* Matrix norms:

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

. Def: The Frobenius (Hilbert-~~Sach~~ Schmidt) norm of $A \in \mathbb{R}^{m\times n}$ is defined as

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

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

$$= \sqrt{tr(A^TA)}$$

$$= \sqrt{tr(AA^T)}$$

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

However, there exist 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 operator between its normed domain and range space.

. Def: Let $A \in \mathbb{R}^{m\times n}$. Then the ~~operator~~ induced matrix norm is defined as

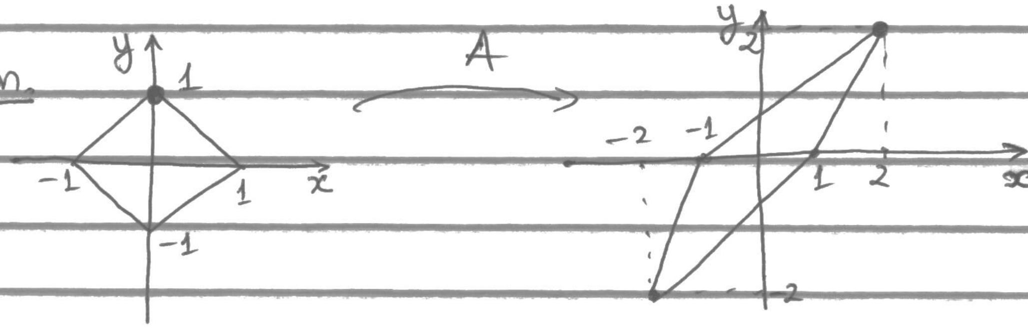
$$\|A\|_p := \sup_{\vec{x}\neq 0} \frac{\|A\vec{x}\|_p}{\|\vec{x}\|_p} = \sup_{\|\vec{x}\|_p=1} \|A\vec{x}\|_p.$$

In other words, $\|A\|_p$ is the smallest constant C satisfying $\|Ax\|_p \le C\|x\|_p \quad \forall x \in \mathbb{R}^n$.

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

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

Solution.



Hence, $\|A\|_1 = \sup_{\|x\|_1 = 1} \|Ax\|_1 = |2| + |2| = 4$

$= |-2| + |-2|$

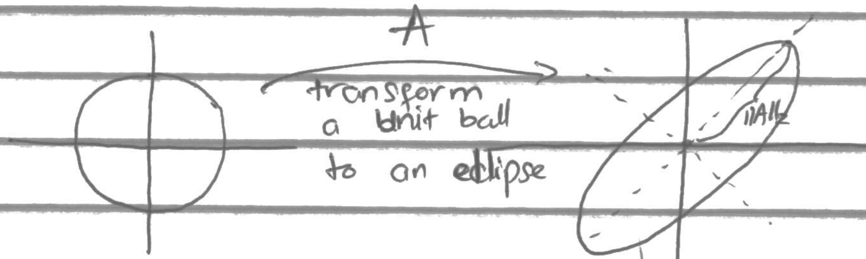
~~In gene~~ achived for $\vec{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ -1 \end{bmatrix}$.

How about $\|A\|_2$?

We will show later that

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

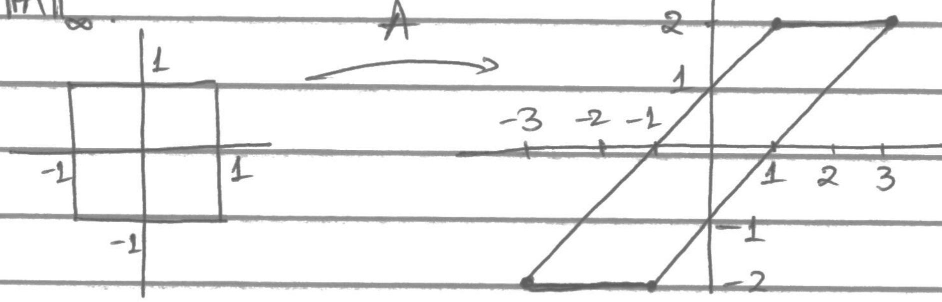
→ the largest eigenvalue of $A^T A$



In this example, $\|A\|_2 \approx 2.9208$

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

$\|A\|_\infty$.



$\rightarrow \quad \|A\|_\infty = 3$ achieved at $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ or $\begin{bmatrix} -1 \\ -1 \end{bmatrix}$.

* the p-norm of a diagonal matrix.

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

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

So, $\|D\|_2 = \max\limits_{1 \le i \le n} |d_i|$.

In general, $\|D\|_p = \max\limits_{1 \le i \le n} |d_i|$. $\forall p \ge 1$.

* The 1-norm of a matrix.

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

$$\|A\|_1 = \max_{1 \le j \le n} \|\vec{a}_j\|_1$$

i.e., max of 1-norm of column vectors.

Pf: For $\vec{x} \neq 0 \in \mathbb{R}^{m}$,

$$\|A\vec{x}\|_1 = \Big\|\sum_{i=1}^{n} x_i \vec{a}_i\Big\|_1$$

$$\overset{\text{triangle ineq}}{\le} \sum_{i=1}^{m} |x_i| \, \|\vec{a}_i\|_1$$

$$\le \Big(\sum_{i=1}^{n} |x_i|\Big) \max_{1\le i\le n} \|\vec{a}_i\|_1$$

$$\cancel{\le}\ \cancel{\|} = \max_{1\le i \le n} \|\vec{a}_i\|_1 \cdot \|\vec{x}\|_1$$

$$\Rightarrow \frac{\|A\vec{x}\|_1}{\|\vec{x}\|_1} \le \max_{1\le i\le n} \|\vec{a}_i\|_1$$

Now can this bound be attained at some $\vec{x}$?

- Ans: Yes!

~~Let~~ ~~$\vec{e}_k$~~ ~~be~~ Take the (k)th column whose has largest 1-norm: ~~let~~ $\|\vec{a}_k\|_1 = \max_{1\le j\le n} \|\vec{a}_j\|_1$

And set $\vec{x} = \vec{e}_k$ $\rightsquigarrow$ the kth vector in the standard basis.

$$\Rightarrow \frac{\|A\vec{e}_k\|_1}{\|\vec{e}_k\|_1} = \frac{\|\vec{a}_k\|_1}{1} = \|\vec{a}_k\|_1$$

* The 2-norm of a matrix

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

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

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

Pf: Consider functions:

$$f(\vec{x}) = \|Ax\|_2^2 = (A\vec{x})^T (A\vec{x}) = \vec{x}^T A^T A \vec{x}.$$

and $$g(\vec{x}) = \|x\|_2^2 = \vec{x}^T \vec{x}.$$

Then consider the following problem:

~~Max f(x)~~

(*) maximize $f(\vec{x})$ subject to $g(\vec{x}) = 1$.

We can use the method of Lagrange multipliers to solve this problem.

In other words, define

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

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

with $g(\vec{x}) = 1$.

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

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

$$\rightarrow \quad 2A^T A\vec{x} - 2\lambda\vec{x} = 0.$$

$$A^T A \vec{x} = \lambda \vec{x}$$

$\vec{x}$ = eigenvector and λ = eigenvalue.

of $A^T A$.

Since $g(\vec{x}) = \vec{x}^T \vec{x} = 1$,

$$\underbrace{\vec{x}^T A^T A \vec{x}}_{\ge 0} = \vec{x}^T (\lambda \vec{x}) = \lambda \vec{x}^T \vec{x} = \underset{\ge 0}{\lambda}$$

Finally, $$\|A\|_2 = \sup_{\|x\|_2 = 1} \|Ax\|_2$$

$$= \left(\sup_{\vec{x}^T \vec{x} = 1} \vec{x}^T A^T A \vec{x} \right)^{1/2} = \sqrt{\lambda_{max}(A^T A)}.$$

* The ∞-norm of a matrix:

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

$\|A\|_\infty = \max_{1\le i\le m} \|\vec{a}_{i\cdot}\|_1$ → the ith row vector of A.

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

Pf: by definition

row vector, column vector

$$\cancel{\|A\|_\infty}\ \|A\vec{x}\|_\infty = \max_{1\le i\le m} |\vec{a}_{i\cdot}\ \vec{x}|$$

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

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

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

$$\Rightarrow \frac{\|Ax\|_\infty}{\|x\|_\infty} \le \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}| = \max_{1\le i\le m} \|a_{i\cdot}\|_1$$

Suppose $\|\vec{x}\|_\infty = 1$, then for which $\vec{x}$ the equality $\|Ax\|_\infty = \max_{1\le i\le m} \|a_{i\cdot}\|_1$ is attained?

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

Then define a vector $\vec{x}$ as

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

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

$|\vec{a}_{k\cdot}\vec{x}| = \|\vec{a}_{k\cdot}\|_1$.

* Why matrix norm is important?

~~A canno~~

A computer cannot represent real numbers exactly unless they are digital numbers (e.g. 0 and 1).

$\Rightarrow$ Suppose we want to solve

$$A\vec{x} = \vec{b}$$

where $A \in \mathbb{R}^{m\times m}$, $\vec{x} \in \mathbb{R}^m$ and $\vec{b} \in \mathbb{R}^m$.

$\Rightarrow$ we have to first encode A, $\vec{x}$, and $\vec{b}$. on the computer.

$$A \longrightarrow A'$$
$$\vec{x} \longrightarrow \vec{x}'$$
$$\vec{b} \longrightarrow \vec{b}'$$

i.e. we solve $A'\vec{x}' = \vec{b}'$.

For simplicity, suppose $\vec{b}' = \vec{b}$ and A is invertible.

$\Rightarrow$ relative error of the solution:

$$\frac{\|\vec{x}' - \vec{x}\|}{\|\vec{x}'\|} = \frac{\|\vec{x}' - A^{-1}\vec{b}\|}{\|\vec{x}'\|}$$

$$= \frac{\|\vec{x}' - A^{-1}A'\vec{x}'\|}{\|\vec{x}'\|}$$

$$= \frac{\|A^{-1}(A - A')\vec{x}'\|}{\|\vec{x}'\|}.$$

$$\le \frac{\|A^{-1}(A - A')\|\|\vec{x}'\|}{\|\vec{x}'\|}$$

$$\le \underbrace{\|A\|\|A^{-1}\|}_{\text{condition number}} \underbrace{\frac{\|A - A'\|}{\|A\|}}_{\text{relative error in matrix}}.$$

condition number: $\kappa(A) = \|A\| \, \|A^{-1}\|$

. If $\kappa(A)$ is large, then A is "bad", i.e., there is a large error in solution $\vec{x}^* = A^{-1}\vec{b}$.

. If A singular, $\kappa(A) = +\infty$.

* Orthogonal Vectors:

Def: . Two vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ are said to be orthogonal if $\vec{x} \cdot \vec{y} = 0$

(The zero vector is orthogonal to any vector)

. Two sets of vectors X, Y are said to be orthogonal if

$$\forall \vec{x} \in X \text{ and } \forall \vec{y} \in Y, \quad \vec{x} \cdot \vec{y} = 0$$

. A set of vectors S is said to be orthogonal if $\forall \vec{x} \in S, \forall \vec{y} \in S, \vec{x} \neq \vec{y}$,

$$\vec{x} \cdot \vec{y} = 0.$$

. A ~~vec~~ set of vectors S is said to be orthonormal if S is orthogonal and $\forall \vec{x} \in S$,

$$\|x\|_2 = 1.$$

(orthonormal = othogonal + normalized)

Thm: The vectors in an orthogonal set S are linearly independent.

Pf: let $S = \{v_1, \dots, v_n\}$.

Suppose they are not lin. indep.

Then $\exists \vec{v}_k \in S$ such that $\vec{v}_k \neq 0$ and

$$\vec{v}_k = \sum_{\substack{i=1 \\ i\neq k}}^{n} c_i \vec{v}_i \quad \text{with} \quad \vec{c} \neq 0.$$

$$\vec{c} = [c_1, \dots, c_{k-1}, c_{k+1}, \dots, c_n]^T.$$

Since S is an orthogonal set,

$$\vec{v}_j \cdot \vec{v}_i = 0 \quad \text{for} \quad j \neq i.$$

$$\Rightarrow \vec{v}_k \cdot \vec{v}_k = \vec{v}_k \cdot \Big(\sum_{\substack{i=1 \\ i\neq k}}^{n} c_i \vec{v}_i \Big).$$

$$= \sum_{\substack{i=1 \\ i\neq k}}^{n} c_i \underbrace{\vec{v}_k \cdot \vec{v}_i}_{=0}.$$

$$= 0$$

$$\Rightarrow \|\vec{v}_k\|_2^2 = 0.$$

$$\Rightarrow \vec{v}_k = 0.$$

contradiction!

* <u>Component of a vector.</u>

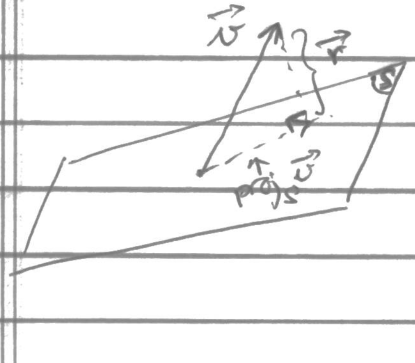
"Inner products can be used to decompose arbitrary vectors into orthogonal components!"

Suppose $S = \{q_1, \dots, q_n\} \subset \mathbb{R}^m$ is an orthonormal set.

Let $\vec{v}$ be an arbitrary vector in $\mathbb{R}^m$.

Let $\vec{r} = \vec{v} - \langle \vec{q}_1, \vec{v} \rangle \vec{q}_1 - \langle \vec{q}_2, \vec{v} \rangle \vec{q}_2 - \dots - \langle \vec{q}_n, \vec{v} \rangle \vec{q}_n$

$\overrightarrow{\text{residual vector}}$ is $\perp$ to $\{\vec{q}_1, \dots, \vec{q}_n\}$.



Why?

$$\langle \vec{q}_j, \vec{r} \rangle = \langle \vec{q}_j, \vec{v} \rangle - \langle \vec{q}_1, \vec{v} \rangle \overset{0}{\cancel{\langle \vec{q}_j, \vec{q}_1 \rangle}}$$

$$- \dots - \langle \vec{q}_n, \vec{v} \rangle \overset{=0}{\cancel{\langle \vec{q}_j, \vec{q}_n \rangle}}$$

$$= \langle \vec{q}_j, \vec{v} \rangle - \langle \vec{q}_j, \vec{v} \rangle \langle \vec{q}_j, \vec{q}_j \rangle.$$

$$= 0.$$

It's true for any $j = 1, \dots, n$.

$$\Rightarrow \vec{v} = \vec{r} + \underbrace{\sum_{i=1}^{n} \langle \vec{q}_i, \vec{v} \rangle \vec{q}_i}$$

$$= \vec{r} + \mathrm{proj}_S \vec{v}$$

$$= \vec{r} + QQ^T \vec{v}$$

where $Q = [\vec{q}_1 \ \vec{q}_2 \ \dots \ \vec{q}_n] \in \mathbb{R}^{m \times n}$.

If $\{\vec{q}_1, \dots, \vec{q}_n\}$ is a basis of $\mathbb{R}^m$, then $n = m$ and $\vec{r} = 0$, i.e., $\vec{v} = \sum_{i=1}^{m} \langle \vec{q}_i, \vec{v} \rangle \vec{q}_i = \sum_{i=1}^{m} (\vec{q}_i \vec{q}_i^T) \vec{v}$

$$\Rightarrow \vec{v} = QQ^T \vec{v}$$

and $QQ^T = I$.

<u>Def:</u> A square matrix $Q \in \mathbb{R}^{m \times m}$ is said to be <u>orthogonal</u> if

$$Q^T = Q^{-1}$$

i.e., $Q^T Q = QQ^T = I$.

<u>E.g.</u> $Q = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} \\ 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & -1/\sqrt{2} \end{bmatrix}$ then $Q^T Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

but

$$QQ^T = \begin{bmatrix} 5/6 & 1/3 & -1/6 \\ 1/3 & 1/3 & 1/3 \\ -1/6 & 1/3 & 5/6 \end{bmatrix}$$

<u>Remark:</u> If $Q = [\vec{q}_1 \ \vec{q}_2 \ \dots \ \vec{q}_n] \in \mathbb{R}^{m \times n}$ with $m > n$ and these vectors are orthonormal, then it is always true that $Q^T Q = I_{n \times n}$ but $QQ^T \neq I_{m \times m}$ unless $m = n$.