Vectors and Matrices Review

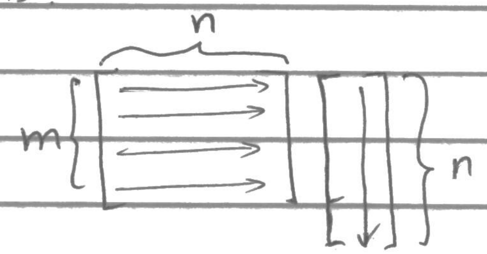
* <u>Matrix – Vector Multiplication:</u>

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

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

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., $\vec{e}_j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow j$

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

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

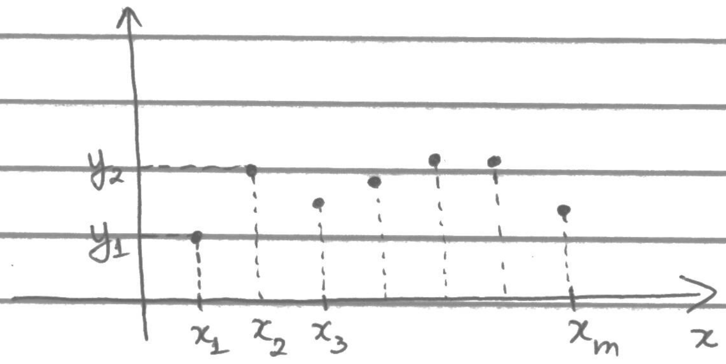
If $F_A = F_B$, Then

$F_A(\vec{e}_j) = \vec{a}_j = F_B(\vec{e}_j) = \vec{b}_j$ 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 $n-1$:

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

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

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

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> !

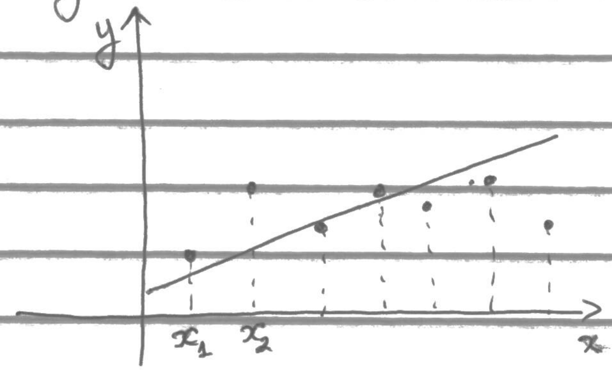
So, $\exists A \in \mathbb{R}^{m\times n}$ for such linear map F.
What is this matrix A?

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

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

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

$\Rightarrow \quad \vec{c}_j = A\vec{b}_j$. 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 $\vec{u}$.

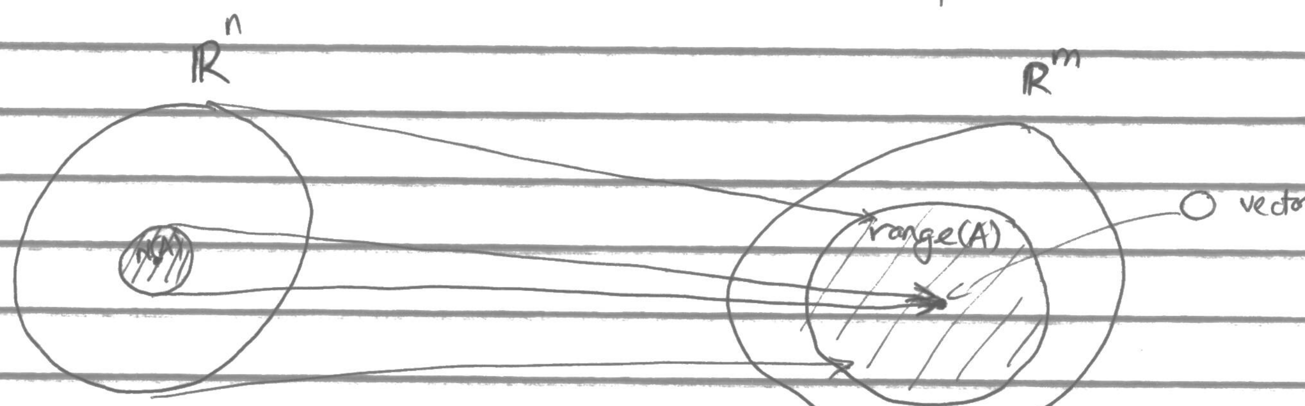
* Range and Nullspace (or Kernel)

Def: 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 $\underline{\text{Im}}(A)$ or $C(A)$
image

It's also called the column space of A.

. $\text{Null}(A) := \{ \vec{x} \in \mathbb{R}^n \mid A\vec{x} = 0 \}$.
is called the nullspace (or kernel) of A



Thm: $\text{range}(A) = \text{span}\{\vec{a}_1, \dots, \vec{a}_n\}$
= a set of all possible linear combination of $\{\vec{a}_1, \dots, \vec{a}_n\}$.

pf: Need to show:

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

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

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

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

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

$$= A\vec{x}$$

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

* <u>Linear Independence and Bases</u>

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

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

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

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

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

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

In $\mathbb{R}^3$, these are <u>linearly dependent</u>

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

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 $\text{rank}(A) = \min\{m, n\}$.

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

<u>Pf:</u> $(\Rightarrow)$ If $\text{rank}(A) = n$, i.e., A is full rank,

$\{\vec{a}_1, \ldots, \vec{a}_n\}$ are lin. indep.

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

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

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

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

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

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

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,

$\Rightarrow \{\vec{a}_1, \dots, \vec{a}_m\}$ 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 $\{\vec{a}_1, \dots, \vec{a}_m\}$

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

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

$$\underbrace{[\vec{e}_1 \mid \vec{e}_2 \mid \dots \mid \vec{e}_m]}_{\underbrace{I}_{m\times m \text{ identity matrix}}} = [A\vec{z}_1 \mid A\vec{z}_2 \mid \dots \mid A\vec{z}_m].$$

$$I = AZ.$$

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

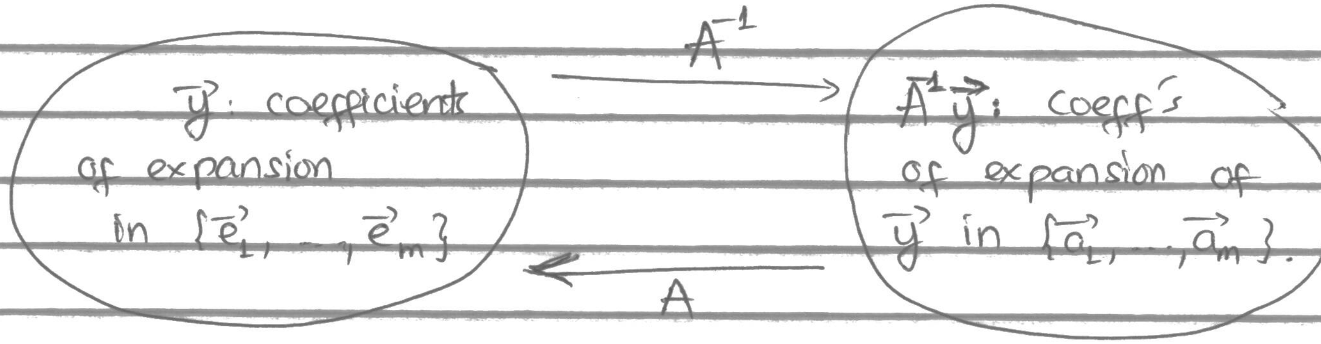
* <u>Matrix inverse times a vector:</u>

$\vec{y} = A\vec{x}$: (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!



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 $\vec{x} \in \mathbb{R}^n$ is defined as</u>

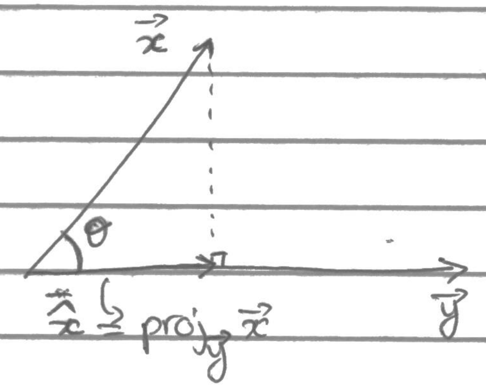
$$\|\vec{x}\|_2 = \sqrt{\langle \vec{x}^T, \vec{x} \rangle} = \sqrt{\sum_{i=1}^{n} |x_i|^2}$$

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 $\vec{x}$ ~~onty~~ onto $\vec{y}$ is

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

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

. **Def:** A **norm** is a function

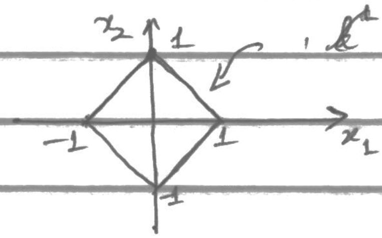
$$\|\cdot\| : \mathbb{R}^n \to \mathbb{R} \quad \text{such that}$$

$\forall \vec{x}, \vec{y} \in \mathbb{R}^n$ 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) $\|\alpha \vec{x}\| = |\alpha| \, \|\vec{x}\|$.

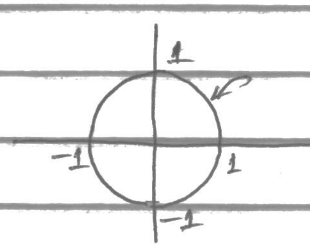
Examples: p-norms (ℓ^p-norms)

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



unit ℓ^1-ball.

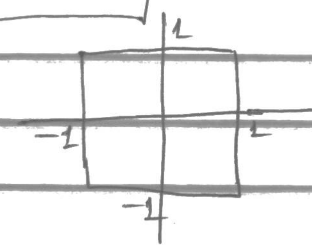
$$\|x\|_2 := \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}$$



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 $\|x\|_2 = 1$?