Fall 2022 | Lecture 7

Convex Optimization

Ariel Procaccia | Harvard University

OPTIMIZATION PROBLEMS

- Casting AI problems as optimization problems has been one of the primary AI trends in the 21st century
- A seemingly remarkable fact:

	Discrete optimization	Continuous optimization
Variable type	Discrete	Continuous
# solutions	Finite	Infinite
Complexity	Exponential	Polynomial

FORMAL DEFINITION

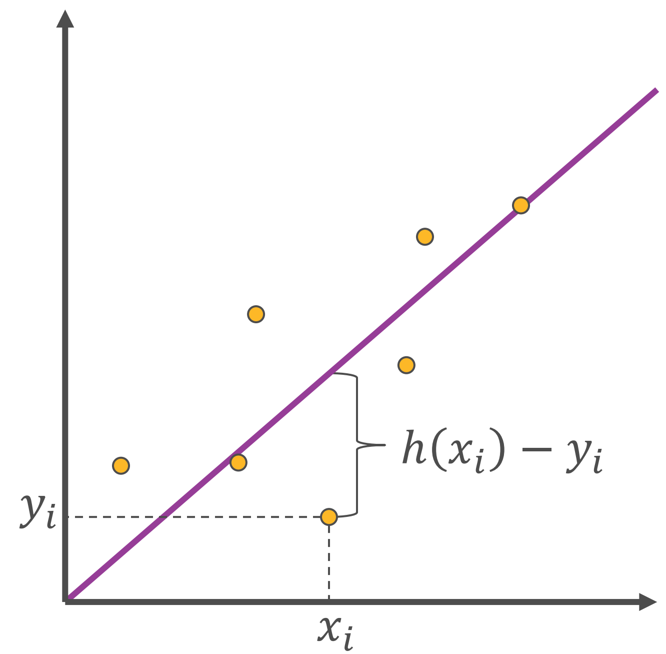
- Interested in problems of the form

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$
$$\text{such that } \boldsymbol{x} \in \mathcal{F}$$

  where:
  - $\boldsymbol{x} \in \mathbb{R}^n$ is the optimization variable
  - $\mathcal{F} \subseteq \mathbb{R}^n$ is the feasible set
- $\boldsymbol{x}^* \in \mathbb{R}^n$ is an optimal solution if $\boldsymbol{x}^* \in \mathcal{F}$ and $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{F}$

EXAMPLE: LEAST-SQUARES FITTING

- Given (x_i, y_i) for $i = 1, \ldots, m$, find $h(x) = ax + b$ that optimizes
$$\min_{a,b} \sum_{i=1}^{m} (ax_i + b - y_i)^2$$
(*a* is slope, *b* is intercept)



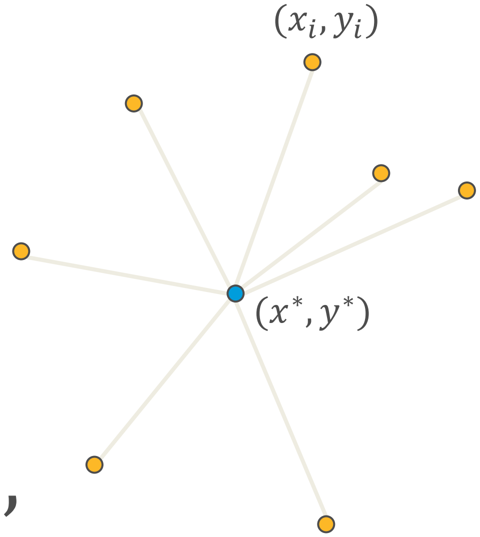
EXAMPLE: WEBER POINT

- Given (x_i, y_i) for $i = 1, \ldots, m$, find the point (x^*, y^*) that minimizes the sum of Euclidean distances:

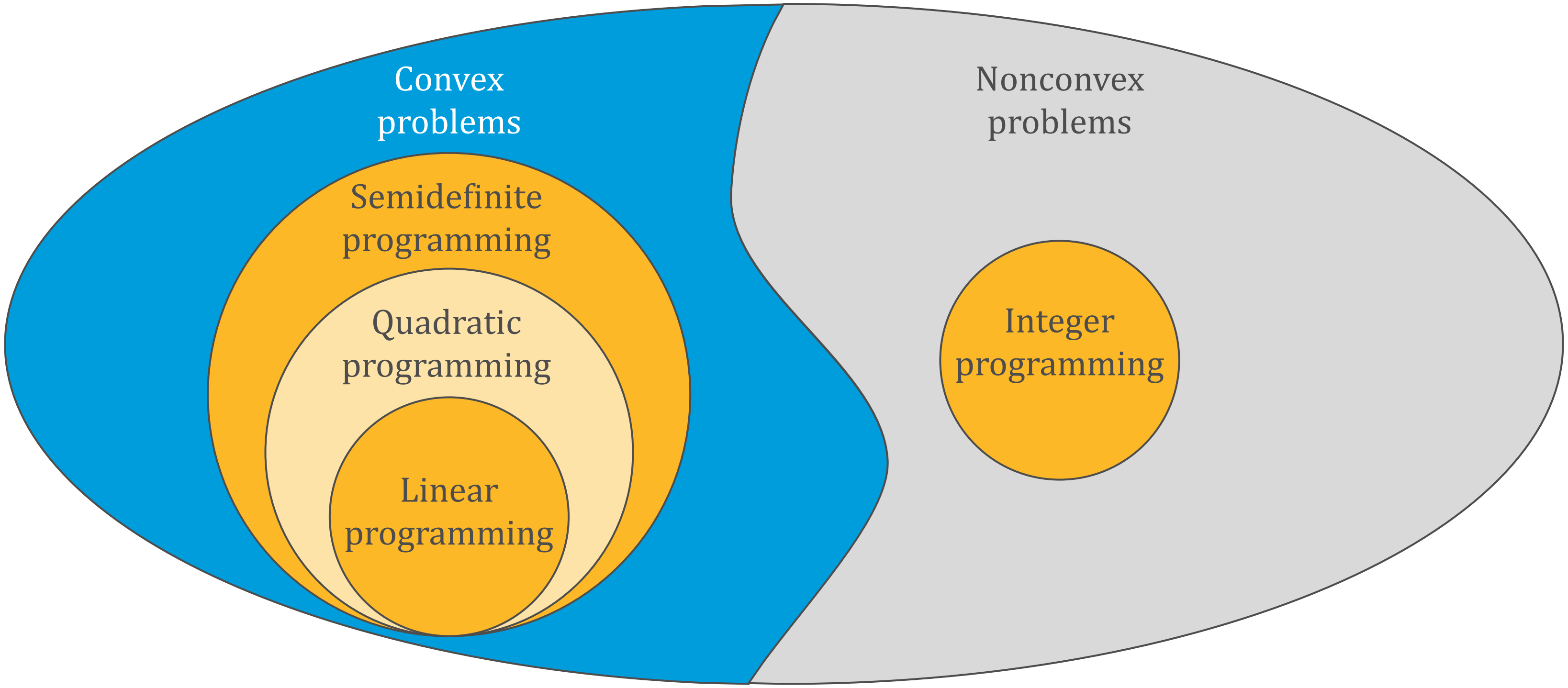
$$\min_{x^*, y^*} \sum_{i=1}^{m} \sqrt{(x^* - x_i)^2 + (y^* - y_i)^2}$$

- We might impose constraints, e.g., require that

$$a \leq x^* \leq b, c \leq y^* \leq d$$



THE OPTIMIZATION UNIVERSE



CONVEX OPTIMIZATION

A convex optimization problem is a specialization of a general optimization problem

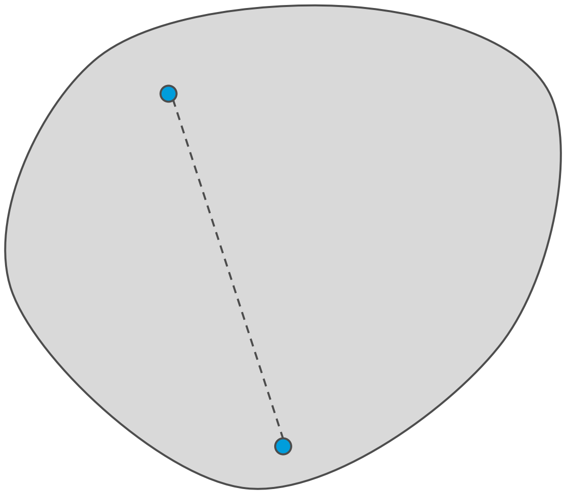
$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$\text{such that } \boldsymbol{x} \in \mathcal{F}$$

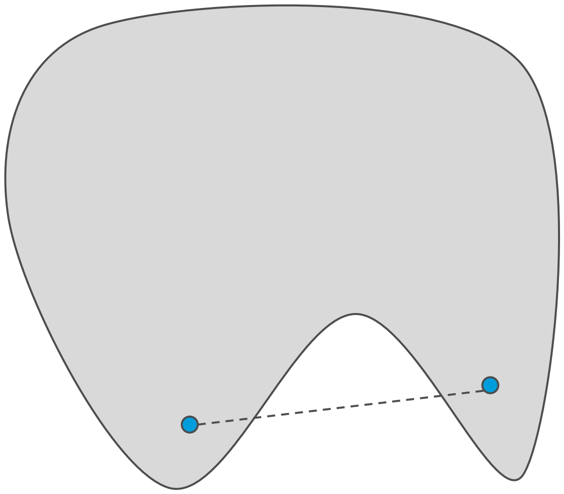
where the target function $f: \mathbb{R}^n \to \mathbb{R}$ is a convex function, and the feasible region $\mathcal{F}$ is a convex set

CONVEX SETS

A set $\mathcal{F} \subseteq \mathbb{R}^n$ is convex if for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{F}$ and $\theta \in [0,1]$, $\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y} \in \mathcal{F}$



Convex set

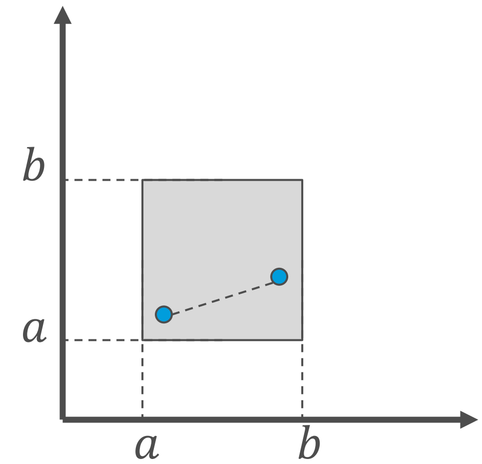


Nonconvex set

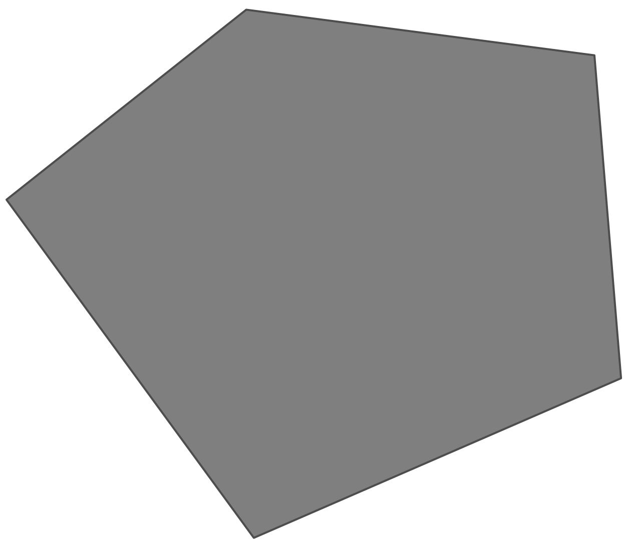
EXAMPLES OF CONVEX SETS

- $\mathcal{F} = \{\boldsymbol{x} \in \mathbb{R}^n : \forall i = 1, \ldots, n,\ a \leq x_i \leq b\}$
- Proof:
  - Let $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{F}$, and $\theta \in [0,1]$
  - For all $i = 1, \ldots, n$, $a \leq x_i$ and $a \leq y_i$, so $\theta x_i + (1-\theta) y_i \geq$

    $$\theta a + (1-\theta) a = a$$
  - Similarly, $\theta x_i + (1-\theta) y_i \leq b$
  - Therefore $\theta \boldsymbol{x} + (1-\theta) \boldsymbol{y} \in \mathcal{F}$ ■



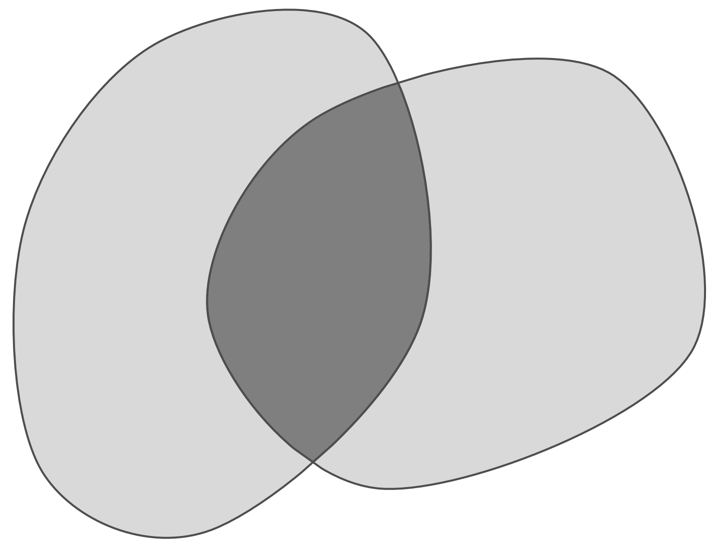
EXAMPLES OF CONVEX SETS



Linear inequalities

$\mathcal{F} = \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} \leq \boldsymbol{b}\}$

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



Intersection of convex sets

$\mathcal{F} = \bigcap_{i=1}^{m} C_i$

$C_1, \ldots, C_m$ are convex

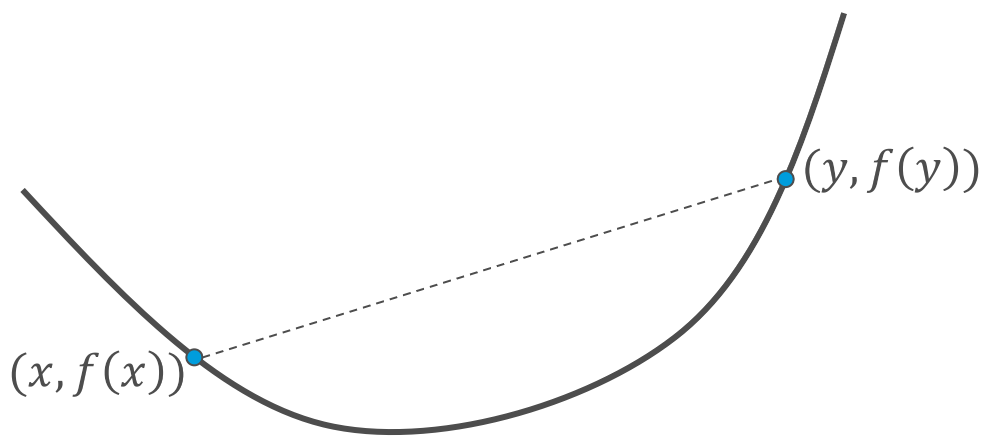
EXAMPLES OF CONVEX SETS

- Poll 1: Which of the following sets are convex?
  1. $\mathcal{F} = \bigcup_{i=1}^{m} C_i$ where $C_1, \dots, C_m$ are convex
  2. $\mathcal{F} = \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} = \boldsymbol{b}\}$ where $A \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$
  3. Both
  4. Neither

CONVEX FUNCTIONS

- A function $f: \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $\theta \in [0,1]$,

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y})$$



- For functions $f: \mathbb{R} \to \mathbb{R}$ that are twice differentiable, equivalent to $f''(x) \geq 0$ for all $x \in \mathbb{R}$
- f is concave if and only if $-f$ is convex

EXAMPLES OF CONVEX FUNCTIONS

- Exponential: $f(x) = e^{ax}$
  - $f''(x) = a^2 e^{ax} \geq 0$ for all $x \in \mathbb{R}$
- Euclidean norm: $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^{n} (x_i)^2}$
  - $$\begin{aligned} \|\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}\|_2 &\leq \|\theta \boldsymbol{x}\|_2 + \|(1-\theta)\boldsymbol{y}\|_2 \\ &= \theta\|\boldsymbol{x}\|_2 + (1-\theta)\|\boldsymbol{y}\|_2 \end{aligned}$$

EXAMPLES OF CONVEX FUNCTIONS

- Poll 2: Which functions are convex?
  1. $f(\boldsymbol{x}) = \sum_{i=1}^{m} a_i f_i(\boldsymbol{x})$ where f_i is convex and $a_i \geq 0$ for $i = 1, \dots, m$
  2. $g(\boldsymbol{x}) = \sqrt{\sum_{i=1}^{n} x_i}$ for $\boldsymbol{x} \geq 0$
  3. Both
  4. Neither

EXAMPLES OF CONVEX PROBLEMS

- Weber point in n dimensions:

$$\min_{\boldsymbol{x}^*} \sum_{i=1}^{m} \left\| \boldsymbol{x}^* - \boldsymbol{x}^{(i)} \right\|_2$$

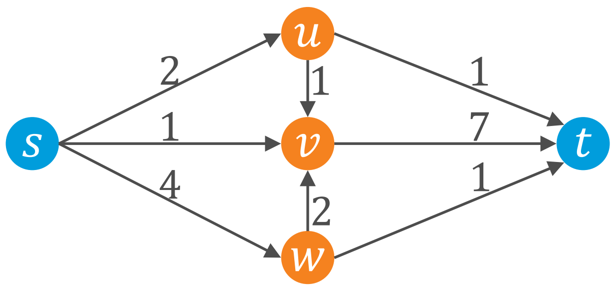
  where $\boldsymbol{x}^* \in \mathbb{R}^n$ is the optimization variable and $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}$ are the problem data

- Linear programming:

$$\begin{aligned} \min_{\boldsymbol{x}} \ & \boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t. } & A\boldsymbol{x} = \boldsymbol{a} \\ & B\boldsymbol{x} \leq \boldsymbol{b} \end{aligned}$$

  where $\boldsymbol{x} \in \mathbb{R}^n$ is the optimization variable, and $\boldsymbol{c} \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, \boldsymbol{a} \in \mathbb{R}^m, B \in \mathbb{R}^{k \times n}, \boldsymbol{b} \in \mathbb{R}^k$ are the problem data

LINEAR PROGRAMMING: EXAMPLE



- In the max flow problem, we are given a directed graph $G = (V, E)$ with a source s and a sink t, and a capacity α_{xy} for each $(x, y) \in E$
- A flow is a function $f: E \to \mathbb{R}^+$ that satisfies $f_{xy} \leq \alpha_{xy}$ for all $(x, y) \in E$, and for all $x \in V \setminus \{s, t\}$, $\sum_{(y,x)\in E} f_{yx} = \sum_{(x,z)\in E} f_{xz}$
- The value of a flow is $\sum_{(s,x)\in E} f_{sx}$
- Poll 3: What is the max flow in the above example?

LINEAR PROGRAMMING: EXAMPLE

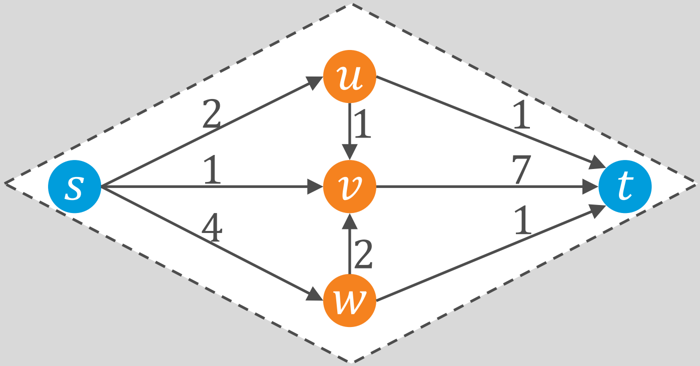
$$\min_{\boldsymbol{x}} \boldsymbol{c}^T\boldsymbol{x}$$
$$\text{s.t. } A\boldsymbol{x} = \boldsymbol{a}$$
$$B\boldsymbol{x} \leq \boldsymbol{b}$$

How does the canonical LP form fit with the max flow example?

$$\max f_{su} + f_{sv} + f_{sw}$$

s.t.

$f_{s,u} \leq 2$	$f_{s,u} \geq 0$	$f_{s,u} = f_{uv} + f_{ut}$
$f_{s,v} \leq 1$	$f_{s,v} \geq 0$	$f_{s,v} + f_{uv} + f_{wv} = f_{vt}$
$f_{s,w} \leq 4$	$f_{s,w} \geq 0$	$f_{sw} = f_{wv} + f_{wt}$
$f_{uv} \leq 1$	$f_{uv} \geq 0$	
$f_{wv} \leq 2$	$f_{wv} \geq 0$	
$f_{u,t} \leq 1$	$f_{u,t} \geq 0$	
$f_{v,t} \leq 7$	$f_{v,t} \geq 0$	
$f_{w,t} \leq 1$	$f_{w,t} \geq 0$	

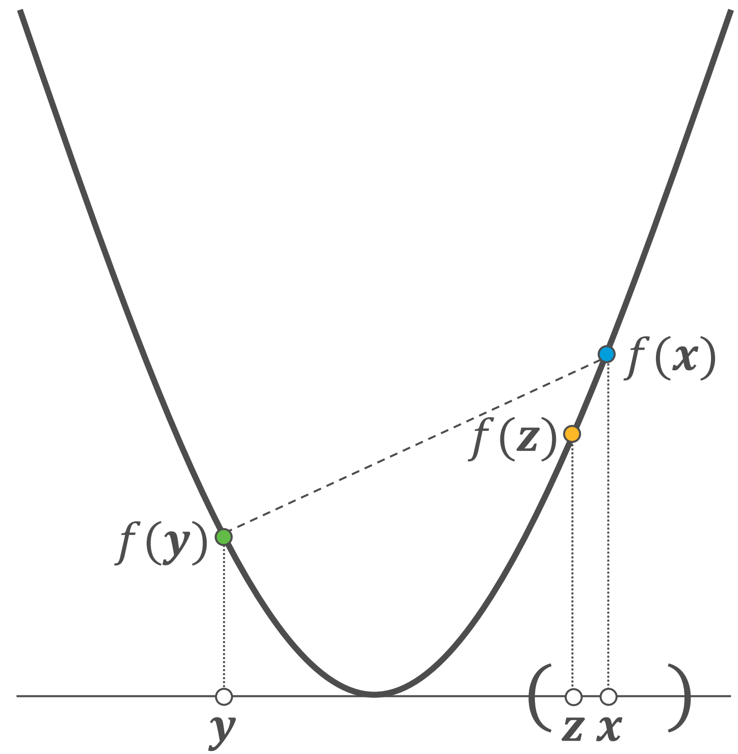


GLOBAL AND LOCAL OPTIMALITY

- A point $\boldsymbol{x} \in \mathbb{R}^n$ is globally optimal if $\boldsymbol{x} \in \mathcal{F}$ and for all $\boldsymbol{y} \in \mathcal{F}$, $f(\boldsymbol{x}) \le f(\boldsymbol{y})$
- A point $\boldsymbol{x} \in \mathbb{R}^n$ is locally optimal if $\boldsymbol{x} \in \mathcal{F}$ and there exists $R > 0$ such that for all $\boldsymbol{y} \in \mathcal{F}$ with $\|\boldsymbol{x} - \boldsymbol{y}\|_2 \le R$, $f(\boldsymbol{x}) \le f(\boldsymbol{y})$
- Theorem: For a convex optimization problem, all locally optimal points are globally optimal

PROOF OF THEOREM

- Suppose $\boldsymbol{x}$ is locally optimal for some R, but not globally optimal
- There is $\boldsymbol{y} \in \mathcal{F}$ such that $f(\boldsymbol{y}) < f(\boldsymbol{x})$
- Define
$$\boldsymbol{z} = (1-\theta)\boldsymbol{x} + \theta\boldsymbol{y}$$
for $\theta = \frac{R}{2\|\boldsymbol{x}-\boldsymbol{y}\|_2}$



PROOF OF THEOREM

- Then:
  - $\boldsymbol{z}$ is feasible (can assume $\|x - y\|_2 > R$)
  - $$\begin{aligned} f(\boldsymbol{z}) &= f\big((1-\theta)\boldsymbol{x} + \theta\boldsymbol{y}\big) \\ &\leq (1-\theta)f(\boldsymbol{x}) + \theta f(\boldsymbol{y}) \\ &< (1-\theta)f(\boldsymbol{x}) + \theta f(\boldsymbol{x}) \\ &= f(\boldsymbol{x}) \end{aligned}$$
  - $\|\boldsymbol{x} - \boldsymbol{z}\|_2 = \left\| \frac{R}{2\|x-y\|_2}(\boldsymbol{x} - \boldsymbol{y}) \right\|_2 = \frac{R}{2} < R$
- Therefore, $\boldsymbol{x}$ is not locally optimal, contradicting our assumption ■

SOLVING CONVEX PROBLEMS

- Convex optimization problems are computationally "easy" to solve
- The gradient descent algorithm follows the objective function downwards until it reaches a global minimum
- It cannot get stuck at a local minimum due to the theorem