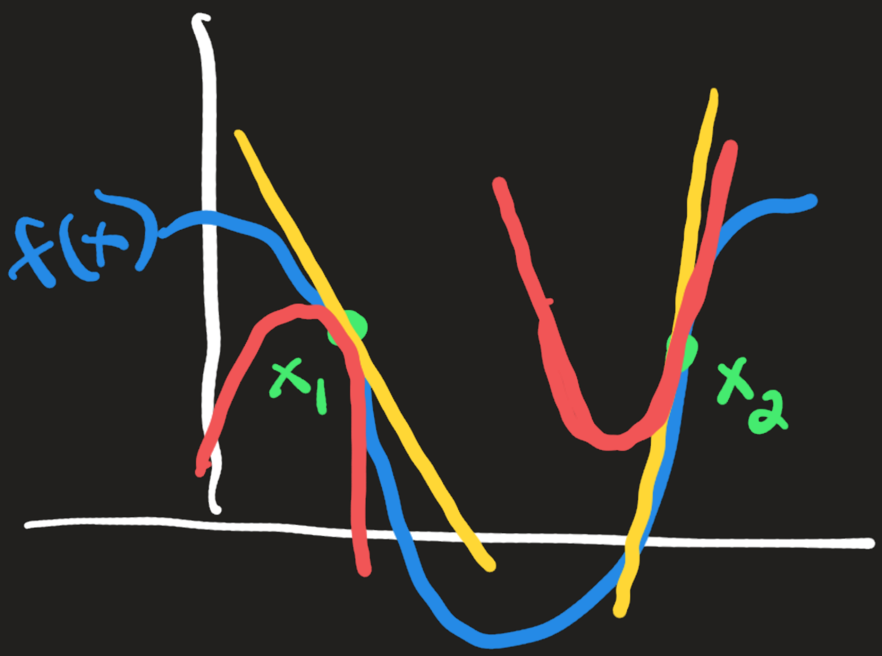
ML and Opt Lec 7

- Review Taylor Series
- Oracle Methods of Optimization
- Convex Sets & Functions ; Convex Optim Probs

Taylor Series

$$g \in o(h) \iff \lim_{x \to 0} \frac{g(x)}{h(x)} = 0$$

Taylor series is a local polynomial approximation, taken at a point, of a differentiable function f



1st-order TS

$$f(x) = f(x_1) + f'(x_1)(x - x_1) + o(|x - x_1|)$$
$$= f(x_2) + f'(x_2)(x - x_2) + o(|x - x_2|)$$

2nd-order TS

$$f(x) = f(x_1) + f'(x_1)(x - x_1) + \frac{1}{2} f''(x_1)(x - x_1)^2 + o(|x - x_1|^2)$$

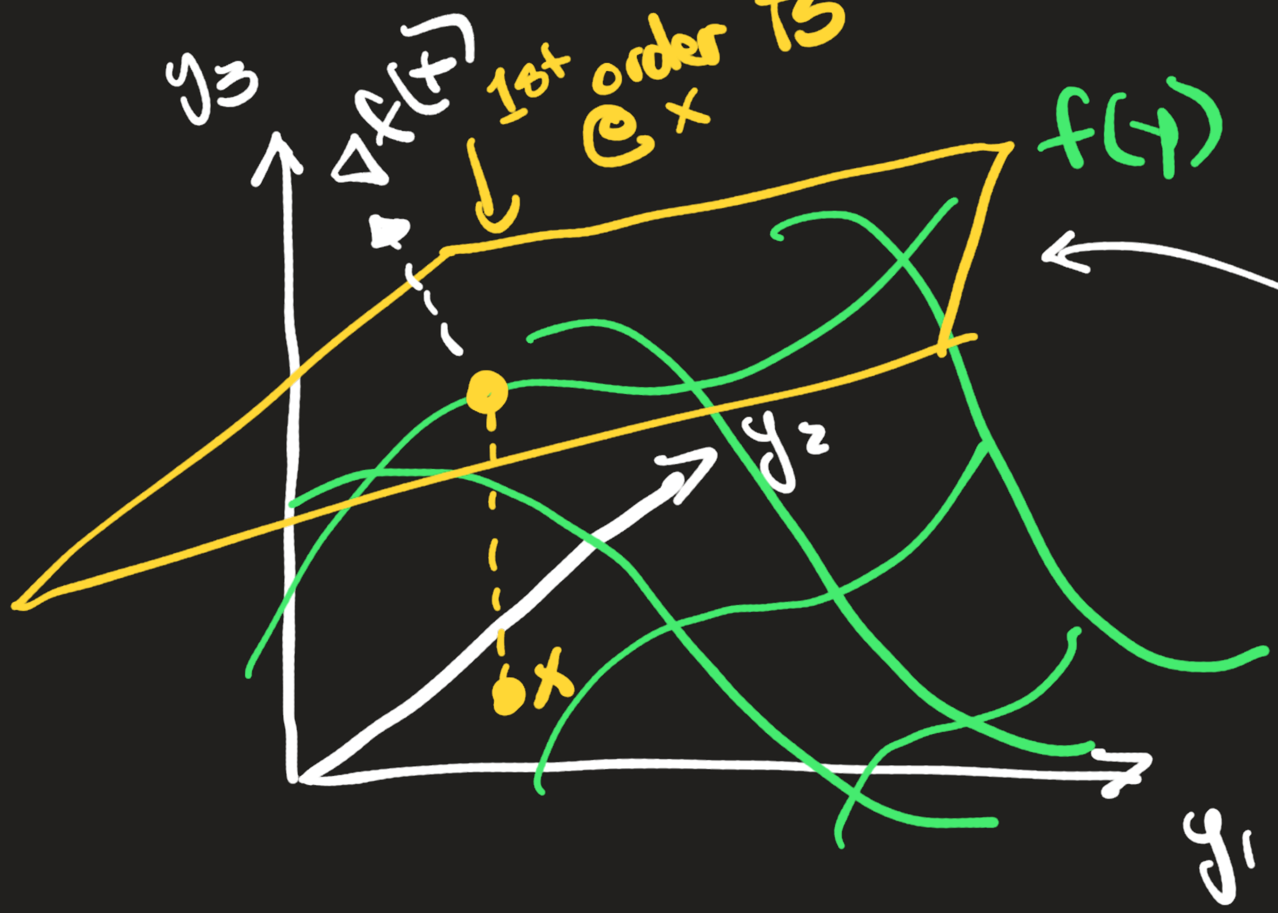
$$= f(x_2) + f'(x_2)(x - x_2) + \frac{1}{2} f''(x_2)(x - x_2)^2 + o(|x - x_2|^2)$$

Multivariate TS

1ST ORDER T.S.

$$f(y) = f(x) + \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(x)(y_i - x_i)$$

$$= f(x) + \langle \nabla f(x), y - x \rangle + o(\|y-x\|_2)$$



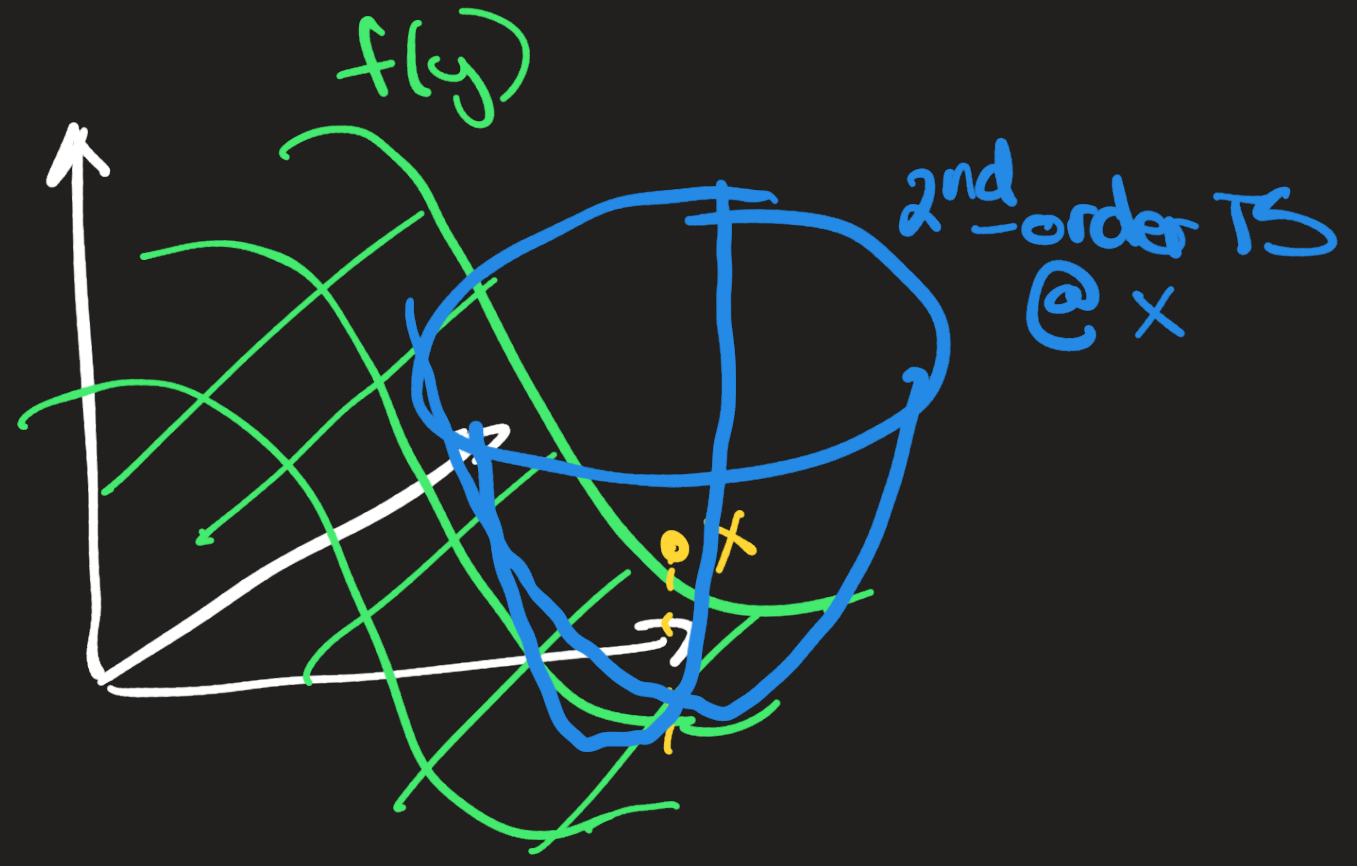
the graph of the 1st order TS is a tangent hyperplane to the graph of f @ x

$\underline{\text{2nd-order TS}}$ (accounts for curvature)

$$f(y) = f(x) + \langle \nabla f(x), y-x \rangle + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2 f(x)}{\partial x_i \partial x_j}(y_i - x_i)\cdot(y_j - x_j)$$

$$= f(x) + \langle \nabla f(x), y-x \rangle + \frac{1}{2}(y-x)^T \nabla^2 f(x)(y-x) + o(\|y-x\|_2^2)$$

(used fact that $v^T M v = \sum_{i=1}^{d}\sum_{j=1}^{d} M_{ij} v_i v_j$)



Oracle Methods of Optimization

To solve a general optim prob numerically, we use iterative methods:

$x_0 \leftarrow$ initial guess of the solution

- compute $f(x_i)$ and/or $\nabla f(x_i)$ and/or $\nabla^2 f(x_i)$
- form a local model
- move to a new point x_{i+1} based on the local model

repeat until "convergence"

zeroth
first
2^{nd} -order

oracles

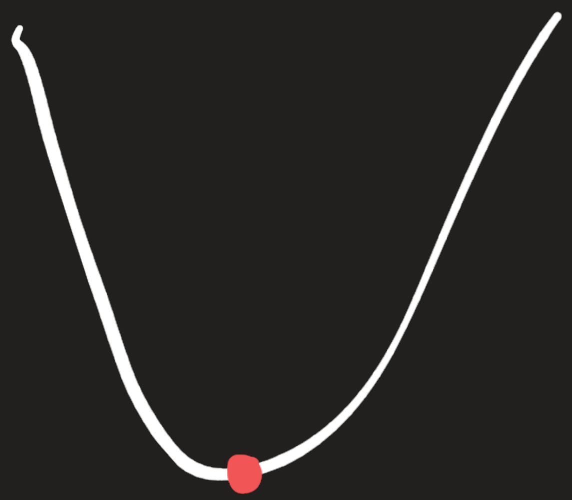
- In general, optimization is NP-Hard so we restrict ourselves to convex optim problems

$$x^* = \underset{x \in C}{\operatorname{argmin}} \; f(x)$$

f – cvx function

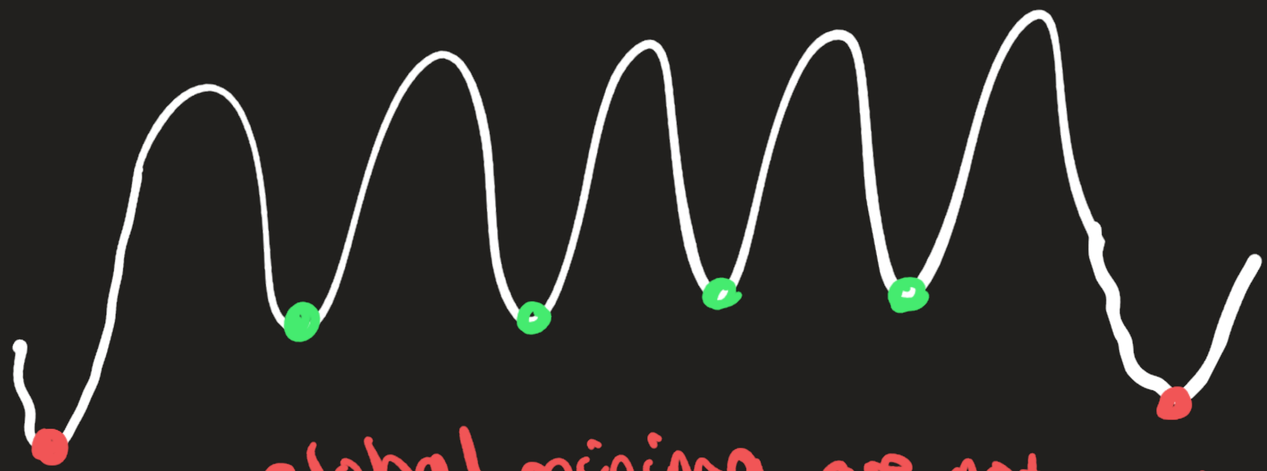
C – cvx set

convex



local minima of convex functions are global minima

nonconvex



global minima are not unique

local minima are not in general, global minima

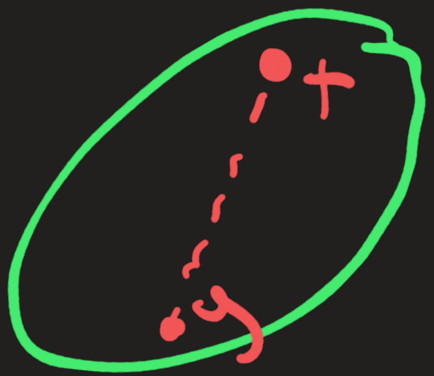
Why convex optim probs?

- very expressive model class
- they are generally tractable
- local minima are global minima so iterative algs are guaranteed to find global minima

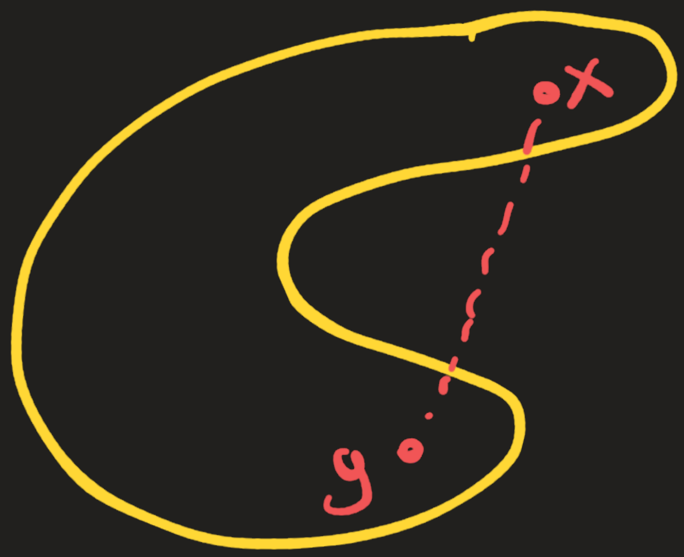
- many ML/statistical estimation probs are naturally convex:
  - Generalized Linear Models (GLMs) fit using MLE give convex optim probs
  - many efficient algs exist, tailored for different settings

Convexity

We say a set C is convex if whenever $x, y \in C$, the line segment $[x,y] = \{ \alpha \in [0,1] : \alpha x + (1-\alpha) y \}$ is contained in C



convex



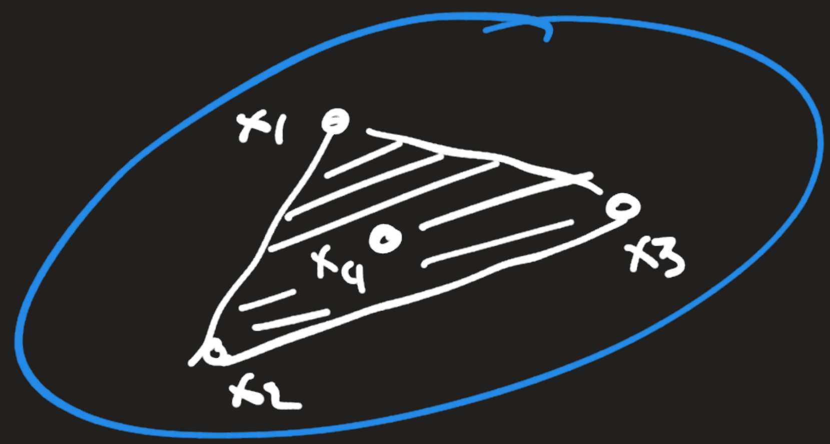
nonconvex

Fact

A set C is convex iff for all $x_1, \dots, x_k \in C$ and any numbers $\theta_1, \dots, \theta_k$ s.t. $\theta_i \geq 0$ for all i and $\sum \theta_i = 1$, the "convex combination"

$$\theta_1 x_1 + \dots + \theta_k x_k \in C$$

Ex



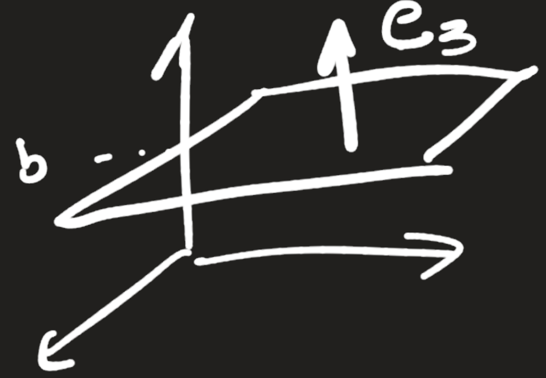
Given $x_1, \dots, x_k$ the set of all convex combinations of these points is called their <u>convex hull</u>. Note this is a convex set

Fact

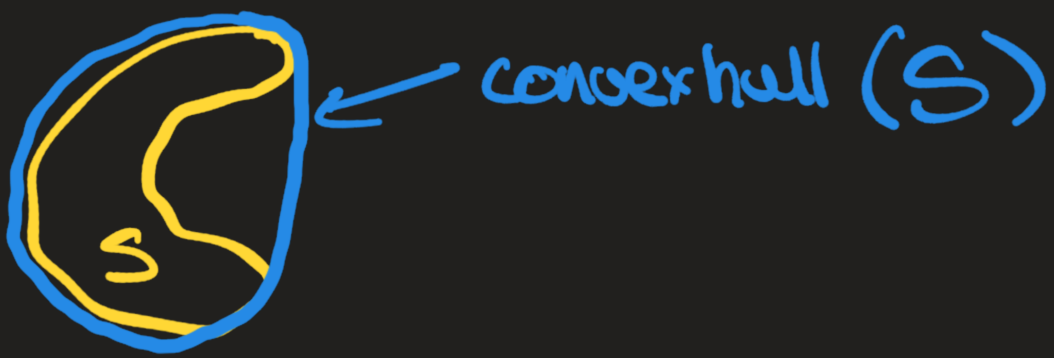
If C is a convex set and X is a r.v. taking values in C, then $\mathbb{E} X \in C$

Examples of cvx sets

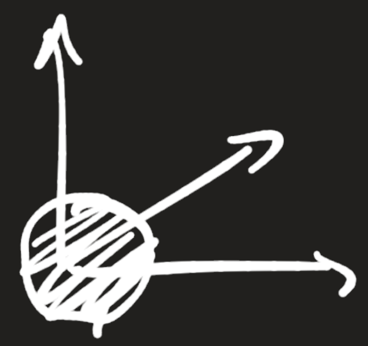
- empty set
- $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}^d$, rays, lines, line segments
- hyperplanes $\{x : a^T x = b\}$



- convex hull of an arb set S, i.e. the smallest convex set containing S



- The Euclidean unit ball $\{x \mid \|x\|_2 \le 1\}$



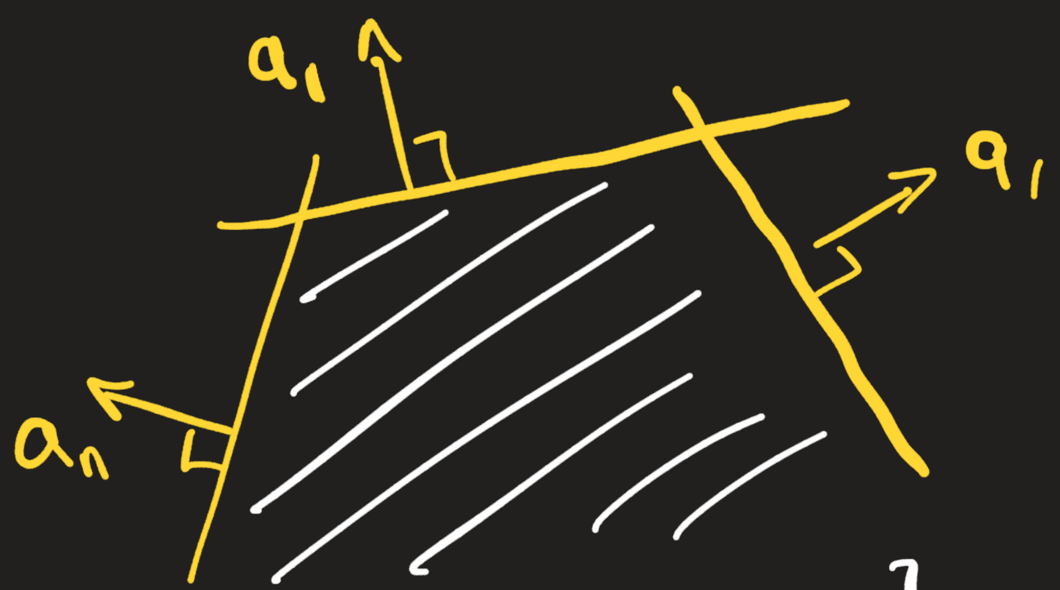
- norm balls $B_{|||\cdot|||}(x, r) = \{y \mid |||x-y||| \le r\}$

$B_{\|\cdot\|_1}(0,1)$

$B_{\|\cdot\|_\infty}(0,1)$

- polyhedrons: sets of solutions to a finite number of linear equalities and inequalities

$$\left\{ x : \begin{array}{c} a_1^T x \le b_1 \\ \vdots \\ a_n^T x \le b_3 \end{array} \right\} = \{ x : Ax \le b \} \qquad A = \begin{bmatrix} a_1^T \\ \vdots \\ a_n^T \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$



$\{x : Ax \le b\}$

- Intersection of arbitrary # of convex set is convex

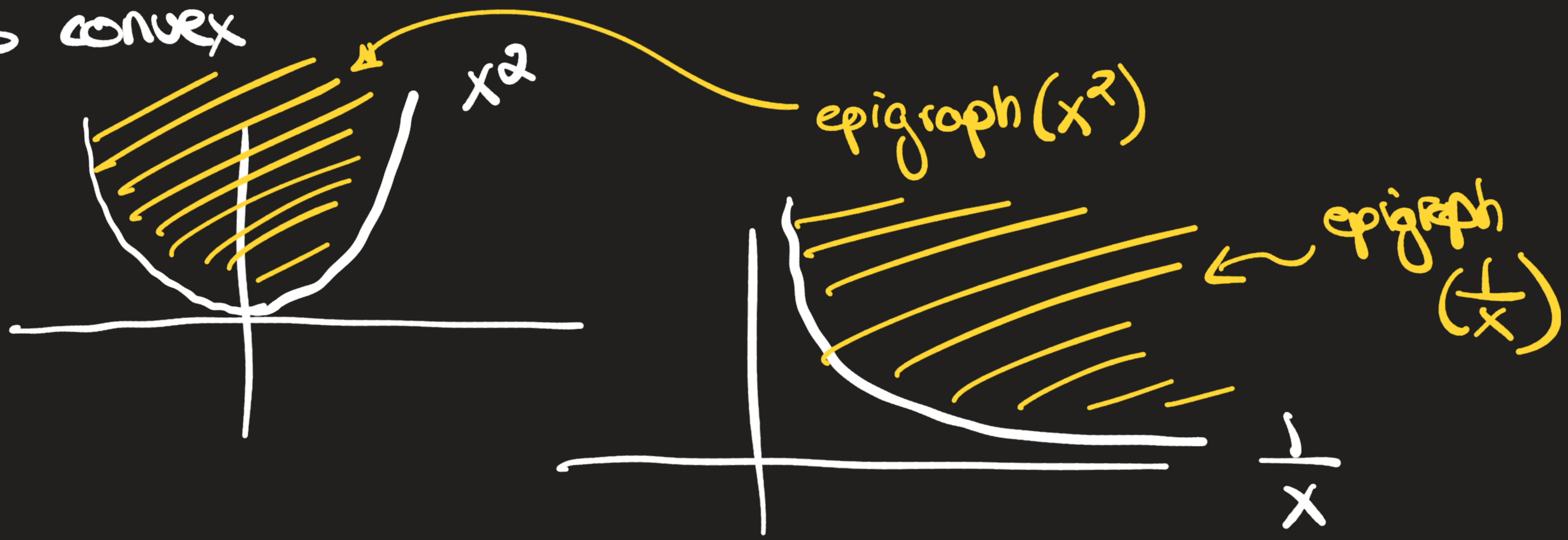
$\bigcap_{i \in I} C_i$ is convex if C_i is convex for all $i \in I$

Convex functions

A function f is convex on its domain, if:

1) $\mathrm{dom}(f)$ is convex

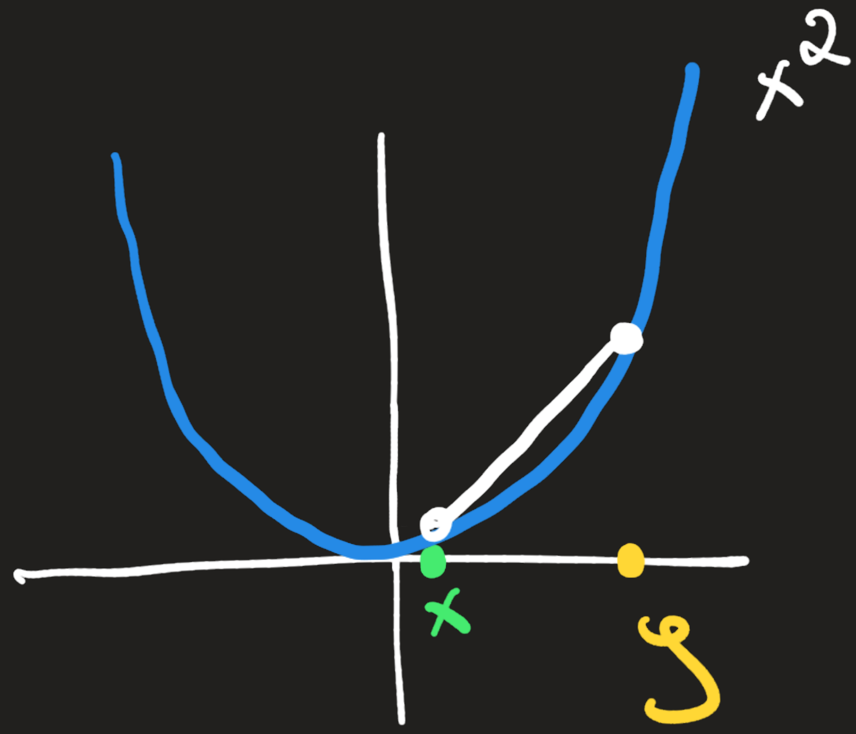
2) $\mathrm{epigraph}(f) = \{(x,t) : t \geq f(x),\ x \in \mathrm{dom}(f)\}$ is convex

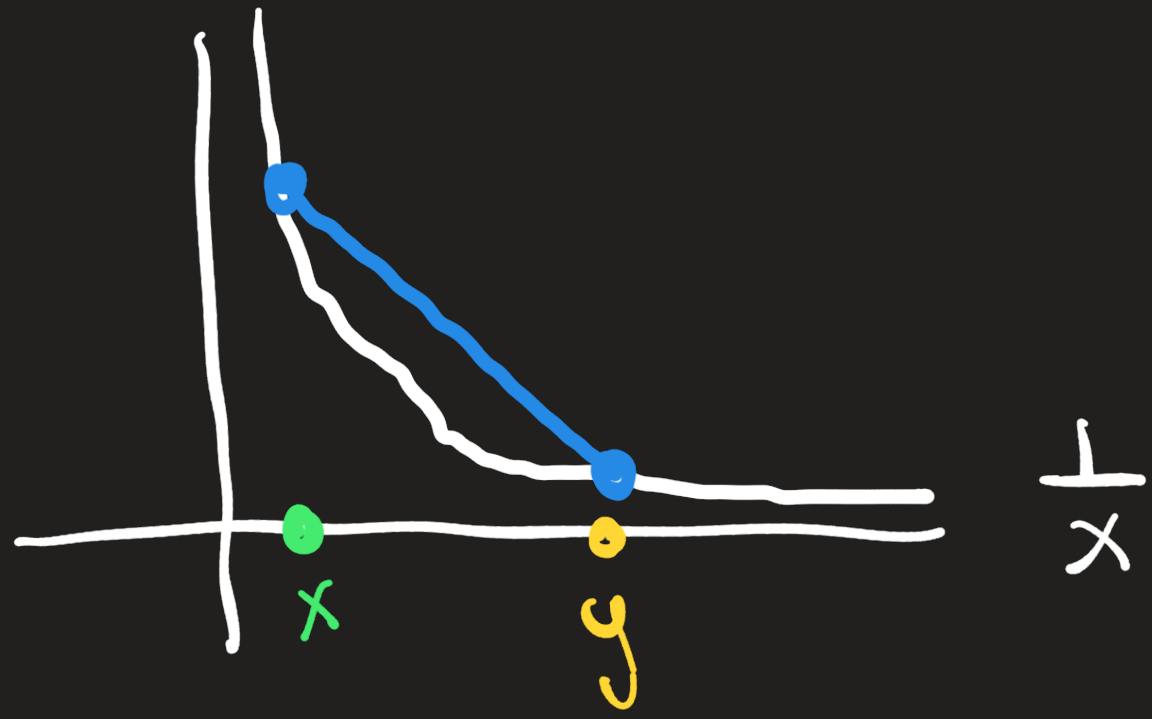


Equivalently, f is convex iff it satisfies:

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

for all $\alpha \in [0,1]$ and $x, y \in C$





f is convex iff it lies below all its secants/chords