Subdifferentials and Their Applications

by

Zili Wu

B.Sc., Xiamen University, 1982.

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Mathematics and Statistics.

We accept this thesis as conforming
[redacted]he required standard.

Dr. Jane J. Ye, Supervisor (Department of Mathematics and Statistics)

[redacted]

Dr. Chris J. Bose, Departmental Member (Department of Mathematics and Statistics)

[redacted]

Dr. Ian F. Putnam, Departmental Member (Department of Mathematics and Statistics)

[redacted]

Dr. Qiji J. Zhu, External Examiner (Department of Mathematics and Statistics, Western Michigan University)

University of Victoria.

Supervisor: Dr. Jane J. Ye

ABSTRACT

To deal with nonsmooth phenomena in mathematics and optimization, various kinds of subdifferentials have been introduced in the literature over the last two decades. Among them are the Clarke generalized gradient, the limiting subgradient, and the proximal subgradient. In this thesis, we provide some conditions that guarantee the nonemptiness of the proximal subgradient, and the sum rule for proximal subgradients to hold. We have also found a class of functions f that can be recovered up to a constant by the proximal subgradient of f or $-f$. Finally we provide some sufficient conditions for the existence of error bounds in terms of the above three subdifferentials.

Examiners:

[redacted]

Dr. Jane J. Ye, Supervisor (Department of Mathematics and Statistics)

[redacted]

Dr. Chris J. Bose, Departmental Member (Department of Mathematics and Statistics)

[redacted]

Dr. Ian F. Putnam, Departmental Member (Department of Mathematics and Statistics)

[redacted] 6/27/97

Dr. Qiji J. Zhu, External Examiner (Department of Mathematics and Statistics, Western Michigan University)

Contents

Acknowledgements

With most sincere appreciation, I thank Dr. Jane J. Ye, my supervisor, for her enthusiastic guidance, encouragement and support in my research work. This thesis could not have been completed without her help, suggestions, advice and patience.

Thanks and appreciation are due to Dr. Chris J. Bose, Dr. Ian F. Putnam and Dr. Qiji J. Zhu for their time in examining my thesis and providing me with enlightening suggestions and remarks.

I gratefully acknowledge the Department of Mathematics and Statistics, University of Victoria, which provided me with the opportunity and support to work on this thesis.

I also thank Dr. Xingfu Zou and Mr. Kelly Choo for their help in my typing of this thesis.

I am grateful to my friends and their family members who have warmly encouraged and helped me since I came to Canada.

Finally, I am greatly indebted to my father, my wife, and my brother and sisters for their love, support and encouragement. I would like to extend heartful thanks to Mr. Zhanghong Wu for his impressive instruction and encouragement.

Chapter 1

Introduction

Subdifferentials are important and useful tools in nonsmooth analysis and optimization (e.g., see [4, 5, 6, 7, 10, 14, 15, 19, 24, 30, 31, 35, 37, 38]). These notions include the Clarke generalized gradient, the limiting subgradient and the proximal subgradient. All these subdifferentials have their own merits and may have applications in different situations. On one hand, the Clarke generalized gradient, being the biggest among the three subdifferentials, may be too big in some applications but usually enjoys good calculus. On the other hand, the proximal subgradient, being the smallest, may be empty and tends to have poor calculus. The main purpose of this thesis is to study the calculus and integration of the proximal subgradient, and the characterization of error estimates for equality and inequality systems in terms of the above three subdifferentials.

In Chapter 2, we study the conditions under which the proximal subgradient is nonempty and the sum rule holds.

In Chapter 3, we discuss the integration problem, i.e., when a function can be recovered up to a constant from its subdifferentials. In R^n, there are some classes of functions which can be recovered by their Clarke generalized gradients.

Rockafellar [37] proved that if the functions f and g are locally Lipschitz and Clarke regular with $\partial f = \partial g$, where ∂f denotes the Clarke generalized gradient of f, then they differ by a constant. Qi [32] showed that a function whose effective domain has a nonempty interior and whose subdifferential is single-valued almost everywhere can be decided by its generalized gradient. Poliquin [31] asserted that a primal lower-nice function f is also recoverable from its generalized gradient (actually in this case its proximal subgradient $\partial^{\pi} f$ coincides with ∂f). However, not every Lipschitz function can be recovered from the knowledge of its Clarke subdifferential. In fact, Rockafellar [37] has constructed a Lipschitz function $f : R^n \to R$ such that $\partial f(x) = [-1,1]^n$ for any $x \in R^n$ but its translate $g(x) = f(x-a)$ does not differ from f by a constant even though $\partial g(x) = \partial f(x)$ for any $x \in R^n$ and any a. This example also shows how "far" the generalized gradient can be from the gradient according to Rademacher's theorem which states that a locally Lipschitz function is differentiable almost everywhere.

For a Lipschitz function, the proximal subgradient would not be too "far" from the gradient since the proximal subgradient of it either consists of the gradient or is empty at the point of differentiability. So a natural question is whether or not a function can be recovered from its proximal subgradient (see Loewen [24]). Recently, Borwein, Girgensohn and Wang [3] have given a negative answer by an example. In Chapter 3, we show that one can use both proximal subgradients of f and $-f$ to recover the function. Let U be an open convex subset in a Hilbert space, and let $f, g : U \to R$ be continuous and $\partial^{\pi} f(x) \cup \partial^{\pi}(-f)(x)$ be nonempty and bounded. We prove that f and g differ by a constant on U if and only if $\partial^{\pi} f(x) = \partial^{\pi} g(x) \quad \forall x$ with $\partial^{\pi} f(x) \neq \emptyset$ and $\partial^{\pi}(-g)(x) \subseteq \partial^{\pi}(-f)(x) \quad \forall x$ with $\partial^{\pi} f(x) = \emptyset$.

In Chapter 4, we present some sufficient conditions for the existence of error bounds and metrical regularity in terms of subdifferentials.

Let C be a nonempty subset of a Banach space X and let f_i be a real-valued function defined on X for each $i = 1, \cdots, r$. Denote

$$S := \{x \in C : f_1(x) \leq 0, \cdots, f_r(x) \leq 0\}$$

and let $z \in S$. The set S is said to have *a global error bound* if there exists a constant $\mu > 0$ such that the distance function satisfies

$$d_S(x) \leq \mu \|F(x)_+\| \quad \forall x \in C, \tag{1.1}$$

where $F(x)_+ = (f_1(x)_+, \cdots, f_r(x)_+)^t$ with $f_i(x)_+ = \max\{f_i(x), 0\}$ for $i = 1, \cdots, r$ and $\|\cdot\|$ is the usual Euclidean norm. The system

$$f_1(x) \leq 0, \cdots, f_r(x) \leq 0$$

is said to be *metrically regular at z relative to C* if there exist positive constants μ and ε such that

$$d_S(x) \leq \mu \|F(x)_+\| \quad \forall x \in C \quad \textit{and} \quad \|x - z\| \leq \varepsilon.$$

The earliest result of an error bound was established by A. J. Hoffman for a linear system in R^n in 1952 [17]. This fundamental result has been extended to an analytic system [25, 26, 28], a polynomial system [18, 27], a convex system [29, 21, 22] and a Lipschitz system [38] in R^n. For an infinite dimensional space, S. M. Robinson [33] and Sien Deng [13] proved that there exists a global error bound for a convex system in a normed space which satisfies the Slater constraint qualification with S being bounded and that inequality (1.1) is also true for a locally Lipschitz convex system in a reflexive Banach space under a condition relating to recession function and unbounded set S. In Chapter 4, we will discuss some Lipschitz (or lower semicontinuous) systems in a Banach space and then use Ioffe's technique to give some sufficient conditions for the existence of error bounds for these systems in terms of the Clarke generalized gradient, the limiting subgradient (as in [38]) and the proximal subgradient. Some parallel results about metric regularity are also provided there.

Throughout this thesis, X is a real Banach space whose open unit ball is denoted by B. For any $x \in X, x+\delta B$ denotes the open ball in X with center at x and radius $\delta > 0$. Let X^* be the dual space of continuous linear functionals on X. For any $x \in X$ and $\xi \in X^*$, $\xi(x)$ is usually written as $\langle \xi, x \rangle$. H denotes a real Hilbert space with the norm induced by

$$\|x\| = \sqrt{\langle x, x \rangle} \quad \text{for each } x \in H,$$

where $\langle \cdot, \cdot \rangle$ is the inner product in H.

Chapter 2

The Generalized Gradient and The Proximal Subgradient

This chapter consists of two sections. In §2.1 some basic concepts that are related to the (Clarke) generalized gradient and that will be used throughout the rest of this thesis are reviewed and some relationships between them are discussed. In §2.2 the calculus of the proximal subgradient is developed in detail. Special attention is paid to the nonemptiness and the sum rule of the proximal subgradients.

2.1 The Clarke Generalized Gradient

Definition 2.1 Let C be a nonempty subset of X. A function $f : C \to R$ is said to be *Lipschitz* (*of rank* L) *on* C if for some nonnegative scalar L one has

$$|f(y_1) - f(y_2)| \leq L\|y_1 - y_2\| \quad \forall y_1, y_2 \in C.$$

We shall say that f is *Lipschitz of rank* L *near* x if $C = x + \delta B$ for some $\delta > 0$. A function f is said to be *locally Lipschitz* on a subset S of X if it is Lipschitz near every point in S.

Definition 2.2 Let $f : X \to R$ be Lipschitz near x. For any vector v in X, *the (Clarke) generalized directional derivative of f at x in the direction v*, denoted $f^\circ(x; v)$, is defined by

$$f^\circ(x; v) = \limsup_{\substack{y \to x \\ t \to 0+}} \frac{f(y + tv) - f(y)}{t}.$$

The Clarke generalized gradient of f at x is the subset of X^* given by

$$\partial f(x) = \{\xi \in X^* : \langle \xi, v \rangle \leq f^\circ(x; v) \quad \forall v \in X\}.$$

Note that since f is Lipschitz near x, the difference quotient whose upper limit gives the Clarke generalized directional derivative always takes values with magnitude at most $L\|v\|$. So $f^\circ(x; v)$ is well defined. Also the function $v \to f^\circ(x; v)$ is positively homogeneous and subadditive on X ([6, Proposition 2.1.1]). Hence by the Hahn-Banach Theorem (e.g., see [11]), there is a linear functional ξ majorized by $f^\circ(x; \cdot)$ and agreeing with it at v. It follows that $\partial f(x)$ is nonempty.

The Clarke generalized gradients have the following basic properties.

Proposition 2.3 [6, Proposition 2.1.2] *Let $f : X \to R$ be Lipschitz of rank L near x. Then $\partial f(x)$ is a nonempty, convex, weak*-compact subset of X^* and $\|\xi\|_* \leq L$ for every ξ in $\partial f(x)$, where $\|\xi\|_*$ is the norm of ξ in X^* defined by*

$$\|\xi\|_* := \sup\{\langle \xi, v \rangle : v \in X, \|v\| \leq 1\}.$$

Proposition 2.4 (Scalar Multiples [6, Proposition 2.3.1]) *Let $f : X \to R$ be Lipschitz near x. Then for any real scalar α, one has*

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

Proposition 2.5 (Finite Sums [6, Proposition 2.3.3]) *If $f_i : X \to R$ is Lipschitz near x for $i = 1, \cdots, m$, then*

$$\partial(\sum_{i=1}^{m} f_i)(x) \subseteq \sum_{i=1}^{m} \partial f_i(x).$$

Proposition 2.6 (Pointwise Maxima [6, Proposition 2.3.12]) *Let $f_i : X \to R$ be Lipschitz near x for $i = 1, \cdots, m$. Assume that for any $y \in X$*

$$f(y) := \max\{f_1(y), \cdots, f_m(y)\}$$

and

$$I(y) := \{1 \leq i \leq m : f_i(y) = f(y)\}.$$

Then

$$\partial f(x) \subset co\{\partial f_i(x) : i \in I(x)\},$$

where $co\, A$ denotes the convex hull of A.

Proposition 2.7 (Lebourg Mean-Value Theorem [20] or [6, Theorem 2.3.7]) *Let x and y be points in X, and suppose that f is Lipschitz on an open set containing the closed line segment $[x, y] := \{tx + (1-t)y : t \in [0, 1]\}$. Then there exists a point u in the open line segment (x, y) such that*

$$f(y) - f(x) \in \langle \partial f(u), y - x \rangle.$$

Let F map X to another Banach space Y. The usual one-sided *directional derivative of F at x in the direction v* is

$$F'(x; v) = \lim_{t \to 0^+} \frac{F(x + tv) - F(x)}{t}$$

provided that this limit exists.

Definition 2.8 Let F map X to another Banach space Y and $\mathcal{L}(X, Y)$ be the space of continuous linear operators from X to Y.

- F is said to be *Gâteaux* (*Fréchet*) *differentiable* at $x \in X$ if there is an operator $DF(x) \in \mathcal{L}(X, Y)$ such that for any v in X

$$\lim_{t \to 0+} \frac{F(x+tv) - F(x)}{t} = DF(x)(v)$$

and the convergence is uniform with respect to v in any finite (bounded) sets. The corresponding operator $DF(x)$ is called *the Gâteaux* (*Fréchet*) *derivative* of F at x.

The condition for F to be Fréchet differentiable at x is equivalent to the assertion that there exists an operator $DF(x) \in \mathcal{L}(X, Y)$ such that for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\|F(x+h) - F(x) - Df(x)(h)\| \leq \varepsilon\|h\| \text{ whenever } h \in X \text{ and } \|h\| \leq \delta$$

(cf. [30, Definition 1.12]).

- F is *strictly* (*Hadamard*) *differentiable* at $x \in X$ if there is an operator $D_sF(x) \in \mathcal{L}(X, Y)$ such that for every $v \in X$

$$\lim_{\substack{y \to x \\ t \to 0+}} \frac{F(y+tv) - F(y)}{t} = D_sF(x)(v)$$

and the convergence is uniform for v in any compact sets. The corresponding operator $D_sF(x)$ is said to be the *strict derivative* of F at x.

- F is *continuously Gâteaux* (*Fréchet*) *differentiable* at $x \in X$ if there exists $\delta > 0$ such that F is Gâteaux (Fréchet) differentiable at each point y in $x + \delta B$ and the mapping $y \to DF(y)$ is continuous at x. In particular, F is *Gâteaux* (*Fréchet*) C^{1+} at $x \in X$ if F is continuously Gâteaux (Fréchet) differentiable at x and the mapping $y \to DF(y)$ is Lipschitz on $x + \delta B$ for some $\delta > 0$.

It is easy to see that pointwise Fréchet (strict) differentiability implies Gâteaux differentiability. The following example shows that the converse may not be true.

Example 2.9 Let $f : R^2 \to R$ be defined by

$$f(x,y) = \begin{cases} \frac{x^3y}{x^4+y^2} & \text{if } x \neq 0 \text{ or } y \neq 0 \\ 0 & \text{if } x = y = 0. \end{cases}$$

It is easy to check that f is Gâteaux differentiable at $(0,0)$, and the Gâteaux derivative at this point is 0. However, since

$$\frac{|f(x,x^2)|}{\|(x,x^2)\|} = \frac{|x^3 \cdot x^2|}{(x^4+x^4)} \times \frac{1}{\sqrt{x^2+x^4}} = \frac{1}{2\sqrt{1+x^2}} \to \frac{1}{2} \quad \text{as } x \to 0,$$

f is not Fréchet differentiable at $(0,0)$.

Although pointwise Gâteaux differentiability does not imply Fréchet differentiability, a continuously Gâteaux differentiable function is always continuously Fréchet differentiable.

Proposition 2.10 (e.g., see [14]) *A function $f : X \to R$ is continuously Gâteaux differentiable at $x \in X$ if and only if f is continuously Fréchet differentiable at this point. Moreover, these derivatives are identical.*

Proof. The sufficiency of this proposition follows from the definitions. Now suppose that f is continuously Gâteaux differentiable at x. To show that f is continuously Fréchet differentiable, it suffices to prove that it is Fréchet differentiable. Under our hypothesis, for any $\varepsilon > 0$, there exists $\delta > 0$ such that the Gâteaux derivative $Df(x)$ exists at each point y in $x + \delta B$ and

$$\|Df(y) - Df(x)\|_* \leq \varepsilon \quad \text{for each } y \in x + \delta B. \tag{2.1}$$

For any $h \in X$ with $\|h\| \leq \delta/2$, let $\Phi(t) = f(x + th)$. Then for any $t \in (0,1)$

$$\begin{aligned} \Phi'(t) &= \lim_{s \to 0} \frac{\Phi(t+s) - \Phi(t)}{s} \\ &= \lim_{s \to 0} \frac{f(x + th + sh) - f(x + th)}{s} \\ &= \langle Df(x + th), h \rangle. \end{aligned}$$

Application of the mean value theorem for functions of one variable to Φ yields

$$\Phi(1) - \Phi(0) = \Phi'(\theta) \quad \text{for some } \theta \in (0,1).$$

Consequently,

$$\begin{aligned} f(x+h) - f(x) - \langle Df(x), h \rangle &= \Phi(1) - \Phi(0) - \langle Df(x), h \rangle \\ &= \Phi'(\theta) - \langle Df(x), h \rangle \\ &= \langle Df(x + \theta h), h \rangle - \langle Df(x), h \rangle \\ &= \langle Df(x + \theta h) - Df(x), h \rangle. \end{aligned}$$

Therefore

$$\begin{aligned} |f(x+h) - f(x) - \langle Df(x), h \rangle| &= |\langle Df(x + \theta h) - Df(x), h \rangle| \\ &\leq \|Df(x + \theta h) - Df(x)\|_* \cdot \|h\| \\ &\leq \varepsilon \|h\| \quad \text{(by virtue of (2.1))}, \end{aligned}$$

which implies that f is Fréchet differentiable at x and that $Df(x)$ is the Fréchet derivative of f at x. ■

Since there is no difference between continuous Gâteaux differentiability and continuous Fréchet differentiability for f at x, f is simply said to be C^1 at x if it is continuously Gâteaux differentiable at x.

Corollary 2.11 *A function $f : X \to R$ is Gâteaux C^{1+} at $x \in X$ if and only if it is Fréchet C^{1+} at x. Hence f is said to be C^{1+} at x if it is Gâteaux C^{1+} at x.*

Proposition 2.12 [6, Corollary of Proposition 2.2.1] *If $F : X \to Y$ is continuously differentiable at x, then F is Lipschitz near x and strictly differentiable at x.*

Proposition 2.13 [6, Proposition 2.2.1] *Let F map a neighborhood of x to Y, and let $\xi \in \mathcal{L}(X, Y)$. Then the following are equivalent:*

(a) *F is strictly differentiable at x and $D_sF(x) = \xi$.*

(b) *F is Lipschitz near x, and for each $v \in X$*

$$\lim_{\substack{y \to x \\ t \to 0+}} \frac{F(y + tv) - F(y)}{t} = \langle \xi, v \rangle.$$

Proposition 2.14 [6, Proposition 2.2.4] *A function $f : X \to R$ is strictly differentiable at x and $D_sf(x) = \xi$ if and only if f is Lipschitz near x and $\partial f(x) = \{\xi\}$.*

To further characterize the strict differentiability, we need the following notion and properties.

Definition 2.15 *A function* $f : X \to R$ is said to be *(Clarke) regular* at x provided

(a) For all $v \in X$, the usual one-sided directional derivative $f'(x; v)$ exists, and

(b) For all $v \in X$, $f'(x; v) = f^{\circ}(x; v)$.

Proposition 2.16 [6, Proposition 2.3.6] *Let* f *be Lipschitz near* x.

(a) *If* f *is strictly differentiable at* x*, then* f *is regular at* x.

(b) *If* f *admits a Gâteaux derivative* $Df(x)$ *and is regular at* x*, then* $\partial f(x) = \{Df(x)\}$.

The following property reveals the close relationship between the strict differentiability and the Gâteaux differentiability and regularity of a function f at x.

Proposition 2.17 *A function* $f : X \to R$ *is strictly differentiable at* x *if and only if* f *is Lipschitz near* x*, Gâteaux differentiable and regular at* x.

Proof. Let $f : X \to R$ be strictly differentiable at x. Then by Proposition 2.14 and the definition of Gâteaux differentiability (Definition 2.8), f is Lipschitz near x and Gâteaux differentiable at x, and hence regular at x by Proposition 2.16.

Conversely suppose that f is Lipschitz near x, Gâteaux differentiable and regular at x. Then it follows from Proposition 2.16 that $\partial f(x) = \{Df(x)\}$. Using Proposition 2.14 again gives that f is strictly differentiable at x. ■

2.2 The Proximal Subgradient

Definition 2.18 Let $f : X \to R \cup \{\infty\}$ be an extended real-valued function and $dom(f) := \{x \in X : f(x) < \infty\}$. f is said to be *lower semicontinuous* (*l.s.c.*) provided

$$f(x) \leq \liminf_{y \to x} f(y) \quad \forall x \in X.$$

This is equivalent to saying that the epigraph of f

$$epi\, f := \{(x, r) \in X \times R : r \geq f(x)\}$$

is closed in $X \times R$ or that the level set $\{x \in X : f(x) \leq r\}$ is closed in X for every $r \in R$.

Example 2.19 Let S be a closed subset of X. Then the *indicator function* defined by

$$\psi_S(x) := \begin{cases} 0 & \text{if } x \in S \\ \infty & \text{if } x \notin S \end{cases}$$

is l.s.c.

Definition 2.20 Let $f : X \to R \cup \{\infty\}$ be *l.s.c.* and $x \in dom(f)$. A vector $\xi \in X^*$ is said to be *a proximal subgradient* of f at x if for some $M > 0$ there exists $\delta > 0$ such that

$$f(y) - f(x) + M\|y - x\|^2 \geq \langle \xi, y - x \rangle \quad \forall y \in x + \delta B.$$

Another way to say this is that

$$\liminf_{y \to x} \frac{f(y) - f(x) - \langle \xi, y - x \rangle}{\|y - x\|^2} > -\infty.$$

The proximal subgradient of f at x, denoted by $\partial^\pi f(x)$, is the set of proximal subgradients of f at x.

Remark 2.21 Rewrite the inequality in the definition 2.20 as

$$f(y) \geq f(x) + \langle \xi, y - x \rangle - M\|y - x\|^2 \quad \forall y \in x + \delta B.$$

Then the graph of the function $y \to f(x)+\langle \xi, y-x\rangle - M\|y-x\|^2$ is a "parabola" with $(x, f(x))$ as its vertex and ξ as its slope at $y = x$. If we call this graph a *supporting parabola* with slope ξ, then ξ is a proximal subgradient of f at x if and only if there exists a supporting parabola with slope ξ such that fits under the graph of f at x.

Remark 2.22 The name of the proximal subgradient was first introduced by Rockafellar [35] for lower semicontinuous functions in R^n. The reason that "proximal" is used is that in R^n (even in a Hilbert space) the proximal subgradient can be defined through the closest point to a set. For a general Banach space, $\partial^\pi f(x)$ as defined is called the 1-Hölder-subdifferential (e.g., see [4]). For simplicity, although an abuse of terminology, we still call $\partial^\pi f(x)$ the proximal subgradient of f in a general Banach space.

Example 2.23

$(a) f(x) = -|x|,$
$$\partial^\pi f(x) = \begin{cases} \{-\frac{x}{|x|}\} & x \neq 0 \\ \emptyset & x = 0, \end{cases}$$

$$\partial^\pi(-f)(x) = \begin{cases} \{\frac{x}{|x|}\} & x \neq 0 \\ [-1, 1] & x = 0, \end{cases} \quad \partial f(x) = \begin{cases} \{-\frac{x}{|x|}\} & x \neq 0 \\ [-1, 1] & x = 0. \end{cases}$$

(b)(See Rockafellar [37])

$f(x) = -|x|^{3/2},$
$$\partial^\pi f(x) = \begin{cases} \{-\frac{3}{2}\frac{x}{|x|^{1/2}}\} & x \neq 0 \\ \emptyset & x = 0, \end{cases}$$

$$\partial^{\pi}(-f)(x) = \begin{cases} \{\frac{3}{2}\frac{x}{|x|^{1/2}}\} & x \neq 0 \\ \{0\} & x = 0, \end{cases} \quad \partial f(x) = \begin{cases} \{-\frac{3}{2}\frac{x}{|x|^{1/2}}\} & x \neq 0 \\ \{0\} & x = 0. \end{cases}$$

$$(c) f(x) = \begin{cases} -|x|^{3/2} & x \leq 0 \\ x & x > 0, \end{cases} \qquad \partial^{\pi} f(x) = \begin{cases} \{-\frac{3}{2}\frac{x}{|x|^{1/2}}\} & x < 0 \\ (0,1] & x = 0 \\ \{1\} & x > 0, \end{cases}$$

$$\partial^{\pi}(-f)(x) = \begin{cases} \{\frac{3}{2}\frac{x}{|x|^{1/2}}\} & x < 0 \\ \emptyset & x = 0 \\ \{-1\} & x > 0, \end{cases} \quad \partial f(x) = \begin{cases} \{-\frac{3}{2}\frac{x}{|x|^{1/2}}\} & x < 0 \\ [0,1] & x = 0 \\ \{1\} & x > 0. \end{cases}$$

As these examples indicate, unlike the Clarke generalized gradient, the proximal subgradient $\partial^{\pi} f(x)$ may be empty for a Lipschitz function f and need not be closed even when $\partial^{\pi} f(x)$ and $\partial f(x)$ are both of infinitely many elements. However $\partial^{\pi} f(x)$ is always convex.

Proposition 2.24 *For all x such that $\partial^{\pi} f(x) \neq \emptyset$, $\partial^{\pi} f(x)$ is convex.*

Proof. Let $\xi_1, \xi_2 \in \partial^{\pi} f(x), \lambda \in [0,1]$. Then

$$\begin{aligned} &\liminf_{y \to x} \frac{f(y) - f(x) - \langle \lambda\xi_1 + (1-\lambda)\xi_2, y - x\rangle}{\|y - x\|^2} \\ &\geq \lambda \liminf_{y \to x} \frac{f(y) - f(x) - \langle \xi_1, y - x\rangle}{\|y - x\|^2} \\ &+(1-\lambda) \liminf_{y \to x} \frac{f(y) - f(x) - \langle \xi_2, y - x\rangle}{\|y - x\|^2} > -\infty, \end{aligned}$$

which explains that $\lambda\xi_1 + (1-\lambda)\xi_2 \in \partial^{\pi} f(x)$. Hence $\partial^{\pi} f(x)$ is convex. ∎

The following proposition indicates that $\partial^{\pi} f(x)$ is always a subset of $\partial f(x)$ if f is Lipschitz near x.

Proposition 2.25 *Let $f : X \to R \cup \{\infty\}$ be Lipschitz near x. Then*

$$\partial^{\pi} f(x) \subseteq \partial f(x).$$

Proof. Let $\xi \in \partial^\pi f(x)$. By definition, there exist $M > 0$ and $\delta > 0$ such that

$$f(y) - f(x) + M\|y - x\|^2 \geq \langle \xi, y - x \rangle \quad \forall y \in x + \delta B.$$

For any $v \in X$ with $\|v\| \neq 0$, let $y = x + tv$ and $t \in (0, \frac{\delta}{\|v\|})$. Then $y \in x + \delta B$ and

$$f(x + tv) - f(x) + Mt^2\|v\|^2 \geq \langle \xi, tv \rangle.$$

Dividing both sides of this inequality by t then letting $t \to 0^+$ we have

$$\limsup_{t \to 0^+} \frac{f(x + tv) - f(x)}{t} \geq \langle \xi, v \rangle.$$

Hence

$$\begin{aligned} f^\circ(x; v) &= \limsup_{\substack{y \to x \\ t \to 0^+}} \frac{f(y + tv) - f(y)}{t} \\ &\geq \limsup_{t \to 0^+} \frac{f(x + tv) - f(x)}{t} \geq \langle \xi, v \rangle. \end{aligned}$$

That is, $\xi \in \partial f(x)$. So $\partial^\pi f(x) \subseteq \partial f(x)$. ■

It is worth discussing the question of nonemptiness of the proximal subgradient. The following propositions provide some answers to this question.

Proposition 2.26 *Let $f : X \to R \cup \{\infty\}$ attain its local minimum at x. Then*

$$0 \in \partial^\pi f(x).$$

Proof. Assume that f attains its local minimum at x. Then there exists $\delta > 0$ such that

$$f(y) \geq f(x) \quad \forall y \in x + \delta B.$$

And hence for any $M > 0$,

$$f(y) - f(x) + M\|y - x\|^2 \geq 0 = \langle 0, y - x \rangle \quad \forall y \in x + \delta B.$$

This proves that $0 \in \partial^{\pi} f(x)$. ■

Lemma 2.27 [24, Corollary 4A.5] *Let $f : X \to R \cup \{\infty\}$ be l.s.c. and Gâteaux differentiable at x. Then*

$$\partial^{\pi} f(x) \subseteq \{Df(x)\},$$

where $Df(x)$ is the Gâteaux derivative of f at x.

Proof. Suppose that $\xi \in \partial^{\pi} f(x)$. By definition, there exist $M > 0$ and $\delta > 0$ such that

$$f(y) - f(x) + M\|y - x\|^2 \geq \langle \xi, y - x \rangle \quad \forall y \in x + \delta B.$$

For any given $v \in X$ and any sufficiently small $t > 0$, substituting $y = x + tv$ into the above inequality, we have

$$\frac{f(x + tv) - f(x)}{t} - \langle \xi, v \rangle \geq -Mt\|v\|^2.$$

It follows that by letting $t \to 0^+$

$$\langle Df(x), v \rangle - \langle \xi, v \rangle \geq 0.$$

Since this holds for all v in X, we must have $\xi = Df(x)$. ■

The following example shows that a Gâteaux differentiable function may have no proximal subgradients.

Example 2.28 Consider the function $f(x) = -|x|^{\frac{3}{2}}$, $x \in R$. It is easy to see that $Df(0) = 0$ and $\partial^\pi f(0) = \emptyset$. Otherwise, suppose $\partial^\pi f(0) = \{0\}$. Then there exist $M > 0$ and $\delta > 0$ such that

$$-|x|^{3/2} + Mx^2 \geq 0 \quad \forall x \in (-\delta, \delta).$$

This is a contradiction since the inequality fails to hold for any $x \in (-\frac{1}{M^2}, 0) \cup (0, \frac{1}{M^2})$.

Note that in this example $f(x)$ is C^1 at 0. This shows that $\partial^\pi f(x)$ may be empty even for C^1 functions. However if a function f is C^{1+} (convex and Gâteaux differentiable) at x, then $\partial^\pi f(x)$ is nonempty and coincides with $\{Df(x)\}$.

Proposition 2.29 *Let $f : X \to R \cup \{\infty\}$ be C^{1+} at x. Then*

$$\partial^\pi f(x) = \{Df(x)\} = \partial f(x).$$

Proof. Since by Propositions 2.25, 2.12 and 2.14

$$\partial^\pi f(x) \subseteq \partial f(x) = \{Df(x)\},$$

by Lemma 2.27, it suffices to show that $Df(x) \in \partial^\pi f(x)$, *i.e.*, for some $M > 0$ and $\delta > 0$

$$f(y) - f(x) + M\|y - x\|^2 \geq \langle Df(x), y - x\rangle \quad \forall y \in x + \delta B.$$

Since f is C^{1+} at x, so is $-f$. By definition and Proposition 2.12, there exist $\delta > 0$ and $L > 0$ such that for any $y_1, y_2 \in x + \delta B$,

$$|(-f)(y_1) - (-f)(y_2)| \leq L\|y_1 - y_2\|,$$

$$\|D(-f)(y_1) - D(-f)(y_2)\|_* \leq L\|y_1 - y_2\|$$

and for any $y \in x + \delta B$, $-f$ is strictly differentiable at y.

For any fixed $y \in x + \delta B$, using Lebourg Mean-Value Theorem (Proposition 2.7) and Proposition 2.14, we have

$$(-f)(y) - (-f)(x) = \langle D(-f)(u), y - x \rangle,$$

where $u := tx + (1-t)y$ for some $t \in (0, 1)$. Thus

$$\|D(-f)(u) - D(-f)(x)\|_* \leq L\|u - x\| < L\|y - x\|,$$

which implies that there exists $\eta \in L\|y - x\|B^*$ such that

$$D(-f)(u) = D(-f)(x) + \eta,$$

where B^* is the open unit ball of X^*. Thus

$$\begin{aligned}(-f)(y) - (-f)(x) &= \langle D(-f)(x) + \eta, y - x \rangle \\ &= \langle D(-f)(x), y - x \rangle + \langle \eta, y - x \rangle \\ &\leq \langle D(-f)(x), y - x \rangle + L\|y - x\|^2.\end{aligned}$$

Note that $D(-f)(x) = -Df(x)$ and $y \in x + \delta B$ is arbitrary. Upon letting $M = L$, we obtain that

$$f(y) - f(x) + M\|y - x\|^2 \geq \langle Df(x), y - x \rangle \quad \forall y \in x + \delta B.$$

That is, $Df(x) \in \partial^\pi f(x)$. The proof of the proposition is completed. ∎

Proposition 2.30 *Let U be an open convex subset of X and $x \in U$. Suppose that $f : U \to R$ is convex and Gâteaux differentiable at x. Then*

$$\partial^\pi f(x) = \{Df(x)\}.$$

Proof. According to Lemma 2.27, we only need to prove that $Df(x) \in \partial^{\pi} f(x)$.

For any $y \in U$ and $t \in (0,1)$, by the convexity of f, we have

$$\begin{aligned} f(x+t(y-x)) - f(x) &\leq tf(y) + (1-t)f(x) - f(x) \\ &= t(f(y) - f(x)). \end{aligned}$$

Thus

$$f(y) - f(x) \geq \frac{f(x+t(y-x)) - f(x)}{t}.$$

From this, letting $t \to 0^+$ yields for any $y \in U$,

$$f(y) - f(x) \geq \lim_{t \to 0^+} \frac{f(x+t(y-x)) - f(x)}{t} = \langle Df(x), y - x \rangle,$$

which implies that $Df(x) \in \partial^{\pi} f(x)$. ∎

Proposition 2.31 *Let U be an open convex subset of X and $f : U \to R$ bounded above on a neighborhood of some point of U. Then f is convex if and only if f is Lipschitz near x and*

$$\partial^c f(x) = \partial^{\pi} f(x) = \partial f(x) \quad \forall x \in U,$$

where $\partial^c f(x) := \{\xi \in X^* : f(y) - f(x) \geq \langle \xi, y - x \rangle$ for all $y \in U\}$ is the subdifferential of f at x in the sense of convex analysis.

Proof. First let f be convex. For any x in U, by [6, Proposition 2.2.6], f is Lipschitz near x. Also by the definition of $\partial^c f(x)$ and $\partial^{\pi} f(x)$ and Proposition 2.25,

$$\partial^c f(x) \subseteq \partial^{\pi} f(x) \subseteq \partial f(x).$$

Furthermore, since $\partial^c f(x)$ is the subdifferential of the convex function f at x, by [6, Proposition 2.2.7],

$$\partial^c f(x) = \partial f(x).$$

It follows that

$$\partial^c f(x) = \partial^\pi f(x) = \partial f(x) \quad \forall x \in U.$$

Conversely if for each x in U, f is Lipschitz near x and $\partial^c f(x) = \partial^\pi f(x) = \partial f(x)$, then for any $x, y \in U$ and $\xi \in \partial f(x), \eta \in \partial f(y)$, one has $\xi \in \partial^c f(x)$ and $\eta \in \partial^c f(y)$, and hence

$$f(y) - f(x) \geq \langle \xi, y - x \rangle,$$

$$f(x) - f(y) \geq \langle \eta, x - y \rangle.$$

Adding the above two inequalities gives

$$\langle \eta - \xi, y - x \rangle \geq 0 \quad \forall x, y \in U \quad \textit{and} \quad \xi \in \partial f(x), \eta \in \partial f(y),$$

which implies that $\partial f(x)$ is monotone. By [6, Proposition 2.2.9], f is convex. The proof is therefore completed. ■

Corollary 2.32 *Let $x \in X$ and $n(y) := \|y - x\|$ $\forall y \in X$. Then*

$$\partial^c n(y) = \partial^\pi n(y) = \partial n(y) \quad \forall y \in X.$$

Proof. First since for any $y_1, y_2 \in X$ and $t \in [0, 1]$

$$\begin{aligned} n(ty_1 + (1-t)y_2) &= \|t(y_1 - x) + (1-t)(y_2 - x)\| \\ &\leq t\|y_1 - x\| + (1-t)\|y_2 - x\| \\ &= tn(y_1) + (1-t)n(y_2), \end{aligned}$$

$n(y)$ is convex.

Besides for any $y \in X$, let $U = y+B$. Then $n(z)$ is bounded on U. Applying Proposition 2.31 to $n(z)$ on U completes the proof. ∎

Proposition 2.33 *Let $f : X \to R$ be Lipschitz near x. Suppose that $\xi \in X^*$ satisfies*

$$\liminf_{\substack{y \to x \\ \eta \in \partial f(y)}} \frac{\langle \eta - \xi, y - x \rangle}{\|y - x\|^2} > -\infty.$$

Then $\xi \in \partial^\pi f(x)$.

Proof. Let $\delta > 0$ be small enough such that f is locally Lipschitz on $x + \delta B$. Then for any $y \in x + \delta B$, by Lebourg Mean-Value Theorem (Proposition 2.7), there exist $\theta \in (0, 1)$ and $\eta \in \partial f(x + \theta(y - x))$ such that

$$f(y) - f(x) = \langle \eta, y - x \rangle.$$

Denote $z = x + \theta(y - x)$. Then

$$\begin{aligned} \frac{f(y) - f(x) - \langle \xi, y - x \rangle}{\|y - x\|^2} &= \frac{\langle \eta - \xi, y - x \rangle}{\|y - x\|^2} \\ = \frac{\langle \eta - \xi, (z - x)/\theta \rangle}{\|z - x\|^2/\theta^2} &= \theta \frac{\langle \eta - \xi, z - x \rangle}{\|z - x\|^2}. \end{aligned}$$

By assumption,

$$\liminf_{\substack{z \to x \\ \eta \in \partial f(z)}} \frac{\langle \eta - \xi, z - x \rangle}{\|z - x\|^2} > -\infty.$$

Thus

$$\liminf_{y \to x} \frac{f(y) - f(x) - \langle \xi, y - x \rangle}{\|y - x\|^2} > -\infty,$$

which implies that $\xi \in \partial^\pi f(x)$. ∎

The following definition in R^n was first given by Rockafellar in [37].

Definition 2.34 A locally Lipschitz function $f : X \to R$ is said to be *hypomonotone* at x if for all ξ in $\partial f(x)$

$$\liminf_{\substack{y \to x \\ \eta \in \partial f(y)}} \frac{\langle \eta - \xi, y - x \rangle}{\|y - x\|^2} > -\infty.$$

A Lipschitz and convex function $f : X \to R$ is hypomonotone at each x in X since for such a function, by [6, Proposition 2.2.9], $\langle \eta - \xi, y - x \rangle \geq 0$ for any $x, y \in X$ and $\xi \in \partial f(x), \eta \in \partial f(y)$.

The following proposition in the case $X = R^n$ was given by Poliquin in [31]. We extend it to a general Banach space using the same method.

Proposition 2.35 Let $f : X \to R$ be hypomonotone at x. Then

$$\partial^\pi f(x) = \partial f(x).$$

Proof. Since f is Lipschitz near x, $\partial^\pi f(x) \subseteq \partial f(x)$. We only need to show that $\partial f(x) \subseteq \partial^\pi f(x)$. Suppose that there exists ξ in $\partial f(x)$ with ξ not in $\partial^\pi f(x)$. Then by Proposition 2.33,

$$\liminf_{\substack{y \to x \\ \eta \in \partial f(y)}} \frac{\langle \eta - \xi, y - x \rangle}{\|y - x\|^2} = -\infty,$$

which contradicts the fact that f is hypomonotone at x. ■

Proposition 2.36 *Let X be a real reflexive Banach space. Suppose that $f_i : X \to R$ is continuous for $i = 1, \cdots, m$ and $f(x) = \max\{f_1(x), \cdots, f_m(x)\}$. Denote*

$$I(x) := \{1 \leq i \leq m : f_i(x) = f(x)\} \quad \forall x \in X.$$

If f_i is Fréchet differentiable at $x \in X$ and $\partial^\pi f_i(x) \neq \emptyset$ for $i \in I(x)$, then

$$\partial^\pi f(x) = co\{Df_i(x) : i \in I(x)\},$$

i.e., $\xi \in \partial^\pi f(x)$ *if and only if*

$$\xi = \sum_{i \in I(x)} \lambda_i Df_i(x) \quad \text{for } \lambda_i \geq 0 \text{ with } \sum_{i \in I(x)} \lambda_i = 1,$$

where $Df_i(x)$ *the is the Fréchet derivative of* f_i *at* x *for* $i \in I(x)$.

Proof. Let us begin with the proof of the sufficiency.

Since for $i \in I(x), f_i$ is Fréchet differentiable at x, it is also Gâteaux differentiable at x. By Lemma 2.27 and the nonemptiness of $\partial^\pi f_i(x)$,

$$\partial^\pi f_i(x) = \{Df_i(x)\} \quad \forall i \in I(x),$$

which implies that there exist $M > 0$ and $\delta > 0$ such that for $i \in I(x)$,

$$f_i(y) - f_i(x) + M\|y - x\|^2 \geq \langle Df_i(x), y - x\rangle \quad \forall y \in x + \delta B.$$

Since for any $i \in I(x)$,

$$f(y) - f(x) \geq f_i(y) - f_i(x),$$

we have $$f(y) - f(x) + M\|y - x\|^2 \geq \langle Df_i(x), y - x\rangle \quad \forall y \in x + \delta B,$$
which implies that $Df_i(x) \in \partial^\pi f(x)$ for $i \in I(x)$.

By Proposition 2.24, $\partial^\pi f(x)$ is convex. So for $\lambda_i \geq 0$ with $\sum_{i \in I(x)} \lambda_i = 1$,

$$\xi = \sum_{i \in I(x)} \lambda_i Df_i(x) \in \partial^\pi f(x).$$

We now prove the necessity. Suppose that $\xi \in \partial^\pi f(x)$, namely for some $M > 0$ and $\delta_1 > 0$,

$$f(y) \geq f(x) - M\|y - x\|^2 + \langle \xi, y - x\rangle \quad \forall y \in x + \delta_1 B \tag{2.2}$$

but $\quad \xi \neq \sum_{i\in I(x)} \lambda_i Df_i(x) \quad$ for any $\lambda_i \geq 0$ with $\sum_{i\in I(x)} \lambda_i = 1$.

Denote

$$S_1 = \{\xi\} \text{ and } S_2 = \{ \sum_{i\in I(x)} \lambda_i Df_i(x) : \lambda_i \geq 0, \sum_{i\in I(x)} \lambda_i = 1\}.$$

Obviously S_1 and S_2 are two disjoint closed convex subsets of X^*. Also S_1 is compact. By the separation theorem [11, Theorem 3.9 (Chapter IV)], S_1 and S_2 can be strictly separated. That is, there exist $0 \neq p^{**} \in X^{**}$ and $\varepsilon > 0$ such that $\|p^{**}\| = 1$ and

$$\langle p^{**}, \xi\rangle \geq \varepsilon + \sup\{\langle p^{**}, \eta\rangle : \eta \in S_2\}.$$

Since X is reflexive, there is $p \in X$ satisfying $\|p\| = \|p^{**}\| = 1$ and

$$\langle p^{**}, x^*\rangle = \langle x^*, p\rangle \quad \forall x^* \in X^*.$$

Thus

$$\langle \xi, p\rangle \geq \varepsilon + \sup\{\langle \eta, p\rangle : \eta \in S_2\}. \tag{2.3}$$

On the other hand, by the continuity of f_i for $i = 1, \cdots, m$, there exists $\delta_2 > 0$ such that

$$f(y) = \max\{f_i(y) : i \in I(x)\} \quad y \in x + \delta_2 B.$$

For all $i \in I(x)$, Fréchet differentiability of f_i at x implies

$$f_i(y) = f_i(x) + \langle Df_i(x), y - x\rangle + \|y - x\|\alpha_i(y - x) \quad \forall y \in X,$$

where $\lim_{y\to x} \alpha_i(y - x) = 0$ for $i \in I(x)$. Therefore for any $y \in x + \delta_2 B$,

$$f(y) \;=\; \max\{f_i(y) : i \in I(x)\}$$

$$
\begin{aligned}
&= \max\{f_i(x) + \langle Df_i(x), y-x\rangle + \|y-x\|\alpha_i(y-x) : i \in I(x)\} \\
&= \max\{f(x) + \langle Df_i(x), y-x\rangle + \|y-x\|\alpha_i(y-x) : i \in I(x)\} \\
&= f(x) + \max\{\langle Df_i(x), y-x\rangle + \|y-x\|\alpha_i(y-x) : i \in I(x)\}.
\end{aligned}
$$

Taking $\delta = \min\{\delta_1, \delta_2\}$ and $y = x + \frac{1}{n}p$, then by inequality (2.2), we have

$$f(x + \frac{1}{n}p) \geq f(x) - \frac{M}{n^2} + \frac{1}{n}\langle \xi, p\rangle \qquad \text{for } n > \frac{1}{\delta},$$

and hence

$$\max\{\langle Df_i(x), p\rangle + \alpha_i(\frac{1}{n}p) : i \in I(x)\} \geq \langle \xi, p\rangle - \frac{M}{n} \qquad \text{for } n > \frac{1}{\delta}.$$

Letting $n \to \infty$ yields

$$\max\{\langle Df_i(x), p\rangle : i \in I(x)\} \geq \langle \xi, p\rangle.$$

Thus

$$
\begin{aligned}
\sup\{\langle \eta, p\rangle : \eta \in S_2\} &= \sup\{\sum_{i\in I(x)} \lambda_i \langle Df_i(x), p\rangle : \lambda_i \geq 0, \sum_{i \in I(x)} \lambda_i = 1\} \\
&\geq \max\{\langle Df_i(x), p\rangle : i \in I(x)\} \geq \langle \xi, p\rangle,
\end{aligned}
$$

which contradicts (2.3). ∎

Corollary 2.37 [1] *Let $f_1, f_2 : R^n \to R$ be differentiable and convex functions. Suppose*

$$f(x) := \max\{f_1(x), f_2(x)\} \quad \forall x \in R^n$$

and $f(\overline{x}) = f_1(\overline{x}) = f_2(\overline{x})$. Then $\xi \in \partial^c f(\overline{x})$ if and only if

$$\xi = \lambda \nabla f_1(\overline{x}) + (1-\lambda)\nabla f_2(\overline{x}) \quad \forall \lambda \in [0,1],$$

where $\partial^c f(\overline{x}) = \{\xi \in R^n : f(x) - f(\overline{x}) \geq \langle \xi, x - \overline{x}\rangle \quad \forall x \in R^n\}$ and $\nabla f_i(\overline{x})$ is the gradient vector of f_i at $\overline{x}$ for $i = 1, 2$.

Proof. Since for each $i = 1, 2,$ f_i is differentiable and convex, f_i is continuous (in fact Lipschitz continuous). And hence f is convex and bounded in a neighborhood of $\overline{x}$. By Proposition 2.31, $\partial^c f(\overline{x}) = \partial^\pi f(\overline{x})$. So by Proposition 2.36, it suffices to show that f_i is Fréchet differentiable at $\overline{x}$ and that $\partial^\pi f_i(\overline{x})$ is nonempty.

Now since for f_i and any $x \in R^n$, there exist a gradient vector $\nabla f_i(x)$ and a function $\alpha_i : R^n \to R$ such that

$$f_i(y) = f_i(x) + \nabla f_i(x)^t(y - x) + \|y - x\|\alpha_i(x, y - x) \quad \forall y \in R^n,$$

where $\lim_{y \to x} \alpha_i(x, y-x) = 0$. This means that the convex function f_i is Gâteaux differentiable at x. Note that for continuous convex functions on finite dimensional spaces, Gâteaux differentiability implies Fréchet differentiability [30]. It follows that f_i is Fréchet differentiable at x. Besides by Proposition 2.31, $\partial^\pi f_i(x) = \partial f_i(x)$, which implies that $\partial^\pi f_i(x)$ is nonempty.

Taking $x = \overline{x}$, we complete the proof. ■

Proposition 2.38 *Let $f_i : X \to R$ be Lipschitz near x for $i = 1, \cdots, m$ and $f(x) = \max\{f_1(x), \cdots, f_m(x)\}$. Denote*

$$I(x) := \{1 \leq i \leq m : f_i(x) = f(x)\} \quad \forall x \in X.$$

Suppose that $\partial^\pi f_i(x) = \partial f_i(x)$ for each $i \in I(x)$. Then

$$\partial f(x) = \partial^\pi f(x) = co\{\partial^\pi f_i(x) : i \in I(x)\}.$$

Proof. Since $\partial^\pi f_i(x) = \partial f_i(x)$ for each $i \in I(x)$, by Propositions 2.25 and 2.6, we have

$$\partial^\pi f(x) \subseteq \partial f(x) \subseteq co\{\partial f_i(x) : i \in I(x)\} = co\{\partial^\pi f_i(x) : i \in I(x)\}.$$

So it suffices to show that

$$co\{\partial^\pi f_i(x) : i \in I(x)\} \subseteq \partial^\pi f(x).$$

Now we suppose that $\xi_i \in \partial^\pi f_i(x)$ for $i \in I(x)$. Then there exist $M > 0$ and $\delta > 0$ such that

$$f_i(y) - f_i(x) + M\|y - x\|^2 \geq \langle \xi_i, y - x\rangle \quad \forall y \in x + \delta B.$$

Thus $$f(y)-f(x)+M\|y-x\|^2 \geq \langle \xi_i, y-x\rangle \quad \forall y \in x+\delta B,$$

which implies that $\xi_i \in \partial^\pi f(x)$ for any $i \in I(x)$. By Proposition 2.24, $\partial^\pi f(x)$ is convex, so for any $\lambda_i \geq 0$ with $\sum_{i\in I(x)} \lambda_i = 1$

$$\sum_{i\in I(x)} \lambda_i \xi_i \in \partial^\pi f(x).$$

This is what we need to prove. ∎

Next we discuss the calculus of the proximal subgradients.

Proposition 2.39 *For any $\lambda > 0$, we have*

$$\partial^\pi(\lambda f)(x) = \lambda \partial^\pi f(x).$$

Proof. Since for any $\lambda > 0$, $\xi \in X^*$

$$\liminf_{y\to x} \frac{(\lambda f)(y) - (\lambda f)(x) - \langle \xi, y - x\rangle}{\|y - x\|^2} > -\infty$$

if and only if

$$\liminf_{y\to x} \frac{f(y) - f(x) + \langle \frac{1}{\lambda}\xi, y - x\rangle}{\|y - x\|^2} > -\infty,$$

$\xi \in \partial^\pi(\lambda f)(x)$ if and only if $\frac{1}{\lambda}\xi \in \partial^\pi f(x)$, *i.e.*, $\xi \in \lambda\partial^\pi f(x)$. ∎

Remark 2.40 The result of Proposition 2.39 may not be true for $\lambda \leq 0$. For example, if $\lambda = 0$ and $f(x) = -|x|, x \in R$, then $\partial^\pi f(0) = \emptyset$ and $\partial^\pi(\lambda f)(0) =$

$\{0\}$; if $\lambda < 0$ and $f(x) = |x|$, then $\partial^\pi(\lambda f)(0) = \emptyset$ and $\lambda\partial^\pi f(0) = [-\lambda, \lambda]$. And hence $\partial^\pi(\lambda f)(0) \neq \lambda\partial^\pi f(0)$.

In the following section we will be adding sets. If A and B are two nonempty subsets of X then

$$A + B := \{a + b : a \in A, b \in B\}.$$

If at least one of A and B is empty, $A + B$ is defined to be empty.

Proposition 2.41 *Let $f, g : X \to R \cup \{+\infty\}$ be l.s.c., $x \in dom(f) \cap dom(g)$ and $\partial^\pi f(x) \neq \emptyset \neq \partial^\pi g(x)$. Then*

$$\partial^\pi f(x) + \partial^\pi g(x) \subseteq \partial^\pi(f + g)(x).$$

Proof. Let $\xi \in \partial^\pi f(x)$ and $\eta \in \partial^\pi g(x)$. By definition, there exist $M_1 > 0, M_2 > 0$ and $\delta > 0$ such that for any $y \in x + \delta B$,

$$f(y) - f(x) + M_1\|y - x\|^2 \geq \langle \xi, y - x\rangle$$

and
$$g(y) - g(x) + M_2\|y - x\|^2 \geq \langle \eta, y - x\rangle.$$

From these inequalities we have

$$(f + g)(y) - (f + g)(x) + (M_1 + M_2)\|y - x\|^2 \geq \langle \xi + \eta, y - x\rangle \quad \forall y \in x + \delta B,$$

i.e.,
$$\xi + \eta \in \partial^\pi(f + g)(x).$$

Hence
$$\partial^\pi f(x) + \partial^\pi g(x) \subseteq \partial^\pi(f + g)(x). \qquad \blacksquare$$

Remark 2.42 The inclusion $\partial^\pi f(x) + \partial^\pi g(x) \subseteq \partial^\pi(f + g)(x)$ was once considered nearly useless. However, it will play a special role in the proofs of Theorems 3.4 and 3.18. Besides Proposition 2.41 asserts that there is no inclusion such as

$$\partial^\pi(f + g)(x) \subseteq \partial^\pi f(x) + \partial^\pi g(x)$$

unless $$\partial^\pi(f+g)(x) = \partial^\pi f(x) + \partial^\pi g(x).$$

The latter is possible for some special cases.

Proposition 2.43 *Let $f, g : X \to R \cup \{\infty\}$ be l.s.c.. Suppose that $\partial^\pi f(x)$ is nonempty and g is C^{1+} at x. Then*

$$\partial^\pi(f \pm g)(x) = \partial^\pi f(x) \pm \partial^\pi g(x).$$

Proof. We first prove

$$\partial^\pi(f+g)(x) = \partial^\pi f(x) + \partial^\pi g(x).$$

Since $\partial^\pi f(x)$ is nonempty and by Proposition 2.29, $\partial^\pi g(x) = \{Dg(x)\}$, we only need to show

$$\partial^\pi(f+g)(x) \subseteq \partial^\pi f(x) + \{Dg(x)\}.$$

Equivalently it is enough to prove that $\xi \in \partial^\pi(f+g)(x)$ implies $\xi - Dg(x) \in \partial^\pi f(x)$. As in the proof of Proposition 2.29, we can use the Mean-Value Theorem and the Lipschitz assumption on Dg to obtain $M_1 > 0$ and $\delta_1 > 0$ such that

$$g(y) - g(x) \leq \langle Dg(x), y - x\rangle + M_1\|y - x\|^2 \quad \forall y \in x + \delta_1 B.$$

If $\xi \in \partial^\pi(f+g)(x)$, then there are some $M_2 > 0$ and $\delta_2 > 0$ such that

$$(f+g)(y) - (f+g)(x) + M_2\|y - x\|^2 \geq \langle \xi, y - x\rangle \quad \forall y \in x + \delta_2 B.$$

Upon taking $\delta = \min\{\delta_1, \delta_2\}$, one arrives at

$$f(y) - f(x) + (M_1 + M_2)\|y - x\|^2 \geq \langle \xi - Dg(x), y - x\rangle \quad \forall y \in x + \delta B,$$

which implies that $\xi - Dg(x) \in \partial^\pi f(x)$.

Next since $-g$ is also C^{1+} at x and $\partial^{\pi}(-g)(x) = \{D(-g)(x)\} = -\partial^{\pi}g(x)$,

$$\partial^{\pi}(f-g)(x) = \partial^{\pi}f(x) + \partial^{\pi}(-g)(x) = \partial^{\pi}f(x) - \partial^{\pi}g(x).$$

The proof of the proposition is therefore completed. ■

Proposition 2.43 has the following easy corollary.

Corollary 2.44 *Let $g : X \to R \cup \{\infty\}$ be C^{1+} at x. Then $\partial^{\pi}f(x)$ is nonempty if and only if $\partial^{\pi}(f+g)(x)$ is nonempty.*

According to this result, the existence of proximal subgradient of f at any point will not change whatever C^{1+} function is added to it. Also if $\partial^{\pi}f(x)$ is empty (nonempty) and $\partial^{\pi}(f+g)(x)$ is nonempty (empty), then g is not C^{1+} at x.

Proposition 2.45 *Let U be an open convex subset of X and $f, g : U \to R$ convex and Gâteaux differentiable at $x \in U$. Then*

$$\partial^{\pi}(f+g)(x) = \partial^{\pi}f(x) + \partial^{\pi}g(x).$$

Proof. It is easy to check that the sum of two functions which are convex on U and Gâteaux differentiable at x is still convex on U and Gâteaux differentiable at x. By Proposition 2.30,

$$\begin{aligned}\partial^{\pi}(f+g)(x) &= \{D(f+g)(x)\} = \{Df(x) + Dg(x)\} \\ &= \{Df(x)\} + \{Dg(x)\} = \partial^{\pi}f(x) + \partial^{\pi}g(x).\end{aligned}$$

■

Proposition 2.46 *Let $f, g : X \to R$ be Lipschitz near x. Suppose $\partial^\pi f(x) = \partial f(x)$ and $\partial^\pi g(x) = \partial g(x)$. Then*

$$\begin{aligned}\partial^\pi (f+g)(x) &= \partial^\pi f(x) + \partial^\pi g(x), \\ \partial^\pi (f+g)(x) &= \partial (f+g)(x), \\ \partial (f+g)(x) &= \partial f(x) + \partial g(x).\end{aligned}$$

Proof. Since $\partial^\pi f(x) = \partial f(x) \neq \emptyset, \partial^\pi g(x) = \partial g(x) \neq \emptyset$, by Propositions 2.41, 2.25 and 2.5,

$$\begin{aligned}\partial^\pi f(x) + \partial^\pi g(x) &\subseteq \partial^\pi (f+g)(x) \subseteq \partial (f+g)(x) \\ &\subseteq \partial f(x) + \partial g(x) = \partial^\pi f(x) + \partial^\pi g(x).\end{aligned}$$

It follows that all equalities in the above inclusions hold. ∎

The following corollary follows immediately from Propositions 2.31 and 2.46.

Corollary 2.47 *Let U be an open convex subset of X. If $f, g : U \to R$ are convex and locally Lipschitz on U, then for any $x \in U$*

$$\partial^\pi (f+g)(x) = \partial^\pi f(x) + \partial^\pi g(x).$$

Proposition 2.48 *Let $f, g : X \to R$ be hypomonotone at $x \in X$. Then $f + g$ is hypomonotone at x and*

$$\partial^\pi (f+g)(x) = \partial^\pi f(x) + \partial^\pi g(x).$$

Proof. For any $y \in X$, by Proposition 2.5,

$$\partial (f+g)(y) \subseteq \partial f(y) + \partial g(y).$$

Now for any fixed $\xi \in \partial(f+g)(x)$, let $\xi = \xi_1+\xi_2, \xi_1 \in \partial f(x)$ and $\xi_2 \in \partial g(x)$. Then

$$\begin{aligned}
&\liminf\{\frac{\langle \eta-\xi, y-x\rangle}{\|y-x\|^2} : \eta \in \partial(f+g)(y), y \to x\} \\
&= \liminf\{\frac{\langle(\eta_1+\eta_2)-(\xi_1+\xi_2), y-x\rangle}{\|y-x\|^2} : \\
&\eta_1+\eta_2 \in \partial(f+g)(y), \eta_1 \in \partial f(y), \eta_2 \in \partial g(y) \text{ and } y \to x\} \\
&\geq \liminf\{\frac{\langle \eta_1-\xi_1, y-x\rangle + \langle \eta_2-\xi_2, y-x\rangle}{\|y-x\|^2} : \\
&\eta_1 \in \partial f(y), \eta_2 \in \partial g(y) \text{ and } y \to x\} \\
&\geq \liminf\{\frac{\langle \eta_1-\xi_1, y-x\rangle}{\|y-x\|^2} : \eta_1 \in \partial f(y) \text{ and } y \to x\} \\
&+ \liminf\{\frac{\langle \eta_2-\xi_2, y-x\rangle}{\|y-x\|^2} : \eta_2 \in \partial g(y) \text{ and } y \to x\} > -\infty.
\end{aligned}$$

This explains that $f+g$ is hypomonotone at x.

By Propositions 2.35, 2.41 and 2.5, we obtain

$$\begin{aligned}
\partial f(x) + \partial g(x) &= \partial^\pi f(x) + \partial^\pi g(x) \subseteq \partial^\pi (f+g)(x) \\
&= \partial(f+g)(x) \subseteq \partial f(x) + \partial g(x).
\end{aligned}$$

Hence we have

$$\partial^\pi (f+g)(x) = \partial^\pi f(x) + \partial^\pi g(x).$$

(This equality can also be obtained by using Propositions 2.35 and 2.46) ∎

According to Propositions 2.39, 2.43, 2.45 and 2.46 we have the following result.

Theorem 2.49 *The equality*

$$\partial^\pi (\sum_{i=1}^{m} \alpha_i f_i)(x) = \sum_{i=1}^{m} \alpha_i \partial^\pi f_i(x)$$

holds if α_i and f_i satisfy one of the following conditions:

(1) $\alpha_1 > 0, \partial^\pi f_1(x) \neq \emptyset$ *and* f_i *is* C^{1+} *at* x *for* $i = 2, \cdots, m$.

(2) $\alpha_i \geq 0, f_i : U \to R$ *is convex and Gâteaux differentiable at* $x \in U$ *for* $i = 1, \cdots, m$.

(3) $\alpha_i > 0, f_i$ *is Lipschitz near* x *and* $\partial^\pi f_i(x) = \partial f_i(x)$ *for* $i = 1, \cdots, m$.

Proposition 2.50 *Let* $g_i : X \to R$ *be Lipschitz near* $x_0 \in X$ *for* $i = 1, \cdots, m$ *and* $f : R^m \to R \cup \{\infty\}$ *be l.s.c.. Suppose that* $g(x_0) := (g_1(x_0), \cdots, g_m(x_0))^t \in$ *dom(f) and* $\lambda_i \geq 0$ $(i = 1, \cdots, m)$ *whenever* $\lambda := (\lambda_1, \cdots, \lambda_m)^t \in \partial^\pi f(g(x_0))$. *Then*

$$\sum_{i=1}^{m} \lambda_i \partial^\pi g_i(x_0) \subseteq \partial^\pi (f \circ g)(x_0).$$

Proof. Let $\lambda \in \partial^\pi f(g(x_0))$. Then for some $M_1 > 0$, there exists $\delta_1 > 0$ such that

$$f(y) - f(g(x_0)) + M_1 \|y - g(x_0)\|^2 \geq \sum_{i=1}^{m} \lambda_i (y_i - g_i(x_0)) \quad \forall y \in g(x_0) + \delta_1 B_m,$$

where $y = (y_1, \cdots, y_m)^t$ and B_m is the open unit ball in R^m.

Let g be Lipschitz of rank L near x_0. Then for δ_1, there is $\delta_2 > 0$ such that

$$|g_i(x) - g_i(x_0)| \leq L\|x - x_0\| < \delta_1/m \quad \forall x \in x_0 + \delta_2 B.$$

If $\xi_i \in \partial^\pi g_i(x_0)$ for $i = 1, \cdots, m$, then there exist $M_2 > 0$ and $(\delta_2 >)\delta > 0$ such that

$$g_i(x) - g_i(x_0) \geq \langle \xi_i, x - x_0 \rangle - M_2 \|x - x_0\|^2 \quad \forall x \in x_0 + \delta B.$$

Thus

$$
\begin{aligned}
f(g(x)) - f(g(x_0)) &+ M_1 m L^2 \|x - x_0\|^2 \\
&\geq f(g(x)) - f(g(x_0)) + M_1 \sum_{i=1}^{m} |g_i(x) - g_i(x_0)|^2 \\
&\geq \sum_{i=1}^{m} \lambda_i (g_i(x) - g_i(x_0)) \\
&\geq \langle \sum_{i=1}^{m} \lambda_i \xi_i, x - x_0 \rangle - \sum_{i=1}^{m} \lambda_i M_2 \|x - x_0\|^2 \quad \forall x \in x_0 + \delta B,
\end{aligned}
$$

which implies that

$$\sum_{i=1}^{m} \lambda_i \xi_i \in \partial^{\pi} (f \circ g)(x_0).$$

Hence

$$\sum_{i=1}^{m} \lambda_i \partial^{\pi} g_i(x_0) \subseteq \partial^{\pi} (f \circ g)(x_0).$$

■

Chapter 3

Integration of The Proximal Subgradient

In the class of continuous functions, we say that a function is integrable if it can be determined, up to an additive constant, by its proximal subgradient. This fundamental problem is considered for two cases in this chapter: continuous functions in Hilbert spaces and Lipschitz functions in Banach spaces.

3.1 The Case in Hilbert Spaces

Lemma 3.1 *Let U be an open convex subset of H and $f : H \to (-\infty, \infty]$ be l.s.c.. Then f is constant on U if and only if*

$$\partial^{\pi} f(x) \subseteq \{0\} \quad \forall x \in U.$$

The result in the special case $H = R^n$ was first proved by Clarke in 1990 [8]. At a meeting in the Canadian Mathematical Society in June 1991 Clarke asked whether there might be a simpler proof than that in [8]. A positive answer was given by Clarke and Redheffer three months later [9]. The statement in a general Hilbert space is just a simple corollary of the following proposition which was showed by Clarke, Stern and Wolenski in [10].

Proposition 3.2 *Let U be an open convex subset of H and $f : U \to (-\infty, \infty]$ be l.s.c.. Then f is Lipschitz of rank $L(\geq 0)$ on U if and only if*

$$\sup\{\|\xi\| : \xi \in \partial^{\pi} f(x)\} \leq L \quad \forall x \in U.$$

Lemma 3.3 *Let A and B be nonempty subsets of H, and let B be bounded. Suppose $A + B \subseteq B$. Then $A = \{0\}$.*

Proof. Suppose that $A + B \subseteq B$ but $A \neq \{0\}$. Then there exists a in A such that $a \neq 0$. Let $P := \{p \in H : \langle p, a\rangle = 0\}$ and consider the orthogonal decomposition (e.g., see [12])

$$H = span\{a\} \oplus P.$$

Any vector b in B can be uniquely written as $b = b_1 a + b_2 p$ for some $b_1, b_2 \in R$ and $p \in P$.

Denote $M := \sup\{b_1 : b_1 a + b_2 p \in B, p \in P\}$. Then since B is bounded,

$$-\infty < M < \infty$$

and there exists b_0 in B such that $b_0 = b_1 a + b_2 p$ for some $p \in P$ and

$$M - \tfrac{1}{2} < b_1 \leq M.$$

Now by assumption, $a + b_0 = (b_1 + 1)a + b_2 p \in B$, but $b_1 + 1 > M + \frac{1}{2}$, which contradicts the definition of M. ∎

Theorem 3.4 *Let U be an open convex subset of H and $f, g : U \to R$ continuous. Suppose that for each x in U, $\partial^{\pi} f(x) \cup \partial^{\pi}(-f)(x)$ is nonempty and* *bounded. Then for some constant c,* $f(x) = g(x) + c \quad \forall x \in U$ *if and only if*

$$\partial^\pi f(x) = \partial^\pi g(x) \quad \forall x \in U \ s.t.\ \partial^\pi f(x) \neq \emptyset \tag{3.1}$$

and

$$\partial^\pi(-g)(x) \subseteq \partial^\pi(-f)(x) \quad \forall x \in U \ s.t.\ \partial^\pi f(x) = \emptyset. \tag{3.2}$$

Proof. The "only if" part of the theorem follows immediately from the definition of the proximal subgradient. We prove the "if" part of this result as follows.

Suppose that (3.1) and (3.2) hold. Consider the function $h : U \to R$ defined by

$$h(x) = f(x) - g(x) \quad \forall x \in U.$$

If $\partial^\pi h(x) = \emptyset$, then $\partial^\pi h(x) \subseteq \{0\}$. Now suppose $\partial^\pi h(x) \neq \emptyset$. Then for any $x \in U \ s.t.\ \partial^\pi f(x) \neq \emptyset$,

$$\begin{aligned} \partial^\pi h(x) + \partial^\pi g(x) &\subseteq \partial^\pi f(x) \quad \text{(by Proposition 2.41)} \\ &= \partial^\pi g(x) \quad \text{(by virtue of (3.1))} \end{aligned}$$

and for any $x \in U \ s.t.\ \partial^\pi f(x) = \emptyset$,

$$\begin{aligned} \partial^\pi h(x) + \partial^\pi(-f)(x) &\subseteq \partial^\pi(-g)(x) \quad \text{(by Proposition 2.41)} \\ &\subseteq \partial^\pi(-f)(x) \quad \text{(by virtue of (3.2))}, \end{aligned}$$

which implies by Lemma 3.3 that

$$\partial^\pi h(x) \subseteq \{0\}.$$

Thus by Lemma 3.1, for some constant c,

$$h(x) = c \qquad \forall x \in U.$$

That is, $$f(x) = g(x) + c \qquad \forall x \in U.$$

The proof of the therem is completed. ■

Corollary 3.5 *Let U_i be an open convex subset of H for $i = 1, \cdots, m$ and $H = \cup_{i=1}^m \overline{U}_i$, where $\overline{U}_i$ is the closure of U_i. Suppose that $f, g : H \to R$ are continuous and that for any x in U_i $(i = 1, \cdots, m)$, $\partial^\pi f(x) \cup \partial^\pi(-f)(x)$ is nonempty and bounded. Then for some constant c, $f(x) = g(x) + c \quad \forall x \in H$ if and only if for any $x \in U_i$ $(i = 1, \cdots, m)$*

$$\partial^\pi f(x) = \partial^\pi g(x) \text{ for } x \text{ with } \partial^\pi f(x) \neq \emptyset \tag{3.3}$$

and $$\partial^\pi(-g)(x) \subseteq \partial^\pi(-f)(x) \text{ for } x \text{ with } \partial^\pi f(x) = \emptyset. \tag{3.4}$$

Proof. The necessity part of the corollary follows immediately from Theorem 3.4.

Now assume that for any $x \in U_i$ $(i = 1, \cdots, m)$, (3.3) and (3.4) hold. Then applying Theorem 3.4 to each U_i, we find some constant c_i for each $i = 1, \cdots, m$ such that

$$f(x) = g(x) + c_i \qquad \forall x \in U_i.$$

By the continuity of f and g, we have

$$f(x) = g(x) + c_i \quad \forall x \in \overline{U_i}.$$

Let $c = f(0) - g(0)$. Then we only need to show that

$$c_i = c \qquad \text{for } i = 1, \cdots, m.$$

For any $i = 1, \cdots, m, U_i$ is convex, so is $\overline{U}_i$ (e.g., see [11]). Let $x \in H$ and consider the line segment $[0, x]$. Obviously $[0, x] \cap \overline{U}_i$ is a closed convex subset of $[0,x]$ and can be written as $[0,x]\cap\overline{U}_i=[x_i,y_i]$ for $i=1,\cdots,m$, where $[x_i,y_i]=\{tx : t\in[s_i,t_i]\}$ for some $0\le s_i\le t_i\le 1$ if $[0,x]\cap\overline{U}_i\neq\emptyset$ and $[x_i,y_i]=\emptyset$ if $[0,x]\cap\overline{U}_i=\emptyset$. Without loss of generality, we suppose that

$$0=s_1\le t_1=s_2\le t_2\cdots\le s_{m-1}\le t_{m-1}=s_m\le t_m=1.$$

Then

$$[0,x]=\cup_{i=1}^{m}[x_i,y_i],$$

and hence by the property of $[x_i,y_i]$ and the continuity of f and g,

$$\begin{aligned} f(x)-g(x) &= f(y_m)-g(y_m)=f(x_m)-g(x_m) \\ &= f(y_{m-1})-g(y_{m-1})=f(x_{m-1})-g(x_{m-1}) \\ &= \cdots\cdots=\cdots\cdots \\ &= f(y_1)-g(y_1)=f(x_1)-g(x_1) \\ &= f(0)-g(0)=c. \end{aligned}$$

In particular, for $x\in\overline{U}_i$,

$$c_i=f(x)-g(x)=c,\ \ i=1,\cdots,m$$

and the proof of the corollary is therefore completed. ∎

According to Theorem 3.4 and Corollary 3.5, if we denote

$$\int_U \partial^\pi f(x)\,dx \text{ and } \int_H \partial^\pi f(x)\,dx$$

the sets of functions which satisfy the conditions in Theorem 3.4 and Corollary 3.5 respectively and differ some constant from f, then

$$\begin{aligned} \int_U \partial^\pi f(x)\,dx &= f(x)+c \qquad \text{for } x\in U, \\ \int_H \partial^\pi f(x)\,dx &= f(x)+c \qquad \text{for } x\in U_i,\ \ i=1,\cdots,m, \end{aligned}$$

where c is an arbitrary constant.

It is worth noting that we can not arrive at the conclusion of Corollary 3.5 without the condition that $f, g : H \to R$ are continuous. Consider the following two functions

$$f(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0, \end{cases} \quad g(x) = \begin{cases} 0 & x \leq 0 \\ 2 & x > 0. \end{cases}$$

These two functions have the same proximal subgradient in $(-\infty, 0)$ and $(0, +\infty)$, yet they differ by constants on pieces of the domain.

Theorem 3.6 *Let U be an open convex subset of H and $f, g : U \to R$ locally Lipschitz. Suppose that $\partial^\pi f(x) \cup \partial^\pi(-f)(x)$ is nonempty for any x in U. Then for some constant c, $f(x) = g(x) + c \quad \forall x \in U$ if and only if*

$$\partial^\pi f(x) \subseteq \partial^\pi g(x) \quad \forall x \in U \text{ s.t. } \partial^\pi f(x) \neq \emptyset \tag{3.5}$$

and

$$\partial^\pi(-g)(x) \subseteq \partial^\pi(-f)(x) \quad \forall x \in U \text{ s.t. } \partial^\pi f(x) = \emptyset. \tag{3.6}$$

Proof. The necessity of the theorem follows directly from the definition of the proximal subgradient.

Now suppose that (3.5) and (3.6) hold. By Propositions 2.25 and 2.3, $\partial^\pi f(x) \cup \partial^\pi(-f)(x)$ is bounded since f is locally Lipschitz. Then for any $x \in U$ *s.t.* $\partial^\pi f(x) \neq \emptyset$, by Proposition 2.41 and inclusion (3.5),

$$\partial^\pi(f - g)(x) + \partial^\pi g(x) \subseteq \partial^\pi f(x) \subseteq \partial^\pi g(x)$$

and for any $x \in U$ *s.t.* $\partial^\pi f(x) = \emptyset$, by Proposition 2.41 and inclusion (3.6),

$$\partial^\pi(f - g)(x) + \partial^\pi(-f)(x) \subseteq \partial^\pi(-g)(x) \subseteq \partial^\pi(-f)(x).$$

Therefore by Lemmas 3.3 and 3.1, for some constant c,

$$f(x) = g(x) + c \quad \forall x \in U.$$

The proof is therefore completed. ■

Remark 3.7 Note that Theorem 3.4 can not be replaced by Theorem 3.6. Thanks are due to Jane Ye and Qiji Zhu for giving the following example

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

It is easy to see that

$$\partial^\pi f(x) = \begin{cases} \{2x \sin \frac{1}{x^2} - \frac{2}{x^3} \cos \frac{1}{x^2}\} & \text{if } x \neq 0 \\ \{0\} & \text{if } x = 0. \end{cases}$$

By Theorem 3.4, one can recover $f(x)$ by its proximal subgradient. However since $f'(x) = 2x \sin \frac{1}{x^2} - \frac{2}{x^3} \cos \frac{1}{x^2}$ goes to infinity as x goes to 0, $f(x)$ is not Lipschitz near 0 and hence Theorem 3.6 is invalid in this case.

Theorem 3.6 has the following corollaries.

Corollary 3.8 *Let $f, g : H \to R$ be locally Lipschitz and $\partial^\pi f(x)$ be nonempty for each x in H. Then $f(x) = g(x) \quad \forall x \in H$ if and only if*

$$f(0) = g(0) \text{ and } \partial^\pi f(x) \subseteq \partial^\pi g(x) \quad \forall x \in H.$$

The result in special case $H = R^n$ gives a positive answer to Loewen's questions in [24] under the condition that $\partial^\pi f(x)$ is nonempty:

Let $f, g : R^n \to R$ be locally Lipschitz functions with $f(0) = g(0)$ and $\partial^\pi f(x) = \partial^\pi g(x)$ for all x. Does it follow that f and g are identical? If so, can one obtain the same conclusion under the weaker hypothesis that $\partial^\pi f(x) \subseteq \partial^\pi g(x)$ for all x?

Corollary 3.9 *Let $H = \cup_{i=1}^{m} \overline{U}_i$ and $\overline{U}_i$ be the closure of the open convex set U_i for $i = 1, \cdots, m$. Suppose that $f, g : H \to R$ are locally Lipschitz and that $\partial^\pi f(x) \cup \partial^\pi(-f)(x)$ is nonempty for any $x \in U_i$ for $i = 1, \cdots, m$. Then for some constant c, $f(x) = g(x) + c \quad \forall x \in H$ if and only if for $i = 1, \cdots, m$*

$$\partial^\pi f(x) \subseteq \partial^\pi g(x) \quad \forall x \in U_i \text{ s.t. } \partial^\pi f(x) \neq \emptyset$$

and
$$\partial^\pi(-g)(x) \subseteq \partial^\pi(-f)(x) \quad \forall x \in U_i \text{ s.t. } \partial^\pi f(x) = \emptyset.$$

Example 3.10 Let U be an open convex subset of H, and let $f, g : U \to R$ be *l.s.c.*. Suppose that for each x in U, $\partial^\pi f(x) \cup \partial^\pi(-f)(x)$ is nonempty and that $\partial^\pi f$ and $\partial^\pi g$ are locally bounded at x, *i.e.*, there exist $M(x) > 0$ and $\delta(x) > 0$ such that $\|\xi\| \leq M(x)$ for any $\xi \in \partial^\pi f(y)$ or $\xi \in \partial^\pi g(y)$ and any y in $x + \delta(x)B$. By Proposition 3.2, both f and g are locally Lipschitz near x in U. According to Theorem 3.6, they differ by a constant when $\emptyset \neq \partial^\pi f(x) \subseteq \partial^\pi g(x) \quad \forall x \in U$.

Example 3.11 Reconsider Example 2.23. Let $U_1 = (-\infty, 0)$ and $U_2 = (0, \infty)$. Then $R = \overline{U}_1 \cup \overline{U}_2$. For any $x \in U_i, \partial^\pi f(x) \neq \emptyset$ for $i = 1, 2$. Within the class of locally Lipschitz functions, these functions can be recovered, up to an additive constant, from the knowledge of their proximal subgradients.

Example 3.12 Let U be an open convex subset of H. Based on Propositions 2.29, 2.31 and 2.35, the following locally Lipschitz functions satisfy $\partial^\pi f(x) \neq \emptyset \quad \forall x \in U$:

(a) $f : U \to R \cup \{\infty\}$ is C^{1+} at each x in U.

(b) $f : U \to R$ is convex.

(c) $f : U \to R$ is hypomonotone at each x in U.

By Theorem 3.6, these functions are recoverable from their proximal subgradients and

$$\int_U \partial^\pi f(x)\,dx = f(x) + c,$$

where c is an arbitrary constant.

In general, any locally Lipschitz function f which satisfies $\partial^\pi f(x) = \partial f(x)$ on U is recoverable from its proximal subgradient and hence from its Clarke generalized gradient. Furthermore for locally Lipschitz functions $f_i : U \to R$ for $i = 1, \cdots, m$, if they satisfy one of the following conditions:

(1) $\alpha_1 > 0, \partial^\pi f_1(x) \neq \emptyset$ and f_i is C^{1+} on U for $i = 2, \cdots, m$.

(2) $\alpha_i > 0$ and $\partial^\pi f_i(x) = \partial f_i(x) \quad \forall x \in U$ for $i = 1, \cdots, m$.

Then by Theorem 2.49,

$$\begin{aligned} \int_U \sum_{i=1}^m \alpha_i \partial^\pi f_i(x)\,dx &= \int_U \partial^\pi (\sum_{i=1}^m \alpha_i f_i)(x)\,dx \\ &= \sum_{i=1}^m \alpha_i f_i(x) + c. \end{aligned}$$

Therefore

$$\int_U \sum_{i=1}^m \alpha_i \partial^\pi f_i(x)\,dx = \sum_{i=1}^m \alpha_i \int_U \partial^\pi f_i(x)\,dx.$$

Example 3.13 Let $U_1 = \{(x,y) : x > 0, y > 0\}$, $U_2 = \{(x,y) : x < 0, y > 0\}$, $U_3 = \{(x,y) : x < 0, y < 0\}$ and $U_4 = \{(x,y) : x > 0, y < 0\}$. Then by Corollary 3.9, the following locally Lipschitz function $f : R^2 \to R$ defined by

$$f(x,y) = |x| - |y|$$

is the uniquely locally Lipschitz function satisfying $f(0,0)=0$ and

$$\partial^\pi f(x,y) = \begin{cases} \{(1,-1)\} & \text{if } (x,y) \in U_1, \\ \{(-1,-1)\} & \text{if } (x,y) \in U_2, \\ \{(-1,1)\} & \text{if } (x,y) \in U_3, \\ \{(1,1)\} & \text{if } (x,y) \in U_4. \end{cases}$$

Remark 3.14 For $f(x,y) = |x| - |y|$ we can find

$$\partial^\pi (f)(0,y) = \begin{cases} [-1,1] \times \{-1\} & \text{if } y > 0 \\ [-1,1] \times \{1\} & \text{if } y < 0 \end{cases}$$

and

$$\partial^\pi (-f)(x,0) = \begin{cases} \{-1\} \times [-1,1] & \text{if } x > 0 \\ \{1\} \times [-1,1] & \text{if } x < 0, \end{cases}$$

which together with Example 3.13 yields

$$\partial^\pi f(x,y) \cup \partial^\pi (-f)(x,y) \neq \emptyset \quad \forall (x,y) \neq (0,0).$$

However since $\partial^\pi f(0,0) = \emptyset$ and $\partial^\pi(-f)(0,0) = \emptyset$, we can not apply Theorem 3.6 to $U = R^2$ and $f(x,y) = |x| - |y|$ to get the above statement in Example 3.13.

This example shows that there does exist a function f which is Lipschitz near x_0 such that both $\partial^\pi f(x_0)$ and $\partial^\pi(-f)(x_0)$ are empty. So we can not use Theorem 3.6 without checking the condition that $\partial^\pi f(x) \cup \partial^\pi(-f)(x)$ is nonempty.

Example 3.15 Let C be a nonempty closed convex set in H, and let P be the nearest point mapping of H onto C, *i.e.*, for any $x \in H, P(x)$ is a point in C such that

$$\|x - P(x)\| = \inf\{\|x - y\| : y \in C\}.$$

(The closest point property [12] indicates that such a point exists uniquely.) Then the following differential equation has a unique C^{1+} solution, up to an additive constant, among all continuous functions $f : H \to R$:

$$\partial^{\pi} f(x) = \{P(x)\} \qquad \forall x \in H. \tag{3.7}$$

Proof. We prove the conclusion in the following steps [30]:

Step 1: The mapping P satisfies the following variational inequality: For all $x \in H$

$$\langle x - P(x), z - P(x) \rangle \leq 0 \quad \text{for all } z \in C.$$

In fact, for $z \in C$ and $0 < t < 1$, since $P(x) \in C$ and C is convex,

$$z_t := tz + (1-t)P(x) \in C,$$

and hence

$$\|x - P(x)\| \leq \|x - z_t\| = \|(x - P(x)) - t(z - P(x))\|.$$

Squaring and then expanding both sides of this inequality yield

$$\|x - P(x)\|^2 \leq \|x - P(x)\|^2 - 2t\langle x - P(x), z - P(x)\rangle + t^2\|z - P(x)\|^2,$$

i.e., $$0 \leq -2t\langle x - P(x), z - P(x)\rangle + t^2\|z - P(x)\|^2.$$

Dividing by t and taking the limit as $t \to 0$, we obtain the above variational inequality.

Step 2: P is a contraction.

For $y \in H$, taking $z = P(y) \in C$ and using variational inequality, we have

$$\langle x - P(x), P(y) - P(x) \rangle \leq 0.$$

Interchanging x and y in this inequality gives

$$\langle y - P(y), P(x) - P(y) \rangle \leq 0.$$

Then adding these two inequalities, we obtain

$$\langle x - y, P(x) - P(y) \rangle \geq \|P(x) - P(y)\|^2 \quad \forall x, y \in H.$$

On the other hand, by Schwarz's inequality,

$$\langle x - y, P(x) - P(y) \rangle \leq \|x - y\| \cdot \|P(x) - P(y)\|.$$

Thus
$$\|P(x) - P(y)\| \leq \|x - y\| \quad \forall x, y \in H.$$

This explains that P is a contraction. And hence P is Lipschitz continuous.

Step 3: The function $g : H \to R$ given by

$$g(x) = \tfrac{1}{2}(\|x\|^2 - \|x - P(x)\|^2)$$

is a solution of differential equation (3.7).

It is easy to see that g is continuous. We show that $P(x)$ is the Fréchet derivative of $g(x)$ as below.

Fix $x \in H$. Then for any $y \in H$, by the definition of P, we have

$$\|(x + y) - P(x + y)\| \leq \|(x + y) - P(x)\|.$$

Hence

$$\begin{aligned} \|(x+y) - P(x+y)\|^2 &\leq \|x+y\|^2 - 2\langle x+y, P(x)\rangle + \|P(x)\|^2 \\ &= \|x+y\|^2 + \|x - P(x)\|^2 - \|x\|^2 - 2\langle y, P(x)\rangle, \end{aligned}$$

which implies that

$$g(x + y) - g(x) - \langle P(x), y \rangle \geq 0.$$

Since $\|x - P(x)\| \leq \|x - P(x+y)\|$, we have

$$\begin{aligned}
g(x+y) &- g(x) - \langle P(x), y\rangle \\
&\leq \frac{1}{2}(\|x+y\|^2 - \|(x+y) - P(x+y)\|^2) \\
&- \frac{1}{2}(\|x\|^2 - \|x - P(x+y)\|^2) - \langle P(x), y\rangle \\
&= \langle x+y, P(x+y)\rangle - \frac{1}{2}\|P(x+y)\|^2 \\
&- \langle x, P(x+y)\rangle + \frac{1}{2}\|P(x+y)\|^2 - \langle P(x), y\rangle \\
&= \langle y, P(x+y) - P(x)\rangle \leq \|y\| \cdot \|P(x+y) - P(x)\| \leq \|y\|^2,
\end{aligned}$$

where the last inequality is from the contraction of P. Therefore

$$0 \leq g(x+y) - g(x) - \langle P(x), y\rangle \leq \|y\|^2 \quad \forall y \in H,$$

which explains that $P(x)$ is the Fréchet derivative of g at x.

Note that $P(x)$ is Lipschitz continuous. By Therem 2.29, g satisfies

$$\partial^\pi g(x) = \{P(x)\}.$$

Step 4: The solution of differential equation (3.7) is unique up to an additive constant.

Let f be any solution of differential equation (3.7) among the class of continuous functions. Then

$$\partial^\pi f(x) = \{P(x)\} = \partial^\pi g(x).$$

By Theorem 3.4, for some constant c,

$$f(x) = g(x) + c \quad \forall x \in H.$$

The proof is completed. ■

Another solution to this problem is to use the following result.

Proposition 3.16 *Let U be an open convex subset of H. Suppose that $g : U \to R$ is C^{1+} at each $x \in U$ and $\partial^\pi f(x) \subseteq \partial^\pi g(x)$ for all x in U. Then $f - g$ is constant on U.*

Proof. Consider

$$h(x) := f(x) - g(x), \ x \in U.$$

By Proposition 2.29, $\partial^\pi g(x)$ is nonempty. So if $\partial^\pi h(x) \neq \emptyset$, then by Proposition 2.41 and assumption, we have

$$\partial^\pi h(x) + \partial^\pi g(x) \subseteq \partial^\pi f(x) \subseteq \partial^\pi g(x).$$

It follows that $\partial^\pi h(x) = \{0\}$ from Lemma 3.3. Therefore by Lemma 3.1, $f - g$ is constant on U. ■

This proposition is an extension of the result of Clarke and Redheffer [9] in which $H = R^n$ and $g \in C^2(U)$.

3.2 The Case in Banach Spaces

Lemma 3.17 *Let U be an open convex subset of X and $h : U \to R$ locally Lipschitz on U. Suppose $\partial h(x) = \{0\}$ for each $x \in U$. Then for some constant c,*

$$h(x) \equiv c \quad \forall x \in U.$$

Proof. For fixed $x_0 \in U$ and any $x \in U$, by Lebourg Mean-Value Theorem, there exists a point u in $(x, x_0) := \{tx + (1-t)x_0 : 0 < t < 1\}$ such that

$$h(x) - h(x_0) \in \langle \partial h(u), x - x_0 \rangle.$$

Note that U is convex,$u \in (x, x_0) \subseteq U$. Thus $\partial h(u) = \{0\}$, and hence

$$h(x) = h(x_0).$$

Since x is arbitrary, $h(x) \equiv c \quad \forall x \in U$ holds for $c = h(x_0)$. ■

Lemma 3.18 *Let A and B be nonempty subsets of X^*, and let B be bounded. Suppose $A + B \subseteq B$. Then*

$$A = \{0\}.$$

Proof. Suppose that there exists a nonzero element a in A. Since

$$a + b \in B \quad \forall b \in B,$$

$$na + b \in B \quad \forall n \in N, b \in B.$$

By the boundness of B, there is a positive M such that

$$\|b\|_* \leq M \quad \forall b \in B.$$

In particular,

$$\|na + b\|_* \leq M \quad \forall n \in N \text{ and } b \in B.$$

On the other hand, since $\|a\|_* \neq 0$, for any $n \in N$ with $n > 2M/\|a\|_*$, we have

$$\|na + b\|_* \geq n\|a\|_* - \|b\|_* \geq n\|a\|_* - M > M,$$

which contradicts the boundness of B. ■

Theorem 3.19 *Let U be an open convex subset of X and $f, g : U \to R$ locally Lipschitz on U. Suppose that $\partial^\pi f(x) \cup \partial^\pi(-f)(x)$ is nonempty for each x in U. Then for some constant c, $f(x) = g(x) + c \quad \forall x \in U$ if and only if*

$$\partial^\pi f(x) \subseteq \partial^\pi g(x) \quad \forall x \in U \text{ s.t. } \partial^\pi f(x) \neq \emptyset, \tag{3.8}$$

$$\partial^\pi(-g)(x) \subseteq \partial^\pi(-f)(x) \quad \forall x \in U \text{ s.t. } \partial^\pi f(x) = \emptyset \tag{3.9}$$

and

$$\partial^\pi(f - g)(x) = \partial(f - g)(x) \quad \forall x \in U. \tag{3.10}$$

Proof. It is easy to see that if $f(x) = g(x) + c \quad \forall x \in U$ for some constant c, then for any x in U

$$\partial^\pi f(x) = \partial^\pi g(x),$$

$$\partial^\pi(-f)(x) = \partial^\pi(-g)(x)$$

and

$$\partial^\pi(f-g)(x) = \{0\} = \partial(f-g)(x).$$

Hence (3.8), (3.9) and (3.10) hold.

Conversely, we suppose that (3.8), (3.9) and (3.10) hold. Then for any $x \in U$ with $\partial^\pi f(x) \neq \emptyset$, we have

$$\begin{aligned} \partial(f-g)(x) + \partial^\pi g(x) &= \partial^\pi(f-g)(x) + \partial^\pi g(x) \quad (\text{by } (3.10)) \\ &\subseteq \partial^\pi f(x) \quad (\text{by Proposition 2.41}) \\ &\subseteq \partial^\pi g(x) \quad (\text{by } (3.8)) \end{aligned}$$

and for any $x \in U$ with $\partial^\pi f(x) = \emptyset$, since $\partial^\pi(-f)(x) \neq \emptyset$

$$\begin{aligned} \partial(f-g)(x) + \partial^\pi(-f)(x) &= \partial^\pi(f-g)(x) + \partial^\pi(-f)(x) \quad (\text{by } (3.10)) \\ &\subseteq \partial^\pi(-g)(x) \quad (\text{by Proposition 2.41}) \\ &\subseteq \partial^\pi(-f)(x) \quad (\text{by } (3.9)). \end{aligned}$$

Note that both $-f$ and g are locally Lipschitz on U. By Propositions 2.25 and 2.3, $\partial^\pi(-f)(x)$ and $\partial^\pi g(x)$ are both bounded. Therefore by Lemma 3.18, we obtain

$$\partial(f-g)(x) = \{0\} \quad \forall x \in U.$$

It follows from Lemma 3.17 that for some constant c,

$$f(x) - g(x) = c \quad \forall x \in U,$$

which is what we need to prove. ∎

Corollary 3.20 *Let U be an open convex subset of X, and $f, g : U \to R$ locally Lipschitz on U. Suppose that $\partial^{\pi} f(x)$ is nonempty for each x in U. Then for some constant c, $f(x) = g(x) + c \quad \forall x \in U$ if and only if*

$$\partial^{\pi} f(x) \subseteq \partial^{\pi} g(x) \text{ and } \partial^{\pi}(f-g)(x) = \partial(f-g)(x) \quad \forall x \in U.$$

Corollary 3.21 *Let U be an open convex subset of X. Suppose that $f, g : U \to R$ are locally Lipschitz on U,*

$$\partial^{\pi} f(x) = \partial f(x) \quad \text{and } \partial^{\pi}(-g)(x) = \partial(-g)(x) \quad \forall x \in U.$$

Then for some constant c, $f(x) = g(x) + c \quad \forall x \in U$ if and only if

$$\partial^{\pi} f(x) \subseteq \partial^{\pi} g(x) \quad \forall x \in U.$$

Proof. Since for any x in U, $\partial^{\pi} f(x) = \partial f(x) \neq \emptyset$ and $\partial^{\pi}(-g)(x) = \partial(-g)(x) \neq \emptyset$,

$$\begin{aligned} \partial^{\pi}(f-g)(x) &\subseteq \partial(f-g)(x) \quad (\text{by Proposition 2.25}) \\ &\subseteq \partial f(x) + \partial(-g)(x) \quad (\text{by Proposition 2.5}) \\ &= \partial^{\pi} f(x) + \partial^{\pi}(-g)(x) \quad (\text{by assumption}) \\ &\subseteq \partial^{\pi}(f-g)(x) \quad (\text{by Proposition 2.41}), \end{aligned}$$

which implies that

$$\partial^{\pi}(f-g)(x) = \partial(f-g)(x) \quad \forall x \in U.$$

The result follows from Corollary 3.20. ■

Corollary 3.22 *Let U be an open convex subset of X. Suppose that $f : U \to R$ is locally Lipschitz on U, $g : U \to R$ is C^{1+} on U and $\partial f(x) = \partial^{\pi} f(x) \subseteq \partial^{\pi} g(x) \quad \forall x \in U$. Then for some constant c,*

$$f(x) = g(x) + c \quad \forall x \in U.$$

Proof. Under our assumption, $-g : U \to R$ is C^{1+}. By Proposition 2.29

$$\partial^{\pi}(-g)(x) = \partial(-g)(x) \quad \forall x \in U.$$

It follows from Corollary 3.21 that for some constant c,

$$f(x) = g(x) + c \quad \forall x \in U.$$

■

Chapter 4

Error Bounds and Metric Regularity

In this chapter we give some sufficient conditions for the existence of error bounds and metric regularity for some Lipschitz (or lower semicontinuous) systems. These conditions are expressed in terms of the Clarke generalized gradient in Banach spaces, the limiting subgradient and the proximal subgradient in R^n. Ioffe's technique [19] whose essential part is an application of Ekeland's variational principle will be repeatly used.

4.1 Error Bounds

Let C be a nonempty subset of X. We define *the distance function associated with* C by

$$d_C(x) = \inf\{\|x - c\| : c \in C\} \quad \forall x \in X.$$

It is easy to check that $d_C : X \to R$ is Lipschitz of rank 1.

For $x \in C$, we denote by $T_C(x)$ *the tangent cone* to C at x which is the set

$$T_C(x) = \{v \in X : d_C^\circ(x, v) = 0\}.$$

By polarity with $T_C(x)$, we define *the normal cone* to C at x to be the set

$$N_C(x) = \{\xi \in X^* : \langle \xi, v \rangle \leq 0 \quad \forall v \in T_C(x)\}.$$

Based on Propositions 2.4.2 and 2.4.12 in [6], $N_C(x)$ can be expressed in terms of the Clarke generalized gradient as

$$N_C(x) = cl\{\cup_{\lambda \geq 0} \lambda \partial d_C(x)\} = \partial \psi_C(x),$$

where cl denotes $weak^*$ closure and ψ_C is the indicator function of C.

The distance function is recalled here largely because of the following important property.

Proposition 4.1 *Let C be a nonempty subset of X and $f : X \to R$ Lipschitz of rank L near x in C. Assume that x minimizes f over C. Then for any $\alpha \geq L$, x is a local unconstrained minimizer of the function $f + \alpha d_C$.*

Proof. Since f is Lipschitz of rank L near x, there exists $\delta > 0$ such that

$$|f(z) - f(y)| \leq L\|z - y\| \quad \forall y, z \in x + \delta B.$$

Now for any y in the ball $x + \frac{\delta}{2}B$, let z be any point in C satisfying

$$\|z - y\| < \frac{\delta}{2}.$$

(Such a point exists in C, for example, x itself lies in C with $\|x - y\| < \delta/2$.) Then z belongs to $x + \delta B$. Since x minimizes f over C,

$$f(x) \leq f(z) \leq f(y) + L\|z - y\|,$$

where the second inequality is from the Lipschitz condition. Thus for any $\alpha \geq L$,

$$\begin{aligned} f(x) &\leq \inf\{f(y) + L\|z-y\| : z \in C \text{ with } \|z-y\| < \delta/2\} \\ &= f(y) + L \cdot \inf\{\|z-y\| : z \in C \text{ with } \|z-y\| < \delta/2\} \\ &= f(y) + L \cdot \inf\{\|z-y\| : z \in C\} \\ &= f(y) + L \cdot d_C(y) \leq f(y) + \alpha d_C(y) \quad \forall y \in x + \frac{\delta}{2}B. \end{aligned}$$

That is, for any $\alpha \geq L, f + \alpha d_C$ attains a local minimum at x. ■

Definition 4.2 Let C be a nonempty closed subset of X and $f_i, |g_j| : X \to R$ be lower semicontinuous for each $i = 1, \cdots, r$ and $j = 1, \cdots, s$. Denote

$$S := \{x \in C : f_1(x) \leq 0, \cdots, f_r(x) \leq 0; g_1(x) = 0, \cdots, g_s(x) = 0\},$$

which is assumed to be nonempty. The set S is said to have a *global error bound* if there exists a constant $\mu > 0$ such that

$$d_S(x) \leq \mu(\|F(x)_+\| + \|G(x)\|) \quad \forall x \in C,$$

where $F(x)_+ = (f_1(x)_+, \cdots, f_r(x)_+)^t$ with $f_i(x)_+ := \max\{f_i(x), 0\}$ for $i = 1, \cdots, r$, $G(x) = (g_1(x), \cdots, g_s(x))^t$ and $\|\cdot\|$ is the usual Euclidean norm. The set S is said to have a *local error bound* if there exist constants $\mu > 0$ and $\delta > 0$ such that

$$d_S(x) \leq \mu(\|F(x)_+\| + \|G(x)\|) \quad \forall x \in C \text{ with } \|(F(x)_+^t, G(x)^t)^t\| < \delta.$$

Throughout this chapter, $F(x)_+$ and $G(x)$ denote the vectors stated above.

Apparently if the set S has a global (local) error bound, then functions f_i and g_j ($1 \leq i \leq r$ and $1 \leq j \leq s$) provide a global (local) error estimate for any point x in C.

Remark 4.3 To illustrate an application of error bounds in optimization problem, consider the following optimization problem with equality and inequality constraints:

$$\begin{aligned} \text{(P) minimize} \quad & h(x) \\ \text{subject to} \quad & f(x) \leq 0 \\ & g(x) = 0 \\ & x \in R^n, \end{aligned}$$

where h is Lipschitz of rank L on R^n, $f : R^n \to R$ and $g : R^n \to R$ are continuous. In this situation the feasible set $S := \{x \in R^n : f(x) \leq 0, g(x) = 0\}$ is closed and it is known that the constrained problem (P) is equivalent to the unconstrained problem:

$$\begin{aligned} \text{minimize} \quad & h(x) + \alpha d_S(x) \\ \text{subject to} \quad & x \in R^n \end{aligned}$$

for any $\alpha > L$ (see [7, Proposition 1.3]). However, the objective function of the unconstrained problem involves the distance function whose generalized gradients are difficult to compute or estimate.

If S has a global error bound, then

$$h(x) + \alpha d_S(x) \leq h(x) + \alpha\mu(f(x)_+ + |g(x)|).$$

If $H(x)$ and $K(x)$ denote the left-hand side and the right-hand side of this inequality respectively, then the above unconstrained problem (i.e., the problem of minimizing $H(x)$) is equivalent to the problem of minimizing $K(x)$. In fact, if x_0 minimizes $H(x)$, then x_0 must be in S. And hence for all x in R^n,

$$K(x_0) = h(x_0) = H(x_0) \le H(x) \le K(x).$$

Conversely suppose x_0 minimizes $K(x)$. Then if x_0 minimizes $H(x)$, we are done. Otherwise there exists $x_1(\neq x_0)$ minimizing $H(x)$. Clearly x_1 is in S, but then

$$K(x_0) \le K(x_1) = H(x_1) < H(x_0) \le K(x_0),$$

which is a contradiction. Hence the constrained problem (P) is equivalent to the problem of minimizing $K(x)$ which is of a simpler form since it is not only unconstrained but also has no distance function. Therefore the necessary optimality condition for problem (P) can be easily derived.

This remark is due to Jane Ye.

In order to present some sufficient conditions for the existence of error bounds, we will use the proof technique of Ioffe [19] whose essential part is an application of the following Ekeland's celebrated variational principle.

Proposition 4.4 (Ekeland Variational Principle [14])

Let (V, d) be a complete metric space, and let $g : V \to R \cup \{+\infty\}$ be a l.s.c. function, $\not\equiv +\infty$ and bounded below. If u is a point in V satisfying

$$g(u) \le \inf_V g + \sigma$$

for some $\sigma > 0$, then for every $\lambda > 0$, there exists a point x in V such that

$$g(x) \le g(u)\ ,\ d(u, x) \le \lambda$$

and $$g(v) + \tfrac{\sigma}{\lambda} d(v, x) > g(x) \quad \forall v \in V \text{ and } v \neq x.$$

Proof. Let us define inductively a sequence $\{u_n\}$ starting with $u_0 = u$. Suppose $u_n \in V$ is known. Now either

(a) $\forall v \neq u_n, g(v) > g(u_n) - \frac{\sigma}{\lambda} d(v, u_n)$. Then set $u_{n+1} = u_n$.

or (b) $\exists v \neq u_n : g(v) \leq g(u_n) - \frac{\sigma}{\lambda} d(v, u_n)$.

Let S_n be the set of all such v in V with inequality (b). Then for any $v \in S_n$

$$g(u_n) \geq g(v) + \frac{\sigma}{\lambda} d(v, u_n) \geq \inf_{S_n} g + \frac{\sigma}{\lambda} d(v, u_n).$$

We can choose $u_{n+1} \in S_n$ such that

$$g(u_{n+1}) - \inf_{S_n} g \leq \frac{1}{2}(g(u_n) - \inf_{S_n} g). \tag{4.1}$$

We claim that $\{u_n\}$ is a Cauchy sequence. Indeed, if case (a) ever occurs, it is stationary, and if not, we have the inequalities

$$\frac{\sigma}{\lambda} d(u_n, u_{n+1}) \leq g(u_n) - g(u_{n+1}) \quad \text{for all} \quad n \in N.$$

Adding them up, we obtain

$$\frac{\sigma}{\lambda} d(u_n, u_p) \leq g(u_n) - g(u_p) \quad \text{for all } n \leq p. \tag{4.2}$$

The sequence $g(u_n)$ is decreasing and bounded below (by $\inf_V g$), hence convergent, so the right-hand side converges to zero as (n, p) tends to infinity. Therefore $\{u_n\}$ is a Cauchy sequence. Since V is complete, $\{u_n\}$ converges to some $x \in V$.

Next we show that x is what we want to find.

Firstly the inequality $g(x) \leq g(u)$ follows from the string of inequalities

$$g(u) \geq g(u_1) \geq \cdots \geq g(u_n) \geq g(u_{n+1}) \geq \cdots$$

and the fact that g is lower semicontinuous

$$g(x) \leq \lim_{n\to\infty} g(u_n).$$

The second inequality $d(u, x) \leq \lambda$ comes from taking $n = 0$ in (4.2):

$$\frac{\sigma}{\lambda} d(u, u_p) \leq g(u) - g(u_p) \leq g(u) - \inf_V g \leq \sigma$$

by assumption and letting $p \to +\infty$.

The last one

$$g(v) + \frac{\sigma}{\lambda} d(v, x) > g(x) \quad \forall v \in V \text{ and } v \neq x$$

must hold also. Otherwise there would be some $v \neq x$ such that

$$g(v) \leq g(x) - \frac{\sigma}{\lambda} d(v, x).$$

Letting $p \to +\infty$ in inequality (4.2), we get

$$\begin{aligned} g(v) &\leq \lim_{p\to\infty} g(u_p) - \frac{\sigma}{\lambda} d(v, x) \\ &\leq g(u_n) - \frac{\sigma}{\lambda} d(u_n, x) - \frac{\sigma}{\lambda} d(v, x) \\ &\leq g(u_n) - \frac{\sigma}{\lambda} d(u_n, v) \end{aligned}$$

and hence $v \in S_n$ for all n. But relation (4.1) can be written as

$$2g(u_{n+1}) - g(u_n) \leq \inf_{S_n} g \leq g(v),$$

which combining with the lower semicontinuity of g leads to

$$g(x) \leq \lim_{n\to\infty} g(u_n) \leq g(v).$$

This contradicts the definition of v. ∎

There are other interesting proofs of the Ekeland principle in [2, 6, 14, 15]. One can also see its wide applications in analysis and optimization there. Roughly speaking, this principle says that there is a "nearby point" which actually minimizes a slightly perturbed function. One limitation on its application is that even when the original function is differentiable, the perturbed function is not. To get rid of this limitation, J. M. Borwein and D. Preiss provided their smooth variational principle in 1987 [4]. Having noted that the former one is not an exact consequence of the latter, Yongxin Li and Shuzhong Shi presented a generalization of these two principles in [23]. However our main purpose is to prove the following important lemma by applying Ekeland's variational principle as Ioffe did in [19].

Lemma 4.5 *Let C be a nonempty closed subset of X and $f : X \to R$ be Lipschitz of rank L on C. Assume that*

$$z \in S := \{y \in C : f(y) \leq 0\}$$

and that for some $\varepsilon > 0$ and $\mu > 0$, there is u in C such that

$$f(u)_+ < \varepsilon(1 + L\mu)^{-1} \; (\|u - z\| < \varepsilon(1 + L\mu)^{-1})$$

and $d_S(u) > \mu f(u)_+$. Then there exist $t > 1$ and $x \in C$ such that

$$0 < f(x) < \varepsilon \; (\|x - z\| < \varepsilon, \, f(x) > 0)$$

and
$$f(v)_+ + \psi_C(v) + (t\mu)^{-1}\|v - x\| \geq f(x) \quad \forall v \in X.$$

Proof. Suppose that for some $\varepsilon > 0$ and $\mu > 0$, the point $u \in C$ satisfies

$$f(u)_+ < \varepsilon(1 + L\mu)^{-1} \; (\text{or } \|u - z\| < \varepsilon\,(1 + L\mu)^{-1})$$

and $d_S(u) > \mu f(u)_+$. Then $u \notin S$ and hence $f(u) > 0$. Clearly we can choose

$t > 1$ such that

$$f(u) \quad < \quad \varepsilon(1+tL\mu)^{-1}\,(\text{ or } \|u-z\| < \varepsilon\,(1+tL\mu)^{-1}) \tag{4.3}$$

and $$d_S(u) \quad > \quad t\mu f(u) := \gamma. \tag{4.4}$$

It is also obvious that

$$f(u) \leq \inf_{y\in C} f(y)_+ + \gamma(t\mu)^{-1}.$$

Now consider the function

$$g(v) := f(v)_+ + \psi_C(v).$$

It is *l.s.c.* since f_+ is Lipschitz and C is closed. Also

$$g(u) \leq \inf_{v\in X} g(v) + \gamma(t\mu)^{-1}.$$

Using Ekeland's variational principle to $g(v), \sigma = \gamma(t\mu)^{-1}$ and $\lambda = \gamma$, we get $x \in C$ such that

$$\|x-u\| \leq \gamma \tag{4.5}$$

and $$g(v) + (t\mu)^{-1}\|v-x\| \geq g(x) \quad \forall v \in X,$$

i.e., $$f(v)_+ + \psi_C(v) + (t\mu)^{-1}\|v-x\| \geq f(x)_+ \geq f(x) \quad \forall v \in X.$$

Hence

$$\begin{aligned}
0 &< f(x) \quad (\text{ by (4.4), (4.5) and the fact } x\in C) \\
&\leq f(u) + L\|x-u\| \quad (\text{ by Lipschitz condition}) \\
&\leq f(u) + L\gamma \quad (\text{ by (4.5)}) \\
&= f(u)(1+Lt\mu) \quad (\text{ by (4.4)}) \\
&< \varepsilon(1+Lt\mu)^{-1}(1+Lt\mu) \quad (\text{ by (4.3)}) \\
&= \varepsilon.
\end{aligned}$$

(or $\qquad f(x) > 0 \quad$ (by (4.4), (4.5) and the fact $x \in C$),

$$\begin{aligned} \|x - z\| &\leq \|x - u\| + \|u - z\| \quad \text{(by triangle inequality)} \\ &\leq \|u - z\| + t\mu f(u) \quad \text{(by (4.4) and (4.5))} \\ &\leq \|u - z\| + t\mu(f(u) - f(z)) \quad \text{(by inequality } f(z) \leq 0) \\ &\leq \|u - z\| + t\mu L\|u - z\| \quad \text{(by Lipschitz condition)} \\ &< \varepsilon \quad \text{(by (4.3)) .)} \end{aligned}$$

This completes the proof. ∎

We are now in a position to give some sufficient conditions for the existence of error bounds.

Theorem 4.6 *Let C be a closed subset of X, and let each $f_i, |g_j| : X \to R$ be Lipschitz of rank L on an open set containing C for $i = 1, \cdots, r$ and $j = 1, \cdots, s$. Define*

$$f(x) = \max\{f_1(x), \cdots, f_r(x); |g_1(x)|, \cdots, |g_s(x)|\}.$$

Suppose

$$S := \{x \in C : f_1(x) \leq 0, \cdots, f_r(x) \leq 0; g_1(x) = 0, \cdots, g_s(x) = 0\}$$

is nonempty. Assume that there exist $\mu > 0$, $0 < \varepsilon \leq +\infty$ and $\alpha \geq L + \mu^{-1}$ such that for any x in C with $0 < f(x) < \varepsilon$ and any ξ in $\partial(f + \alpha d_C)(x)$, we have

$$\|\xi\|_* \geq \mu^{-1}.$$

Then for any x in C with $f(x)_+ < \varepsilon(1 + L\mu)^{-1}$, we have

$$d_S(x) \leq \mu f(x)_+ \leq \mu(\|F(x)_+\| + \|G(x)\|).$$

Proof. We prove this theorem in three steps as follows.

Step 1: The conclusion holds for the case $r = 1$ and $s = 0$. In this case

$$f(x) = f_1(x) \text{ and } S = \{x \in C : f(x) \leq 0\}.$$

Suppose that there were $u \in C$ such that

$$f(u)_+ < \varepsilon(1 + L\mu)^{-1} \quad \text{but} \quad d_S(u) > \mu f(u)_+.$$

Then by Lemma 4.5, there exist $t > 1$ and $x \in C$ such that $0 < f(x) < \varepsilon$ and

$$f(v)_+ + \psi_C(v) + (t\mu)^{-1}\|v - x\| \geq f(x) \quad \forall v \in X.$$

That is, the function

$$\varphi(v) := f(v)_+ + (t\mu)^{-1}\|v - x\|$$

attains its minimum on C at x. Since

$$|f(v_1)_+ - f(v_2)_+| \leq |f(v_1) - f(v_2)| \quad \forall v_1, v_2 \in X,$$

$f(v)_+$ is Lipschitz of rank L, and hence $\varphi(v)$ is Lipschitz of rank not exceeding $L + (t\mu)^{-1}$. By Propositions 4.1 and 2.26, for any $\alpha \geq L + \mu^{-1} > L + (t\mu)^{-1}$, we have

$$0 \in \partial^\pi(\varphi + \alpha d_C)(x) = \partial^\pi(f_+ + (t\mu)^{-1}h + \alpha d_C)(x), \tag{4.6}$$

where $h(v) = \|v - x\|$.

Note that $0 < f(x)$ and f is Lipschitz near x. There exists a positive constant δ such that

$$f(y)_+ = f(y) \quad \forall y \in x + \delta B.$$

According to the definition of the proximal subgradient, we have

$$\partial^\pi(f_+ + (t\mu)^{-1}h + \alpha d_C)(x) = \partial^\pi(f + (t\mu)^{-1}h + \alpha d_C)(x). \tag{4.7}$$

Thus

$$0 \in \partial^\pi(f + (t\mu)^{-1}h + \alpha d_C)(x) \quad (\text{by (4.6) and (4.7)})$$

$$\begin{aligned}
&\subseteq \partial(f+(t\mu)^{-1}h+\alpha d_C)(x) \quad (\text{by Proposition 2.25}) \\
&\subseteq \partial(f+\alpha d_C)(x)+(t\mu)^{-1}\partial h(x) \quad (\text{by Propositions 2.4 and 2.5}) \\
&\subseteq \partial(f+\alpha d_C)(x)+(t\mu)^{-1}\overline{B}_* \quad (\text{by Proposition 2.3}),
\end{aligned}$$

where $\overline{B}_*$ is the closed unit ball in X^*. The above inclusions yield the existence of $\xi \in \partial(f+\alpha d_C)(x)$ such that

$$\|\xi\|_* \leq (t\mu)^{-1} < \mu^{-1},$$

which together with the relation $x \in C$ and $0 < f(x) < \varepsilon$ contradicts the assumptions.

Step 2: The result is true when $r > 1$ and $s = 0$. Under this condition,

$$f(x) = \max\{f_1(x), \cdots, f_r(x)\}$$

and $\qquad S = \{x \in C : f_1(x) \leq 0, \cdots, f_r(x) \leq 0\} = \{x \in C : f(x) \leq 0\}.$

Denote $I(x) := \{1 \leq i \leq m : f_i(x) = f(x)\}$. Then for any $i \in I(x)$,

$$f(x)_+ = f_i(x)_+ \leq \|F(x)_+\|.$$

Since f is Lipschitz of rank L, by Step 1

$$d_S(x) \leq \mu f(x)_+ \leq \mu\|F(x)_+\| \quad \forall x \in C \text{ with } f(x)_+ < \varepsilon(1+L\mu)^{-1}.$$

Step 3: The theorem holds in the case $r \geq 0$ and $s \geq 1$.

Note that

$$S = \{x \in C : f_1(x) \leq 0, \cdots, f_r(x) \leq 0; |g_1(x)| \leq 0, \cdots, |g_s(x)| \leq 0\}.$$

By Step 2, for any $x \in C$ with $f(x)_+ < \varepsilon(1+L\mu)^{-1}$, we have

$$d_S(x) \;\leq\; \mu f(x)_+ \leq \mu\|(f_1(x)_+, \cdots, f_r(x)_+; |g_1(x)|, \cdots, |g_s(x)|)^t\|$$

$$= \mu(\|(f_1(x)_+, \cdots, f_r(x)_+)^t\|^2 + \|(|g_1(x)|, \cdots, |g_s(x)|)^t\|^2)^{1/2}$$
$$= \mu(\|F(x)_+\|^2 + \|G(x)\|^2)^{1/2}$$
$$\leq \mu(\|F(x)_+\| + \|G(x)\|).$$

The proof is therefore completed. ■

Before we apply Theorem 4.6 to a system of convex inequalities, we recall the following notions.

Definition 4.7 Let C be a nonempty closed convex subset of X. *The recession cone* of C, denoted by C^∞, is the set

$$C^\infty = \{x \in X : \exists \{\mu_i\} \subseteq (0, +\infty) \ \&\ \{x_i\} \subseteq C \ s.t. \lim_{i\to\infty} \mu_i = 0 \text{ and } \lim_{i\to\infty} \mu_i x_i = x\}.$$

According to [34, Theorem 2A(c)], C^∞ can be equivalently expressed as

$$C^\infty = \{x \in X : C + \{x\} \subseteq C\}.$$

For example, when C is a nonempty bounded closed and convex set,

$$C^\infty = \{0\}.$$

For a continuous and convex function $f : X \to R$, since its epigraph

$$epi\, f := \{(x, r) : x \in X \text{ and } r \geq f(x)\}$$

is a closed convex subset of $X \times R$, one can use the recession cone of $epi\, f$ to define *the recession function of* f, denoted by f^∞, *i.e.*,

$$epi(f^\infty) = (epi\, f)^\infty.$$

For example, for the function $f(x) = \|x\|$, $epi\, f = \{(x, r) : x \in X, \|x\| \leq r\}$, so

$$\begin{aligned}(epi\, f)^\infty &= \{(y, s) \in X \times R : epi\, f + \{(y, s)\} \subseteq epi\, f\} \\ &= \{(y, s) \in X \times R : \|y\| \leq s\} = epi\, f,\end{aligned}$$

which implies that $f^\infty(x) = \|x\|$. For more examples, we refer to [36].

As in Corollary 3.1 of [38], specializing Theorem 4.6 to convex systems, we now derive sufficient conditions for the existence of global error bounds for convex systems in terms of the recession cone of C and the recession function of f. This result was given by Deng [13] under the assumption that X is a reflexive Banach space and functions are convex and locally Lipschitz. In the following corollary, X is a Banach space.

Corollary 4.8 *Let C be a closed convex subset of X and each $f_i : X \to R$ be convex and Lipschitz of rank L on an open set containing C for $i = 1, \cdots, r$. Assume that*

$$S := \{x \in C : f_i(x) \leq 0, i = 1, \cdots, r\} \neq \emptyset$$

and denote $f(x) := \max\{f_i(x) : i = 1, \cdots, r\}$ and $F(x) := (f_1(x), \cdots, f_r(x))^t$. Suppose that there exist a unit vector $\hat{u} \in C^\infty$ and a constant $\mu > 0$ such that

$$f_i^\infty(\hat{u}) \leq -\mu^{-1} \qquad \text{for } i = 1, \cdots, r.$$

Then

$$d_S(x) \leq \mu f(x)_+ \leq \mu \|F(x)_+\| \quad \forall x \in C.$$

Proof. For any $x \in C$ and any $\alpha \geq 0$, by Proposition 2.5 and equality

$$N_C(x) = cl(\cup_{\lambda \geq 0} \lambda \partial d_C(x)),$$

we have

$$\partial(f + \alpha d_C)(x) \subseteq \partial f(x) + \alpha \partial d_C(x) \subseteq \partial f(x) + N_C(x).$$

So by Theorem 4.6, it is enough to show that for any $x \in C$ with $0 < f(x)$, any $\xi \in \partial f(x)$ and any $\eta \in N_C(x)$, we have

$$\|\xi + \eta\|_* \geq \mu^{-1}.$$

Note that C is convex. By [6, Proposition 2.4.4],

$$N_C(x) = \{\eta \in X^* : \langle \eta, y - x\rangle \leq 0 \quad \forall y \in C\}.$$

Since $\hat{u} \in C^\infty, x + \hat{u} \in C$. Then for any $\eta \in N_C(x)$,

$$\langle \eta, \hat{u}\rangle = \langle \eta, x + \hat{u} - x\rangle \leq 0.$$

And hence for any $\xi \in \partial f(x)$ and $\eta \in N_C(x)$,

$$\|\xi + \eta\|_* \geq \langle \xi + \eta, -\hat{u}\rangle \geq -\langle \xi, \hat{u}\rangle.$$

On the other hand, for any $\xi \in \partial f(x)$, by Proposition 2.6, there exist $\alpha_i \geq 0$ and $\xi_i \in \partial f_i(x)$ for each $i \in I(x)$ such that

$$\xi = \sum_{i\in I(x)} \alpha_i \xi_i \qquad \text{with} \qquad \sum_{i\in I(x)} \alpha_i = 1.$$

Recall that each f_i is convex and Lipschitz for each $i = 1, \cdots, r$, by Proposition 2.31, we have

$$\partial f_i(x) = \{\xi_i \in X^* : f_i(y) - f_i(x) \geq \langle \xi_i, y - x\rangle \quad \forall y \in X\}.$$

Applying [34, Corollary 3C(c)] to f_i, we obtain

$$\begin{aligned} f_i^\infty(\hat{u}) &= \sup\{f_i(y + \hat{u}) - f_i(y) : y \in X\} \\ &\geq \sup\{\langle \xi_i, \hat{u}\rangle : \xi_i \in \partial f_i(y), y \in X\} \\ &\geq \langle \xi_i, \hat{u}\rangle \qquad \forall \xi_i \in \partial f_i(x) \text{ and } x \in C. \end{aligned}$$

Hence

$$\begin{aligned} \|\xi + \eta\|_* &\geq -\langle \xi, \hat{u}\rangle = -\langle \sum_{i\in I(x)} \alpha_i \xi_i, \hat{u}\rangle \\ &= -\sum_{i\in I(x)} \alpha_i \langle \xi_i, \hat{u}\rangle \geq -\sum_{i\in I(x)} \alpha_i f_i^\infty(\hat{u}) \geq \mu^{-1}. \end{aligned}$$

This completes the proof. ∎

Remark 4.9 Under the condition of Corollary 4.8, C must be an unbounded set. Otherwise suppose that C is bounded, then $C^\infty = \{0\}$, which is a contradiction since we assume $0 \neq \hat{u} \in C^\infty$.

There are some limitations of its application for Theorem 4.6 because the Clarke generalized gradient often contains many (even infinitely many) elements. In R^n these conditions can be slightly weakened in terms of the proximal subgradient and the limiting subgradient.

Definition 4.10 Let $f : R^n \to R \cup \{\infty\}$ be *l.s.c.* and $x \in dom(f)$. *The limiting subgradient* of f at x, denoted by $\hat{\partial} f(x)$, is the set

$$\hat{\partial} f(x) := \{\lim_{k\to\infty} \xi_k : \xi_k \in \partial^\pi f(x_k), x_k \to x, f(x_k) \to f(x)\}.$$

The singular limiting subgradient of f at x, denoted by $\hat{\partial}^\infty f(x)$, is the set

$$\hat{\partial}^\infty f(x) := \{\lim_{k\to\infty} t_k \xi_k : \xi_k \in \partial^\pi f(x_k), x_k \to x, f(x_k) \to f(x), t_k \downarrow 0\}.$$

For instance, for the function $f(x) = -|x|, x \in R$, since $\partial^\pi f(0) = \emptyset$ and

$$\partial^\pi f(x) = \{-\frac{x}{|x|}\} \quad \text{for} \quad x \neq 0,$$

$\hat{\partial} f(0) = \{-1, 1\}$ and $\hat{\partial}^\infty f(0) = \{0\}$.

Proposition 4.11 [7, Proposition 1.2] *A l.s.c. function $f : R^n \to R \cup \{\infty\}$ is Lipschitz near x if and only if*

$$\hat{\partial}^\infty f(x) = \{0\}.$$

In that case we have $\partial f(x) = co\, \hat{\partial} f(x)$; in general we have

$$\partial f(x) = cl\, co\, \{\hat{\partial} f(x) + \hat{\partial}^\infty f(x)\},$$

where $\partial f(x)$ is the Clarke generalized gradient as defined in [7].

Proposition 4.12 *If $\hat{\partial}^{\infty} f(x) \neq \emptyset$, then $0 \in \hat{\partial}^{\infty} f(x)$.*

Proof 1. Let $\xi \in \hat{\partial}^{\infty} f(x)$. Then there exist sequences $\{t_k\} \subseteq (0, \infty)$, $\{x_k\} \subseteq R^n$ and $\{\xi_k\} \subseteq R^n$ such that

$$\xi = \lim_{k \to \infty} t_k \xi_k, \quad \xi_k \in \partial^{\pi} f(x_k), x_k \to x, f(x_k) \to f(x) \quad \text{and} \quad t_k \to 0.$$

If there exists a subsequence $\{k_i\}$ of $\{k\}$ such that $\xi_{k_i} = 0$, then

$$0 = \lim_{i \to \infty} t_{k_i} \xi_{k_i} = \lim_{k \to \infty} t_k \xi_k \in \hat{\partial}^{\infty} f(x).$$

If there is no such a subsequence, then taking

$$\bar{t}_k = \begin{cases} \frac{1}{k^2} & \text{if} \quad \|\xi_k\| \leq k; \\ \frac{1}{\|\xi_k\|^2} & \text{if} \quad \|\xi_k\| > k, \end{cases}$$

we have
$$\bar{t}_k \to 0^+ \quad \text{as} \quad k \to +\infty.$$

Thus we can pick up a subsequence of $\{\bar{t}_k\}$, which we label the same, such that $\bar{t}_k \downarrow 0$ as $k \to +\infty$. And hence

$$0 = \lim_{k \to \infty} \bar{t}_k \xi_k \in \hat{\partial}^{\infty} f(x).$$

Proof 2. Let $\xi \in \hat{\partial}^{\infty} f(x)$. Then there exist sequences $\{t_k\} \subseteq (0, \infty)$, $\{x_k\} \subseteq R^n$ and $\{\xi_k\} \subseteq R^n$ such that

$$\xi = \lim_{k \to \infty} t_k \xi_k, \quad \xi_k \in \partial^{\pi} f(x_k), x_k \to x, f(x_k) \to f(x) \quad \text{and} \quad t_k \to 0.$$

And hence

$$0 = \lim_{k \to \infty} t_k^2 \xi_k \in \hat{\partial}^{\infty} f(x).$$

■

The second proof is due to Chris Bose.

Remark 4.13 For any lower semicontinuous function f, from the above notions and propositions, we have

$$\partial^{\pi} f(x) \subseteq \hat{\partial} f(x) \subseteq \partial f(x).$$

These inclusions may also be proper. For example, for the function $f(x) = -|x|, \partial^{\pi} f(0) = \emptyset, \hat{\partial} f(0) = \{-1, 1\}$ and $\partial f(0) = [-1, 1]$.

Proposition 4.14 [7, Proposition 1.5] *Let $f, g : R^n \to R \cup \{\infty\}$ be l.s.c., $x \in (dom\, f) \cap (dom\, g)$ and $\hat{\partial}^{\infty} f(x) \cap -\hat{\partial}^{\infty} g(x) = \{0\}$. Then we have*

(a) $\hat{\partial}(f + g)(x) \subseteq \hat{\partial} f(x) + \hat{\partial} g(x)$,

(b) $\hat{\partial}^{\infty}(f + g)(x) \subseteq \hat{\partial}^{\infty} f(x) + \hat{\partial}^{\infty} g(x)$,

(c) $\partial(f + g)(x) \subseteq \partial f(x) + \partial g(x)$.

Remark 4.15 If we review the proof of [7, Proposition 1.5], then we see that the condition $\hat{\partial}^{\infty} f(x) \cap -\hat{\partial}^{\infty} g(x) = \{0\}$ in Proposition 4.14 can be replaced by $\hat{\partial}^{\infty} f(x) \cap -\hat{\partial}^{\infty} g(x) \subseteq \{0\}$. Thus if one of f and g is Lipschitz near x, then these sum rules hold. Furthermore it is worth pointing out that for general lower semicontinuous functions, the proximal subgradient does not have "exact" sum rule as the limiting subgradient and the Clarke generalized gradient do. We only have the so-called "fuzzy sum rule" stated below that is, however, the basis of the proofs of [7, Proposition 1.5] and Theorem 4.18.

Proposition 4.16 [7, Proposition 1.4] *Let $f, g : R^n \to R \cup \{\infty\}$ be l.s.c., $x \in (dom\, f) \cap (dom\, g)$ and $\xi \in \partial^{\pi}(f + g)(x)$. Then for any $\delta > 0$, there exist $x_1, x_2 \in x + \delta B$ such that*

$$f(x_1) \in f(x) + \delta B_1, g(x_2) \in g(x) + \delta B_1,$$

and
$$\xi \in \partial^\pi f(x_1) + \partial^\pi g(x_2) + \delta B,$$

where $B_1 = (-1, 1)$ *and* B *is the open unit ball in* R^n.

Next we begin to improve the conditions in Theorem 4.6 in the case where $X = R^n$ by making use of these sum rules.

Theorem 4.17 *Let* C *be a nonempty closed subset of* R^n *and* $S = \{x \in C : f(x) \leq 0\} \neq \emptyset$, *where* $f : R^n \to R$ *is Lipschitz of rank* L *on an open set containing* C. *Assume that there exist* $\mu > 0$ *and* $0 < \varepsilon \leq +\infty$ *such that*

$$\|\xi\| \geq \mu^{-1}$$

whenever $\xi \in \hat{\partial}(f + \alpha d_C)(x)$ *for some* $\alpha \geq L + \mu^{-1}$ *and any* $x \in C$ *with* $0 < f(x) < \varepsilon$. *Then we have*

$$d_S(x) \leq \mu f(x)_+$$

whenever $x \in C$ *with* $f(x)_+ < \varepsilon(1 + L\mu)^{-1}$.

Proof. Suppose that there were $u \in C$ such that

$$f(u)_+ < \varepsilon(1 + L\mu)^{-1} \quad \text{but} \quad d_S(u) > \mu f(u)_+.$$

Then by Lemma 4.5, there exist $t > 1$ and $x \in C$ such that $0 < f(x) < \varepsilon$ and

$$f(v)_+ + \psi_C(v) + (t\mu)^{-1}\|v - x\| \geq f(x) \quad \forall v \in R^n.$$

This implies that the function

$$\varphi(v) = f(v)_+ + (t\mu)^{-1}h(v)$$

attains its minimum on C at x, where $h(v) = \|v - x\|$. Since $\varphi(v)$ is Lipschitz of rank not exceeding $L + (t\mu)^{-1}$, by Propositions 4.1 and 2.26, for any $\alpha \geq L + (t\mu)^{-1}$, we have

$$0 \in \partial^\pi(\varphi + \alpha d_C)(x) \subseteq \hat{\partial}(\varphi + \alpha d_C)(x).$$

Then by Proposition 4.14, Remark 4.13 and Proposition 2.3,

$$\begin{aligned} 0 &\in \hat{\partial}(f_+ + (t\mu)^{-1}h + \alpha d_C)(x) \subseteq \hat{\partial}(f_+ + \alpha d_C)(x) + (t\mu)^{-1}\hat{\partial}h(x) \\ &\subseteq \hat{\partial}(f_+ + \alpha d_C)(x) + (t\mu)^{-1}\overline{B} = \hat{\partial}(f + \alpha d_C)(x) + (t\mu)^{-1}\overline{B}, \end{aligned}$$

where $\overline{B}$ stands for the closed unit ball in R^n and the equality is due to the Lipschitz condition of f and the fact $0 < f(x)$. This inclusion implies that for any $\alpha \geq L + \mu^{-1}$, there exists $\xi \in \hat{\partial}(f + \alpha d_C)(x)$ such that

$$\|\xi\| \leq (t\mu)^{-1} < \mu^{-1},$$

which is contrary to the assumption. ∎

Theorem 4.18 *Let C be a closed subset of R^n and $S = \{x \in C : f(x) \leq 0\} \neq \emptyset$, where $f : R^n \to R$ is Lipschitz of rank L on an open set containing C. Assume that there exist $\mu > 0$ and $0 < \varepsilon \leq \infty$ such that*

$$\|\xi\| \geq \mu^{-1}$$

whenever $\xi \in \partial^\pi(f + \psi_C)(x)$ for any $x \in C$ with $0 < f(x) < \varepsilon$. Then we have

$$d_S(x) \leq \mu f(x)_+$$

whenever $x \in C$ with $f(x)_+ < \varepsilon(1 + L\mu)^{-1}$.

Proof. Suppose that there were $u \in C$ such that

$$f(u)_+ < \varepsilon(1 + L\mu)^{-1} \quad \text{but} \quad d_S(u) > \mu f(u)_+.$$

Then by Lemma 4.5, there exist $t > 1$ and $x \in C$ such that $0 < f(x) < \varepsilon$ and

$$f(v)_+ + \psi_C(v) + (t\mu)^{-1}\|v - x\| \geq f(x) \quad \forall v \in R^n.$$

This means that the function

$$\varphi(v) = f(v)_+ + \psi_C(v) + (t\mu)^{-1}h(v)$$

attains its minimum on R^n at x, where $h(v) = \|v - x\|$. By Proposition 2.26,

$$0 \in \partial^\pi(f_+ + \psi_C + (t\mu)^{-1}h)(x). \tag{4.8}$$

Since $0 < f(x) < \varepsilon$ and f is Lipschitz continuous near x, for $\varepsilon_1 = \min\{f(x), \varepsilon - f(x)\} > 0$, there exists $\delta_1 > 0$ such that

$$-f(x) \leq -\varepsilon_1 < f(y) - f(x) < \varepsilon_1 \leq \varepsilon - f(x) \quad \forall y \in x + \delta_1 B,$$

which explains that

$$0 < f(y) < \varepsilon \quad \forall y \in x + \delta_1 B. \tag{4.9}$$

Let $\delta := \min\{\delta_1, f(x)/2, (1 - t^{-1})\mu^{-1}\}$. Then $x + \delta B \subseteq x + \delta_1 B$. According to inclusion (4.8) and Proposition 4.16, there are $x_1, x_2 \in x + \delta B$ such that

$$\begin{aligned}(f + \psi_C)(x_1) &\in ((f + \psi_C)(x) - \delta, (f + \psi_C)(x) + \delta) \\ &= (f(x) - \delta, f(x) + \delta)\end{aligned}$$

and
$$0 \in \partial^\pi (f + \psi_C)(x_1) + (t\mu)^{-1} \partial^\pi h(x_2) + \delta B.$$

The first inclusion implies that

$$x_1 \in C \cap (x + \delta B) \subseteq C \cap (x + \delta_1 B)$$

and hence by inequality (4.9), $0 < f(x_1) < \varepsilon$. And the second one explains that there exists $\xi \in \partial^\pi (f + \psi_C)(x_1)$ satisfying

$$\|\xi\| < (t\mu)^{-1} + \delta \leq (t\mu)^{-1} + (1 - t^{-1})\mu^{-1} = \mu^{-1}$$

since $\partial^\pi h(x_2) \subseteq \overline{B}$ (the closed unit ball in R^n). This contradicts the assumption.

■

Remark 4.19 Based on Steps 2 and 3 in the proof of Theorem 4.6, Theorems 4.17 and 4.18 can be extended to a Lipschitz inequalities and equalities system:

$$S := \{x \in C : f_1(x) \le 0, \cdots, f_r(x) \le 0; g_1(x) = 0, \cdots, g_s(x) = 0\}.$$

We omit the statement and the proof here. In particular, for the case $\varepsilon = \infty$, Theorem 4.18 is allowed to be extended to a lower semicontinuous system in which f_i, $|g_j|$ is merely l.s.c. for $i = 1, \cdots, r; j = 1, \cdots, s$. We first prove the following theorem.

Theorem 4.20 *Let C be a closed subset of R^n and $S = \{x \in C : f(x) \le 0\} \neq \emptyset$, where $f : R^n \to R \cup \{\infty\}$ is l.s.c.. Assume that there exists $\mu > 0$ such that*

$$\|\xi\| \ge \mu^{-1}$$

whenever $\xi \in \partial^\pi (f + \psi_C)(x)$ for any $x \in C$ with $0 < f(x) < \infty$. Then we have

$$d_S(x) \le \mu f(x)_+ \quad \forall x \in C.$$

Proof. Suppose that there were $u \in C$ such that

$$d_S(u) > \mu f(u)_+.$$

Then $u \notin S$ and hence $0 < f(u) < \infty$. Besides we can choose $t > 1$ such that

$$d_S(u) > t\mu f(u) := \gamma, \tag{4.10}$$

which explains that

$$f(u) \le \inf_{v \in C} f(v)_+ + \gamma(t\mu)^{-1}.$$

Consider the function

$$g(v) := f(v)_+ + \psi_C(v).$$

Since C is closed and f_+ is easily checked to be *l.s.c.*, g is *l.s.c.*. Also

$$g(u) \le \inf_{v \in R^n} g(v) + \gamma(t\mu)^{-1}.$$

Applying Ekeland's variational principle (Proposition 4.4) to g with $\sigma = \gamma(t\mu)^{-1}$ and $\lambda = \gamma$, we find $x \in C$ satisfying

$$\|x - u\| \leq \gamma \tag{4.11}$$

and

$$g(v) + (t\mu)^{-1}h(v) \geq g(x) \quad \forall v \in R^n, \tag{4.12}$$

where $h(v) = \|v - x\|$.

From inequalities (4.10), (4.11) and (4.12), we have

$$x \in C, x \notin S \quad \text{and} \quad f(x)_+ = g(x) < \infty,$$

i.e.,

$$x \in C \text{ and } 0 < f(x) < \infty. \tag{4.13}$$

On the other hand, inequality (4.12) implies that the function

$$g(v) + (t\mu)^{-1}h(v)$$

attains its minimum on R^n at x. Hence by Proposition 2.26,

$$0 \in \partial^\pi(g + (t\mu)^{-1}h)(x). \tag{4.14}$$

Since f is *l.s.c.* and $0 < f(x)$, there exists $\delta_1 > 0$ such that

$$0 < f(y) \qquad \forall y \in x + \delta_1 B.$$

Let $\delta := \min\{\delta_1, f(x)/2, (1 - t^{-1})\mu^{-1}\}$. Then $x + \delta B \subseteq x + \delta_1 B$. By Proposition 4.16 and inequality (4.14), there exist x_1 and x_2 both in $x + \delta B$ such that

$$\begin{aligned}(f_+ + \psi_C)(x_1) &\in ((f_+ + \psi_C)(x) - \delta, (f_+ + \psi_C)(x) + \delta) \\ &= (f(x) - \delta, f(x) + \delta)\end{aligned}$$

and

$$0 \in \partial^\pi(f_+ + \psi_C)(x_1) + (t\mu)^{-1}\partial^\pi h(x_2) + \delta B,$$

where B is the open unit ball in R^n. The first inclusion means that

$$x_1 \in C \cap (x + \delta B) \subseteq C \cap (x + \delta_1 B),$$

hence $0 < f(x_1) < \infty$. The second one implies that there exists

$$\xi \in \partial^{\pi}(f + \psi_C)(x_1)$$

such that

$$\|\xi\| < (t\mu)^{-1} + \delta \leq (t\mu)^{-1} + (1 - t^{-1})\mu^{-1} = \mu^{-1},$$

which contradicts the assumption. ■

Corollary 4.21 *Let C be a closed subset of R^n and each $f_i, |g_j| : R^n \to R \cup \{\infty\}$ be l.s.c. for $i = 1, \cdots, r$ and $j = 1, \cdots, s$. Assume that*

$$S := \{x \in C : f_1(x) \leq 0, \cdots, f_r(x) \leq 0; g_1(x) = 0, \cdots, g_s(x) = 0\} \neq \emptyset$$

and denote

$$f(x) = \max\{f_1(x), \cdots, f_r(x); |g_1(x)|, \cdots, |g_s(x)|\}.$$

Suppose that there exists $\mu > 0$ such that

$$\|\xi\| \geq \mu^{-1}$$

whenever $\xi \in \partial^{\pi}(f + \psi_C)(x)$ for any $x \in C$ with $0 < f(x) < \infty$. Then we have

$$d_S(x) \leq \mu f(x)_+ \leq \mu(\|F(x)_+\| + \|G(x)\|) \quad \forall x \in C.$$

Proof. For any $x \in R^n$, since $f_1, \cdots, f_r$ and $|g_1|, \cdots, |g_s|$ are l.s.c.,

$$\begin{aligned} \liminf_{x_m \to x} f(x_m) &= \liminf_{x_m \to x} \max\{f_1(x_m), \cdots, f_r(x_m); |g_1(x_m)|, \cdots, |g_s(x_m)|\} \\ &\geq \liminf_{x_m \to x} f_i(x_m) \text{ (and } \liminf_{x_m \to x} |g_j(x_m)|) \\ &\geq f_i(x) \text{ (and } |g_j(x)|) \quad \forall i = 1, \cdots, r \text{ (and } j = 1, \cdots, s), \end{aligned}$$

and hence

$$\liminf_{x_m \to x} f(x_m) \geq f(x) \quad \forall x \in R^n,$$

which implies that f is *l.s.c.*. By Theorem 4.20, we have

$$d_S(x) \leq \mu f(x)_+ \leq \mu(\|F(x)_+\| + \|G(x)\|) \quad \forall x \in C.$$ ■

Example 4.22 Consider the function $g : R \to R$ given by

$$g(x) = \begin{cases} 1 - |x|, & \text{if } x \text{ is a rational number;} \\ -1 + |x|, & \text{if } x \text{ is a irrational number.} \end{cases}$$

Take $C = R$. Then $S = \{x \in R : g(x) = 0\} = \{-1, 1\}, \psi_C(x) = 0$ and

$$|g(x)| = |1 - |x|| = \begin{cases} 1 - |x|, & \text{if } |x| \leq 1; \\ |x| - 1, & \text{if } |x| > 1 \end{cases}$$

is $l.s.c.$ (in fact it is Lipschitz of rank 1). It is easy to find

$$\begin{aligned} \partial^\pi |g(x)| &= \{-1\} \quad \text{for} \quad x < -1 \quad \text{or} \quad 0 < x < 1, \\ \partial^\pi |g(x)| &= \{1\} \quad \text{for} \quad -1 < x < 0 \quad \text{or} \quad 1 < x \quad \text{and} \\ \partial^\pi |g(0)| &= \emptyset. \end{aligned}$$

For any $x \in C$ with $g(x) \neq 0$, since any

$$\xi \in \partial^\pi(|g| + \psi_C)(x) = \partial^\pi |g(x)| \subseteq \{-1, 1\},$$

we have $\|\xi\| = 1$. Thus by Corollary 4.21,

$$d_S(x) \leq |g(x)| = |1 - |x|| \quad \forall x \in R.$$

4.2 Metric Regularity

Now we turn to discuss the metric regularity.

Definition 4.23 Let C be a nonempty subset of X and $f_i, |g_j| : X \to R$ lower semicontinuous for each $i = 1, \cdots, r$ and $j = 1, \cdots, s$. Let

$$z \in S := \{x \in C : f_1(x) \leq 0, \cdots, f_r(x) \leq 0; g_1(x) = 0, \cdots, g_s(x) = 0\}.$$

The system:

$$f_1(x) \le 0, \cdots, f_r(x) \le 0; g_1(x) = 0, \cdots, g_s(x) = 0$$

is said to be *metrically regular at z relative to C* if there exist positive constants μ and ε such that

$$d_S(x) \le \mu(\|F(x)_+\| + \|G(x)\|)$$

whenever $x \in C$ and $\|x - z\| \le \varepsilon$.

It is easy to see that if S has a global arror bound or if each f_i and $|g_j|$ ($1 \le i \le r$ and $1 \le j \le s$) are continuous and S has a local error bound, then the corresponding system is metrically regular at any point of S.

Theorem 4.24 *Let C be a closed subset of X, and let each $f_i, |g_j| : X \to R$ be Lipschitz of rank L on an open set containing C for $i = 1, \cdots, r$ and $j = 1, \cdots, s$. Assume that*

$$z \in S = \{x \in C : f_1(x) \le 0, \cdots, f_r(x) \le 0; g_1(x) = 0, \cdots, g_s(x) = 0\}$$

and
$$f(x) = \max\{f_1(x), \cdots, f_r(x); |g_1(x)|, \cdots, |g_s(x)|\}.$$

Suppose that there exist $\mu > 0$ and $0 < \varepsilon \le \infty$ such that

$$\|\xi\|_* \ge \mu^{-1}$$

whenever $\xi \in \partial(f + \alpha d_C)(x)$ for some $\alpha \ge L + \mu^{-1}$ and any $x \in C$ with $\|x - z\| < \varepsilon$ and $f(x) > 0$. Then we have

$$d_S(x) \le \mu f(x)_+ \le \mu(\|F(x)_+\| + \|G(x)\|)$$

whenever $x \in C$ with $\|x - z\| < \varepsilon(1 + L\mu)^{-1}$.

Proof. Suppose that there were $u \in C$ such that

$$\|u - z\| < \varepsilon(1 + L\mu)^{-1} \quad \text{and} \quad d_S(u) > \mu f(u)_+.$$

Then by Lemma 4.5, there exist $t > 1$ and $x \in C$ such that

$$\|x - z\| < \varepsilon, f(x) > 0$$

and $$f(v)_+ + \psi_C(v) + (t\mu)^{-1}\|v - x\| \geq f(x) \quad \forall v \in X.$$

The rest is similar to the proof of Theorem 4.6. ∎

Corollary 4.25 *Let C be a closed subset of X and $|g| : X \to R$ be Lipschitz of rank L on an open set containing C. Assume that $z \in S := \{x \in C : g(x) = 0\}$ and that there exist $\mu > 0$ and $0 < \varepsilon \leq \infty$ such that*

$$\|\xi\|_* \geq \mu^{-1}$$

whenever $\xi \in \partial(|g| + \alpha d_C)(x)$ for some $\alpha \geq L + \mu^{-1}$ and any $x \in C$ with $\|x - z\| < \varepsilon$ and $g(x) \neq 0$. Then we have

$$d_S(x) \leq \mu |g(x)|$$

whenever $x \in C$ with $\|x - z\| < \varepsilon(1 + L\mu)^{-1}$.

Remark 4.26 Note that

$$\partial(f + \alpha d_C)(x) \subseteq \partial f(x) + \alpha \partial d_C(x) \subseteq \partial f(x) + N_C(x).$$

We can use $\partial f(x) + \alpha \partial d_C(x)$ or $\partial f(x) + N_C(x)$ to replace $\partial(f + \alpha d_C)(x)$ in Theorem 4.24 and Corollary 4.25. In particular when we use $\partial|g|(x) + \alpha \partial d_C(x)$ or $\partial|g(x)| + N_C(x)$ to replace $\partial(|g| + \alpha d_C)(x)$ in Corollary 4.25, we obtain Ioffe's Theorem 1 and Corollary 1.1 in [19]. Finally the systems in Theorem 4.24 and Corollary 4.25 are all metrically regular at z under the corresponding conditions.

Remark 4.27 For $X = R^n$, referring to the proof of Theorem 4.17, we can use Lemma 4.5 to prove that the sets $\partial(f + \alpha d_C)(x)$ in Theorem 4.24 can be replaced by $\hat{\partial}(f + \alpha d_C)(x)$.

Theorem 4.28 *Let C be a nonempty closed subset of R^n and $f : R^n \to R$ be Lipschitz of rank L on an open set containing C. Assume that $z \in S := \{x \in C : f(x) \leq 0\}$ and that there exist $\mu > 0$ and $0 < \varepsilon \leq \infty$ such that*

$$\|\xi\| \geq \mu^{-1}$$

whenever $\xi \in \partial^\pi(f + \psi_C)(x)$ for any $x \in C$ with $\|x - z\| < \varepsilon$ and $f(x) > 0$. Then we have

$$d_S(x) \leq \mu f(x)_+$$

whenever $x \in C$ with $\|x - z\| < \varepsilon(1 + L\mu)^{-1}$.

Proof. Suppose that there were $u \in C$ such that

$$\|u - z\| < \varepsilon(1 + L\mu)^{-1} \quad \text{and} \quad d_S(u) > \mu f(u)_+.$$

Then by Lemma 4.5, there exist $t > 1$ and $x \in C$ such that

$$\|x - z\| < \varepsilon, f(x) > 0$$

and $$f(v)_+ + \psi_C(v) + (t\mu)^{-1}\|v - x\| \geq f(x) \quad \forall v \in R^n.$$

Since f is Lipschitz near x and $f(x) > 0$, there exist $\varepsilon - \|x - z\| > \delta_1 > 0$ such that

$$0 < f(y) \quad \forall y \in x + \delta_1 B.$$

Let $\delta = \min\{\delta_1, f(x)/2, (1 - t^{-1})\mu^{-1}\}$. Then $x + \delta B \subseteq x + \delta_1 B$ and for any $x_1 \in x + \delta B$,

$$\|x_1 - z\| \leq \|x_1 - x\| + \|x - z\| < \delta_1 + \|x - z\| < \varepsilon.$$

The rest is similar to the proof of Theorem 4.18. ∎

Remark 4.29 Theorem 4.28 can be extended to the same Lipschitz system as in Theorem 4.24.

Bibliography

[1] Mokhtar S. Bazaraa, Hanif D. Sherali and C. M. Shetty, *Nonlinear Programming Theory and Algorithms*, John Wiley & Sons, Inc., Singapore, 1993.

[2] J. M. Borwein, Stability and regular points of inequality systems, *Journal of Optimization Theory and Applications* **48** (1986), 9-52.

[3] J. M. Borwein, R. Girgensohn and X. Wang, On the Construction of Hölder and Proximal Subderivatives, CECM, preprint.

[4] J. M. Borwein and D. Preiss, A smooth variational principle with applications to subdifferentiability and to differentiability of convex functions, *Transactions of the American Mathematical Society* **303** (1987), 517-527.

[5] J. M. Borwein and Q. J. Zhu, Viscosity solutions and viscosity subderivatives in smooth banach spaces with applications to metric regularity, *SIAM Journal on Control and Optimization* **34** (1996), 1568-1591.

[6] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York, 1983.

[7] F. H. Clarke, *Methods of Dynamic and Nonsmooth Optimization,* CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadephia, Vol.**57**, 1989.

[8] F. H. Clarke, An indirect method in the calculus of variations, *Transactions of the American Mathematical Society* **336** (1993), 655-673.

[9] F. H. Clarke and R. M. Redheffer, The proximal subgradient and constancy, *Canadian Mathematical Bulletin* **36** (1993), 30-32.

[10] F. H. Clarke, R. J. Stern and P. R. Wolenski, Subgradient criteria for monotonicity, the Lipschitz condition, and convexity, *Canadian Journal of Mathematics* **45** (1993), 1167-1183.

[11] John B. Conway, *A Course in Functional Analysis,* 2nd Edition, Graduate Texts in Mathematics, Vol. **96**, Springer-Verlag, New York, 1990.

[12] Lokenath Debnath and Piotr Minkusinski, *Introduction to Hilbert Spaces with Applications,* Academic Press, San Diego, 1990.

[13] Sien Deng, Computable error bounds for convex inequality systems in reflexive Banach spaces, *SIAM Journal on Optimization* **7** (1997), 274-279.

[14] I. Ekeland, On the variational principle, *Journal of Mathematical Analysis and Applications* **47** (1974), 324-353.

[15] I. Ekeland, Nonconvex minimization problems, *Bulletin of the American Mathematical Society* **1** (1979), 443-474.

[16] Gerald B. Folland, *Real Analysis Modern Techniques and Their Applications,* Wiley-Interscience, New York, 1984.

[17] A. J. Hoffman, On approximate solutions of systems of linear inequalities, *Journal of Research of the National Bureau of Standards* **49** (1952), 263-265.

[18] L.Hormander, On the division of ditributions by polynomials, *Arkiv for Matematik* **3** (1958), 555-568.

[19] A. D. Ioffe, Regular points of Lipschitz functions, *Transactions of the American Mathematical Society* **251** (1979), 61-69.

[20] G. Lebourg, Valeur moyenne pour gradient généralisé, *Co.R. Acad. Sci. Paris* **281** (1975), 125-144.

[21] A. S. Lewis and J. S. Pang, Error bounds for convex inequality systems, preprint.

[22] Wu Li, Abadie's constraint qualification, metric regularity, and error bounds for differentiable convex inequalities, preprint.

[23] Yongxin Li and Shuzhong Shi, A generalization of Ekeland's ϵ- and of Borwein-Preiss' smooth ϵ- variational principles, preprint.

[24] Philip D. Loewen, *Optimal Control Via Nonsmooth Analysis,* CRM PROCEEDING & LECTURE NOTES, American Mathematical Society, Providence, 1993.

[25] M. S. Lojasiewicz, Division d'une distribution par une function analytique de variables réelles, *Comptes de Rendus de Séance, Paris* **246** (1958), 683-686.

[26] M. S. Lojasiewicz, Sur la probleme de la division, *Studia Mathematica* **18** (1959), 87-136.

[27] Xiao-Dong Luo and Zhi-Quan Luo, Extension of Hoffman's error bound to polynomial systems, *SIAM Journal on Optimization* **4** (1994), 383-392.

[28] Zhi-Quan Luo and Jong-Shi Pang, Error bounds for analytic systems and their applications, *Mathematical Programming* **67** (1995), 1-28.

[29] O. L. Mangasarian, A condition number for differentiable convex inequalities, *Mathematics of Operations Research* **10** (1985), 175-179.

[30] Robert R. Phelps, *Convex Functions, Monotone Operators and Differentiability,* 2nd Edition, *Lecture Notes in Mathematics, Vol.***1364**, Springer-Verlag, Berlin, 1993.

[31] René A. Poliquin, Integration of subdifferentials of nonconvex functions, *Nonlinear Analysis, Theory, Methods & Applications* **17** (1991), 385-398.

[32] Liqun Qi, The maximal normal operator space and integration of subdifferentials of nonconvex functions, *Nonlinear Analysis, Theory, Methods & Applications* **13** (1989), 1003-1011.

[33] S. M. Robinson, An application of error bounds for convex programming in a linear space, *SIAM Journal on Control* **13** (1975), 271-273.

[34] R. T. Rockafellar, Level sets and continuity of conjugate convex functions, *Transactions of the American Mathematical Society* **123** (1966), 46-63.

[35] R. T. Rockafellar, Proximal subgradients, marginal values, and augmented Lagrangians in nonconvex optimization, *Mathematics of Operations Research* **6** (1981), 424-436.

[36] R. T. Rockafellar, *Convex Analysis,* Princeton University Press, New Jersey, 1970.

[37] R. T. Rockafellar, Favorable classes of Lipschitz-continuous functions in subgradient optimization, *Progress in Nondifferentiable Optimization,* (Edited by E. Nurminski), 125-144. IIASA Collaborative Proceedings Series, International Institute of Applied Systems Analysis, Laxenburg, Austria 1982.

[38] Jane J. Ye, New Uniform Parametric Error Bounds, submitted to *Journal of Optimization Theory and Applications.*

VITA

Surname: Wu	Given Names: Zili
Place of Birth: P.R.China	Date of Birth: October 23, 1957

Educational Institutions Attended:

University of Victoria:	1995 to 1997
Xiamen University:	1985 to 1986
Xiamen University:	1978 to 1982

Degree Awarded:

B.Sc.	Xiamen University	1982
Graduate Certificate	Xiamen University	1986

Honours and Awards:

University of Victoria Fellowship	1995-1997
First Prize for Excellence in Teaching of Jimei Navigation Institute	September 1991
Excellent Teacher of Colleges and Universities of the Ministry of Communications, PRC	September 1988
People's Prize	1978-1982

Publications:

1.Zili Wu, Quantitative Korovkin type theorem for Linear operators on m-variate function, *Annals of Mathematical Researches* **26**(1993), 94-99.

2.Wenzhong Chen and Zili Wu, Direct theorem on approximation of two-dimentional exponential-type operators on $(C, A)_\alpha$, *Journal of Xiamen University (Natural Science)* **31** (1992), 215-19.

3.Wenzhong Chen and Zili Wu, Global approximation theorems for some exponential-type operators on space of two-variate function, *Journal of Xiamen University (Natural Science)* **31**(1992), 6-11.

4.Wenzhong Chen and Zili Wu, An approximation theorem for linear operators on H_ω^* in space of two-variate function, *Journal of Xiamen University (Natural Science)* **30**(1991), 345-48.

5.Wenzhong Chen and Zili Wu, Approximation theorems for linear operators on space of two-variate function, *Mathematica Applicata* **4**(1991), 36-43.

6.Zili Wu, Some propertise of two dimentional Kantorovich operators, *Journal of Jimei Navigation Institute* **8**(1990), 74-83.