Fenchel Duality

3.1 Subgradients and Convex Functions

Ryota Iwamoto

February 14, 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.

Definition (Indicator Funciton)

The indicator function of a set C of $\mathbb{E}$, denoted by δ_C, is defined by

$$\delta_C(x) = \begin{cases} 0 & if\, x \in C, \\ \infty & otherwise. \end{cases}$$

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

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

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

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

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 $f(x + y) \leq f(x) + f(y)$ 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 $-f(x) \leq f(-x)$ for all x in $\mathbb{E}$. The *lineality space* of a sublinear function f is the set

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

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.

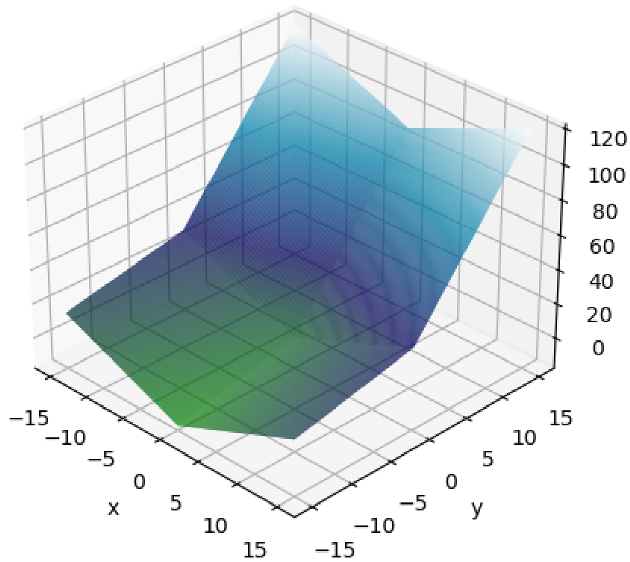
Definition (Sublinear)

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

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

Figure:

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



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 $f(0) = f(0 \cdot x) = 0 \cdot f(x) = 0$

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

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

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. $\square$

Definition (Lineality Space)

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

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

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

(I) We show that

(i) $\forall x, y \in linf$, $x + y \in linf$.
(ii) $\forall x \in linf$, $\alpha \in \mathbb{R}$, $\alpha x \in linf$.

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

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

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

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

This inequality means that $-f(x+y) = f(-x-y)$.
(ii) We show that $\forall x \in \mathrm{lin} f$, $\alpha \in \mathbb{R}$,

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

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), $\mathrm{lin} f$ is a subspace of $\mathbb{E}$.
(II) We show that for a sublinear function f it holds that

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

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 $\mathrm{lin} f$ 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 $core(C)$) 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).

Definition (the core of a set)

Let $C \subset \mathbb{E}$. The **core** of the set C is defined as

$$core(C) := \{x \in \mathbb{E} \mid \forall d \in \mathbb{E}, \exists t_0 > 0 \, s.t. \, \forall t(0 < t < t_0), x + td \in C\} \ .$$

Exercise 2 (Core versus interior)

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

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

Prove $0 \in core(D) \backslash int(D)$

Figure:

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 $core(domf)$, 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

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

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

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

$\because \forall t_1, t_2 (t_1 \leq t_2) \in \mathbb{R}$, $g(d; t_1) \leq g(d; t_2)$.

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

$$+\infty > g(d; s) \geq g(d; t) \downarrow f'(\bar{x}; d) \geq g(d; -s) > -\infty. \tag{1}$$

Again bu 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. $\because$ Like that, $\forall d \in \mathbb{E}, \lambda \geq 0, t > 0$,

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

$\square$

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

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

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 $\mathrm{dom} f$. 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*

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

(Exercise 19). We say f is *essentially strictly convex* if it is strictly convex on any convex subset of $\mathrm{dom}\partial f$.
The following very easy observation suggests the fundamental significance of subgradients in optimization.

Definition (Subgtarients 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

$$\partial f(\bar{x}) \coloneqq \{\phi \in \mathbb{E} \mid f(x) \geq f(\bar{x}) + \langle \phi, x - \bar{x} \rangle, \forall x \in \mathbb{E}\}.$$

Exercise 19 (Domain of subdifferential)

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

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

Prove $0 \in core(D) \backslash int(D)$