Lecture 8: Convex functions

Rajat Mittal *

IIT Kanpur

As already discussed, convex optimization is to optimize a convex function over a set of convex constraint functions. Today, the discussion will be about the convex functions, their properties and their relation with convex sets.

1 Definition

A function is called convex if the line segment connecting any two points on the graph lies above the graph. Formally, a function $f : \mathbb{R}^n \to \mathbb{R}$ is called convex iff

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y),\ \forall x, y \in \mathbb{R}^n,\ 0 < \theta < 1$$

We have assumed the domain to be the entire $\mathbb{R}^n$ space. In general, if the domain is a convex subset of $\mathbb{R}^n$, and satisfies the above mentioned property, it is called a convex function. The canonical example of a convex function is $f(x) = x^2$. You should check that this function is convex.

Note that since the domain is convex, if we restrict the function on any line passing through the domain, the restricted function will be convex. Conversely, if the function is convex on all the lines passing through the domain, then it is convex on the whole domain.

A function f is called concave if it satisfies the above mentioned inequality in opposite direction ($\geq$ instead of $\leq$). It is clear that if f is convex then $-f$ is concave.

Exercise 1. Characterize the functions which are both convex and concave.

1.1 Relation to convex sets

A convex set is the set which contains all possible convex combinations of its points. What is the relation between convex sets and convex functions? The *epigraph* of the function is all points which lie above the graph of the function.

Formally, given a function $f : \mathbb{R}^n \to \mathbb{R}$, the epigraph is the set of all points

$$\{(y, x) : y \geq f(x),\ x \in \mathbb{R}^n,\ y \in \mathbb{R}\}$$

So, if the function is convex, then its epigraph is a convex set and vice versa. This gives a geometric interpretation of a function being convex. The opposite of epigraph (points which lie below the graph of a function) is known as *hypograph.*

1.2 Relation to derivatives

Lets take a look at the graph of the function $f(x) = x^2$. Draw a tangent at every point of the graph. Notice that the tangent always lies below the graph.

The intuition can be formalized. A function is convex iff for any two x, y in the domain

$$f(y) \geq f(x) + f'(x)(y - x)$$

This definition holds for functions of one variable. In the case of more variables, $f'(x)$ will be a vector and a dot product with $y - x$ will be taken.

---

* Thanks to books from Boyd and Vandenberghe, Dantzig and Thapa, Papadimitriou and Steiglitz

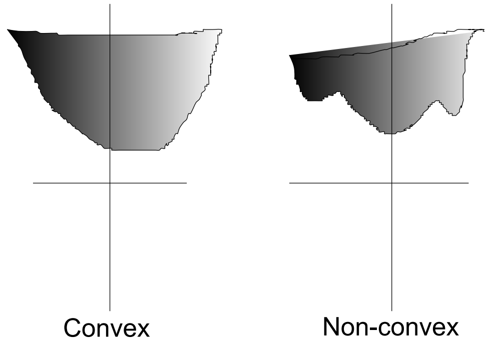

Fig. 1. Epigraph for functions

Another equivalent definition can be given in terms of second derivative can be given. A function is convex iff its second derivative $f''(x)$ is non-negative. For functions with multiple variables, *Hessian* should be a *positive semidefinite matrix*. For a function $f : \mathbb{R}^n \to \mathbb{R}$, the Hessian is an $n \times n$ matrix with $(i, j)^{th}$ entry to be the partial derivative with respect to x_i and then x_j. These concepts will be explained later in the course.

A function f will be strictly convex iff it satisfies

$$f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y), \ \forall x, y \in \mathbb{R}^n, \ 0 < \theta < 1$$

Notice the strict inequality in this case. In terms of second derivative, the second derivative should be strictly positive.

1.3 Examples

Lets take a look at some other classes of functions which are convex.

- Affine functions: Any function of the form $Ax + b$ is convex. The proof is a simple exercise.
- Exponential functions: A function of the form e^{ax} is convex.
- The absolute value function is convex. Easiest way is to look at the epigraph.
- Any norm defined on the vector space is convex. Why?
- The function $f(x) = \frac{1}{x^2}$ is convex on the positive real line.

Exercise 2. Give an example of the function which is monotonically increasing but not convex.

2 Global optimality and local optimality

For a general function there are two kind of optimal points. There is *global optimum*, which is the general notion of being optimal as compared to any other point on the domain. Hence a point x^* is globally optimal (say minimum) for a function f iff $f(x) \geq f(x^*)$ for any x in domain of f. In general it is very hard to find such points.

On the other hand, there is a concept of *local optimum*, where the point has minimum value among all values in the neighborhood. Hence, a point x^* is a local optimum point for a function f iff for some ϵ,

$$f(x^*) \leq f(x), \; \forall x \in B_\epsilon(x^*)$$

Here $B_\epsilon(x^*)$ is the ball of radius ϵ around x^*. As compared to global optimum, it is much easier to check for a local optimum. For instance, we can check that the derivative in every direction is positive and continuous.

The special property of convex functions is that a locally optimal point is globally optimal too. This again makes it much easier to optimize a convex function. The proof is given for a single variable function, but the theorem works for multi-variable function too.

Proof. Suppose point x^* is locally optimal. Since the function is convex, for any point y,

$$f(y) \geq f(x^*) + f'(x^*)(y - x^*).$$

This can be written as,

$$f(y) \geq f(x^*) + (\lim_{t \to 0} \frac{f(x^* + t(y - x^*)) - f(x^*)}{t(y - x)})(y - x),$$

using the definition of first derivative. This can be simplified to,

$$f(y) \geq f(x^*) + (\lim_{t \to 0} \frac{f(x^* + t(y - x^*)) - f(x^*)}{t}).$$

Clearly the second term on the right is positive because x^* is a local optimum point. This implies, for any y in the domain,

$$f(y) \geq f(x^*)$$

So x^* is globally optimal too. □

3 Jensen's inequality

One of the most important properties of convex function is *Jensen's inequality*. Given a random variable X and a convex function f,

$$E(f(X)) \geq f(E(X))$$

This can be proved using the straightforward extension of the basic inequality for convex functions.

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

This innocent looking inequality can be used to prove other inequalities by using different convex functions.

- AM-GM: Since the function $\log x$ is concave (it satisfies Jensen's inequality in opposite direction). Applying Jensen's inequality on it for a random variable which takes values a and b with equal probability.

  $$\begin{aligned} \log \frac{a+b}{2} &\leq \frac{\log a + \log b}{2} \\ \Rightarrow \quad \frac{a+b}{2} &\leq \sqrt{e^{\log a + \log b}} \\ \Rightarrow \quad \frac{a+b}{2} &\leq \sqrt{ab} \end{aligned}$$

  This shows the arithmetic mean-geometric mean inequality.

- Suppose $x_1, \cdots, x_k$ are some positive numbers. Say $p_1, \cdots, p_k$ is a probability distribution ($\sum_i p_i = 1,\ p_i \geq 0$). Then we need to show,

$$\mathit{\Pi}_i x_i^{p_i} \leq \sum_i p_i x_i.$$

Proof. Assume that $x_i = e^{y_i}$ (why can we assume that?). Then the inequality turns into,

$$\mathit{\Pi}_i e^{\sum_i y_i p_i} \leq \sum_i p_i e^{y_i}.$$

This is Jensen's inequality for function $f(x) = e^x$. □

4 Constructing convex functions from other convex functions

There are many ways in which you can combine convex functions to obtain another convex function.

- Summation of two convex functions is convex. This can be extended to any conic combination of convex functions,

$$\theta_1 f_1 + \cdots + \theta_k f_k,$$

is convex if $\theta_i \geq 0$ and $\sum_i \theta_i = 1$.

Exercise 3. Give two convex functions whose multiplication is not convex.

- Maximum: Given functions $\{f_i\}$, all of which are convex, then their point-wise maximum $f(x) = \max\{f_1(x), \cdots, f_k(x)\}$, is also convex.
- Lets consider a composition of two convex functions for the single variable case. In general, it need not be convex. But if the inner function is non-decreasing, then the composition is convex. This can be proved by checking the first and second derivative (applying the chain rule).

$$\begin{aligned} f &= f_1(f_2) \\ \Rightarrow \quad f' &= f_1'(f_2) f_2' \\ \Rightarrow \quad f'' &= f_1''(f_2)(f_2')^2 + f_1'(f_2) f_2'' \end{aligned}$$

So f'' is greater than zero if f_1, f_2 are convex and f_1' is non-decreasing.