A TUTORIAL ON CONVEX SUBDIFFERENTIAL CALCULUS

April 24, 2015

Nguyen Mau Nam

Abstract. In this paper, we provide an easy way to access calculus rules for the subdifferential in the sense of convex analysis in finite dimensions.

Key words. subgradient, subdifferential, coderivative, generalized differentiation.

AMS subject classifications. 49J52, 49J53, 90C31.

1 Introduction

In the 1960’s, Jean-Jacques Moreau and R. Tyrrell Rockafellar independently introduced a generalized differentiation concept for convex functions that are not necessarily differentiable called the subdifferential. It then becomes one of the most central concepts of the field convex analysis. The subdifferential in the sense of convex analysis plays a crucial role in many applications of convex analysis to optimization from both theoretical and numerical aspects. The choice of convex functions to start with is based on the fact that the class of convex functions has several favorable properties for applications to optimization. For example, for a convex function defined on $\mathbb{R}^n$, any local minimizer is an absolute minimizer. If the derivative of a differentiable convex function vanishes at a point, then that point is an absolute minimizer of the function.

In contrast to the classical derivative which takes single values, the subdifferential of a convex function at a point is a set in general. The set-valued property of the subdifferential mapping makes it challenging for developing subdifferential calculus. However, the subdifferential in the sense of convex analysis possesses nice geometric representations. Thus, it is possible to develop its calculus based on geometric properties of sets, namely, convex separation. Convex subdifferential calculus was first developed in the works of R. Tyrrell Rockafellar in his book “Convex Analysis” [10] published in 1970. Since then, several monographs have been written to have a more complete picture of convex analysis in both finite and infinite dimensions; see [1, 2, 4, 7, 9, 13] and the references therein.

In the time when convex analysis has become more and more important for applications to many fields such as computational statistics, machine learning and sparse optimization, we revisit the convex subdifferential and provide an easy way to excess major convex subdifferential calculus rules in finite dimensions. The subdifferential calculus is based on the proper separation property with the use of relative interiors of convex sets. Some part of the paper has been outlined in the exercises of our book “An Easy Path to Convex Analysis and Applications”. This paper provides a complete development of of the calculus serving as supplemental material for the book.

1 Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, PO Box 751, Portland, OR 97207, United States (mau.nam.nguyen@pdx.edu). The research of Nguyen Mau Nam was partially supported by the NSF under grant #1411817 and the Simons Foundation under grant #208785.
2 Basic Properties of Convex Sets

In this section, we introduce some basic concepts and properties of convex sets. The detailed proofs for all results presented here and in the next section can be found in [9]. Throughout the paper, we consider the Euclidean space $\mathbb{R}^n$ of all $n$-tuples of real numbers with the inner product given by

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i,$$

where $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$. The Euclidean norm induced by this inner product is

$$\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

We often identify each element $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ with the column vector $x = [x_1, \ldots, x_n]^T$.

The line connecting two points $a$ and $b$ in $\mathbb{R}^n$ is given by

$$L[a, b] := \{ \lambda a + (1 - \lambda) b \mid \lambda \in \mathbb{R} \},$$

and the line segment/interval connecting $a$ and $b$ is given by

$$[a, b] := \{ \lambda a + (1 - \lambda) b \mid \lambda \in [0, 1] \}.$$

A subset $\Omega$ of $\mathbb{R}^n$ is called convex if $\lambda x + (1 - \lambda) y \in \Omega$ for all $x, y \in \Omega$ and $\lambda \in (0, 1)$. From this definition, we see that $\Omega$ is convex if and only if the line segment connecting any two points of $\Omega$ is a subset of $\Omega$.

A mapping $B: \mathbb{R}^n \to \mathbb{R}^p$ is called affine if there exist an $n \times p$ matrix $A$ and an element $b \in \mathbb{R}^p$ such that

$$B(x) = Ax + b$$

for all $x \in \mathbb{R}^n$. The convexity of sets is preserved under affine mappings as described in the next proposition.

**Proposition 2.1** Let $B: \mathbb{R}^n \to \mathbb{R}^p$ be an affine mapping.

(i) If $\Omega$ is a convex subset of $\mathbb{R}^n$, then the direct image $B(\Omega)$ is a convex subset of $\mathbb{R}^p$.

(ii) If $\Theta$ is a convex subset of $\mathbb{R}^p$, then the inverse image $B^{-1}(\Theta)$ is a convex subset of $\mathbb{R}^n$.

From the definition, it is obvious that if $\{\Omega_i\}_{i \in I}$ is a collection of convex sets in $\mathbb{R}^n$, then the intersection $\bigcap_{i \in I} \Omega_i$ is convex. In particular, the intersection of any two convex sets is a convex set. This property motivates the definition of the convex hull of an arbitrary subset of $\mathbb{R}^n$. Given a subset $\Omega \subset \mathbb{R}^n$, define the convex hull of $\Omega$ by

$$\text{co } \Omega := \bigcap \{ C \mid C \text{ is convex and } \Omega \subset C \}.$$

Equivalently, the convex hull of a set $\Omega$ is the smallest convex set containing $\Omega$.

The following important result is a direct consequence of the definition.
Proposition 2.2  For any subset $\Omega$ of $\mathbb{R}^n$, its convex hull admits the representation

$$\text{co } \Omega = \left\{ \sum_{i=1}^{m} \lambda_i w_i \mid \sum_{i=1}^{m} \lambda_i = 1, \lambda_i \geq 0, w_i \in \Omega, m \in \mathbb{N} \right\}.$$ 

A subset $\Omega$ of $\mathbb{R}^n$ is called affine if for any $x, y \in \Omega$ and for any $\lambda \in \mathbb{R}$, one has

$$\lambda x + (1 - \lambda)y \in \Omega.$$ 

Geometrically, $\Omega$ is affine if and only if the line connecting any two points $a, b \in \Omega$ is a subset of $\Omega$. From the definition, we can show that the intersection of a collection of affine sets is an affine set.

The affine hull of $\Omega$ is given by

$$\text{aff } (\Omega) := \bigcap \left\{ A \mid A \text{ is affine and } \Omega \subset A \right\}.$$ 

Proposition 2.3  For any subset $\Omega$ of $\mathbb{R}^n$, its affine hull admits the representation

$$\text{aff } (\Omega) = \left\{ \sum_{i=1}^{m} \lambda_i \omega_i \mid \sum_{i=1}^{m} \lambda_i = 1, \omega_i \in \Omega, m \in \mathbb{N} \right\}.$$ 

We are now ready to present the concept of relative interior, which plays a central role in developing convex subdifferential calculus.

Definition 2.4  An element $x$ belongs to the relative interior of $\Omega$, $x \in \text{ri } (\Omega)$, if there exists $\delta > 0$ such that

$$B(x, \delta) \cap \text{aff } (\Omega) \subset \Omega,$$

where $B(x, \delta)$ is the closed ball centered at $x$ with radius $\delta$.

The following simple proposition is useful in that follows and serves as an example for better understanding of the relative interior of a convex set.

Proposition 2.5  Let $\Omega$ be a nonempty convex set. Suppose that $\bar{x} \in \text{ri } \Omega$ and $\bar{y} \in \Omega$. Then there exists $t > 0$ such that

$$\bar{x} + t(\bar{x} - \bar{y}) \in \Omega.$$ 

Proof.  Choose $\delta > 0$ such that

$$B(x, \delta) \cap \text{aff } (\Omega) \subset \Omega.$$ 

Note that $\bar{x} + t(\bar{x} - \bar{y}) = (1 + t)\bar{x} + (1 - t)\bar{y} \in \text{aff } (\Omega)$ for all $t \in \mathbb{R}$ as it is an affine combination of $\bar{x}$ and $\bar{y}$. Choose $t > 0$ sufficiently small such that $\bar{x} + t(\bar{x} - \bar{y}) \in B(x, \delta)$. Then $\bar{x} + t(\bar{x} - \bar{y}) \in B(x, \delta) \cap \text{aff } (\Omega) \subset \Omega$.  

Given two elements $a, b \in \mathbb{R}^n$, define the half-open interval

$$[a, b) := \{ta + (1 - t)b \mid 0 < t \leq 1\}.$$
**Theorem 2.6** Let $\Omega$ be a nonempty convex subset of $\mathbb{R}^n$. Then

(i) $\text{ri} (\Omega) \neq \emptyset$, and

(ii) $(a, b) \subset \text{ri} (\Omega)$ for any $a \in \text{ri} (\Omega)$ and $b \in \bar{\Omega}$.

The theorem below gives a convenient way to compute the relative interior of the direct image of a convex set via an affine mapping in terms of the relative interior of the set.

**Theorem 2.7** Let $B : \mathbb{R}^n \to \mathbb{R}^p$ be an affine mapping and let $\Omega$ be a convex subset of $\mathbb{R}^n$. Then

$$B(\text{ri} \, \Omega) = \text{ri} \, B(\Omega).$$

Consequently, we obtain the corollary below which will be used in what follows. Recall that the difference of two subsets $A$ and $B$ of $\mathbb{R}^n$ is defined by

$$A - B := \{a - b \mid a \in A \text{ and } b \in B\}.$$

**Corollary 2.8** Let $\Omega_1$ and $\Omega_2$ be convex subsets of $\mathbb{R}^n$. Then

$$\text{ri} \, (\Omega_1 - \Omega_2) = \text{ri} \, \Omega_1 - \text{ri} \, \Omega_2.$$

**Proof.** Consider the linear mapping $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ given by $B(x, y) = x - y$. Form the cartesian product $\Omega = \Omega_1 \times \Omega_2$. Then

$$B(\Omega) = \Omega_1 - \Omega_2.$$

Consequently,

$$\text{ri} \, (\Omega_1 - \Omega_2) = \text{ri} \, B(\Omega) = B(\text{ri} \, \Omega) = B(\text{ri} \, (\Omega_1 \times \Omega_2)) = B(\text{ri} \, \Omega_1 \times \text{ri} \, \Omega_2) = \text{ri} \, \Omega_1 - \text{ri} \, \Omega_2.$$

In the proof we use the fact that $\text{ri} \, (\Omega_1 \times \Omega_2) = \text{ri} \, \Omega_1 \times \text{ri} \, \Omega_2$, which can be proved using the definition. $\square$

Given a set $\Omega \subset \mathbb{R}^n$, the *distance function* associated with $\Omega$ is defined by

$$d(x; \Omega) := \inf \{\|x - \omega\| \mid \omega \in \Omega\}.$$

For each $x \in \mathbb{R}^n$, the *Euclidean projection* from $x$ to $\Omega$ is defined by

$$\pi(x; \Omega) := \{\omega \in \Omega \mid \|x - \omega\| = d(x; \Omega)\}.$$

We can show that if $\Omega$ is a nonempty closed convex subset of $\mathbb{R}^n$, then $\pi(\bar{x}; \Omega)$ is a singleton for every $\bar{x} \in \mathbb{R}^n$.

The following proposition plays a crucial role in proving major results on convex separation.

**Proposition 2.9** Let $\Omega$ be a nonempty convex subset of $\mathbb{R}^n$ and let $\bar{\omega} \in \Omega$. Then we have

$$\bar{\omega} = \pi(\bar{x}; \Omega) \text{ if and only if } \langle \bar{x} - \bar{\omega}, \omega - \bar{\omega} \rangle \leq 0 \text{ for all } \omega \in \Omega.$$

(2.1)
Proof. Let us prove (2.1) under the assumption that $\bar{\omega} \in \pi(\bar{x}; \Omega)$. Fix any $t \in (0, 1)$ and $\omega \in \Omega$. Then $t \omega + (1 - t)\bar{\omega} \in \Omega$ and by the definition,

$$||\bar{x} - \omega||^2 \leq ||\bar{x} - [t \omega + (1 - t)\bar{\omega}]||^2.$$ 

This implies

$$||\bar{x} - \bar{\omega}||^2 \leq ||\bar{x} - [\bar{\omega} + t(\omega - \bar{\omega})]||^2 = ||\bar{x} - \bar{\omega}||^2 - 2t \langle \bar{x} - \bar{\omega}, \omega - \bar{\omega} \rangle + t^2 ||\omega - \bar{\omega}||^2.$$ 

Consequently,

$$2 \langle \bar{x} - \bar{\omega}, \omega - \bar{\omega} \rangle \leq t ||\omega - \bar{\omega}||^2.$$ 

Letting $t \to 0^+$ gives (2.1).

Now suppose that (2.1) is satisfied. For any $\omega \in \Omega$, one has

$$||\bar{x} - \omega||^2 = ||\bar{x} - \bar{\omega} + \bar{\omega} - \omega||^2 = ||\bar{x} - \bar{\omega}||^2 + 2 \langle \bar{x} - \bar{\omega}, \bar{\omega} - \omega \rangle + ||\bar{\omega} - \omega||^2 \geq ||\bar{x} - \omega||^2.$$ 

Thus, $||\bar{x} - \bar{\omega}|| \leq ||\bar{x} - \omega||$ for all $\omega \in \Omega$, and hence $\bar{\omega} = \pi(\bar{x}; \Omega)$. □

3 Basic Properties of Convex Functions

In this section, we discuss the definition and basic properties of convex functions in $\mathbb{R}^n$. We work with extended-real-valued functions which take values in $(-\infty, \infty] = \mathbb{R} \cup \{\infty\}$. The following conventions are used for arithmetic calculations on $(-\infty, \infty]$:

$$\alpha + \infty = \infty + \alpha = \infty \text{ for all } \alpha \in \mathbb{R},$$

$$\infty + \infty = \infty,$$

$$\alpha \cdot \infty = \infty \cdot \alpha \text{ for all } \alpha > 0,$$

$$\infty \cdot \infty = \infty,$$

$$0 \cdot \infty = \infty \cdot 0 = 0.$$

Definition 3.1 Let $\Omega$ be a nonempty convex set and let $f : \mathbb{R}^n \to (-\infty, \infty]$ be an extended-real-valued function. The function $f$ is called convex if

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \text{ for all } x, y \in \mathbb{R}^n \text{ and } \lambda \in (0, 1).$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be an extended-real-valued function. The domain and epigraph of $f$ are defined respectively by

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

$$\text{epi } f := \{(x, \alpha) \in \mathbb{R}^{n+1} \mid x \in \mathbb{R}^n \text{ and } \alpha \geq f(x)\}.$$ 

We can prove that $f$ is a convex function on $\mathbb{R}^n$ if and only if its epigraph is a convex set in $\mathbb{R}^{n+1}$. Moreover, if $f : \mathbb{R}^n \to (-\infty, \infty]$ is a convex function, then its domain is a convex subset of $\mathbb{R}^n$.

To illustrate these concepts, we consider the following simple example.
Example 3.2 Consider the extended-real-valued function given by

\[ f(x) = \begin{cases} 
0 & \text{if } |x| \leq 1, \\
\infty & \text{otherwise.} 
\end{cases} \]

Based on the definition, we can show that \( f \) is a convex function with \( \text{dom } f = [-1, 1] \) and \( \text{epi } f = [-1, 1] \times [0, \infty) \).

The class of convex functions is favorable for optimization applications. The proposition below gives an important feature of convex function.

Let \( f : \mathbb{R}^n \to (-\infty, \infty] \) be a convex function and let \( \bar{x} \in \text{dom } f \). Recall that \( f \) has a local minimum at \( \bar{x} \) if there exists \( \delta > 0 \) such that

\[ f(x) \geq f(\bar{x}) \text{ for all } x \in B(\bar{x}; \delta). \]

If the inequality above holds for all \( x \in \mathbb{R}^n \), we say that \( f \) has an absolute minimum at \( \bar{x} \).

Proposition 3.3 Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a convex function. Then \( f \) has a local minimum at \( \bar{x} \) if and only if \( f \) has an absolute minimum at \( \bar{x} \).

Proof. We only need to prove the implication since the converse is trivial. Suppose that \( f \) has a local minimum at \( \bar{x} \). Then there exists \( \delta > 0 \) with

\[ f(u) \geq f(\bar{x}) \text{ for all } u \in B(\bar{x}; \delta). \]

For any \( x \in \mathbb{R} \), one has that \( x_k = (1 - \frac{1}{k})\bar{x} + \frac{1}{k}x \to \bar{x} \). Thus, \( x_k \in B(\bar{x}; \delta) \) when \( k \) is sufficiently large. It follows that

\[ f(\bar{x}) \leq f(x_k) \leq (1 - \frac{1}{k})f(\bar{x}) + \frac{1}{k}f(x). \]

This implies

\[ \frac{1}{k}f(\bar{x}) \leq \frac{1}{k}f(x), \]

and hence \( f(\bar{x}) \leq f(x) \). Therefore, \( f \) has an absolute minimum at \( \bar{x} \).

We are now ready to present a generalized differentiation concept for convex functions called the subdifferential. This concept was introduced in the 1960’s by Jean-Jacques Moreau and R. Tyrrell Rockafellar in order to deal with convex functions that are not necessarily differentiable. This concept turns out to be very useful in applications to optimization from both theoretical and numerical aspects.

Definition 3.4 An element \( v \in \mathbb{R}^n \) is called a subgradient of a convex function \( f : \mathbb{R}^n \to \mathbb{R} \) at \( \bar{x} \in \mathbb{R}^n \) if it satisfies the inequality

\[ f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle \text{ for all } x \in \mathbb{R}^n. \]

The set of all subgradients of \( f \) at \( \bar{x} \) is called the subdifferential of the function at \( \bar{x} \) and is denoted by \( \partial f(\bar{x}) \).
Directly from the definition, one has the following subdifferential Fermat rule:

\[ f \text{ has an absolute minimum at } \bar{x} \text{ if and only if } 0 \in \partial f(\bar{x}). \]

Let \( f : \mathbb{R}^n \to (-\infty, \infty] \) be an extended-real-valued function and let \( \bar{x} \in \text{dom } f \). We say that \( f \) is Fréchet differentiable or simply differentiable at \( \bar{x} \) if there exists an element \( v \in \mathbb{R}^n \) such that

\[
\lim_{x \to \bar{x}} \frac{f(x) - f(\bar{x}) - \langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} = 0.
\]

If \( f \) is differentiable at \( \bar{x} \), we can show that the element \( v \) is and is called the gradient of \( f \) at \( \bar{x} \) denoted by \( \nabla f(\bar{x}) \).

The proposition below shows that the subdifferential of a convex function at a given point reduces to the gradient at that point when the function is differentiable.

**Proposition 3.5** Suppose that \( f : \mathbb{R}^n \to \mathbb{R} \) is convex and Fréchet differentiable at \( \bar{x} \). Then

\[
\langle \nabla f(\bar{x}), x - \bar{x} \rangle \leq f(x) - f(\bar{x}) \text{ for all } x \in \mathbb{R}^n.
\] (3.1)

Moreover, \( \partial f(\bar{x}) = \{\nabla f(\bar{x})\} \).

**Proof.** Since \( f \) is Fréchet differentiable at \( \bar{x} \), by definition, for any \( \epsilon > 0 \), there exists \( \delta > 0 \) such that

\[
-\epsilon\|x - \bar{x}\| \leq f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x} \rangle \leq \epsilon\|x - \bar{x}\| \text{ whenever } \|x - \bar{x}\| < \delta.
\]

Define

\[
\psi(x) = f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \epsilon\|x - \bar{x}\|.
\]

Then \( \psi(x) \geq \psi(\bar{x}) = 0 \) for all \( x \in B(\bar{x}; \delta) \). Since \( \varphi \) is a convex function, \( \psi(x) \geq \psi(\bar{x}) \) for all \( x \in \mathbb{R}^n \). Thus,

\[
\langle \nabla f(\bar{x}), x - \bar{x} \rangle \leq f(x) - f(\bar{x}) + \epsilon\|x - \bar{x}\| \text{ for all } x \in \mathbb{R}^n.
\]

Letting \( \epsilon \to 0 \), one obtains (3.1).

Equality (3.1) implies that \( \nabla f(\bar{x}) \in \partial f(\bar{x}) \). Take any \( v \in \partial f(\bar{x}) \), one has

\[
\langle v, x - \bar{x} \rangle \leq f(x) - f(\bar{x}) \text{ for all } x \in \mathbb{R}^n.
\]

The Fréchet differentiability of \( f \) also implies that for any \( \epsilon > 0 \), there exists \( \delta > 0 \) such that

\[
\langle v - \nabla f(\bar{x}), x - \bar{x} \rangle \leq \epsilon\|x - \bar{x}\| \text{ whenever } \|x - \bar{x}\| < \delta.
\]

It follows that \( \|v - \nabla f(\bar{x})\| \leq \epsilon \), which implies \( v = \nabla f(\bar{x}) \) since \( \epsilon > 0 \) is arbitrary. Therefore, \( \partial f(\bar{x}) = \{\nabla f(\bar{x})\} \).

\[\square\]
Remark 3.6 Let \( f : \mathbb{R}^n \to (-\infty, \infty] \) be a convex function. If \( f \) is differentiable at \( \bar{x} \), we can also see that (3.1) is satisfied from the following simple calculation:

\[
\langle \nabla f(\bar{x}), x - \bar{x} \rangle = \lim_{t \to 0^+} \frac{f(\bar{x} + t(x - \bar{x})) - f(\bar{x})}{t} = \lim_{t \to 0^+} \frac{f(tx + (1-t)\bar{x}) - f(\bar{x})}{t} \leq \lim_{t \to 0^+} \frac{tf(x) + (1-t)f(\bar{x}) - f(\bar{x})}{t} = f(x) - f(\bar{x}).
\]

The subdifferential formula for the norm function is given in the example below.

Example 3.7 Let \( p(x) = \|x\| \), the Euclidean norm function on \( \mathbb{R}^n \). Then

\[
\partial p(x) = \begin{cases} 
\mathbb{B} & \text{if } x = 0, \\
\{ x \|x\} & \text{otherwise.}
\end{cases}
\]

Since the function \( p \) is differentiable with \( \nabla p(x) = \frac{x}{\|x\|} \) for \( x \neq 0 \), it suffices to prove the formula for \( x = 0 \). By definition, an element \( v \in \mathbb{R}^n \) is a subgradient of \( p \) at \( 0 \) if and only if

\[
\langle v, x \rangle = \langle v, x - 0 \rangle \leq p(x) - p(0) = \|x\| \text{ for all } x \in \mathbb{R}^n.
\]

For \( x := v \in \mathbb{R}^n \), one has \( \langle v, v \rangle \leq \|v\| \). This implies \( \|v\| \leq 1 \) or \( v \in \mathbb{B} \). Moreover, if \( v \in \mathbb{B} \), by the Cauchy-Schwarz inequality,

\[
\langle v, x - 0 \rangle = \langle v, x \rangle \leq \|v\|\|x\| \leq \|x\| = p(x) - p(0) \text{ for all } x \in \mathbb{R}^n.
\]

It follows that \( v \in \partial p(0) \). Therefore, \( \partial p(0) = \mathbb{B} \).

The proposition below allows us to represent the relative interior of the graph of a convex function in terms of the relative interior of its domain; see, e.g., [7, Proposition 1.1.9].

Proposition 3.8 Let \( f : \mathbb{R}^n \to (-\infty, \infty] \) be a convex function. Then

\[
\text{ri (epi } f) = \{ (x, \lambda) \mid x \in \text{ri (dom } f), \lambda > f(x) \}.
\]

4 Convex Separation

In this section, we introduce convex separation theorems, which will be important in developing generalized differentiation calculus rules. Although all theorems are presented in the setting of \( \mathbb{R}^n \), they possess far generalizations to infinite dimensional settings. We begin with a result on strict separation of a convex set and a point outside the set.

Proposition 4.1 Let \( \Omega \) be a nonempty closed convex set and let \( \bar{x} \notin \Omega \). Then there exists a vector \( v \neq 0 \) such that

\[
\sup \{ \langle v, x \rangle \mid x \in \Omega \} < \langle v, \bar{x} \rangle.
\]
Proof. Denote $\bar{\omega} := \pi(x; \Omega)$, let $v := \bar{x} - \bar{\omega}$ and fix any $x \in \Omega$. By Proposition 2.9,
$$\langle v, x - \bar{\omega} \rangle = \langle \bar{x} - \bar{\omega}, x - \bar{\omega} \rangle \leq 0.$$ It follows that 
$$\langle v, x - \bar{\omega} \rangle = \langle v, x - \bar{x} + \bar{x} - \bar{\omega} \rangle = \langle v, x - \bar{x} + v \rangle \leq 0.$$ The last inequality implies 
$$\langle v, x \rangle \leq \langle v, \bar{x} \rangle - \|v\|^2.$$ Therefore, 
$$\sup_{x \in \Omega} \{\langle v, x \rangle \} = \langle v, \bar{x} \rangle,$$ which completes the proof. □

Remark 4.2 It is not hard to show that the closure $\overline{\Omega}$ of a convex set $\Omega$ is convex. If $\Omega$ is a nonempty convex set in $\mathbb{R}^n$ and $\bar{x} \not\in \overline{\Omega}$, applying Proposition 4.1 for the convex set $\overline{\Omega}$ gives nonzero element $v \in \mathbb{R}^n$ such that 
$$\sup_{x \in \Omega} \{\langle v, x \rangle \} \leq \langle v, \bar{x} \rangle.$$ This property can be proved in a more general setting for convex sets in a subspace of $\mathbb{R}^n$ instead of in $\mathbb{R}^n$.

Proposition 4.3 Let $L$ be a subspace of $\mathbb{R}^n$ and let $\Omega \subset L$ be a nonempty convex set with $\bar{x} \in L$ and $\bar{x} \not\in \overline{\Omega}$. Then there exists $v \in L$, $v \neq 0$, such that 
$$\sup_{x \in \Omega} \{\langle v, x \rangle \} < \langle v, \bar{x} \rangle.$$ Proof. By Proposition 4.1, there exists $w \in \mathbb{R}^n$ such that 
$$\sup_{x \in \Omega} \{\langle w, x \rangle \} < \langle w, \bar{x} \rangle.$$ It is well-known that $\mathbb{R}^n$ can be represented as $\mathbb{R}^n = L \oplus L^\perp$, where 
$$L^\perp := \{u \in \mathbb{R}^n \mid \langle u, x \rangle = 0 \text{ for all } x \in L\}.$$ Thus, we have the representation $w = u + v$, where $u \in L^\perp$ and $v \in L$. For any $x \in \Omega \subset L$, one has $\langle u, x \rangle = 0$ and 
$$\langle v, x \rangle = \langle u, x \rangle + \langle v, x \rangle = \langle u + v, x \rangle \leq \sup_{x \in \Omega} \{\langle w, x \rangle \} < \langle w, \bar{x} \rangle = \langle u + v, \bar{x} \rangle = \langle u, \bar{x} \rangle + \langle v, \bar{x} \rangle = \langle v, \bar{x} \rangle.$$ It follows that $\sup_{x \in \Omega} \{\langle v, x \rangle \} < \langle v, \bar{x} \rangle$, which also implies $v \neq 0$. The proof is now complete. □

Lemma 4.4 Let $\Omega$ be a nonempty convex subset of $\mathbb{R}^n$. Suppose that $0 \in \overline{\Omega} \setminus \text{ri} \Omega$. Then there exists a sequence $\{x_k\} \subset \mathbb{R}^n$ such that $x_k \not\in \overline{\Omega}$ for all $k \in \mathbb{N}$ and $x_k \to 0$ as $k \to \infty$. 

9
Proof. Suppose $0 \in \overline{\Omega} \setminus \text{ri}\Omega$. By Theorem 2.6(i), the relative interior of $\Omega$ is nonempty, so there exists $x_0 \in \text{ri}\Omega$. Then $-tx_0 \notin \overline{\Omega}$ for all $t > 0$. Indeed, by contradiction suppose that $-tx_0 \in \overline{\Omega}$ for some $t > 0$. It follows from Theorem 2.6(ii) that

$$0 = \frac{t}{1+t}x_0 + \frac{1}{1+t}(-tx_0) \in \text{ri}\Omega.$$  

This is a contradiction because $0 \notin \text{ri}\Omega$ by the assumption. Thus $-tx_0 \notin \Omega$ for all $t > 0$.

Let $x_k := -\frac{x_0}{k}$. Then $x_k \notin \Omega$ for every $k$ and $x_k \to 0$ as $k \to \infty$. □

We continue with another important theorem on proper separation.

Definition 4.5 Two nonempty convex sets $\Omega_1$ and $\Omega_2$ are said to be properly separated if there exists a nonzero vector $v \in \mathbb{R}^n$ such that

$$\sup\{\langle v, x \rangle \mid x \in \Omega_1\} \leq \inf\{\langle v, y \rangle \mid y \in \Omega_2\}$$

and

$$\inf\{\langle v, x \rangle \mid x \in \Omega_1\} < \sup\{\langle v, y \rangle \mid y \in \Omega_2\}.$$  

Proposition 4.6 Let $\Omega$ be a nonempty convex set in $\mathbb{R}^n$. Then $0 \notin \text{ri}\Omega$ if and only if the sets $\Omega$ and $\{0\}$ can be properly separated, i.e., there exists $v \in \mathbb{R}^n$, $v \neq 0$, such that

$$\sup\{\langle v, x \rangle \mid x \in \Omega\} \leq 0$$

and

$$\inf\{\langle v, x \rangle \mid x \in \Omega\} < 0.$$  

Proof. We consider two cases.

Case 1: $0 \notin \overline{\Omega}$. By Remark 4.2 with $\bar{x} = 0$, there exists $v \neq 0$ such that

$$\sup\{\langle v, x \rangle \mid x \in \Omega\} < \langle v, \bar{x} \rangle = 0.$$  

It follows that $\Omega$ and $\{0\}$ are properly separated.

Case 2: $0 \in \overline{\Omega} \setminus \text{ri}\Omega$. By Lemma 4.4, there exists a sequence $\{x_k\} \subset \mathbb{R}^n$ with $x_k \notin \overline{\Omega}$ for every $k$ and $x_k \to 0$ as $k \to \infty$. Let $L := \text{aff}\Omega$. Since every subspace of $\mathbb{R}^n$ is closed, $L$ is a closed affine set. Observe also that $L$ is a subspace of $\mathbb{R}^n$ since $0 \in \overline{\Omega} \subset \overline{\mathbb{R}^n} = \mathbb{R}^n$. By Proposition 4.3, there exists a sequence $\{v_k\} \subset L$ with $v_k \neq 0$ for all $k$ and

$$\sup\{\langle v_k, x \rangle \mid x \in \Omega\} < 0.$$  

Then $\langle v_k, x \rangle < 0$ for all $x \in \Omega$. Let $w_k := \frac{v_k}{\|v_k\|}$ and observe that $\|w_k\| = 1$ for all $k \in \mathbb{N}$. We can assume without loss of generality that $w_k \to v \in L$ with $\|v\| = 1$ as $k \to \infty$. Then

$$\sup\{\langle v, x \rangle \mid x \in \Omega\} \leq 0.$$  

To show that the condition

$$\inf\{\langle v, x \rangle \mid x \in \Omega\} < 0$$

would contradict the assumption that $0 \notin \text{ri}\Omega$. □
is satisfied, it suffices to show that there exists \( x \in \Omega \) with \( \langle v, x \rangle < 0 \). Suppose by contradiction that \( \langle v, x \rangle \geq 0 \) for all \( x \in \Omega \). Since \( \sup \{ \langle v, x \rangle \mid x \in \Omega \} \leq 0 \), it follows that \( \langle v, x \rangle = 0 \) for all \( x \in \Omega \). Since \( v \in L = \text{aff} \Omega \), we can write \( v = \sum_{i=1}^{m} \lambda_i \omega_i \), where \( \sum_{i=1}^{m} \lambda_i = 1 \) and \( \omega_i \in \Omega \) for each \( i = 1, \ldots, m \). Then
\[
\| v \|^2 = \langle v, v \rangle = \sum_{i=1}^{m} \lambda_i \langle v, \omega_i \rangle = 0.
\]
This is a contradiction because \( \| v \| = 1 \). Thus, \( \Omega \) and \( \{ 0 \} \) are properly separated.

Now suppose that \( \Omega \) and \( \{ 0 \} \) are properly separated. Then there exists \( 0 \neq v \in \mathbb{R}^n \) such that
\[
\sup \{ \langle v, x \rangle \mid x \in \Omega \} \leq 0,
\]
and there exists \( \bar{x} \in \Omega \) with \( \langle v, \bar{x} \rangle < 0 \). Suppose by contradiction that \( 0 \in \text{ri} \Omega \). By Proposition 2.5,
\[
0 + t(0 - \bar{x}) = -t\bar{x} \in \Omega \text{ for some } t > 0.
\]
This implies
\[
\langle v, -t\bar{x} \rangle \leq \sup \{ \langle v, x \rangle \mid x \in \Omega \} \leq 0.
\]
Then \( \langle v, \bar{x} \rangle \geq 0 \), which is a contradiction. Therefore, \( 0 \notin \text{ri} \Omega \). \( \square \)

**Theorem 4.7** Let \( \Omega_1 \) and \( \Omega_2 \) be two nonempty convex subsets of \( \mathbb{R}^n \). Then \( \Omega_1 \) and \( \Omega_2 \) are properly separated if and only if \( \text{ri} (\Omega_1) \cap \text{ri} (\Omega_2) = \emptyset \).

**Proof.** Define \( \Omega := \Omega_1 - \Omega_2 \) and note that \( \text{ri} \Omega_1 \cap \text{ri} \Omega_2 = \emptyset \) if and only if
\[
0 \notin \text{ri} (\Omega_1 - \Omega_2) = \text{ri} \Omega_1 - \text{ri} \Omega_2.
\]
First, suppose that \( \text{ri} \Omega_1 \cap \text{ri} \Omega_2 = \emptyset \). Then \( 0 \notin \text{ri} (\Omega_1 - \Omega_2) = \text{ri} \Omega \). By Proposition 4.6, the sets \( \Omega \) and \( \{ 0 \} \) are properly separated, so there exists \( v \in \mathbb{R}^n \) such that \( \langle v, x \rangle \leq 0 \) for all \( x \in \Omega \), and there exists \( y \in \Omega \) such that \( \langle v, y \rangle < 0 \). For any \( \omega_1 \in \Omega_1 \) and \( \omega_2 \in \Omega_2 \), one has \( x := \omega_1 - \omega_2 \in \Omega \), and hence
\[
\langle v, \omega_1 - \omega_2 \rangle = \langle v, x \rangle \leq 0.
\]
This implies \( \langle v, \omega_1 \rangle \leq \langle v, \omega_2 \rangle \). Choose \( \bar{\omega}_1 \in \Omega_1 \) and \( \bar{\omega}_2 \in \Omega_2 \) such that \( y = \omega_1 - \omega_2 \). Then
\[
\langle v, \bar{\omega}_1 - \bar{\omega}_2 \rangle = \langle v, y \rangle < 0,
\]
which implies \( \langle v, \bar{\omega}_1 \rangle < \langle v, \bar{\omega}_2 \rangle \). Therefore, \( \Omega_1 \) and \( \Omega_2 \) are properly separated.

Next, suppose that \( \Omega_1 \) and \( \Omega_2 \) are properly separated. We can easily see that \( \Omega = \Omega_1 - \Omega_2 \) and \( \{ 0 \} \) are properly separated. Applying Proposition 4.6 again yields
\[
0 \notin \text{ri} \Omega = \text{ri} (\Omega_1 - \Omega_2) = \text{ri} \Omega_1 - \text{ri} \Omega_2.
\]
Therefore,
\[
\text{ri} \Omega_1 \cap \text{ri} \Omega_2 = \emptyset.
\]
The proof is now complete. \( \square \)
5 Normal Cone Intersection Rules

In this section, we develop convex subdifferential calculus for a particular class of convex functions called the indicator functions to convex sets.

Let $\Omega \subset \mathbb{R}^n$ and let $\bar{x} \in \Omega$. The normal cone to $\Omega$ at $\bar{x}$ is defined by

$$N(\bar{x}; \Omega) := \{v \in \mathbb{R}^n \mid \langle v, x - \bar{x} \rangle \leq 0 \text{ for all } x \in \Omega\}.$$ 

Although the normal cone to a convex set is defined geometrically, it can be represented as the subdifferential of a convex function. Define the indicator function associated with a set $\Omega$ by

$$\delta(x; \Omega) = \begin{cases} 0 & \text{if } x \in \Omega, \\ \infty & \text{otherwise}. \end{cases}$$

**Proposition 5.1** Let $\Omega$ be a nonempty convex set and let $\bar{x} \in \Omega$. Then

$$\partial \delta(\bar{x}; \Omega) = N(\bar{x}; \Omega).$$

**Proof.** It follows from the definition that $v \in \partial \delta(\bar{x}; \Omega)$ if and only if

$$\langle v, x - \bar{x} \rangle \leq \delta(x; \Omega) - \delta(\bar{x}; \Omega) = \delta(x; \Omega) \text{ for all } x \in \mathbb{R}^n.$$ 

Since $\delta(x; \Omega) = 0$ whenever $x \in \Omega$, one has

$$\langle v, x - \bar{x} \rangle \leq 0 \text{ for all } x \in \Omega,$$

which implies $v \in N(\bar{x}; \Omega)$. We have proved that the inclusion $\partial \delta(\bar{x}; \Omega) \subset N(\bar{x}; \Omega)$. The proof of the opposite inclusion is straightforward. 

In the theorem below, we show that the normal cone to the intersection of two convex sets in $\mathbb{R}^n$ at a point can be represented as the sum of the normal cones to the sets at the point if they intersect significantly.

**Theorem 5.2** Let $\Omega_1$ and $\Omega_2$ be nonempty convex sets in $\mathbb{R}^n$ such that $\text{ri } \Omega_1 \cap \text{ri } \Omega_2 \neq \emptyset$. Then

$$N(\bar{x}, \Omega_1 \cap \Omega_2) = N(\bar{x}, \Omega_1) + N(\bar{x}, \Omega_2) \text{ for all } \bar{x} \in \Omega_1 \cap \Omega_2.$$ 

(5.1)

**Proof.** Fix $v \in N(\bar{x}; \Omega_1 \cap \Omega_2)$ and get by the definition that

$$\langle v, x - \bar{x} \rangle \leq 0 \text{ for all } x \in \Omega_1 \cap \Omega_2.$$ 

Denote $\Theta_1 := \Omega_1 \times [0, \infty)$ and $\Theta_2 := \{(x, \lambda) \mid x \in \Omega_2, \lambda \leq \langle v, x - \bar{x} \rangle\}$. It follows from Proposition 3.8 that $\text{ri } \Theta_1 = \text{ri } \Omega_1 \times (0, \infty)$ and

$$\text{ri } \Theta_2 = \{(x, \lambda) \mid x \in \text{ri } \Omega_1, \lambda < \langle v, x - \bar{x} \rangle\}.$$ 

12
It is easy to check, arguing by contradiction, that \( \text{ri } \Theta_1 \cap \text{ri } \Theta_2 = \emptyset \). Applying Theorem 4.7, we find \( 0 \neq (w, \gamma) \in \mathbb{R}^n \times \mathbb{R} \) such that
\[
\langle w, x \rangle + \lambda_1 \gamma \leq \langle w, y \rangle + \lambda_2 \gamma \quad \text{for all } (x, \lambda_1) \in \Theta_1, \ (y, \lambda_2) \in \Theta_2. \tag{5.2}
\]
Moreover, there are \( (\bar{x}, \bar{\lambda}_1) \in \Theta_1 \) and \( (\bar{y}, \bar{\lambda}_2) \in \Theta_2 \) satisfying
\[
\langle w, \bar{x} \rangle + \bar{\lambda}_1 \gamma < \langle w, \bar{y} \rangle + \bar{\lambda}_2 \gamma.
\]
Observe that \( \gamma \leq 0 \) since otherwise we can get a contradiction by applying (5.2) with \( (\bar{x}, k) \in \Theta_1 \) for all \( k > 0 \) and \( (\bar{x}, 0) \in \Theta_2 \). We will verify that \( \gamma < 0 \). By contradiction, assume that \( \gamma = 0 \) and then get
\[
\langle w, x \rangle \leq \langle w, y \rangle \quad \text{for all } x \in \Omega_1, \ y \in \Omega_2,
\]
\[
\langle w, \bar{x} \rangle < \langle w, \bar{y} \rangle \quad \text{with } \bar{x} \in \Omega_1, \ \bar{y} \in \Omega_2.
\]
Thus, the sets \( \Omega_1 \) and \( \Omega_2 \) are properly separated, and hence it follows from Theorem 4.7 that \( \text{ri } \Omega_1 \cap \text{ri } \Omega_2 = \emptyset \). This is a contradiction, which shows that \( \gamma < 0 \).

Let \( \mu := -\gamma > 0 \) and deduce from (5.2), by using \((x, 0) \in \Theta_1 \) when \( x \in \Omega_1 \), and \((\bar{x}, 0) \in \Theta_2 \), that
\[
\langle w, x \rangle \leq \langle w, \bar{x} \rangle \quad \text{for all } x \in \Omega_1.
\]
This implies that \( w \in N(\bar{x}; \Omega_1) \), and hence \( \frac{w}{\mu} \in N(\bar{x}; \Omega_1) \). To proceed further, we get from (5.2), due to \((\bar{x}, 0) \in \Theta_1 \) and \((y, \alpha) \in \Theta_2 \) for all \( y \in \Omega_2 \) with \( \alpha := \langle v, y - \bar{x} \rangle \), that
\[
\langle w, \bar{x} \rangle \leq \langle w, y \rangle + \gamma \langle v, y - \bar{x} \rangle \quad \text{whenever } y \in \Omega_2.
\]
Dividing both sides by \( \gamma \) gives us
\[
\left\langle \frac{w}{\gamma}, \bar{x} \right\rangle \geq \left\langle \frac{w}{\gamma}, y \right\rangle + \langle v, y - \bar{x} \rangle \quad \text{whenever } y \in \Omega_2.
\]
This clearly implies the inequality
\[
\left\langle \frac{w}{\gamma} + v, y - \bar{x} \right\rangle \leq 0 \quad \text{for all } y \in \Omega_2,
\]
and thus \( \frac{w}{\gamma} + v = -\frac{w}{\mu} + v \in N(\bar{x}; \Omega_2) \). Letting finally \( v_1 := \frac{w}{\mu} \in N(\bar{x}; \Omega_1) \) and \( v_2 := -\frac{w}{\mu} + v \in N(\bar{x}; \Omega_2) \) gives us \( v = v_1 + v_2 \), which ensures the validity of (5.1). \( \square \)

**Example 5.3** Consider two convex sets
\[
\Omega_1 := \{(x, \lambda) \in \mathbb{R}^2 \mid \lambda \geq x^2 \} \quad \text{and } \Omega_2 := \{(x, \lambda) \mid \lambda \leq -x^2 \}.
\]
For \( \bar{x} = (0, 0) \in \mathbb{R}^n \),
\[
N(\bar{x}; \Omega_1) = \{0\} \times (-\infty, 0], \ N(\bar{x}; \Omega_2) = \{0\} \times [0, \infty), \text{ and } N(\bar{x}; \Omega_1 \cap \Omega_2) = \mathbb{R}^2.
\]
Thus, \( N(\bar{x}; \Omega_1) + N(\bar{x}; \Omega_2) = \{0\} \times \mathbb{R} \neq N(\bar{x}; \Omega_1 \cap \Omega_2) \). Note that \( \text{ri } \Omega_1 \cap \text{ri } \Omega_2 = \emptyset \).
It has been proved in [9, Corollary 2.16] that the normal cone intersection rule (5.1) is satisfied at $\bar{x}$ under the qualification condition

$$N(\bar{x}; \Omega_1) \cap [-N(\bar{x}; \Omega_2)] = \{0\}.$$  

The following example illustrate the relation between the qualification condition and the relative interior assumption in Theorem 5.2.

**Example 5.4** Consider two convex sets in $\mathbb{R}^2$ given by $\Omega_1 := \mathbb{R} \times \{0\}$ and $\Omega_2 := (-\infty, 0] \times \{0\}$. Obviously,

$$\text{ri } \Omega_1 \cap \text{ri } \Omega_2 \neq \emptyset.$$  

For $\bar{x} = (0, 0)$, this condition does not imply that $N(\bar{x}; \Omega_1) \cap [-N(\bar{x}; \Omega_2)] = \{0\}$. Indeed,

$$N(\bar{x}; \Omega_1) = \{0\} \times \mathbb{R} \text{ and } N(\bar{x}; \Omega_2) = [0, \infty) \times \mathbb{R},$$

so $N(\bar{x}; \Omega_1) \cap [-N(\bar{x}; \Omega_2)] = \{0\} \times \mathbb{R}$. Therefore, Theorem 5.1 is applicable in this case but [9, Corollary 2.16] is not applicable.

Now suppose that $N(\bar{x}; \Omega_1) \cap [-N(\bar{x}; \Omega_2)] = \{0\}$ for some $\bar{x} \in \Omega_1 \cap \Omega_2$. Then we can show that

$$\text{ri } \Omega_1 \cap \text{ri } \Omega_2 \neq \emptyset.$$  

Indeed, by contradiction, suppose that $\text{ri } \Omega_1 \cap \text{ri } \Omega_2 = \emptyset$. Then they are properly separated by Theorem 4.7. In particular, there exists $v \neq 0$ such that

$$\langle v, x \rangle \leq \langle v, y \rangle \text{ for all } x \in \Omega_1, y \in \Omega_2.$$  

Since $\bar{x} \in \Omega_2$, one has $\langle v, x \rangle \leq \langle v, \bar{x} \rangle$, or equivalently, $\langle v, x - \bar{x} \rangle \leq 0$ for all $x \in \Omega_1$, and hence $v \in N(\bar{x}; \Omega_1)$. Similarly, we can show that $-v \in N(\bar{x}; \Omega_2)$. Therefore, $0 \neq v \in N(\bar{x}; \Omega_1) \cap [-N(\bar{x}; \Omega_2)]$, a contradiction.

6 Subdifferential Sum Rules

In this section, we apply the normal cone intersection rule from Theorem 5.2 in order to develop the subdifferential sum rule. This approach is based on the following proposition which allows us to represent the subdifferential of a convex function in terms of the normal cone to the epigraph of the function.

**Proposition 6.1** Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a convex function and let $\bar{x} \in \text{dom } f$. Then

$$\partial f(\bar{x}) = \{v \in \mathbb{R}^n \mid (v, -1) \in N((\bar{x}, f(\bar{x})); \text{epi } f)\}.$$  

**Proof.** Fix any $v \in \partial f(\bar{x})$. Then

$$\langle v, x - \bar{x} \rangle \leq f(x) - f(\bar{x}) \text{ for all } x \in \mathbb{R}^n.$$  

(6.1)
Let us show that $(v, -1) \in N((\bar{x}, f(\bar{x})); epi f)$. Fix any $(x, \lambda) \in epi f$. Then $\lambda \geq f(x)$, and hence
\[
((v, -1), (x, \lambda) - (\bar{x}, f(\bar{x}))) = (v, x - \bar{x}) + (-1)(\lambda - f(\bar{x}))
\]
\[
= (v, x - \bar{x}) - (\lambda - f(\bar{x}))
\]
\[
\leq (v, x - \bar{x}) - (f(x) - f(\bar{x})) \leq 0.
\]
Note that the last inequality holds by (6.1).

Now suppose that $(v, -1) \in (v, -1) \in N((\bar{x}, f(\bar{x})); epi f)$. Fix any $x \in dom f$. Then $(x, f(x)) \in epi f$, and hence
\[
((v, -1), (x, f(x)) - (\bar{x}, f(\bar{x}))) \leq 0.
\]
This implies
\[
(v, x - \bar{x}) - (f(x) - f(\bar{x})) \leq 0.
\]
Consequently,
\[
(v, x - \bar{x}) \leq f(x) - f(\bar{x}).
\]
Note that this inequality holds obviously if $x \notin dom f$, i.e., $f(x) = \infty$. Therefore, $v \in \partial f(\bar{x})$.
\[\square\]

Now we are ready to establish the convex subdifferential sum rule. The theorem below tells us that the subdifferential sum rule (6.3) is satisfied the relative interiors of the domains of the function intersect each other significantly in the sense of (6.2).

**Theorem 6.2** Let $f_1, f_2: \mathbb{R}^n \to (-\infty, \infty]$ be convex functions satisfying the condition

\[
ri (dom f_1) \cap ri (dom f_2) \neq \emptyset. \quad (6.2)
\]

Then for all $\bar{x} \in dom f_1 \cap dom f_2$ we have

\[
\partial (f_1 + f_2)(\bar{x}) = \partial f_1(\bar{x}) + \partial f_2(\bar{x}). \quad (6.3)
\]

**Proof.** First, we show that $\partial f_1(\bar{x}) + \partial f_2(\bar{x}) \subset \partial (f_1 + f_2)(\bar{x})$. Fix any $v \in \partial f_1(\bar{x}) + \partial f_2(\bar{x})$. Then $v = v_1 + v_2$, where $v_1 \in \partial f_1(\bar{x})$ and $v_2 \in \partial f_2(\bar{x})$. It follows from the definition that $\langle v_1, x - \bar{x} \rangle \leq f_1(x) - f_1(\bar{x})$ and $\langle v_2, x - \bar{x} \rangle \leq f_2(x) - f_2(\bar{x})$ for all $x \in \mathbb{R}^n$. Then
\[
\langle v_1 + v_2, x - \bar{x} \rangle = \langle v_1, x - \bar{x} \rangle + \langle v_2, x - \bar{x} \rangle
\]
\[
\leq f_1(x) - f_1(\bar{x}) + f_2(x) - f_2(\bar{x})
\]
\[
\leq f_1(x) + f_2(x) - (f_1(\bar{x}) + f_2(\bar{x}))
\]
\[
\leq (f_1 + f_2)(x) - (f_1 + f_2)(\bar{x}).
\]
Thus, $v = v_1 + v_2 \in \partial (f_1 + f_2)(\bar{x})$.

Let us prove the opposite inclusion. Fix any $v \in \partial (f_1 + f_2)(\bar{x})$. By the definition, we have
\[
\langle v, x - \bar{x} \rangle \leq (f_1 + f_2)(x) - (f_1 + f_2)(\bar{x}). \quad (6.4)
\]
Define the following sets:
\[
\Omega_1 := \{(x, \lambda_1, \lambda_2) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} | \lambda_1 \geq f_1(x)\} = \text{epi } f_1 \times \mathbb{R}
\]
\[
\Omega_2 := \{(x, \lambda_1, \lambda_2) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} | \lambda_2 \geq f_2(x)\}.
\]
We can easily verify using the definition that \((v, -1, -1) \in N((\bar{x}, f_1(\bar{x}), f_2(\bar{x})), \Omega_1 \cap \Omega_2)\).

Let us show that \(\text{ri } \Omega_1 \cap \text{ri } \Omega_2 \neq \emptyset\). By Proposition 3.8,
\[
\text{ri } \Omega_1 = \{(x, \lambda_1, \lambda_2) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} | x \in \text{ri } (\text{dom } f_1), \lambda_1 > f_1(x)\} = \text{ri } (\text{epi } f_1) \times \mathbb{R}
\]
\[
\text{ri } \Omega_2 = \{(x, \lambda_1, \lambda_2) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} | x \in \text{ri } (\text{dom } f_2), \lambda_2 > f_2(x)\}.
\]
Choose \(z \in \text{ri } (\text{dom } f_1) \cap \text{ri } (\text{dom } f_2)\). It is not hard to see that
\[
(z, f_1(z) + 1, f_2(z) + 1) \in \text{ri } \Omega_1 \cap \text{ri } \Omega_2 \neq \emptyset.
\]
By Theorem 5.2, we have
\[
N((\bar{x}, f_1(\bar{x}), f_2(\bar{x})), \Omega_1 \cap \Omega_2) = N((\bar{x}, f_1(\bar{x}), f_2(\bar{x})), \Omega_1) + N((\bar{x}, f_1(\bar{x}), f_2(\bar{x})), \Omega_2).
\]

From the structures of the normal cones to the sets \(\Omega_1\) and \(\Omega_2\), one has
\[
(v, -1, -1) = (v_1, -\gamma_1, 0) + (v_2, 0, -\gamma_2),
\]
where \((v_1, -\gamma_1) \in N((\bar{x}, f_1(\bar{x})), \text{epi } f_1)\) and \((v_2, -\gamma_2) \in N((\bar{x}, f_2(\bar{x})), \text{epi } f_2)\). It follows that
\[
v = v_1 + v_2, \gamma_1 = \gamma_2 = 1.
\]
Then \(v_1 \in \partial f_1(\bar{x})\) and \(v_2 \in \partial f_2(\bar{x})\). Thus, \(v \in \partial f_1(\bar{x}) + \partial f_2(\bar{x})\). Therefore, \(\partial (f_1 + f_2)(\bar{x}) \subseteq \partial f_1(\bar{x}) + \partial f_2(\bar{x})\). The proof is now complete. \(\square\)

**Corollary 6.3** If \(\bar{x} \in \text{ri } (\text{dom } f)\), then \(\partial f(\bar{x}) \neq \emptyset\).

**Proof.** Define the function
\[
g(x) := f(x) + \delta_\{\bar{x}\}(x) = \begin{cases} f(\bar{x}) & \text{if } x = \bar{x}, \\ \infty & \text{otherwise} \end{cases}
\]
via the indicator function of \(\{\bar{x}\}\). Then epi \(g = \{\bar{x}\} \times [f(\bar{x}), \infty)\), and hence \(N((\bar{x}, g(\bar{x})); \text{epi } g) = \mathbb{R}^n \times (-\infty, 0]\). It follows that \(\partial g(\bar{x}) = \mathbb{R}^n\) and \(\partial \delta_\{\bar{x}\}(\bar{x}) = \mathbb{R}^n\). For the function \(h(x) = \delta_\{\bar{x}\}(x)\), by Proposition 3.8,
\[
\text{ri } (\text{dom } h) = \{\bar{x}\}.
\]
Thus, \(\text{ri } (\text{dom } f) \cap \text{ri } (\text{dom } h) \neq \emptyset\). Employing the subdifferential sum rule from Theorem 6.2 gives us
\[
\mathbb{R}^n = \partial g(\bar{x}) = \partial f(\bar{x}) + \mathbb{R}^n,
\]
which ensures that \(\partial f(\bar{x}) \neq \emptyset\). \(\square\)
7 Subdifferential Chain Rules

As we see from the previous section, the normal cone intersection rule in Theorem 5.2 plays an important role in proving the subdifferential sum rule. In this section, we use Theorem 5.2 again to establish the subdifferential chain rule. To proceed with deriving chain rules, we present first the following geometric lemma.

**Lemma 7.1** Let $B : \mathbb{R}^n \to \mathbb{R}^p$ be an affine mapping given by $B(x) = Ax + b$, where $A$ is an $n \times p$ matrix and $b \in \mathbb{R}^p$. For any $(\bar{x}, \bar{y}) \in \text{gph} B$, we have

$$N((\bar{x}, \bar{y}); \text{gph} B) = \{(u, v) \in \mathbb{R}^n \times \mathbb{R}^p \mid u = -A^\top v\}.$$  

**Proof.** It is clear that the set $\text{gph} B$ is convex and $(u, v) \in N((\bar{x}, \bar{y}) \text{gph} B)$ if and only if

$$\langle u, x - \bar{x} \rangle + \langle v, B(x) - B(\bar{x}) \rangle \leq 0 \text{ whenever } x \in \mathbb{R}^p. \quad (7.1)$$

It follows directly from the definitions that

$$\langle u, x - \bar{x} \rangle + \langle v, B(x) - B(\bar{x}) \rangle = \langle u, x - \bar{x} \rangle + \langle v, A(x) - A(\bar{x}) \rangle = \langle u, x - \bar{x} \rangle + \langle A^\top v, x - \bar{x} \rangle = \langle u + A^\top v, x - \bar{x} \rangle,$$

which implies the equivalence of (7.1) to $\langle u + A^\top v, x - \bar{x} \rangle \leq 0$, and so to $u = -A^\top v$. \hfill \Box

**Theorem 7.2** Let $f : \mathbb{R}^p \to \overline{\mathbb{R}}$ be a convex function and let $B : \mathbb{R}^n \to \mathbb{R}^p$ be as in Lemma 7.1 with $B(\bar{x}) \in \text{dom } f$ for some $\bar{x} \in \mathbb{R}^p$. Denote $\bar{y} := B(\bar{x})$ and assume that the range of $B$ contains a point of $\text{ri}(\text{dom } f)$. Then we have the subdifferential chain rule

$$\partial(f \circ B)(\bar{x}) = A^\top \partial f(\bar{y}) = \{A^\top v \mid v \in \partial f(\bar{y})\}. \quad (7.2)$$

**Proof.** Fix $v \in \partial(f \circ B)(\bar{x})$ and form the subsets of $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}$ by

$$\Omega_1 := \text{gph} B \times \mathbb{R} \text{ and } \Omega_2 := \mathbb{R}^n \times \text{epi } f.$$  

Then

$$\text{ri } \Omega_1 = \Omega_1 = \text{gph} B \times \mathbb{R}, \quad \text{ri } \Omega_2 = \{(x, y, \lambda) \mid x \in \mathbb{R}^n, y \in \text{ri}(\text{dom } f), \lambda > f(y)\}.$$  

The assumption made guarantees that $\text{ri } \Omega_1 \cap \text{ri } \Omega_2 \neq \emptyset$.

It follows from the definition of the subdifferential and the normal cone that $(v, 0, -1) \in N((\bar{x}, \bar{y}, \bar{z}); \Omega_1 \cap \Omega_2)$, where $\bar{z} := f(\bar{y})$. Indeed, fix any $(x, y, \lambda) \in \Omega_1 \cap \Omega_2$. Then $y = B(x)$ and $\lambda \geq f(y)$, and so $\lambda \geq f(B(x))$. Thus

$$\langle v, x - \bar{x} \rangle + 0(y - \bar{y}) + (-1)(\lambda - \bar{z}) \leq \langle v, x - \bar{x} \rangle - [f(B(x)) - f(B(\bar{x}))] \leq 0.$$  

17
Employing the intersection rule gives us
\[(v, 0, -1) \in N((\bar{x}, \bar{y}, \bar{z}); \Omega_1) + N((\bar{x}, \bar{y}, \bar{z}); \Omega_2),\]
which reads that \((v, 0, -1) = (v, -w, 0) + (0, w, -1)\) with \((v, -w) \in N((\bar{x}, \bar{y}); \text{gph} B)\) and \((w, -1) \in N((\bar{y}, \bar{z}); \text{epi} f)\). Then we get
\[v = A^\top w \text{ and } w \in \partial f(\bar{y}),\]
which implies in turn that \(v \in A^\top(\partial f(\bar{y}))\) and thus verifies the inclusion \(\subset\) in (7.2). The opposite inclusion follows directly from the definition of the subdifferential. \(\Box\)

8 \ Subdifferentials of Optimal Value Function

In this section, we study subdifferential rule for a class of convex function called the optimal value/marginal function defined by a convex objective function and a set-valued mapping. We say that \(F : \mathbb{R}^n \rightharpoonup \mathbb{R}^p\) is a set-valued mapping if \(F(x)\) is a subset of \(\mathbb{R}^p\) for every \(x \in \mathbb{R}^n\). The domain and the graph of \(F\) are defined by
\[
\text{dom} F := \{x \in \mathbb{R}^n \mid F(x) \neq \emptyset\},
\]
\[
\text{gph} F := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^p \mid y \in F(x)\}.
\]

Let \(F : \mathbb{R}^n \rightharpoonup \mathbb{R}^p\) be a set-valued mapping with convex graph. Given \((\bar{x}, \bar{y}) \in \text{gph} F\), the coderivative of \(F\) at \((\bar{x}, \bar{y})\) is a set-valued mapping from \(\mathbb{R}^p\) to \(\mathbb{R}^n\) defined by
\[
D^* F(\bar{x}, \bar{y})(v) := \{u \in \mathbb{R}^n \mid (u, -v) \in N((\bar{x}, \bar{y}); \text{gph} F)\}.
\]

As an example, consider the affine mapping \(B\) given in Lemma 7.1. Then \(B\) is a set-valued mapping with convex graph. The normal cone formula given in Lemma 7.1 yields
\[D^* B(\bar{x}, \bar{y})(v) = A^\top v, \text{where } \bar{y} = B(\bar{x}).\]

Now we derive a formula for calculating the subdifferential of the optimal value/marginal function given in the form
\[
\mu(x) := \inf \{\varphi(x, y) \mid y \in F(x)\} \tag{8.1}
\]
via the coderivative of the set-valued mapping \(F\) and the subdifferential of the function \(\varphi\). First we derive the following useful estimate.

**Lemma 8.1** Let \(\mu(\cdot)\) be the optimal value function \((8.1)\) generated by a convex-graph mapping \(F : \mathbb{R}^n \rightharpoonup \mathbb{R}^p\) and a convex function \(\varphi : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}\). Suppose that \(\mu(x) > -\infty\) for all \(x \in \mathbb{R}^n\), fix some \(\bar{x} \in \text{dom} \mu\), and consider the solution set
\[
S(\bar{x}) := \{\bar{y} \in F(\bar{x}) \mid \mu(\bar{x}) = \varphi(\bar{x}, \bar{y})\}.
\]
If \(S(\bar{x}) \neq \emptyset\), then for any \(\bar{y} \in S(\bar{x})\) we have
\[
\bigcup_{(u, v) \in \partial \varphi(\bar{x}, \bar{y})} [u + D^* F(\bar{x}, \bar{y})(v)] \subset \partial \mu(\bar{x}). \tag{8.2}
\]
Proof. It is not hard to show that $\mu : \mathbb{R}^n \to (-\infty, \infty]$ is a convex function under the assumptions made. Pick $w$ from the left-hand side of (8.2) and find $(u, v) \in \partial \varphi(\bar{x}, \bar{y})$ with

$$w - u \in D^* F(\bar{x}, \bar{y})(v).$$

It gives us $(w - u, -v) \in N((\bar{x}, \bar{y}); gph F)$ and thus

$$\langle w - u, x - \bar{x} \rangle - \langle v, y - \bar{y} \rangle \leq 0 \text{ for all } (x, y) \in gph F,$$

which shows that whenever $y \in F(x)$ we have

$$\langle w, x - \bar{x} \rangle \leq \langle u, x - \bar{x} \rangle + \langle v, y - \bar{y} \rangle \leq \varphi(x, y) - \varphi(\bar{x}, \bar{y}) = \varphi(x, y) - \mu(\bar{x}).$$

This implies in turn the estimate

$$\langle w, x - \bar{x} \rangle \leq \inf_{y \in F(x)} \varphi(x, y) - \mu(\bar{x}) = \mu(x) - \mu(\bar{x}),$$

which justifies the inclusion $w \in \partial \mu(\bar{x})$, and hence completes the proof of (8.2). \qed

Theorem 8.2 Consider the optimal value function (8.1) under the assumptions of Lemma 8.1. For any $\bar{y} \in S(\bar{x})$, we have the equality

$$\partial \mu(\bar{x}) = \bigcup_{(u, v) \in \partial \varphi(\bar{x}, \bar{y})} \left[ u + D^* F(\bar{x}, \bar{y})(v) \right]$$

provided the validity of the qualification condition

$$\text{ri}(\text{dom } \varphi) \cap \text{ri}(\text{gph } F) \neq \emptyset.$$

Proof: Taking Lemma 8.1 into account, it is only required to verify the inclusion “$\subset$” in (8.3). To proceed, pick $w \in \partial \mu(\bar{x})$ and $\bar{y} \in S(\bar{x})$. For any $x \in \mathbb{R}^n$, we have

$$\langle w, x - \bar{x} \rangle \leq \mu(x) - \mu(\bar{x}) = \mu(x) - \varphi(\bar{x}, \bar{y})$$

$$\leq \varphi(x, y) - \varphi(\bar{x}, \bar{y}) \text{ for all } y \in F(x).$$

This implies that, whenever $(x, y) \in \mathbb{R}^n \times \mathbb{R}^p$, the following inequality holds:

$$\langle w, x - \bar{x} \rangle + \langle 0, y - \bar{y} \rangle \leq \varphi(x, y) + \delta((x, y); gph F) - \left[ \varphi(\bar{x}, \bar{y}) + \delta((\bar{x}, \bar{y}); gph F) \right].$$

Denote further $f(x, y) := \varphi(x, y) + \delta((x, y); gph F)$ and deduce from the subdifferential sum rule in Theorem 6.2 under the qualification condition (8.4) that the inclusion

$$(w, 0) \in \partial f(\bar{x}, \bar{y}) = \partial \varphi(\bar{x}, \bar{y}) + N((\bar{x}, \bar{y}); gph F)$$

is satisfied. This shows that $(w, 0) = (u_1, v_1) + (u_2, v_2)$ with $(u_1, v_1) \in \partial \varphi(\bar{x}, \bar{y})$ and $(u_2, v_2) \in N((\bar{x}, \bar{y}); gph F)$. This yields $v_2 = -v_1$, so $(u_2, -v_1) \in N((\bar{x}, \bar{y}); gph F)$. Finally, we get $u_2 \in D^* F(\bar{x}, \bar{y})(v_1)$ and therefore

$$w = u_1 + u_2 \in u_1 + D^* F(\bar{x}, \bar{y})(v_1),$$

which completes the proof of the theorem.
Theorem 8.3 (revisit). Let $f: \mathbb{R}^p \to \mathbb{R}$ be a convex function and let $B: \mathbb{R}^n \to \mathbb{R}^p$ be as in Lemma 7.1 with $B(\bar{x}) \in \text{dom } f$ for some $\bar{x} \in \mathbb{R}^p$. Denote $\bar{y} := B(\bar{x})$ and assume that the range of $B$ contains a point of $\text{ri(dom } f)$. Then we have the subdifferential chain rule

$$\partial(f \circ B)(\bar{x}) = A^\top (\partial f(\bar{y})) = \{ A^\top v \mid v \in \partial f(\bar{y}) \}. \quad (8.5)$$

Proof. Consider $F(x) = \{ B(x) \}$ and $\varphi(x, y) = f(y)$. Then $\mu(x) = (f \circ B)(x)$. Moreover,

$$\text{ri(dom } \varphi) = \mathbb{R}^n \times \text{ri(dom } f), \text{ri(gph } F) = \text{gph } B.$$ 

The assumption made guarantees that the qualification condition 8.4 is satisfied. Thus,

$$\partial \mu(\bar{x}) = \bigcup_{v \in \partial \varphi(\bar{y})} \left[ D^* F(\bar{x}, \bar{y})(v) \right] = A^\top (\partial f(\bar{y})).$$

The proof is now complete. \qed

References

[1] H. H. Bauschke and P. L. Combettes: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, Springer, 2011.
[2] D. P. Bertsekas, A. Nedić and A. E. Ozdaglar: Convex Analysis and Optimization, Athena Scientific, 2003.
[3] S. Boyd and L. Vandenberghe: Convex Optimization, Cambridge University Press, 2004.
[4] J. M. Borwein and A. S. Lewis: Convex Analysis and Nonlinear Optimization, Springer, 2000.
[5] J.M. Borwein, Q.J. Zhu: Techniques of Variational Analysis, Springer, New York (2005).
[6] F.H. Clarke: Optimization and Nonsmooth Analysis, Wiley, New York (1983).
[7] J.-B. Hiriart-Urruty, C. Lemaréchal: Convex Analysis and Minimization Algorithms I. Fundamentals, Springer-Verlag, Berlin (1993).
[8] B.S. Mordukhovich: Variational Analysis and Generalized Differentiation, I: Basic Theory, Springer, Berlin (2006).
[9] B. S. Mordukhovich and N. M. Nam: An Easy Path to Convex Analysis and Applications, Morgan & Claypool Publishers, 2014.
[10] R. T. Rockafellar: Convex Analysis, Princeton University Press, Princeton, NJ, 1970.
[11] R.T. Rockafellar, R-J. Wets: Variational Analysis, Springer, Berlin (1998).
[12] W. Schirotzek: Nonsmooth Analysis, Springer, Berlin (2007).
[13] C. Zălinescu: Convex analysis in general vector spaces, World Scientific, Singapore (2002).