Optimality Conditions in Convex Optimization Revisited

Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur
Kanpur-208016
India

C. S. Lalitha
Department of Mathematics
University of Delhi
Delhi-208016
India

FIRST DRAFT

KanGAL Report Number 2010002

Abstract

The phrase convex optimization refers to the minimization of a convex function over a convex set. However the feasible convex set need not be always described by convex inequalities. In this article we consider a convex feasible set which are described by inequality constraints which are locally Lipschitz and not necessarily convex and need not be smooth. We show that if the Slater's constraint qualification and a simple non-degeneracy condition is satisfied then the Karush-Kuhn-Tucker type optimality condition is both necessary and sufficient.

1 Introduction

This article is motivated by the recent paper of Lasserre [4]. In this paper Lasserre considers a smooth convex function to be minimized over a convex set. However unlike the traditional setting where the convex feasible set of a convex optimization problem is often described by convex inequalities, in [4] the convex feasible set is described by inequality constraints which are smooth but not necessarily convex. It is well know that if the inequality constraints are convex and differentiable and the Slater constraint qualification is satisfied then the Karush-Kuhn-Tucker (KKT) optimality conditions are both necessary and sufficient. Lasserre [4] showed that even if the convex feasible set is not described by convex inequality constraints, the Slater constraint qualification along with a mild non-degeneracy conditions renders the KKT conditions both necessary and sufficient. In order to motivate we describe the work of Lasserre [4] in slightly more detailed manner.
Consider the problem of minimizing a convex function $f : \mathbb{R}^n \to \mathbb{R}$ over a convex set K. The convex set K is described as follows

$$K = \{x \in \mathbb{R}^n : g_i(x) \leq 0, \quad i = 1, \dots, m\}, \tag{1}$$

where each g_i is a smooth function but not necessarily convex. As shown in [**?**] it is simple to observe that the following set in $\mathbb{R}^2$,

$$K = \{x \in \mathbb{R}^2 : 1 - x_1 x_2 \leq 0, \quad x_1 \geq 0, \quad x_2 \geq 0\},$$

is convex the constraint the function $1 - x_1 x_2$ is not convex though smooth. In order to prove the necessity and sufficiency of the KKT conditions in such a case Lasserre [4] considered the following non-degeneracy condition. The convex feasible set K is said to satisfy the non-degeneracy condition if for all $i = 1, \ldots, m$ we have

$$\nabla g_i(x) \neq 0, \quad \text{whenever} \quad x \in K \quad \text{and} \quad g_i(x) = 0.$$

The main result in Lasserre [4] can be stated as follows.

Theorem 1.1 *Let us consider the minimization of a smooth convex function f over a convex set K given by (1) where the functions g_i are smooth but need not be convex. Assume that the Slater condition and the non-degeneracy condition holds. Then the KKT condition is both necessary and sufficient.*

Thus Lasserre [4] concludes that as far as KKT conditions in smooth convex optimization is concerned it the convexity of the feasible set is a more important feature than its representation by smooth convex inequalities. In this article we consider the case when f is a nondifferentiable convex function and the convex set K is described by locally Lipschitz inequality constraints which are not necessarily differentiable. It is natural to ask to what extent the framework developed by Lasserre [4] can be extended to this case. We will show that Lasserre's framework can be extended to the nonsmooth setting if we consider the locally Lipschitz function representing the set K to be regular in the sense of Clarke [3]. We will introduce a suitable non-degeneracy condition in the nonsmooth setting in order to prove that the nonsmooth KKT condition is both necessary and sufficient. We also would like to point out before hand that the necessary optimality condition in our setting is of mixed type since it is represented through the subdifferential of f and the Clarke subdifferential of $g_i's$. For details on the subdifferential of a convex function see for example Rockafellar [5], Bertsekas [1], Borwein and Lewis [2]. For details We will present our main results with examples in the next section. We will end this section by stating some notations that will be used in the sequel. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function then $\partial f(x)$ denotes the subdifferential of f at x. The Clarke subdifferential of a locally Lipschitz function $g : \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$ is given as $\partial^\circ g(x)$. The directional derivative of g at x in the direction v is denoted as $g'(x, v)$ and the Clarke directional derivative of a locally Lipschitz function g at x in the direction v is denoted as $g^\circ(x, v)$. For more details on the Clarke directional derivative and its relationship with the Clarke subdifferential see Clarke [3]. A locally Lipschitz function g is said to be *regular* in the sense of Clarke (see Clarke [3]) at a point x if g is directionally differentiable at x in all the directions v and $g^\circ(x, v) = g'(x, v)$ in all directions v. It is important to note that both the subdifferential of a convex function and the Clarke subdifferential of a locally Lipschitz function are compact convex sets. Further it is important to note that some important class of locally Lipschitz functions are regular. For example consider the function

$$f(x) = \max\{f_1(x), \ldots, f_m(x)\}$$

where each f_i is a smooth function. Then f is a locally Lipschitz function regular in the sense of Clarke [3].

2 Main Results

We would again like to recall that we are studying the problem of minimizing a nondifferentiable convex function over a convex set K which is represented through locally Lipschitz inequality constraints ie

$$K = \{x \in \mathbb{R}^n : g_i(x) \leq 0, \quad i = 1, \ldots, m\}, \tag{2}$$

where each g_i is a locally Lipschitz function which need not be differentiable. For example consider the set K_1 given as

$$K_1 = \{x \in \mathbb{R} : \max\{x^3, x\} \leq 0\}.$$

The set $K_1 = \{x : x \leq 0\}$ and hence convex. Further note that the function $\max\{x^3, x\}$ is a regular function in the sense of Clarke. The notion of regular functions as we will see will play a pivotal role here. We begin by introducing the nonsmooth degeneracy condition which we call as Assumption(A).

Definition 2.1 *Consider the set K given by (2) where each g_i is a locally Lipschitz function. The set K is said to satisfy the Assumption (A) if for all $i = 1, \ldots, m$,*

$$0 \notin \partial^\circ g_i(x), \quad \textit{whenever} \quad x \in K \quad \textit{and} \quad g_i(x) = 0.$$

Let us now provide an example to show where such a condition is fulfilled and another example showing where it is not fulfilled. Consider the following set

$$K_2 = \{x \in \mathbb{R} : \max\{x^3, x\} - 1 \leq 0\}$$

Observe that $K_2 = \{x \in \mathbb{R} : x \leq 1\}$. Let us set $g(x) = \max\{x^3, x\} - 1$. Then $g(1) = 0$ and $\partial^\circ g(1) = [1, 3]$. Thus the Assumption (A) holds for K_2. Also observe that $g(x)$ is regular in the sense of Clarke [3]
Now consider the set K_3 given as

$$K_3 = \{x \in \mathbb{R} : \min\{x^2, x\} \leq 0\}.$$

It is clear that $K_3 = \{x \in \mathbb{R} : x \leq 0\}$. Let us now set $g(x) = \min\{x^2, x\}$. Then $g(0) = 0$ and $\partial^\circ g(0) = [0, 1]$. Hence Assumption(A) is not satisfied for K_3.
We will now state the following characterization of a convex set in terms of the Clarke directional derivative.

Proposition 2.2 *Let the set K be given by (2), i.e. represented by locally Lipschitz inequality constraints. Assume that each g_i is regular in the sense of Clarke. Let the Slater constraint qualification hold i.e. there exists $\hat{x}$ such that $g_i(\hat{x}) < 0$ for all $i = 1, \ldots, m$. Further assume that K satisfies Assumption(A). Then K is convex if and only if for every $i = 1, \ldots, m$,*

$$g_i^\circ(x, y - x) \leq 0, \quad \textit{for all} \quad x, y \in K \quad \textit{with} \quad g_i(x) = 0. \tag{3}$$

Proof : Let us first assume that K is convex. On the contrary assume that (3) does not hold. Hence there exists $r \in \{1, \ldots, m\}$ and $x, y \in K$ such that $g_r(x) = 0$ and $g_r^\circ(x, y - x) > 0$. Since g_r is regular in the sense of Clarke we have $g_r'(x, y - x) > 0$. This shows that $g_r(x + \lambda(y - x)) > 0$ for all $\lambda > 0$ sufficiently small. Since K is convex $x + \lambda(y - x) \in K$ for $\lambda \in (0, 1)$ sufficiently small. This is a contradiction since $g_r(x + \lambda(x - y)) > 0$ shows that $x + \lambda(y - x) \notin K$.
Conversely assume that (3) is satisfied and we have to show that K is a convex set. Since the Slater constraint qualification holds we conclude that K has an interior. Now consider any boundary point of $x \in K$. Thus there exists an $j \in \{1, \ldots, m\}$ such that $g_j(x) = 0$. Since (3) holds we have $g_j^\circ(x, y - x) \leq 0$ for all $y \in K$. Now from Clarke [3] we know that for any $\xi_j \in \partial^\circ g_j(x)$

$$\langle \xi_j, y - x \rangle \leq 0 \quad \forall y \in K.$$

Since Assumption (A) holds we see $\xi_j \neq 0$ and hence there a non-trivial supporting hyperplane to K at x. Hence from Theorem 1.3.3 of Schneider we have that K is convex. Hence the result $\square$
We are now in a position to state our main result.

Theorem 2.3 *Let us consider the problem of minimizing the convex function $f : \mathbb{R}^n \to \mathbb{R}$ over the the convex set K. Let us assume that K is given by (2) and each g_i is a locally Lipschitz function and regular in the sense of Clarke. Further assume that the Slater constraint qualification holds and the set K satisfies the Assumption (A). Then $\bar{x} \in K$ is a global minimum of f over K if and only if there exists scalars $\lambda_i \geq 0$ such that*

i) $0 \in \partial f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \partial^\circ g_i(\bar{x})$

ii) $\lambda_i g_i(\bar{x}) = 0, \quad \forall i = 1, \dots, m$

Proof : Let $\bar{x} \in K$ be a minimizer of f over K. Since the convex function f is locally Lipschitz we know from Clarke [3] that there exists $\lambda_0 \geq 0, \lambda_1 \geq 0, \dots, \lambda_m \geq 0$ not all simultaneously zero, such that

i) $0 \in \lambda_0 \partial^\circ f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \partial^\circ g_i(\bar{x})$

ii) $\lambda_i g_i(\bar{x}) = 0 \quad \forall i = 1, \dots, m.$

Since $\partial^\circ f(\bar{x}) = \partial f(\bar{x})$ (see Clarke [3]) we have

i) $0 \in \lambda_0 \partial f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \partial^\circ g_i(\bar{x})$

ii) $\lambda_i g_i(\bar{x}) = 0 \quad \forall i = 1, \dots, m.$

We shall now show that using the Slater constraint qualification and the Assumption (A) we will show that $\lambda_0 > 0$. To begin with let us observe that using support function calculus we can write the optimality conditions above as

i) $\lambda_0 f'(\bar{x}, h) + \sum_{i=1}^{m} \lambda_i g_i^\circ(\bar{x}, h) \geq 0 \quad \forall h \in \mathbb{R}^n$

ii) $\lambda_i g_i(\bar{x}) = 0 \quad \forall i = 1, \dots, m.$

Let us assume that λ_0. Hence from i) immediately above we have

$$\sum_{i=1}^{m} \lambda_i g_i^\circ(\bar{x}, h) \geq 0 \quad \forall h \in \mathbb{R}^n. \tag{4}$$

Consider the set $I = \{i \in \{1, \dots, m\} : \lambda_i > 0\}$. This set is of course non-empty since $\lambda_0 = 0$. Since Slater constraint qualification holds there exists $\hat{x}$ such that $g_i(\hat{x}) < 0$ for all $i = 1, \dots m$. Now since each g_i is continuous there exists $\delta > 0$ such that for all $x \in B(\hat{x}, \delta)$, $g_i(x) < 0$ for all $i = 1, \dots, m$. Now setting $h = x - \bar{x}$ in (4) where $x \in B(\hat{x}, \delta)$ we conclude that

$$\sum_{i \in I} \lambda_i g_i^\circ(\bar{x}, x - \bar{x}) \geq 0.$$

Since $\lambda_i > 0$ when $i \in I$ we have $g_i(\bar{x}) = 0$ when $i \in I$ and since K is convex we conclude using Proposition 2.2 that for all $i \in I$, $g_i^\circ(\bar{x}, x - \bar{x}) = 0$ for all $x \in B(\hat{x}, \delta)$. This shows that $0 \in \partial^\circ g_i(\bar{x})$ for all $i \in I$ and hence this contradicts Assumption (A). This shows that $\lambda_0 > 0$ and without loss of generality we can take $\lambda_0 = 1$ and thus establishing the necessary part.
For sufficiency of the above conditions we proceed as follows. On the contrary assume that $\bar{x}$ is not the global minimum and hence there exists $z \in K$ such that $f(\bar{x}) > f(z)$. Now using the convexity of f we have the following,

$$0 > f(z) - f(\bar{x}) \geq f'(\bar{x}, z - \bar{x}).$$

This shows using the optimality conditions

$$0 > -\sum_{i \in I} \lambda_i g_i^\circ(\bar{x}, z - \bar{x}) \geq 0,$$

where we arrive at the last inequality using the Proposition 2.2 and the fact that $\lambda_i > 0$ for all $i \in I$. Hence we arrive at a contradiction. This proves that $\bar{x}$ is the global minimizer. □

References

[1] D. P. Bertsekas, *Convex Analysis and Optimization*, Athena Scientific, 2003.

[2] J. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization*, Springer, 2000.

[3] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley Interscience, 1983.

[4] J. B. Lasserre, On representations of the feasible set in convex optimization, *Optimization Letters*, Vol 4, 2010, pp 1-5.

[5] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.