On optimality conditions in robust optimization problems with locally Lipschitz constraints[1]

Zhe Hong[2] and Do Sang Kim[3]

[2]Department of Applied Mathematics
Pukyong National University, Busan 48513, Korea
E-mail: hongchul1990@gmail.com
[3]Department of Applied Mathematics
Pukyong National University, Busan 48513, Korea
E-mail: dskim@pknu.ac.kr

Abstract. In this paper, we study a convex optimization problem which minimizes a convex function over a convex feasible set defined by finitely many locally Lipschitz constraints (not necessarily convex or differentiable) in the face of data uncertainty. Under a non-degeneracy condition and the Slater constraint qualification, we present Karush–Kuhn–Tucker optimality conditions for the robust convex optimization problem. Moreover, we apply the obtained results to study the KKT optimality conditions for a quasi ϵ-solution to the robust convex optimization problem.

1 Introduction

A standard constrained convex programming problem is minimizing a convex function over a convex feasible set C, which is usually given by convex inequality constraints, that is,

$$\min_{x\in\mathbb{R}^n} \ f(x) \ \text{ subject to } \ x \in C, \tag{P}$$

where $f:\mathbb{R}^n \to \mathbb{R}$ and $C := \{x \in \mathbb{R}^n : g_i(x) \leq 0, i = 1,\dots,m\}$, here $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1,\dots,m$ are convex functions.

However, the convex feasible set C of problem (P) may not be described by convex inequality constraints. In 2010, Lasserre [11] studied a convex optimization problem, whose objective function is differentiable and convex, and constraint functions are differentiable but not necessarily convex (surely the feasible set shall be convex), and obtained the Karush–Kuhn–Tucker (KKT) optimality conditions (both necessary and sufficient) with the help of the Slater constraint qualification and an additional non-degeneracy condition. In 2013, Dutta and Lalitha [10] extended the study to a nonsmooth scenario involving the locally Lipschitz functions, say concretely, they considered a convex optimization problem, whose objective function is convex (not necessarily differentiable)

[1]This paper is based on the published one "Approximate optimality conditions for robust convex optimization without convexity of constraints. Linear and Nonlinear Analysis **5** (2019), no.1, 173–182" written by Z. Hong, L.G. Jiao and D.S. Kim.

and the constraint functions are locally Lipschitz (not necessarily convex or differentiable). They showed that if the Slater constraint qualification and a simple non-degeneracy condition were satisfied then the KKT type optimality condition was both necessary and sufficient.

In this paper, we mainly apply some results of Sisarat *et al.* [14] to study the KKT optimality conditions for a quasi ϵ-solution to the robust convex optimization problem.

Consider the following convex optimization problem:

$$\min\ f(x)\ \text{ s.t. }\ g_i(x) \leq 0,\ i = 1, \ldots, m, \tag{CP}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function and $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, are locally Lipschitz functions such that the set $S_i := \{x \in R^n : g_i(x) \leq 0\}$ is convex, and then the feasible set $S = \cap_{i=1}^m S_i$ is also convex.

The convex optimization problem (CP) in the face of data uncertainty in the constraints can be written by the following problem:

$$\min\ f(x)\ \text{ s.t. }\ g_i(x, v_i) \leq 0, i = 1, \ldots, m, \tag{UCP}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, $g_i : \mathbb{R}^n \times \mathbb{R}^q \to \mathbb{R}$, $g_i(\cdot, v_i)$ is a locally Lipschitz function and $g_i(x, \cdot)$ is a concave function, and $v_i \in \mathbb{R}^q$ is an uncertain parameter which belongs to the compact convex set $\mathcal{V}_i \subset \mathbb{R}^q$, $i = 1, \ldots, m$.

In this work, we treat the robust approach for (UCP), which is the worst case approach for (UCP); see, for example, [1–4, 13]. Now, we associate with (UCP) its robust counterpart:

$$\min\ f(x)\ \text{ s.t. }\ g_i(x, v_i) \leq 0,\ \forall v_i \in \mathcal{V}_i,\ i = 1, \ldots, m. \tag{RCP}$$

Denote by $F := \{x \in \mathbb{R}^n : g_i(x, v_i) \leq 0,\ \forall v_i \in \mathcal{V}_i,\ i = 1, \ldots, m\}$ as the feasible set of (RCP), and assume here the feasible set F is convex. Set $F = \bigcap_{i=1}^m \bigcap_{v_i \in \mathcal{V}_i} F_i(v_i)$, where $F_i(v_i) := \{x \in \mathbb{R}^n : g_i(x, v_i) \leq 0\}$, $v_i \in \mathcal{V}_i$, $i = 1, \ldots, m$.

Let $x \in F$, $I := \{1, \ldots, m\}$ and define functions $\psi_i \colon \mathbb{R}^n \to \mathbb{R}$ by $\psi_i(x) := \max\{g_i(x, v_i) : v_i \in \mathcal{V}_i\}$, $i \in I$. Let $I(x) := \{i \in I : \psi_i(x) = 0\}$. We put for each $i \in I(x)$,

$$\mathcal{V}_i(x) := \{v_i \in \mathcal{V}_i : g_i(x, v_i) = \psi_i(x)\}.$$

Definition 1.1 *Consider the problem (RCP). We say that the non-degeneracy condition holds at $x \in F$ if for all $i \in I(x)$ and all $v_i \in \mathcal{V}_i(x)$*

$$0 \notin \partial^\circ g_i(x, v_i).$$

The feasible set F is said to satisfy the non-degeneracy condition if it holds for every $x \in F$.

Remark 1.1 *This condition was introduced firstly by Lasserre [11] in the case that g_i is differentiable. Motivated by this idea, Dutta and Lalitha [10] extended the non-degeneracy condition to the nonsmooth case.*

2 Preliminaries

In this section, we recall some notations and give preliminary results for next sections. Throughout this paper, $\mathbb{R}^n$ denotes the n-dimensional Euclidean space with the inner product $\langle \cdot, \cdot \rangle$ and the associated Euclidean norm $\|\cdot\|$. We say that a set Γ in $\mathbb{R}^n$ is *convex* whenever $\mu a_1 + (1-\mu)a_2 \in \Gamma$ for all $\mu \in [0,1]$, $a_1, a_2 \in \Gamma$. We denote the domain of f by $\operatorname{dom} f$, that is, $\operatorname{dom} f := \{x \in \mathbb{R}^n : f(x) < +\infty\}$. f is said to be *convex* if for all $\lambda \in [0,1]$,

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

for all $x, y \in \mathbb{R}^n$. The function f is said to be *concave* whenever $-f$ is convex. The (convex) subdifferential of f at $x \in \mathbb{R}^n$ is defined by

$$\partial f(x) = \begin{cases} \{x^* \in \mathbb{R}^n \mid \langle x^*, y - x \rangle \leq f(y) - f(x),\ \forall y \in \mathbb{R}^n\}, & \text{if } x \in \operatorname{dom} f, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Let $g : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function, that is, for each $x \in \mathbb{R}^n$, there exist an open neighborhood U and a constant $L > 0$ such that for all y and z in U,

$$|g(y) - g(z)| \leq L\|y - z\|.$$

Definition 2.1 *For each $d \in \mathbb{R}^n$, the* Clarke directional derivative *of g at $x \in \mathbb{R}^n$ in the direction d, denoted by $g^\circ(x; d)$, is given by*

$$g^\circ(x; d) \;=\; \limsup_{h \to 0,\ t \to 0+} \frac{g(x + h + td) - g(x + h)}{t}.$$

We also denote the usual one-sided directional derivative *of g at x by $g'(x; d)$. Thus*

$$g'(x; d) = \lim_{t \to 0+} \frac{g(x + td) - g(x)}{t},$$

whenever this limit exists.

Definition 2.2 *The* Clarke subdifferential *of g at x, denoted by $\partial^\circ g(x)$, is the (nonempty) set of all ξ in $\mathbb{R}^n$ satisfying the following condition:*

$$g^\circ(x; d) \geq \langle \xi, d \rangle, \quad \textit{for all} \;\; d \in \mathbb{R}^n.$$

We summarize some fundamental results in the calculus of the Clarke subdifferential (for more details, see [5–8, 12]):

- $\partial^\circ g(x)$ is a nonempty, convex, compact subset of $\mathbb{R}^n$;
- The function $d \mapsto g^\circ(x; d)$ is convex;
- For every d in $\mathbb{R}^n$, one has

$$g^\circ(x; d) = \max\{\langle \xi, d \rangle : \xi \in \partial^\circ g(x)\}. \tag{2.1}$$

Let $\mathcal{V} \subset \mathbb{R}^q$ be a compact set and let $g\colon \mathbb{R}^n \times \mathcal{V} \to \mathbb{R}$ be a given function. Here after all, we assume that the following assumptions hold:

- (A1) $g(x, v)$ is upper semicontinuous in (x, v).
- (A2) g is locally Lipschitz in x, uniformly for v in $\mathcal{V}$, that is, for each $x \in \mathbb{R}^n$, there exist an open neighborhood U of x and a constant $L > 0$ such that for all y and z in U, and $v \in \mathcal{V}$,

$$|g(y, v) - g(z, v)| \leq L\|y - z\|.$$

- (A3) $g_x^\circ(x, v; \cdot) = g_x'(x, v; \cdot)$, the derivatives being with respect to x.

We define a function $\psi\colon \mathbb{R}^n \to \mathbb{R}$ by

$$\psi(x) := \max\{g(x, v) : v \in \mathcal{V}\},$$

and observe that our assumptions (A1)-(A2) imply that ψ is defined and finite (with the maximum defining ψ attained) on $\mathbb{R}^n$. Let

$$\mathcal{V}(x) := \{v \in \mathcal{V} : g(x, v) = \psi(x)\},$$

then for each $x \in \mathbb{R}^n$, $\mathcal{V}(x)$ is a nonempty closed set.

The following lemma, which is a nonsmooth version of Danskin's theorem [9] for max-functions, makes connection between the functions $\psi'(x; d)$ and $g_x^\circ(x, v; d)$.

Lemma 2.1 *Under the assumptions* (A1)–(A3), *the usual one-sided directional derivative* $\psi'(x; d)$ *exists, and satisfies*

$$\begin{aligned}\psi'(x; d) = \psi^\circ(x; d) &= \max\{g_x^\circ(x, v; d) : v \in \mathcal{V}(x)\} \\ &= \max\{\langle \xi, d\rangle : \xi \in \partial_x^\circ g(x, v),\ v \in \mathcal{V}(x)\}.\end{aligned}$$

The following result will be useful in the sequel.

Lemma 2.2 [13] *In addition to the basic assumptions* (A1)–(A3), *suppose that* $\mathcal{V}$ *is convex, and that* $g(x, \cdot)$ *is concave on* $\mathcal{V}$, *for each* $x \in U$. *Then the following statements hold*:

(i) *The set* $\mathcal{V}(x)$ *is convex and compact.*

(ii) *The set*

$$\partial_x^\circ g(x, \mathcal{V}(x)) := \{\xi : \exists v \in \mathcal{V}(x) \ \ \textit{such that} \ \ \xi \in \partial_x^\circ g(x, v)\}$$

is convex and compact.

(iii) $\partial^\circ \psi(x) = \{\xi : \exists v \in \mathcal{V}(x) \ \ \textit{such that} \ \ \xi \in \partial_x^\circ g(x, v)\}$.

It is worth noting that the concavity of $g_i(x, \cdot)$ plays an important role, since our main results (Theorem 3.1, 3.2 and 3.3) shall be obtained with the aid of the above Lemma 2.2.

3 Optimality Conditions

First, the Slater constraint qualification along with the non-degeneracy condition gives the following equivalent characterization of the convex set F under the robust counterpart scenario.

Theorem 3.1 *Let F be given in the problem* (RCP). *Assume that each g_i satisfies the assumptions* (A1)–(A3). *Moreover, assume that the non-degeneracy condition holds at $x \in F$, and the Slater constraint qualification also holds, that is, there exists $x_0 \in \mathbb{R}^n$ such that $g_i(x_0, v_i) < 0$, for all $v_i \in \mathcal{V}_i$, $i = 1, \ldots, m$. Then F is convex if and only if for all $i \in I(x)$, there exists $\bar{v}_i \in \mathcal{V}_i(x)$ such that*

$$g_{i_x}^{\circ}\big(x, \bar{v}_i; y - x\big) \leq 0, \text{ for all } x, y \in F.$$

The following result is a robust KKT optimality theorem for (RCP), which is a robust version of [10, Theorem 2.4].

Theorem 3.2 *Let us consider the problem* (RCP). *Assume that each g_i satisfies the assumptions* (A1)–(A3). *Moreover assume that the non-degeneracy condition holds at $\bar{x} \in F$, and the Slater constraint qualification also holds. Then $\bar{x} \in F$ is an optimal solution of f over F if and only if there exist $\bar{\lambda}_i \geq 0$ and $\bar{v}_i \in \mathcal{V}_i(\bar{x})$, $i = 1, \ldots, m$, such that*

$$0 \in \partial f(\bar{x}) + \sum_{i=1}^{m} \bar{\lambda}_i \partial_x^{\circ} g_i(\bar{x}, \bar{v}_i),$$
$$0 = \bar{\lambda}_i g_i(\bar{x}, \bar{v}_i), \ i = 1, \ldots, m.$$

The following example examines the validness of our main results whenever non-degeneracy condition is satisfied.

Example 3.1 *Consider the following convex optimization problem with data uncertainty:*

$$\begin{array}{lll} (\mathrm{RCP})^2 & \min & -x \\ & s.t. & x \in F^2 := \{x \in \mathbb{R} : \max\{vx^3, vx\} - 2 \leq 0, \ \forall v \in \mathcal{V}\}, \end{array}$$

where $\mathcal{V} := [1, 2]$. Let $f(x) = -x$ and $g(x, v) = \max\{vx^3, vx\} - 2$. Then we can easily see that $F^2 = (-\infty, 1]$ is the robust feasible set of $(\mathrm{RCP})^2$ and $\bar{x} = 1$ is an optimal solution of $(\mathrm{RCP})^2$. Clearly, g satisfies the assumptions (A1)–(A3), *and the Slater condition holds for $(\mathrm{RCP})^2$. Moreover, $\mathcal{V}(\bar{x}) = \{2\}$, and so for $\bar{v} := 2 \in \mathcal{V}(\bar{x})$, $0 \notin \partial^{\circ} g(\bar{x}, \bar{v}) = [2, 6]$, i.e., the non-degeneracy condition holds. Let $0 \leq \bar{\lambda} \leq \frac{1}{2}$. Then we have*

$$0 \in \partial f(\bar{x}) + \bar{\lambda} \partial_x^{\circ} g(\bar{x}, \bar{v}) = \{-1\} + \bar{\lambda}[2, 6],$$
$$0 = \bar{\lambda} g(\bar{x}, \bar{v}).$$

So, Theorem 3.2 holds.

Definition 3.1 *Given* $\epsilon \geq 0$*, a point* $\bar{x} \in F$ *is said to be a quasi* ϵ*-solution of problem* (RCP)*, if*

$$f(\bar{x}) \leq f(x) + \sqrt{\epsilon}\|x - \bar{x}\|, \quad \forall x \in F.$$

By employing Theorem 3.2, we give the following robust KKT optimality theorem for a quasi ϵ-solution in (RCP).

Theorem 3.3 *Let us consider the problem* (RCP). *Assume that each* g_i *satisfies the assumptions* (A1)–(A3). *Moreover assume that the non-degeneracy condition holds at* $\bar{x} \in F$*, and the Slater constraint qualification also holds. Then* $\bar{x} \in F$ *is a quasi* ϵ*-solution of* (RCP) *if and only if there exist* $\bar{\lambda}_i \geq 0$ *and* $\bar{v}_i \in \mathcal{V}_i(\bar{x})$*,* $i = 1, \ldots, m$*, such that*

$$0 \in \partial f(\bar{x}) + \sum_{i=1}^{m} \bar{\lambda}_i \partial_x^{\circ} g_i(\bar{x}, \bar{v}_i) + \sqrt{\epsilon}\mathbb{B},$$
$$0 = \bar{\lambda}_i g_i(\bar{x}, \bar{v}_i),\ i = 1, \ldots, m,$$

where $\mathbb{B}$ *stands for the unit ball.*

References

[1] A. Beck and A. Ben-Tal, Duality in robust optimization: primal worst equals dual best. *Oper. Res. Lett.* **37**(2009) , 1–6.

[2] A. Ben-Tal, D. den Hertog and J. Vial, Deriving robust counterparts of nonlinear uncertain inequalities. *Math. Program.* **149**(2015), 265–299.

[3] A. Ben-Tal and A. Nemirovski, Robust convex optimization. *Math. Oper. Res.* **23**(1998), 769–805.

[4] A. Ben-Tal and A. Nemirovski, Selected topics in robust convex optimization. *Math. Program.* **112**(2008), 125–158.

[5] F.H. Clarke, Generelized gradients and applications. *Trans. Amer. Math. Soc.* **205**(1975), 247–262.

[6] F.H. Clarke, A new approach to Lagrange multipliers. *Math. Oper. Res.* **1**(1976), 165–174.

[7] F.H. Clarke, Generalized gradients of Lipschitz functions. *Adv. in Math.* **40**(1981), 52–67.

[8] F.H. Clarke, *Optimization and Nonsmooth Analysis.* Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1990)

[9] J.M. Danskin, *The Theory of Max-Min and its Application to Weapons Allocation Problems.* Springer-Verlag, New York (1967)

[10] J. Dutta and C.S. Lalitha, Optimality conditions in convex optimization revisited. *Optim. Lett.* **7**(2013), 221–229.

[11] J.B. Lasserre, On representations of the feasible set in convex optimization. *Optim. Lett.* **4**(2010), 1–5.

[12] G. Lebourg, Generic differentiability of Lipschitzian functions. *Trans. Amer. Math. Soc.* **256**(1979), 125–144.

[13] G.M. Lee and P.T. Son, On nonsmooth optimality theorems for robust optimization problems. *Bull. Korean Math. Soc.* **51** (2014), 287–301.

[14] N. Sisarat, R. Wangkeeree and G.M. Lee, Some characterizations of robust solution sets for uncertain convex optimization problems with locally Lipschitz inequality constraints. *J. Ind. Manag. Optim.* **16**(2020), 469–493.