CMU 15-781

Lecture 13: Convex Optimization

Teacher:
Gianni A. Di Caro

OPTIMIZATION PROBLEMS

- Casting AI problems as optimization problems has been one of the primary trends of the last 15 years
- A seemingly remarkable fact:

	Search problems	Optimization problems
Variable type	Discrete	Continuous
# solutions	Finite	Infinite
Complexity	Exponential	Polynomial (*Convex* class)

Formal definition

- Optimization problems are of the form

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

  - $f: \mathbb{R}^n \longmapsto \mathbb{R}$ is the objective function
  - $\boldsymbol{x} \in \mathbb{R}^n$ is the optimization vector variable
  - $\mathcal{F} \subseteq \mathbb{R}^n$ is the feasible set (constraints)
- $\boldsymbol{x}^* \in \mathbb{R}^n$ is an optimal solution (global minimum) if $\boldsymbol{x}^* \in \mathcal{F}$ and $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{F}$
- *Mathematical programming* problem

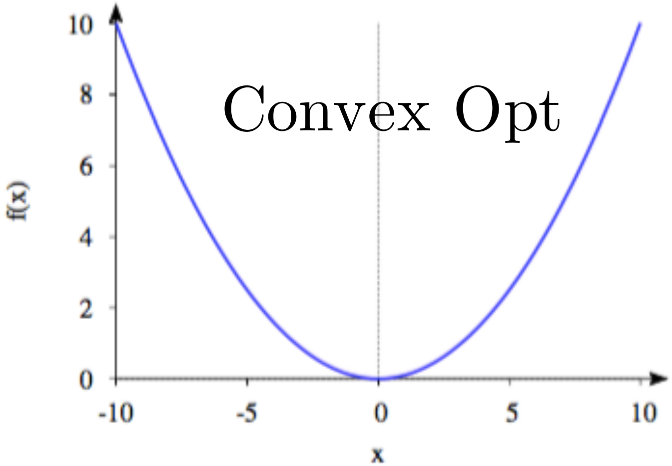
Properties

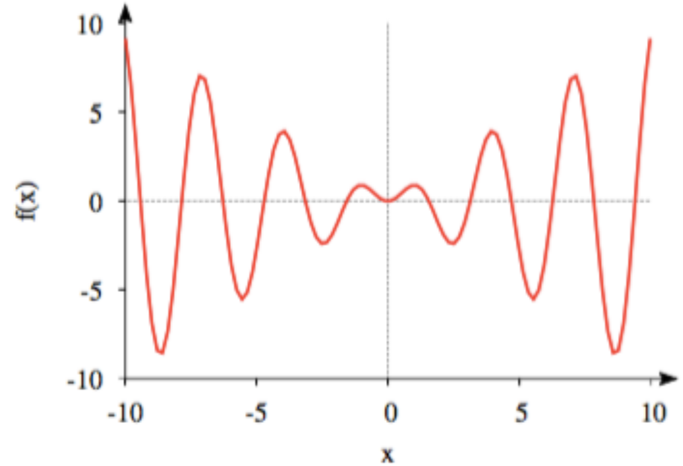
- Given an optimization problem:

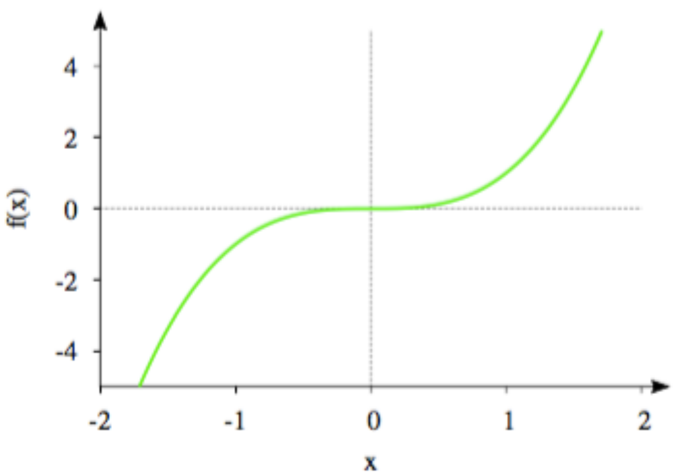
$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$
$$\text{such that } \boldsymbol{x} \in \mathcal{F}$$

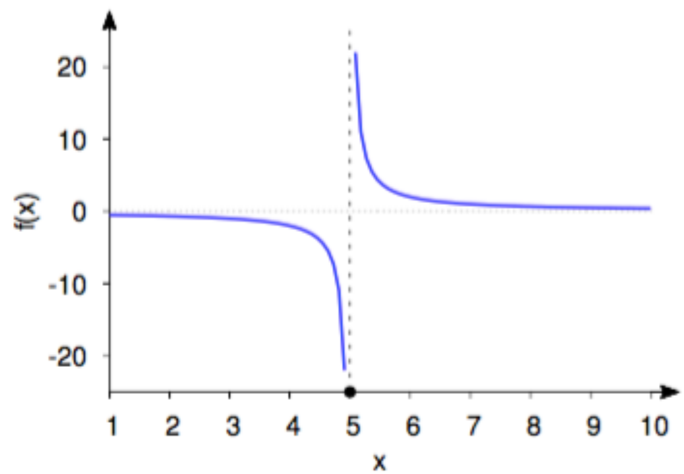
- $\min_{\boldsymbol{x}} f(\boldsymbol{x})$ is equivalent to $\max_{\boldsymbol{x}} -f(\boldsymbol{x})$
- If $\mathcal{F} = \emptyset$ the problem has no solution (unfeasible)
- If $\mathcal{F}$ is an open set, only the inf (sup) is guaranteed but not min (max)
- The problem is unbounded if $f \rightarrow -\infty$

Unconstrained 1D Example cases

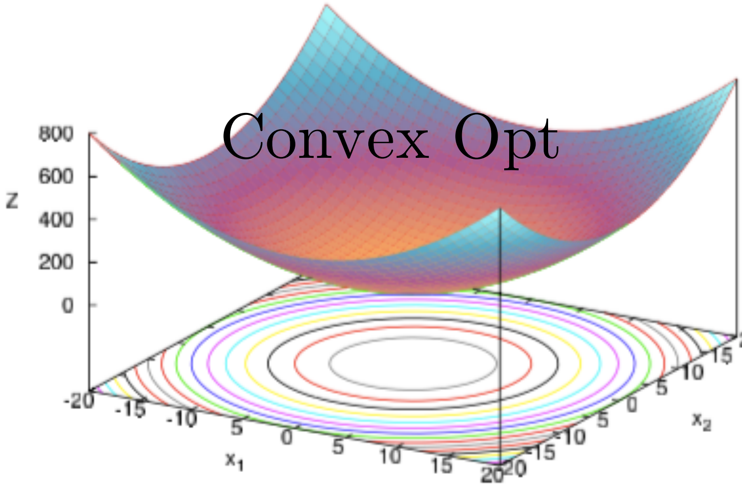


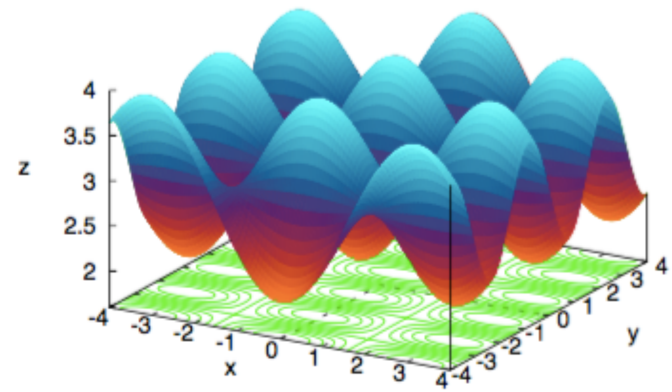


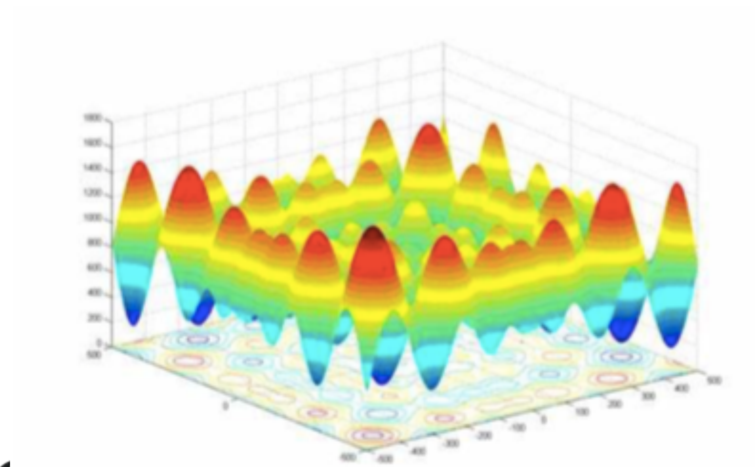


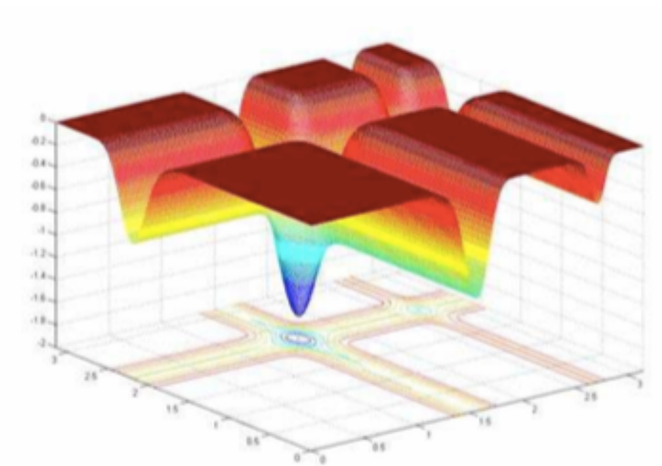


Unconstrained 3D Example cases









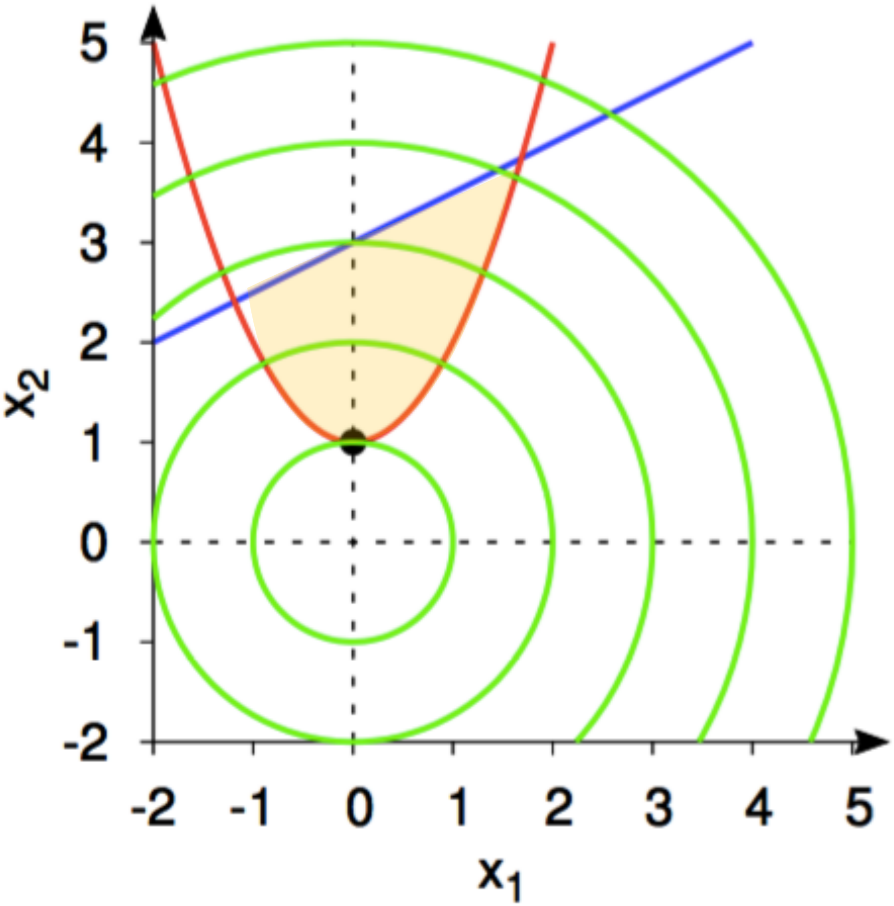
(Constrained) Example cases of Mathematical Programming

	Linear	Convex	Reals	Certainty
$\min_{\vec{x}}$	$2x_1 + x_2 - 4x_3$	$2x_1 + x_2 - 4x_3$	$2x_1 + x_2 - 4x_3$	$2x_1 + x_2 - 4x_3$
s.t.	$x_1 + x_2 \leqslant 5$	$x_1^4 + x_2 \leqslant 5$	$x_1 + x_2 \leqslant 5$	$x_1 + x_2 \leqslant 5$
	$x_1, x_2, x_3 \geqslant 0$	$x_1 + x_3 \geqslant 0$	$x_1, x_2, x_3 \geqslant 0$	$x_1, x_2, x_3 \in \{0, 1\}$
$\min_{\vec{x}}$	$2x_1 + x_2 - 4x_3^3$	$2x_1 + x_2 + 4x_3^3$	$2x_1 + x_2 - 4x_3$	$2x_1 + x_2 - \mathbb{E}_\omega Q(x_3, \omega)$
s.t.	$x_1 + \sqrt{x_2} \leqslant 5$	$x_1 + sin(x_2) \leqslant 5$	$x_1 + x_2 \leqslant 5$	$x_1 + x_2 \leqslant 5$
	$x_1, x_2, x_3 \geqslant 0$	$x_1 + x_3 \geqslant 0$	$x_1, x_2, x_3 \in \mathbb{Z}^+$	$x_1, x_2, x_3\ 0, \ \omega \sim U[0, 10]$
	Non-linear	Non-convex	Zeals	Stochastic

Example of constrained MP

$$\min_{\underline{x}} \; z = x_1^2 + x_2^2$$

$$\begin{aligned} s.t. \quad & x_1 - 2x_2 + 6 \geqslant 0 \\ & -x_1^2 + x_2 - 1 \geqslant 0 \\ & x_1, x_2 \geqslant 0 \end{aligned}$$

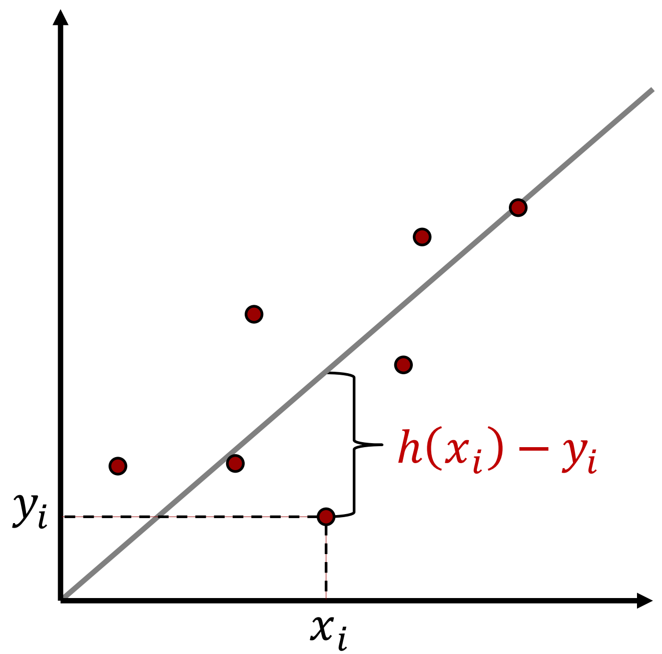


Example: least-squares fitting

- 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)

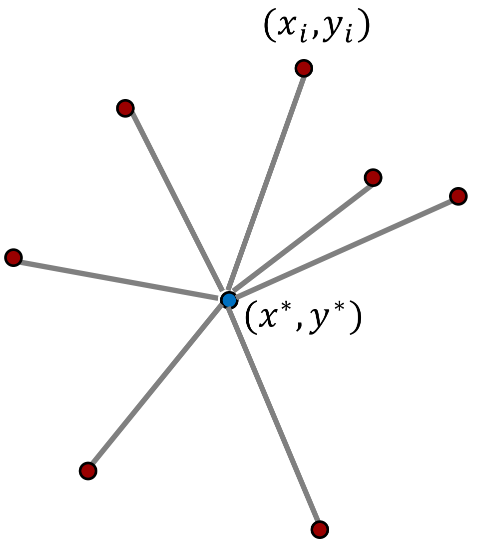


Example: Weber point

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

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

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



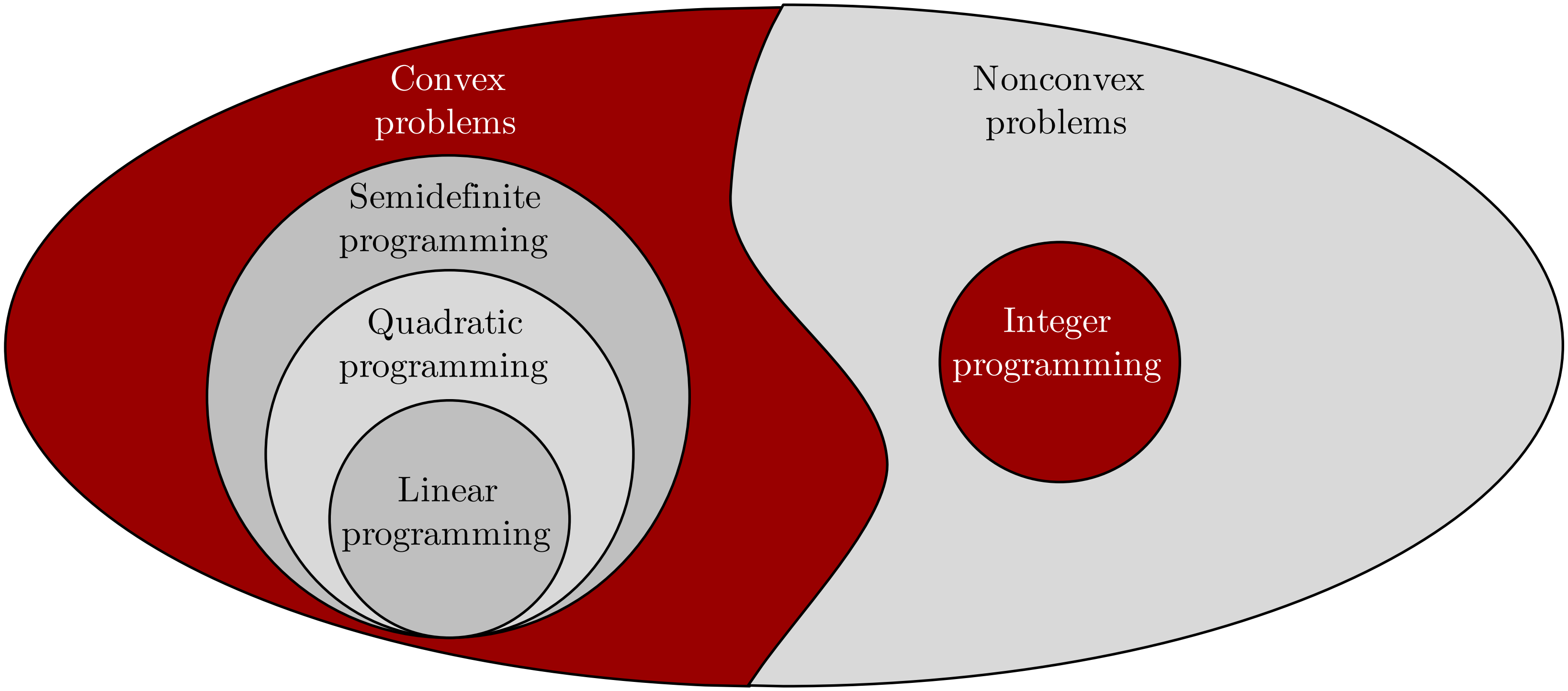
Machine learning

- 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

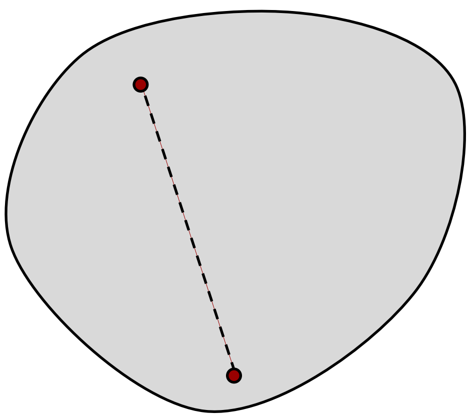
- A convex optimization problem is a special case of a general optimization problem

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

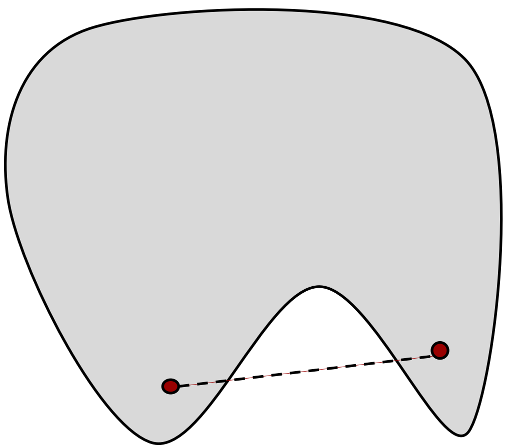
where the target function $f: \mathbb{R}^n \to \mathbb{R}$ is a convex function, and the feasible region $\mathcal{F}$ is a convex set

Convex sets

- A set $\mathcal{F} \subseteq \mathbb{R}^n$ is convex if for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{F}$ and $\theta \in [0,1]$, $\theta \boldsymbol{x} + (1 - \theta) \boldsymbol{y} \in \mathcal{F}$
- A set is convex if, given two points in it, it contains all their possible linear (convex) combinations



Convex set



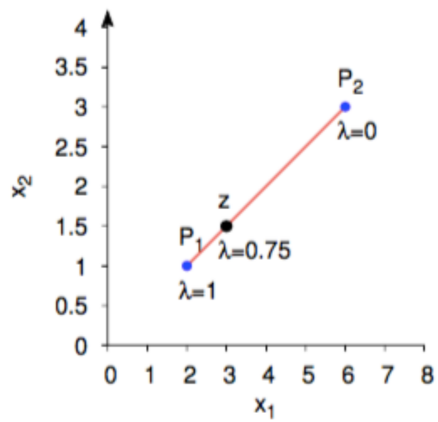
Nonconvex set

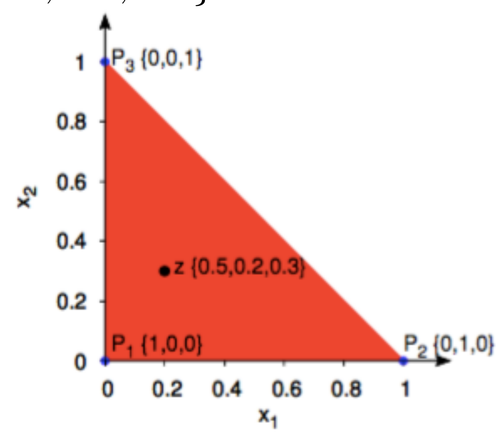
Convex combination

- Given k points $\boldsymbol{P}_i \in \mathbb{R}^n$, $i = 1, \ldots, k$, a point $z \in \mathbb{R}^n$ is a **convex combination of the points** P_i if:

$$z = \sum_{i=1}^{k} \lambda_i \boldsymbol{P}_i, \quad \lambda_i \geq 0 \ \forall i, \quad \sum_{i=1}^{k} \lambda_i = 1$$

- If $k = 2 \quad \rightarrow z = \lambda \boldsymbol{P}_1 + (1-\lambda)\boldsymbol{P}_2, \quad \lambda_1 = \lambda,\ \lambda_2 = (1-\lambda)$
- *Example:* $k = 2, \quad \boldsymbol{P}_1 = (2,1),\ \boldsymbol{P}_2 = (6,3),\ \lambda = 0.75 \rightarrow z = (3, 1.5)$
- *Example:* $k = 3, \quad \boldsymbol{P}_1 = (0,0), \boldsymbol{P}_2 = (1,0), \boldsymbol{P}_3 = (0,1), \lambda_i = \{0.5, 0.2, 0.3\} \rightarrow$ $z = (0.2, 0.3)$

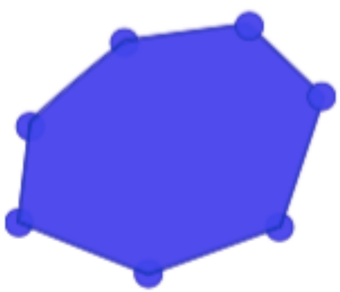


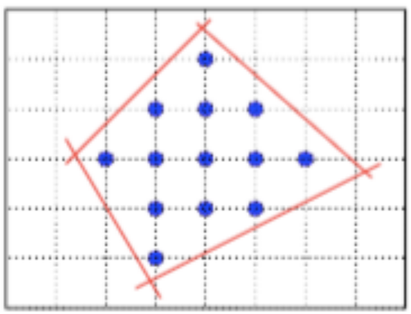


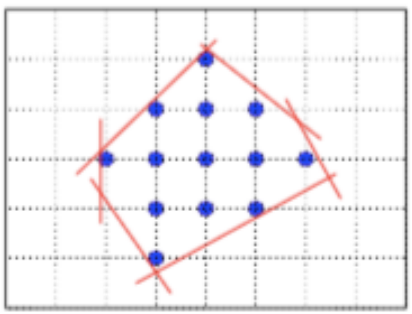
Convex Hull

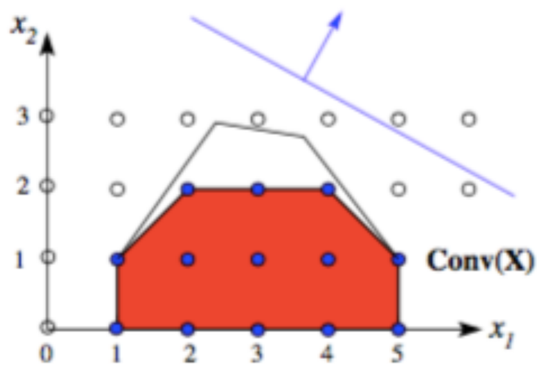
- Given a set P of k points of $\mathbb{R}^n$, $P = \{\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_k\}$, the smallest convex set, $conv(P)$, that includes P is the **convex hull**, $P \subseteq conv(P)$
- $conv(P)$ is the set of all convex combinations of the points in P:

$$conv(P) = \{z \in \mathbb{R}^n \; : \; z = \sum_{i=1}^{k} \lambda_i \boldsymbol{P}_i, \quad \forall \lambda_i, \; i = 1, \ldots, k \,|\, \lambda_i \geq 0 \wedge \sum_{i=1}^{k} \lambda_i = 1\}$$







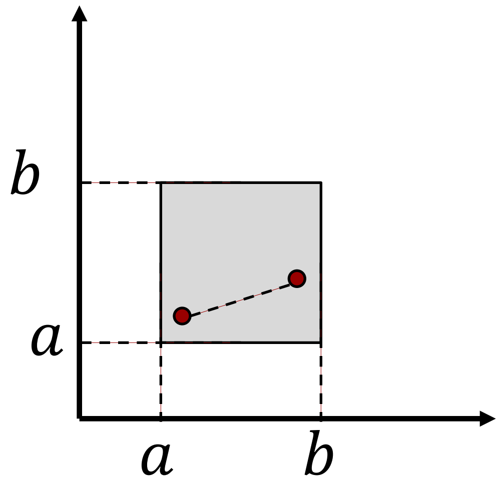


Examples of convex sets

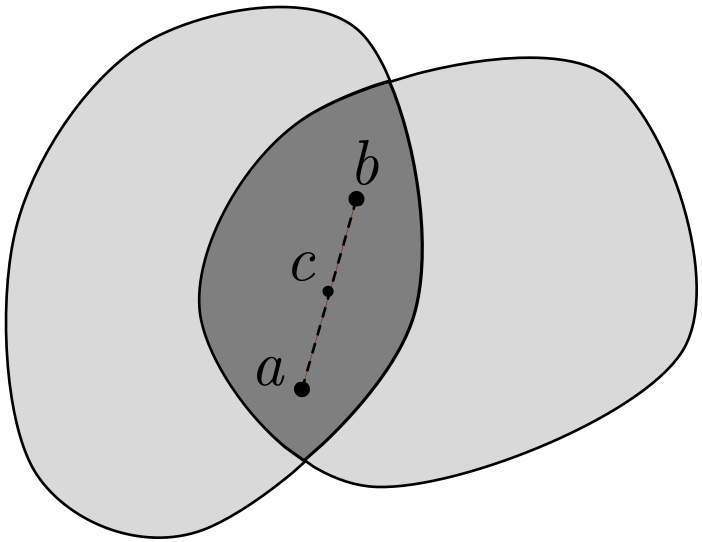
- $\mathcal{F} = \{\boldsymbol{x} \in \mathbb{R}^n : \forall i = 1, \dots, n,\ a \le x_i \le b\}$
- Proof:
  - Let $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{F}$, and $\theta \in [0,1]$
  - For all $i = 1, \dots, n,$

    $a \le x_i$ and $a \le y_i$, so

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



Intersection of convex sets



Intersection of convex sets

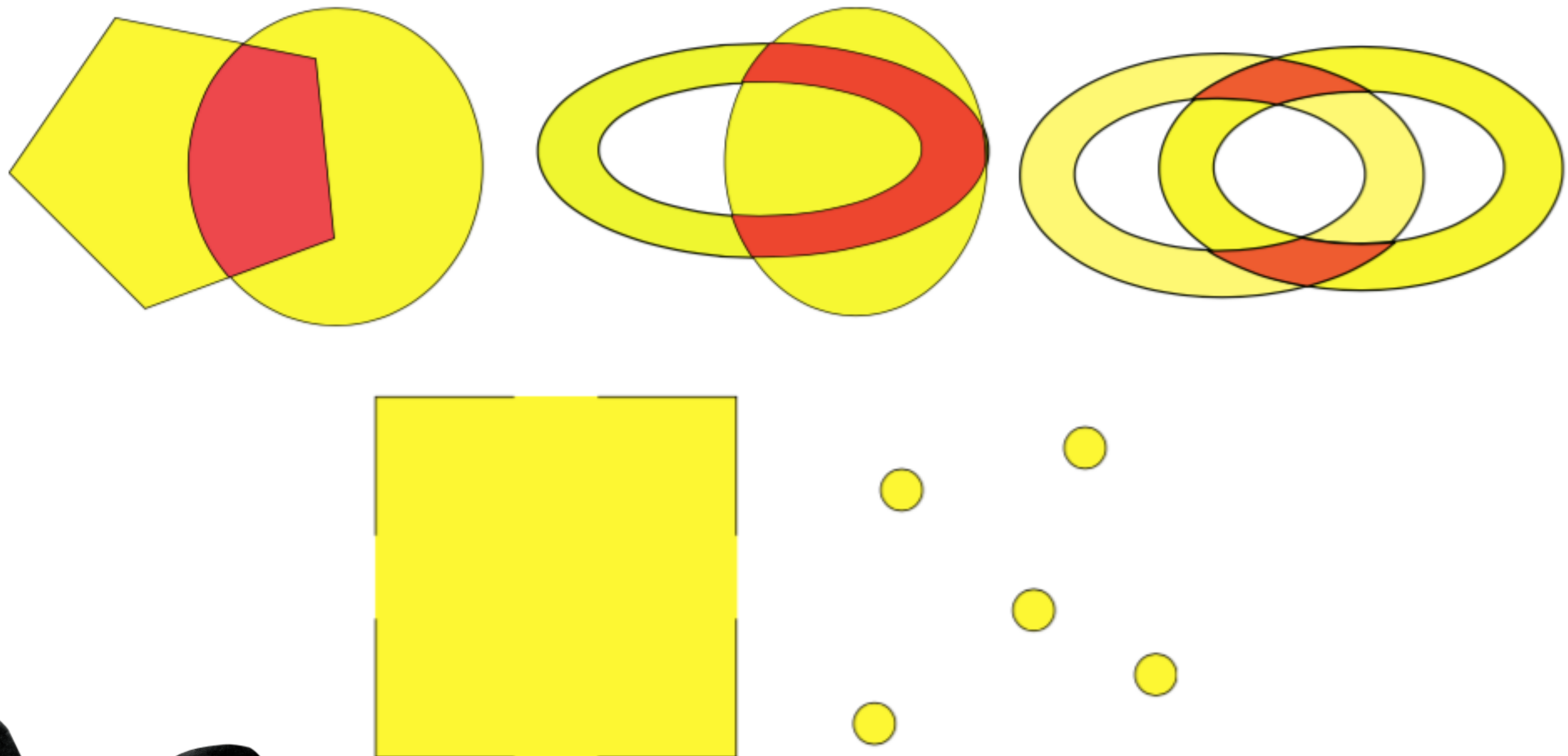
$$\mathcal{F} = \bigcap_{i=1}^{m} C_i$$

$C_1, \dots, C_m$ are convex

Proof (by contradiction):

- Let's prove it first for two convex sets A and B.
- Let a and b be two points belonging to C = A ∩ B (and, therefore, to both A and B).
- Let's assume there is a third point c on the line between that a and b, such that $c \notin$ C, meaning that C is not convex.
- But, for the convexity of A, every point on the line a-b must be in A, and the same holds for B → c must be in C!
- For m intersecting sets the same reasoning can be applied in pairs

Examples of (non)convex sets

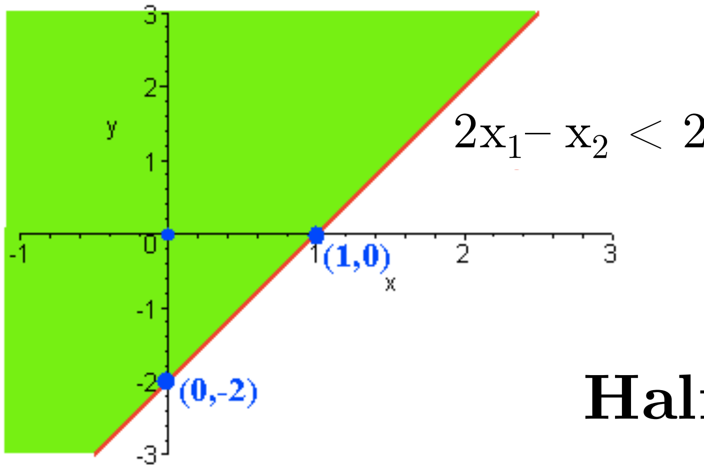


Examples of convex sets

- 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

LINEAR INEQUALITIES



Linear inequalities

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

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

Half-space ($\leq$ closed, $<$ open)

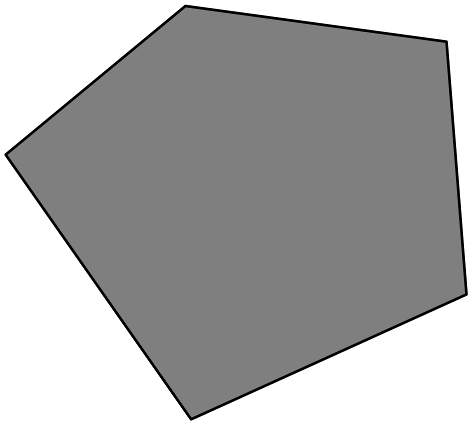
Convex (obvious by geometrical considerations):

Two points x and y in $\mathcal{F}$: a$x \leq$ b, a$y \leq$ b

$\theta x + (1-\theta)y \in \mathcal{F}$? $\rightarrow \theta x + (1-\theta)y \leq$ b/a

$$\theta x + (1-\theta)y \leq \theta\left(\frac{b}{a}\right) + (1-\theta)\left(\frac{b}{a}\right) = \text{b/a}$$

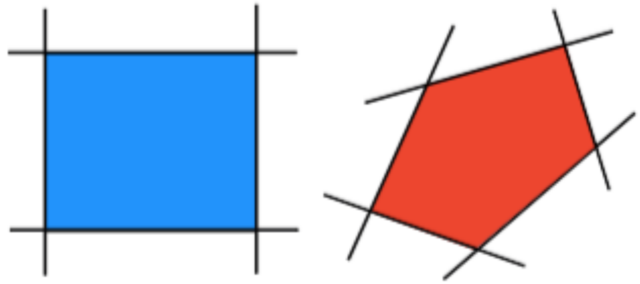
Systems of linear inequalities



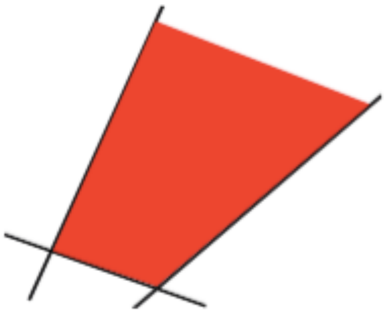
Linear inequalities

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

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



Polytopes



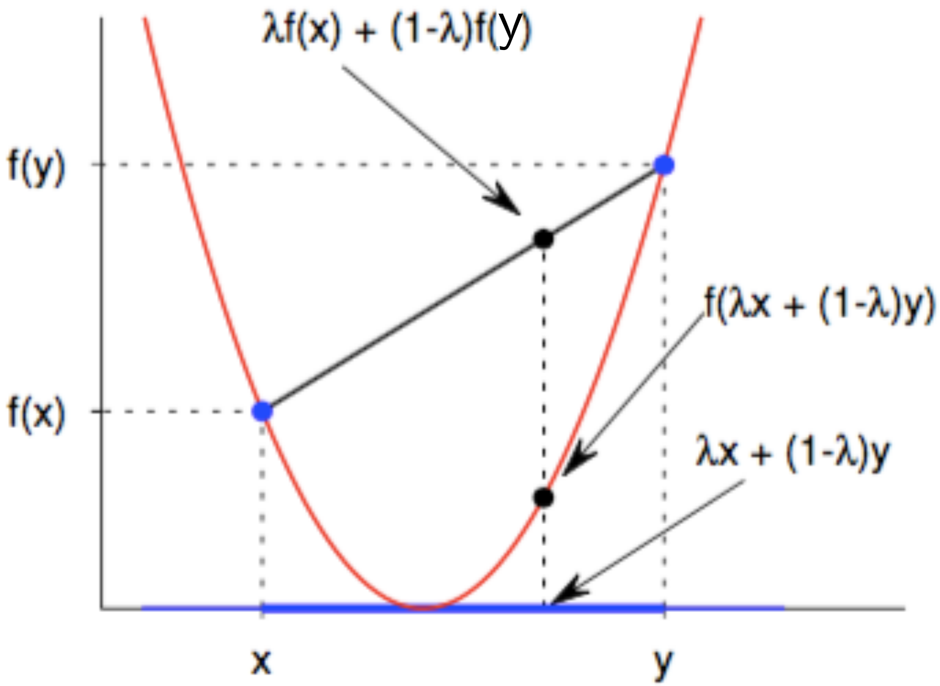
Polyhedron

- Every half-space inequality defines a convex set
- Their intersection is convex

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 $\lambda \in [0,1]$

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



The graph of f is always below (or on) the line segment $\lambda f(\boldsymbol{x}) + (1\text{-}\lambda) f(\boldsymbol{y})$ connecting $(\boldsymbol{x}, f(\boldsymbol{x}))$ to $(\boldsymbol{y}, f(\boldsymbol{y}))$

The line interpolation between any two points in the domain, always over estimates the value of the function

For $f: \mathbb{R} \to \mathbb{R}$, this equals to $f'' > 0$

Examples of convex problems

- Exponential: $f(x) = e^{ax}$
  - $f''(x) = a^2 e^{ax} \geq 0$ for all $x \in \mathbb{R}$
- Euclidean (L2) 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 $\boldsymbol{x}$

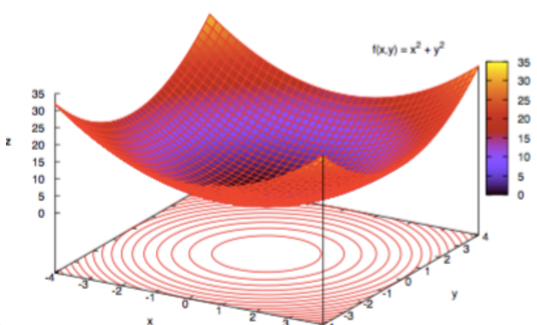
Affine transformation

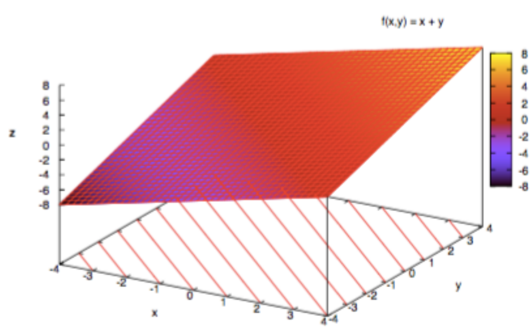
Examples of convex problems

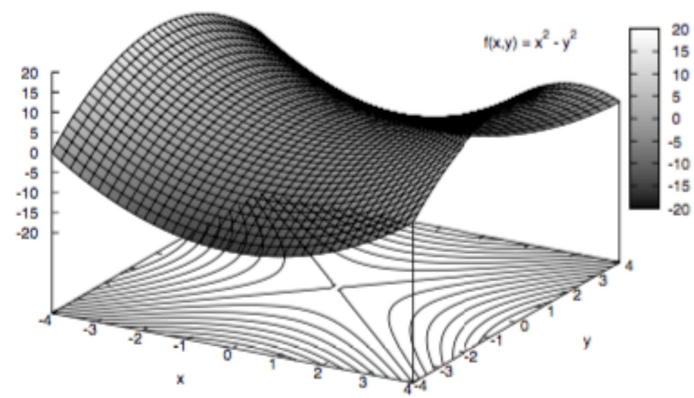
- Sublevel sets (isolines): If f is convex,

$$\{x \in \mathbb{R}^n : f(\boldsymbol{x}) \leq c\} \text{ is a convex set}$$

$$f(\lambda x+(1-\lambda)y) \leq \lambda f(x)+(1-\lambda)f(y) \leq \lambda c+(1-\lambda)c = c$$







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 $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}$ are problem data

- A convex optimization problem (why?)

Affine transformation over a convex function (Euclidean norm) + Linear combination which is also convex

Examples of convex problems

- Linear programming:

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

  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

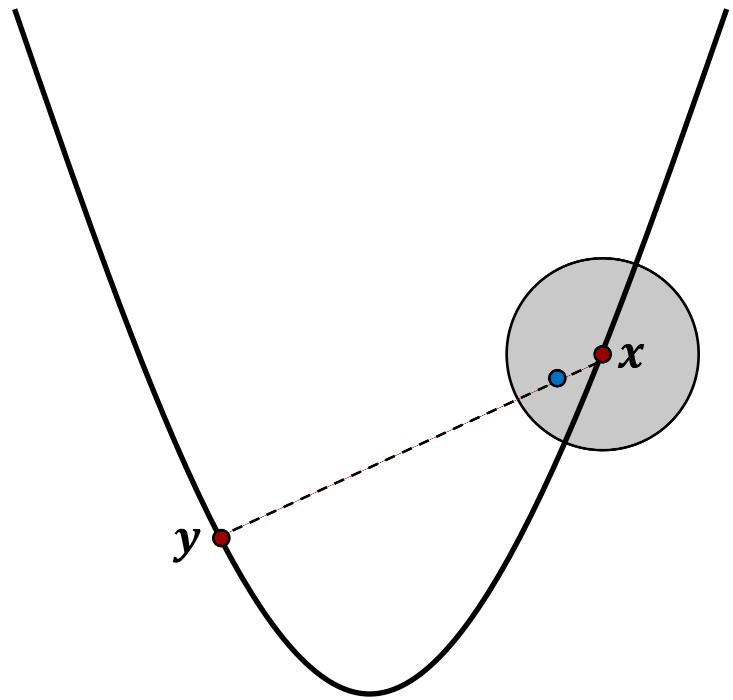
- A point $\boldsymbol{x} \in \mathbb{R}^n$ is globally optimal (global minimum) if $\boldsymbol{x} \in \mathcal{F}$ and for all $\boldsymbol{y} \in \mathcal{F}$, $f(\boldsymbol{x}) \leq f(\boldsymbol{y})$
- A point $\boldsymbol{x} \in \mathbb{R}^n$ is locally optimal if $\boldsymbol{x} \in \mathcal{F}$ and there exists $R > 0$ small 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 (one, or infinite global optima)

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



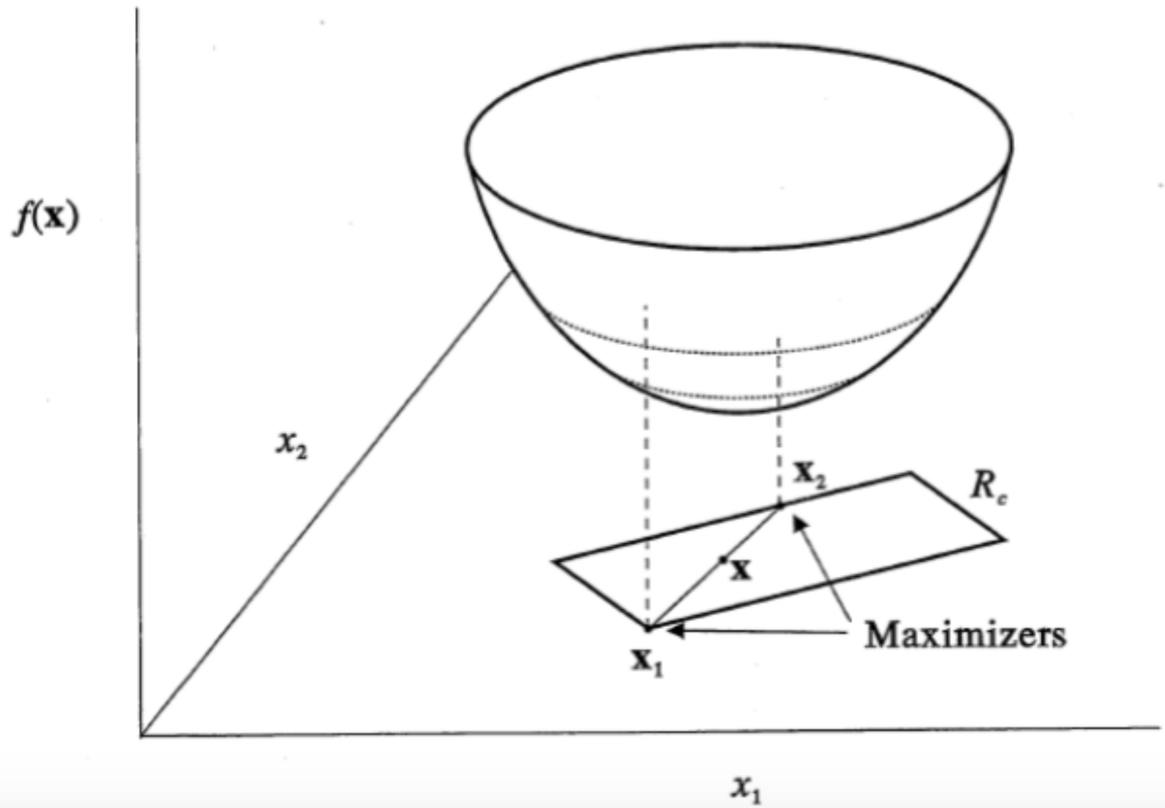
Proof of theorem

- Then:
  - $\boldsymbol{z}$ is feasible (for small enough R)
  - $f(\boldsymbol{z}) = f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \le \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y})$
    $< \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{x}) = f(\boldsymbol{x})$
  - $\|\boldsymbol{x} - \boldsymbol{z}\|_2 = \left\| \frac{R}{2\|x-y\|_2} (\boldsymbol{x} - \boldsymbol{y}) \right\|_2 = \frac{R}{2} < R$

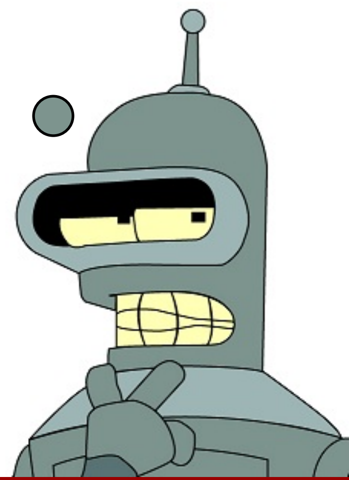
  it's inside thee R ball!
- Therefore, $\boldsymbol{x}$ is not locally optimal, contradicting our assumption ■

Maxima of convex functions

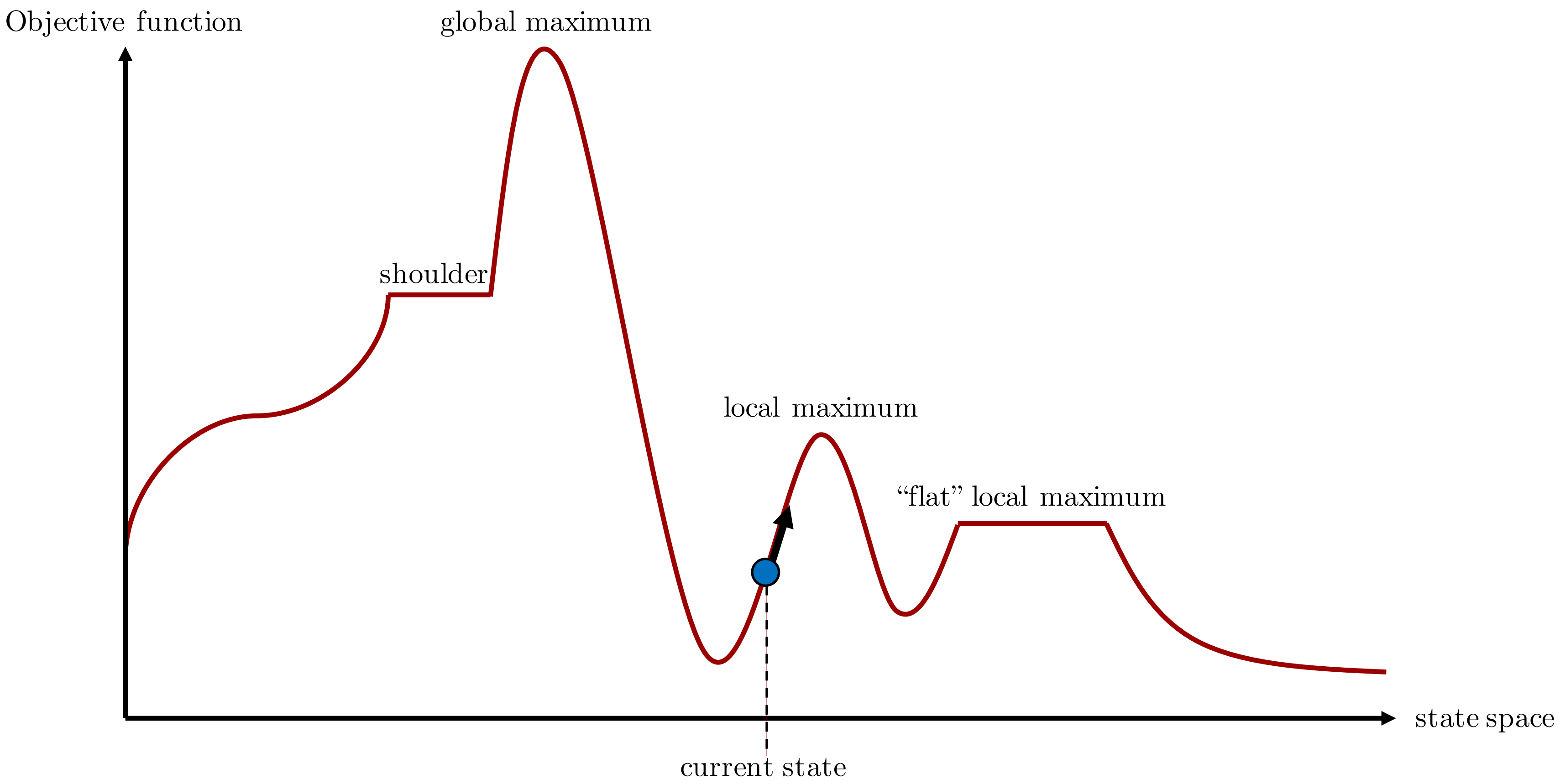
On the frontier of the domain



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

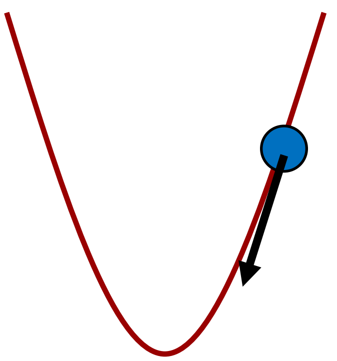


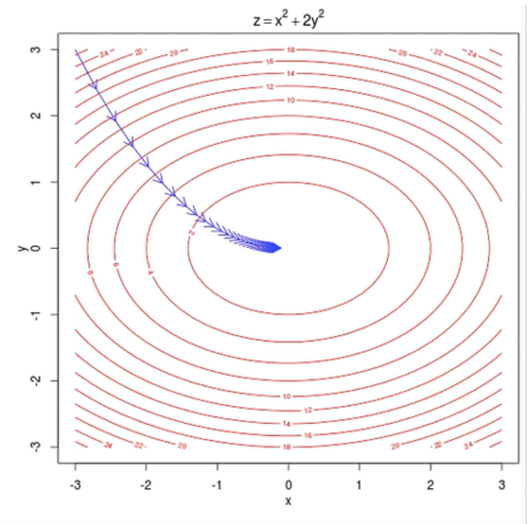
Reminder: hill-climbing search



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 $i = 1, \dots, m,$

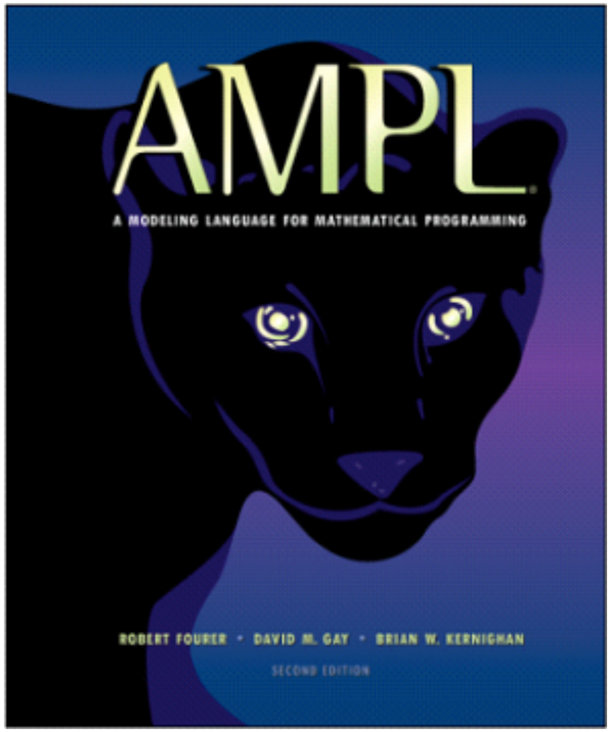
$$\min_{\boldsymbol{x}} \sum_{i=1}^{m} \left\| \boldsymbol{x} - \boldsymbol{a}^{(i)} \right\|_2 \;\; \text{s.t. } x_1 + x_2 = 0$$

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
```

AMPL: A SET OF SOLVERS + NICE MODELING LANGUAGE




```
set ORIG;   # origins
set DEST;   # destinations

set LINKS within {ORIG,DEST};

param supply {ORIG} >= 0;   # amounts available at origins
param demand {DEST} >= 0;   # amounts required at destinations

   check: sum {i in ORIG} supply[i] = sum {j in DEST} demand[j];

param cost {LINKS} >= 0;   # shipment costs per unit
var Trans {LINKS} >= 0;    # units to be shipped

minimize Total_Cost:
   sum {(i,j) in LINKS} cost[i,j] * Trans[i,j];

subject to Supply {i in ORIG}:
   sum {(i,j) in LINKS} Trans[i,j] = supply[i];

subject to Demand {j in DEST}:
   sum {(i,j) in LINKS} Trans[i,j] = demand[j];
```

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!