Inner Product & Norms

Note Title

★ Inner Product

<u>Def.</u> The inner product between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ is defined as

$$\mathbf{x}^T \mathbf{y} = \sum_{i=1}^{m} x_i y_i \in \mathbb{R}$$

and is also written as

$$\mathbf{x} \cdot \mathbf{y}, \ (\mathbf{x}, \mathbf{y}), \text{ or } \langle \mathbf{x}, \mathbf{y} \rangle .$$

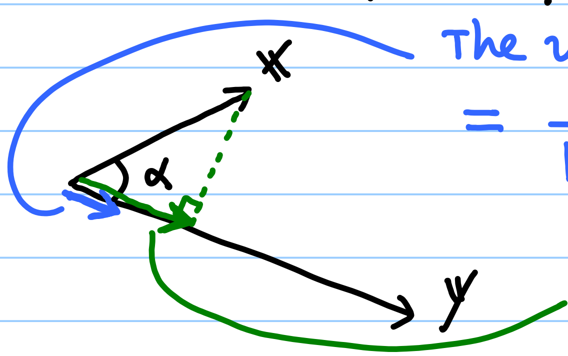
The ℓ^2-norm of $\mathbf{x} \in \mathbb{R}^m$ is defined as

$$\|\mathbf{x}\|_2 := \sqrt{\mathbf{x}^T \mathbf{x}} = \sqrt{\sum_{i=1}^{m} |x_i|^2},$$

which is the Euclidean length of $\mathbf{x}$. This is often written as $\|\mathbf{x}\|$.

The angle α between $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$, can be computed by

$$\cos \alpha = \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\| \, \|\mathbf{y}\|}$$



The unit vector along $\mathbf{y}$

$$= \frac{\mathbf{y}}{\|\mathbf{y}\|}$$

Hence, the projection of $\mathbf{x}$ onto $\mathbf{y}$ is

$$\operatorname{proj}_{\mathbf{y}} \mathbf{x} = \left(\|\mathbf{x}\| \cos\alpha\right) \frac{\mathbf{y}}{\|\mathbf{y}\|}$$

$$= \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{y}\|^2} \mathbf{y}$$

★ Vector Norms

→ to quanify (or measure) the size (or length) of a vector

Def. A norm is a function
$\|\cdot\| : \mathbb{R}^m \to \mathbb{R}$ s.t.
$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^m, \ \forall \alpha \in \mathbb{R}$

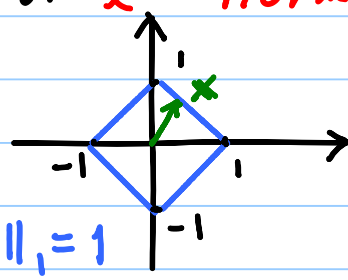
(1) $\|\mathbf{x}\| \geq 0$ and $\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0}$

(2) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ The triangle inequality

(3) $\|\alpha \mathbf{x}\| = |\alpha|\, \|\mathbf{x}\|$

Examples p-norms (or ℓ^p-norms)

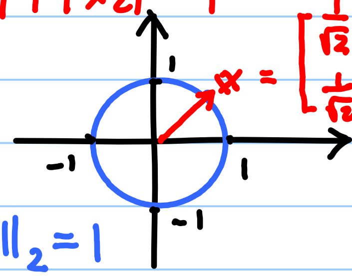
$$\|\mathbf{x}\|_1 := \sum_{i=1}^{m} |x_i|$$



$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $\|\mathbf{x}\|_1 = 1$

$|x_1| + |x_2| = 1$

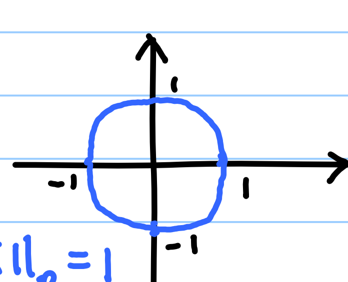
$$\|\mathbf{x}\|_2 := \left(\sum_{i=1}^{m} |x_i|^2 \right)^{1/2}$$



$\mathbf{x} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

$\|\mathbf{x}\|_2 = 1$

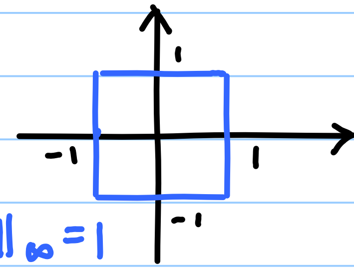
$$\|\mathbf{x}\|_p := \left(\sum_{i=1}^{m} |x_i|^p \right)^{1/p}$$



$\|\mathbf{x}\|_p = 1$

$p > 2$

$$\| \mathbf{x} \|_\infty := \max_{1 \le i \le m} |x_i|$$



$\| \mathbf{x} \|_\infty = 1$

<u>Exercise</u>: what is the vector $\mathbf{x} \in \mathbb{R}^2$ that achieves $\max \| \mathbf{x} \|_1$ subject to $\| \mathbf{x} \|_2 = 1$?

☆ <u>Matrix Norms</u>

- One can view an $m \times n$ matrix as a vector of length mn, then use one of the vector norms.

<u>Def.</u> The Frobenius (Hilbert-Schmidt) norm of $A \in \mathbb{R}^{m \times n}$ is defined as

$$\| A \|_F := \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}$$

$$= \left(\sum_{j=1}^{n} \| a_j \|_2^2 \right)^{1/2}$$

$$= \sqrt{\operatorname{tr}(A^T A)}$$

$$= \sqrt{\operatorname{tr}(A A^T)}$$

<u>Def.</u> For $X \in \mathbb{R}^{m \times n}$, $\operatorname{tr}(X) := \sum_{i=1}^{\min(m,n)} x_{ii}$ is called the trace of X.

- However, $\exists$ different types of matrix norms called **induced matrix norms** (often called **operator norms**), which are defined in terms of the behavior of a matrix as an <u>operator</u> between its normed domain and range space.

<u>Def.</u> Let $A \in \mathbb{R}^{m \times n}$. Then the **induced matrix** (or **operator**) **norm** is defined as

$$\|A\|_p := \sup_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \mathbf{x} \neq \mathbf{0}}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$$

$$= \sup_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \|\mathbf{x}\|_p = 1}} \|A\mathbf{x}\|_p$$

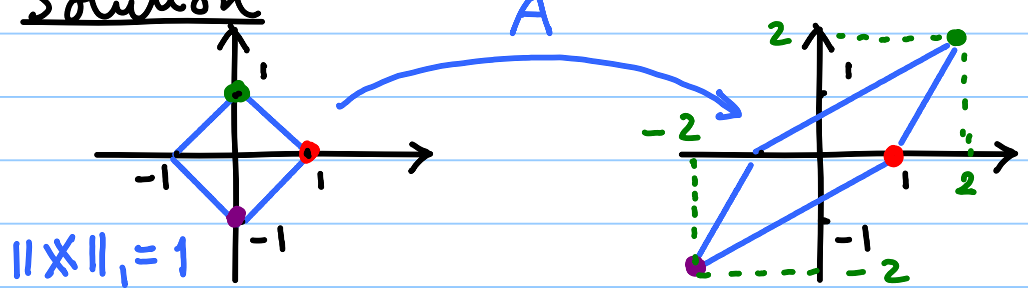
In other words, $\|A\|_p$ is the smallest constant C satisfying

$$\|A\mathbf{x}\|_p \leq C\,\|\mathbf{x}\|_p \quad \forall\, \mathbf{x} \in \mathbb{R}^n.$$

<u>Example</u> Consider $A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} : \mathbb{R}^2 \to \mathbb{R}^2$.

Compute $\|A\|_1$, $\|A\|_2$, $\|A\|_\infty$.

<u>Solution</u>



Hence, $\sup \|A\mathbf{x}\|_1 = \max \|A\mathbf{x}\|_1$
$= |2|+|2| = |-2|+|-2| = 4$

achieved for $\mathbf{x} = [0,1]^T, [0,-1]^T$.

In fact,

$$\begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \rightarrow \left\| \begin{bmatrix} 2 \\ 2 \end{bmatrix} \right\|_1 = 2+2=4$$

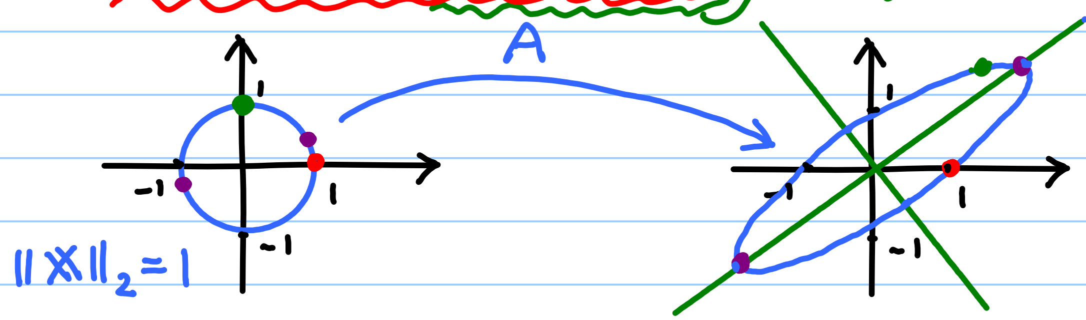
$$\begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix} \rightarrow \left\| \begin{bmatrix} -2 \\ -2 \end{bmatrix} \right\|_1 = |-2|+|-2| = 4.$$

How about $\|A\|_2$?

$\Rightarrow$ as I'll prove later,

$$\|A\|_2 = \sqrt{\lambda_{max}(A^TA)}$$

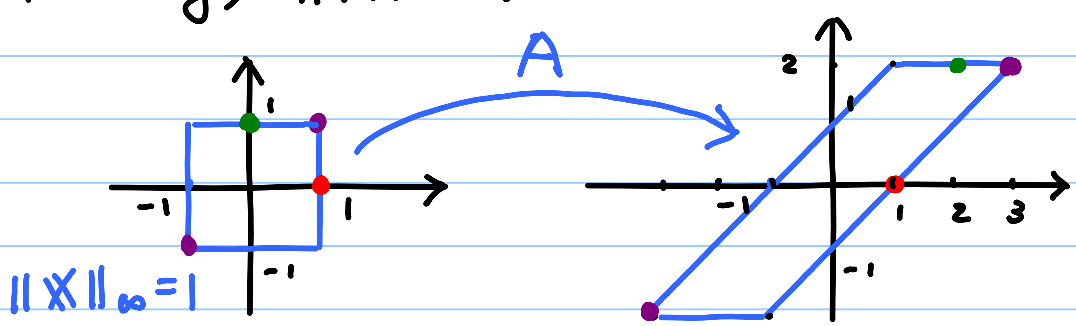
The largest eigenvalue of A^TA.



In this case $\|A\|_2 \approx 2.9208$

= the length of the major semi axis of the ellipsis.

Finally, $\|A\|_\infty$.



From this figure, we can see
$\|A\|_\infty = 3$.

In fact, $\begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x+2y \\ y \end{bmatrix}$

So, $$\|A\|_\infty = \max_{\substack{|x|\le 1 \\ |y|\le 1}} \left(|x+2y|, |2y| \right)$$
$$= \max_{\substack{|x|\le 1 \\ |y|\le 1}} |x+2y|$$
$$= 3 \quad \text{at } \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ or } \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$
//

- The p-norm of a diagonal matrix

Say $$D = \begin{bmatrix} d_1 & & & O \\ & d_2 & & \\ & & \ddots & \\ O & & & d_m \end{bmatrix}$$

Then, D maps the unit sphere in $\mathbb{R}^m$ (denoted by S^{m-1}) to a hyperellipsoid whose semiaxes are $|d_1|, \cdots, |d_m|$.

So, $$\|D\|_2 = \max_{1\le i\le m} |d_i|$$

In fact, $$\|D\|_p = \max_{1\le i\le m} |d_i|$$
for $\forall p \ge 1$. //

- <u>The 1-norm of a matrix</u>

$A \in \mathbb{R}^{m\times n}$

$$\|A\|_1 = \max_{1\le j\le n} \|a_j\|_1$$

i.e., max. of 1-norms of col. vec's.

(Proof) Suppose $\mathbf{x} \in \mathbb{R}^n$

Then
$$\|A\mathbf{x}\|_1 = \Big\|\sum_{j=1}^{n} x_j a_j\Big\|_1$$
$$\le \sum_{j=1}^{n} |x_j|\, \|a_j\|_1$$

via the triangle ineq.

$$\le \max_{1\le j\le n} \|a_j\|_1 \cdot \sum_{j=1}^{n} |x_j|$$
$$= \max_{1\le j\le n} \|a_j\|_1 \cdot \|\mathbf{x}\|_1$$

So
$$\frac{\|A\mathbf{x}\|_1}{\|\mathbf{x}\|_1} \le \max_{1\le j\le n} \|a_j\|_1$$

?

Now can this <u>bound</u> be attained at some $\mathbf{x}$? $\Rightarrow$ Yes!

Let $\|a_k\|_1 = \max_{1\le j\le n} \|a_j\|_1$

Then set $\mathbf{x} = e_k$

$$\Rightarrow \frac{\|Ae_k\|_1}{\|e_k\|_1} = \frac{\|a_k\|_1}{1} = \|a_k\|_1$$

//

- **The 2-norm of a matrix**

$A \in \mathbb{R}^{m \times n}$

$$\|A\|_2 = \sqrt{\lambda_{max}(A^T A)}$$

where $\lambda_{max}(A^T A)$ is the largest (positive) eigenvalue of $A^T A$.

(Proof) Note the def. of $\|A\|_2$, i.e.,

$$\|A\|_2 = \sup_{\|\mathbf{x}\|_2 = 1} \|A\mathbf{x}\|_2$$

Consider functions:

$$f(\mathbf{x}) := \|A\mathbf{x}\|_2^2 = (A\mathbf{x})^T(A\mathbf{x})$$
$$= \mathbf{x}^T A^T A \mathbf{x}.$$
$$g(\mathbf{x}) := \|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}$$

Then consider the following problem.

(*) Maximize $f(\mathbf{x})$ subject to $g(\mathbf{x}) = \underline{1}$.

$\Rightarrow$ This can be solved by the method of Lagrange multipliers (MAT 21C)

In other words, define

$$h(\mathbf{x}, \lambda) := f(\mathbf{x}) - \lambda(g(\mathbf{x}) - 1)$$

The solution to (*) $\Longleftrightarrow$ $\dfrac{\partial h}{\partial x_i} = 0, \ 1 \le i \le n$

with $g(\mathbf{x}) = 1$

Can show that $\frac{\partial h}{\partial x_i} = 0 \quad 1 \le i \le n$

leads to $\frac{\partial h}{\partial \mathbf{x}} = \mathbf{0}$

i.e., $2 A^T A \mathbf{x} - 2 \lambda \mathbf{x} = \mathbf{0}$

$$A^T A \mathbf{x} = \lambda \mathbf{x} \Rightarrow$$

$\mathbf{x}$: eigenvector

λ : eigenvalue of $A^T A$

Now $g(\mathbf{x}) = \mathbf{x}^T \mathbf{x} = 1$ So

$$\underbrace{\mathbf{x}^T A^T A \mathbf{x}}_{\ge 0} = \lambda \underbrace{\mathbf{x}^T \mathbf{x}}_{=1} = \lambda$$

so this is also ≥ 0

Finally,

$$\|A\|_2 = \sup_{\|\mathbf{x}\|_2 = 1} \|A\mathbf{x}\|_2$$

$$= \left(\sup_{\mathbf{x}^T\mathbf{x} = 1} \mathbf{x}^T A^T A \mathbf{x} \right)^{\frac{1}{2}}$$

$$= \sqrt{\lambda_{max}(A^T A)} \quad //$$

- The ∞-norm of a matrix

$A \in \mathbb{R}^{m \times n}$

$$\|A\|_\infty = \max_{1 \le i \le m} \| a_{i\cdot} \|_1$$

$a_{i\cdot}$: ith row vector of A

Note : Let $\mathbf{x} \in \mathbb{R}^k = \mathbb{R}^{k \times 1}$

Then $\mathbf{x}^T \in \mathbb{R}^{1 \times k}$ = a row vector with k entries

$$\|\mathbf{x}^T\|_1 = \|\mathbf{x}\|_1 = \sum_{j=1}^{k} |x_j|$$

also, note $A = \begin{bmatrix} a_{1\cdot} \\ \vdots \\ a_{m\cdot} \end{bmatrix}$

(Proof) $\|A\mathbf{x}\|_\infty \overset{\text{definition}}{=} \max_{1\le i\le m} |a_{i\cdot}\mathbf{x}|$

$$= \max_{1\le i\le m} \left| \sum_{j=1}^{n} a_{ij} x_j \right|$$

$$\le \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|\,|x_j|$$

$$\le \|\mathbf{x}\|_\infty \cdot \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|$$

So, $\dfrac{\|A\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} \le \max_{1\le i\le m} \|a_{i\cdot}\|_1$

Suppose $\|\mathbf{x}\|_\infty = 1$. then for which $\mathbf{x}$, the equality

$$\|A\mathbf{x}\|_\infty = \max_{1\le i\le m} \|a_{i\cdot}\|_1$$

is attained ?

$\Rightarrow$ Let $\|a_{k\cdot}\|_1 = \max_{1\le i\le m} \|a_{i\cdot}\|_1$

Then define $\mathbf{x}$ as

$$x_j = \begin{cases} 1 & \text{if } a_{kj} \ge 0 \\ -1 & \text{if } a_{kj} < 0. \end{cases}$$

Clearly $\|\mathbf{x}\|_\infty = 1$ and

$$|a_{i\cdot}\mathbf{x}| = \left| \sum_{j=1}^{n} a_{ij} x_j \right|$$

$$\le \sum_{j=1}^{n} |a_{ij}|\, \underbrace{|x_j|}_{=1}$$

$$= \sum_{j=1}^{n} |a_{ij}|$$

$$= \| a_{i\cdot} \|_1 \qquad 1 \le i \le m$$

But if $i = k$, this becomes an equality, and the max. is achieved!

$$\| A \mathbb{x} \|_\infty = \max_{1 \le i \le m} | a_{i\cdot} \mathbb{x} |$$

$$= | a_{k\cdot} \mathbb{x} |$$

$$= \| a_{k\cdot} \|_1$$

$$= \max_{1 \le i \le m} \| a_{i\cdot} \|_1$$

//