Optimality Conditions for Nonsmooth Convex Optimization

Sangkyun Lee

Oct 22, 2014

Let us consider a convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}}$ is the extended real field, $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$, which is proper ($f$ never admits $-\infty$ and is not identically equal to $+\infty$). Convexity of extended-real valued functions are defined in the same way as the real-valued counterparts, but with additional operations defined on $\overline{\mathbb{R}}$ (e.g. $\infty + \alpha = \infty$ for $\alpha > 0$).

This article discusses the optimality conditions of minimizing extended-real, nonsmooth convex functions. For the purpose we need two ingredients: (i) the continuity of convex functions and (ii) the existence of directional derivatives of convex functions.

1. Terminology

A proper function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *convex* if

$$f(\alpha x + (1-\alpha)x') \le \alpha f(x) + (1-\alpha) f(x'), \ \ \forall x, x' \in \mathbb{R}^n, \forall \alpha \in (0,1)$$

where the inequality is considered in $\overline{\mathbb{R}}$ (defining additional necessary operations such as $\alpha\infty = \infty$, $\infty + \infty = \infty$, $\infty \le \infty$, etc.)

The *(effective) domain* of a function f is

$$\operatorname{dom} f := \{x \in \mathbb{R}^n : f(x) < +\infty\},$$

which is nonempty when f is proper.

The *relative interior* of a convex set $C \subset \mathbb{R}^n$ is the interior of C for the topology relative to the affine hull of C, i.e.

$$\operatorname{ri} C = \{x \in \operatorname{aff} C : \exists \delta > 0 \text{ s.t. } (\operatorname{aff} C) \cap B(x, \delta) \subset C\}.$$

Here $B(x, \delta)$ is a closed ball with radius δ centered at x. To see the reason why we need this notion, imagine a "flat" filled circle in $\mathbb{R}^3$ as for our convex set C. The interior of C is empty, since C does not contain any 3-D ball centered at a point from C. The affine hull $\operatorname{aff} C$ is the affine plane containing C, relative to which we can consider another notion of interior using 2-D balls within the plane. When $C \ne \emptyset$, then $\operatorname{ri} C \ne \emptyset$ ([HUL96], Theorem 2.1.3).

The *epigraph* of f is the set

$$\operatorname{epi} f := \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : f(x) \le r\},$$

which is nonempty if f is proper. f is convex if and only if $\operatorname{epi} f$ is a convex set in $\mathbb{R}^n \times \mathbb{R}$.

2. Continuity of Convex Functions

Lemma 1 ([HUL96], Lemma IV.3.1.1) *For a proper convex function f, suppose that there exist $x_0, \delta, m,$ and M such that*

$$m \le f(x) \le M, \ \ \forall x \in B(x_0, 2\delta).$$

Then f is Lipschitz continuous on $B(x_0, \delta)$, that is,

$$|f(x') - f(x)| \le \frac{M-m}{\delta}\|x' - x\| \;\; \forall x, x' \in B(x_0, \delta).$$

Proof For two different points x, x' in $B(x_0, \delta)$, take

$$x'' = x' + \delta\frac{x' - x}{\|x' - x\|} \in B(x_0, 2\delta).$$

This implies that x' lies on the line segment $[x, x'']$,

$$x' = \frac{\|x' - x\|}{\delta + \|x' - x\|}x'' + \frac{\delta}{\delta + \|x' - x\|}x.$$

Using the convexity of f and the given conditions, it leads to

$$f(x') - f(x) \le \frac{\|x' - x\|}{\delta + \|x' - x\|}[f(x'') - f(x)] \le \frac{\|x' - x\|}{\delta}(M - m).$$

A similar inequality can be obtained by exchanging x and x', to show the claim. ■

Theorem 2 ([HUL96], Theorem IV.3.1.2, Modified) *For a proper convex f, there exists a compact convex subset S of* $\mathrm{ri}\,\mathrm{dom}\,f$ *such that with $L = L(S) \ge 0$,*

$$|f(x') - f(x)| \le L\|x - x\| \;\; \forall x, x' \in S.$$

Proof First, note that $\mathrm{ri}\,\mathrm{dom}\,f \neq \emptyset$. Since the theorem statement ignores points outside of $\mathrm{aff}\,\mathrm{dom}\,f$, we can discuss in $\mathbb{R}^d$ instead of in $\mathbb{R}^n$ where d is the dimension of $\mathrm{dom}\,f$[1]. Then we may assume $\mathrm{ri}\,\mathrm{dom}\,f = \mathrm{int}\,\mathrm{dom}\,f$ to simply the proof.

Under the assumption, $\mathrm{dom}\,f$ ($\mathrm{ri}\,\mathrm{dom}\,f$ without the assumption, for example) contains $d+1$ affinely independent points $v_0, v_1, \ldots, v_d$. Consider a simplex defined as a convex hull of these points, $\Delta := \mathrm{conv}\,\{v_0, v_1, \ldots, v_d\} \subset \mathrm{dom}\,f$, having x_0 in its interior. Then we can find $\delta > 0$ such that $B(x_0, 2\delta) \subset \Delta$. Then any $x \in B(x_0, 2\delta)$ can be written $x = \sum_{i=0}^d \alpha_i v_i$ for some $\alpha_i \ge 0$, $\sum_{i=0}^d \alpha_i = 1$, so that the convexity of f gives

$$f(x) \le \sum_{i=0}^d \alpha_i f(v_i) \le M := \max\{f(v_0), \ldots, f(v_d)\}.$$

On the other hand, from the convexity of f tells that for all $y \in B(x_0, 2\delta)$,

$$f(x) \ge f(y) + g(y)^T(x - y) \ge m := \min_{z \in B(x_0, 2\delta)}\{f(z) + g(z)^T(z - x)\}.$$

Note that the minimum exists since $B(x_0, 2\delta)$ is a compact set. The claim follows from Lemma 1. ■

This result implies that a proper convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is continuous (even Lipschitz continuous) at x "well inside" of $\mathrm{ri}\,\mathrm{dom}\,f$. However, if x approaches the relative boundary $\mathrm{rbd}\,\mathrm{dom}\,f := \mathrm{cl}\,\mathrm{dom}\,f \setminus \mathrm{ri}\,\mathrm{dom}\,f$, continuity may break down[2].

1. The dimension of a set C is defined as the dimension of a subspace parallel to $\mathrm{aff}\,C$.
2. Relative closure is just closure, since affine hull is always closed. If f is finite everywhere, the continuity at boundaries of the domain of f is easier to check using definitions, however it becomes nontrivial in $\overline{\mathbb{R}}$.

3. Directional Derivatives and Subdifferential

3.1 Subdifferential is Nonempty

We define a vector g as a *subgradient* for a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ at $x \in \mathrm{ri}\ \mathrm{dom} f$ if

$$f(y) \geq f(x) + g^T(y - x), \quad \forall y \in \mathbb{R}^n.$$

The above is called the *subgradient inequality.* The set of all subgradients defines the *subdifferential*, that is,

$$\partial f(x) := \{g \in \mathbb{R}^n : f(y) \geq f(x) + g^T(y - x) \ \ \forall y \in \mathbb{R}^n\}.$$

Theorem 3 ([HUL96], Proposition IV.1.2.1) *For a proper convex function f, $\partial f(x)$ at any $x \in \mathrm{ri}\ \mathrm{dom} f$ is nonempty.*

This theorem is intuitively understandable, but its proof requires the existence of separating hyperplanes for convex sets involving several theorems. We skip the proof here, but interested readers are advised to check the reference above.

3.2 Existence of Directional Derivatives

For any function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, the *one-sided directional derivative* of f at x (where $f(x)$ is finite) with respect to a direction $d \in \mathbb{R}^n$ is defined as the limit

$$f'(x; d) := \lim_{\epsilon \downarrow 0} \frac{f(x + \epsilon d) - f(x)}{\epsilon}.$$

if the limit exists (we allow $+\infty$ or $-\infty$ as limits). Note that

$$-f'(x; -d) = \lim_{\epsilon \uparrow 0} \frac{f(x + \epsilon d) - f(x)}{\epsilon}$$

Theorem 4 ([Roc97] Theorem 23.1) *Let f be convex and be finite near x. For each direction d, the difference quotient in the definition of $f'(x; d)$ is nondecreasing function of $\epsilon > 0$, so that $f'(x; d)$ exists and*

$$f'(x; d) = \inf_{\epsilon > 0} \frac{f(x + \epsilon d) - f(x)}{\epsilon}.$$

Moreover, $f'(x; d)$ is a positively homogeneous convex function of d, with $f'(x; 0) = 0$ and $-f'(x; -d) \leq f'(x; d)$ for all d.

Proof We can express the difference quotient as $\epsilon^{-1}h(\epsilon d)$ for $\epsilon > 0$, where $h(d) = f(x + d) - f(x)$. The convex set $\mathrm{epi}\, h$ is obtained by translating $\mathrm{epi}\, f$ so that $(x, f(x))$ is moved to $(0, 0)$. On the other hand, $\epsilon^{-1}h(\epsilon d) = (h\epsilon^{-1})(d)$, where $h\epsilon^{-1}$ is defined as the convex function whose epigraph is $\epsilon^{-1}\mathrm{epi}\, h$. Since $\mathrm{epi}\, h$ contains the origin, $\epsilon^{-1}\mathrm{epi}\, h$ increases as ϵ^{-1} increases. That is, for each d, the difference quotient $(h\epsilon^{-1})(d)$ decreases as ϵ decreases. Furthermore, $(h\epsilon^{-1})(d)$ is bounded below since f is convex,

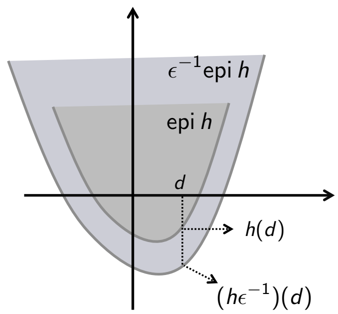
$$(h\epsilon^{-1})(d) \geq g^T d$$

for some subgradient $g \in \partial f(x)$. Then by the monotone convergence theorem, the limit $f'(x; d)$ exists and

$$\inf_{\epsilon > 0} (h\epsilon^{-1})(d) = f'(x; d).$$

Furthermore, this indicates that $f'(x; d)$ is a finite-valued convex function in d. It is also positively homogeneous in d since for any $\alpha > 0$,

$$f'(x; \alpha d) = \inf_{\epsilon \downarrow 0} \frac{f(x + \epsilon(\alpha d)) - f(x)}{\epsilon} = \alpha \inf_{\epsilon\alpha \downarrow 0} \frac{f(x + (\epsilon\alpha)d) - f(x)}{\epsilon\alpha} = \alpha f'(x; d).$$



$f'(x;0) = 0$ is obvious from definition. From convexity,

$$\frac{1}{2}f'(x;d) + \frac{1}{2}f'(x;-d) \geq f'\left(x; \frac{d}{2} - \frac{d}{2}\right) = 0$$

which gives the final claim. ∎

3.3 An Alternative Definition of Subdifferential

The previous theorem allows us to characterize a subdifferential using directional derivatives.

Theorem 5 *The subdifferential $\partial f(x)$ for $x \in \operatorname{ri}\operatorname{dom} f$ has an equivalent definition,*

$$\partial f(x) := \{g \in \mathbb{R}^n : g^T d \leq f'(x;d) \ \ \forall d \in \mathbb{R}^n\}.$$

Proof For any direction d and $\epsilon > 0$ such that $y = x + \epsilon d \in \mathbb{R}^n$, the subgradient inequality implies that

$$g^T d \leq \frac{f(x+\epsilon d) - f(x)}{\epsilon}.$$

That is,

$$g^T d \leq \inf_{\epsilon > 0} \frac{f(x+\epsilon d) - f(x)}{\epsilon} = f'(x;d).$$

The converse is trivial ($d = y - x$ and $\epsilon = 1$). ∎

This reveals that the subdifferential $\partial f(x)$ is a compact convex set.

4. Optimality Conditions

4.1 Unconstrained Case

Let us consider the optimization problem

$$\min_{x \in \mathbb{R}^n} \quad f(x) \tag{1}$$

where $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper and convex.

Theorem 6 *The following statements are equivalent:*

(i) *x^* is a solution of* (1).

(ii) $0 \in \partial f(x^*)$.

(iii) $f'(x; d) \geq 0 \;\; \forall d \in \mathbb{R}^n$.

Proof [(ii) $\Rightarrow$ (i)] Suppose that $g = 0 \in \partial f(x^*)$. Then the subgradient inequality

$$f(x) \geq f(x^*) + g^T(x - x^*) = f(x^*), \;\; \forall x \in \operatorname{dom}(f),$$

implies that x^* is a solution.

[(i) $\Rightarrow$ (iii)] Suppose that x^* is a solution, i.e., $f(x^*) \leq f(x)$ for all $x \in \operatorname{dom}(f)$. For any direction $d \in \mathbb{R}^n$ such that $x^* + \epsilon d \in \operatorname{dom} f$, every subgradient $g \in \partial f(x^*)$ satisfies that

$$g^T d \leq \inf_{\epsilon > 0} \frac{f(x^* + \epsilon d) - f(x^*)}{\epsilon} = f'(x^*; d)$$

where $f'(x^*; d) \geq 0$ since x^* is a minimizer.

[(iii) $\Rightarrow$ (ii)] The inequality $h(g) = g^T d \leq f'(x^*; d)$ defines a halfspace for each d, which contains the origin since $f'(x^*; d) \geq 0$. The subdifferential

$$\partial f(x^*) := \{g : g^T d \leq f'(x^*; d) \;\; \forall d \in \mathbb{R}^n\}$$

is the intersection of all such halfspaces, and therefore should contain the origin. ■

The above is my own proof.

4.2 Constrained Case

Let us first define geometric objects useful for describing optimality conditions in this case.

Definition 7 (Tangent Cone) *Let $X \subset \mathbb{R}^n$ be a nonempty set. A direction $d \in \mathbb{R}^n$ is a* tangent direction *to X at $x \in X$ if there exists a sequence $\{x_k\} \in S$ and a scalar sequence $\{t_k\}$ such that*

$$x_k \to x, \;\; t_k \to 0, \;\; \frac{x_k - x}{t_k} \to d \;\; \text{as} \;\; k \to \infty.$$

The set of all such directions is called the tangent cone *to X at x, denoted by $T_X(x)$.*

- $T_X(x)$ is indeed a cone.
- $T_X(x)$ always contains the origin.
- $T_X(x)$ is always closed.
- If X is a convex set, $T_X(x)$ is also convex.

Definition 8 (Normal Cone) *Let $X \subset \mathbb{R}^n$ be a nonempty set. The direction $s \in \mathbb{R}^n$ is called* normal *to X at $x \in X$ if*

$$\langle s, x' - x \rangle \leq 0 \;\; \forall x' \in X.$$

The set of all such directions is called the normal cone *to X at x, denoted by $N_X(x)$.*

Definition 9 (Polar Cone) *For a cone $K \subset \mathbb{R}^n$, the* polar cone *of K is defined by*

$$K^\circ = \{s \in \mathbb{R}^n : \langle s, d \rangle \leq 0, \; \forall d \in K\}.$$

It is easy to check the following:

- $N_X(x) = (T_X(x))^\circ$.
- $T_X(x) = (N_X(x))^\circ$ if X is convex[3].

Definition 10 (Indicator Function) *$I_S(x) : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is an indicator function for a nonempty set $S \subset \mathbb{R}^n$ such that*

$$I_S(x) := \begin{cases} 0 & \text{if } x \in S \\ +\infty & o.w. \end{cases}$$

Note that indicator functions are convex.

Definition 11 (Support Function) *Let S be a nonempty set in $\mathbb{R}^n$. The function $\sigma_S : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by*

$$\sigma_S(x) = \sup\{\langle s, x\rangle : s \in S\}$$

is called the support function *of S.*

Note that from the continuity and linearity of the function $\langle s, \cdot\rangle$ maximized over S, $\sigma_S = \sigma_{\operatorname{cl} S} = \sigma_{\operatorname{conv} S}$.

Proposition 12 ([HUL96], Example V.2.3.1) *For a closed convex cone $K \subset \mathbb{R}^n$, the support function $\sigma_K(d)$ becomes*

$$\sigma_K(d) = \begin{cases} 0 & \text{if } \langle s, d\rangle \le 0 \text{ for all } s \in K \\ +\infty & o.w. \end{cases}$$

That is, $\sigma_K(d)$ is the indicator function of the polar cone K°. Since $K^{\circ\circ} = K$, the support function of K° is the indicator of K.

Theorem 13 ([HUL96], Theorem V.3.3.3 (i)) *Let σ_1 and σ_2 be the support functions of the nonempty closed convex sets S_1 and S_2. For positive scalars t_1 and t_2, $t_1\sigma_1 + t_2\sigma_2$ is the support function of $\operatorname{cl}(t_1S_1 + t_2S_2)$.*

Proof Let $S = \operatorname{cl}(t_1S_1 + t_2S_2)$ which is a closed convex set. By definition, its support function is

$$\sigma_S(d) = \sup\{\langle t_1s_1 + t_2s_2, d\rangle : s_1 \in S_1, s_2 \in S_2\}.$$

In the above expression, s_1 and s_2 run independently in their index sets S_1 and S_2. Therefore, we have

$$\sigma_S(d) = t_1 \sup_{s \in S_1} \langle s, d\rangle + t_2 \sup_{s \in S_2} \langle s, d\rangle.$$

■

The following separation theorem is also useful.

Theorem 14 ([HUL96], Theorem III.4.1.1) *Let $C \subset \mathbb{R}^n$ be nonempty closed convex, and let $x \notin C$. Then there exists $s \in \mathbb{R}^n$, $s \neq 0$ such that*

$$\langle s, x\rangle > \sup_{y \in C} \langle s, y\rangle.$$

3. For a convex cone C, $(C^\circ)^\circ = \operatorname{cl} C$ (the closure of C), and thus if C is convex and closed then $(C^\circ)^\circ = C$ (polar cone theorem).

Proof Let $P_C(x)$ be the Euclidean projection of x onto C, that is, $P_C(x) = \arg\min_{y \in C} \frac{1}{2}\|y - x\|_2^2$. From the optimality condition of smooth convex minimization (see e.g. lecture notes of Numerical Optimization), $P_C(x)$ is the projection if and only if

$$\langle x - P_C(x), y - P_C(x) \rangle \leq 0 \ \ \forall y \in C.$$

Let $s := x - P_C(x) \neq 0$. The above inequality gives that

$$0 \geq \langle s, y - x + s \rangle = \langle s, y \rangle - \langle s, x \rangle + \|s\|^2.$$

That is,

$$\langle s, x \rangle - \|s\|^2 \geq \langle s, y \rangle \ \ \forall y \in C.$$

■

Now we consider the optimization over a nonempty convex set $X \subset \mathbb{R}^n$,

$$\min_{x \in X} \ f(x) \tag{2}$$

where $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex on X.

Theorem 15 ([HUL96], Theorem VII.1.1.1) *The following statements are equivalent:*

(i) x^* *is a solution of* (2).

(ii) $f'(x^*, x - x^*) \geq 0$ *for all* $x \in X$.

(iii) $f'(x^*, d) \geq 0$ *for all* $d \in T_X(x^*)$.

(iv) $0 \in \partial f(x^*) + N_X(x^*)$.

Proof [(i) ⇒ (ii) ⇒ (iii)] Choose a $x \in X$. By convexity, $x^* + t(x - x^*) \in X$ for all $t \in [0, 1]$. If (i) is true, then (ii) follows since

$$\frac{f(x^* + t(x - x^*)) - f(x)}{t} \geq 0 \ \ \forall t \in (0, 1]$$

and $f'(x^*, x - x^*)$ is the limit of this quantity as $t \downarrow 0$. Letting $d = x - x^*$ and using positive homogeneity, we have $f'(x; d) \geq 0$ for all $d = \alpha(x - x^*)$ with $x \in X$ and $\alpha > 0$ (this objective is called the cone generated by $X - \{x^*\}$). $T_X(x^*)$ is the closure of this cone ([HUL96] Proposition III.5.2.1), and (iii) follows from the continuity of the finite convex function $f'(x^*; \cdot)$.

[(iii) ⇒ (i)] For any $x \in X$, we have $x - x^* \in T_X(x^*)$, and therefore (iii) implies that

$$0 \leq f'(x^*, x - x^*) \leq f(x) - f(x^*)$$

where the second inequality is due to Theorem 4. This implies (i).

[(iii) ⇔ (iv)] Since $f'(x; \cdot)$ is finite everywhere (Theorem 4), we can rewrite (iii) as

$$f'(x^*; d) + I_{T_X(x^*)}(d) \geq 0 \ \ \forall d \in \mathbb{R}^n.$$

The indicator $I_{T_X(x^*)}(d)$ is $\sigma_{N_X(x^*)}$, the support of the polar cone $N_X(x^*) = (T_X(x^*))^\circ$. Also, $f'(x^*; d)$ is the support of $\partial f(x^*)$ by definition in Theorem 5. Using Theorem 13, (iii) is equivalent to the condition that

$$0 \leq \sigma_{\partial f(x^*)}(d) + \sigma_{N_X(x^*)}(d) = \sigma_{\partial f(x^*) + N_X(x^*)}(d) \ \ \forall d \in \mathbb{R}^n.$$

The sum S of the compact convex set $\partial f(x^*)$ and the closed convex set $N_X(x^*)$ is a closed convex set. Suppose that $0 \notin S$. By Theorem 14, then there exists a nonzero vector $d_0 \in \mathbb{R}^n$ such that

$$\langle 0, d_0 \rangle > \sup\{\langle s, d_0 \rangle : s \in S\} = \sigma_S(d_0)$$

This is a contradiction. Therefore $0 \in S = \partial f(x^*) + N_X(x^*)$.

■

References

[HUL96] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex Analysis and Minimization Algorithms I*. Springer-Verlag, 1996.

[Roc97] R. T. Rockafellar. *Convex Analysis*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, 1997.