Vectors and Matrices Review

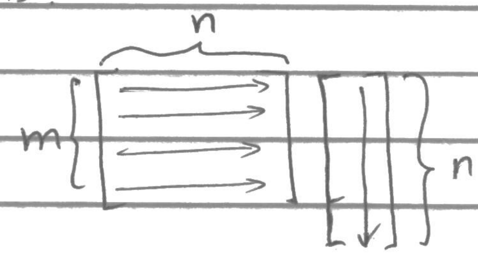
* <u>Matrix - Vector Multiplication:</u>

$\vec{x} \in \mathbb{R}^n$

$A = [a_{ij}] \in \mathbb{R}^{m \times n}$

m rows, n columns.

$\vec{y} = A\vec{x} \in \mathbb{R}^m$.



Very important to understand that $\vec{y}$ is <u>a linear combination of the column vectors</u> of A.

$$A = \begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \\ | & | & & | \end{bmatrix} \quad \text{and} \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$\vec{a}_j$ is the jth column of A

$$\begin{aligned} \vec{y} &= A\vec{x} \\ &= x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_n\vec{a}_n . \end{aligned}$$

<u>Thm:</u> Let $F_A : \mathbb{R}^n \to \mathbb{R}^m$ be a map defined as

$$F_A(\vec{x}) = A\vec{x}.$$

Then, F_A is a <u>linear map</u>

i.e., $\forall \vec{u}, \vec{v} \in \mathbb{R}^n$ and $\forall \alpha \in \mathbb{R}$.

$$\begin{cases} F_A(\vec{u} + \vec{v}) = F_A(\vec{u}) + F_A(\vec{v}) \\ F_A(\alpha \vec{u}) = \alpha F_A(\vec{u}). \end{cases}$$

<u>Conversely</u>, for <u>any</u> linear map $F: \mathbb{R}^n \to \mathbb{R}^m$, there exists a <u>unique</u> matrix $A \in \mathbb{R}^{m \times n}$ such that $F = F_A$.

Proof:

($\Rightarrow$) Show that F_A is linear.

For any $\vec{u}, \vec{v} \in \mathbb{R}^m$ and $\alpha \in \mathbb{R}$

- $F_A(\vec{u}+\vec{v}) = A(\vec{u}+\vec{v}) = A\vec{u} + A\vec{v} = F_A(\vec{u}) + F_A(\vec{v})$
- $F_A(\alpha\vec{u}) = A(\alpha\vec{u}) = \alpha A\vec{u} = \alpha F_A(\vec{u})$.

($\Leftarrow$) Let F be a linear map.

Let $\{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$ be the standard basis of $\mathbb{R}^n$, i.e., $\vec{e}_j = \begin{array}{c} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \\ \leftarrow j \\ \\ \\ \\ \end{array}$

Set $F(\vec{e}_j) = \vec{a}_j \in \mathbb{R}^m$.

Let $A = [\vec{a}_1 \ \vec{a}_2 \dots \vec{a}_m] \in \mathbb{R}^{m\times n}$

Now pick any $\vec{x} \in \mathbb{R}^n$, we can always write

$$\vec{x} = x_1\vec{e}_1 + \dots + x_n\vec{e}_n.$$

$$\Rightarrow \quad F(\vec{x}) = x_1 F(\vec{e}_1) + \dots + x_n F(\vec{e}_n)$$

$$= x_1\vec{a}_1 + \dots + x_n\vec{a}_n.$$

$$\Rightarrow \quad F(\vec{x}) = A\vec{x} = F_A(\vec{x}).$$

Uniqueness, let $A, B \in \mathbb{R}^{m\times n}$.

~~$F_A(\vec{e}_j) = \vec{a}_j$~~

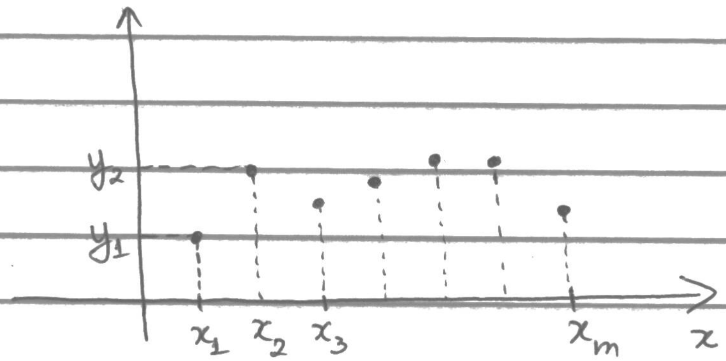
If $F_A = F_B$, Then

$F_A(\vec{e}_j) = \vec{a}_j = F_B(\vec{e}_j) = \vec{b}_j$. for $1 \le j \le n$

$\Rightarrow \quad A = B$.

Example: A Vandermonde matrix

Let $\{x_1, \dots, x_m\}$ be a set of sample points.



(Assume $x_i \neq x_j$ if $i \neq j$)

Consider a space of polynomials of degree at most $n-1$:

$$\mathcal{P}_{n-1}[x] := \{ p(x) = c_0 + c_1 x + \cdots + c_{n-1} x^{n-1}, \quad c_j \in \mathbb{R},\ j = 0, 1, \dots, n-1 \}.$$

It's clear that $\mathcal{P}_{n-1}[x]$ is a linear (vector) space since $\forall\ p, q \in \mathcal{P}_{n-1}[x]$,

$$p + q \in \mathcal{P}_{n-1}[x]$$

and $\alpha p \in \mathcal{P}_{n-1}[x] \quad \forall\, \alpha \in \mathbb{R}$.

Hence, a map from a coefficient vector

$$\vec{c} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix} \in \mathbb{R}^n$$

to vectors of sampled polynomial values

$$\vec{y} = \begin{bmatrix} p(x_1) \\ p(x_2) \\ \vdots \\ p(x_m) \end{bmatrix} \in \mathbb{R}^m$$

is <u>linear</u>!

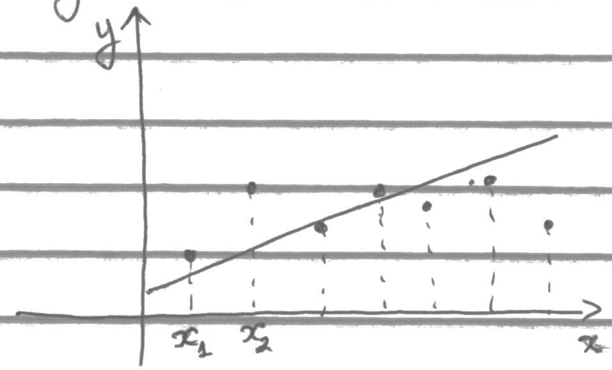
So, $\exists\, A \in \mathbb{R}^{m \times n}$ for such linear map F.
what is this matrix A?

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^{n-1} \end{bmatrix} \in \mathbb{R}^{m \times n}.$$

called the $m \times n$ Vandermonde matrix.

$$\vec{y} = A\vec{c} \iff \begin{bmatrix} p(x_1) \\ p(x_2) \\ \vdots \\ p(x_m) \end{bmatrix} = \begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_m & \cdots & x_m^{n-1} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix}$$

This matrix is often used in the least squares polynomial fitting to a set of measurements or noisy data.



In the case of a line fitting, $(n = 2)$.

But you may have many points, i.e., m large.

Then you might want to find a line s.t.
such that the size of $\underbrace{\vec{y} - A\vec{c}}_{\text{residual error}}$ is small.

In the case of line fitting, $\vec{c} = \begin{bmatrix} c_0 \\ c_1 \end{bmatrix}$.

* Matrix - Matrix Multiplication:

$C = AB$

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

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

Note that

$$[\vec{c}_1 \dots \vec{c}_m] = [\vec{a}_1 \dots \vec{a}_k][\vec{b}_1 \dots \vec{b}_n].$$

$\Rightarrow \vec{c}_j = A\vec{b}_j$. for $1 \le j \le n$.

$\Rightarrow$ each $\vec{c}_j$ is a linear combination of column vectors of A with the coefficients vector b_j.

Eg. 1) Outer Product.

Let $\vec{u} \in \mathbb{R}^m = \mathbb{R}^{m\times 1}$.

$\vec{v} \in \mathbb{R}^n = \mathbb{R}^{n\times 1}$

Then the outer product between $\vec{u}$ and $\vec{v}$ is

$$\vec{u}\,\vec{v}^T = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix} [v_1 \dots v_n] = \begin{bmatrix} u_1 v_1 & \dots & u_1 v_n \\ u_2 v_1 & \dots & u_2 v_n \\ \vdots & & \vdots \\ u_m v_1 & \dots & u_m v_n \end{bmatrix} \in \mathbb{R}^{m\times n}$$

This matrix has rank 1 because $\vec{u}\vec{v}^T = [v_1\vec{u} \dots v_n\vec{u}]$

i.e., each column is just a scalar multiple of the same vector $\vec{u}$.

* <u>Range and Nullspace</u> (or Kernel)

<u>Def:</u> Let A be an $m \times n$ matrix.

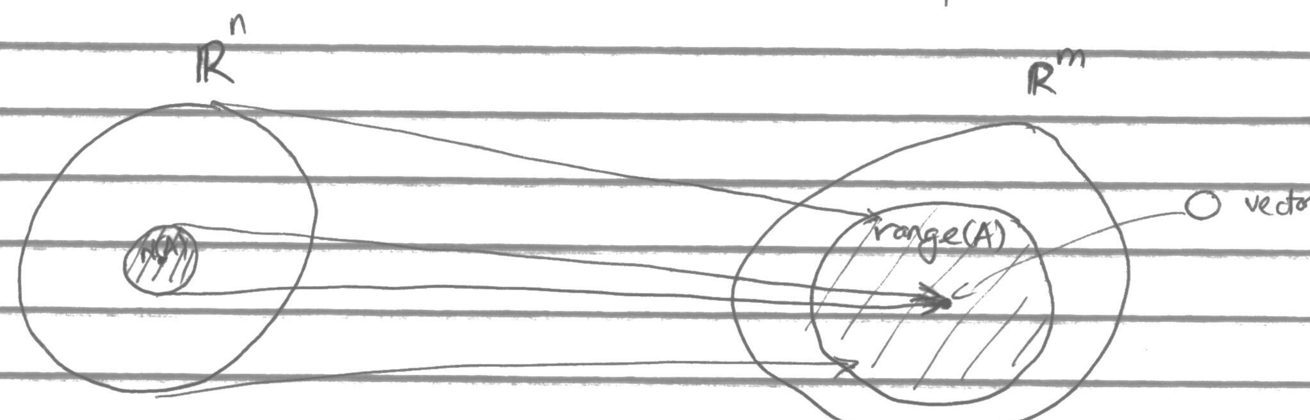
- $\text{range}(A) := \{\vec{y} \in \mathbb{R}^m \mid \vec{y} = A\vec{x},\ \vec{x} \in \mathbb{R}^n\}$.

often written as $\text{Ran}(A)$ or $\underset{\text{image}}{\underline{\text{Im}}(A)}$ or $C(A)$

It's also called the <u>column space</u> of A.

- $\text{Null}(A) := \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = 0\}$.

is called the <u>nullspace</u> (or kernel) of A



<u>Thm:</u> $\text{range}(A) = \text{span}\{\vec{a}_1, \ldots, \vec{a}_n\}$

= a set of all possible linear combination of $\{\vec{a}_1, \ldots, \vec{a}_n\}$.

<u>pf:</u> Need to show:

1) $\text{range}(A) \subset \text{span}\{\vec{a}_1, \ldots, \vec{a}_n\}$.

and 2) $\text{span}\{\vec{a}_1, \ldots, \vec{a}_n\} \subset \text{range}(A)$.

1). Pick any $\vec{y} \in \text{range}(A)$. Then $\exists\, \vec{x} \in \mathbb{R}^n$ such that $\vec{y} = A\vec{x} = x_1\vec{a}_1 + \ldots + x_n\vec{a}_n$.

$\Rightarrow \vec{y} \in \text{span}\{\vec{a}_1, \ldots, \vec{a}_n\}$.

2) Take any $\vec{y} \in \text{span}\{\vec{a}_1, \ldots, \vec{a}_n\}$.
Then $\vec{y} = x_1\vec{a}_1 + \ldots + x_n\vec{a}_n$ for some scalars $x_1, \ldots, x_n$.

$$= A\vec{x}$$

$$\Rightarrow \vec{y} \in \text{range}(A).$$

* Linear Independence and Bases

Def: • The vectors $\vec{a}_1, \ldots, \vec{a}_n$ in $\mathbb{R}^m$ are called linearly independent if

$$x_1\vec{a}_1 + \ldots + x_n\vec{a}_n = 0 \iff x_j = 0, \quad 1 \leq j \leq n.$$

• A set of m linearly independent vectors in $\mathbb{R}^m$ is called a basis in $\mathbb{R}^m$.

$\Rightarrow$ A matrix representation of a basis in $\mathbb{R}^m$ is an $m \times m$ matrix. Note that any vector in $\mathbb{R}^m$ can be written as a linear combination of the m basis vectors in $\mathbb{R}^m$.

Def: The dimension of span $\{\vec{a}_1, \ldots, \vec{a}_n\}$ is the maximal number of linearly independent vectors among $\{\vec{a}_1, \ldots, \vec{a}_n\}$.

E.g. $\vec{a}_1 = (1,1,1)^T$, $\vec{a}_2 = (1,1,0)^T$ and $\vec{a}_3 = (0,0,1)^T$.

In $\mathbb{R}^3$, these are linearly dependent

$$\vec{a}_1 = \vec{a}_2 + \vec{a}_3.$$

And $\dim \text{span}\{\vec{a}_1, \vec{a}_2, \vec{a}_3\} = 2$

and $\text{span}\{\vec{a}_1, \vec{a}_2, \vec{a}_3\} = \text{span}\{\vec{a}_2, \vec{a}_3\}$.

We cannot write <u>any vector</u> in $\mathbb{R}^3$ by a lin. combi. of $\{\vec{a}_2, \vec{a}_3\}$.

* <u>Rank:</u>

<u>Def:</u> . The <u>column rank</u> of A

$:= \dim(\text{range}(A))$

$=$ number of lin. indep. column vectors of A.

. The ~~rank~~ <u>row rank</u> of A

$:= \dim(\text{range}(A^T))$

$=$ number of lin indep. row vectors of A

. $\text{rank}(A) = \dim(\text{range}(A))$.

. ~~$A^{m\times n}$ is~~

. $A \in \mathbb{R}^{m\times n}$ is said to be of <u>full rank</u> if $\text{rank}(A) = \min\{m, n\}$.

$$\boxed{\begin{aligned}&\text{Thm: } A \in \mathbb{R}^{m\times n},\ m \geq n, \text{ is of full rank}\\ &\Leftrightarrow \forall \vec{x}, \vec{y} \in \mathbb{R}^n,\ \vec{x} \neq \vec{y},\\ &\qquad A\vec{x} \neq A\vec{y}\end{aligned}}$$

<u>Pf:</u> ($\Rightarrow$) If $\text{rank}(A) = n$, i.e., A is full rank, $\{\vec{a}_1, \dots, \vec{a}_n\}$ are lin. indep.

$\Rightarrow \forall \vec{x}, \vec{y} \in \mathbb{R}^n$ such that $\vec{x} \neq \vec{y}$, $\vec{z} = \vec{x} - \vec{y} \neq 0$.

$A\vec{z} = z_1\vec{a}_1 + \dots + z_n\vec{a}_n \neq 0$.

$\Rightarrow A\vec{x} \neq A\vec{y}$.

($\Leftarrow$) Suppose A is not of full rank, i.e., $\{\vec{a}_1, \dots, \vec{a}_n\}$ are lin. dependent.

$$\Rightarrow \exists\ \vec{c} \in \mathbb{R}^n,\ \vec{c}\neq 0 \text{ such that } \cancel{A\vec{c}\neq 0}$$

$$\sum_{j=1}^{n} c_j \vec{a}_j = 0 \quad \text{or} \quad A\vec{c} = 0.$$

Set $\vec{y} = \vec{x} + \vec{c} \neq \vec{x}$

Then $A\vec{y} = A(\vec{x}+\vec{c}) = A\vec{x} + A\vec{c} = A\vec{x}$.

contradiction!

* Inverse:

. Def: A is said to be nonsingular or invertible $\Leftrightarrow$ A is square and of full rank.

. If $A \in \mathbb{R}^{m\times m}$ nonsingular,

$\Rightarrow \{\vec{a}_1, \ldots, \vec{a}_m\}$ form a basis of $\mathbb{R}^m$.

$\Rightarrow$ The canonical basis vector $\vec{e}_j \in \mathbb{R}^m$ can also be written as a lin. combi. of $\{\vec{a}_1, \ldots, \vec{a}_m\}$

$\exists z_{ij}$ $$\vec{e}_j = \sum_{i=1}^{m} z_{ij}\vec{a}_i ,$$

$$\vec{e}_j = A\vec{z}_j \quad \text{where } \vec{z}_j = (z_{1j}, \ldots, z_{mj})^T.$$

$$[\vec{e}_1 \mid \vec{e}_2 \mid \ldots \mid \vec{e}_m] = [A\vec{z}_1 \mid A\vec{z}_2 \mid \ldots \mid A\vec{z}_m].$$

$$\underbrace{I}_{m\times m \text{ identity matrix}} = AZ.$$

Such matrix $Z \in \mathbb{R}^{m\times m}$ is called the inverse of A and written as A^{-1}.

Any nonsingular matrix has a unique inverse, and $AA^{-1} = A^{-1}A = I$.

* **Thm:** (Equivalences of a nonsingular matrix)

For $A \in \mathbb{R}^{m \times m}$, the following statements are equivalent:

a) A has an inverse A^{-1}.

b) $\text{rank}(A) = m$.

c) $\text{range}(A) = \mathbb{R}^m$.

d) $\text{Null}(A) = \{0\}$.

e) 0 is not an eigenvalue of A.

f) 0 is not a singular value of A.

g) $\det(A) \neq 0$.

* <u>Matrix inverse times a vector:</u>

$\vec{y} = A\vec{x}$. $\quad$ (A nonsingular)

$\Rightarrow \quad \vec{x} = A\vec{y}$.

This means that $A^{-1}y$ represents an <u>expansion coefficients</u> of $\vec{y}$ in the basis of columns of A.

$\Rightarrow$ Multiplication by A^{-1} is a <u>change of basis</u> operation!

$\vec{y}$: coefficients of expansion in $\{\vec{e}_1, \ldots, \vec{e}_m\}$ $\xrightarrow{A^{-1}}$ $A^{-1}\vec{y}$: coeff's of expansion of $\vec{y}$ in $\{\vec{a}_1, \ldots, \vec{a}_m\}$.

$A^{-1}\vec{y}$: coeff's of expansion of $\vec{y}$ in $\{\vec{a}_1, \ldots, \vec{a}_m\}$ $\xrightarrow{A}$ $\vec{y}$: coefficients of expansion in $\{\vec{e}_1, \ldots, \vec{e}_m\}$

Inner Product and Norms

* **Inner Product:**

- **Def:** The <u>inner product</u> between two vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ is defined as

$$\langle \vec{x}, \vec{y} \rangle = \vec{x} \cdot \vec{y} = \vec{x}^T \vec{y} = \sum_{i=1}^{n} x_i y_i \quad (\in \mathbb{R})$$

The <u>ℓ^2-norm of $\vec{x} \in \mathbb{R}^n$</u> is defined as

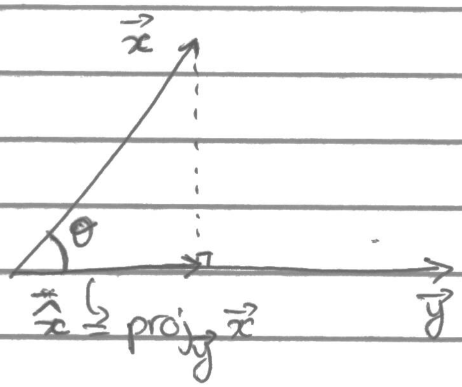
$$\|\vec{x}\|_2 = \sqrt{\langle \vec{x}^T, \vec{x} \rangle} = \sqrt{\sum_{i=1}^{n} |x_i|^2}$$

We also denote it by $\|\vec{x}\|$

(This is the <u>Euclidean length</u> of $\vec{x}$.)

The <u>angle</u> θ between $\vec{x}, \vec{y} \in \mathbb{R}^n$ can be computed by

$$\cos\theta = \frac{\langle \vec{x}, \vec{y} \rangle}{\|\vec{x}\| \, \|\vec{y}\|}.$$



The <u>projection</u> of $\vec{x}$ ~~onto~~ onto $\vec{y}$ is

$$\text{proj}_{\vec{y}} \vec{x} = \|\vec{x}\| \cos\theta \, \frac{\vec{y}}{\|\vec{y}\|} = \frac{\vec{x}^T \vec{y}}{\|\vec{y}\|^2} \vec{y}.$$

* Vector norms: to quantify (or measure) the size (or length) of a vector.

. Def: A norm is a function

$$\|\cdot\| : \mathbb{R}^n \to \mathbb{R} \quad \text{such that}$$

$\forall \vec{x}, \vec{y} \in \mathbb{R}^n$ and $\forall \alpha \in \mathbb{R}$

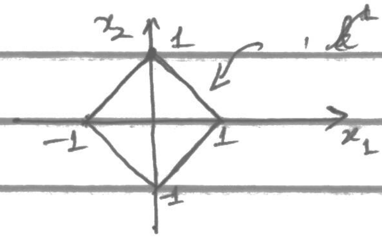
1) $\|\vec{x}\| \geq 0$ and $\|\vec{x}\| = 0 \iff \vec{x} = 0$.

2) $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ (the triangle inequality)

3) $\|\alpha \vec{x}\| = |\alpha| \, \|\vec{x}\|$.

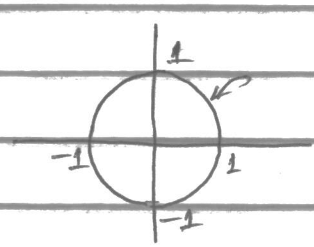
Examples: p-norms (ℓ^p-norms)

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



unit ℓ^1-ball.

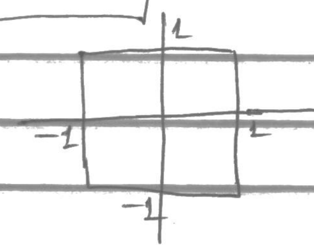
$$\|x\|_2 := \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}$$



unit ℓ^2-ball.

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}, \quad 1 \leq p < \infty$$

$$\|x\|_\infty := \max_{1 \leq i \leq n} |x_i|$$



Exercise: what is the vector $\vec{x} \in \mathbb{R}^2$ that achives $\max \|x\|_1$ subject to $\|x\|_2 = 1$?

* 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-~~Sack~~ Schmidt) norm of $A \in \mathbb{R}^{m\times n}$ is defined as

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

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

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

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

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

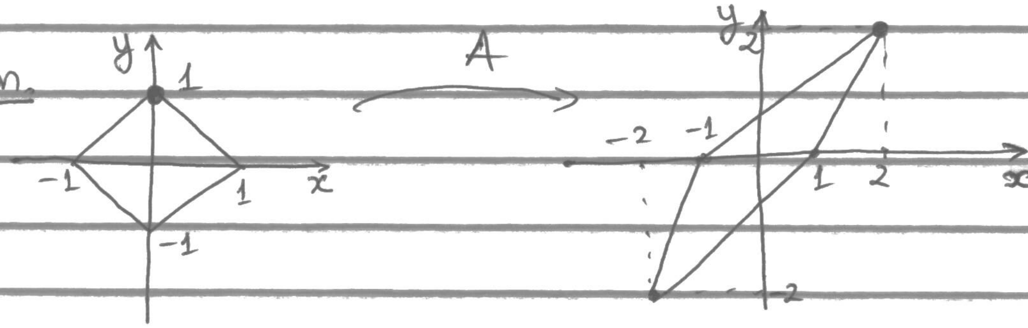
$$\|A\|_p := \sup_{\vec{x}\neq 0} \frac{\|A\vec{x}\|_p}{\|\vec{x}\|_p} = \sup_{\|\vec{x}\|_p=1} \|A\vec{x}\|_p .$$

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

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

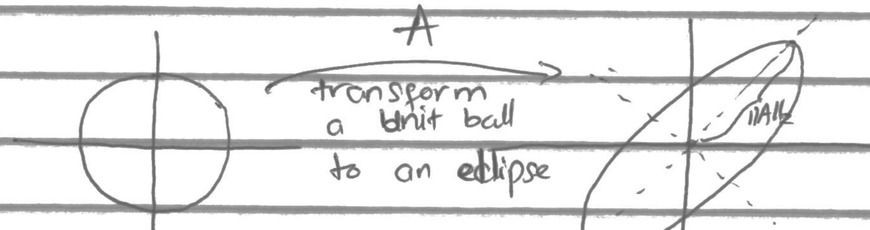
~~In gene~~ achived for $\vec{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ -1 \end{bmatrix}$.

How about $\|A\|_2$?

We will show later that

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

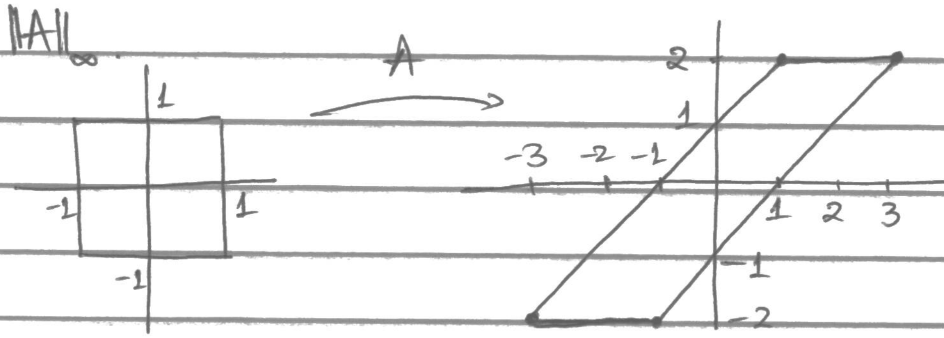
→ the largest eigenvalue of $A^T A$.



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 $|d_1|, \dots, |d_n|$

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

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

* <u>the 1-norm of a matrix</u>.

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

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

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

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

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

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

$$\leq \left(\sum_{i=1}^{n} |x_i|\right) \max_{1 \leq i \leq n} \|\vec{a}_i\|_1$$

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

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

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

- Ans: Yes!

~~Let~~ $\cancel{\vec{e}_k}$ ~~be~~ Take the (k)th column whose has largest 1-norm: ~~let~~

$$\|\vec{a}_k\|_1 = \max_{1 \leq j \leq n} \|\vec{a}_j\|_1$$

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

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

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

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

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

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

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

$\vec{x}$ = eigenvector and λ = eigenvalue. of A^TA.

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

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

$$\longrightarrow \quad \ge 0$$

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

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

* The ∞-norm of a matrix:

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

$\|A\|_\infty = \max\limits_{1\le i\le m} \|\vec{a}_{i\cdot}\|_1$ $\longrightarrow$ 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

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

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

triangle inequality $\le \max\limits_{1\le i\le m} \sum\limits_{j=1}^{n} |a_{ij}|\,|x_j|$

$\le \|\vec{x}\|_\infty \max\limits_{1\le i\le m} \sum\limits_{j=1}^{n} |a_{ij}|$

$\Rightarrow \dfrac{\|Ax\|_\infty}{\|x\|_\infty} \le \max\limits_{1\le i\le m} \sum\limits_{j=1}^{n} |a_{ij}| = \max\limits_{1\le i\le m} \|a_{i\cdot}\|_1$

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

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

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~~ 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 "<u>bad</u>", i.e., there is a large error in solution $\vec{x}^* = A^{-1}\vec{b}$.

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

---

* <u>Orthogonal Vectors</u>:

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

(The zero vector is orthogonal to any vector)

. Two <u>sets</u> of vectors X, Y are said to be <u>orthogonal</u> 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 <u>orthogonal</u> 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 <u>orthonormal</u> if S is orthogonal and $\forall \vec{x} \in S$,

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

(orthonormal = othogonal + normalized)

<u>Thm</u>: The vectors in an orthogonal set S are <u>linearly independent</u>.

<u>Pf</u>: let $S = \{v_1, \ldots, v_n\}$.

Suppose they are <u>not lin. indep.</u>

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,

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

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

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

$$= 0$$

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

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

contradiction!

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

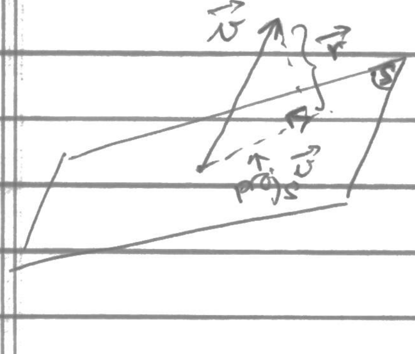
"Inner products can be used to decompose arbitrary vectors into orthogonal components!".

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?

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

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

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

$$= 0.$$

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

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

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

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

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 $m = n$.