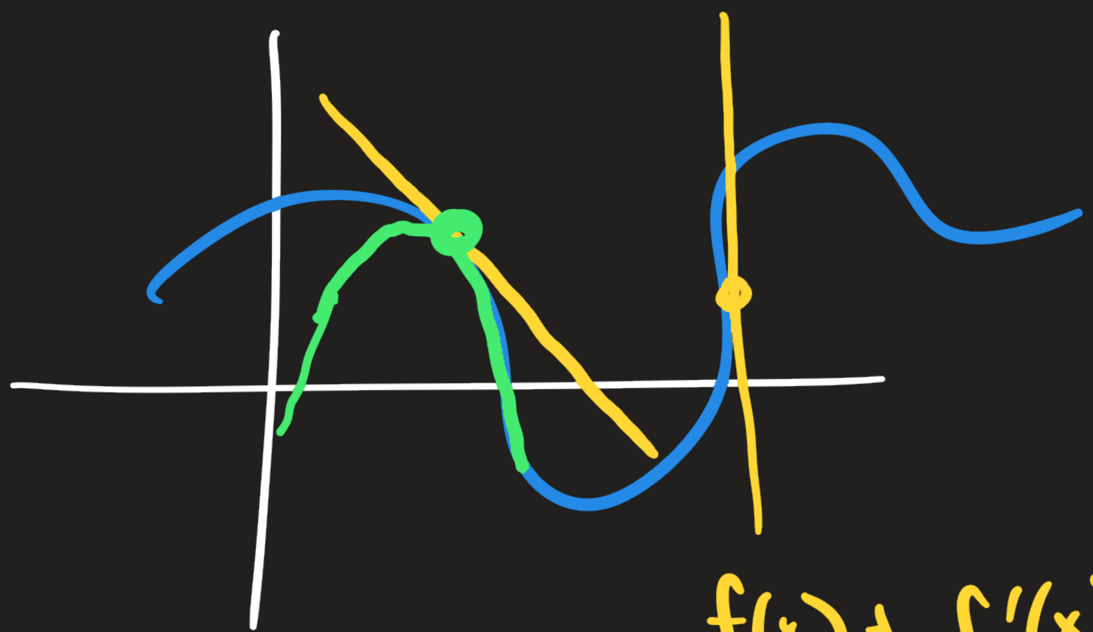
Lecture 6 ML & Opt

- Review of Taylor Series
- Oracle Models of Optimization
- Optimization as Minimization
- Convex Sets, Functions; Convex Optim Probs
- First & 2^{nd} order characterizations of Convexity
- Examples

Taylor Series

Taylor Series is a **local** polynomial **approximation** about a point, x, of an arbitrary smooth function.



First-order TS expansion of f at x

$$f(x)+f'(x)\cdot h \approx f(x+h)$$

$$f(y+h) \approx f(y)+f'(y)\cdot h+\frac{1}{2}f''(y)h^2$$

Second-order TS expansion of f at y

Multivariate TS

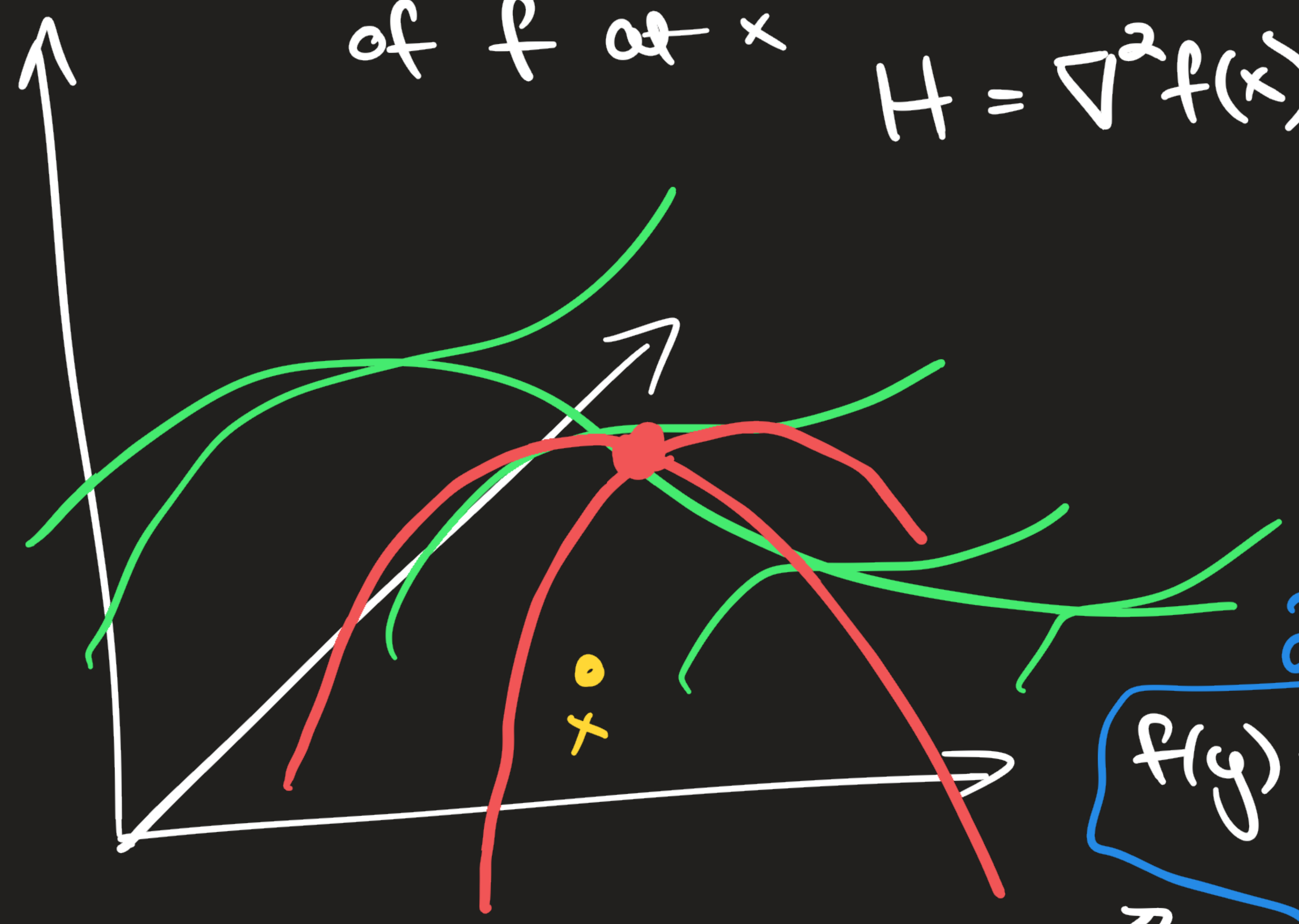
$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle$$

First-order TS

$$\nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2} \right]$$

$$f(y) \approx f(x) + \frac{\partial f(x)}{\partial x_1} \cdot (y_1 - x_1) + \frac{\partial f(x)}{\partial x_2} \cdot (y_2 - x_2)$$

Define the Hessian, matrix of 2^{nd} derivatives of f at x



$$H = \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(x)}{\partial x_2^2} \end{bmatrix}$$

is symmetric

2^{nd}-order TS

$$\boxed{f(y) \approx f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T H (y-x)}$$

$$f(y) \approx f(x) + \frac{\partial f(x)}{\partial x_1}(y_1 - x_1) + \frac{\partial f(x)}{\partial x_2}(y_2 - x_2)$$
$$+ \frac{1}{2}\left[\frac{\partial^2 f(x)}{\partial x_1^2}(y_1 - x_1)^2 + 2\frac{\partial^2 f}{\partial x_1 \partial x_2}(y_1 - x_1)(y_2 - x_2) + \frac{\partial^2 f}{\partial x_2^2}(y_2 - x_2)^2\right]$$

In general, $f: \mathbb{R}^d \to \mathbb{R}$

$$f(y) \approx f(x) + \langle \nabla f(x), y-x \rangle + \frac{1}{2} \langle y-x, H(y-x) \rangle$$

where

$H \in \mathbb{R}^{d \times d}$ (symmetric) and $H_{ij} = \dfrac{\partial^2 f(x)}{\partial x_i \partial x_j}$

Oracle Models of Optimization

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

$x_0 \leftarrow$ initial guess of the solution

compute $f(x_0)$ and/or $\nabla f(x_0)$ and/or $\nabla^2 f(x_0)$

form a local model

move to a new point based on the local model

repeat until "convergence"

Convex Optim

- We've seen we reduce ML to optim probs (RERM and MLE or MAP)

- We also know that we only want our numerical soln f_T to satisfy

$$R_n(f_T) \leq R_n(\hat{f}_{\mathcal{F}}) + O(\varepsilon)$$

so we only care about solving optim probs to ε-suboptimality

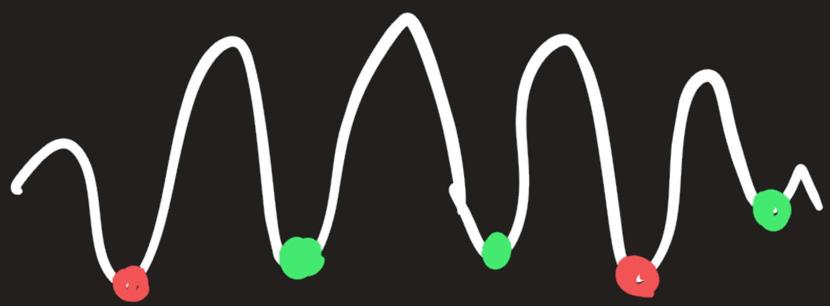
- In general optim is NP-Hard so we restrict ourselves to <u>convex optim problems</u> (for now)

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

f - cvx function

C - cvx set

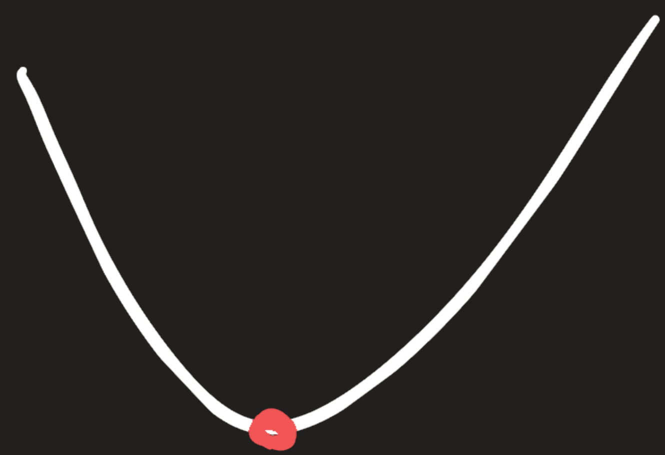
<u>nonconvex</u>



global minima are not unique

local minima are <u>not</u> always global minima

<u>convex</u>



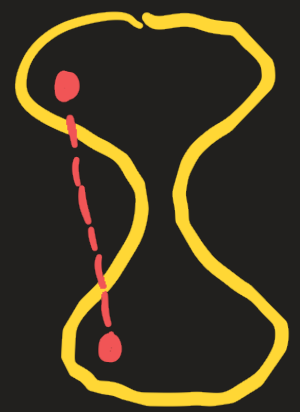
local minima <u>are</u> global minima

- Why restrict ourselves (for now) to 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 probs are naturally convex:
    - GLMs fit using MLE give convex optim probs
    - many efficient alg 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

C – cvx set

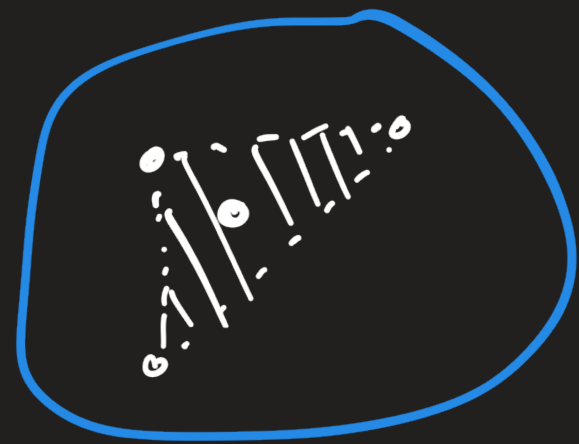


not a cvx set

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. $\theta_k \geq 0$ for all i and $\sum_i \theta_i = 1$, the convex combination

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

Ex

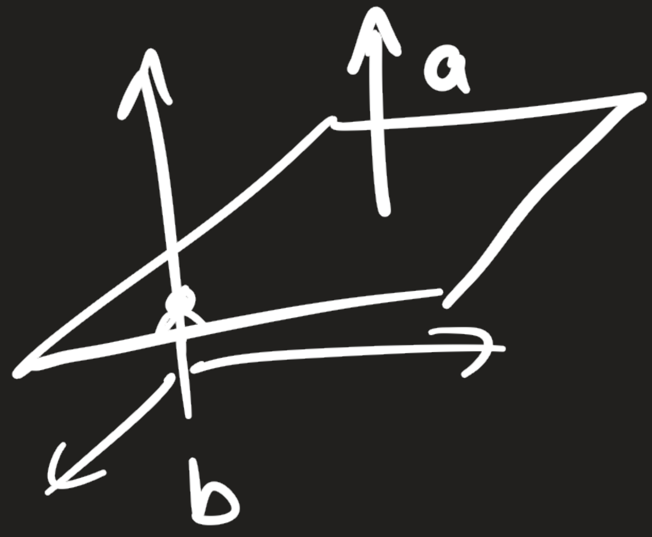


Given $x_1, \dots, x_k$ the set of all convex combinations of these pts is called their <u>convex hull</u>

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

Ex of convex sets

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



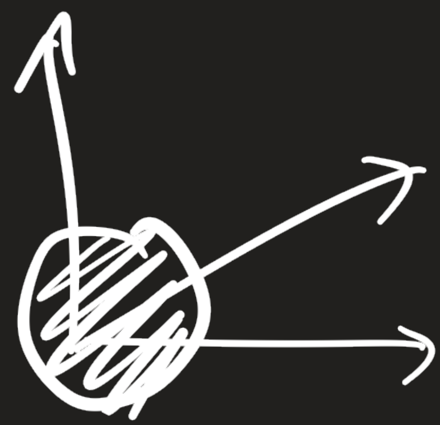
- positive orthant

  $x \geq 0$

NNLS : $\min_{x \geq 0} \|Ax-b\|_2^2$

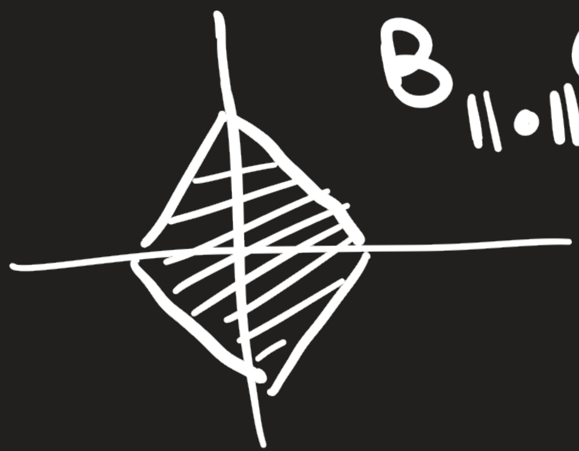
- convex hull of an arb. set S, i.e. the smallest convex set containing S

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



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

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



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

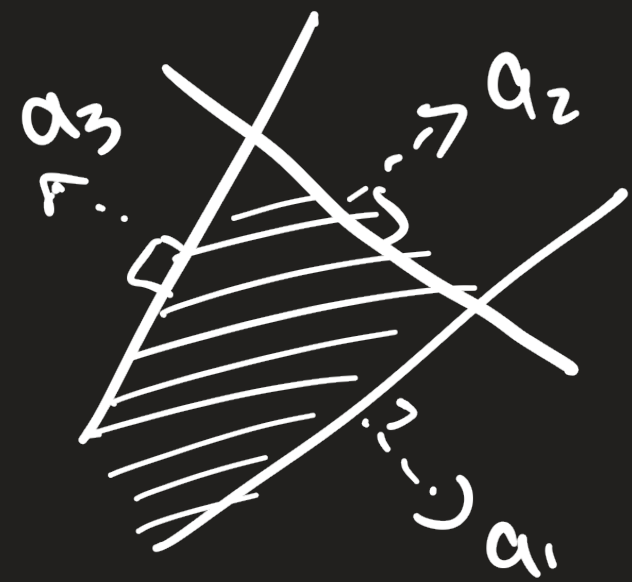
$B_{\|\cdot\|_2}(0,1)$

$$\|x\|_\infty = \max_i |x_i| = \lim_{p \to \infty} \|x\|_p$$

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

$$\left\{ x : \begin{array}{l} a_1^T x \le b_1 \\ a_2^T x \le b_2 \\ a_3^T x \le b_3 \\ \vdots \end{array} \right\} = \{ x : Ax \le b \}$$

$$A = \begin{bmatrix} a_1^T \\ \vdots \\ a_n^T \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$



- Intersection of arbitrary # of convex sets is convex

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

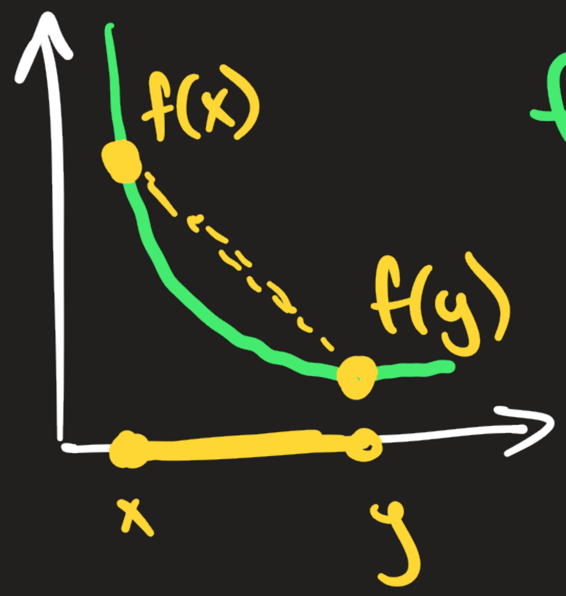
Convex function

A function f defined on a convex set C

is called 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 $x, y \in C$



$f(x) = \frac{1}{x}$ is convex on $\mathbb{R}_{++}$

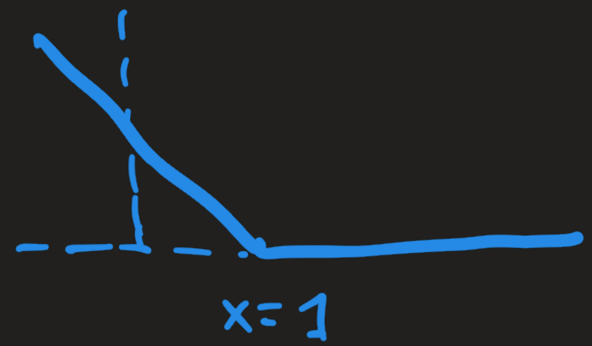
Graphically: a function is convex if it lies under its chords

Strict convexity

$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y)$ is satisfied for all $x, y \in \mathrm{dom}(f)$ and $\alpha \in (0,1)$

$f = 0$

not strictly convex



$(1-x)_+$

x^2

strictly cvx!

Most important consequence of convexity

If a function f is convex on a convex set C, then any local minimum of f on C is a global minimum.

Fact

If f is strictly convex on C then the minimizer of f on C (if any exists) is unique.