Fenchel Duality

3.1 Subgradients and Convex Functions

Ryota Iwamoto

February 22, 2023

We use the book; Convex Analysis and Nonlinear Optimization (author: J.M.BORWEIN and A.S.LEWIS), pp.33-36.

> We have already seen, in the First order sufficient condition (2.1.2), one benefit of convexity in optimization: critical points of convex functions are global minimizers. In this section we extend the types of functions we consider in two important ways:
>
> (i) We do not require f to be differentiable.
> (ii) We allow f to take the value $+\infty$.

This book Chapter 2 explains a optimization of convex functions with good conditions, which is differentiable and not including infinity points. In this section, we consider extended functions like being not differentiable and allowed to take the value $+\infty$.

> Our derivation of first order conditions in Section 2.3 illustrates the utility of considering nonsmooth functions even in the context of smooth problems. Allowing the value $+\infty$ lets us rephrase a problem like
>
> $$\inf \{g(x) \mid x \in C\}$$
>
> as $\inf (g + \delta_C)$, where the indicator function $\delta_C(x)$ is 0 for x in C and $+\infty$ otherwise.

In the case of having $g_1, \ldots, g_n$, these domains are usually different. However if we use $\delta_C(x)$, these domains is equal to $\mathbb{E}$.

Here We consider the definition of indicator function.

The domain of a function $f : \mathbb{E} \to (-\infty, +\infty]$ is the set

$$\text{dom} f = \{x \in \mathbb{E} \mid f(x) < +\infty\}.$$

We say f is convex if it is convex on its domain, and proper if its domain is nonempty. We call a function $g : \mathbb{E} \to [-\infty, +\infty)$ concave if $-g$ is convex, although for reasons of simplicity we will consider primarily convex functions. If a convex function f satisfies the stronger condition

$$f(\lambda x + \mu y) \le \lambda f(x) + \mu f(y), \ for \ all \ x, y \in \mathbb{E}, \ \lambda, \mu \in \mathbb{R}_+$$

we say f is *sublinear*. If $f(\lambda x) = \lambda f(x)$ for all x in $\mathbb{E}$ and λ in $\mathbb{R}_+$ then f is *positively homogeneous*: in particular this implies $f(0) = 0$. (Recall the convention $0 \cdot (+\infty) = 0$.) If $f(x + y) \le f(x) + f(y)$ for all x and y in $\mathbb{E}$ then we say f is *subadditive*. It is immediate that if the function f is sublinear then $-f(x) \le f(-x)$ for all x in $\mathbb{E}$. The *lineality space* of a sublinear function f is the set

$$\text{lin} f = \{x \in \mathbb{E} \mid -f(x) = f(-x)\}.$$

In many cases of convex analysis, we need to consider minimization problems. For the reason, we exclude $-\infty$ from a range of function.

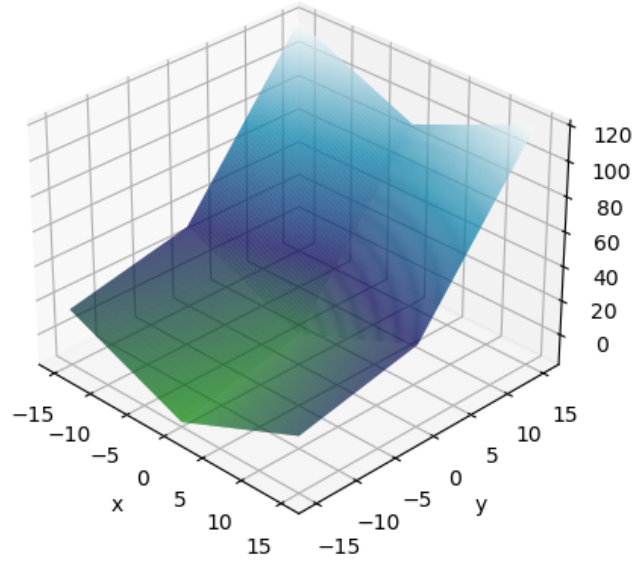We describe some definitions and the figure below.

Definition (Sublinear)

A function f is **sublinear** if this f satisfies the condition

$$f(\lambda x + \mu y) \le \lambda f(x) + \mu f(y), \ for \ all \ x, y \in \mathbb{E}, \ \lambda, \mu \in \mathbb{R}_+.$$

Figure:

$$f(x, y) = \begin{cases} 2|x| + y & \text{if } y \le 0, \\ 2|x| + 6y & \text{if } y > 0. \end{cases}$$

Definition (Positively Homogeneous and Subadditive)

A function f is **positively homogeneous** if this f satisfies the condition

$$f(\lambda x) = \lambda f(x), \ for \ all \ x \in \mathbb{E}, \ \lambda \in \mathbb{R}_+.$$

Also a function f is **subadditive** if this f satisfies the condition

$$f(x + y) \leq f(x) + f(y), \ for \ all \ x, y \in \mathbb{E}.$$

Proposition

If a proper function f is sublinear then $-f(x) \leq f(-x)$ for all x in $\mathbb{E}$.

Proof. We show that it holds $-f(x) \leq f(-x)$ for all x in $\mathbb{E}$.

By the assumption of sublinear of f, we can use the inequality

$$f(\lambda x + \mu y) \leq \lambda f(x) + \mu f(y), \ for \ all \ x, y \in \mathbb{E}, \ \lambda, \mu \in \mathbb{R}_+.$$

For any x and y, we take $y = -x$ since $\mathbb{E}$ is a vector space, and also take $\lambda = \mu = 1$. These values and the above assumption, we get the inequality

$$f(x + (-x)) \leq f(x) + f(-x).$$

By $f(0) = 0$,

$\because$ Since f is proper, $\exists x_0 \in \mathbb{E} \ s.t. \ f(x_0) < +\infty$. Then $f(0) = f(0 \cdot x) = 0 \cdot f(x) = 0$

3

$$0 \leq f(x) + f(-x).$$

Therefore we got $-f(x) \leq f(-x)$ for all x in $\mathbb{E}$. $\qquad\qquad\square$

Proposition (the former of Proposition 3.1.1)

If the function f is sublinear if and only if f is positively homogeneous and subadditive.

Proof. We show that

$$f \text{ is sublinear} \Leftrightarrow \text{f is positively homogeneous and subadditive.}$$

($\Rightarrow$) By the definition of sublinear of f, for any y we take $y = 0$ and $\mu = 0$. Then f is positively homogeneous. Like that, the subadditive of f holds; you should put $\lambda = \mu = 1$ to the inequality of sublinear.

($\Leftarrow$) By the definition of subadditive and positively homogeneous of f, $\forall x, y \in \mathbb{E}$, $\lambda, \mu \in \mathbb{R}_+$,

$$f(\lambda x + \mu y) \leq f(\lambda x) + f(\mu y) = \lambda f(x) + \mu f(y).$$

Therefore we completed to prove it. $\qquad\qquad\square$

Definition (Lineality Space)

The *lineality space* of a sublinear function f, denoted by $\mathrm{lin} f$, is the set

$$\mathrm{lin} f = \{x \in \mathbb{E} \mid -f(x) = f(-x)\}.$$

> The following result (whose proof is left as an exercise) shows this set is a subspace.

Proposition 3.1.1 (Sublinearity)

A function $f : \mathbb{E} \to (-\infty, +\infty]$ is sublinear if and only if it is positively homogeneous and subadditive. For a sublinear function f, the lineality space $\mathrm{lin} f$ is the largest subspace of $\mathbb{E}$ on which f is linear.

Proof. The former proposition has already been proved. Accordingly we show the latter one, that the way of proof is two steps;

    (I) $\mathrm{lin} f$ is a subspace of $\mathbb{E}$.

    (II) $\mathrm{lin} f$ is the "largest" subspace of $\mathbb{E}$.

(I) We show that

(i) $\forall x, y \in \lin f, x + y \in \lin f.$

(ii) $\forall x \in \lin f, \alpha \in \mathbb{R}, \alpha x \in \lin f$.

(i) We prove that $\forall x, y \in \lin f,$

$$-f(x + y) = f(-x - y).$$

For any x and y, by the definition of linearity space, it holds that

$$-f(x) = f(-x) \text{ and } -f(y) = f(-y).$$

And, Since f is sublinear, we have $f(x + y) \le f(x) + f(y)$. Using the result of the above proposition about a sublinear property, we can get the inequality:

$$-f(x + y) \ge -f(x) - f(y) = f(-x) + f(-y) \ge f(-x - y) = f(-(x + y)) \ge -f(x + y).$$

This inequality means that $-f(x + y) = f(-x - y)$.

(ii) We show that $\forall x \in \lin f, \alpha \in \mathbb{R},$

$$-f(\alpha x) = f(-\alpha x).$$

By the definition of positively homogeneous of f and the definition of lineality space,

$$-f(\alpha x) = -\alpha f(x) = \alpha f(-x) = f(-\alpha x).$$

For the result of (i) and (ii), $\lin f$ is a subspace of $\mathbb{E}$.

(II) We show that for a sublinear function f it holds that

$$\forall V \subset \mathbb{E}, \text{ where } V \text{ is subspace of } \mathbb{E}, V \subset \lin f$$

For any x in V, $-x \in V$ holds because V is a subspace of $\mathbb{E}$. Since f is linear,

$$0 = f(0) = f(x + (-x)) = f(x) + f(-x).$$

Therefore $\lin f$ is the largest subspace of $\mathbb{E}$. □

As in the First order sufficient condition (2.1.2), it is easy to check that if the point $\bar{x}$ lies in the domain of the convex function f then the directional derivative $f'(\bar{x}; \cdot)$ is well-defined and positively homogeneous, taking values in $[-\infty, +\infty]$. The core of a set C (written $\mathrm{core}(C)$) is the set of points x in C such that for any direction d in $\mathbb{E}$, $x + td$ lies in C for all small real t. This set clearly contains the interior of C, although it may be larger (Exercise 2).
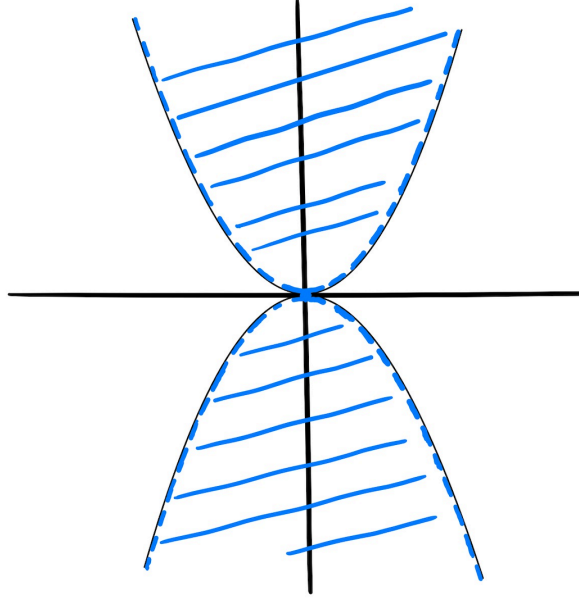
Exercise 2 (Core versus interior)

Considering the set in $\mathbb{R}^2$

$$D = \{(x, y) \mid y = 0 \; or \; |y| \geq x^2\} \;.$$

Prove $0 \in \text{core}(D) \backslash \text{int}(D)$

Figure(core(D)):

Proposition 3.1.2 (Sublinearity of the directional derivative)

If the function $f : \mathbb{E} \to (-\infty, +\infty]$ is convex then for any points $\bar{x}$ in $\text{core}(\text{dom} f)$, the directional derivative $f'(\bar{x}; \cdot)$ is everywhere finite and sublinear.

Proof. For d in $\mathbb{E}$ and nonzero t in $\mathbb{R}$, define

$$g(d; t) = \frac{f(\bar{x} + td) - f(\bar{x})}{t}.$$

By convexity we deduce, for $0 < t \leq s \in \mathbb{R}$, the inequality

$$g(d; -s) \leq g(d; -t) \leq g(d; t) \leq g(d; s).$$

$\because \forall t_1, t_2 (0 < t_1 \leq t_2)$, $\underset{(1)}{\underline{g(d; t_1) \leq g(d; t_2)}}$, $\underset{(2)}{\underline{g(d; -t_2) \leq g(d; -t_1)}}$, and $\underset{(3)}{\underline{g(d; -t_1) \leq g(d; t_1)}}$.

(1) We use the convexity of f.

(2) It's leaded from (1) $\times (-1)$.

(3) We can get since $f(\bar{x})$ is the middle points of $f(\bar{x} + td)$ and $f(\bar{x} - td)$.

Since $\bar{x}$ lies in core(dom$f$), for small $s > 0$ both $g(d; -s)$ and $g(d; s)$ are finite, so as $t \downarrow 0$ we have

$$+\infty > g(d; s) \geq g(d; t) \downarrow f'(\bar{x}; d) \geq g(d; -s) > -\infty.$$

Again by convexity we have, for any directions d and e in $\mathbb{E}$ and real $t > 0$,

$$g(d + e; t) \leq g(d; 2t) + g(e; 2t).$$

$\because$

$$\begin{aligned}
g(d + e; t) &= \frac{f(\bar{x} + t(d + e)) - f(\bar{x})}{t} \\
&= \frac{2f(\bar{x} + 2t(\frac{1}{2}d + \frac{1}{2}e)) - 2f(\bar{x})}{2t} \\
&= \frac{2f(\frac{1}{2}(\bar{x} + 2td) + \frac{1}{2}(\bar{x} + 2te)) - 2f(\bar{x})}{2t} \\
&\leq \frac{f(\bar{x} + 2td) - f(\bar{x})}{2t} + \frac{f(\bar{x} + 2te) - f(\bar{x})}{2t} \\
&= g(d; 2t) + g(e; 2t)
\end{aligned}$$

Now letting $t \downarrow 0$ gives subadditivity of $f'(\bar{x}; \cdot)$. The positively homogeneity is easy to check. $\because$ If $\lambda = 0$, it's clear. $\forall d \in \mathbb{E}, \lambda > 0, t > 0$,

$$\begin{aligned}
g(\lambda d; t) &= \frac{f(\bar{x} + t\lambda d) - f(\bar{x})}{t} \\
&= \frac{\lambda(f(\bar{x} + t\lambda d) - f(\bar{x}))}{\lambda t} \\
&= \lambda g(d; \lambda t)
\end{aligned}$$

$\square$

The idea of the derivative is fundamental in analysis because it allows us to approximate a wide class of functions using *linear functions*. In optimization we are concerned specifically with the minimization of functions, and hence often a *one-sided approximation* is sufficient. In place of the gradient we therefore consider *subgradients*, those elements ϕ of $\mathbb{E}$ satisfying

$$\langle \phi, x - \bar{x} \rangle \leq f(x) - f(\bar{x}) \text{ for all points } x \in \mathbb{E} \qquad (3.1.4)$$

We denote the set of subgradients (called the *subdifferential*) by $\partial f(\bar{x})$, defining $\partial f(\bar{x}) = \emptyset$ for $\bar{x}$ not in dom$f$. The subdifferential is always a closed convex set. We can think of $\partial f(\bar{x})$ as the value at $\bar{x}$ of the "multifunction" or "set-valued map" $\partial f : \mathbb{E} \to \mathbb{E}$. The importance of such maps is another of our themes. We define its *domain*

$$\text{dom}\partial f = \{x \in \mathbb{E} \mid \partial f(x) \neq \emptyset\}$$

(Exercise 19). We say f is *essentially strictly convex* if it is strictly convex on any convex subset of dom∂f.

The following very easy observation suggests the fundamental significance of subgradients in optimization.

Definition (Subgradient and Subdifferential)

Let $f : \mathbb{E} \to (-\infty, +\infty]$ and $\bar{x}$ in $\mathbb{E}$. If it holds that

$$\exists \phi \in \mathbb{E} \ s.t. \ f(x) \geq f(\bar{x}) + \langle \phi, x - \bar{x} \rangle, \forall x \in \mathbb{E}$$

then ϕ is called the subgradient of f at $\bar{x}$.

Also, the set of all subgradients of f at $\bar{x}$ is called the subdifferential of f at $\bar{x}$, denoted by $\partial f(\bar{x})$, that is
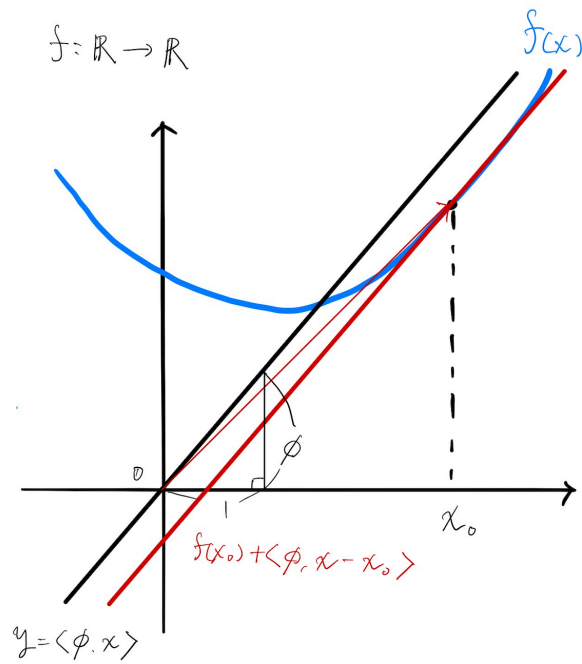
$$\partial f(\bar{x}) := \{\phi \in \mathbb{E} \mid f(x) \geq f(\bar{x}) + \langle \phi, x - \bar{x} \rangle, \forall x \in \mathbb{E}\}.$$

Observation(1):

We consider the inequality as the relation between two functions. The graph is written as the shifted one of $y = \langle \phi, x \rangle$ with the direction of $(x_0, f(x_0))$.
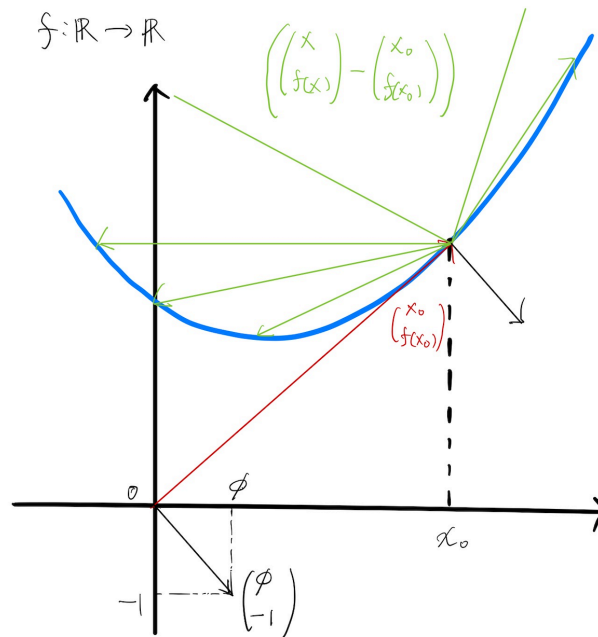
$$y = f(x)$$
$$y = \langle \phi, x \rangle + f(x_0)$$

Figure:

Observation(2):

We consider the inequality as the inner product.

$$f(x) \geq f(\bar{x}) + \langle \phi, x - \bar{x} \rangle$$
$$\langle \phi, x - \bar{x} \rangle + f(\bar{x}) - f(x) \leq 0$$
$$\langle \phi, x - \bar{x} \rangle + (-1)(f(x) - f(\bar{x})) \leq 0$$
$$\left\langle \begin{pmatrix} \phi \\ -1 \end{pmatrix}, \begin{pmatrix} x \\ f(x) \end{pmatrix} - \begin{pmatrix} x_0 \\ f(x_0) \end{pmatrix} \right\rangle \leq 0$$

Figure:

$f : \mathbb{R} \to \mathbb{R}$

$\left(\begin{pmatrix} x \\ f(x) \end{pmatrix} - \begin{pmatrix} x_o \\ f(x_o) \end{pmatrix} \right)$

$\begin{pmatrix} x_o \\ f(x_o) \end{pmatrix}$

0

ϕ

x_o

-1

$\begin{pmatrix} \phi \\ -1 \end{pmatrix}$

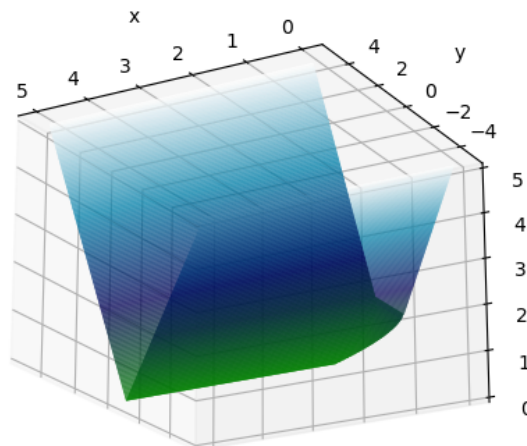**Exercise 19 (Domain of subdifferential)**

If the function $f : \mathbb{R}^2 \to (-\infty, +\infty]$ is defined by

$$f(x_1, x_2) = \begin{cases} \max\{1 - \sqrt{x_1}, |x_2|\} & \text{if } x_1 \geq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

prove that f is convex but that $\mathrm{dom}\,\partial f$ is not convex.

Figure:

> **Proposition 3.1.5 (Subgradients at optimality)**
>
> For any proper functions $f : \mathbb{E} \to (-\infty, +\infty]$, the points $\bar{x}$ is a (global) minimizer of f if and only if the condition $0 \in \partial f(\bar{x})$ holds.

Proof. The definition that $\bar{x} \in \mathbb{E}$ is a global minimizer of f is

$$\forall x \in \mathbb{E}, f(\bar{x}) \le f(x). \tag{1}$$

Also $0 \in \partial f$ means that

$$\forall x \in \mathbb{E}, f(x) \ge f(\bar{x}) + \langle 0, x - \bar{x} \rangle. \tag{2}$$

We prove $(1) \Rightarrow (2)$ and $(1) \Leftarrow (2)$.

$(\Rightarrow)$ Paying attention to the property of proper of f, $\forall x \in \mathbb{E}$

$$f(\bar{x}) \le f(x)$$
$$0 \le f(x) - f(\bar{x})$$
$$\langle 0, x - \bar{x} \rangle \le f(x) - f(\bar{x}).$$

It means that $0 \in \partial f(\bar{x})$ holds.

$(\Leftarrow)$ By (2), like above, $\forall x \in \mathbb{E}$

$$\langle 0, x - \bar{x} \rangle \le f(x) - f(\bar{x})$$
$$0 \le f(x) - f(\bar{x})$$
$$f(\bar{x}) \le f(x).$$

That is, $\bar{x}$ is a global minimizer of f.

Therefore we have got $(1) \Leftrightarrow (2)$.

$\square$

> Alternatively put, minimizers of f correspond exactly to "zeroes" of ∂f. The derivative is a local property whereas the subgradient definition (3.1.4) describes a global property. The main result of this section shows that the set of subgradients of a convex function is usually *nonempty*, and that we can describe it locally in terms of the directional derivative. We begin with another simple exercise.

> **Proposition 3.1.6 (Subgradients and directional derivatives)**
>
> If the function $f : \mathbb{E} \to (-\infty, +\infty]$ is convex and the point $\bar{x}$ lies in $\mathrm{dom} f$, then an element ϕ of $\mathbb{E}$ is a subgradient of f at $\bar{x}$ if and only if it satisfies $\langle \phi, \cdot \rangle \le f'(\bar{x}; \cdot)$.

(1) ϕ is a subgradient of f at $\bar{x}$.

(2) $\forall d \in \mathbb{E}, \langle \phi, d \rangle \leq f'(\bar{x}; d)$.

Proof. We show $(1) \Rightarrow (2)$ and $(1) \Leftarrow (2)$.

  $(\Rightarrow)$

  $(\Leftarrow)$

$\square$

> The idea behind the construction of a subgradient for a function f that we present here is rather simple. recursively construct a decreasing sequence of sublinear functions which, after translation, minorize f. At each step we guarantee one extra direction of linearity. The basic step is summarized in the following exercise.

Lemma 3.1.7

Suppose that the function $p : \mathbb{E} \to (-\infty, +\infty]$ is sublinear and that the points $\bar{x}$ lies in core(dom p). Then the function $q(\cdot) = p'(\bar{x}; \cdot)$ satisfies the conditions

  (i) $q(\lambda \bar{x}) = \lambda p(\bar{x})$ for all real λ,

  (ii) $q \leq p$, and

  (iii) $\mathrm{lin}q \supset \mathrm{lin}p + \mathrm{span}\{\bar{x}\}$.

Proof. We show these relations.

  (i) Since $\bar{x} \in \mathrm{core}(\mathrm{dom}p)$, for a given λ

$$\exists \hat{t} > 0 \; s.t. \; \forall t \in (0, \hat{t}), \bar{x} + t(\lambda \bar{x}) \in \mathrm{dom}p \text{ and } \bar{x} + t(-\lambda \bar{x}) \in \mathrm{dom}p. \tag{1}$$

Also given $\lambda \in \mathbb{R}$, it holds that

$$\exists \widetilde{t} > 0 \; s.t. \; \forall t \in (-\widetilde{t}, \widetilde{t}), 1 + \lambda t > 0 \text{ and } 1 - \lambda t < 0. \tag{2}$$

For the t which satisfies (1) and (2), i.e., for $t = \min\{\hat{t}, \widetilde{t}\} > 0$,

$$\begin{aligned}
q(\lambda \bar{x}) = p'(\bar{x}; \lambda \bar{x}) &= \lim_{t \downarrow 0} \frac{p(\bar{x} + \lambda \bar{x} t) - p(\bar{x})}{t} \\
&= \lim_{t \downarrow 0} \frac{p((1 + \lambda t)\bar{x}) - p(\bar{x})}{t} \\
&= \lim_{t \downarrow 0} \frac{(1 + \lambda t)p(\bar{x}) - p(\bar{x})}{t} \quad \because f \text{ is sublinear} \\
&= \lambda p(\bar{x}).
\end{aligned}$$

  (ii)$\forall x \in \mathbb{E}$, there exist $t > 0$ satisfying (1) that is described with not $\bar{x}$ but x,

$$q(x) = p'(\bar{x}; x) = \lim_{t \downarrow 0} \frac{p(\bar{x} + tx) - p(\bar{x})}{t}$$

$$\leq \lim_{t \downarrow 0} \frac{p(\bar{x}) + tp(x) - p(\bar{x})}{t} \quad \because f \text{ is sublinear}$$

$$= p(x).$$

(iii) We show that

$$\forall y \in \lim p + \mathrm{span}\{\bar{x}\}, y \in \lim q.$$

Since q is sublinear, it holds that $-q(y) \leq q(-y)$. Accordingly we prove the reverse, $-q(y) \geq q(-y)$.

As $y \in \lim p + \mathrm{span}\{\bar{x}\}$,

$$\exists x \in \lim p, \alpha \in \mathbb{R} \ s.t. \ y = x + \alpha\bar{x}.$$

Also since x is in $\lim p$, $-p(x) = p(-x)$ holds by the definition of lineality space.

Then,

$$q(-y) = q(-(x + \alpha\bar{x}))$$

$$\leq q(-x) + q(-\alpha\bar{x}) \because q \text{ is sublinear}$$

$$\leq p(-x) + -p(\alpha\bar{x}) \because \text{Lemma3.1.7 (i), (ii)}$$

$$= -p(x) + -p(\alpha\bar{x}) \because x \in \lim p$$

$$\leq -q(x) - q(\alpha\bar{x}) \because \text{Lemma3.1.7 (ii)}$$

$$= (-1)(q(x) + q(\alpha\bar{x}))$$

$$= (-1)(q(x + \alpha\bar{x}))$$

$$= -q(y).$$

Therefore it holds that $\lim q \supset \lim p + \mathrm{span}\{\bar{x}\}$.

$\square$

> With this tool we are now ready for the main result, which gives conditions guaranteeing the existence of a subgradient. Proposition 3.1.6 showed how to identify subgradients from directional derivatives; this next result shows how to move in the reverse direction.

> **Theorem 3.1.8 (Max formula)**
>
> If the function $f : \mathbb{E} \to (-\infty, +\infty]$ is convex then any points $\bar{x}$ in core(dom$f$) and any direction $d \in \mathbb{E}$ satisfying
>
> $$f'(\bar{x}; d) = \max\{\langle \phi, d \rangle \mid \phi \in \partial f(\bar{x})\} \qquad (3.1.9)$$
>
> In particular, the subdifferential $\partial f(\bar{x})$ is nonempty.

Proof. In view of Proposition 3.1.6 we simply have to show that for any fixed d in $\mathbb{E}$ there is a subgradient ϕ satisfying $\langle \phi, d \rangle = f'(\bar{x}; d)$.

Choose a bases $\{e_1, e_2, \ldots, e_n\}$ for $\mathbb{E}$ with $e_1 = d$ if d is nonzero. Now define a sequence of functions $p_0, p_1, \ldots, p_n$ recursively by $p_0(\cdot) = f'(\bar{x}; \cdot)$, and $p_k(\cdot) = p'_{k-1}(e_k; \cdot)$ for $k = 1, 2, \ldots, n$. We essentially show that $p_n(\cdot)$ is the required subgradient. (Remark: If $p_n(\cdot) = \langle \phi, \cdot \rangle$ and $p_0(\cdot) = p_n(\cdot)$, we can get $f'(\bar{x}; \cdot) = p_0(\cdot) = p_n(\cdot) = \langle \phi, \cdot \rangle$.)

First note that, by Proposition 3.1.2, each p_k is everywhere finite and sublinear. By part (iii) of Lemma 3.1.7 we know

$$\text{lin } p_k \supset \text{lin } p_{k-1} + \text{span } \{e_k\} \text{ for } k = 1, 2, \ldots, n.$$

so p_n is linear.

$\therefore$

For $k = 1, 2, \ldots, n$,

$$\text{lin } p_k \supset \text{lin } p_{k-1} + \text{span } \{e_k\}$$
$$\text{lin } p_k \supset \text{lin } p_{k-2} + \text{span } \{e_{k-1}\} + \text{span } \{e_k\}$$
$$\vdots$$
$$\text{lin } p_k \supset \text{lin } p_0 + \text{span } \{e_1\} + \cdots + \text{span } \{e_k\}.$$

That is, lin $p_k = \mathbb{E}$.

Thus there is an element ϕ of $\mathbb{E}$ satisfying $\langle \phi, \cdot \rangle = p_n(\cdot)$. Part (ii) of Lemma 3.1.7 implies $p_n \le p_{n-1} \le \cdots \le p_0$, so certainly, by Proposition 3.1.6, any point x in E satisfies

$$p_n(x - \bar{x}) \le p_0(x - \bar{x}) = f'(\bar{x}; x - \bar{x}) \le f(x) - f(\bar{x}).$$

$\therefore$

To identify ϕ is a subgradient, we show the inequality:

$$f(x) \le f(\bar{x}) + \langle \phi, x - \bar{x} \rangle.$$

By the definition of the directional derivative,

$$f'(\bar{x}; x - \bar{x}) = \lim_{x \downarrow 0} \frac{f(\bar{x} + t(x - \bar{x})) - f(\bar{x})}{t} \leq \frac{f(\bar{x} + 1(x - \bar{x})) - f(\bar{x})}{1} = f(x) - f(\bar{x}).$$

Thus ϕ is a subgradient. If d is zero then we have $p_n = 0 = f'(\bar{x}; 0)$.

Finally, if d is nonzero then by part (i) of Lemma 3.1.7 we see

$$p_n(d) \leq p_0(d) = p_0(e_1) = -p_0'(e_1; -e_1) =$$
$$-p_1(-e_1) = -p_1(-d) \leq -p_n(-d) = p_n(d).$$

whence $p_n(d) = p_0(d) = f'(\bar{x}; d)$.

$\square$