KTH, SCHOOL OF ELECTRICAL ENGINEERING

On Optimality Condition for Constrainted Convex Optimization Problem

Yuchao Li

September 11, 2017

1 PROBLEM STATEMENT

A *constrainted convex optimization problem* can be formulated as

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x \in X \end{aligned} \tag{1.1}$$

where x is the *decision vector*, $f(x)$ is the *objective function* which is *convex* and *differentiable* and X is the *feasible set*. Denote the optimal decision vector as $x^\star$. According to **Lecture Note 2**, it must satisfy

$$\langle \nabla f(x^\star), y - x^\star \rangle \geq 0 \qquad \forall y \in X. \tag{1.2}$$

which is both a necessary and sufficient condition. Elaborate the necessity and sufficiency of the statement.

2 ELABORATION

First, we prove the sufficiency. According to [1] and **Lecture Note 2**, for any convex function $f(x)$, it must holds that

$$f(y) \geq f(x) + (y - x)^T \nabla f(x) \qquad \text{for all } x, y \in \text{dom } f \tag{2.1}$$

Since $(y - x^\star)^T \nabla f(x^\star) \geq 0$, we have

$$f(y) - f(x^\star) \geq (y - x^\star)^T \nabla f(x^\star) \implies f(y) - f(x^\star) \geq 0 \quad \forall y \in X$$

Namely $x^\star$ returns optimal value of $f(x)$.
Second, we prove the necessity. We start by introducing a proposition.

Proposition. *If $f(x)$ is differentiable, then*

$$\lim_{\lambda\to 0}\frac{f(x+\lambda(y-x))-f(x)}{\lambda}=(y-x)^T\nabla f(x). \tag{2.2}$$

Proof. Since $f(x)$ has total derivative, according to [2], for small enough λ, we have

$$f(x+\lambda(y-x))-f(x)=\lambda(y-x)^T\nabla f(x)+o(\rho)$$

where $o(\rho)$ denotes *higher order infinitesimal* of ρ and $\rho=|\lambda|\|y-x\|$. Then we have

$$\frac{f(x+\lambda(y-x))-f(x)}{\lambda}=(y-x)^T\nabla f(x)+\frac{o(\rho)}{\lambda} \tag{2.3}$$

Since we have

$$\lim_{\lambda\to 0}\frac{o(\rho)}{\lambda}=0$$

then take limits on both sides of Eq. (2.3), the proof is concluded. □

We know that $f(y)\geq f(x^\star)$ for all $y\in X$. Suppose for some $y\in X$, we have

$$(y-x^\star)^T\nabla f(x^\star)<0. \tag{2.4}$$

Denote $w=(1-\theta)x^\star+\theta y$ where $\theta\in(0,1]$, then we have

$$(w-x^\star)^T\nabla f(x^\star)=((1-\theta)x^\star+\theta y-x^\star)^T\nabla f(x^\star)=\theta(y-x^\star)^T\nabla f(x^\star)<0.$$

Again we denote $z=(1-\theta)x^\star+\theta w$, we would have

$$(z-x^\star)^T\nabla f(x^\star)=\theta^2(y-x^\star)^T\nabla f(x^\star)<0.$$

This iteration can go on and on, which indicates that if there is some y fullfil Eq. (2.4), then all points on the line segment $x^\star y$ (except $x^\star$ itself) fullfil Eq. (2.4). According to the proposition above, we have

$$\lim_{n\to\infty}\frac{f(x^\star+\theta^n(y-x^\star))-f(x^\star)}{\theta^n}=(y-x^\star)^T\nabla f(x^\star)<0. \tag{2.5}$$

Denote $\beta=(y-x^\star)^T\nabla f(x^\star)$, then $\beta<0$ and for $0<\sigma<-\beta$, there must exists a N so that for all $n>N$, $|(f(x^\star+\theta^n(y-x^\star))-f(x^\star))/\theta^n-\beta|<\sigma$. Namely, there are some $u=x^\star+\theta^n(y-x^\star)\in X$ that $f(u)<f(x^\star)$ which contradicts the assumption. Therefore, there is no such y to make $(y-x^\star)^T\nabla f(x^\star)<0$, which concludes the proof.

References

[1] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[2] Mathematics Department, *Advanced mathematics (Chineses)*, Higher Education Press, 2007.