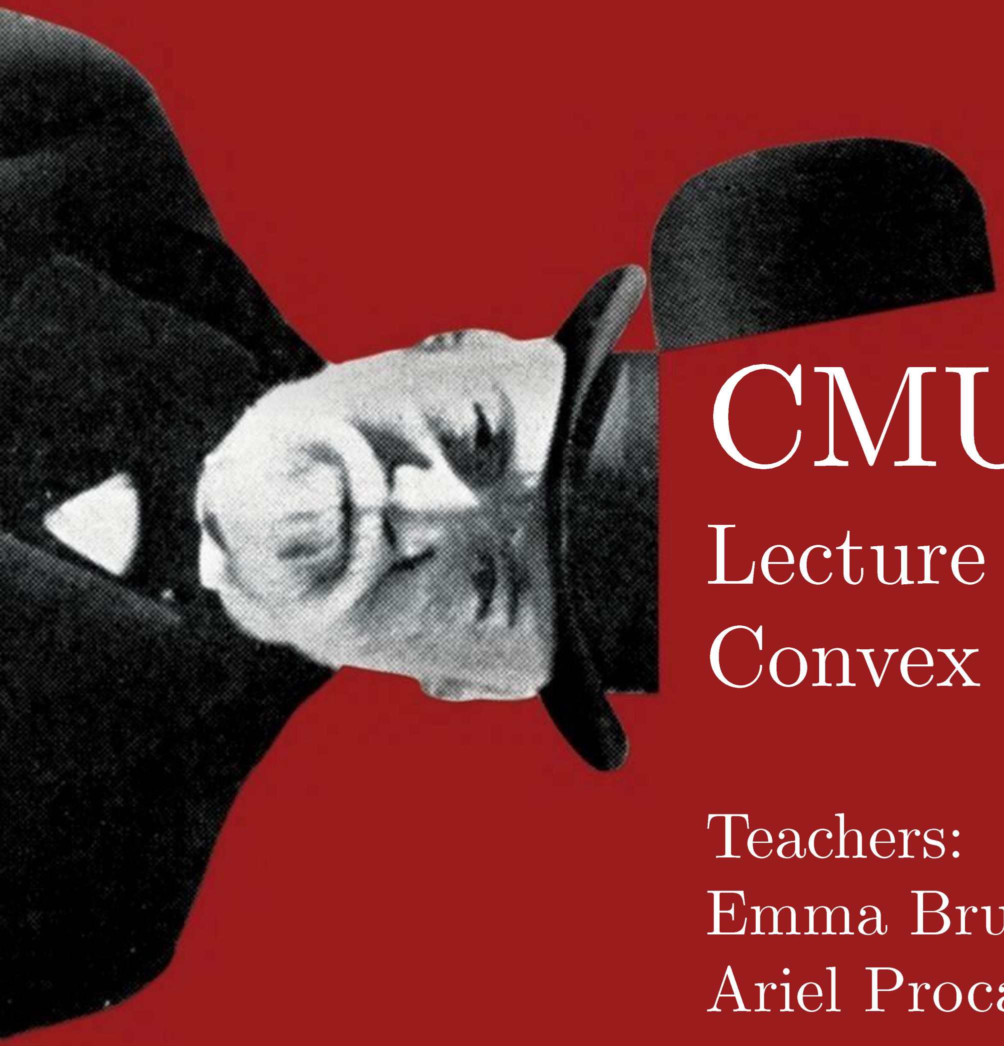


CMU 15-781

Lecture 13: Convex Optimization

Teachers:
Emma Brunskill
Ariel Procaccia (this time)

Optimization problems

- Casting AI problems as optimization problems has been one of the primary trends of the last 15 years
- 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

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

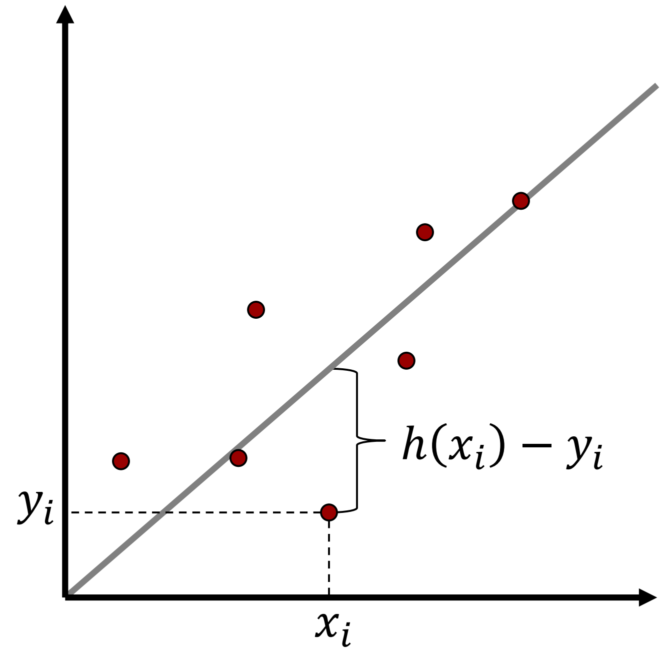
where:

  - $\boldsymbol{x} \in \mathbb{R}^n$ 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

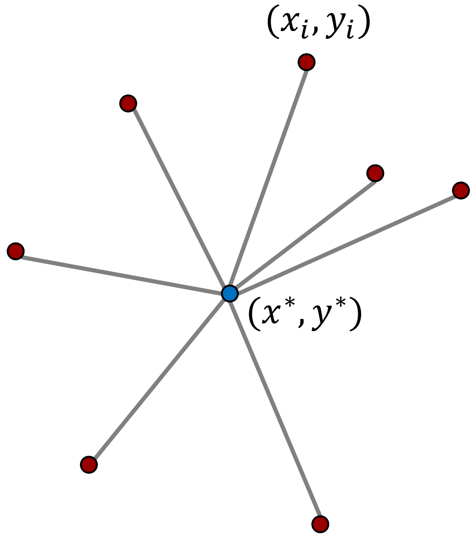
- Given (x_i, y_i) for $i = 1, \dots, m$, find $h(x) = ax + b$ that optimizes
$$\min_{a,b} \sum_{i=1}^{m} (ax_i + b - y_i)^2$$
(a is slope, b is intercept)



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:

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



- Many modifications, e.g., might want $a \le x^* \le b, c \le y^* \le d$

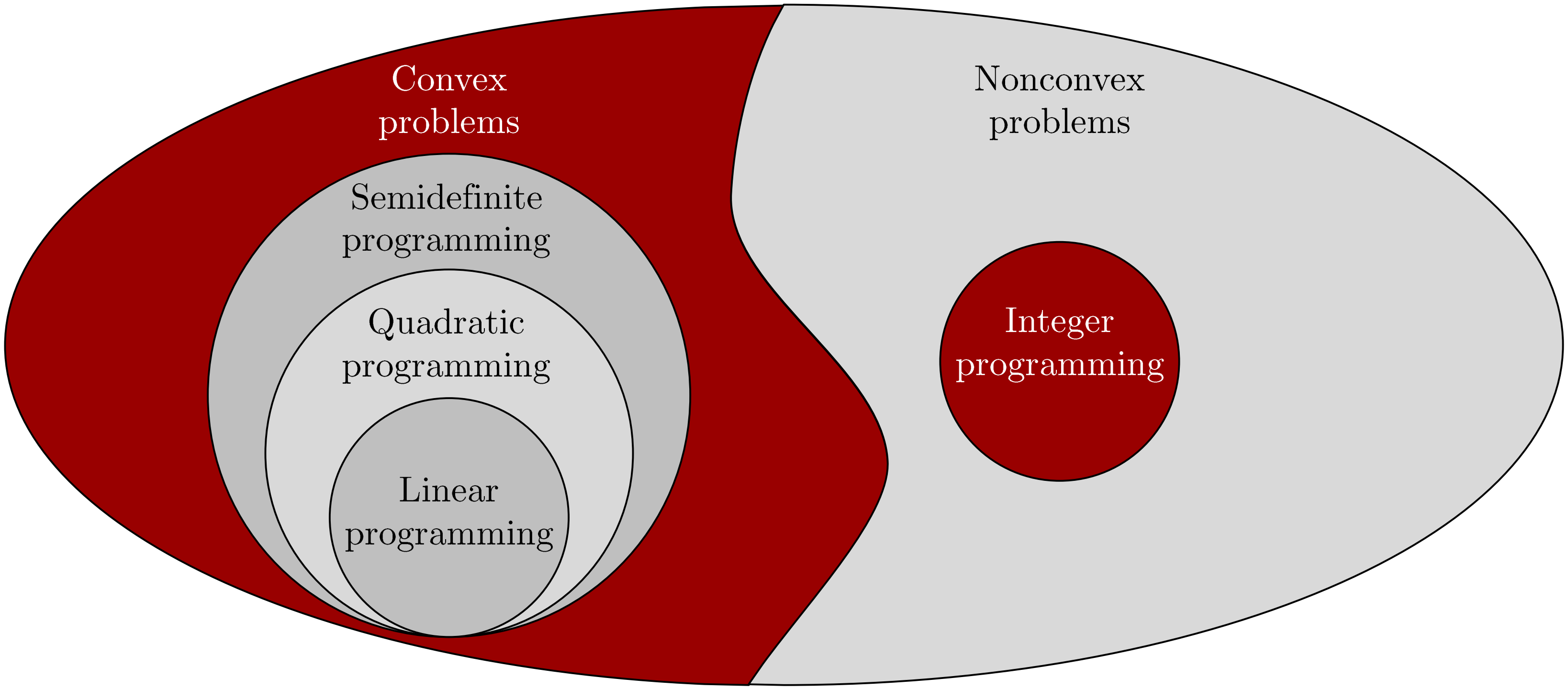
Machine learning

- Many machine learning problems can be described as minimizing a loss function

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^n} \sum_{i=1}^{m} L\left(\sum_{j=1}^{n} \alpha_j x_j^{(i)}, y^{(i)}\right)$$

  - $\boldsymbol{x}^{(i)} \in \mathbb{R}^n$ are input features
  - $y^{(i)} \in \mathbb{R}$ (regression) or $y^{(i)} \in \{0,1\}$ (classification) are outputs
  - $\boldsymbol{\alpha} \in \mathbb{R}^n$ are model parameters

The optimization universe



Convex optimization

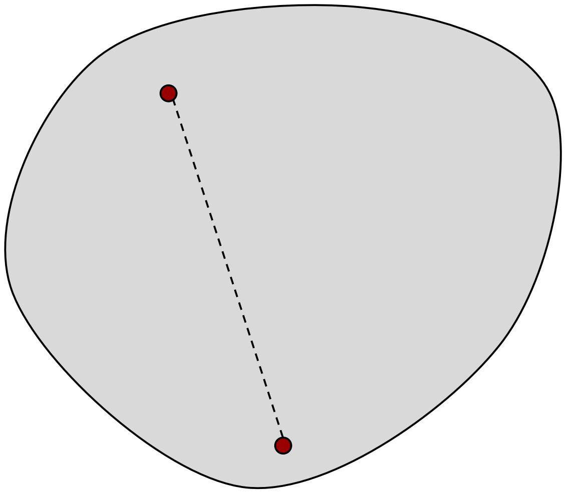
- A convex optimization problem is a specialization of a general optimization problem

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

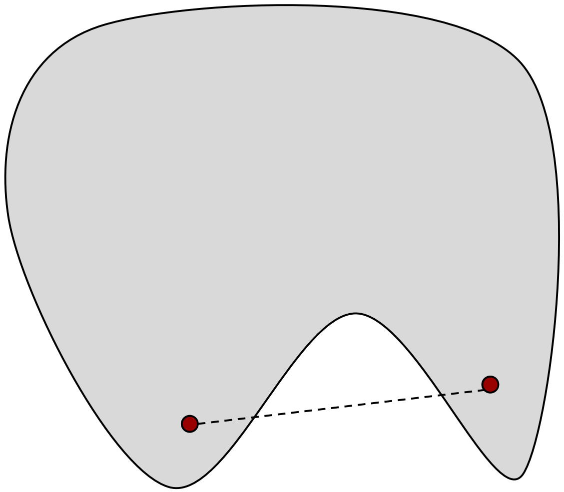
where the target function $f: \mathbb{R}^n \to \mathbb{R}$ is a convex function, and the feasible region $\mathcal{F}$ is a convex set

Convex sets

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



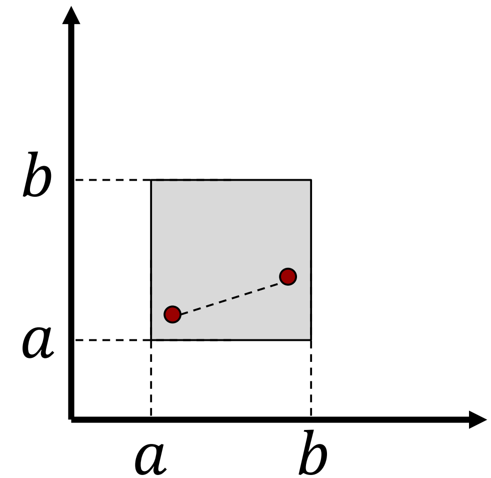
Convex set



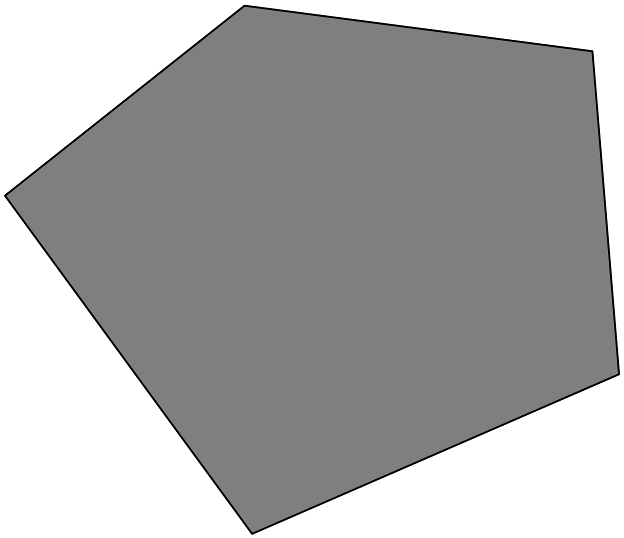
Nonconvex set

Examples of convex sets

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



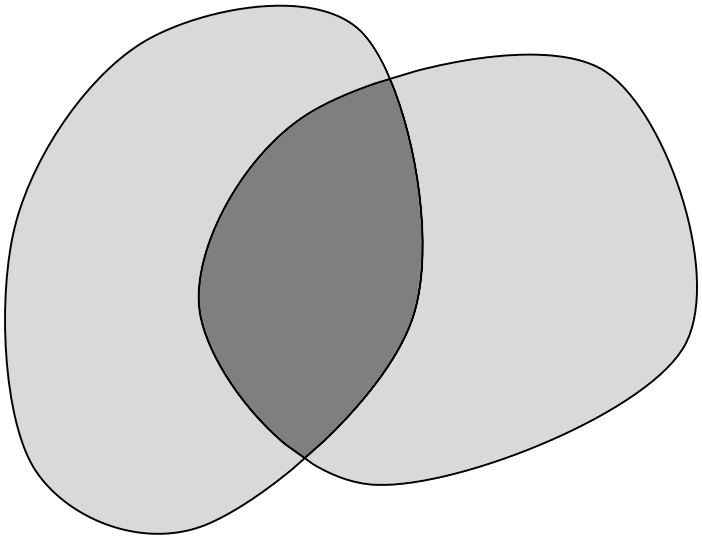
Examples of convex sets



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

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

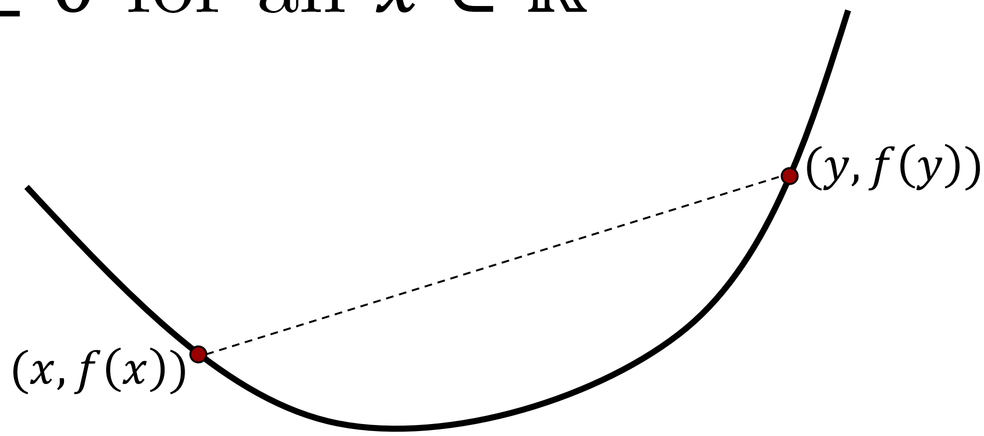
$C_1, \ldots, C_m$ are convex

Examples of convex sets

- Poll 1: Which of the following sets are convex?
  1. $\mathcal{F} = \cup_{i=1}^{m} C_i$ where $C_1, \ldots, C_m$ 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 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 $f: \mathbb{R} \to \mathbb{R}$, equivalent to $f''(x) \geq 0$ for all $x \in \mathbb{R}$



Examples of convex problems

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

Examples of convex problems

- 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 $i = 1, \ldots, m$
  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 optimization variable and $\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(m)}$ are problem data
- A convex optimization problem (why?)

Examples of convex problems

- Linear programming:

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

  where $\boldsymbol{x} \in \mathbb{R}^n$ is 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 problem data

- A convex optimization problem (why?)

Global and local optimality

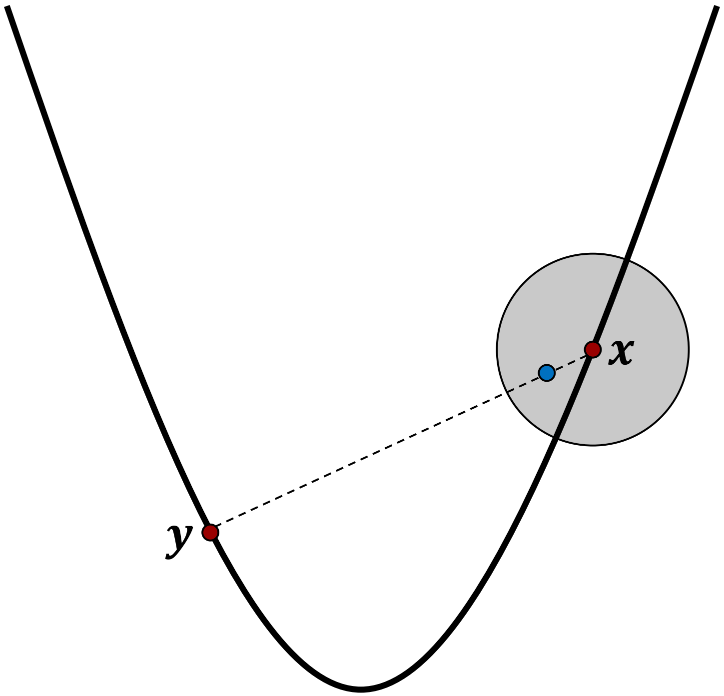
- A point $\boldsymbol{x} \in \mathbb{R}^n$ is globally optimal if $\boldsymbol{x} \in \mathcal{F}$ and for all $\boldsymbol{y} \in \mathcal{F}$, $f(\boldsymbol{x}) \leq f(\boldsymbol{y})$
- A point $\boldsymbol{x} \in \mathbb{R}^n$ is locally optimal if $\boldsymbol{x} \in \mathcal{F}$ and there exists $R > 0$ such that for all $\boldsymbol{y} \in \mathcal{F}$ with $\|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq R$, $f(\boldsymbol{x}) \leq f(\boldsymbol{y})$
- 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 $\boldsymbol{y}$ such that $f(\boldsymbol{y}) < f(\boldsymbol{x})$
- Define

$$\boldsymbol{z} = \theta \boldsymbol{x} + (1 - \theta)\boldsymbol{y}$$

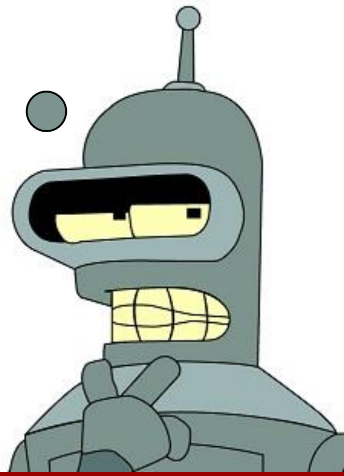
for $\theta = 1 - \frac{R}{2\|\boldsymbol{x} - \boldsymbol{y}\|_2}$



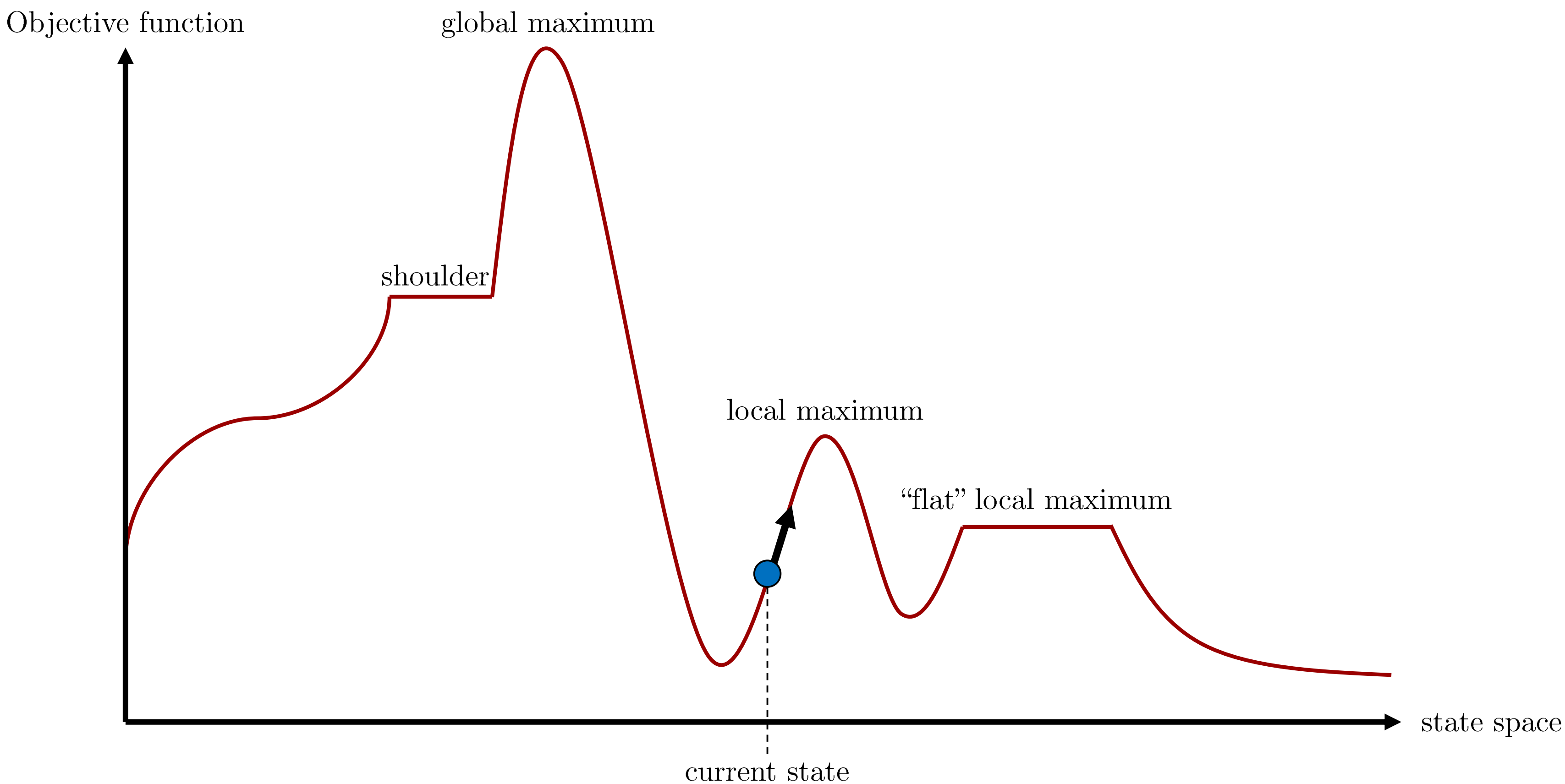
Proof of theorem

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

How could this theorem help us in solving convex optimization problems?

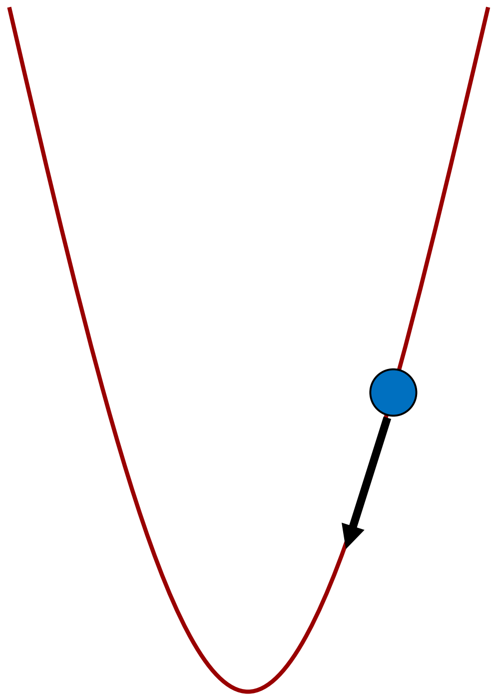


Reminder: hill-climbing search



Solving convex problems

- Convex optimization problems can be solved in polynomial time
- For unconstrained problems, use gradient descent
- Constrained problems require a projection operator that, given $\boldsymbol{x}$, returns the "closest" $\boldsymbol{y} \in \mathcal{F}$



Solving convex problems

- There are a wide range of tools that can take optimization problems in "natural" forms and compute a solution
- Examples include: CVX (MATLAB), YALMIP (MATLAB), AMPL (custom language), GAMS (custom language), cvxpy (Python)

Solving convex problems

Given $\boldsymbol{a}^{(i)} \in \mathbb{R}^2$ for $i = 1, \dots, m$,

$$\min_{\boldsymbol{x}} \sum_{i=1}^{m} \left\| \boldsymbol{x} - \boldsymbol{a}^{(i)} \right\|_2 \text{ s.t. } x_1 + x_2 = 0$$

Constrained Weber Point

```
import cvxpy as cp
import numpy as np

n = 2
m = 10
A = np.random.randn(m,n)
x = cp.Variable(n)
f = sum([cp.norm(x - A[i,:],2) for i in range(m)])
constraints = [sum(x) == 0]
result = cp.Problem(cp.Minimize(f), constraints).solve()
print x.value
```

Summary

- Terminology:
  - Convex optimization problem
  - Convex set
  - Convex function
  - Local and global optimum
- Big ideas:
  - In convex problems, every locally optimal solution is globally optimal!