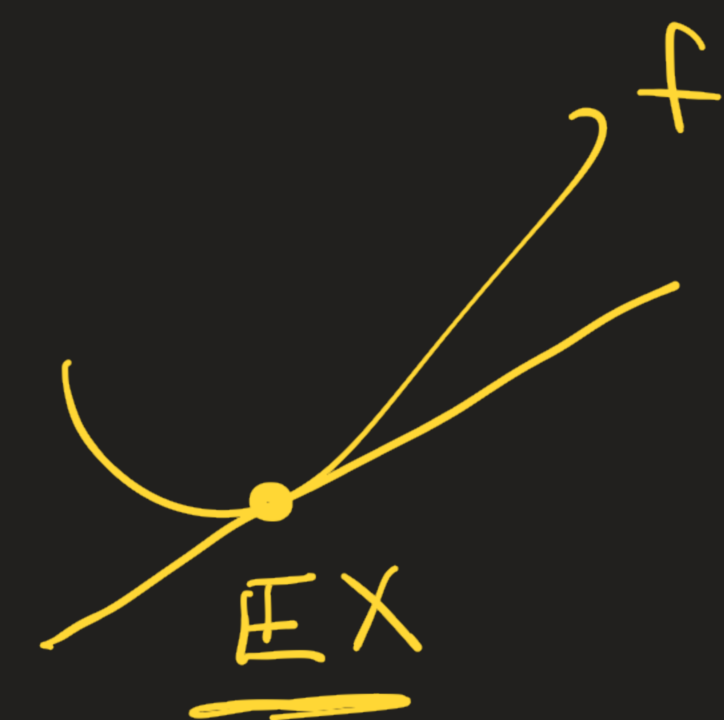
Jensen's Ineq If f is a convex function on a set C and X is a random variable that takes values in C, then

$$\mathbb{E} f(X) \geq f(\mathbb{E}X)$$

if g is concave (i.e. $f = -g$ is cvx) then

$$\mathbb{E} g(X) \leq g(\mathbb{E}X)$$



Prf (assume f is differentiable)

By the first-order condition, for any $x \in C$ we have

$$f(x) \geq f(\mathbb{E}X) + \nabla f(\mathbb{E}X)^T (x - \mathbb{E}X)$$

Now take the expectation when $x \sim X$

$$\Rightarrow \underline{\mathbb{E}f(X)} \geq \mathbb{E}\left[f(\mathbb{E}X) + \nabla f(\mathbb{E}X)^T(X - \mathbb{E}X)\right]$$

$$= f(\mathbb{E}X) + \mathbb{E}\left[\underbrace{\nabla f(\mathbb{E}X)^T(X - \mathbb{E}X)}_{= \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(\mathbb{E}X)(X - \mathbb{E}X)_i}\right]$$

$$= f(\mathbb{E}X) + \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(\mathbb{E}X)\, \underbrace{\mathbb{E}\left((X - \mathbb{E}X)_i\right)}_{= (\mathbb{E}X - \mathbb{E}X)_i = 0}$$

$$= f(\mathbb{E}X) + \nabla f(\mathbb{E}X)^T(\mathbb{E}X - \mathbb{E}X)$$

$$= \underline{f(\mathbb{E}X)}$$

$\blacksquare$

<u>Ex</u>

$$f(x) = \text{logsumexp}(x)$$

Jensen's $\Rightarrow$ $\mathbb{E}\,\text{logsumexp}(X) \geq \text{logsumexp}(\mathbb{E}X)$

Examples of cvx function

$$f(x) = x^2 \Rightarrow \frac{df}{dx} = 2x \Rightarrow \frac{d^2f}{dx^2} = 2 \geq 0 \text{ everywhere}$$

$\Rightarrow f(x)$ is convex on $\mathbb{R}$

e^{-x}

$$f(x) = e^{\lambda x} \Rightarrow \frac{df}{dx} = \lambda e^{\lambda x} \Rightarrow \frac{d^2f}{dx} = \lambda^2 e^{\lambda x} \geq 0$$

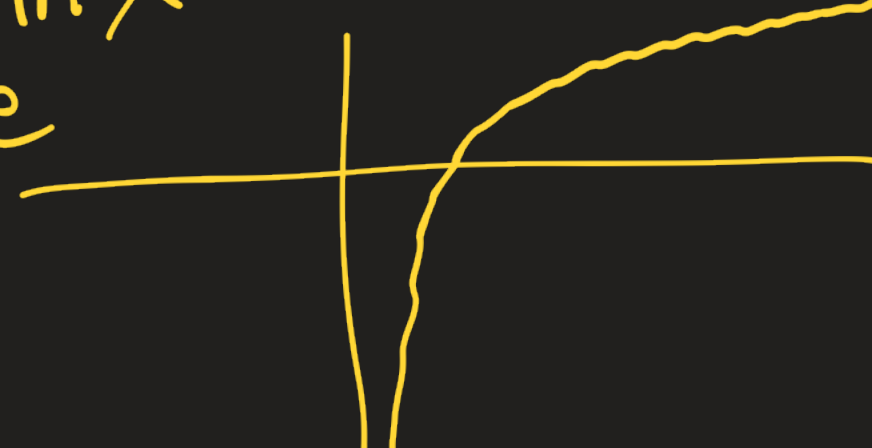
everywhere

e^x

$\Rightarrow f(x)$ is convex on $\mathbb{R}$

$f(x) = \ln x$ is concave

$\ln(x)$



$$\frac{df}{dx} = \frac{1}{x} \Rightarrow \frac{d^2f}{dx^2} = -\frac{1}{x^2} \leq 0$$

$\Rightarrow f$ is concave on $(0, \infty)$

$f(x) = x_+$

Check this satisfies

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$

for all $x, y \in \mathbb{R}$ $\Rightarrow$ f is convex on $\mathbb{R}$

$f(x) = \frac{1}{2}\|x\|_2^2$

$$\nabla f(x) = x \quad \Rightarrow \quad \nabla^2 f = I$$

Any norm of a vector (matrix) is convex.

Recall a norm $|||\cdot|||$ of a vector is a function that satisfies:

$d(x,y) = |||x-y|||$

1) $|||x||| \geq 0$ for all vectors x

2) $|||x||| = 0 \iff x = 0$

3) $|||\alpha x||| = |\alpha|\, |||x|||$ for any $\alpha \in \mathbb{R}$

4) $|||x+y||| \leq |||x||| + |||y|||$

Fact any norm is convex

Consq: $\|x\|_1$ is convex

- $f(x) = Ax + b$ (affine functions) are convex

b/c $f(\alpha x + (1-\alpha)y) = \alpha f(x) + (1-\alpha) f(y)$

(check!)

- $f(x) = \frac{1}{2} x^T A x + b^T x + c$ (quadratic functions)

$\Rightarrow \nabla f(x) = Ax + b$

$\Rightarrow \nabla^2 f(x) = A$

so f is convex iff $A \succeq 0$

Operations that preserve convexity

- If $f_1, \dots, f_n$ are convex and $\alpha_1, \dots, \alpha_n \geq 0$ then

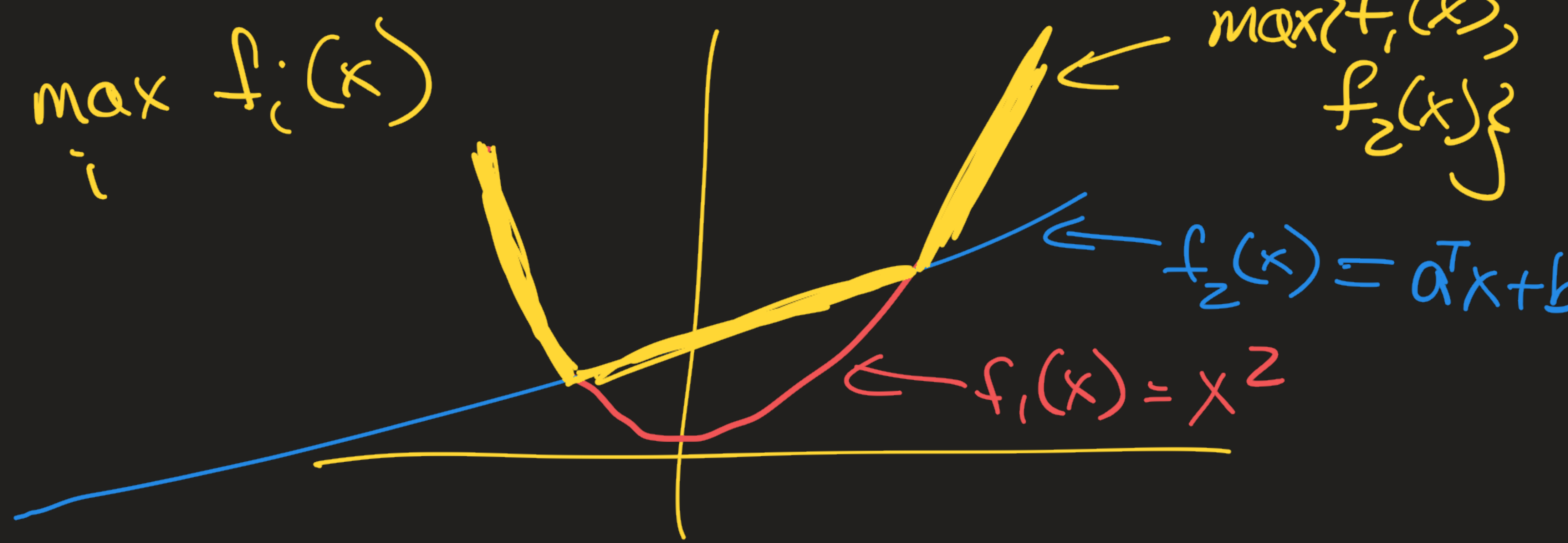
$$f = \sum \alpha_i f_i \text{ is convex}$$

(use the definition of cvxty)

- If f is convex then $f(Ax+b)$ is convex

- The pointwise maximum of convex fcns is cvx

$$f(x) = \max_i f_i(x)$$



Prf

We want to show $f(\alpha x+(1-\alpha)y) \leq \alpha f(x)+(1-\alpha)f(y)$

Note

$$f(\alpha x+(1-\alpha)y) = \max_i f_i(\alpha x+(1-\alpha)y)$$

by cvxty of the f_i

$$\leq \max_i \left[\alpha f_i(x) + (1-\alpha) f_i(y)\right]$$

$$\leq \max_i \left[\alpha f_i(x)\right] + \max_i \left[(1-\alpha) f_i(y)\right]$$

$$= \alpha \max_i f_i(x) + (1-\alpha) \max_i f_i(y)$$

$$= \alpha f(x) + (1-\alpha) f(y)$$

$\blacksquare$

Q: is $(x_+)^2$ cvx?

is $|||x|||^2$ cvx where $|||\cdot|||$ is a norm?

Fact if f is a convex & non-decreasing function ($x \geq y \Rightarrow f(x) \geq f(y)$) and g is a convex function, then $(f \circ g)(x) = f(g(x))$ is a convex function

Prf

Consider $(f \circ g)(\alpha x + (1-\alpha)y) = f(g(\alpha x + (1-\alpha)y))$

use the fact that g is convex to see

$$g(\alpha x + (1-\alpha)y) \leq \alpha g(x) + (1-\alpha)g(y)$$

and use that f is non-decreasing to see

$$f(g(\alpha x + (1-\alpha)y)) \leq f(\alpha g(x) + (1-\alpha)g(y))$$

now f is convex, so we have

$$f(\alpha g(x) + (1-\alpha) g(y)) \le \alpha f(g(x)) + (1-\alpha) f(g(y))$$
$$= \alpha (f \circ g)(x) + (1-\alpha)(f \circ g)(y)$$

Q: does this imply that the square of any convex function is convex?

No: because x^2 is only non-decreasing on $\mathbb{R}^+$

Fact The square of any positive convex function is convex

Cor: the square of any norm is cvx

Examples of convex sets

- empty set

$\mathbb{R}$, $\mathbb{R}^d$, $\mathbb{R}_+$, $\mathbb{R}_+^d$, lines, line segments, rays, subspaces, $\mathbb{C}$, $\mathbb{C}^d$

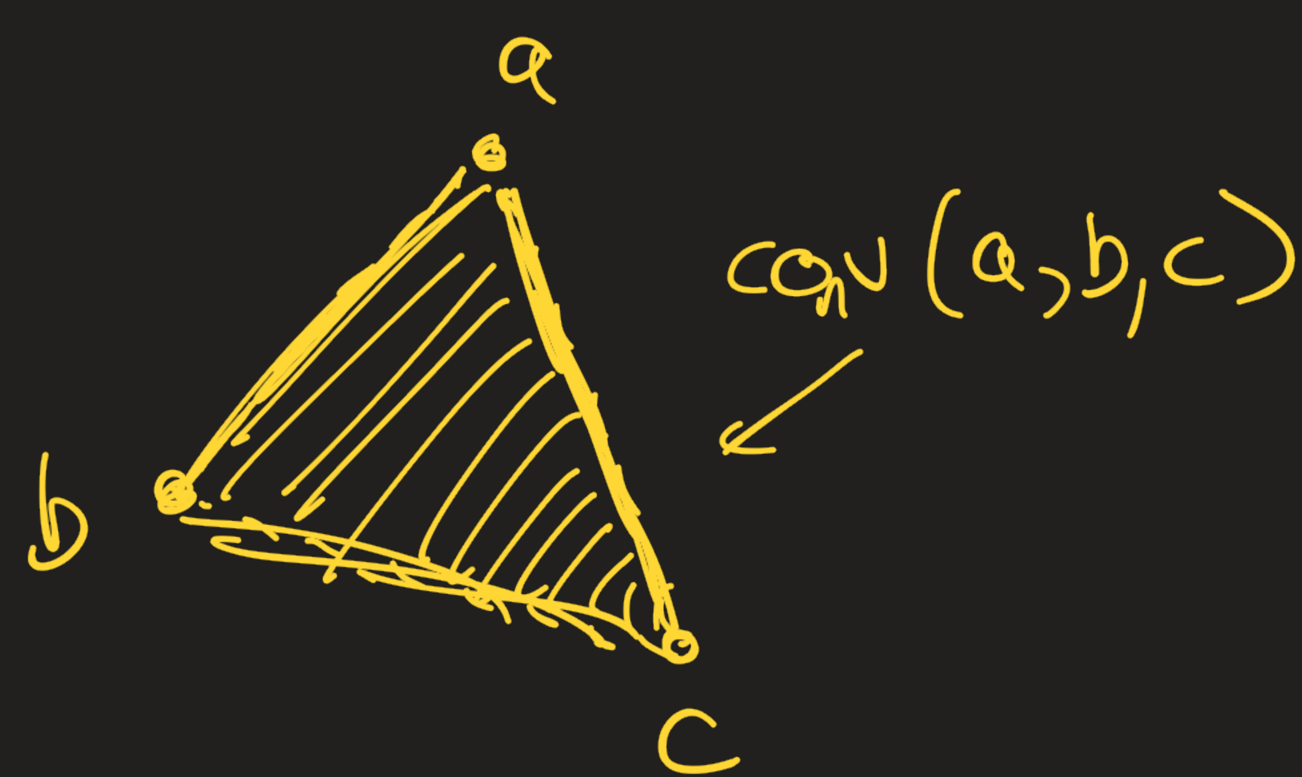
- hyperplanes: $\{x \mid a^T x = 0\} = H_a$

- half-spaces: $\{x \mid a^T x \leq b\} = H^{\leq}_{a,b}$

$a^T x \leq 0$

- convex hull of set S

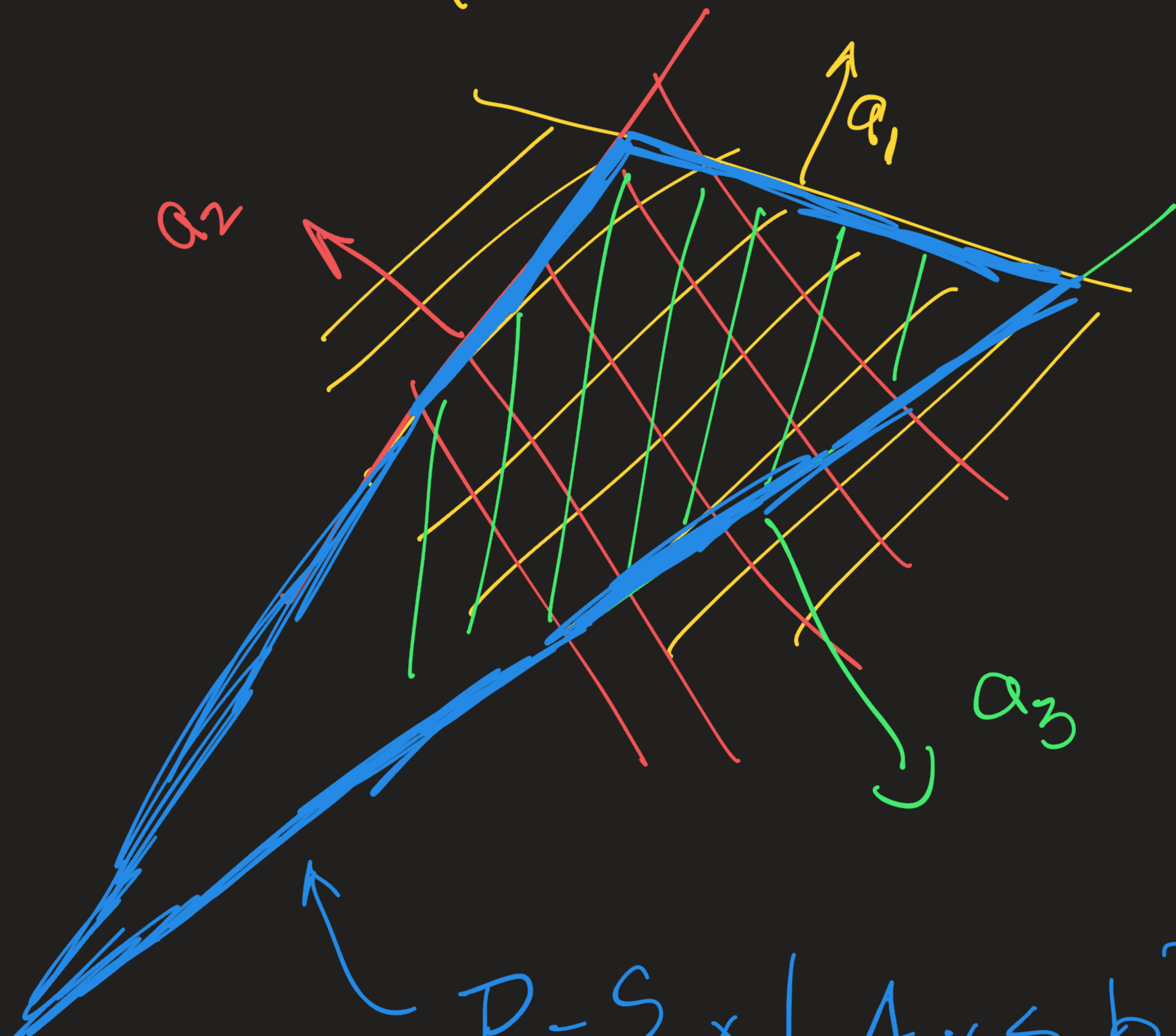
$$\mathrm{conv}(S) = \{ \theta_1 x_1 + \cdots + \theta_k x_k : x_i \in S, \; \theta_i \geq 0, \; \sum \theta_i = 1 \}$$



- norm balls $B_{\|\|\cdot\|\|}(x_c, r) = \{ x \mid \|\|x - x_c\|\| \leq r \}$

$B_{\|\cdot\|_2}(0, r)$

- polyhedrons: set of solutions to a finite number of linear equalities & ineqs.



$a_1^T x \leq b_1$

$a_2^T x \leq b_2$

$a_3^T \leq b_3$

$P = \{ x \mid Ax \leq b \}$

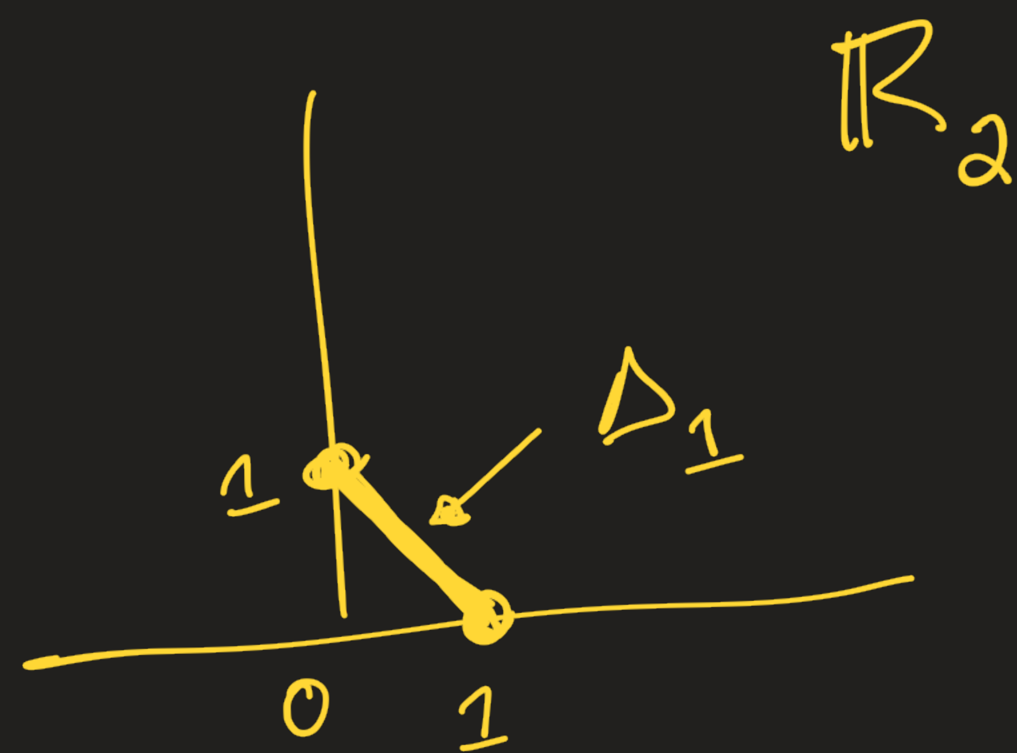
where $A = \begin{bmatrix} a_1^T \\ a_2^T \\ a_3^T \end{bmatrix}$

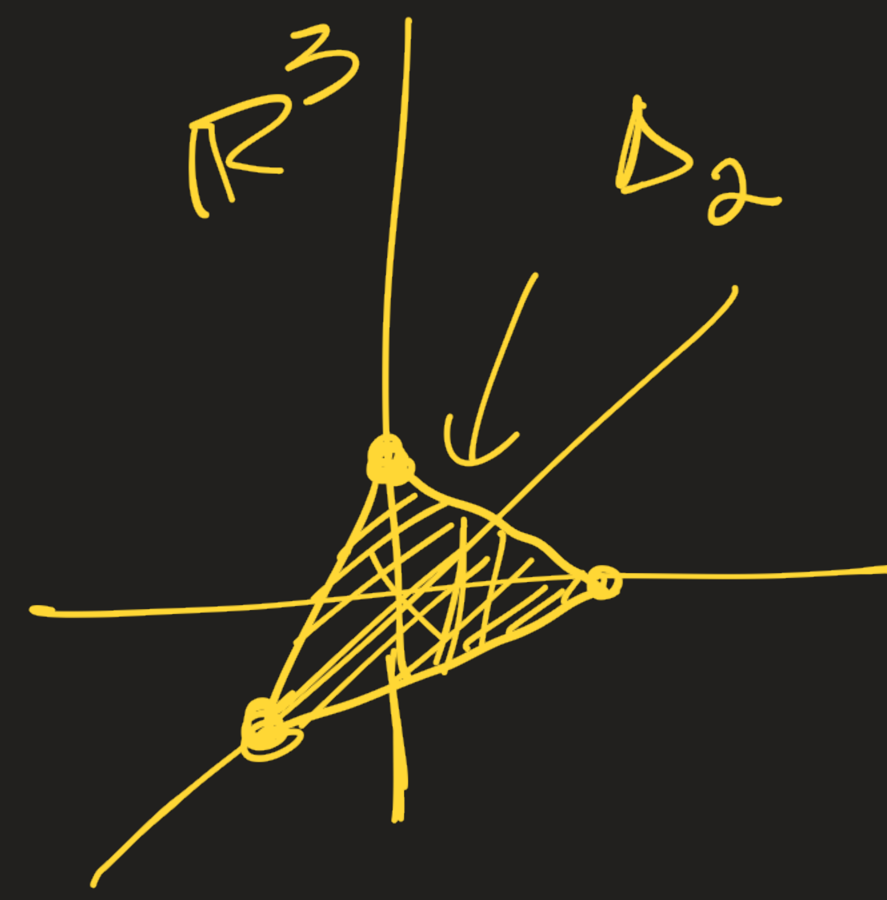
$b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$

- Intersection of arb # of convex sets is convex

$$\bigcap_{i \in I} C_i \text{ is convex}$$

- probability simplex: $\Delta_{d-1} = \{\theta \mid \theta \in \mathbb{R}^d \text{ probability vectors}\}$





Examples of Convex Optim Prob

1) Linear programming

$$\min_{x} \; c^T x$$
$$Ax \leq b$$

b/c $x \mapsto c^T x$ is cvx (linear funcs are cvx)

and

$\{x \mid Ax \leq b\}$ is cvx (polyhedra are cvx)

2) Least-squares

$$x^* = \operatorname{argmin}_{x \in \mathbb{R}^d} \|Ax - b\|_2^2$$

and $\mathbb{R}^d$ is cvx

b/c $Ax - b$ is affine (so cvx)

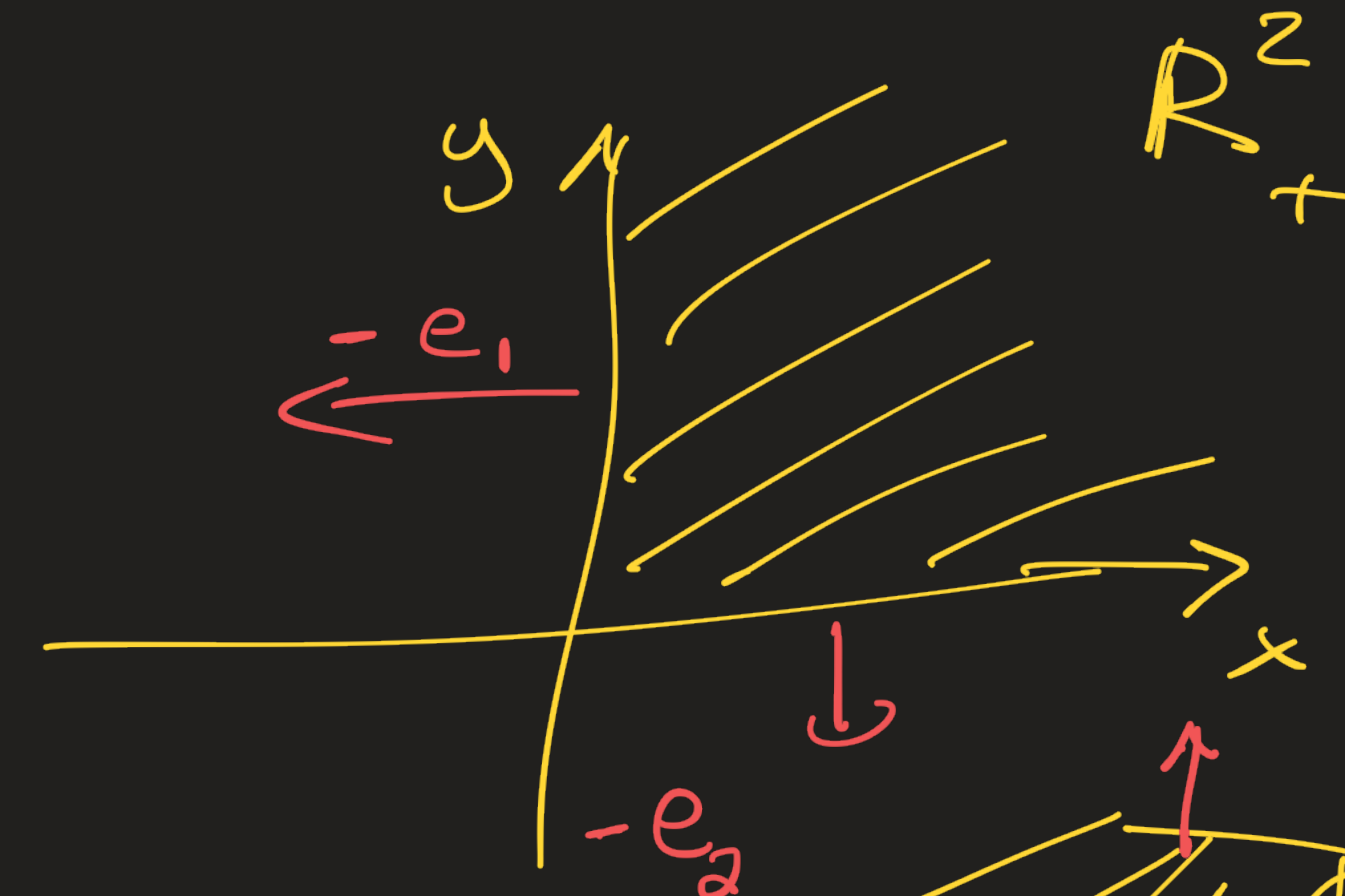
and $\|\cdot\|_2^2$ is cvx

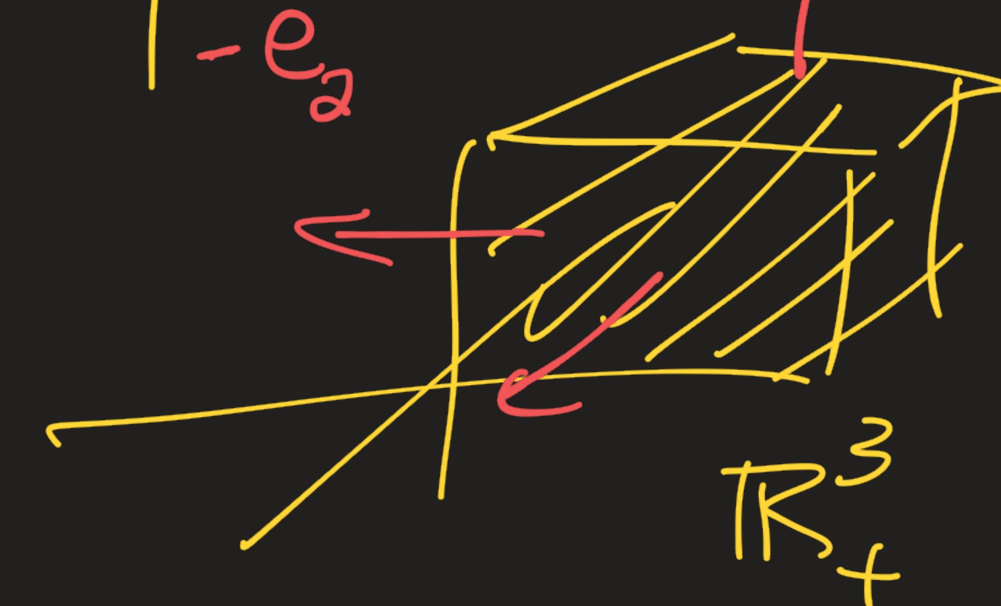
and cvx fun of a affine fun is cvx

3) Nonnegative least squares

$$x^* = \underset{x}{\operatorname{argmin}} \|Ax - b\|_2^2$$

$$\boxed{x \geq 0}$$





$$\{x \mid x \geq 0\}$$

$$= \left\{x \;\middle|\; \begin{matrix} x_1 \geq 0 \\ \vdots \\ x_d \geq 0 \end{matrix}\right\}$$

$$= \left\{x \;\middle|\; \begin{matrix} e_1^T x \geq 0 \\ \vdots \\ e_d^T x \geq 0 \end{matrix}\right\}$$

$$= \left\{x \;\middle|\; \begin{matrix} -e_1^T x \leq 0 \\ \vdots \\ -e_d^T x \leq 0 \end{matrix}\right\} = \{x \mid \underbrace{-I}_{A}\, x \leq \underbrace{0}_{b}\}$$ is a polyhedron

2) Regularized least-squares, e.g.

$$\min_x \underbrace{\underbrace{\|Ax-b\|_2^2}_{\text{cvx}} + \lambda\underbrace{\|x\|_1}_{\text{cvx}}}_{\text{cvx b/c is positive weighted sum of cvx}} \qquad (\lambda \geq 0)$$

optimizing over all of $\mathbb{R}^d$, which is cvx

3) Constrained least squares, e.g.

$$\min_x \underbrace{\|Ax-b\|_2^2}_{\text{cvx}}$$

$$\underbrace{\|x\|_\infty \le \tau}_{B_\infty(0,\tau) \text{ is a convex set}}$$

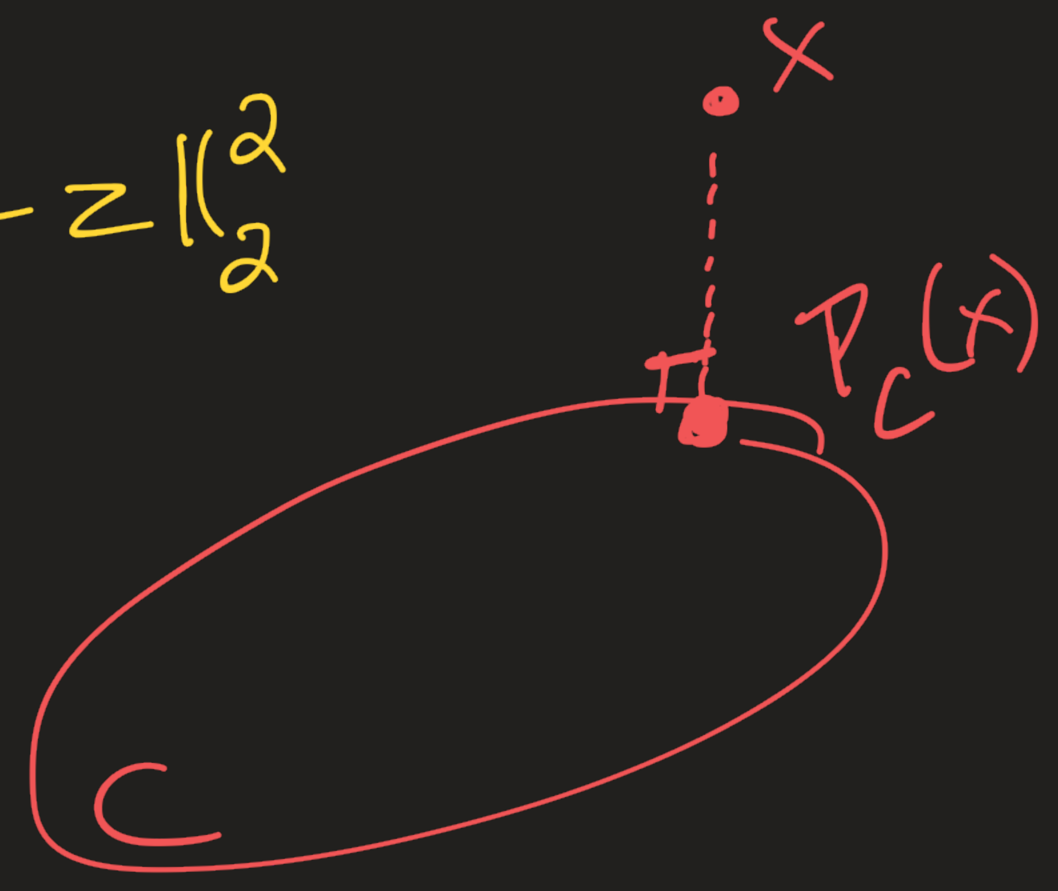
4)

$$\min \|Ax-b\|_2^2$$

$$\left.\begin{array}{l} \|x\|_\infty \le \tau \\ \|x\|_2 \le \gamma \end{array}\right\} \begin{array}{l} \text{two convex sets (both norm balls)} \\ \text{so their intersection is cvx} \end{array}$$

4) Projection onto a convex set C

$$P_C(x) = \underset{z \in C}{\operatorname{argmin}} \|x - z\|_2^2$$



$P_C(x)$ is the closest point in C to x

$z \mapsto x - z$ is affine function

$\| \|_2^2$ is convex

so $z \mapsto \|x - z\|_2^2$ is convex

and C is convex

Min-max regression

Suppose we want to solve a regression problem and minimize the worst pointwise error

find x so $b_i \approx a_i^T x$

$$\min_x \max_i |b_i - a_i^T x| = \min_x \max_i |(Ax-b)_i|$$

$$= \min_{x \in \mathbb{R}^d} \|Ax - b\|_\infty$$

where $\|z\|_\infty = \max_i |z_i|$ is the inf-norm/max-norm/ℓ_∞-norm

RERM

In particular: - if $\ell(u,v)$ is cvx in its first argument

$$\ell(\alpha x+(1-\alpha)y, v) \le \alpha \ell(x,v) + (1-\alpha)\ell(y,v)$$

- our hypotheses are linear of the form

$$w_0 + w^T x$$

or

$$wx + b$$

- and the regularizer Ω is convex

Then

$$F(w) = \frac{1}{n}\sum_{i=1}^{n} \ell\left(w_0 + w^T x_i, y_i\right) + \Omega(w_0, w)$$

is convex in w_0, w