

CS 49/149: 21st Century Algorithms (Fall 2018): Lecture 4

Date: 25th September, 2018

Topic: A quick introduction to convex sets and functions

Scribe: Devina Kumar

Disclaimer: These notes have not gone through scrutiny and in all probability contain errors. Please email errors to Devina.Kumar.GR@dartmouth.edu.

1 Introduction to Convexity

We first define some terms.

convex function

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if $\forall x, y, f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$ for $\lambda \in [0, 1]$

convex set

$S \subset \mathbb{R}^n$ is a convex set if $\forall x, y \in S, \lambda x + (1 - \lambda)y \in S$

Let us try to think about what the definition of a convex function might mean geometrically. Suppose we have two points, x and y , and a convex function f . We plot $(x, f(x))$ and $(y, f(y))$ and then connect them with a line. If f is convex, then between x and y , the line will lie above the graph of f .

Examples of convex functions. Now that we've defined what a convex function is, let's take a look at a couple examples of convex functions.

$$f(x) = x^2$$

$$\begin{aligned} & (\lambda x + (1 - \lambda)y)^2 \leq \lambda x^2 + (1 - \lambda)y^2 \\ & \lambda^2 x^2 + 2\lambda(1 - \lambda)xy + (1 - \lambda)^2 y^2 \leq \lambda x^2 + (1 - \lambda)y^2 \\ & \lambda^2 x^2 + 2\lambda(1 - \lambda)xy + (1 - \lambda)^2 y^2 \leq \lambda x^2 + y^2 - \lambda y^2 \\ & \lambda^2 x^2 - \lambda x^2 + 2\lambda(1 - \lambda)xy + (1 - \lambda)^2 y^2 - y^2 + \lambda y^2 \leq 0(\lambda^2 - \lambda)x^2 - 2(\lambda^2 - \lambda)xy + (\lambda^2 - \lambda)y + 2 \leq 0 \\ & (\lambda^2 - \lambda)(x - y)^2 \leq 0 \end{aligned}$$

$$(fx) = a^T x - b$$

$$\begin{aligned} f(\lambda x + (1 - \lambda)y) &= a^T(\lambda x + (1 - \lambda)y) - b \\ &= \lambda a^T x + (1 - \lambda)a^T y - (\lambda b + (1 - \lambda)b) \\ &= \lambda(a^T x - b) + (1 - \lambda)(a^T y - b) \\ &= \lambda f(x) + (1 - \lambda)f(y) \end{aligned}$$

We showed that these functions were convex by substituting the function into our definition of convexity and showing that the inequality held true. However, in some cases, we can prove convexity by looking at the second derivative of the function in question.

Theorem 1. If $f : \mathbb{R} \rightarrow \mathbb{R}$ and it is twice differentiable, then f is convex if and only if $\frac{\delta^2 f}{\delta x} \geq 0$ everywhere.

Proof. Consider i, j, k in a function space where $i < j < k$. By definition of a convex function, the slope between i and j is \leq the slope between j and k . If this were not true, then the line segment between i and k would be under the function for some interval, which contradicts the definition of a convex function. Thus, we observe that the slope is staying the same or increasing from i to k , meaning that the second derivative is nonnegative.

Now suppose that the second derivative is nonnegative. Let us look at the interval between i and k . Suppose that at some j between i and k , there is a point $(j, f(j))$ that is above the line segment between $(i, f(i))$ and $(k, f(k))$. From the Mean Value Theorem, we know that between i and j , there exists a point where the first-order derivative is equal to the average rate of change between i and j . We note that the average rate of change between i and j is greater than that between i and k . However, if our second derivative is nonnegative, then this would not hold, because once the function would reach the point $(j, f(j))$, the derivative would increase, meaning that we would end up at a point above $(k, f(k))$ (which we could only reach if the derivative decreased at some point). Thus, if the derivative is always increasing, we will not have a $(j, f(j))$ that is above the line segment between $(i, f(i))$ and $(k, f(k))$, meaning that the function is convex. \square

We can use this theorem to easily show convexity for other functions, such as $f(x) = x \ln(x)$ for $x \in [0, \text{inf}]$. We can also use some other theorems when doing analysis of convex functions.

Theorem 2. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if and only if function $g : \mathbb{R} \rightarrow \mathbb{R}$ where $g(z) = f(x + zy)$ is convex for $x \in \text{dom}(f)$ and $y \in \mathbb{R}^n$.

Theorem 3. The sum of convex functions is also convex.

With these properties in mind, let us look at some more functions. Consider $f(x) = \max(x_1, x_2, \dots, x_n)$ where $x \in \mathbb{R}^n$. Is this convex?

Suppose that we have some x, y , and λ that contradict the inequality that defines convexity. So, we have some points x_i, y_i where $i \in 1, \dots, n$ such that $f(\lambda x_i + (1 - \lambda)y_i) > \lambda f(x) + (1 - \lambda)f(y)$. However, let us take a closer look at the function. We can say $f(\lambda x_i + (1 - \lambda)y_i) = \lambda x_i + (1 - \lambda)y_i$ (because the function is just the maximum). However, we note that $\lambda x_i + (1 - \lambda)y_i$ is $\leq \lambda \max(x) + (1 - \lambda)\max(y)$, which is $\lambda f(x) + (1 - \lambda)f(y)$. Thus, the condition $f(\lambda x_i + (1 - \lambda)y_i) \leq \lambda f(x) + (1 - \lambda)f(y)$ holds and the function is convex.

What about $f^*(x) = \max(|x_1|, |x_2|, \dots, |x_n|)$? Well, let us take $f^*((\lambda x_i + (1 - \lambda)y_i))$ where x_i, y_i satisfy $f^*((\lambda x_i + (1 - \lambda)y_i)) > \lambda f^*(x) + (1 - \lambda)f^*(y)$. In this case, we get that $f^*((\lambda x_i + (1 - \lambda)y_i)) = |\lambda x_i + (1 - \lambda)y_i|$. Now, using the triangle inequality, we get that $|\lambda x_i + (1 - \lambda)y_i| \leq |\lambda x_i| + |(1 - \lambda)y_i|$. Now, using the fact that norms are absolutely scaleable, we can say that $|\lambda x_i| + |(1 - \lambda)y_i| = (\lambda)|x_i| + (1 - \lambda)|y_i|$. Thus, the condition $f^*((\lambda x_i + (1 - \lambda)y_i)) \leq \lambda f^*(x) + (1 - \lambda)f^*(y)$ holds and the function is convex. In fact, we can apply this to norms in general to see that all norms are convex functions.

Theorem 4. All norms are convex functions.

Now, let us consider a set of points related to a convex function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. This set of points is the set of points lying on or above the graph of f . We refer to this set of points as the epigraph of f , or $\text{epi}(f)$. We define $S_f = \text{epi}(f) \subseteq \mathbb{R}^{n+1} = \{(x, t) : t \geq f(x)\}$.

CLAIM: f is a convex function if and only if $\text{epi}(f) \subseteq \mathbb{R}^{n+1}$ is a convex set.

Let us take two points (x, t) and (y, s) in $\text{epi}(f)$ and $\lambda \in [0, 1]$. We define (w, z) as $\lambda(x, t) + (1 - \lambda)(y, s)$. We thus have:

$$\begin{aligned} z &= \lambda t + (1 - \lambda)s \\ &\geq f(x) + (1 - \lambda)f(y) \text{ (because the points are in the epigraph)} \\ &\geq f(\lambda x + (1 - \lambda)y) \\ &= f(w) \end{aligned}$$

So, we have shown that the points line in the epigraph. ▢

Question: Show that if the epigraph is convex, the function is convex as well.

1.1 Hahn-Banach Theorem

We continue our discussion about convex sets by introducing the Hahn-Banach Theorem. Suppose we have some convex set S and some point z that lies outside of S . Then, we can slice the space so that all of S lies on one side of the slice, and z lies on the other. This is the intuition behind the Hahn-Banach Theorem, which we formalize below.

Theorem 5. Let S be in some \mathbb{R}^n be a closed convex set and let z not be in S . Then, there exists $w \in \mathbb{R}^n, \theta \in \mathbb{R}$ such that:

1. $w^T x \geq \theta$ for all $x \in S$
2. $w^T z < \theta$

This concept, also referred to as the separation theorem, supports the notion of the supporting hyperplane. Suppose we have a convex set S , and a point z that lies on the boundary of S . Then, there is a $w \in \mathbb{R}^n$ such that $w^T x \geq w^T z$ for all $x \in S$. This vector that defines the supporting hyperplane defines a subgradient, which can be thought of as a linear approximation of f at a point z that underestimates a function.

Definition: $g \in \mathbb{R}^n$ is a subgradient of f at x if

$$\begin{aligned} f(y) &\geq f(x) + (y - x)^T g \text{ for all } y \\ &= f(x) + \sum_{j=1}^n g_j(y_j - x_j) \end{aligned}$$

Corollary 1. If f is convex, then f has a subgradient at all points of x . The set of subgradients of f at x is denoted by $\delta f(x)$.

Consider the definition of a separating hyperplane. If f is convex, then let us consider the epigraph of f . From the separating hyperplane, we know that for the points on the boundary of the convex set $\text{epi}(f)$, there is a hyperplane that can separate the set from points outside the set, and there is a hyperplane that hits the set at a boundary point x and the vector defining that hyperplane is a subgradient of f at x .

1.2 Convex Optimization

Convex optimization has to do with minimizing a convex function $f(x)$ where $x \in S$ and S is a convex set. These types of problems are very common. For example, linear regression is a special case of convex optimization. Unconstrained convex optimization generally involves $\min(f(x)), x \in \mathbb{R}^n$. The nice thing about convexity, however, is that x is a local optimum, so $\exists \epsilon \forall z$ such that $\|z - x\|_2 \leq \epsilon$ means $f(x) \leq f(z)$.

CLAIM: x is a global optimum

Proof. Now, suppose that there is a point z such that $f(z) < f(x)$. Because our set S is convex, we can say that $\lambda x + (1 - \lambda)z \in S$. Because f is convex, we get the following.

$$\begin{aligned} f(\lambda x + (1 - \lambda)z) &\leq \lambda f(x) + (1 - \lambda)f(z) \\ &< \lambda f(x) + (1 - \lambda)f(x) = f(x) \end{aligned}$$

However, we notice that $\lambda x + (1 - \lambda)z$ approaches x as λ approaches 1, which would then contradict $f(\lambda x + (1 - \lambda)z) < f(x)$ because it would mean that x is not a local optimum. Thus, x is a global optimum. \square

Now, say that we are at an arbitrary point x where $f(x)$ may or may not be a local optimum. Because $f(x)$ is convex, there is a linear function that underapproximates this function.

CLAIM: if f is convex, $\nabla f(x)$ is a subgradient.

Proof. We want to show for the y values we are interested in, $(y-x)^T \nabla f(x) = \lim_{\delta \rightarrow 0} \frac{f(x + \delta(y-x)) - f(x)}{\delta}$.

$$\lim_{\delta \rightarrow 0} \frac{f(x + \delta(y-x)) - f(x)}{\delta} \leq \frac{\delta f(y) + (1 - \delta)f(x) - f(x)}{\delta}$$

From this, we can cancel terms and divide to get $f(y) - f(x)$. This implies that $f(y) - f(x) \geq (y-x)^T \nabla f(x)$, which then implies $f(y) \geq f(x) + (y-x)^T \nabla f(x)$, thus showing that $\nabla f(x)$ is a subgradient. \square



Question: Is $\nabla f(x)$ the only subgradient?