 + 2\|v\| \cdot \|w\| + \|w\|^2 = (\|v\| + \|w\|)^2$. $\square$

Remark As with CS, the Δ inequality looks more impressive when written for specific norms. For L^2 norm for instance :

$$\sqrt{\int_a^b (f(x)+g(x))^2 \, dx} \leq \sqrt{\int_a^b f(x)^2 \, dx} + \sqrt{\int_a^b g(x)^2 \, dx}$$

which holds for any continuos functions on $[a,b]$.

More on Norms

Norms are used to measure the length or magnitude of the elements of the underlying vector space. Every inner product induces the norm. But we can define the norm directly, without inner product. In fact, many norms used in applications don't have the associated inner products.

Def Let V be a vector space. A norm on V is a function $\|\cdot\| : V \to \mathbb{R}$ that satisfies the following properties :

1) Positive: $\|v\| \geq 0$ and $\|v\| = 0 \iff v = 0$
2) Homogeneous : $\|\alpha v\| = |\alpha| \cdot \|v\| \quad \forall \alpha \in \mathbb{R}$
3) Δ inequality : $\|v+w\| \leq \|v\| + \|w\|$.

Remark : Norms induced by inner products do satisfy these properties

Let us give several important examples of norms that do not come from $\langle \cdot, \cdot \rangle$.

① $V = \mathbb{R}^n$

- 1-norm : $\|v\|_1 = \sum_{i=1}^n |v_i|$ (also called taxicab norm)
- p-norm : $\|v\|_p = \left(\sum_{i=1}^n |v_i|^p \right)^{1/p}$, $p \geq 1$

(2-norm is the Euclidean norm) $\longleftarrow$ the only p-norm that comes from $\langle \cdot, \cdot \rangle$

- ∞-norm $\|\cdot\|_\infty = \lim_{p \to \infty} \|\cdot\|_p$

$\|v\|_\infty = \max\{|v_1|, \ldots, |v_n|\}$

② There are analogous norms on $V = C[a,b]$

- L^p norm : $\|f\|_p = \left(\int_a^b |f(x)|^p \, dx \right)^{1/p}$, $(p \geq 1)$
- L^∞ norm : $\|f\|_\infty = \max \{ |f(x)|, \ x \in [a,b] \}$

Remark L^2 norm is the only L^p norm that comes from $\langle \cdot, \cdot \rangle$

Unit Vectors, Spheres, and Balls

Let V be a normed vector space.

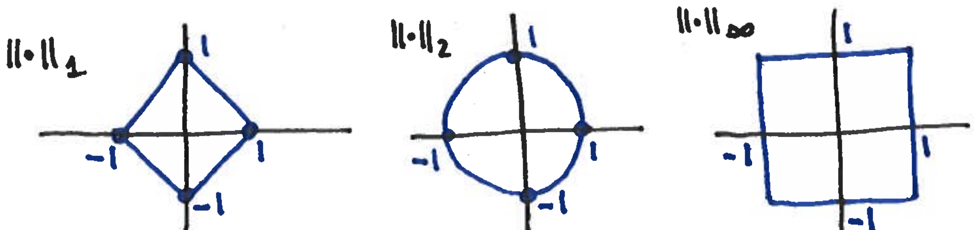
Def If $\|v\| = 1$, then $v \in V$ is called a unit vector.

Def The unit sphere in V is $S_1 = \{ v \in V : \|v\| = 1 \}$

Def The unit ball in V is $B_1 = \{ v \in V : \|v\| \leq 1 \}$

Remark If $v \in V$ and $v \neq 0$, we can always "normalize" it $\tilde{v} = \frac{v}{\|v\|} \Rightarrow \|\tilde{v}\| = 1$

Here is how unit spheres look like for 1-, 2-, and ∞-norm in $\mathbb{R}^2$



$|x| + |y| = 1$ $\quad x^2 + y^2 = 1$ $\quad \max\{|x|, |y|\} = 1$

This picture suggests that althoug the norms are different (take different values on V), they are "equivalent", meaning that are, in a sertain sence, close to each other. For instance, if $\|v\|_1$ is small, then so is $\|v\|_2$ and $\|v\|_\infty$.

The following theorem formalizes this idea.

this is important

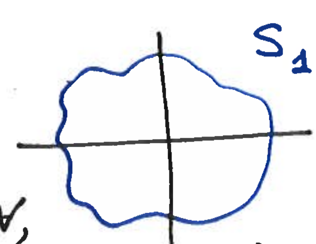
Th (Equivalence of Norms) Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be two norms on a finite-dimensional vector space V (for example, $\mathbb{R}^n$). Then $\exists\ \alpha, \beta > 0$ such that for every $v \in V$

$$\boxed{\alpha \|v\|_1 \leq \|v\|_2 \leq \beta \|v\|_1} \qquad (\#)$$

This result very import for analytical purposes: calculations may be simplified drastially, while the conslusion will be valid for any norm.

intuition why this is importang.

Sketch of Proof:



Let S_1 denote the unit sphere with respect to norm $\|\cdot\|_1$

And let $\alpha = \min \{ \|u\|_2 : u \in S_1 \}$

$\beta = \max \{ \|u\|_2 : u \in S_1 \}$

max and min exist because ~~in a finite dim V,~~ ~~S_1 is compact~~

a) $\|\cdot\|$ is a continuous function

b) in finite-dim V, S_1 is compact.

consequen. of $\triangle$ inequality.

If $v \neq 0$ (if $v = 0 \Rightarrow$ (#) is obvious)

$$\|v\|_2 = \Big\| \underbrace{\frac{v}{\|v\|_1}}_{u \in S_1} \cdot \|v\|_1 \Big\|_2 = \|v\|_1 \cdot \|u\|_2 \ \substack{\geq \alpha \cdot \|v\|_1 \\ \leq \beta \|v\|_1}$$

$\square$