Harvard CS 182: Artificial Intelligence	September 26, 2022

Lecture 7: Convex Optimization

Lecturer: Ariel Procaccia *Authors: Max Guo, Janani Sekar*

1 Introduction

We're continuing along the trend of investigating different problem-solving and optimization techniques. Although convex optimization seems more like applied math, it will be useful for both giving context for our next lecture on integer programming and because convex optimization techniques are used in machine learning all the time.

1.1 Optimization Problems

Casting AI problems as optimization problems has been one of the primary AI trends in the 21st century. One motivation is because the continuous versions of some problems are, a bit counter intuitively, easier than their respective discrete problems (see Table 1).

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

Table 1: Discrete vs Continuous Optimization Problems

Note that discrete optimization problems are NP-hard to solve, while continuous optimization problems can be solved more easily — we'll discuss this later into the course. Let's define optimization problems formally:

Definition 1 (Optimization Problem) *Consider a function f. Let $\mathcal{F} \subseteq \mathbb{R}^n$ be the feasible set, and call $\mathbf{x} \in \mathbb{R}^n$ be the optimization variable. In an optimization problem, we wish to find $\min_{\mathbf{x}} f(\mathbf{x})$ such that $\mathbf{x} \in \mathcal{F}$. We call $\mathbf{x}^* \in \mathbb{R}^n$ an optimal solution if $\mathbf{x}^* \in \mathcal{F}$ and $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{F}$.*

1.2 Examples of Optimization Problems

Example 1 (Least-Squares Fitting) *Given* (x_i, y_i) *for* $i = 1, \ldots, m$, *find* $h(x) = ax + b$ *that optimizes:*

$$\min_{a,b} \sum_{i=1}^{m} (ax_i + b - y_i)^2$$

where a *is the slope, and* b *is the intercept.*

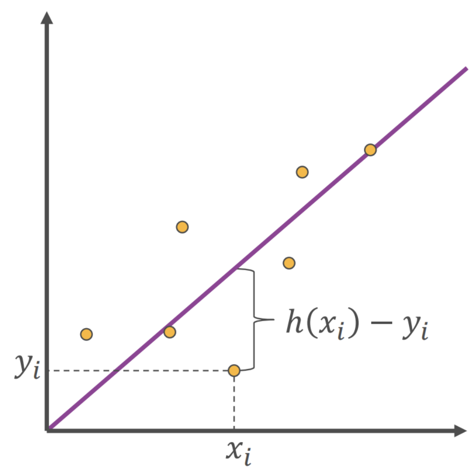


Figure 1: Least-Squares Fitting

Example 2 (Weber Point) *Given* (x_i, y_i) *for* $i = 1, \ldots, m$, *find the point* (x^*, y^*) *that minimizes the sum of Euclidean distances:*

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

There are also many modifications we can make to this problem, including adding constraints such as $a \leq x^* \leq b, c \leq y^* \leq d$.

2 Convex Optimization

Instead of thinking of optimization problems as discrete versus continuous, we can think of optimization problems as convex versus non convex problems.

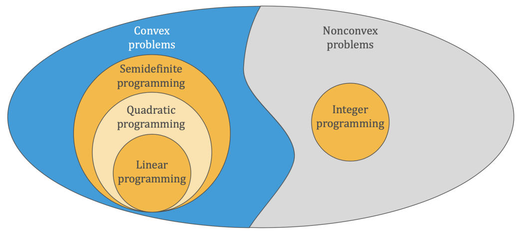


Figure 2: The Optimization Universe

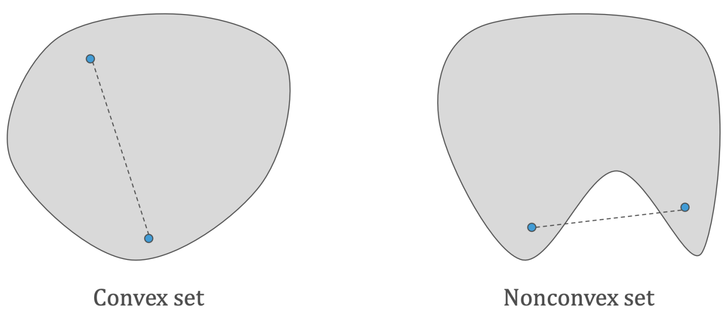
Let's define what a convex optimization problem is:

Definition 2 (Convex Optimization Problem) *A convex optimization problem is a specialization of a general optimization problem* $\min_x f(\mathbf{x})$ *such that* $\mathbf{x} \in \mathcal{F}$ *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.*

Let's define what *convex functions* and *convex sets* are.

2.1 Convex Sets

Definition 3 (Convex Sets) *A set* $\mathcal{F} \subseteq \mathbb{R}^n$ *is convex if for all* $\mathbf{x}, \mathbf{y} \in \mathcal{F}$ *and* $\theta \in [0, 1]$, $\theta\mathbf{x} + (1 - \theta)\mathbf{y} \in \mathcal{F}$.



Intuitively the definition of convex sets is: for any two points in the set, if I draw the line between the two points, every point in the line must be in the set. Now we'll provide some examples of convex sets:

Example 3 *Prove that*

$$\mathcal{F} = \{\mathbf{x} \in \mathbb{R}^n : \forall i = 1, \ldots, n, a \leq x_i \leq b\}$$

is a convex set.

Proof: Let $\mathbf{x}, \mathbf{y} \in \mathcal{F}$, 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, $\theta \mathbf{x} + (1-\theta)\mathbf{y} \in \mathcal{F}$.
■

We also see that sets defined by linear inequalities are, by a similar argument to the above, convex sets:

Example 4 *Suppose we have an $m \times n$ matrix A with entries $a_{ij}, 1 \leq i \leq m, 1 \leq j \leq n$. Suppose we have a $\vec{b} \in \mathbb{R}^m$. Then the set defined by:*

$$\{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid \sum_{j=1}^{n} a_{ij} x_j \leq b_i \forall 1 \leq i \leq m\}$$

is a convex set.

Proof: Let $\vec{x}$ and $\vec{y}$ be two arbitrary vectors that satisfy the conditions. In other words, $A\vec{x} \leq \vec{b}$ and $A\vec{y} \leq b$. Then:

$$\begin{aligned} A(\theta\vec{x} + (1-\theta)\vec{y}) &= \theta A\vec{x} + (1-\theta)A\vec{y} \\ &\leq \theta\vec{b} + (1-\theta)\vec{b} \\ &= \vec{b}. \end{aligned}$$

■

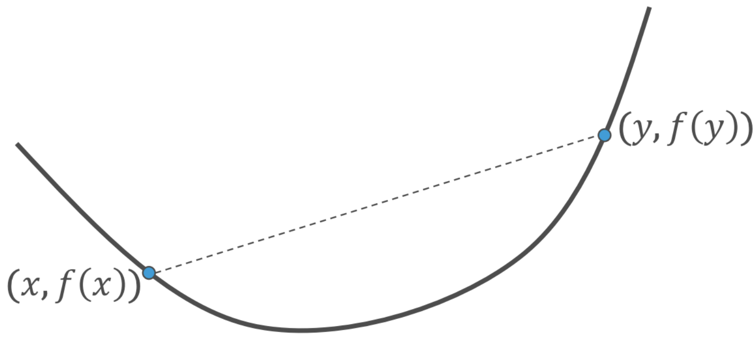
Moreover, *intersection of convex sets are convex.* An argument works likes this: for any two points within the intersection of several convex sets, the line between them is (because each of the sets is convex) also withing each of the convex sets, so the line must be in the intersection, which proves the intersection is a convex set.

Note that it is *not* true that the union of convex sets is convex.

2.2 Convex Functions

Definition 4 (Convex Function) *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\theta \in [0,1]$,*

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



For functions that $f : \mathbb{R} \to \mathbb{R}$ that are twice differentiable, this definition is equivalent to saying that $f''(x) \geq 0$ for all $x \in \mathbb{R}$ (this may have been the definition you have seen before in calculus courses!)

Finally, f is concave if and only if $-f$ is convex.

Now let's give some examples for convex functions:

Example 5

- *Exponential.* $f(x) = e^{ax}$. *We can show this is convex via the second derivative:*
$$f''(x) = a^2 e^{ax} \geq 0 \quad \text{for all } x \in \mathbb{R}.$$
- *Euclidean Norm.* $f(\mathbf{x}) = ||\mathbf{x}||_2 = \sqrt{\sum_{i=1}^{n}(x_i)^2}$. *We can show this is convex via the definition (and via the triangle inequality):*
$$\begin{aligned} ||\theta\mathbf{x} + (1-\theta)\mathbf{y}||_2 &\leq ||\theta\mathbf{x}||_2 + ||(1-\theta)\mathbf{y}||_2 \\ &= \theta||\mathbf{x}||_2 + (1-\theta)||\mathbf{y}||_2 \end{aligned}$$

We also have that weighted sums of convex functions are convex:

Example 6 *Let $f(x) = \sum_{i=1}^{m} a_i f_i(x)$, where f_i is convex and $a_i \geq 0$ for all $i = 1, \ldots, m$. Then f is convex.*

Proof:

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \sum_i a_i f_i(\theta x + (1-\theta)y) \\ &\leq \sum_i a_i(\theta f_i(x) + (1-\theta)f_i(y)) \\ &= \sum_i a_i f_i(x) + (1-\theta)\sum_i a_i f_i(y) \\ &= \theta f(x) + (1-\theta)f(y) \end{aligned}$$

■

2.3 Convex Optimization Problems

Let's give a few examples of convex optimization problems.

Example 7 (Weber point) *Recall that the Weber point problem in n dimensions finds the value of*

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

where $\mathbf{x}^* \in \mathbb{R}^n$ is the optimization variable and $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}$ are the problem data. This is a convex optimization problem because we have a convex function here (we showed norms are convex and that sums of convex functions are convex, and the inside is a translation which does not affect the convexity), and we are working over a convex set (the entirety of $\mathbb{R}^n$ in this case).

Example 8 (Linear Programming) *The linear programming problem can be formulated as finding*

$$\min_{\mathbf{x}} \mathbf{c}^T \mathbf{x} \qquad \text{such that } A\mathbf{x} = \mathbf{a} \text{ and } B\mathbf{x} \leq \mathbf{b},$$

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

Verbally, we are trying to maximize a (convex) linear objective function subject to linear constraints (which gives us a convex set).

To make this concrete, we can look at a more specific linear programming problem.

Example 9 (Max Flow) *In the max flow problem, we are given a directed graph $G = (V, E)$, with two specific vertices, a source s, a sink t. Every edge has a capacity a_{xy} for each $(x, y) \in E$. The flow is a function:*

$$f : E \to \mathbb{R}^+$$

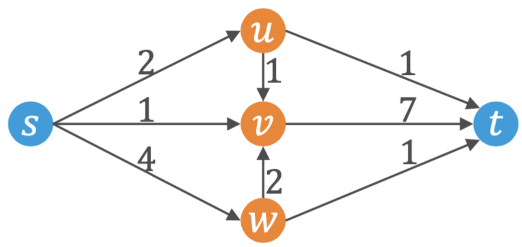
that satisfies:

$$f_{xy} \leq a_{xy} \text{ for all } (x, y) \in E, \text{ and } \sum_{(y,x) \in E} f_{yx} = \sum_{(x,x) \in E} f_{xx} \text{ for all } x \in V \backslash \{s, t\}$$

The value of a flow is:

$$\sum_{(s,x) \in E} f_{sx}$$

Verbally, this is saying that for every vertex excluding s and t, the total flow coming into the vertex should be equal to the total flow coming out of the vertex. The value of the flow is the total flow from the source.



We can fit the max flow problem in the figure above to the format of a canonical linear program:

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

Along every edge, linear constraints require the flow to be ≥ 0 and $\leq$ the edge's capacity. Finally, there are three flow conservation constraints based on the graph:

$$f_{s,u} = f_{u,v} + f_{u,t}$$

$$f_{s,w} = f_{w,v} + f_{w,t}$$

$$f_{s,v} = f_{u,v} + f_{w,v} = f_{v,t}$$

Now let's move on to the "optimization" part of convex optimization problems.

2.4 Optimality

Definition 5 (Global Optima) *A point $\mathbf{x} \in \mathbb{R}^n$ is globally optimal if $\mathbf{x} \in \mathcal{F}$ and for all $\mathbf{y} \in \mathcal{F}$, $f(\mathbf{x}) \leq f(\mathbf{y})$.*

Definition 6 (Local Optima) *A point $\mathbf{x} \in \mathbb{R}^n$ is locally optimal if $\mathbf{x} \in \mathcal{F}$ and there exists $R > 0$ such that for all $\mathbf{y} \in \mathcal{F}$ with $||\mathbf{x} - \mathbf{y}||_2 \leq R$, $f(\mathbf{x}) \leq f(\mathbf{y})$.*

The following theorem is a huge reason for why we care about convexity:

Theorem 7 *For a convex optimization problem, all locally optimal points are globally optimal.*

In other words, if we have a function we are trying to optimize, and we are using techniques that find us local optima, then given a convex function and having found a local optima we can be confident that we've actually found a global optima and fully solved our optimization problem!

Let's prove the theorem:

Proof: Suppose $\mathbf{x}$ is locally optimal for some R but not globally optimal. Then there is a point $\mathbf{y}$ such that $f(\mathbf{y}) < f(\mathbf{x})$. Define $\mathbf{z} = (1-\theta)\mathbf{x} + \theta\mathbf{y}$ for $\theta = \frac{R}{2||\mathbf{x}-\mathbf{y}||_2}$. We can assume that $||\mathbf{x}-\mathbf{y}||_2 > R$, since otherwise $\mathbf{y}$ would already contradict $\mathbf{x}$'s local optimality. Then $\mathbf{z}$ is feasible, since

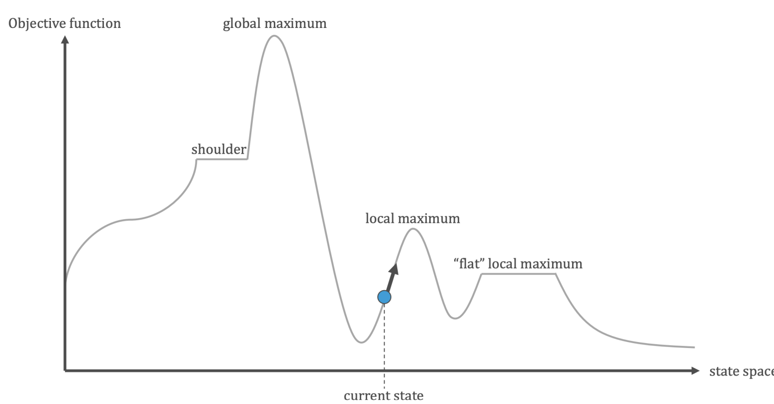
$$||\mathbf{x}-\mathbf{z}||_2 = \left|\left|\frac{R}{2||\mathbf{x}-\mathbf{y}||_2}(\mathbf{x}-\mathbf{y})\right|\right|_2 = \frac{R}{2} < R.$$

However,

$$\begin{aligned} f(\mathbf{z}) &= f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \\ &\le \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{y}) \\ &< \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{x}) \\ &= f(\mathbf{x}) \end{aligned}$$

Hence, $\mathbf{x}$ is not locally optimal, contradicting our assumption. ■

Why is this good for optimization? In general, we might have functions with the following features:



In convex optimization problems, we just follow the objective functions upwards/downwards, and when we've reached a local optima we are guaranteed that it's a global optima!

For unconstrained problems, this looks like using *gradient descent*, which we will talk about briefly later when we get to machine learning. Constrained problems requires a projection operator that, given a $\mathbf{x}$, returns the "closest $\mathbf{y} \in \mathcal{F}$.