SC 607: Optimization **Spring 2019**

Lecture 15: March 8 2019

Instructor: Ankur A. Kulkarni *Scribes: Arun Kumar Miryala, Anuraag Tummanapally*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

15.1 Review

In the previous lecture, a necessary and sufficient condition for x^* to be a minimizer for convex optimization problem is derived. A condition on local minimizer x^*, where x^* is constrained over the tangent cone set $T(x^*, S)$, S need not be convex is found. Let us recall some of the concepts that were discussed.

Theorem 15.1. *Let S be a convex set and let f be a continuously differentiable (C^1) convex function. Consider the following convex optimization problem:*

$$\text{minimize } f(x), \tag{15.1}$$
$$\text{subject to } x \in S \tag{15.2}$$

Then x^* is minimizer if and only if $\nabla f(x^*)^\top (y - x^*) \geqslant 0$, for all $y \in S$.

Definintion 15.1. *For $x^* \in S$, the tangent cone $T(x^*; S)$ is defined as:*

$$T(x^*, S) := \left\{ d \;\middle|\; \exists\, x_k \subseteq S : x_k \to x^* \;\&\; t_k \downarrow 0 \text{ s.t. } d = \lim_{k \to \inf} \frac{x_k - x^*}{t_k} \right\}. \tag{15.3}$$

Tangent cone of a set captures the precise shape of the set.

Theorem 15.2. *Consider the optimization problem:*

$$\begin{gathered} \text{Minimize } f(x) \\ \text{subject to } x \in S \end{gathered}$$

where $f \in C^1$. If x^* is a local minimum, then $\nabla f(x^*)^\top d \geqslant 0$, $\forall d \in T(x^*, S)$.

15.2 Introduction

In this lecture, we will derive a necessary and sufficient condition for x^* to be a minimizer for convex optimization problem and also find a condition on local minimizer x^*, where x^* is constrained over the tangent cone set $T(x^*, S)$, where S need not be convex.

15.3 Minimizer of Convex Optimiztion Problem

Proof. (Sufficiency)
Suppose that $x^* \in S$ satisfies
$$\nabla f(x^*)^\top (y - x^*) \geqslant 0, \ \forall\, y \in S. \tag{15.4}$$
By the Lemma 14.1,
$$f(y) \geqslant f(x^*) + \nabla f(x^*)^\top (y - x^*), \ \forall\, x^*, y. \tag{15.5}$$
Combining (15.4) and (15.5) gives
$$f(y) \geqslant f(x^*), \ \forall\, y \in S. \tag{15.6}$$
This implies that x^* is a global minima.

(Necessity)
Let x^* be a global minima, which implies $f(x^*) \leqslant f(y), \forall\, y \in S$. Suppose $\nabla f(x^*)^\top (y - x^*) < 0$, for some $y \in S$.
Let $z(t) = x^* + t(y - x^*)$. We will show that $f(z(t)) < f(x^*)$ for small enough t.
Consider
$$\frac{f(z(t)) - f(x^*)}{t} = \frac{f(x^* + t(y - x^*)) - f(x^*)}{t} \tag{15.7}$$
Taking limit as $t \to 0$ gives,
$$\begin{aligned} \lim_{t \to 0} \frac{f(x^* + t(y - x^*)) - f(x^*)}{t} &= \nabla f(x^*)^\top (y - x^*) \\ &< 0 \end{aligned} \tag{15.8}$$
from initial assumption. Hence, there exits t small enough such that $f(z(t)) < f(x^*)$. But this contradicts that x^* is a global minima. Therefore, $\nabla f(x^*)^\top (y - x^*) \geqslant 0, \ \forall\, y \in S$. □

Remarks.

1. *Geometrically, the above condition imposes that gradient of the function f at x^* should make an acute angle with vector $(y - x^*)$, $\forall\, y \in S$.*
2. *If f is C^1 and convex function and x^* satisfies $\nabla f(x^*)^\top (y - x^*) \geqslant 0, \ \forall\, y \in S$, then x^* is global minima over S.*
3. *If S is convex set, f is C^1 and x^* is a local minima, then $\nabla f(x^*)^\top (y - x^*) \geqslant 0, \ \forall\, y \in S$.*
4. *In particular, if $S = \mathbb{R}^n$ and x^* is local minima, then $\nabla f(x^*) = 0$.*

15.4 Tangent Cone

15.4.1 A more relaxed necessary condition for local minima over non-convex set S

Before we state the next theorem, we provide some preliminary background regarding the $O(t)$ and $o(t)$ notations.

Consider two sequences $\{a_k\}$ and $\{b_k\}$. We know that $\lim_{k\to\infty} \frac{a_k}{b_k} = L \in (0,\infty) \implies$ For large enough k, $\frac{a_k}{b_k} \approx L$, i.e., $a_k = Lb_k + o(b_k) \implies \frac{a_k}{b_k} = L + \frac{o(b_k)}{b_k}$. Thus, for $\lim_{k\to\infty} \frac{a_k}{b_k} = L \in (0,\infty)$ to even hold in the first place, we must have $\frac{o(b_k)}{b_k} \downarrow 0$ as $b_k \downarrow 0$.

Note that $o(t) :=$ term such that $\frac{o(t)}{t} \to 0$ as $t \to 0$ and $O(t) :=$ term such that $\frac{O(t)}{t} \to$ some finite constant $\neq 0$ as $t \to 0$. With slight abuse of notation, $o(1) :=$ a quantity $\to 0$ as $t \to 0$ and $O(1) :=$ a quantity $\to$ some constant $\neq 0$ as $t \to 0$.

With some simple calculations, we can show that: $o(t) \times o(t) \leqslant o(t)$, $o(t) \times O(t) = o(t)$, $O(t) \times O(t) = O(t^2)$. So also, $o(t) + o(t) = o(t)$, $o(t) + O(t) = O(t)$, $O(t) + O(t) = O(t)$.

With this background, we now come to the main statement of the theorem.

Proof. Suppose not, i.e., $\exists\, d \in T(x^*, S)$ such that $\nabla f(x^*)^\top d < 0$. $\exists\,\{x_k\} \to x^*$ and $t_k \downarrow 0$ such that $t_k d = (x_k - x^*) + o(t_k)$. Rearranging the terms, we get, $x_k = x^* + t_k d + o(t_k)$. Recall Taylor's theorem: $f(x+p) = f(x) + \nabla f(x+tp)^\top p$ for some $t \in (0,1)$. We thus obtain:

$$f(x_k) = f(x^* + t_k d + o(t_k)) = f(x^*) + \nabla f(x^* + \delta(t_k d + o(t_k)))^\top (t_k d + o(t_k)), \text{ where } \delta \in (0,1).$$

Note that, as $k \to \infty$, $t_k \downarrow 0$, hence the $\delta(.)$ term in $\nabla f(.)$ vanishes. This leaves us with $\nabla f(x^*)$, which is in fact $O(1)$. Since $O(1) \times o(t_k) = o(t_k)$, we get $f(x_k) = f(x^*) + \nabla f(x^* + \delta(t_k d + o(t_k)))^\top t_k d + o(t_k)$. This can be written as $f(x_k) = f(x^*) + \nabla f(x^*)^\top t_k d + \Big[\nabla f(x^* + \delta(t_k d + o(t_k))) - \nabla f(x^*)\Big]^\top t_k d + o(t_k)$. Once again notice that the term $[.]$ vanishes as $t_k \downarrow 0$, and therefore, $[.] t_k d$ becomes $o(t_k)$.

So, we are finally left with:

$$f(x_k) = f(x^*) + \nabla f(x^*)^\top t_k d + o(t_k) = f(x^*) + t_k \left[\nabla f(x^*)^\top d + \frac{o(t_k)}{t_k}\right].$$

We know that $\frac{o(t_k)}{t_k} \downarrow 0$ as $t_k \downarrow 0$. Also, we started our proof with the assumption that $\nabla f(x^*)^\top d < 0$. This makes the term $\left(t_k \left[\nabla f(x^*)^\top d + \frac{o(t_k)}{t_k}\right]\right) < 0$ as $t_k \downarrow 0$, which in turn implies that for large k, $f(x_k) < f(x^*)$. This is contradictory to the fact of x^* being the local minimum. Hence, our initial assumption was wrong and indeed $\nabla f(x^*)^\top d \geqslant 0$, $\forall d \in T(x^*, S)$. $\square$