UNIVERSITY OF ALBERTA

PROX-REGULAR FUNCTIONS IN HILBERT SPACES

BY

SANATH KUMARA BORALUGODA ©

A thesis

submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

IN

MATHEMATICS

DEPARTMENT OF MATHEMATICS

EDMONTON, ALBERTA

FALL 1998

0-612-34740-0

ABSTRACT

The prox-regular functions, a broad class of nonsmooth functions of interest in variational analysis and optimization, recently introduced by Poliquin and Rockafellar in finite-dimensional spaces, are further studied in Hilbert spaces. The key properties of prox-regular functions in $\mathbb{R}^n$ which include a subgradient characterization of prox-regularity, a Lipschitzian property of the graph of the subdifferential mapping of a prox-regular function, and smooth ($\mathcal{C}^{1+}$) and convexity (lower-$\mathcal{C}^2$) properties of its envelope functions are extended to an arbitrary Hilbert space. Subgradient and proto-derivative characterizations are also given in separable Hilbert spaces, for the convexity and the strong convexity of envelope functions. A partial extension in Hilbert space is given to the connection between the second-order Mosco epi-derivatives of prox-regular functions and the proto-derivatives of their subdifferentials.

Two new issues of prox-regular functions are taken up. First, the smoothness property of envelope functions is used to solve the fundamental problem of identifying nonsmooth functions (up to an additive constant) from their subdifferentials for a large class of prox-regular functions in Hilbert space. Second, the basic calculus rules such as addition of prox-regular functions, and a more general form of chain rule (composition of a prox-regular function with a $\mathcal{C}^{1+}$ mapping) are developed in finite-dimensional spaces.

TO:

MY PARENTS AND MY TEACHERS

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my supervisor Professor René Poliquin for his guidance, support and encouragement throughout the course of my studies at the University of Alberta. I am also grateful to him for introducing me to *Nonsmooth Analysis* .

In addition, I would like to thank the Professors Jack Macki, Henry Van Roessel, Wieslaw Krawcewicz, and Jane Ye for their valuable advice and interest in my research.

Finally, I gratefully acknowledge the funding I have received from the Canadian government under the Commonwealth Scholarship Program as well as the financial support from the University of Alberta.

CONTENTS

CHAPTER 1. INTRODUCTION **1**

1.1 Background and Motivation 1

1.2 Hilbert Space Extensions and New Issues 6

1.3 Notation 10

CHAPTER 2. PROX-REGULAR FUNCTIONS IN HILBERT SPACES **13**

2.1 Proximal Analysis 14

2.2 Prox-Regular Functions 22

2.3 Subgradient Characterization of Prox-Regularity 26

2.4 Regularity Properties of Moreau Envelopes 33

2.5 Convexity of Moreau Envelopes 43

2.6 Second-Order Theory 55

CHAPTER 3. INTEGRATION OF PROX-REGULAR FUNCTIONS **65**

3.1 Integration Problem 65

3.2 Main Result 67

3.3 Necessity of the Assumptions 70

CHAPTER 4. CALCULUS OF PROX-REGULAR FUNCTIONS **73**

4.1 The Chain Rule 73

4.2 Some Applications 79

REFERENCES **82**

List of Symbols

$\mathbb{R}$	The real numbers
$\overline{\mathbb{R}}$	The extended real numbers
$\mathbb{R}_+$	The set of positive real numbers
$\mathbb{N}$	The natural numbers
$\mathbb{B}$	The open unit ball
$\overline{\mathbb{B}}$	The closed unit ball
$\mathbb{B}(x;r)$	The open ball of radius $r > 0$, centered at x
$\overline{\mathbb{B}}(x;r)$	The closed ball of radius $r > 0$, centered at x
$\lvert x \rvert$	The Hilbertian norm
X^*	The dual space of X
$\langle x, y \rangle$	The canonical inner product
$\bigoplus$	The direct sum
$\nabla F(x)$	The Jacobian of the mapping $F : \mathbb{R}^n \to \mathbb{R}^m$
$\nabla F(x)^*$	The adjoint of $\nabla F(x)$
$\mathcal{C}^1$	The class of continuously Fréchet differentiable mappings
$\mathcal{C}^{1+}$	The class of $\mathcal{C}^1$ mappings with locally Lipschitz derivative
$I_{n\times n}$	The identity matrix of size n
$C \setminus D$	The complement of D relative to C
$\operatorname{dom} f$	The effective domain of f
$\operatorname{epi} f$	The epigraph of f
$\operatorname{gph} \Gamma$	The graph of Γ

$\partial_p f$	The proximal subdifferential of the function f
∂f	The limiting (proximal) subdifferential of the function f
$\partial^\infty f$	The limiting singular subdifferential of the function f
e_λ	The Moreau envelope
P_λ	The proximal mapping
$\xrightarrow{w}$	The weak convergence
$\xrightarrow{f}$	The f-attentive convergence
$\xrightarrow{e}$	The epigraphical convergence
$\xrightarrow{m}$	The Mosco epi-convergence
$\xrightarrow{pk}$	The Painlevé-Kuratowski convergence
$\operatorname{argmin}_{x\in C} f(x)$	The set of minimizers for f over C
$\operatorname{argmax}_{x\in C} f(x)$	The set of maximizers for f over C
a.e.	almost everywhere
l.s.c.	lower semicontinuous
w.r.t.	with respect to

CHAPTER 1

INTRODUCTION

1.1. Background and Motivation

Nonsmooth analysis is one of the most attractive and promising areas in modern mathematics. A systematic study of local behavior of nondifferentiable (not necessarily differentiable) functions and set-valued mappings (multifunctions) is accomplished in such a framework. In recent years, it has grown rapidly in connection with the study of problems of functional analysis, optimization, optimal design, mechanics and plasticity, differential equations, and control theory. Recently, Terry Rockafellar, a pioneer in this area, has given a more appropriate title, *variational analysis*, to reflect this breadth (cf. [**46**]).

It is well known that the subgradients of convex functions have very favorable properties, and have been the basic impetus to develop more general theory of nonsmooth analysis. Evidently, identifying nonconvex functions with properties that closely resemble the properties of convex functions is advantageous for the possible development of both the subgradient theory and computation.

In this thesis, we focus on one such class of functions; namely *prox-regular* functions in Hilbert space. These functions were first introduced in 1996, by Poliquin and Rockafellar in [**29**], and thoroughly investigated in [**29**] and [**30**].

However, their analysis is confined to finite-dimensional spaces and does not deal with important issues like integration of subdifferentials and the calculus rules of prox-regular functions. Our further investigation not only deals with such issues but extends most of the important properties in [**29**] to an arbitrary Hilbert space.

First, we introduce the prox-regular functions and discuss the key facts developed in [**29**]. A concept that is essential in defining prox-regular functions is that of the *proximal subgradient.*

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ (i.e., extended-real-valued function) and $\bar{x}$ be a point where f is finite. A vector $\bar{v}$ in $\mathbb{R}^n$ is said to be a proximal subgradient of f at $\bar{x}$ provided that there exist scalars $\varepsilon > 0$ and $r > 0$ such that

$$f(x) \geq f(\bar{x}) + \langle \bar{v}, x - \bar{x} \rangle - \frac{r}{2}|x - \bar{x}|^2 \quad \text{for all} \quad x \in \mathbb{B}(\bar{x}; \varepsilon),$$

where $\mathbb{B}(\bar{x}; \varepsilon)$ is the open ball of radius $\varepsilon > 0$, centered at $\bar{x}$. The set of such $\bar{v}$, if any, is denoted $\partial_p f(\bar{x})$ and is referred to as the *proximal subdifferential.*

A limiting form of $\partial_p f$ is defined by

$$\partial f(\bar{x}) := \{\ \lim v_k \ :\ v_k \in \partial_p f(x_k),\ x_k \to \bar{x} \ \text{ with } \ f(x_k) \to f(\bar{x})\ \},$$

which is referred to as the *limiting (proximal) subdifferential.*

Another useful limiting subdifferential is defined by

$$\partial^\infty f(\bar{x}) := \{\lim t_k v_k \ :\ t_k \searrow 0,\ v_k \in \partial_p f(x_k),\ x_k \to \bar{x} \text{ with } f(x_k) \to f(\bar{x})\},$$

which is referred to as the *limiting singular subdifferential.*

A lower semicontinuous (l.s.c.) function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be *prox-regular* at $\bar{x}$, a point where f is finite, for the subgradient $\bar{v} \in \partial f(\bar{x})$, if there exist parameters $\varepsilon > 0$ and $r \geq 0$ such that for every point $(x, v) \in \operatorname{gph} \partial f$ obeying $|x - \bar{x}| < \varepsilon$,

$|f(x) - f(\bar{x})| < \varepsilon$, and $|v - \bar{v}| < \varepsilon$, one has the local estimate

$$f(x') \geq f(x) + \langle v, x' - x \rangle - \frac{r}{2}|x' - x|^2 \quad \text{for all} \quad x' \in \mathbb{B}(\bar{x}; \varepsilon).$$

When this holds for all $\bar{v} \in \partial f(\bar{x})$, f is said to be prox-regular at $\bar{x}$.

The class of prox-regular functions can be described as a very broad class of nonsmooth functions of interest in variational analysis and optimization, which admits effective generalizations of many of the subdifferential properties of extended-valued convex functions.

Now we summarize the key facts of prox-regular functions developed in [**29**].

(a) Subgradient characterization of prox-regularity

The following subgradient characterization of prox-regularity, established in [**29**], deserves special attention in several respects.

A l.s.c. function f is prox-regular at $\bar{x}$ for $\bar{v}$ if and only if $\bar{v}$ is a proximal subgradient of f at $\bar{x}$ and, under a suitable localization (f-attentive), the multifunction $\partial f + rI$ is monotone.

In most cases, this subgradient characterization can be used as a handy tool to test the prox-regularity of a function. For example, all $\mathcal{C}^{1+}$ functions (differentiable with locally Lipschitz gradient), all l.s.c. proper convex functions, all lower-$\mathcal{C}^2$ functions (locally the sum of the function and a positive multiple of the norm square is convex), and all primal-lower-nice functions (see Definition 2.2.1) are in turn prox-regular too.

As it was pointed out in [**29**], the above "pre-monotonicity" property of ∂f is sufficient for a full range of desirable subdifferentiable properties. For example, when f is prox-regular at $\bar{x}$ for $\bar{v}$, the graph of ∂f coincides, under a suitable localization near $(\bar{x}, \bar{v})$, with a *Lipschitz manifold* of dimension n in $\mathbb{R}^n \times \mathbb{R}^n$, a property previously detected only for convex function and their very close allies [**39**]. Further, it plays a key role in establishing the smoothness and convexity properties of envelope functions of a prox-regular function.

(b) Regularity and convexity properties of envelope functions

For a proper, l.s.c. function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and parameter $\lambda > 0$, the *Moreau envelope* function is defined by

$$e_\lambda(x) := \inf_{x'} \{ f(x') + \frac{1}{2\lambda}|x' - x|^2 \}.$$

These functions not only approximate but provide a kind of regularization of f. For a l.s.c function f (may take ∞ values and exhibit discontinuities) minorized by some quadratic function, it is known that, for λ small enough, e_λ is finite and locally Lipschitz continuous, and approximates f in the sense that e_λ increases pointwise to f as $\lambda \searrow 0$ (see the book by Attouch [**1**]).

As a companion to the envelope function e_λ we have the *proximal mapping* $P_\lambda : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined by

$$P_\lambda(x) := \operatorname*{argmin}_{x'} \{ f(x') + \frac{1}{2\lambda}|x' - x|^2 \},$$

that relate to the numerical techniques like the proximal point algorithm in the minimization of f.

For a prox-regular function, a strong connection between the function and its envelope functions and the proximal mappings was established in [**29**]:

If f is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v}$ (see Definition 2.2.3), then for any $\lambda \in (0, 1/r)$, where r is a parameter in the definition of prox-regularity, there is a convex neighborhood V of $\bar{x}$ such that

- the mapping P_λ is single-valued and Lipschitz continuous on V with $P_\lambda = (I + \lambda T)^{-1}$, where T is a localization of ∂f around $(\bar{x}, \bar{v})$.

- the function e_λ is a differentiable function with locally Lipschitz gradient ($\mathcal{C}^{1+}$) and lower-$\mathcal{C}^2$ on V with

$$e_\lambda + \frac{r}{2(1-\lambda r)}|\cdot|^2 \ \text{ convex } \ , \ \nabla e_\lambda = \frac{1}{\lambda}[I - P_\lambda] = [\lambda I + T^{-1}]^{-1}.$$

These are very important findings of prox-regular functions not only from a variational analysis point of view but for the possible development of numerical methods for minimizing e_λ, which in effect would open a new approach to minimizing f despite its nonsmoothness. In fact, these are the properties one would expect only of convex functions and alike. For instance, when f is convex e_λ is convex too, and actually of class $\mathcal{C}^{1+}$, and the above formulas hold. Moreover, the proximal mapping P_λ can be used not only to parameterize the graph of ∂f but in connection with convex minimization algorithms such as the proximal point algorithm, see [**36**].

(c) Second-order Theory

In addition to the desirable functional and subdifferentiable properties outlined above, prox-regular functions have particularly satisfactory second-order behaviour in $\mathbb{R}^n$. In [**29**], a perfect equivalence between second-order epi-differentiability of f at $\bar{x}$ for $\bar{v}$ and the proto-differentiability of a suitable localization of ∂f at $(\bar{x}, \bar{v})$ was established with a natural formula relating these two derivatives. This generalizes the classical idea of obtaining second derivatives by differentiating first derivatives, which was previously known only for convex functions and strongly amenable functions; see [**23**], [**25**] and [**42**].

Moreover, the additional hypothesis that the second-order epi-derivative function $f''_{\bar{x},\bar{v}}$ is finite on a neighborhood of the origin suffices to establish the second-order expansion (possibly with a nonquadratic second-order term) formula

$$f(x) = f(\bar{x}) + \langle \bar{v}, x - \bar{x}\rangle + f''_{\bar{x},\bar{v}}(x - \bar{x}) + o|x - \bar{x}|^2.$$

1.2. Hilbert Space Extensions and New Issues

Here, we state and discuss the principle results of our investigation. In the first part of the thesis, we extend the results $(a), (b)$ and (c) to Hilbert spaces. In the remaining part, we focus on two new issues of prox-regular functions.

(A) Extension of subgradient characterization

We begin by extending the subgradient characterization of prox-regularity, described in (a), to an arbitrary Hilbert space. This extension enables us not only to enhance the territory of the prox-regular class but to obtain many desirable subdifferential properties, including the Lipschitz manifold property of the graph of ∂f. The *smooth variational principle* is used as a basic tool in establishing this result.

(B) Extension of regularity and convexity properties

We prove that all the results stated under (b) are true in an arbitrary Hilbert space setting, including the C^{1+} smoothness (Fréchet sense) and the convexity properties of envelope functions. This clearly allows, as mentioned under (b), in the possible development of subgradient theory and computation to tackle some Hilbert space problems as well. We give one particular example, in the theory of partial differential equations, to highlight this point.

In [**47**], Strömberg studied the following Cauchy problem:

$$\frac{\partial}{\partial t}u(x,t) + \frac{1}{2}\left|\frac{\partial}{\partial x}u(x,t)\right|^2 = 0 \quad x \in X, t > 0,$$
$$u(x,0) = f(x) \quad x \in X,$$

where X is an arbitrary Hilbert space.

He proved ([**47**], Proposition 3) that when X is an arbitrary Hilbert space and $f + \frac{1}{2T}|\cdot|^2$ is convex, where $T > 0$, (i.e., f is lower-C^2 everywhere and hence prox-regular everywhere too) then

$$u(x,t) = e_t(x) = \inf_{x'} \{ f(x') + \frac{1}{2t}|x' - x|^2 \}$$

is a solution to the above Hamilton-Jacobi equation at each point (x, t) in $X \times (0, T)$..

Now we can say more here:

For a prox-regular function f (with parameters ε and r) at $\bar{x}$ for $\bar{v}$, we know that $e_t(x)$ is $\mathcal{C}^{1+}$ around $\bar{x}$ for small enough t, and hence, there is a neighborhood V of $\bar{x}$ such that $u(x, t) = e_t(x)$ is a local solution to the above Cauchy problem at every point (x, t) in $V \times (0, 1/r)$. Note here that f may allow infinite values and exhibit discontinuities, and hence there is much flexibility for setting up the initial condition for u.

In addition to the lower-$\mathcal{C}^2$ property of e_λ, conditions were given in **[29]** under which e_λ itself is convex or strongly convex. We also find extensions to these results in separable Hilbert spaces. In achieving these results, an extended version of Rademacher's theorem, a concept of null sets in Banach spaces, and a criterion for integrability of Banach space valued functions are employed as well.

(C) Extension of second-order theory

The extension of the second-order theory of prox-regular functions to Hilbert spaces is not that promising. We establish the following partial extension for the generalized second-order differentiation:

Let f be prox-regular at $\bar{x}$ for $\bar{v}$. If f is twice Mosco epi-differentiable at $\bar{x}$ for $\bar{v}$, then a suitable localization of the subgradient mapping ∂f is proto-differentiable, and the natural derivative formula holds. The extended results of smoothness and convexity properties of e_λ play an important role in achieving this result. A

Hilbert space example is given to show that the second-order expansion of a prox-regular function fails to exist even for a convex function with finite second-order epi-derivative everywhere.

(D) Integration of prox-regular functions

A fundamental problem in nonsmooth analysis is to identify functions that can be recovered up to an additive constant, from the knowledge of their subgradients. More precisely, a function f is deemed integrable if whenever $\partial_{\#}g(x) = \partial_{\#}f(x)$ for all x then f and g differ only by an additive constant. Here $\partial_{\#}$ refers to a subdifferential which can be taken in many different ways (eg. Clarke subdifferential, Mordukhovich subdifferential, Fréchet subdifferential, Ioffe approximate subdifferential, proximal and limiting subdifferential).

Probably the most well known and the oldest result in this area is that the convex functions are integrable (in the above sense) even in a Banach Space; see [**33**]. However, very few other examples were known. For convex functions all types of known subdifferentials are reduced to the subdifferential in convex analysis, but in nonconvex cases the type of the subdifferential used plays a key role. The proximal subdifferential has been successful in identifying some nonconvex functions (up to an additive constant). This was done by Poliquin [**24**] for the p.l.n. functions defined on $I\!R^n$, and later extended to Hilbert spaces by Thibault and Zagrodny [**48**].

The contribution we make to the integration problem is to identify a large class of prox-regular functions that can be recovered from the knowledge of their limiting (proximal) subgradients. More precisely, we prove in an arbitrary Hilbert space that if two functions, which have the same limiting subgradients locally, are prox-regular and subdifferentially continuous relative to a pair $(\bar{x}, \bar{v})$ then the functions differ by a constant in a local neighborhood of $(\bar{x}, \bar{v})$. We also construct an example to show that our integration result covers a much broader class of functions than that of the p.l.n. case [**24**]. The central tool that we employ here is the smoothness property of the envelopes of prox-regular functions.

(E) Calculus of prox-regular functions

In [**29**], a large core of examples were given to show the magnitude and applicability of the prox-regular class. However, the lack of calculus rules has been a hindrance to the constructive development of this class. We overcome this difficulty by developing basic calculus rules for prox-regular functions. A master key to our calculus is the following chain rule.

Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be $\mathcal{C}^{1+}$ near $\bar{x}$ and $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ with $g(F(\bar{x}))$ finite and a natural constraint qualification is satisfied at $F(\bar{x})$. We prove that for a fixed $\bar{v} \in \partial(g \circ F)(\bar{x})$, if g is prox-regular at $F(\bar{x})$ for all $y \in g(F(\bar{x}))$ with $\nabla F(\bar{x})^* y = \bar{v}$, then the composite function $g \circ F$ is prox-regular at $\bar{x}$ for $\bar{v}$. Here $\nabla F(\bar{x})^*$ denotes the adjoint of the Jacobian matrix ∇F at $\bar{x}$.

As an easy application of the above chain rule we have the following sum rule: Let $f = f_1 + f_2$, $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $\bar{x} \in \operatorname{dom} f$ and the only vector $y_i \in \partial^\infty f_i(\bar{x})$ with $y_1 + y_2 = 0$ is $y_1 = y_2 = 0$, where $\partial^\infty f$ denotes the limiting singular subdifferential of f. Assume also that f_i are prox-regular for all $v_i \in \partial f_i(\bar{x})$ such that $v_1 + v_2 = \bar{v}$. Then f is prox-regular at $\bar{x}$ for $\bar{v}$.

Another consequence of our chain rule is the identification of new examples of integrable functions (in the sense of (D)) on $\mathbb{R}^n$.

This thesis is organized as follows. In **Chapter 2**, we extend the main results of prox-regular functions in $\mathbb{R}^n$ to Hilbert spaces. That includes a subgradient characterization of prox-regular functions, regularity and convexity properties of its envelope functions and some second-order properties. In **Chapter 3** we present an integration result, and in **Chapter 4** we give the calculus rules of prox-regular functions with some of their consequences.

1.3. Notation

The terminology and notation we adopt here is the standard one of *convex* and *variational* analysis (cf. [**14**], [**35**], [**46**]). We'll be working in a real Hilbert space X with norm $|\cdot|$. The *open unit ball* in X is denoted by $I\!B$, its closure by $\overline{I\!B}$. The open ball of radius $r > 0$, centered at x, is denoted by $I\!B(x;r)$, and its closure by $\overline{I\!B}(x;r)$.

A quite useful convention in optimization theory, which we'll also adopt, is to allow functions to be *extended-real-valued*, i.e. to take values in $\overline{I\!R} = [-\infty, \infty]$. We employ extended arithmetic with the convention (oriented toward minimization)

$$\infty + (-\infty) = (-\infty) + \infty = \infty, \quad 0 \cdot \infty = 0 = 0 \cdot (-\infty).$$

The extended-real line $\overline{I\!R}$ has all the properties of a compact interval. Every subset $R \subset I\!R$ has a *supremum* (least upper bound) in $\overline{I\!R}$, which is denoted by $\sup R$, and likewise an *infimum* (greatest lower bound), $\inf R$.

For an extended-real-valued function f on a set C, we also introduce notions for the sets of points x where the minimum or maximum of f over C is regarded as being attained :

$$\operatorname*{argmin}_C f := \operatorname*{argmin}_{x\in C} f(x) := \begin{cases} \{x \in C | f(x) = \inf_C f\} & \text{if } \inf_C f \neq \infty, \\ \emptyset & \text{if } \inf_C f = \infty, \end{cases}$$

$$\operatorname*{argmax}_C f := \operatorname*{argmax}_{x\in C} f(x) := \begin{cases} \{x \in C | f(x) = \sup_C f\} & \text{if } \sup_C f \neq -\infty, \\ \emptyset & \text{if } \sup_C f = -\infty. \end{cases}$$

For a function $f : X \to \overline{I\!R}$ we define the following:

The *effective domain* of f is denoted by

$$\operatorname{dom} f := \{x \in X \mid f(x) < +\infty\}$$

and its *epigraph*

$$\operatorname{epi} f := \{(x, \alpha) \in X \times I\!R \mid f(x) \leq \alpha\}.$$

We call f a *proper* function if $f(x) < \infty$ for at least one $x \in X$, and $f(x) > -\infty$ for all $x \in X$, or in otherwords, if $\operatorname{dom} f$ is a nonempty set on which f is finite; otherwise it is *improper*.

The function $f : X \to \overline{I\!R}$ is *lower semicontinuous* (l.s.c.) at $\bar{x}$ if

$$\liminf_{x \to \bar{x}} f(x) \geq f(\bar{x})$$

and lower semicontinuous on X if this holds for every $\bar{x} \in X$. The l.s.c. of f at $\bar{x}$ is clearly equivalent to saying that for all $\varepsilon > 0$, there exists $\delta > 0$ so that $y \in I\!B(\bar{x}; \delta)$ implies $f(y) \geq f(\bar{x}) - \varepsilon$.

Let S be a subset of X. The *indicator function* of S, denoted by $I_S(\cdot)$, is the extended-valued function defined by

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

The *inner product* of v and x is denoted $\langle v, x \rangle$, a notation which is also employed when X is a Banach space for the evaluation, at $x \in X$, of the linear functional $v \in X^*$ (the space of continuous linear functionals defined on X).

The notation $x = w\text{-}\lim_{k \to \infty} x_k$ or $x_k \xrightarrow{w} x$ means that the sequence $\{x_k\}$ converges weakly to x in X.

Let Y be another Hilbert space. A *set-valued map* (multifunction) T from X to Y, written as $T : X \rightrightarrows Y$, is characterized by its graph, $\operatorname{gph} T$, the subset of the product space $X \times Y$ defined by

$$\operatorname{gph} T := \{(x, y) \in X \times Y \mid y \in T(x)\}.$$

The *domain* and *range* of $T : X \rightrightarrows Y$ are taken to be the sets

$$\operatorname{dom} T := \{\, x \mid T(x) \neq \emptyset \,\}, \qquad \operatorname{rge} T := \{\, y \mid \exists x \text{ with } y \in T(x) \,\}.$$

The *inverse* T^{-1} of T is the set-valued map from Y to X, defined by

$$x \in T^{-1}(y) \iff y \in T(x) \iff (x, y) \in \operatorname{gph} T.$$

CHAPTER 2

PROX-REGULAR FUNCTIONS IN HILBERT SPACES

Analysis of prox-regular functions is based on proximal analysis in Hilbert space. For this reason, in section 2.1, we review the basic concepts in proximal analysis. In section 2.2 we define the prox-regularity of a function in Hilbert space along with the subdifferential continuity. Section 2.3 establishes the subgradient characterization of prox-regularity. In section 2.4 we obtain the regularity properties ($\mathcal{C}^{1+}$ smoothness) of Moreau envelopes of a prox-regular function. We also identify a localization of gph ∂f of a prox-regular function f with a Lipschitz manifold in $X \times X$. Section 2.5 deals with the convexity properties of Moreau envelopes. We show that for a prox-regular function the Moreau envelope function is lower-$\mathcal{C}^2$ (i.e., locally the sum of the function and a positive multiple of the norm square is convex). Further the conditions are given in separable Hilbert space setting, under which e_λ itself is convex or strongly convex. In section 2.6 we give a partial extension to the second-order property [**29**], Theorem 6.1. We prove in a Hilbert space that when f is prox-regular and twice Mosco epi-differentiable at $\bar{x}$ for $\bar{v}$ then a localization of the subgradient mapping ∂f is proto-differentiable at $\bar{x}$ for $\bar{v}$, with a natural formula relating these two derivatives. A Hilbert space example is given to show that the second-order expansion of a prox-regular function fails to exist even for a convex function with finite second-order Mosco epi-derivative everywhere.

2.1. Proximal Analysis

The *proximal subgradient*, a generalized notion of classical derivative, turns out to be a powerful tool in characterizing a variety of functional properties in *nonsmooth* analysis. A powerful body of theory of proximal subgradients and their counterparts, the *proximal analysis*, is now available. We refer the interested reader to the recent book of Clarke, Ledyaev, Stern and Wolenski **[14]** for a coherent and comprehensive exposition of proximal analysis.

First, we gather a basic tool kit from proximal analysis for our task ahead. Recall that X denotes a real Hilbert space and $\overline{\mathbb{R}}$ represents the extended real line.

Definition 2.1.1. *(proximal subgradients) Let $f : X \to \overline{\mathbb{R}}$ and $\bar{x}$ be a point where f is finite. A vector $\bar{v}$ is a proximal subgradient of f at $\bar{x}$, if there exist $\varepsilon > 0$ and $r > 0$ such that*

$$f(x) \geq f(\bar{x}) + \langle \bar{v}, x - \bar{x} \rangle - \frac{r}{2}|x - \bar{x}|^2 \quad \text{for all} \quad x \in \mathbb{B}(\bar{x}; \varepsilon),$$

where $\mathbb{B}(\bar{x}; \varepsilon)$ denotes the open ball of radius $\varepsilon > 0$, centered at $\bar{x}$.
The set of all such $\bar{v}$ is denoted by $\partial_p f(\bar{x})$, and is referred to as the proximal subdifferential.

The existence of a proximal subgradient $\bar{v}$ at $\bar{x}$ thus corresponds to the existence of a "local quadratic support" to f at $\bar{x}$. This means the possibility of approximating f from below (thus in a one-sided manner) by a function whose graph is a parabola. The point $(\bar{x}, f(\bar{x}))$ is a contact point between the graph of f and the parabola, and $\bar{v}$ is the slope of the parabola at that point. Compare this with the usual derivative, in which the graph of f is approximated by an affine function.

It follows immediately from the definition that the proximal subdifferential, $\partial_p f(\bar{x})$, is convex, however it is not necessarily open, closed, or nonempty.

Example 2.1.2. *Let $f : \mathbb{R} \to \mathbb{R}$. One can easily verify that the following:*

$$(a) \quad f(x) = -|x|, \qquad \partial_p f(x) = \begin{cases} -\frac{x}{|x|} & \text{if } x \neq 0, \\ \emptyset & \text{if } x = 0. \end{cases}$$

$$(b) \quad f(x) = -|x|^{3/2}, \qquad \partial_p f(x) = \begin{cases} -\frac{3}{2}\frac{x}{|x|^{1/2}} & \text{if } x \neq 0, \\ \emptyset & \text{if } x = 0. \end{cases}$$

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

Note that in (b), f is a differentiable (C^1) function but it has no proximal subgradients at $x = 0$, and in (c), the subdifferential set $\partial_p f(x)$ at $x = 0$ is not open or closed.

Before developing further properties of proximal subgradients, we need to recall some facts about classical derivatives.

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

$$F'(x, v) := \lim_{t \downarrow 0} \frac{F(x + tv) - F(x)}{t}$$

when this limit exits. F is said to admit a *Gâteaux derivative* at x, an element in the space $\mathcal{L}(X, Y)$ of continuous linear operators from X to Y denoted $DF(x)$, provided that for every v in X, $F'(x, v)$ exists and equals $DF(x)v$. This is equivalent to saying that the difference quotient converges for each v, that one has

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

and that the convergence is uniform with respect to v in finite sets (the last condition is automatically true). If the word "finite" in the preceding sentence is replaced by "compact", the derivative is known as *Hadamard*; for "bounded" we obtain the *Fréchet* derivative. When $X = \mathbb{R}^n$, Hadamard and Fréchet differentiability are equivalent; when F is Lipschitz near x, then Hadamard and Gâteaux differentiabilities coincide.

It turns out that the differential concept most naturally linked to the theory of limiting subgradients is that of *strict differentiability* (cf. [**12**], proposition 2.2.4). We shall say that F admits a strict derivative at x, an element of $\mathcal{L}(X, Y)$ denoted $D_sF(x)$, provided that for each v, the following holds:

$$\lim_{\substack{x' \to x \\ t \downarrow 0}} \frac{F(x' + tv) - F(x')}{t} = D_sF(x)v$$

and provided the convergence is uniform for v in compact sets. (This last condition is automatic if F is Lipschitz near x).

The first proposition relates $\partial_p f$ to classical differentiability. Recall that a function f is said to be *proper* if $f(x) < \infty$ for at least one $x \in X$, and $f(x) > -\infty$ for all $x \in X$.

Proposition 2.1.3. *Let $f : X \to \overline{\mathbb{R}}$ be l.s.c., proper and $U \subset X$ be open.*

(a) Assume that f is Gâteaux differentiable at $x \in U$. Then $\partial_p f(x) \subseteq \{Df(x)\}$.

(b) If $f \in C^2(U)$, then $\partial_p f(x) = \{Df(x)\}$ for all $x \in U$.

(c) If f is convex, then $v \in \partial_p f(x)$ if and only if

$$f(y) \geq f(x) + \langle v, y - x\rangle \quad \text{for all} \quad y \in X. \tag{2.1.1}$$

In other words, when f is convex $\partial_p f(x)$ coincides with the subdifferential of convex analysis (the set of vectors v satisfying 2.1.1).

Proof. See Clarke, Ledyave, Stern and Wolenski [**14**], Corollary 2.6. □

One of the primary aims of subgradient theory is the analysis of optimality. The classical *rule of Fermat* states that a function's derivative must vanish at a local minimum. This rule has the following extension to our nonsmooth setting.

Proposition 2.1.4. *(Fermat's rule generalized) Let $f : X \to \overline{\mathbb{R}}$ be l.s.c. and proper.*

(a) If f has a local minimum at $\bar{x}$, then $0 \in \partial_p f(\bar{x})$.

(b) Conversely, if f is convex and $0 \in \partial_p f(\bar{x})$, then $\bar{x}$ is a global minimum of f.

Proof.

(a) The definition of a local minimum says that there exists $\varepsilon > 0$ so that $f(x) \geq f(\bar{x})$ for all $x \in \mathbb{B}(\bar{x};\varepsilon)$, which satisfies the definition 2.1.1 with $\bar{v} = 0$ and $r = 0$, and hence $0 \in \partial_p f(\bar{x})$.

(b) Under the hypothesis, (2.1.1) holds with $v = 0$. Thus $f(x) \geq f(\bar{x})$ for all $x \in X$, which says that $\bar{x}$ is a global minimum of f. □

Nonsmooth calculus has been developed in varying degrees of generality. The price to pay for the greatest generality is heavy in terms of technicality. Here, we will not survey proximal calculus extensively, however for our purposes we record the basic sum rule. First note that we cannot expect a calculus sum rule of the form

$$\partial_p f(x) + \partial_p g(x) \;=\; \partial_p(f+g)(x)$$

to hold in much generality. The inclusion $\partial_p f(x) + \partial_p g(x) \subseteq \partial_p(f+g)(x)$ can be established easily, but unfortunately, it is nearly useless. To see that the reverse inclusion is not always true, simply take $f(x) = |x|$ and $g(x) = -|x|$ and compare the subgradient sets at 0.

We observe that the sum rule just mentioned is trivial if one summand is C^2.

Proposition 2.1.5. *Let $f : X \to \overline{I\!R}$ be l.s.c., proper, and let $\bar{x} \in X$ where f is finite. Suppose further that g is C^2 in a neighborhood of $\bar{x}$. Then*

$$\partial_p(f+g)(\bar{x}) = \partial_p f(\bar{x}) + Dg(\bar{x}).$$

Proof. Use the fact that for a C^2 function g the defining functional inequality of proximal subgradient can be applied to both g and $-g$ with their gradients. □

Even though the exact sum rule fails in general, the following result known as "fuzzy sum rule" holds in surprising generality.

Theorem 2.1.6. *(fuzzy sum rule) Let $x_0 \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$, and let $v \in \partial_p(f_1 + f_2)(x_0)$. Suppose that either:*

(a) f_1 and f_2 are weakly lower semicontinuous (automatically the case if X is finite dimensional); or

(b) one of the functions is Lipschitz near x_0.

Then, for any $\varepsilon > 0$, there exist (for $i = 1, 2$) points $x_i \in I\!B(x_0; \varepsilon)$ with $|f_i(x_0) - f_i(x_i)| < \varepsilon$ such that

$$v \in \partial_p f_1(x_1) + \partial_p f_2(x_2) + \varepsilon I\!B,$$

where $I\!B$ is the open unit ball.

Proof. See Clarke, Ledyave, Stern and Wolenski [**14**], Theorem 8.3. □

We now record another important fact about proximal subgradients in Hilbert spaces; the set $\operatorname{dom}(\partial_p f)$ of points in $\operatorname{dom} f$ at which at least one proximal subgradient exists is dense in $\operatorname{dom} f$.

Theorem 2.1.7. *(density theorem) Let $f : X \to \overline{I\!R}$ be l.s.c., proper, and bounded below. Then the following set is a dense subset of $\operatorname{gph} f$:*

$$S := \{(x, f(x));\, x \in \operatorname{dom} f \quad\text{and}\quad \partial_p f(x) \neq \emptyset\}$$

In particular, $\operatorname{dom}(\partial_p f)$ *is dense in* $\operatorname{dom} f$.

Proof. See Clarke, Ledyave, Stern and Wolenski [**14**], Theorem 3.1. □

One of the drawbacks of proximal subgradients is that the set $\partial_p f(x)$ would seem potentially empty for many x, and that leads to poor calculus. A remedy comes through perturbing the base point and leads one to define limiting subgradients.

With respect to any function $f : X \to \overline{I\!R}$, we'll say that a sequence of points x_k in X *converges in the f-attentive sense* to $\bar{x}$, written $x_k \xrightarrow{f} \bar{x}$, when not only $x_k \to \bar{x}$ but $f(x_k) \to f(\bar{x})$ (cf. [**46**]):

$$x_k \xrightarrow{f} \bar{x} \iff x_k \to \bar{x} \text{ with } f(x_k) \to f(\bar{x}).$$

Of course, f-attentive convergence is the same as ordinary convergence of x to $\bar{x}$ wherever f is continuous.

Definition 2.1.8. *(limiting subgradients) Let $f : X \to \overline{I\!R}$ and $\bar{x}$ a point where f finite.*

(a) A vector $\bar{v}$ is a limiting (proximal) subgradient of f at $\bar{x}$, if for some sequence v_k such that $v_k \in \partial_p f(x_k)$ and $x_k \xrightarrow{f} \bar{x}$ one has $\bar{v} = w\text{-}\lim_{k\to\infty} v_k$.

That is, we consider the set of all vectors $\bar{v}$ that can be expressed as the weak limit (which is what "w-lim" signifies) of some sequence $\{v_k\}$, where $v_k \in \partial_p f(x_k)$ for each k, and where $x_k \to \bar{x}$, $f(x_k) \to f(\bar{x})$. The set of all such $\bar{v}$ is denoted by $\partial f(\bar{x})$, and is referred to as the limiting proximal subdifferential.

(b) A vector $\bar{v}$ is a limiting singular subgradient of f at $\bar{x}$, if for some sequences $t_k \searrow 0$, v_k such that $v_k \in \partial_p f(x_k)$ and $x_k \xrightarrow{f} \bar{x}$ one has $\bar{v} = w\text{-}\lim_{k\to\infty} t_k v_k$.

The set of all such $\bar{v}$ is denoted by $\partial^\infty f(\bar{x})$, and is referred to as the limiting singular subdifferential.

The limiting subdifferential, $\partial f(x)$, which contains $\partial_p f(x)$ is not necessarily open, convex or nonempty but it is sequentially weakly closed. Moreover, when $X = I\!R^n$, $\partial f(x)$ is a closed set, and if f is Lipschitz near x, then $\partial f(x) \neq \emptyset$.

Example 2.1.9. *Let* $f : I\!R \to I\!R$.

(a) $f(x) = -|x|, \qquad \partial f(0) = \{-1, +1\}, \quad \partial^\infty f(0) = \{0\}$.

(b) $f(x) = -\sqrt{|x|}, \quad \partial f(0) = \emptyset, \quad \partial^\infty f(0) = (-\infty, \infty)$.

Although a type of closure operation was used in defining $\partial f(\bar{x})$, it is a fact that this set may fail to be closed when X is infinite dimensional or when f fails to be Lipschitz. These facts make the limiting calculus most appealing in the presence of Lipschitz hypothesis or in finite dimensions. Here is a sharper form of (nonfuzzy) sum rule.

Proposition 2.1.10. *(sum rule) If one of* f_1, f_2 *is Lipschitz near* $x \in X$, *then*

$$\partial(f_1 + f_2)(x) \subseteq \partial f_1(x) + \partial f_2(x).$$

Proof. See Clarke, Ledyave, Stern and Wolenski [**14**], Proposition 10.1. ◻

In parallel with proximal subdifferentials, we might be led to believe that $\partial f_1(x) + \partial f_2(x) \subseteq \partial(f_1 + f_2)(x)$ and hence the equality holds in proposition 2.1.10. However, this is not the case, as seen by the following example.

Example 2.1.11. *Let* $f_i : I\!R \to I\!R$, $i = 1, 2$ *defined by* $f_i(x) := (-1)^i |x|$. *Then* $\partial f_1(0) = \{-1, 1\}$, $\partial f_2(0) = [-1, 1]$ *and* $\partial(f_1 + f_2)(0) = \{0\}$. *Hence* $\partial f_1(0) + \partial f_2(0) \not\subset \partial(f_1 + f_2)(0)$.

However, there are supplementary hypothesis (such as regularity) under which equality does hold in Proposition 2.1.10 (cf. [**14**], Chapter 2).

Finally, we present a limiting form of Chain Rule and Sum Rule in finite dimensional context, which will be useful in Chapter 4.

Theorem 2.1.12. *(chain rule) Let $f(x) := g(F(x))$, where $F : \mathbb{R}^n \to \mathbb{R}^m$ is C^1 on some neighborhood of $\bar{x}$, while $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is l.s.c., proper with $F(\bar{x})$ in* $\mathrm{dom}(f)$. *Assume further that the only vector $y \in \partial^\infty g(F(\bar{x}))$ with $\nabla F(\bar{x})^* y = 0$ is $y = 0$, where the Jacobian matrix for F at $\bar{x}$ is denoted by $\nabla F(\bar{x})$, and its adjoint by $\nabla F(\bar{x})^*$. Then*

$$\partial f(\bar{x}) \subseteq \nabla F(\bar{x})^* \partial g(F(\bar{x})).$$

Proof. See Rockafellar and Wets [**46**], Theorem 10.6. □

As an easy application of the chain rule we have the following sum rule.

Corollary 2.1.13. *(sum rule) Suppose $f = f_1 + \cdots + f_m$ for proper, l.s.c. functions $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ and let $\bar{x} \in \mathrm{dom}\, f$. Assume also that the only combination of vectors $v_i \in \partial^\infty f_i(\bar{x})$ with $v_1 + \cdots + v_m = 0$ is $v_1 = \cdots = v_m = 0$. Then*

$$\partial f(\bar{x}) \subseteq \partial f_1(\bar{x}) + \cdots + \partial f_m(\bar{x})$$

$$\partial^\infty f(\bar{x}) \subseteq \partial^\infty f_1(\bar{x}) + \cdots + \partial^\infty f_m(\bar{x}).$$

Proof. See [**46**] Corollary 10.9. □

For a separable function it is easy to verify the following subgradient formula.

Proposition 2.1.14. *Let $f(x) = f_1(x_1) + \cdots + f_m(x_m)$ for l.s.c. functions $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, where $x \in \mathbb{R}^n$ is expressed as $(x_1, \ldots, x_m)$ with $x_i \in \mathbb{R}^{n_i}$ and $n_1 + \cdots + n_m = n$. Then at any point $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_m)$ where f is finite one has*

$$\partial f(\bar{x}) = \partial f_1(\bar{x}_1) \times \cdots \times \partial f_m(\bar{x}_m)$$

$$\partial^\infty f(\bar{x}) \subseteq \partial^\infty f_1(\bar{x}_1) \times \cdots \times \partial^\infty f_m(\bar{x}_m).$$

Proof. See [**46**] Proposition 10.5. □

The limiting proximal subgradients play a key role in defining prox-regular functions and the development throughout the thesis.

2.2. Prox-regular Functions

Prox-regular functions emerge as a generalization to the primal-lower-nice functions (p.l.n.), earlier introduced by Poliquin [**24**], in connection with recovering a function from its subgradient mapping. First we introduce the p.l.n. functions in Hilbert spaces.

Recall that, in our notation, X represents an arbitrary Hilbert space while ∂f denotes the limiting proximal subdifferential on X.

Definition 2.2.1. *(primal-lower-nice property) A l.s.c. function $f : X \to \overline{\mathbb{R}}$ is primal-lower-nice (p.l.n.) at $\bar{x}$, a point where f is finite, if there exist scalars $R > 0$, $c > 0$ and $\varepsilon > 0$ such that*

$$f(x') \geq f(x) + \langle v, x' - x\rangle - \frac{r}{2}|x' - x|^2 \quad \textit{for all} \quad x' \in \mathbb{B}(\bar{x}; \varepsilon)$$

whenever $r > R$, $|v| < cr$, $v \in \partial f(x)$ and $|x - \bar{x}| < \varepsilon$.

Before stating the definition of prox-regularity, we recall that f is *locally lower semicontinuous* at $\bar{x}$ if f is l.s.c. relative to the set $\{x \mid |x - \bar{x}| < \varepsilon, f(x) < \alpha\}$ for some $\varepsilon > 0$ and $\alpha > f(\bar{x})$. This is equivalent to the epigraph of f being closed relative to a neighborhood of $(\bar{x}, f(\bar{x}))$. Such a neighborhood is all that counts when the focus is on subgradients of f at $\bar{x}$.

Definition 2.2.2. *(prox-regularity of functions) A function $f : X \to \overline{\mathbb{R}}$ is prox-regular at $\bar{x}$ for $\bar{v}$ if f is finite and locally l.s.c. at $\bar{x}$ with $\bar{v} \in \partial f(\bar{x})$, and there exist scalars $\varepsilon > 0$ and $r \geq 0$ such that*

$$f(x') \geq f(x) + \langle v, x' - x\rangle - \frac{r}{2}|x' - x|^2 \quad \textit{for all} \quad x' \in \mathbb{B}(\bar{x}; \varepsilon)$$

whenever $v \in \partial f(x)$, $|v - \bar{v}| < \varepsilon$, $|x - \bar{x}| < \varepsilon$, $|f(x) - f(\bar{x})| < \varepsilon$.

When this holds for all $\bar{v} \in \partial f(\bar{x})$, f is said to be prox-regular at $\bar{x}$.

The class of prox-regular functions is much broader than that of p.l.n. functions. We see this directly from the definitions that the functional inequality for p.l.n. functions has to hold for all subgradients and do so with a linear growth condition, whereas for prox-regular functions the inequality only has to hold for subgradients close to a fixed $\bar{v}$ and just a neighborhood making not only x close to $\bar{x}$ but $f(x)$ close to $f(\bar{x})$, i.e. the localization of the subgradient mapping is in terms of an f-attentive neighborhood of $(\bar{x}, \bar{v})$. In particular, prox-regularity requires every limiting proximal subgradient v near $\bar{v}$ associated with an evolution point $(x, f(x))$ near $(\bar{x}, f(\bar{x}))$ to be a proximal subgradient, and all such proximal subgradients to share a common quadratic rate constant r.

For many functions the local property of f-attentiveness is automatic, because closeness of subgradients already ensures closeness of function values, then the condition on function values of Definition 2.2.2 is redundant. This leads to the following definition.

Definition 2.2.3. *(subdifferential continuity) A function $f : X \to \overline{I\!R}$ is subdifferentially continuous at $\bar{x}$ for $\bar{v}$, where $\bar{v} \in \partial f(\bar{x})$, if for every $\delta > 0$ there exist $\varepsilon > 0$ such that $|f(x) - f(\bar{x})| < \delta$ whenever $|x - \bar{x}| < \varepsilon$ and $|v - \bar{v}| < \varepsilon$ with $v \in \partial f(x)$. If this holds for all $\bar{v} \in \partial f(\bar{x})$, f is said to be subdifferentially continuous at $\bar{x}$.*

Next example shows that how a prox-regular function can fail to be subdifferentially continuous at $\bar{x} \in \operatorname{dom} f$.

Example 2.2.4. *Let $f : I\!R \to I\!R$.*

$$f(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x \le 0, \end{cases} \qquad \partial f(x) = \partial_p f(x) = \begin{cases} [0,\infty) & \text{if } x = 0, \\ \{0\} & \text{if } x \ne 0. \end{cases}$$

Obviously f is l.s.c. everywhere. It's easy to see too that f is prox-regular everywhere. The graph of ∂f has a vertical branch at $(\bar{x}, \bar{v}) = (0,0)$, though. As $(x_k, 0) \to (\bar{x}, \bar{v})$ with $x_k > 0$ we have $f(x_k) \equiv 1$, so $f(x_k) \not\to f(\bar{x}) = 0$. Hence f fails to be subdifferentially continuous at $\bar{x}$ for that $\bar{v}$.

In [**29**], Poliquin and Rockafellar showed that many important functions are subdifferentially continuous on $I\!R^n$. For p.l.n. functions this property holds in Hilbert space as well.

Proposition 2.2.5. *If $f : X \to \overline{I\!R}$ is p.l.n. at $\bar{x}$, then for all x in a neighborhood of $\bar{x}$ it is subdifferentially continuous at x for any $v \in \partial f(x)$.*

Proof. The proof given in [**29**], Proposition 2.2 can be carried over to Hilbert spaces as the only requirement there was the norm be given by an inner product. ◻

The scope and importance of the class of prox-regular functions in Hilbert space is readily appreciated from the fact that it includes not only all C^{1+} functions, all l.s.c., proper, convex functions, and all p.l.n. functions, but all strongly amenable functions.

Definition 2.2.6. *(strong amenability) A function $f : X \to \overline{I\!R}$ is strongly amenable at $\bar{x}$ if $f(\bar{x})$ is finite and there is an open neighborhood U of $\bar{x}$ on which f has a representation as $g \circ F$ with F a C^2 mapping from U to another Hilbert space Y and g a proper, l.s.c., convex function on Y such that the constraint qualification*

$$I\!R_+(\operatorname{dom} g - F(\bar{x})) - DF(\bar{x})(X) = Y$$

holds. Here $DF(\bar{x})$ denotes the Fréchet derivative of F at $\bar{x}$.

Note that in the preceding definition we adopt an extended version of the alternate form of the constraint qualification in [**29**], Definition 2.4 to the setting of an infinite-dimensional Hilbert space, cf.[**15**].

Proposition 2.2.7. *If $f : X \to \overline{I\!R}$ is strongly amenable at $\bar{x}$, then f is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$.*

Proof. Apply [**15**], Theorem 2.4 to conclude that f is p.l.n. at $\bar{x}$, and hence in particular it is prox-regular and subdifferentially continuous (from Proposition 2.2.5) at $\bar{x}$ for any $\bar{v} \in \partial f(\bar{x})$. □

Strongly amenable functions are omnipresent in optimization theory and variational analysis. In fact the problems most commonly encountered in optimization theory can be reformulated in terms of these functions. see [**10**], [**16**], [**18**]-[**20**], [**24**]-[**26**], [**29**]-[**32**] and [**40**]-[**42**].

The analysis of prox-regularity can be greatly simplified by normalizing to the case where $\bar{x} = 0$ and $\bar{v} = 0$, along with $f(\bar{x}) = 0$, as seen next.

Remark 2.2.8. *(perturbation of prox-regularity) Let $f : X \to \overline{I\!R}$ be prox-regular at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ and consider the perturbed function*

$$\tilde{f}(x) := f(x + \bar{x}) - f(\bar{x}) - \langle \bar{v}, x \rangle.$$

We then have $0 \in \partial \tilde{f}(0)$, along with $\tilde{f}(0) = 0$. It follows easily from the definition of prox-regularity for f that $\tilde{f}$ too is prox-regular at $\bar{x} = 0$ for $\bar{v} = 0$.

2.3. Subgradient Characterization of Prox-Regularity

Our first result establishes the subgradient characterization of prox-regularity in Hilbert space setting, which paves the way to the impending analysis. We show that f is prox-regular at $\bar{x}$ for $\bar{v}$ if and only if $\bar{v}$ is a proximal subgradient of f at $\bar{x}$ and, under suitable localization the multifunction $\partial f + rI$ is monotone (Γ is monotone if whenever $u_i \in \Gamma(x_i)$, $i = 1, 2$, then $\langle u_1 - u_2, x_1 - x_2 \rangle \geq 0$), where $r > 0$ constant and I is the identity mapping. This "pre-monotonicity" property is sufficient for a full range of desirable subdifferentiable properties. For example, when f is prox-regular at $\bar{x}$ for $\bar{v}$, the graph of ∂f coincides under a suitable localization near $(\bar{x}, \bar{v})$ (Definition 2.3.1), with a *Lipschitz manifold* in $X \times X$ (see Theorem 2.4.7).

Definition 2.3.1. *An f-attentive localization of ∂f around $(\bar{x}, \bar{v})$, is a (generally set-valued) mapping $T : X \rightrightarrows X$ whose graph in $X \times X$ is the intersection of* $\operatorname{gph} \partial f$ *with the product of an f-attentive neighborhood of $\bar{x}$ and an ordinary neighborhood of $\bar{v}$; this contrasts with an ordinary localization, in which the f-attentive neighborhood of $\bar{x}$ is relaxed to an ordinary neighborhood. More specifically for an $\varepsilon > 0$, the f-attentive ε-localization of ∂f around $(\bar{x}, \bar{v})$, is the mapping $T : X \rightrightarrows X$ defined by*

$$T(x) = \begin{cases} \{v \in \partial f(x) \mid |v - \bar{v}| < \varepsilon\} & \text{if } |x - \bar{x}| < \varepsilon \text{ and } |f(x) - f(\bar{x})| < \varepsilon, \\ \emptyset & \text{otherwise.} \end{cases} \tag{2.3.1}$$

Next we present a minimization principle due to Borwein and Preiss [**9**], which plays a key role in establishing the subgradient characterization of prox-regularity.

Theorem 2.3.2. *(smooth variational principle) Let $f : X \to \overline{I\!R}$ be l.s.c. and bounded below, and let $\varepsilon > 0$. Suppose that x_0 is a point satisfying $f(x_0) <$*

$\inf_{x\in X} f(x) + \varepsilon$. *Then, for any* $\lambda > 0$ *there exist points* y *and* z *with*

$$|z - x_0| < \lambda, \quad |y - z| < \lambda, \quad f(y) \le f(x_0)$$

and having the property that the function

$$x \to f(x) + \frac{\varepsilon}{\lambda^2}|x - z|^2$$

has a unique minimum at $x = y$.

Proof. See Clarke, Ledyave, Stern and Wolenski [**14**], Theorem 4.2. $\square$

The following consequence of the above variational principle will be useful in the proof of next theorem.

Remark 2.3.3. *Let* $\{x_k\}$ *be a minimizing sequence of* $\inf_{x\in X} f(x)$, *i.e. there exists* $\varepsilon_k \searrow 0$ *such that* $f(x_k) \le \inf_{x\in X} f(x) + \varepsilon_k$. *Then there exists another minimizing sequence* $\{y_k\}$ *such that* $|y_k - x_k| < 4\sqrt{\varepsilon_k}$ *with* $0 \in \partial f(y_k) + \sqrt{\varepsilon_k}\overline{B}$.
To see this, for each k, *take* $\varepsilon := \varepsilon_k$ *and* $\lambda := 2\sqrt{\varepsilon_k}$ *in Theorem 2.3.2. Then corresponding to the minimizing sequence* $\{x_k\}$ *there exist sequences* $\{y_k\}$ *and* $\{z_k\}$ *with* $|z_k - x_k| < 2\sqrt{\varepsilon_k}$, $|y_k - z_k| < 2\sqrt{\varepsilon_k}$ *(these two inequalities imply* $|y_k - x_k| < 4\sqrt{\varepsilon_k}$*),* $f(y_k) \le f(x_k)$ *(implies* $\{y_k\}$ *also a minimizing sequence), and the function*

$$x \to f(x) + \frac{\varepsilon_k}{(2\sqrt{\varepsilon_k})^2}|x - z_k|^2$$

has a unique minimum at $x = y_k$. *The latter implies, by the Fermat's rule 2.1.4,*

$$0 \in \partial_p f(y_k) + \frac{2}{4}(y_k - z_k),$$

$$0 \in \partial f(y_k) + \sqrt{\varepsilon_k}\overline{B}.$$

Now we establish the subgradient characterization of prox-regularity in Hilbert space which is obtained in finite-dimensional spaces by Poliquin and Rockafellar. See [**29**], Theorem 3.2.

Theorem 2.3.4. *(subgradient characterization of prox-regularity) When $f : X \to \overline{I\!R}$ is locally l.s.c. at $\bar{x}$, the following are equivalent.*

(a) The function f is prox-regular at $\bar{x}$ for $\bar{v}$, where $\bar{v} \in \partial f(\bar{x})$.

(b) The vector $\bar{v}$ is a proximal subgradient to f at $\bar{x}$, and there is an f-attentive ε-localization T of ∂f at $(\bar{x}, \bar{v})$ with a constant $r > 0$ such that $T + rI$ is monotone, i.e.,

$$\langle v_1 - v_0, x_1 - x_0 \rangle \geq -r|x_1 - x_0|^2 \text{ when } v_i \in T(x_i),\ i = 0, 1. \tag{2.3.2}$$

Proof. (a) $\Rightarrow$ (b). Take ε and r from Definition 2.2.2 of prox-regularity, and for the same ε let T be the f-attentive ε-localization of ∂f as in (2.3.1). As noted, the prox-regularity condition implies for every $(x, v) \in \operatorname{gph} T$ that v is a proximal subgradient of f at x, and this applies in particular to $(\bar{x}, \bar{v})$. Indeed, for any two pairs (x_0, v_0) and (x_1, v_1) in $\operatorname{gph} T$ we have

$$\begin{aligned} f(x_1) &\geq f(x_0) + \langle v_0, x_1 - x_0 \rangle - \frac{r}{2}|x_1 - x_0|^2, \\ f(x_0) &\geq f(x_1) + \langle v_1, x_0 - x_1 \rangle - \frac{r}{2}|x_0 - x_1|^2. \end{aligned}$$

In adding these inequalities together, we get the inequality in (2.3.2).

(b)$\Rightarrow$(a). There is no loss of generality in supposing $\bar{x} = 0$, $\bar{v} = 0$, and $f(0) = 0$ (cf. Remark 2.2.8). Further, we may assume that f be l.s.c. on X with bounded domain, since that can be manufactured out of the local l.s.c. property by adding some indicator function to f. Let $\bar{\varepsilon}$ and $\bar{r}$ be parameter values such that when T is the f-attentive $\bar{\varepsilon}$-localization of ∂f the property in (b) holds for $\bar{r}$. We first establish two claims.

Claim 1. *There exist $\varepsilon \in (0, \bar{\varepsilon})$ and $r > \bar{r}$ such that $\varepsilon r < \bar{\varepsilon}$ and $z \in X$ with $|z| < \frac{\varepsilon}{4}$ we have*

$$\operatorname*{argmax}_{|x|\le\varepsilon}\left\{\langle rz,x\rangle-\frac{r}{2}|x|^2-f(x)\right\}\subset\frac{3\varepsilon}{4}\overline{\mathbb{B}},$$

Proof of Claim 1. With our setting $\bar{x}=0,\bar{v}=0,f(0)=0$ and $\bar{v}\in\partial_p f(\bar{x})$ we may also assume that f has a lower bound $-(\bar{r}/2)\varepsilon^2$ on $\varepsilon\overline{\mathbb{B}}$ for any $0<\varepsilon<\bar{\varepsilon}$. Then by [**24**], Lemma 3.2 (Note that [**24**],Lemma 3.2 is stated in terms of finite dimensional space, but the only essential requirement is that the norm be given by an inner product) we have

$$\left.\begin{array}{l}\text{for any } 0<\varepsilon<\bar{\varepsilon} \text{ and for } r\ge\dfrac{32(-m)}{3\varepsilon^2}, \text{ where } m:=\inf\limits_{|x|\le\varepsilon}\{f(x)\}\\ \text{one has } \operatorname*{argmax}\limits_{|x|\le\varepsilon}\left\{\langle rz,x\rangle-\dfrac{r}{2}|x|^2-f(x)\right\}\subset\dfrac{3\varepsilon}{4}\overline{\mathbb{B}},\quad \text{for all } |z|<\dfrac{\varepsilon}{4}.\end{array}\right\}\quad(2.3.3)$$

Now restrict $\varepsilon\in(0,\bar{\varepsilon})$ such that $(16/3)\bar{r}\varepsilon<\bar{\varepsilon}$, $\varepsilon(\bar{r}+1)<\bar{\varepsilon}$ and set $r:=\max\{\frac{32(-m)}{3\varepsilon^2},(\bar{r}+1)\}$. For m, we have $m\ge(-\bar{r}/2)\varepsilon^2$ by the lower bound $(-\bar{r}/2)\varepsilon^2$ of f on $\varepsilon\overline{\mathbb{B}}$. We then have

$$\text{either } \varepsilon r=\frac{32(-m)}{3\varepsilon}\le\frac{16}{3}\bar{r}\varepsilon<\bar{\varepsilon} \text{ or } \varepsilon r=\varepsilon(\bar{r}+1)<\bar{\varepsilon},$$

as required by the claim. Then Claim 1 follows from (2.3.3).

Claim 2. *There exist $\varepsilon_1>0$ and $r_1>0$ such that if $v=r_1(z-x)$ is in $\partial f(x)$ with $|x|<(\varepsilon_1/4)$, $|z|<(\varepsilon_1/4)$ and $f(x)<(\varepsilon_1/4)$ then*

$$f(x')\ \ge\ f(x)+\langle v,\,x'-x\rangle-\frac{r}{2}|x'-x|^2 \text{ for all } x'\in\varepsilon_1\overline{\mathbb{B}}.$$

Proof of Claim 2. Let $0<\varepsilon_1<\min\{\bar{\varepsilon},16/3\}$ and $r_1>\bar{r}$ with $\varepsilon_1 r_1<\bar{\varepsilon}$, where r_1 is given by Claim 1 with $\varepsilon=\varepsilon_1$. Let $v=r_1(z-x)\in\partial f(x)$ with $|x|<(\varepsilon_1/4)$, $|z|<(\varepsilon_1/4)$ and $f(x)<(\varepsilon_1/4)$. Notice that

$$|v|\le r_1(|z|+|x|)<r_1\left(\frac{\varepsilon_1}{4}+\frac{\varepsilon_1}{4}\right)=\frac{r_1\varepsilon_1}{2}<\bar{\varepsilon}.$$

Consider the following optimization problem:

$$(\mathcal{P}): \quad \sup_{|x'|\le\varepsilon_1} \Big\{ \langle r_1 z, x'\rangle - \frac{r_1}{2}|x'|^2 - f(x') \Big\}.$$

Notice that this supremum is a finite number because f is bounded below on $\varepsilon_1 \overline{I\!B}$. Let $\{x_k\}$ be any maximizing sequence of $(\mathcal{P})$, i.e., there exists $\{\eta_k\}$ a sequence of nonnegative numbers converging to 0 such that

$$\langle r_1 z, x_k\rangle - \frac{r_1}{2}|x_k|^2 - f(x_k) \ge \sup_{|x'|\le\varepsilon_1} \Big\{ \langle r_1 z, x'\rangle - \frac{r_1}{2}|x'|^2 - f(x') \Big\} - \eta_k.$$

Equivalently,

$$-\langle r_1 z, x_k\rangle + \frac{r_1}{2}|x_k|^2 + f(x_k) \le \inf_{|x'|\le\varepsilon_1} \Big\{ -\langle r_1 z, x'\rangle + \frac{r_1}{2}|x'|^2 + f(x') \Big\} + \eta_k,$$

where $\{x_k\}$ is a minimizing sequence of $(-\mathcal{P})$. By Claim 1, we may assume without loss of generality that $\{x_k\} \subset (3/4)\varepsilon_1 \overline{I\!B}$. Then by Remark 2.3.3, there exists $\{w_k\}$, another minimizing sequence of $(-\mathcal{P})$ same as maximizing sequence of $(\mathcal{P})$ such that $|w_k - x_k| \le 4\sqrt{\eta_k}$ and eventually

$$0 \in \partial\Big[-\langle r_1 z, \cdot\rangle + \frac{r_1}{2}|\cdot|^2 + f(\cdot)\Big](w_k) + \sqrt{\eta_k}\,\overline{I\!B}.$$

We may rewrite the above inclusion with $\{y_k\} \in \overline{I\!B}$ so that

$$(r_1 z - r_1 w_k) - \sqrt{\eta_k}\, y_k \in \partial f(w_k).$$

Eventually $|(r_1 z - r_1 w_k) - \sqrt{\eta_k}\, y_k| < \bar{\varepsilon}$. To see this notice that

$$\begin{aligned}
|(r_1 z - r_1 w_k) - \sqrt{\eta_k}\, y_k| &\le r_1|z| + r_1|w_k| + \sqrt{\eta_k} \\
&< r_1\frac{\varepsilon_1}{4} + r_1\Big(\frac{3\varepsilon_1}{4} + 4\sqrt{\eta_k}\Big) + \sqrt{\eta_k} \\
&= r_1\varepsilon_1 + (1 + 4r_1)\sqrt{\eta_k}.
\end{aligned}$$

Hence, eventually $|(r_1 z - r_1 w_k) - \sqrt{\eta}_k y_k| < \bar{\varepsilon}$ because $\{\eta_k\}$ is converging to 0 and $r_1\varepsilon_1 < \bar{\varepsilon}$.

Now we show that $|f(w_k)| < \bar{\varepsilon}$ eventually.

Indeed, since $\{w_k\}$ is a maximizing sequence of $(\mathcal{P})$, there exists $\{\eta'_k\}$ a sequence of nonnegative numbers converging to 0 such that

$$\eta'_k + \langle r_1 z, w_k\rangle - \frac{r_1}{2}|w_k|^2 - f(w_k) \geq \sup_{|x'|\leq\varepsilon_1}\left\{\langle r_1 z, x'\rangle - \frac{r_1}{2}|x'|^2 - f(x')\right\} \geq 0.$$

The last inequality is a consequence of $f(0) = 0$. We then have

$$\begin{aligned} f(w_k) &\leq \langle r_1 z, w_k\rangle - \frac{r_1}{2}|w_k|^2 + \eta'_k \\ &\leq r_1|z||w_k| + \eta'_k \\ &< \frac{r_1\varepsilon_1}{4}(\frac{3\varepsilon_1}{4} + 4\sqrt{\eta}_k) + \eta'_k \\ &< \frac{3}{16}\bar{\varepsilon}\varepsilon_1 + r_1\varepsilon_1\sqrt{\eta}_k + \eta'_k. \end{aligned}$$

Hence, we eventually have that $|f(w_k)| < \bar{\varepsilon}$ because the sequences η_k and η'_k are converging to 0 and $\varepsilon_1 < 16/3$. Note here that the inequality $f(w_k) > -\bar{\varepsilon}$ comes from the l.s.c. of f at 0.

We have shown that $|(r_1 z - r_1 w_k) - \sqrt{\eta}_k y_k| < \bar{\varepsilon}$ with $(r_1 z - r_1 w_k) - \sqrt{\eta}_k y_k \in \partial f(w_k)$, $|w_k| < \bar{\varepsilon}$ and $|f(w_k)| < \bar{\varepsilon}$. Now, if $v = r_1(z-x) \in \partial f(x)$ with $|x| < (\varepsilon_1/4)$, $|z| < (\varepsilon_1/4)$, $|f(x)| < (\varepsilon_1/4)$, and hence $|v| < \bar{\varepsilon}$. By (2.3.2) we then have

$$\langle (r_1 z - r_1 w_k) - \sqrt{\eta}_k y_k - v, w_k - x\rangle \geq -\bar{r}|w_k - x|^2$$

$$(\bar{r} - r_1)|w_k - x|^2 \geq \sqrt{\eta}_k\langle y_k, w_k - x\rangle.$$

Letting $\eta_k \searrow 0$, we conclude that $\{w_k\}$ converges to x (recall that $r_1 > \bar{r}$). Because $\{w_k\}$ is a maximizing sequence of $(\mathcal{P})$ we conclude that the supremum is attained at x. This is because

$$\limsup_{k\to\infty}\left\{\langle r_1 z, w_k\rangle - \frac{r_1}{2}|w_k|^2 - f(w_k)\right\} \leq \langle r_1 z, x\rangle - \frac{r_1}{2}|x|^2 - f(x).$$

We have shown that

$$\langle r_1 z, x'\rangle - \frac{r_1}{2}|x'|^2 - f(x') \le \langle r_1 z, x\rangle - \frac{r_1}{2}|x|^2 - f(x)$$

for all $x' \in \varepsilon_1 \overline{I\!B}$. Rearranging this inequality we get

$$\begin{aligned} f(x') &\ge f(x) + \langle r_1 z, x' - x\rangle + \frac{r_1}{2}|x|^2 - \frac{r_1}{2}|x'|^2 \\ &= f(x) + \langle r_1(z - x), x' - x\rangle - \frac{r_1}{2}|x' - x|^2 \end{aligned}$$

for all $x' \in \varepsilon_1 I\!B$. This completes the proof of Claim 2.

Now to finish off the proof just let $\tilde{\varepsilon} = \varepsilon_1/8$, $\tilde{r} = r_1$. Then if $v \in \partial f(x)$ with $|v| < \tilde{\varepsilon}$, $|x| < \tilde{\varepsilon}$, $|f(x)| < \tilde{\varepsilon}$ and for $z = v/\tilde{r} + x$ we have

$$|z| \le \frac{|v|}{r_1} + |x| < |v| + |x| < 2\tilde{\varepsilon} = \frac{\varepsilon_1}{4}.$$

Notice here that we have used $r_1 > \bar{r} + 1 > 1$ which is true by our choice of r_1 . Then from Claim 2, whenever $|x| < \tilde{\varepsilon}$, $|f(x)| < \tilde{\varepsilon}$, $|v| < \tilde{\varepsilon}$ with $v \in \partial f(x)$ we have

$$f(x') \ge f(x) + \langle v, x' - x\rangle - \frac{\tilde{r}}{2}|x' - x|^2 \text{ for all } x' \in \tilde{\varepsilon}\overline{I\!B}.$$

This tells us that f is prox-regular at $\bar{x} = 0$ for $\bar{v} = 0$ (with parameters $\tilde{\varepsilon}$ and $\tilde{r}$).

□

Remark 2.3.5. *The implication (a) ⇒ (b) is true in general Banach space setting as one sees this readily from the proof.*

Remark 2.3.6. *The proof of the Theorem in finite dimensions ([**29**], Theorem 3.2) heavily depends on the existence of minimizers of l.s.c. function on a compact set. Our proof here relies on a more general technique - the smooth variational principle.*

2.4. Regularity Properties of Moreau Envelopes

For a proper, l.s.c. function $f : X \to \overline{I\!R}$ and parameter value $\lambda > 0$, the *Moreau envelope* function, e_λ and the *proximal mapping*, P_λ are defined by

$$e_\lambda(x) := \inf_{x'} \left\{ f(x') + \frac{1}{2\lambda}|x' - x|^2 \right\},$$
$$P_\lambda(x) := \operatorname*{argmin}_{x'} \left\{ f(x') + \frac{1}{2\lambda}|x' - x|^2 \right\}.$$

The primary aim of studying the envelope functions e_λ and the proximal mappings P_λ associated with a function f is to learn more about the behaviour of f around a point $\bar{x}$ when f is prox-regular at $\bar{x}$ for a vector $\bar{v} \in \partial f(\bar{x})$. For example, the nice properties of e_λ and P_λ of a prox-regular function f (Theorem 2.4.4) with the already established subgradient characterization of f (Theorem 2.3.4) reveal the major fact that the graph of ∂f coincides, under a suitable localization, with a Lipschitz manifold in $X \times X$ (Theorem 2.4.7).

We proceed to establish the smoothness properties of the Moreau envelopes of prox-regular functions.

Let $f : X \to \overline{I\!R}$ be prox-regular at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$. Then $\bar{v}$ is actually a proximal subgradient of f at $\bar{x}$. In order to simplify our analysis, by Remark 2.2.8., without any loss of generality we normalize to the case $\bar{x} = 0$, $\bar{v} = 0$ and $f(0) = 0$. Since our primary interest of f and ∂f depend only on the local geometry of $\operatorname{epi} f$ around $(\bar{x}, f(\bar{x}))$, we may further, if necessary, add to f the indicator of some ball with center at $\bar{x}$ to make $\operatorname{dom} f$ be bounded. By taking the radius of that ball small enough we can get the quadratic inequality for $\bar{v} \in \partial_p f(\bar{x})$ to hold for all x. Thus we work under the baseline assumptions that

$$\left.\begin{array}{l} f \text{ is locally l.s.c. at } 0 \text{ with } f(0) = 0, \text{ and} \\ r > 0 \text{ is such that } f(x) \geq -\frac{r}{2}|x|^2 \text{ for all } x \end{array}\right\} \qquad (2.4.1)$$

which imply that

$$e_\lambda(0) = 0 \text{ and } P_\lambda(0) = \{0\} \text{ when } \lambda \in (0, 1/r). \tag{2.4.2}$$

In order to follow the steps of Poliquin and Rockafellar, we next extend the results of Propositions 4.2 and 4.3 of [**29**] to Hilbert spaces. First, we require a lemma.

Lemma 2.4.1. *([**29**], Lemma 4.1) Under assumptions (2.4.1), consider any $\lambda \in (0, 1/r)$ and let $\mu = (1 - \lambda r)^{-1}$. For any $\rho > 0$,*

$$f(x') + \frac{1}{2\lambda}|x' - x|^2 \leq e_\lambda(x) + \rho \quad \Longrightarrow \quad \begin{cases} |x'| \leq 2\mu|x| + \sqrt{2\lambda\mu\rho}, \\ f(x') \leq \frac{1}{2\lambda}|x|^2 + \rho, \\ f(x') \geq -\frac{r}{2}\left(2\mu|x| + \sqrt{2\lambda\mu\rho}\right)^2. \end{cases} \tag{2.4.3}$$

Proof. The same proof of [**29**], Lemma 4.1 can be carried over to this Hilbertian case, since the only requirement there was the norm be given by an inner product. □

Proposition 2.4.2. *Under assumptions (2.4.1), consider any $\lambda \in (0, 1/r)$. For any $\varepsilon > 0$ there is a neighborhood V of $\bar{x} = 0$ such that*

(a) e_λ is Lipschitz continuous on V with constant ε and bounded below by a quadratic function,

(b) $|x'| < \varepsilon$, $|f(x')| < \varepsilon$ and $\lambda^{-1}|x - x'| < \varepsilon$ for all $x' \in P_\lambda(x)$ when $x \in V$.

Proof. (The proof given here differs from that of Poliquin and Rockafellar [**29**], Proposition 4.2 (a) and (c) because the argument given there relies on the existence of minimizers of a l.s.c. function over a closed bounded set, which is not true in the case of Hilbert spaces). Let $\mu = (1 - \lambda r)^{-1}$ and $\varepsilon' \in (0, \varepsilon)$. Choose $\delta > 0$ and $\rho > 0$ small enough that $(2\varepsilon' + 3\delta)/\lambda \leq \varepsilon$ and

$$2\mu\delta + \sqrt{2\lambda\mu\rho} \leq \varepsilon', \quad \frac{1}{2\lambda}\delta^2 + \rho \leq \varepsilon', \quad \frac{r}{2}\left(2\mu\delta + \sqrt{2\lambda\mu\rho}\right)^2 \leq \varepsilon', \quad \frac{\delta(1 + 2\mu)}{\lambda} \leq \varepsilon',$$

and let $V := \{x \mid |x| \leq \delta\}$ and $C := \{x \mid |x| \leq \varepsilon'\}$.

(a) Let any x and y belong to V. For any $\rho > 0$, by the definition of $e_\lambda(y)$ as an infimum, there exists x' such that

$$f(x') + \frac{1}{2\lambda}|x' - y|^2 \leq e_\lambda(y) + \rho.$$

Then by Lemma 2.4.1 we have $|x'| \leq 2\mu|y| + \sqrt{2\lambda\mu\rho} \leq 2\mu\delta + \sqrt{2\lambda\mu\rho} \leq \varepsilon'$, which implies $x' \in C$.

Thus we have

$$\begin{aligned} e_\lambda(x) - e_\lambda(y) &\leq f(x') + \frac{1}{2\lambda}|x' - x|^2 - f(x') - \frac{1}{2\lambda}|x' - y|^2 + \rho \\ &= \frac{1}{2\lambda}|x - y|^2 - \frac{1}{\lambda}\langle x - y, x' - y\rangle + \rho \\ &\leq \frac{1}{2\lambda}|x - y|^2 + \frac{1}{\lambda}|x - y||x' - y| + \rho \\ &\leq K|x - y| + \rho, \end{aligned} \tag{2.4.4}$$

where K is chosen so that $K := (1/\lambda)\sup\{|x| + 2|z - x|;\, x \in V,\, z \in C\} < \infty$. Indeed, we have $K \geq (1/\lambda)\{|y| + 2|x' - y|\}$ for all $y \in V$ and $x' \in C$ and hence

$$|x - y|K \geq \frac{1}{\lambda}\{|x - y||y| + 2|x - y||x' - y|\}.$$

We also have that $|x - y|K \geq \frac{1}{\lambda}|x||x - y|$ because $K \geq \frac{1}{\lambda}|x|$ for all x in V. In adding these inequalities together, we get the inequality in (2.4.4):

$$\begin{aligned} |x - y|K &\geq \frac{1}{2\lambda}\{|x - y|(|x| + |y|) + 2|x - y||x' - y|\} \\ &\geq \frac{1}{2\lambda}|x - y|^2 + \frac{1}{\lambda}|x - y||x' - y|. \end{aligned}$$

And this constant K cannot be bigger than ε:

$$\begin{aligned} K &= \frac{1}{\lambda}\sup\{|x| + 2|z - x|;\, x \in V, z \in C\} \\ &\leq \frac{1}{\lambda}\sup\{|x| + 2(|z| + |x|);\, x \in V, z \in C\} \\ &\leq \frac{1}{\lambda}(3\delta + 2\varepsilon') \leq \varepsilon. \end{aligned}$$

Reversing the roles of x and y, and then letting $\rho\searrow 0$ in (2.4.4) shows that e_λ is Lipschits of rank ε on V.

The asserted lower bound for e_λ follows from

$$\begin{aligned}
e_\lambda(x) &= \inf_{x'}\{f(x') + \frac{1}{2\lambda}|x'-x|^2\} \\
&\geq \inf_{x'}\{-\frac{r}{2}|x'|^2 + \frac{1}{2\lambda}|x'-x|^2\} \\
&= -\frac{\frac{1}{2\lambda}\frac{r}{2}}{\frac{1}{2\lambda}-\frac{r}{2}}|x|^2 \\
&= -\frac{r}{2(1-r\lambda)}|x|^2.
\end{aligned}$$

(b) When $x'\in P_\lambda(x)$, then Lemma 2.4.1 is true for every $\rho>0$ which implies

$$\begin{aligned}
|x'| &\leq 2\mu|x| \leq 2\mu\delta \leq \varepsilon' < \varepsilon, \\
f(x') &\leq \frac{1}{2\lambda}|x|^2 \leq \frac{1}{2\lambda}\delta^2 \leq \varepsilon' < \varepsilon, \\
f(x') &\geq -\frac{r}{2}\left(2\mu|x| + \sqrt{2\lambda\mu\rho}\right)^2 \geq -\varepsilon' > -\varepsilon,
\end{aligned}$$

and also

$$\begin{aligned}
\frac{1}{\lambda}|x-x'| &\leq \frac{1}{\lambda}(|x|+|x'|) \\
&\leq \frac{1}{\lambda}(1+2\mu)|x| \\
&\leq \frac{\delta}{\lambda}(1+2\mu) \leq \varepsilon' < \varepsilon.
\end{aligned}$$

□

Proposition 2.4.3. *Under assumptions (2.4.1), there exists for each $\lambda\in(0,1/r)$ a neighborhood V of $\bar{x}=0$ on which*

(a) $\partial e_\lambda(x)\subset\{\lambda^{-1}(x-x')\,|\,x'\in P_\lambda(x)\}$ and $P_\lambda(x)\neq\emptyset$, where $x\in V$,

(b) $x'\in P_\lambda(x)\Longrightarrow\lambda^{-1}(x-x')\in\partial f(x')$, i.e., $x'\in(I+\lambda\partial f)^{-1}(x)$.

Proof. Verification of (b) is easy. We begin with that. Recall that the existence of a proximal subgradient at x' corresponds to the existence of a "local quadratic support" to f at x' (see the Definition 2.1.1 and the remarks thereafter). When $x' \in P_\lambda(x)$ we have

$$f(x'') + \frac{1}{2\lambda}|x'' - x|^2 \geq f(x') + \frac{1}{2\lambda}|x' - x|^2 \text{ for all } x'',$$

so that $f(x'') - f(x') \geq q(x'')$ for the quadratic function $q(x'') = \left(|x' - x|^2 - |x'' - x|^2\right)/2\lambda$. We have $q(x') = 0$ and $Dq(x') = \lambda^{-1}(x - x')$, so q forms a local quadratic support to f at x'. Thus $\lambda^{-1}(x - x') \in \partial_p f(x')$. In particular, we have (b).

Now we verify (a). (Our proof here differs from that of Poliquin and Rockafellar [**29**], Proposition 4.3 (a), as their arguments require the compactness of closed, bounded sets. Another difficulty is to work with "weak-limits" required by the limiting proximal subdifferentials). We fix $\lambda \in (0, 1/r)$ and choose a neighborhood V of 0 with the properties in Proposition 2.4.2. First note that the Lipschitz property of e_λ on V (2.4.2(a)) ensures that the limiting proximal subdifferential $\partial e_\lambda(x)$ is nonempty for all x belong to V (see, Loewen[**21**], Cor. 4C.9). Consider any point $x \in V$ and any $v \in \partial e_\lambda(x)$. Then $v = w\text{-}\lim_{k\to\infty} v_k$ for some sequence $v_k \in \partial_p e_\lambda(x_k)$ and $x_k \to x$ with $e_\lambda(x_k) \to e_\lambda(x)$. For each k, there are positive numbers M_k and δ_k such that

$$e_\lambda(w) \geq e_\lambda(x_k) + \langle v_k, w - x_k\rangle - M_k|w - x_k|^2 \qquad \forall\, w \in x_k + \delta_k \mathbb{B}. \tag{2.4.5}$$

Choose any $t_k > 0$ so small that $t_k < \delta_k$ and $M_k t_k < 1/k$. This allows us to set $w = x_k + t_k u$, where $u \in \mathbb{B}$ in (2.4.5): the result can be written as

$$\langle v_k, t_k u\rangle \leq e_\lambda(x_k + t_k u) - e_\lambda(x_k) + M_k t_k^2. \tag{2.4.6}$$

By the definition of $e_\lambda(x_k)$ as an infimum, there exists y_k such that

$$e_\lambda(x_k) \leq f(y_k) + \frac{1}{2\lambda}|y_k - x_k|^2 \leq e_\lambda(x_k) + t_k^2. \tag{2.4.7}$$

Also notice that $e_\lambda(x_k + t_k u) \le f(y_k) + \frac{1}{2\lambda}|y_k - (x_k + t_k u)|^2$. Thus, (2.4.6) gives us

$$\langle v_k, t_k u\rangle \le \frac{1}{2\lambda}|y_k - (x_k + t_k u)|^2 - \frac{1}{2\lambda}|y_k - x_k|^2 + (M_k + 1)t_k^2,$$

which we can expand on the right and then rewrite as

$$\begin{aligned}\langle v_k - \frac{1}{\lambda}(x_k - y_k), u\rangle &\le \frac{t_k}{2\lambda} + M_k t_k + t_k \\ &< \frac{t_k}{2\lambda} + \frac{1}{k} + t_k. \end{aligned} \tag{2.4.8}$$

Claim. *There exists a subsequence $\{y_{k'}\}$ of $\{y_k\}$ that converges strongly to $y' :=$ $(x - \lambda v)$.*

Proof of Claim. It is immediate from (2.4.8) that $\{y_k\}$ is norm bounded. Then there exists a subsequence $\{y_{k'}\}$ of $\{y_k\}$ that converges weakly to $y' := x - \lambda v$. This again follows from (2.4.8) just replacing y_k with $y_{k'}$. Now to see $\{y_{k'}\}$ actually converges to y' strongly, rewrite (2.4.8) as

$$\langle v_{k'} - \frac{1}{\lambda}x_{k'} - (v - \frac{1}{\lambda}x) + (v - \frac{1}{\lambda}x) + \frac{y_{k'}}{\lambda}, u\rangle < \frac{t_{k'}}{2\lambda} + \frac{1}{k'} + t_{k'}$$

and, eventually

$$\langle (v - \frac{1}{\lambda}x) + \frac{y_{k'}}{\lambda}, u\rangle < \eta_{k'}, \text{ where } \eta_{k'} \searrow 0,$$

that implies $|(v - \frac{1}{\lambda}x) + \frac{y_{k'}}{\lambda}| < \eta_{k'}$ eventually since $u \in I\!B$ is arbitrary. So $\{y_{k'}\}$ is strongly converging to $y' = x - \lambda v$, as required.

Restricting to the subsequence $\{y_{k'}\}$ in (2.4.7) we have

$$e_\lambda(x_{k'}) \le f(y_{k'}) + \frac{1}{2\lambda}|y_{k'} - x_{k'}|^2 \le e_\lambda(x_{k'}) + t_{k'}^2.$$

Since f is locally l.s.c. at 0, taking the lower limit (*liminf*) of the above inequality as $k' \to \infty$ confirms that $e_\lambda(x) = f(y') + \frac{1}{2\lambda}|y' - x|^2$, where $y' = x - \lambda v$. Thus we have proved that $P_\lambda(x) \neq \emptyset$ for all $x \in V$ and since $v = \frac{x-y'}{\lambda}$ the inclusion in (a) is valid too. ∎

When we assume f to be prox-regular, the above propositions with Theorem 2.3.4 entail the $\mathcal{C}^{1+}$ smoothness of e_λ and the local single-valuedness of P_λ as seen by the next theorem. The proof of the next theorem in Hilbert space easily follows, in fact, it is quite the same as in the finite-dimensional case ([**29**], Theorem 4.4).

Theorem 2.4.4. *Suppose that f is prox-regular at $\bar{x} = 0$ for $\bar{v} = 0$ with respect to ε and r, in particular with (2.4.1) holding. Let T be the f-attentive ε-localization of ∂f around $(0,0)$. Then for each $\lambda \in (0, 1/r)$ there is a neighborhood V of $\bar{x} = 0$ such that, on V, the mapping P_λ is single-valued and Lipschitz continuous with constant $1/(1-\lambda r)$ and*

$$P_\lambda(x) \;=\; (I + \lambda T)^{-1}(x) \;=\; [\textit{singleton}],$$

while the function e_λ is of class C^{1+} with $De_\lambda(0) = 0$ and

$$De_\lambda(x) = \frac{x - P_\lambda(x)}{\lambda} = \lambda^{-1}\left[I - [I + \lambda T]^{-1}\right](x).$$

Proof. Choose V open and small enough that the properties in Propositions 2.4.2 and 2.4.3 hold on V. Then for $x \in V$ we have $P_\lambda(x)$ nonempty by 2.4.3(a), while $\partial e_\lambda(x)$ is nonempty by 2.4.2(a) and satisfies the inclusion in 2.4.3(a). In this inclusion and the one in 2.4.3(b) we can replace ∂f by T because of 2.4.2(b). Aiming at the formulas claimed here for $P_\lambda(x)$ and $De_\lambda(x)$, we first show that $(I + \lambda T)^{-1}$ cannot be multivalued and P_λ is Lipschitz continuous on V.

Suppose that $x_i \in (I + \lambda T)^{-1}(x) \cap V$, $i = 0, 1$. Then $(x - x_i)/\lambda \in T(x_i)$. Invoking the prox-regularity of f, we have the monotonicity of $T + rI$ by Theorem 2.3.4 and therefore

$$\left\langle \left[\frac{x - x_1}{\lambda}\right] - \left[\frac{x - x_0}{\lambda}\right], x_1 - x_0 \right\rangle \geq -r|x_1 - x_0|^2,$$

hence $-\lambda^{-1}|x_1 - x_0|^2 \geq -r|x_1 - x_0|^2$. Then $(1-\lambda r)|x_1 - x_0|^2 \leq 0$, so $x_1 = x_0$.

To show P_λ is Lipschitz continuous, let $x_i' \in P_\lambda(x_i)$ with $x_i \in V$, $i = 0, 1$. We have

$$\left\langle \left[\frac{x_1 - x_1'}{\lambda}\right] - \left[\frac{x_0 - x_0'}{\lambda}\right], x_1' - x_0' \right\rangle \geq -r|x_1' - x_0'|^2,$$

so $\langle x_1 - x_0, x_1' - x_0' \rangle \geq (1-\lambda r)|x_1' - x_0'|^2$, i.e., $|x_1 - x_0| \geq (1-\lambda r)|x_1' - x_0'|$. This can be written in the form $|x_1' - x_0'| \leq [1/(1-\lambda r)]\,|x_1 - x_0|$.

Thus we have $P_\lambda(x) = (I + \lambda T)^{-1}(x)$ and the limiting proximal subdifferential ∂e_λ reduces to a single valued mapping on V, i.e., $\partial e_\lambda(x) = \frac{x - P_\lambda(x)}{\lambda}$. Then by [**12**], Proposition 2.2.4, $\partial e_\lambda(x)$ coincides with the strict derivative of e_λ, $D_s e_\lambda(x)$ on V, i.e., $\partial e_\lambda(x) = D_s e_\lambda(x) = \frac{x - P_\lambda(x)}{\lambda}$. Because $P_\lambda(x)$ is Lipschitz continuous, e_λ is actually of class C^{1+} on V. However, when X is finite-dimensional, the limiting subdifferential reduces to a singleton on an open set is necessary and sufficient for the corresponding Lipschitz function to be C^1 ([**12**], Corollary to Proposition 2.2.4).

□

The following lemma helps us to write the derivative formula in Theorem 2.4.4 in a useful form.

Lemma 2.4.5. *For any mapping $T : X \rightrightarrows X$ and any $\lambda > 0$, one has the identity*

$$\lambda^{-1}[I - (I + \lambda T)^{-1}] = (\lambda I + T^{-1})^{-1}.$$

Proof. The proof in the Hilbert space setting follows exactly as that in [**29**], Lemma 4.5. □

Proposition 2.4.6. *In Theorem 2.4.4, the derivative formula can be expressed equivalently as:*

$$De_\lambda(x) = \left[\lambda I + T^{-1}\right]^{-1}(x).$$

Proof. Simply combine Theorem 2.4.4 with Lemma 2.4.5. $\square$

Next we establish the aforementioned Lipschitzian property of a graph of a subdifferential mapping of a prox-regular function. For that matter, we adopt from Rockafellar [**39**] the notion of Lipschitz manifold to suit our Hilbert space settings.

Let Y be another Hilbert space. A set $M \subset X \times Y$ is a *Lipschitz manifold* around a point $(\bar{x}, \bar{y})$ in M if there is an open neighborhood U of $(\bar{x}, \bar{y})$ and a one-to-one mapping between U and an open subset O of $X \times Y$, continuously differentiable(Fréchet) in both directions, under which $U \cap M$ is identified with $O \cap \operatorname{gph} F$ for some Lipschitz continuous mapping F from an open subset of X into Y.

Theorem 2.4.7. *If the function $f : X \to \overline{I\!R}$ is prox-regular at $\bar{x}$ for a vector $\bar{v} \in \partial f(\bar{x})$, then for any $\varepsilon > 0$ the graph of the f-attentive ε-localization of ∂f at $(\bar{x}, \bar{v})$ is a Lipschitz manifold around $(\bar{x}, \bar{v})$ in $X \times X$. When f is subdifferentially continuous, this can be said of the graph of ∂f itself.*

Proof. For simplicity we can normalize to $\bar{x} = 0$ and $\bar{v} = 0$ (cf. 2.2.8): geometrically this just amounts to a translation of $\operatorname{gph} \partial f$ and its localizations. The formula in Proposition 2.4.6 then identifies $\operatorname{gph} T$ with the graph of the Lipschitz continuous mapping De_λ near $\bar{x}$ under a certain linear change of coordinates around $(\bar{x}, \bar{v})$. $\square$

As a consequence of this we deduce that the monotonicity of the subgradient mapping $T + rI$ in Theorem 2.3.4 is in fact "locally maximal";

Definition 2.4.8. *A mapping $S : X \rightrightarrows X$ is locally maximal monotone relative to $(\bar{x}, \bar{v}) \in \operatorname{gph} S$ if there is a neighborhood U of $(\bar{x}, \bar{v})$ in $X \times X$ such that, for* *every monotone mapping* $S' : X \rightrightarrows X$ *with* $\operatorname{gph} S' \supset \operatorname{gph} S$*, one has* $U \cap \operatorname{gph} S' = U \cap \operatorname{gph} S$.

Proposition 2.4.9. *If the function* $f : X \to \overline{I\!R}$ *is prox-regular at* $\bar{x}$ *for* $\bar{v} \in \partial f(\bar{x})$ *with parameter values* $\varepsilon > 0$ *and* $r > 0$*, the* f*-attentive* ε*-localization* T *of* ∂f *at* $(\bar{x}, \bar{v})$ *has the property that* $T + rI$ *is not just monotone but locally maximal monotone relative to* $(\bar{x}, \bar{v} + r\bar{x})$*. When* f *is subdifferentially continuous, this can be said of* $\partial f + rI$*.*

Proof. (The proof is quite the same as in [**29**], Proposition 4.8 with slight modification to Hilbertian settings). We can suppose $(\bar{x}, \bar{v}) = (0, 0)$. The elements $(x, v) \in \operatorname{gph} T$ correspond one-to-one to those of $\operatorname{gph} S$ for $S = T + rI$ under $(x, v) \longleftrightarrow (x, v + rx)$, this being affine in both directions. Hence by Theorem 2.4.7, $\operatorname{gph} S$ is a Lipschitz manifold around $(0, 0)$. The same is then true for the graph of the mapping $P = (I + S)^{-1}$; the correspondence between $\operatorname{gph} S$ and $\operatorname{gph} P$ is given by $(x, y) \longleftrightarrow (x + y, x)$. The monotonicity of S implies that P is nonexpansive (hence Lipschitz continuous) relative to its domain D in X. Some neighborhood of $(0, 0)$ in $\operatorname{gph} P$ thus corresponds one-to-one to a subset of D containing 0 under a mapping that is Lipschitz continuous in both directions. Since $\operatorname{gph} P$ is a Lipschitz manifold around $(0, 0)$, it follows that a subset of D containing 0 corresponds in such a way to an open subset of X, and therefore that D is a neighborhood of 0. For any monotone mapping S' with $\operatorname{gph} S' \supset \operatorname{gph} S$, the mapping $P' = (I + S')^{-1}$, whose graph corresponds one-to-one with that of S', is nonexpansive too, and $\operatorname{gph} P' \supset \operatorname{gph} P$. Therefore, P' can do no more than coincide with P on a neighborhood of 0. This means that the graph of S' must agree with that of S on a neighborhood of $(0, 0)$, and hence that S is locally maximal monotone with respect to $(0, 0)$. $\square$

2.5. Convexity of Moreau Envelopes

In this section we investigate the local properties of convexity of the envelope functions e_λ of prox-regular functions. We prove (in a Hilbert space) that in some local neighborhood the sum of e_λ and a positive multiple of norm square is convex. Further (in a separable Hilbert space), the conditions are given under which e_λ itself is convex or strongly convex.

Let $\Gamma : X \rightrightarrows X$. Recall that

- Γ is *monotone* if $\langle u_1 - u_2, x_1 - x_2 \rangle \geq 0$ whenever $u_i \in \Gamma(x_i)$.
- Γ is *strongly monotone* if $\Gamma - \mu I$ is monotone for some $\mu > 0$.

Lemma 2.5.1. *Let $T : X \rightrightarrows X$ be any set-valued mapping. Suppose that $T = \sigma I + M$ where M is monotone and σ is any value in $\mathbb{R}$ (positive, negative, zero). Let $\lambda > 0$ be small enough that $1 + \lambda\sigma > 0$. Then the mapping S_λ given by either side of the identity in Lemma 2.4.5 can be expressed by*

$$S_\lambda = \frac{\sigma}{1+\lambda\sigma} I + M' \text{ with } M'(w) = \frac{1}{1+\lambda\sigma}\Big(\frac{\lambda}{1+\lambda\sigma} I + M^{-1}\Big)^{-1}\Big(\frac{1}{1+\lambda\sigma} w\Big),$$

this mapping M' being monotone. Thus, when $\lambda > 0$ is sufficiently small,

$$T - \sigma I \text{ monotone} \qquad \Longrightarrow \qquad S_\lambda - \frac{\sigma}{1+\lambda\sigma} I \text{ monotone.}$$

Proof. The proof in the Hilbert space setting follows exactly as that in [**29**], Lemma 5.1. ☐

Theorem 2.5.2. *Suppose that f is prox-regular at $\bar{x} = 0$ for $\bar{v} = 0$ with respect to ε and r, in particular with (2.4.1) holding, and let $\lambda \in (0, 1/r)$. Then on some neighborhood of 0 the function*

$$e_\lambda + \frac{r}{2(1-\lambda r)} |\cdot|^2$$

is nonnegative and convex.

Proof. Prox-regularity of f at $\bar{x} = 0$ for $\bar{v} = 0$ implies the monotonicity of the mapping $T + rI$ (cf.Theorem 2.3.4). Then by taking $\sigma = -r$ in Lemma 2.5.1, we have $S_\lambda + r(1 - \lambda r)^{-1}I$ monotone, where S_λ is the mapping given by the identity in Lemma 2.4.5. But this is the derivative mapping of the function in question. Hence, this function is convex. The nonnegativity assertion follows from Proposition 2.4.2 (a), where we proved $e_\lambda \geq -\frac{r}{2(1-\lambda r)}|\cdot|^2$. □

Corollary 2.5.3. *If f is prox-regular at $\bar{x} = 0$ for $\bar{v} = 0$, and λ is sufficiently small, then on some neighborhood of the origin e_λ is a lower-C^2 function, hence in particular prox-regular itself.*

Proof. From the Theorem 2.5.2, we know that the function $f_0 := e_\lambda + \frac{r}{2(1-\lambda r)}|\cdot|^2$ is finite, convex on some neighborhood of 0 for $\lambda \in (0, 1/r)$, which in turn satisfies the characterization of lower-C^2 property for e_λ. Prox-regularity of e_λ follows from Theorem 2.3.4 because $\{0\} = \partial_p e_\lambda(0)$, and the mapping $De_\lambda + r(1 - \lambda r)^{-1}I$ is monotone around 0. □

In order to obtain a characterization of the convexity of e_λ, first we need to introduce a concept of "null" sets in infinite-dimensional spaces. For our purposes the most useful generalization of a null set is that of "Haar-null" set introduced by J.P.R. Christensen in [**11**].

Definition 2.5.4. *(Haar-null set) A Borel subset N of a separable Banach space E is called a Haar-null set if there exists a probability measure μ on the σ-algebra of Borel subsets of E so that $\mu(N + x) = 0$ for all $x \in E$.*

We recall some results of Christensen [**11**] about the notion of Haar-null sets.

Proposition 2.5.5. *Let E be a separable Banach space. Then we have the following.*

(a) If $E = \mathbb{R}^n$, then $H \subset E$ is Haar-null in E if and only if H is Lebesgue-negligible in $\mathbb{R}^n$.

(b) If $(H_n)_{n\in\mathbb{N}}$ is a countable family of Haar-null sets in E, then the set $H = \cup_{n\in\mathbb{N}} H_n$ is Haar-null in E.

(c) If H is Haar-null in E, then $E \setminus H$ is dense in E.

(d) Let B be a separable Banach space and H be a Haar-null subset in $B \times \mathbb{R}^n$. Then for almost every b in B, that is except for a Haar-null subset in B, the section

$$H(b) = \{z \in \mathbb{R}^n \mid (b, z) \in H\}$$

is Lebesgue-negligible subset in $\mathbb{R}^n$.

Proof. See the book by Christensen [**11**] or Borwein and Moors [**7**], Proposition 2.1. $\square$

We will need the following infinite-dimensional version of Rademachar's theorem due to Christensen, which states that locally Lipschitz mapping from a separable Banach Space to a separable reflexive Banach space is differentiable almost all points in the sense of Haar measure.

Proposition 2.5.6. *Let E be a separable Banach space and F be a separable reflexive Banach space. Let f be a locally Lipschitz mapping from E into F. Then f is Gâteaux differentiable on a subset D_f with $E \setminus D_f$ Haar-null in E.*

Proof. See [**11**], Theorem 7.5. $\square$

We will also need the following lemma in which we characterize the monotonicity of Lipschitz mappings in separable Hilbert spaces.

Lemma 2.5.7. *Let X be a separable Hilbert space. Suppose P is a Lipschitz continuous mapping from an open convex set $O \subset X$ into X. Then P is monotone on O if and only if the Gâteaux derivative $DP(y)$ is positive semidefinite wherever it exists for y in O.*

Proof. First assume that P is monotone on O. Let y in O such that the Gâteaux derivative $DP(y)$ exists. Then, from the definition of DP and the monotonicity of P, it follows that, for any $\eta \in X$,

$$\langle \eta, DP(y)\eta \rangle \;=\; \lim_{t \downarrow 0} \frac{1}{t^2} \langle t\eta, P(y+t\eta) - P(y) \rangle \;\geq\; 0.$$

So we get the positive semidefiniteness of $DP(y)$ as desired.

Conversely, assume that the Gâteaux derivative $DP(y)$ is positive semidefinite wherever it exists for y in O. Then, by Proposition 2.5.6, there exists a subset M of O on which P is Gâteaux differentiable and its Gâteaux derivative $DP(y)$ is positive semidefinite, and such that $O \setminus M$ is Haar-null in O. It suffices to prove

$$\langle P(y+v) - P(y), v \rangle \geq 0$$

for all $y,\ y+v \in O$.

If $v = 0$ then the result is trivial. Let us consider the case $v \neq 0$. As for each $y \in O$ the function $s \mapsto y + sv$ from $[0,1]$ into O is derivable, the function $s \mapsto P(y+sv)$ is derivable at each $s \in [0,1]$ such that $y + sv \notin N := O \setminus M$. Since X is a Hilbert space, we may write $X = G \oplus \mathbb{R}\cdot v$, as a direct sum of $\mathbb{R}\cdot v$ and a subspace G. Restricting to the subset O of X we write $O = G' \oplus R\cdot v$, where $G' \subset G$ and $R \subset \mathbb{R}$. Then, writing

$$y + sv \;=\; (y_1, y_2 v + sv), \quad \text{where} \quad y_1 \in G' \quad \text{and} \quad y_2 \in \mathbb{R},$$

it follows that the function $s \mapsto P(y+sv)$ is derivable at each $s \in [0,1]$ such that $y_2 v + sv \notin N(y_1)$, where $N(y_1)$ denotes the section of N at y_1,i.e.,

$$N(y_1) \;=\; \{y_2 v + sv \in R\cdot v \,|\, (y_1, y_2 v + sv) \in N\}.$$

By Proposition 2.5.5(d), there exists a subset $L' \subset G'$ with $G' \setminus L'$ Haar-null in G' and $N(y_1)$ is Lebesgue-negligible in Rv for each $y \in O$ such that $y_1 \in L'$. Therefore for each $y \in O$ such that $y_1 \in L'$ the function $s \mapsto \langle v, P(y+sv)\rangle$ is derivable for almost every s in $[0,1]$ and its derivative is given by $\langle v, DP(y+sv)v\rangle$. Thus for every $y \in O$ such that $y_1 \in L'$, applying the Fundamental Theorem of Calculus for the Lipschitz (implying absolute continuity) function $s \mapsto \langle v, P(y+sv)\rangle$, we obtain

$$\langle P(y+v) - P(y), v\rangle \;=\; \int_0^1 \langle v, DP(y+sv)v\rangle ds \qquad (2.5.1)$$

For such y and y_1, we then have

$$\langle P(y+v) - P(y), v\rangle \;\geq\; 0,$$

by our assumption and (2.5.1). Moreover the definition of a Haar-null set and Proposition 2.5.5 (b) and (c) imply that $M \cap (L' \times R\cdot v)$ is dense in $G' \oplus R\cdot v = O$. Therefore the required inequality is verified for each $y \in M \cap (L' \times R\cdot v)$ and by the continuity of P, is true for all y in O. This completes the proof of lemma. □

The convexity of e_λ itself has a full characterization in terms of subgradient mapping of f and its *proto-derivative.* To state it, we recall the following generalized notion of differentiation of set-valued mappings in terms of set convergence. A family of sets $C_n \subset X$ *Painlevé-Kuratowski* (PK) converges to C, denoted by $C_n \xrightarrow{pk} C$ if

$$\limsup C_n \;=\; \liminf C_n \;=\; C.$$

Here $\limsup C_n$ is the set of all accumulation points of sequences from the sets C_n and $\liminf C_n$ is the set of limit points of such sequences. For more on (PK) convergence see [**3**], [**6**], [**17**], [**41**], [**46**] and the reference therein.

We say that set-valued mapping $T : X \rightrightarrows X$ is *proto-differentiable* at a point x for an element $v \in T(x)$ if graphs of the set-valued mappings

$$\Delta_{x,v,t}T : \xi \mapsto \big[T(x + t\xi) - v\big]/t,$$

regarded as a family indexed by $t > 0$, Painlevé-Kuratowski (PK) converge as $t \searrow 0$. If so, the limit mapping is denoted by $T'^{(pk)}_{x,v}$ and called the *proto-derivative* of T at x for v; see [**6**], [**17**], [**19**], [**20**], [**41**], [**46**]. This proto-derivative mapping assigns to each $\xi \in X$ a subset $T'^{(pk)}_{x,v}(\xi)$ of X, which could be empty for some choices of ξ.

The following known results of proto-derivatives of set-valued mappings will be useful in the next several results of this section (cf. C. DO [**17**]).

Let $\Gamma : X \rightrightarrows X$ and $z \in \Gamma(x)$.

- Γ is monotone $\iff$ Γ^{-1} is monotone.
- Γ is monotone $\implies$ the proto-derivative mapping $\Gamma'^{(pk)}_{x,z}$ is monotone.
- Γ is proto-differentiable at x relative to z $\iff$ Γ^{-1} is proto-differentiable at z relative to x. One has $\left(\Gamma^{-1}\right)'^{(pk)}_{z,x} = \left(\Gamma'^{(pk)}_{x,z}\right)^{-1}$.
- Γ is locally single-valued and Hadamard differentiable at x $\implies$ it is proto-differentiable at x and $\Gamma'^{(pk)}_{x,\Gamma(x)} = D\Gamma(x)$, the Hadamard derivative of Γ at x.

A well known result of convex functions will be required:

- A Gâteaux differentiable function $f : X \to I\!R$ is convex $\iff$ the derivative mapping is monotone.

We now extend the characterization of convexity of e_λ of **[29]**, Proposition 5.4, to separable Hilbert spaces. The Lemma 2.5.7 plays a key role in establishing it.

Proposition 2.5.8. *Let $f : X \to \overline{I\!R}$, where X is a separable Hilbert space. Suppose that f is prox-regular at $\bar{x} = 0$ for $\bar{v} = 0$ with respect to ε and r, in particular with (2.4.1) holding, and let $\lambda \in (0, 1/r)$. Let T be the f-attentive ε-localization T of ∂f around $(0,0)$. Then the following conditions are equivalent:*

(a) The function e_λ is convex on a neighborhood of 0.

(b) There is a neighborhood U of $(0,0)$ such that if T_0 is the localization of T obtained by intersecting the graph of T with U, then $T_0^{-1} + \lambda I$ is monotone.

(c) There is a neighborhood U of $(0,0)$ such that at all points $(x,v) \in U \cap \operatorname{gph} T$ where T is proto-differentiable, the proto-derivative mapping $T'^{(pk)}_{x,v} : X \rightrightarrows X$ is such that $(T'^{(pk)}_{x,v})^{-1} + \lambda I$ is monotone.

(d) Same as (c) but with restriction to the points (x,v) where in addition the graph of $T'^{(pk)}_{x,v}$ is a linear subspace of $X \times X$.

Proof. The equivalence between (a) and (b) is easy to establish. Indeed, we have e_λ convex locally around 0 if and only if its the derivative mapping $De_\lambda = S_\lambda$ is monotone locally around the point $(0,0)$ in its graph, or equivalently, S_λ^{-1} has such local monotonicity. By Proposition 2.4.6 we have $S_\lambda = (\lambda I + T^{-1})^{-1}$, which means that ${S_\lambda}^{-1} = \lambda I + T^{-1}$. This gives the equivalence between (a) and (b). The local monotonicity of S_λ^{-1} implies that of its proto-derivative mappings where they exist. Proto-derivative mappings for S_λ^{-1} have the form $\lambda I + (T'^{(pk)}_{x,v})^{-1}$ in terms of proto-derivative mappings for T, and their monotonicity thus corresponds to the mappings $(T'^{(pk)}_{x,v})^{-1} + \lambda I$ being monotone. Thus we have (b) implies (c). Since (d) is a special case of (c), we also have (c) implies (d). We must show now that (d) implies (a). Condition (d) means that the mapping $(T'^{(pk)}_{x,v})^{-1} + \lambda I$ is monotone, or equivalently, $(De_\lambda)'^{(pk)}_{x+\lambda v} = \left((T'^{(pk)}_{x,v})^{-1} + \lambda I\right)^{-1}$ is monotone at points (x, v) near (0,0) where the proto-derivative of T exists as a linear mapping (may be set-valued). Because the mapping $S_\lambda = De_\lambda$ is Lipschitz continuous around 0, it is Gâteaux (hence Hadamard) differentiable a.e.(w.r.t. a Haar-null set in a neighborhood of 0), and hence in particular proto-differentiable with the proto derivative being the Gâteaux (same as Hadamard) derivative (a continuous linear operator)(cf. [**17**], Corollary 3.6). Thus, we have $(De_\lambda)'^{(pk)}_{x+\lambda v} = D(De_\lambda(x+\lambda v)) = DS_\lambda(x + \lambda v)$, for almost all points of $x + \lambda v$ near 0 (w.r.t. a Haar-null set). Then the monotonicity of $(De_\lambda)'^{(pk)}_{x+\lambda v}$ translate into the positive semidefiniteness of the Gâteaux derivative mapping $DS_\lambda(x+\lambda v)$. Then by Lemma 2.5.7, this is equivalent to the monotinicity of $S_\lambda = De_\lambda$ on a neighborhood of 0. This yields (a), and the proof is complete. $\square$

For strong monotonicity of e_λ, we have the following sufficient condition.

Proposition 2.5.9. *Suppose that f is prox-regular at $\bar{x} = 0$ for $\bar{v} = 0$ with respect to ε and r, and let $\lambda \in (0, 1/r)$. Let T be the f-attentive ε-localization T of ∂f around $(0,0)$. Suppose T is strongly monotone with modulus $\mu > 0$, i.e., $T - \mu I$ is monotone. Then, on some neighborhood of 0, one has the strong convexity of e_λ with modulus $\mu/(1 + \lambda\mu)$, i.e., the convexity of*

$$e_\lambda - \frac{\mu}{2(1 + \lambda\mu)}|\cdot|^2.$$

Proof. This follows from Lemma 2.5.1 for $\sigma = \mu$, because the derivative mapping of the function in question is $S_\lambda - [\mu/(1 + \lambda\mu)]I$ with S_λ the mapping given by the identity in Lemma 2.4.5. $\square$

Next we characterize the strong monotonicity of T in terms of its proto-derivative.

First we establish two lemmas. To state it, we recall a criterion for integrability of Banach-space-valued functions that can be defined by considering an associated one-dimensional integral (cf. Berger [5]).

Suppose a function $x(t)$ is defined on a measure space $(T, \mu, \sigma(T))$ with range in a Banach space X. Then a definition of integrability of $x(t)$ by duality is as follows.

Definition 2.5.10. *We say that $x(t)$ is integrable if there is an element $I_E(x) \in X$ for each element E of the σ-ring $\sigma(T)$ such that*

$$\langle x^*, I_E(x)\rangle = \int_E \langle x^*, x(t)\rangle d\mu \qquad \text{(in the Lebesgue sense)}$$

for each $x^* \in X^*$. We set $\int_E x(t)d\mu = I_E(x)$.

Lemma 2.5.11. *Let the vector function $t \mapsto p(t)$ from $[0,1]$ into a Hilbert space X be integrable on $[0,1]$. We then have*

$$\int_0^1 |p(t)|^2 dt \geq |\int_0^1 p(t)dt|^2.$$

Proof. Since the function $\varphi := |\cdot|^2$ is convex and continuous everywhere on X, the subgradient set $\partial\varphi(x)$ is nonempty for all x in X. Hence there exists v in X such that

$$\varphi(x) \geq \varphi(x_0) + \langle v, x - x_0\rangle \quad \text{for all} \quad x,$$

where $x_0 = \int_0^1 p(t)dt \in X$.

Setting $x = p(t)$ and integrating the above inequality we get

$$\int_0^1 \varphi(p(t))dt \geq \varphi(x_0) + \langle v, \int_0^1 p(t)dt - x_0\rangle \quad \text{i.e.,}$$

$$\int_0^1 |p(t)|^2 dt \geq |\int_0^1 p(t)dt|^2.$$

$\square$

Next lemma is the key to characterize the strong monotonicity of T, in which we extend the results of [**29**], Lemma 5.6 to separable Hilbert spaces.

Lemma 2.5.12. *Suppose P is a Lipschitz continuous mapping from an open convex set O of a separable Hilbert space X into X. Then the following conditions are equivalent for any $\alpha > 0$.*

(a) $P^{-1} - \alpha I$ is monotone.

(b) For all $y \in O$ where P is proto-differentiable, the proto-derivative mapping $P_y^{\prime(pk)}$ is such that $(P_y^{\prime(pk)})^{-1} - \alpha I$ is monotone.

(c) For all $y \in O$ where P is Gâteaux differentiable, the Gâteaux derivative $DP(y)$ satisfies

$$\langle \eta, DP(y)\eta \rangle \ \geq \ \alpha \big| DP(y)\eta \big|^2 \text{ for all } \eta \in X.$$

Proof. Condition (a) implies condition (b) through the fact that the proto-derivative of a monotone mapping, if it exists, is another monotone mapping. Since for Lipschitz mappings Gâteaux and Hadamard derivatives coincide, when P is Gâteaux differentiable, it is proto differentiable and the proto-derivative coincides with its Gâteaux (same as Hadamard) derivative (cf. [**17**], Corollary 3.6), and hence we have condition (b) implies condition (c). We must show now that condition (c) implies condition (a). Condition (a) means that

$$\big\langle P(y+v) - P(y), v \big\rangle \geq \alpha \big| P(y+v) - P(y) \big|^2 \quad \text{for all} \quad y, y+v \in O.$$

Since P is Lipschitz on O, from (2.5.1), we have

$$\langle P(y+v) - P(y), v \rangle \ = \ \int_0^1 \langle v, DP(y+sv)v \rangle ds.$$

where $y \in O$ and $y_1 \in L'$ as in Lemma 2.5.7. This implies

$$\begin{aligned}\langle P(y+v) - P(y), v\rangle &\geq \alpha \int_0^1 \left|DP(y+sv)v\right|^2 ds \\ &\geq \alpha\left|\int_0^1 DP(y+sv)vds\right|^2 \\ &= \alpha\left|P(y+v) - P(y)\right|^2.\end{aligned}$$

where the inequalities are based on the assumptions in (c) and the Lemma 2.5.11, respectively. Then the result follows for all $y, y+v \in O$, as proved in Lemma 2.5.7.
□

In the following, we extend the characterization of strong monotonicity of T, given in [**29**], Proposition 5.7, to separable Hilbert spaces.

Proposition 2.5.13. *Let $f : X \to \overline{\mathbb{R}}$, where X is a separable Hilbert space. Suppose that f is prox-regular at $\bar{x}$ for $\bar{v}$ with respect to ε and r. Let T be the f-attentive ε-localization T of ∂f around $(\bar{x}, \bar{v})$. Then the following conditions on T and a value $\mu > 0$ are equivalent:*

(a) T is strongly monotone with modulus μ locally around the point $(\bar{x}, \bar{v}) \in$ gph T.

(b) There is a neighborhood U of $(\bar{x}, \bar{v})$ such that at all points $(x, v) \in U \cap \operatorname{gph} T$ where T is proto-differentiable, the proto-derivative mapping $T'^{(pk)}_{x,v} : X \rightrightarrows X$ is strongly monotone with modulus μ.

(c) Same as (b) but with restriction to the points (x, v) where in addition the graph of $T'^{(pk)}_{x,v}$ is a linear subspace of $X \times X$.

Proof. Without any loss of generality we may reduce to the case $\bar{x} = 0 = \bar{v}$ with (2.4.1) holding (see Section 2.4). We have (a) implies (b), applying the fact, proto-derivative of a monotone mapping, if it exists, is another monotone mapping, for $T - \mu I$. Since (c) is a special case of (b), we also have (b) implies (c). We must show now that (c) implies (a). Consider any $\rho > r$ such that $\rho + \mu > 0$, where r is a local constant from the definition of prox-regularity. Let $\alpha = \rho + \mu$, $P = (T + \rho I)^{-1}$, and $M = T + rI$. Since M is a maximal monotone mapping in graph around $(0,0)$ (cf. Proposition 2.4.9), and $P = \left(M + (\rho - r)I\right)^{-1}$ with $\rho - r > 0$, by [**2**], Theorem 3.5.9, P is Lipschitz continuous on some neighborhood of 0. Condition (c) means that $T'^{(pk)}_{x,v} - \mu I$ monotone, or equivalently, $(T + \rho I)'^{(pk)}_{x,v+\rho x} - \alpha I = (P'^{(pk)}_{v+\rho x})^{-1} - \alpha I$ is monotone at points $(x,v) \in \operatorname{gph} T$ near $(0,0)$ where the proto-derivative of T exists as a linear mapping (may be set-valued). Because the mapping P is Lipschitz continuous around 0, it is Gâteaux (hence Hadamard) differentiable a.e.(w.r.t. a Haar-null set) around 0, and hence in particular proto-differentiable with the proto-derivative being the Gâteaux derivative (continuous linear operator) (cf. [**17**], Corollary 3.6). Then by Lemma 2.5.12, condition (c) is equivalent to the monotonicity of $P^{-1} - \alpha I = (T + \rho I) - \alpha I = T - \mu I$ at points $(x,v) \in \operatorname{gph} T$ near (0,0). □

Corollary 2.5.14. *Let $f : X \to \overline{I\!R}$, where X is a separable Hilbert space. Suppose that f is prox-regular at $\bar{x}$ for $\bar{v}$ with respect to ε and r. Let T be the f-attentive ε-localization T of ∂f around $(\bar{x}, \bar{v})$. Then the following conditions on T are equivalent:*

(a) T is monotone locally around the point $(\bar{x}, \bar{v}) \in \operatorname{gph} T$.

(b) There is a neighborhood U of $(\bar{x}, \bar{v})$ such that at all points $(x,v) \in U \cap \operatorname{gph} T$ where T is proto-differentiable, the proto-derivative mapping $T'^{(pk)}_{x,v} : X \rightrightarrows X$ is monotone.

(c) Same as (b) but with restriction to the points (x,v) where in addition the graph of $T'^{(pk)}_{x,v}$ is a linear subspace of $X \times X$.

Proof. Apply Proposition 2.5.13 to $T_\mu = T + \mu I$ for all $\mu > 0$. □

Remark 2.5.15. *If the function f is also subdifferentially continuous, then all results in this section concerning T as an f-attentive localization of ∂f at $(\bar{x}, \bar{v})$ can be restated in terms of T being an ordinary localization.*

2.6. Second-Order Theory

It's time now for a closer look at the classical idea of obtaining second derivatives by differentiating first derivatives. How might this fit into the framework of "generalized second-order" differentiation of prox-regular functions ? We answer this question in Theorem 2.6.4.

First we recall some terminology :

A family of functions $f_n : X \to \overline{I\!R}$ *Mosco epi-converges* to f, denoted by $f_n \xrightarrow{m} f$, if f_n *strongly and weakly epi-converges* to f, i.e., the epigraph of f_n (PK) converges to the epigraph of f in both the weak and strong topologies. See [**1**], [**6**], [**17**], and [**19**]. In other words, we have for all x

$$f(x) \leq \liminf f_n(x_n) \qquad \text{whenever} \qquad x_n \xrightarrow{w} x$$

and

$$\text{there exists} \qquad x_n \to x \qquad \text{with} \qquad f(x) \geq \limsup f_n(x_n).$$

We will say that f_n Mosco epi-converges to f on $C \subset X$ if for all $x \in C$

$$f(x) \leq \liminf f_n(x_n) \quad \text{whenever} \quad x_n \xrightarrow{w} x \quad \text{and} \quad \{x_n\} \subset C$$

and

$$\text{there exists} \quad x_n \to x \quad \text{with} \quad \text{with} \quad \{x_n\} \subset C \quad \text{and} \quad f(x) \geq \limsup f_n(x_n).$$

Recall that a function f is *twice Mosco epi-differentiable* at $\bar{x}$ for a vector $\bar{v} \in \partial f(\bar{x})$ if the second-order difference quotient functions $\Delta^2_{\bar{x},\bar{v},t} f : X \to \overline{I\!R}$, defined by

$$\Delta^2_{\bar{x},\bar{v},t} f(\xi) = \big[f(\bar{x} + t\xi) - f(\bar{x}) - t\langle \bar{v}, \xi\rangle\big] \big/ \tfrac{1}{2} t^2 \text{ for } t > 0,$$

Mosco epi-converge to a proper function as $t \searrow 0$. The Mosco epi-limit is then the *second Mosco epi-derivative* function $f''^{(m)}_{\bar{x},\bar{v}} : X \to \overline{I\!R}$. see [**6**], [**17**], [**19**]. This function, when it exists, is sequentially weakly l.s.c., proper and positively homogeneous of degree 2.

When X is finite-dimensional, the weak convergence in the definition is replaced with strong convergence, and hence we drop the prefix "Mosco" in the terminology. We simply say epi-convergence and epi-differentiable appropriately in the definition. For more on epi-derivatives see [**27**], [**40**], [**46**].

In this section, we establish the connection between the epi-differentiability of a prox-regular function and the proto-differentiability of its subdifferential mapping with a natural formula relating these two derivatives, in the context of Hilbert spaces.

We will need the following results:

Proposition 2.6.1. *Let $\varphi_n : X \to \overline{I\!R}$ be a family of l.s.c. functions equi-bounded below near $\bar{x}$ (i.e. $\inf\limits_{n \in I\!N} \inf\limits_{x \in B(\bar{x},r)} \{\varphi_n(x)\} > -\infty$) with $\{\varphi_n(\bar{x})\}$ bounded. Assume further that $\{\varphi_n\}$ Mosco epi-converges to φ on some neighborhood of $\bar{x}$. Then there exist $0 < r_1 < r_2$ such that for all $\lambda > 0$ small enough $(\varphi_n + \delta_{B(\bar{x},r_2)})_\lambda$ Mosco-epi converges to $(\varphi + \delta_{B(\bar{x},r_2)})_\lambda$ on $I\!B(\bar{x}, r_1)$, where $(\varphi_n + \delta_{B(\bar{x},r_2)})_\lambda$ denotes the Moreau λ-envelope of $\varphi_n + \delta_{B(\bar{x},r_2)}$.*

Proof. See Levi, Poliquin and Thibault ([**19**], Proposition 3.3). □

Proposition 2.6.2. *(sum rule) Let $f : X \to \overline{I\!R}$ be twice Mosco epi-differentiable at x for $v \in \partial f(x)$, and g be any C^2 function on X with the mapping $\xi \to \langle D^2 g(x)\xi, \xi\rangle$ is weakly lower semicontinuous. Then the function $h = f + g$ is twice*

Mosco epi-differentiable at x with

$$h''^{(m)}_{x,u}(\xi) = f''^{(m)}_{x,v}(\xi) + \langle D^2 g(x)\xi, \xi\rangle,$$

where $u = v + Dg(x)$, $v \in \partial f(x)$.

Proof. See author's M.Sc. thesis ([**6**], Proposition 3.2.5) $\square$

Theorem 2.6.3. *(Attouch's theorem) Let $\{\varphi_n\}$, φ be a sequence of l.s.c. proper convex functions on X. Then $\varphi_n \xrightarrow{m} \varphi$ if and only if the following conditions hold:*

(i) $\operatorname{gph}\partial\varphi_n \xrightarrow{pk} \operatorname{gph}\partial\varphi$.

(ii) $\exists(\xi,\eta) \in \operatorname{gph}\partial\varphi$, $\exists(\xi_n,\eta_n) \in \operatorname{gph}\partial\varphi_n$ *such that* $(\xi_n,\eta_n) \to (\xi,\eta)$ *and* $\varphi_n(\xi_n) \to \varphi(\xi)$.

Proof. See Attouch's book ([**1**], Theorem 3.66). $\square$

In $\mathbb{R}^n$, Poliquin and Rockafellar established the relationship between the second-order epi-derivative of a prox-regular function and the proto-derivative of its subgradient mapping ([**29**],Theorem 6.1). Our next theorem gives a partial extension of that result in the context of a Hilbert space.

Theorem 2.6.4. *Assume that $f : X \to \overline{\mathbb{R}}$ is prox-regular at $\bar{x}$ for $\bar{v}$ with constants ε and r. Let T be the f-attentive ε-localization of ∂f around $(\bar{x},\bar{v})$. If f is twice Mosco epi-differentiable at $\bar{x}$ for $\bar{v}$, then T is proto-differentiable at $\bar{x}$ for $\bar{v}$. One has*

$$T'^{(pk)}_{\bar{x},\bar{v}}(\xi) = \partial\big[\tfrac{1}{2} f''^{(m)}_{\bar{x},\bar{v}}\big](\xi) \quad \text{for all } \ \xi.$$

The converse is true when X is a finite-dimensional space.

Proof. Without loss of generality we can suppose that $\bar{x} = 0$, $\bar{v} = 0$, $f(0) = 0$ with (2.4.1) holding (see Section 2.4). In addition we may assume, without any loss of generality, f to be l.s.c. on X with the domain of f is included in the closed ball of radius ε, since that can be manufactured out of the local l.s.c. property by adding the indicator function of the set $\overline{\mathbb{B}}(0,\varepsilon)$ to f. Consider any $\lambda \in (0, 1/r)$ and the function

$$\hat{e}_\lambda(x) := e_\lambda(x) + \frac{r}{2(1-\lambda r)}|x|^2. \tag{2.6.1}$$

There is a neighborhood of 0 on which this function is $\mathcal{C}^{1+}$ by Theorem 2.4.4 and convex by Theorem 2.5.2, the derivative mapping being

$$D\hat{e}_\lambda \;=\; De_\lambda + \frac{r}{1-\lambda r}I. \tag{2.6.2}$$

Let

$$f_{\bar{x},\bar{v},t}(\xi) := \frac{f(\bar{x}+t\xi) - f(\bar{x}) - t\langle \bar{v}, \xi\rangle}{t^2}, \qquad \text{where} \quad t > 0.$$

Because f is prox-regular at $\bar{x} = 0$ for $\bar{v} = 0$, we have $f(\bar{x}+t\xi) - f(\bar{x}) - \langle \bar{v}, \xi\rangle \geq -\frac{r}{2}|t\xi|^2$ for all ξ, and hence $f_{\bar{x},\bar{v},t}(\xi) = \frac{f(t\xi)}{t^2} \geq -\frac{r}{2}|\xi|^2$. Then, there exists $\rho > 0$ and t small enough such that the functions $f_{\bar{x},\bar{v},t}$ are equi-bounded below on $\overline{\mathbb{B}}(0,\rho)$ and $\operatorname{dom} f_{\bar{x},\bar{v},t} \subset \overline{\mathbb{B}}(0,\rho)$ with $f_{\bar{x},\bar{v},t}(0) = 0$.

Since we assumed that f is twice Mosco epi-differentiable at $\bar{x} = 0$ for $\bar{v} = 0$, i.e., $f_{\bar{x},\bar{v},t} \xrightarrow{m} \frac{1}{2}f''^{(m)}_{\bar{x},\bar{v}}$, applying Proposition 2.6.1 there exists r_1, $0 < r_1 < \rho$, such that for all λ small enough, the Moreau λ-envelopes

$$\left(f_{\bar{x},\bar{v},t}\right)_\lambda \xrightarrow{m} \left(\tfrac{1}{2}f''^{(m)}_{\bar{x},\bar{v}}\right)_\lambda \quad \text{on} \quad \mathbb{B}(0, r_1).$$

Observe that

$$\begin{aligned}
\left(f_{\bar{x},\bar{v},t}\right)_{\lambda}(\xi) &= \inf_{\xi'}\left\{f_{\bar{x},\bar{v},t}(\xi') + \frac{1}{2\lambda}\left|\xi'-\xi\right|^2\right\} \\
&= \inf_{\xi'}\left\{\frac{f(t\xi')}{t^2} + \frac{1}{2\lambda}\left|\xi'-\xi\right|^2\right\} \\
&= \frac{1}{t^2}\inf_{\xi'}\left\{f(t\xi') + \frac{1}{2\lambda}\left|t\xi'-t\xi\right|^2\right\} \\
&= \frac{1}{t^2}e_{\lambda}(t\xi) \\
&= \frac{e_{\lambda}(\bar{x}+t\xi) - e_{\lambda}(\bar{x}) - t\langle De_{\lambda}(\bar{x}),\xi\rangle}{t^2},
\end{aligned}$$

which Mosco epi-converge to $\frac{1}{2}\left(e_{\lambda}\right)^{\prime\prime(m)}_{\bar{x},\bar{v}}$. In otherwords, for λ small enough, e_{λ} is twice Mosco epi-differentiable at $\bar{x}=0$ for $\bar{v}=0$ with

$$\tfrac{1}{2}\left(e_{\lambda}\right)^{\prime\prime(m)}_{\bar{x},\bar{v}} = \left(\tfrac{1}{2}f^{\prime\prime(m)}_{\bar{x},\bar{v}}\right)_{\lambda} \quad \text{on} \quad \mathbb{B}(0,r_1).$$

It follows from the formulas (2.6.1) and (2.6.2) and the sum rule (Proposition 2.6.2) that $\hat{e}_{\lambda}$ is twice Mosco epi-differentiable at $\bar{x}=0$ for $\bar{v}=0$ with

$$(\hat{e}_{\lambda})^{\prime\prime(m)}_{\bar{x},\bar{v}}(\xi) = (e_{\lambda})^{\prime\prime(m)}_{\bar{x},\bar{v}}(\xi) + \frac{r}{(1-\lambda r)}|\xi|^2 \quad \text{on} \quad \mathbb{B}(0,r_1) \tag{2.6.3}$$

Convexity of $\hat{e}_{\lambda}$ ensures (cf. [**17**], Theorem 3.9) that the twice Mosco epi-differentiability of $\hat{e}_{\lambda}$ at $\bar{x}=0$ for $\bar{v}=0$ is equivalent to the proto-differentiability of $D\hat{e}_{\lambda}$ at $\bar{x}=0$ for $\bar{v}=0$ with

$$\begin{aligned}
\partial\left[\tfrac{1}{2}(\hat{e}_{\lambda})^{\prime\prime(m)}_{\bar{x},\bar{v}}\right](\xi) &= (D\hat{e}_{\lambda})^{\prime(pk)}_{\bar{x},\bar{v}}(\xi) \\
&= (De_{\lambda})^{\prime(pk)}_{\bar{x},\bar{v}}(\xi) + \frac{r}{(1-\lambda r)}\xi \quad \text{on} \quad \mathbb{B}(0,r_1)
\end{aligned} \tag{2.6.4}$$

Hence we have the proto-differentiability of De_{λ} at $\bar{x}=0$ for $\bar{v}=0$.

But De_{λ} has been identified locally with $[\lambda I + T^{-1}]^{-1}$ in Proposition 2.4.6. The graph of the latter mapping is the image of the graph of T under the invertible linear transformation $(x,v)\mapsto(x+\lambda v, v)$ from $X\times X$ onto itself. Since proto-differentiability at $\bar{x}$ for $\bar{v}$ is a geometric property of graphs at $(\bar{x}, \bar{v})$ that is maintained when graphs are subjected to an invertible linear transformation, and the proto-derivative mappings themselves then correspond under the same transformation, we deduce that the proto-differentiable of De_λ at $\bar{x}$ for $\bar{v}$ is equivalent to that of T on $\mathbb{B}(0, r_1)$, in which event there is the formula

$$(De_\lambda)'^{(pk)}_{\bar{x},\bar{v}} = \left[\lambda I + S^{-1}\right]^{-1} \text{ with } S = T'^{(pk)}_{\bar{x},\bar{v}} \text{ on } \mathbb{B}(0, r_1). \tag{2.6.5}$$

Since the proto-derivative mapping is positively homogenuous, the above equivalence (and 2.6.5) is true everywhere. This complete the first part of the proof.

We now turn to verify the derivative formula in the theorem. The positive homogeneity of the derivative mappings involved in the formulas (2.6.3), (2.6.4) and (2.6.5) imply that they are actually valid everywhere and hence combining them yield

$$\partial\left[\tfrac{1}{2}(e_\lambda)''^{(m)}_{\bar{x},\bar{v}} + \frac{r}{2(1-\lambda r)}|\cdot|^2\right](\xi) = \left[\lambda I + S^{-1}\right]^{-1}(\xi) + \frac{r}{(1-\lambda r)}\xi \quad \text{for all } \xi.$$

Thus

$$\partial\left[\tfrac{1}{2}(e_\lambda)''^{(m)}_{\bar{x},\bar{v}}\right](\xi) = \left[\lambda I + S^{-1}\right]^{-1}(\xi) \text{ with } S = T'^{(pk)}_{\bar{x},\bar{v}}, \tag{2.6.6}$$

and hence

$$\operatorname{gph}\partial\left(\tfrac{1}{2}(e_\lambda)''^{(m)}_{\bar{x},\bar{v}}\right) \xrightarrow{pk} \operatorname{gph} T'^{(pk)}_{\bar{x},\bar{v}} \text{ as } \lambda \searrow 0. \tag{2.6.7}$$

The convexity of Mosco epi-limit of convex functions (cf. [**17**], Proposition 2.2) implies that of $\frac{1}{2}(e_\lambda)''^{(m)}_{\bar{x},\bar{v}}(\xi) + \frac{r}{2(1-\lambda r)}|\xi|^2$ (through 2.6.3). Hence the functions $\frac{1}{2}(e_\lambda)''^{(m)}_{\bar{x},\bar{v}}(\xi) + \frac{r}{2}|\xi|^2$ are convex for sufficiently small λ, which indeed increase to $\frac{1}{2}f''^{(m)}_{\bar{x},\bar{v}}(\xi) + \frac{r}{2}|\xi|^2$ as $\lambda \searrow 0$. Thus

$$\tfrac{1}{2}(e_\lambda)''^{(m)}_{\bar{x},\bar{v}}(\xi) + \frac{r}{2}|\xi|^2 \xrightarrow{m} \tfrac{1}{2}f''^{(m)}_{\bar{x},\bar{v}}(\xi) + \frac{r}{2}|\xi|^2 \text{ as } \lambda \searrow 0.$$

Then, by Attouch's theorem (Theorem 2.6.3) for convex functions,

$$\operatorname{gph}\left\{\partial\big(\tfrac{1}{2}(e_\lambda)''^{(m)}_{\bar{x},\bar{v}}\big)+rI\right\}\xrightarrow{pk}\operatorname{gph}\left\{\partial\big(\tfrac{1}{2}f''^{(m)}_{\bar{x},\bar{v}}\big)+rI\right\}\quad\text{as}\quad\lambda\searrow 0. \tag{2.6.8}$$

From (2.6.7) and (2.6.8) we conclde that $T'^{(pk)}_{\bar{x},\bar{v}}(\xi)=\partial\big[\tfrac{1}{2}f''^{(m)}_{\bar{x},\bar{v}}\big](\xi)$ for all ξ, as required.

When X is finite-dimensional, the converse of Propositions 2.6.1 (see [**25**], Proposition 2.1) and 2.6.2 (see [**40**], Proposition 2.10) are true, and hence the proof given here can easily be reversed. □

Corollary 2.6.5. *Assume that $f: X\to\overline{\mathbb{R}}$ is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v}$ with constants ε and r. If f is twice Mosco epi-differentiable at $\bar{x}$ for $\bar{v}$, then ∂f is proto-differentiable at $\bar{x}$ for $\bar{v}$. One has*

$$(\partial f)'^{(pk)}_{\bar{x},\bar{v}}(\xi)=\partial\big[\tfrac{1}{2}f''^{(m)}_{\bar{x},\bar{v}}\big](\xi)\quad\text{for all}\quad\xi.$$

The converse is true when X is a finite-dimensional space.

Proof. Just apply the theorem noting that the f-attentiveness in the localization of ∂f to T is superfluous here. □

For a convex, C^2 function the above derivative formula agrees with the classical results of second derivatives.

Corollary 2.6.6. *For a convex, C^2 function $f: X\to\mathbb{R}$ one has*

$$\begin{aligned} f''^{(m)}_{x,Df(x)}(\xi) &= \langle D^2 f(x)\xi,\xi\rangle,\\ (Df)'^{(pk)}_{x,Df(x)}(\xi) &= D^2 f(x)\xi,\quad \xi\in X \end{aligned}$$

and hence the derivative formula in Theorem 2.6.4 holds.

Proof. See [**17**], Proposition 4.1. □

The proof of Theorem 2.6.4 has revealed additional facts concerning f and the second-order properties of its Moreau envelopes e_λ, which we record next.

Theorem 2.6.7. *Suppose that f is prox-regular at $\bar{x} = 0$ for $\bar{v} = 0$ with respect to ε and r, in particular with (2.4.1) holding, and let $\lambda \in (0, 1/r)$. If f is twice Mosco epi-differentiable at 0 for 0, then e_λ has this property. One then has $\frac{1}{2}(e_\lambda)''^{(m)}_{0,0}$ as the Moreau λ-envelope of $\frac{1}{2}f''^{(m)}_{0,0}$, and the function $f''^{(m)}_{0,0} + r|\cdot|^2$ is nonnegative and convex with*

$$\partial\big[\tfrac{1}{2}(e_\lambda)''^{(m)}_{0,0}\big] = \big[\lambda I + S^{-1}\big]^{-1} = \lambda^{-1}\big[I - (I + \lambda S)^{-1}\big] \quad \text{for} \quad S := \partial\big[\tfrac{1}{2}f''^{(m)}_{0,0}\big].$$

Proof. Follows readily from the proof of Theorem 2.6.4. ■

Corollary 2.6.8. *Suppose that f is prox-regular at $\bar{x} \in \operatorname{argmin} f$ for $\bar{v} = 0$ with respect to ε and r. Let T be the f-attentive ε-localization of ∂f around $(\bar{x}, \bar{v})$. Assume that there is a neighborhood U of $(\bar{x}, \bar{v})$ and $\bar{\lambda} > 0$ such that at all points $(x, v) \in U \cap \operatorname{gph} T$, and for all $0 < \lambda < \bar{\lambda}$, f is twice Mosco epi-differentiable at x for v, then e_λ has this property at $x + \lambda v$ for v. One then has*

$$\partial\big[\tfrac{1}{2}(e_\lambda)''^{(m)}_{x+\lambda v, v}\big] = \big[\lambda I + S^{-1}\big]^{-1} = \lambda^{-1}\big[I - (I + \lambda S)^{-1}\big] \quad \text{for} \quad S := \partial\big[\tfrac{1}{2}f''^{(m)}_{x,v}\big].$$

Proof. Assume that $\bar{x} = 0$ with $f(0) = 0$. Consider $(\tilde{x}, \tilde{v}) \in \operatorname{gph} T$ and the function $\tilde{f}(x) := f(x + \tilde{x}) - f(\tilde{x}) - \langle \tilde{v}, x \rangle$. There is a neighborhood U of $(0, 0)$ and $R > r$ such that for all points $(\tilde{x}, \tilde{v}) \in U \cap \operatorname{gph} T$, we have $\tilde{f}(x) \geq -(R/2)|x|^2$ for all x (see [**29**], Corollary 6.6).

It is very easy to verify that for $0 < \lambda < (1/R)$

$$\tilde{e}_\lambda(w) = e_\lambda(w + \tilde{x} + \lambda\tilde{v}) - \langle w, \tilde{v} \rangle - f(\tilde{x}) - (\lambda/2)|\tilde{v}|^2,$$

here $\tilde{e}_\lambda$ is the Moreau λ-envelope of $\tilde{f}$. From this we conclude that $(\tilde{e}_\lambda)''^{(m)}_{0,0} = (e_\lambda)''^{(m)}_{\tilde{x}+\lambda\tilde{v},\tilde{v}}$. Finally note that $De_\lambda(\tilde{x}+\lambda\tilde{v}) = \tilde{v}$. (because $\tilde{v} \in \partial f(\tilde{x})$), $\tilde{f}''^{(m)}_{0,0} = f''^{(m)}_{\tilde{x},\tilde{v}}$, and that $\tilde{f}$ is prox-regular at 0 for 0 with respect to ε and R, in particular with

(2.4.1) holding ($\bar{f}$ in place of f and R in place of r). Now simply apply Theorem 2.6.7 to the function $\bar{f}$. ◻

Not every property of prox-regular functions in finite-dimensional spaces has a Hilbert space extension. We recall in Theorem 2.6.9 that a prox-regular function in finite-dimensional spaces has a second-order expansion. We conclude Chapter 2 by giving an example which illustrates that even for convex functions in Hilbert space this property does not hold.

Theorem 2.6.9. *([29], Theorem 6.7) Let X be a finite dimensional space. Suppose $f : X \to \overline{I\!R}$ is prox-regular at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ with constants ε, r, and also that f is twice epi-differentiable at $\bar{x}$ for $\bar{v}$. If the second-order epi-derivative function $f''_{\bar{x},\bar{v}}$ is finite on a neighborhood of 0, it must actually be a lower-$\mathcal{C}^2$ function. Then f must itself be lower-$\mathcal{C}^2$ around $\bar{x}$, differentiable at $\bar{x}$ with $Df(\bar{x}) = \bar{v}$, and the second-order difference quotient functions $\Delta^2_{\bar{x},\bar{v},t} f$ not only epi-converge to $f''_{\bar{x},\bar{v}}$, but converge uniformly on all bounded sets. In other words, one has the expansion*

$$f(x) \;=\; f(\bar{x}) + \langle \bar{v},\, x - \bar{x}\rangle + f''_{\bar{x},\bar{v}}(x - \bar{x}) + o\big(|x - \bar{x}|^2\big).$$

The following example shows that the extension of above theorem to Hilbert spaces fails.

Example 2.6.10. *(Borwein and Noll [8]) Let $f : l_2 \to I\!R$ be the convex function*

$$f(x) = \sum_{n=1}^{\infty} f_n(x_n), \qquad x = (x_n) \in l_2, \tag{2.6.9}$$

where $f_n(\eta) = n^{-\alpha}|\eta|$ $\left(\frac{1}{2} < \alpha < 1\right)$.

First notice that convexity of f ensures the prox-regularity of f everywhere in l_2. Consider the point $x = (n^{-2}) \in l_2$. Then f is twice Mosco epi-differentiable at x with

$$f_x^{\prime\prime(m)}(\xi) = \frac{1}{2}\sum_{n=1}^{\infty} f_n''(x_n)\xi_n^2 = 0 \quad \text{for all} \quad \xi \in l_2.$$

However, the second-order difference quotient fails to converge to $f_x^{\prime\prime(m)}$. In fact, if $\Delta^2_{x,v,t}f \to 0$ pointwise, then f had to be Lipschitz smooth at x (cf. Borwein and Noll [8], Proposition 2.2). Borwein and Noll showed that this is not the case. See Borwein and Noll [8], Example 2, pp 62 for details.

CHAPTER 3

INTEGRATION OF PROX-REGULAR FUNCTIONS

3.1. Integration Problem

In this chapter, we study the fundamental problem of determining functions that can be recovered up to an additive constant, from the knowledge of their subgradients. More precisely, a function f is deemed integrable if whenever $\partial_{\#}g(x) = \partial_{\#}f(x)$ for all x then f and g differ only by an additive constant. Here $\partial_{\#}$ refers to a subdifferential, which can be taken in many different ways (e.g. Dini subdifferential, Clarke subdifferential, b-subdifferential, Michel-Penot subdifferential, Mordukhovich subdifferential, Ioffe approximate subdifferential, Frechet subdifferential, and proximal subdifferential).

The scope of the (non-differentiable) functions that are deemed integrable seems somewhat restricted. It is clear that not every function can be recovered, up to an additive constant, from its subgradients. We only need to look at the following two functions.

Example 3.1.1.

Let

$$f(x) = \begin{cases} 0 & \text{if } x \le 0, \\ 1 & \text{if } x > 0, \end{cases} \qquad g(x) = \begin{cases} 0 & \text{if } x \le 0, \\ 2 & \text{if } x > 0, \end{cases}$$

then

$$\partial f(x) = \partial g(x) = \begin{cases} \{0\} & \text{if } x \neq 0, \\ [0, \infty) & \text{if } x = 0. \end{cases}$$

These two functions have the same subgradients everywhere yet they differ by different constants in the pieces of the (connected) domain.

It has been conjectured that the locally Lipschitzian functions can be recovered from their proximal subgradients. This is due to a theorem of Rademacher, a locally Lipschitzian function is differentiable almost everywhere. Recently, this was proven negative by Benoist.

Example 3.1.2. *(Benoist [4])*
For every countable dense set $D \subset \mathbb{R}$, there exists infinitely many Lipschitzian functions f, differing by more than a constant, such that

$$\partial_p f(x) = \begin{cases} (-1, +1) & \text{if } x \in D, \\ \emptyset & \text{if } x \notin D. \end{cases}$$

However, it was proved in [**38**] that this undesirable situation does not arise for some important classes of locally Lipschitzian functions such as the upper regular, semismooth and separably regular functions.

Probably the most well known and the oldest result in this area concerns convex functions. If two l.s.c. convex functions (defined on Banach spaces) have the same subgradients, then they differ by a constant; see Rockafellar [**33**]. However, very few other examples were known.

The first work outside the field of locally Lipschitzian functions was done by Poliquin for the p.l.n. functions. If two functions are p.l.n. at $\bar{x}$ and have the same subgradients, then on a neighborhood of $\bar{x}$ the functions differ by a constant. See [**24**]. Later this result was extended to Hilbert spaces by Thibault and Zagrodny

[48]. The contribution we make to the integration problem is to identify a large class of prox-regular functions, which differ only by a constant, from the knowledge of their limiting subgradients. We establish the integration result in an arbitrary Hilbert space, and certainly it applies to a much wider territory than that of p.l.n. case (See Example 3.3.2). The central tool in achieving this integration result is the smoothness property of the Moreau envelopes of prox-regular functions that we established in Chapter 2.

3.2. Main Result

We prove that if two functions, which have the same subgradients locally, are prox-regular and subdifferentially continuous relative to a pair $(\bar{x}, \bar{v})$ then the functions differ by a constant in a local neighborhood of $(\bar{x}, \bar{v})$. More precisely, we have:

Theorem 3.2.1. *Let $f_i : X \to \overline{I\!R}$ be prox-regular at $\bar{x}$ for $\bar{v} \in \partial f_i(\bar{x})$, $i = 1, 2$. Assume that there exists a neighborhood of $\bar{x}$ such that both f_1 and f_2 have the same limiting subgradients and f_1 is subdifferentially continuous at $\bar{x}$ for $\bar{v}$. Then f_2 is subdifferentially continuous at $\bar{x}$ for $\bar{v}$, and there is a k in $I\!R$ such that $f_1(x) = f_2(x) + k$ for all x near $\bar{x}$ with v in $\partial f_i(x)$ close to $\bar{v}$.*

Proof. Without loss of generality (cf. 2.4.1) we normalize to the case $\bar{x} = 0$, $\bar{v} = 0$ with

$$\left.\begin{array}{r} f_i \text{ is locally l.s.c. at } 0 \text{ with } f_i(0) = 0, \text{ and } r > 0 \\ \text{is such that } f_i(x) \geq -\dfrac{r}{2}|x|^2 \text{ for all } x, \text{ and } i = 1, 2 \end{array}\right\} \tag{3.2.1}$$

which imply that

$$e^i_\lambda(0) = 0 \text{ and } P^i_\lambda(0) = \{0\} \text{ when } \lambda \in (0, 1/r) \text{ and } i = 1, 2, \tag{3.2.2}$$

where e^i_λ and P^i_λ are the Moreau envelope function and the proximal mapping of f_i, respectively.

We may further assume that there exists $\varepsilon > 0$ such that f_1 and f_2 are prox-regular at $\bar{x} = 0$ for $\bar{v} = 0$ with respect to the same r with (3.2.1) holding. For $i = 1, 2$ let T_i be the f_i-attentive ε-localization of ∂f_i around $(0, 0)$. Then, by Theorem 2.4.4, for each $\lambda \in (0, 1/r)$ and $i = 1, 2$ there exists $\delta > 0$ such that, on $V := \{x\,;\,|x| < \delta\}$, the mappings P_λ^i are single-valued and Lipschitz continuous with constant $1/(1 - \lambda r)$ and

$$P_\lambda^i(x) \;=\; (I + \lambda T_i)^{-1}(x) \;=\; [\text{singleton}], \tag{3.2.3}$$

while the functions e_λ^i is of class C^{1+} with $De_\lambda^i(0) = 0$ and

$$De_\lambda^i(x) = \frac{x - P_\lambda^i(x)}{\lambda} = \lambda^{-1}\left[I - [I + \lambda T_i]^{-1}\right](x), \tag{3.2.4}$$

and the properties in Propositions 2.4.2 and 2.4.3 hold.

Decreasing ε further if necessary, we can arrange that f_1 and f_2 have the same subgradients on $\varepsilon I\!B$, where $\varepsilon > 0$ comes from the definition of prox-regularity of f_i.

Claim 1. *For each $\lambda \in (0, 1/r)$, we have $P_\lambda^1(x) = P_\lambda^2(x) =$ [singleton], and $e_\lambda^1(x) = e_\lambda^2(x)$ on V.*

Proof of Claim 1. First notice that the proximal mappings P_λ^i, $i = 1, 2$ are single-valued on V by (3.2.3). Let any x in V and $x_i = P_\lambda^i(x)$, $i = 1, 2$. Then by Propositions 2.4.2(b) and 2.4.3(b) we have $|x_1| < \varepsilon$, $|f_1(x_1)| < \varepsilon$ and $|v_1| < \varepsilon$, where $v_1 = \frac{1}{\lambda}(x - x_1) \in \partial f_1(x_1)$. With the same reasoning $x_2 = P_\lambda^2(x)$ gives $|x_2| < \varepsilon$ and $|v_2| < \varepsilon$, where $v_2 = \frac{1}{\lambda}(x - x_2) \in \partial f_2(x_2)$. Since $|x_2| < \varepsilon$ we have $v_2 \in \partial f_2(x_2) = \partial f_1(x_2)$. Since f_1 is subdifferentially continuous at $\bar{x} = 0$ for $\bar{v} = 0$, we may also assume that $|f_1(x_2)| < \varepsilon$. Thus applying Theorem 2.3.4 for the pairs (x_1, v_1) and (x_2, v_2) we get

$$\left\langle \left[\frac{x - x_1}{\lambda}\right] - \left[\frac{x - x_2}{\lambda}\right], x_1 - x_2 \right\rangle \geq -r\left|x_1 - x_2\right|^2,$$

hence $-\lambda^{-1}\left|x_1 - x_2\right|^2 \geq -r\left|x_1 - x_2\right|^2$. Then $(1 - \lambda r)|x_1 - x_2|^2 \leq 0$, so $x_1 = x_2$. Therefore, we have $P_\lambda^1(x) = P_\lambda^2(x)$ and by (3.2.4), $De_\lambda^1(x) = De_\lambda^2(x)$ on V. Thus we conclude $e_\lambda^1(x) = e_\lambda^2(x)$ since $e_\lambda^i(0) = 0$ when $\lambda \in (0, 1/r)$ and $i = 1, 2$ by (3.2.2).

Claim 2. *For all x in $\operatorname{dom} \partial f_1 \cap (\delta/4)I\!B$ and v in $\partial f_1(x)$ with δ small enough such that $|v| < (\delta/4) < \varepsilon$, and λ small enough we have $P_\lambda^1(z_\lambda) = P_\lambda^2(z_\lambda) = \{x\}$, where $z_\lambda = x + \lambda v$.*

Proof of Claim 2. Take any x in $\operatorname{dom} \partial f_1 \cap (\delta/4)I\!B$ and restrict $\lambda < 3$. Then

$$|z_\lambda| \leq |x| + \lambda|v| < \frac{\delta}{4} + \lambda\frac{\delta}{4} = (1 + \lambda)\frac{\delta}{4} < 4(\frac{\delta}{4}) = \delta,$$

so z_λ belongs to V.

Let $\tilde{x}$ be an element of $P_\lambda^1(z_\lambda) = P_\lambda^2(z_\lambda)$ (equality due to Claim 1). Then by Propositions 2.4.2(b) and 2.4.3(b) we have $|\tilde{x}| < \varepsilon$, $|f_1(\tilde{x})| < \varepsilon$ and $|\tilde{v}| < \varepsilon$, where $\tilde{v} = \frac{1}{\lambda}(z_\lambda - \tilde{x}) \in \partial f_1(\tilde{x})$. By our hypothesis $v = \frac{z_\lambda - x}{\lambda} \in \partial f_1(x)$ with $|v| < (\delta/4) < \varepsilon$ and $|x| < (\delta/4) < \varepsilon$. Since f_1 is subdifferentially continuous at $\bar{x} = 0$ for $\bar{v} = 0$, we may also assume that $|f_1(x)| < \varepsilon$. Thus applying Theorem 2.3.4 for the pairs $(\tilde{x}, \tilde{v})$ and (x, v) we get

$$\left\langle \left[\frac{z_\lambda - \tilde{x}}{\lambda}\right] - \left[\frac{z_\lambda - x}{\lambda}\right], \tilde{x} - x \right\rangle \geq -r\left|\tilde{x} - x\right|^2,$$

hence $-\lambda^{-1}\left|\tilde{x} - x\right|^2 \geq -r\left|\tilde{x} - x\right|^2$. Then $(1 - \lambda r)|\tilde{x} - x|^2 \leq 0$, so $\tilde{x} = x$. Thus we have $P_\lambda^1(z_\lambda) = P_\lambda^2(z_\lambda) = \{x\}$ as claimed.

Claim 3. *If x belongs to* $\operatorname{dom} \partial f_1$ *and x is near $\bar{x} = 0$ with subgradients $v \in \partial f_1(x) = \partial f_2(x)$ and close to $\bar{v} = 0$, we have $f_1(x) = f_2(x)$.*

Proof of Claim 3. Take any x in $\operatorname{dom} \partial f_1 \cap (\delta/4)\mathbb{B}$ and v in $\partial f_1(x)$ with $|v| < (\delta/4) < \varepsilon$. Restricting λ as in Claim 2, we have $z_\lambda = x + \lambda y$ in V. Then by Claims 1 and 2, we get $P_\lambda^1(z_\lambda) = P_\lambda^2(z_\lambda) = \{x\}$ and $e_\lambda^1(z_\lambda) = e_\lambda^2(z_\lambda)$. This means

$$f_1(x) + \frac{1}{2\lambda}|x - z_\lambda|^2 = f_2(x) + \frac{1}{2\lambda}|x - z_\lambda|^2,$$

and hence $f_1(x) = f_2(x)$. This completes the Claim and hence the Theorem. $\square$

3.3. Necessity of the Assumptions

The following examples show that the assumptions in Theorem 3.2.1 are necessary. Further, Example 3.3.2 shows that Theorem 3.2.1 covers a much broader class of functions than that of p.l.n. case [**24**].

Example 3.3.1. *(necessity of subdifferential continuity)*

Let

$$f_1(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0, \end{cases} \qquad f_2(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 2 & \text{if } x > 0, \end{cases}$$

then

$$\partial f_1(x) = \partial_p f_1(x) = \partial f_2(x) = \partial_p f_2(x) = \begin{cases} \{0\} & \text{if } x \neq 0, \\ [0, \infty) & \text{if } x = 0. \end{cases}$$

These two functions are prox-regular but not subdifferentially continuous at $\bar{x} = 0$ for $\bar{v} = 0$ (cf. Example 2.2.4). We see that they do not differ by a constant in any neighborhood of $(\bar{x}, \bar{v})$. This explains the necessity of the subdifferential continuity of the functions in Theorem 3.2.1.

Example 3.3.2. *(necessity of the closeness of the subgradients)*

Let

$$f_1(x) = \begin{cases} 0 & \text{if } x \le 0, \\ \sqrt{x} & \text{if } x > 0, \end{cases} \qquad f_2(x) = \begin{cases} 0 & \text{if } x \le 0, \\ 1 + \sqrt{x} & \text{if } x > 0, \end{cases}$$

then

$$\partial f_1(x) = \partial_p f_1(x) = \partial f_2(x) = \partial_p f_2(x) = \begin{cases} \{0\} & \text{if } x < 0, \\ [0, \infty) & \text{if } x = 0, \\ \frac{1}{2\sqrt{x}} & \text{if } x > 0. \end{cases}$$

First, we claim that both f_1 and f_2 are prox-regular and subdifferentially continuous at $\bar{x} = 0$ for $\bar{v} = 0$. To see this , take $\varepsilon = \frac{1}{4}$ and for $i = 1, 2$, T_i be the f_i-attentive ε-localization of ∂f_i around $(\bar{x}, \bar{v})$. It is easy to calculate, for $i = 1, 2$

$$T_i(x) = \begin{cases} \{0\} & \text{if } -\frac{1}{4} < x < 0, \\ [0, \frac{1}{4}) & \text{if } \quad x = 0, \\ \emptyset & \text{if } \quad 0 < x < \frac{1}{4}. \end{cases}$$

Then the prox-regularity of f_i, $i = 1, 2$, follows from the monotonicity of T_i via Theorem 2.3.4. Since f_1 is continuous it remains to verify that f_2 is subdifferentially continuous at $\bar{x} = 0$ for $\bar{v} = 0$. Indeed, for any sequence $(x_n, v_n) \to (0, 0)$ with $v_n \in \partial f_2(x_n)$ eventually we have $f_2(x_n) = 0 = f_2(0)$. Thus, f_2 is also subdifferentially continuous at $\bar{x} = 0$ for $\bar{v} = 0$. Yet f_1 and f_2 differ by different constants on any neighborhood of $\bar{x} = 0$. However, when we restrict to , say with $\varepsilon = \frac{1}{4}$, not only $|x - \bar{x}| < \varepsilon$ but $|v - \bar{v}| < \varepsilon$ with $v \in \partial f_1(x) = \partial f_2(x)$, then such x has to be in $(-\varepsilon, 0]$ and we have $f_1(x) = 0 = f_2(x)$ for all x in $(-\varepsilon, 0]$. This justifies that the requirement of taking not only x close to $\bar{x}$ but also the subgradients v close to $\bar{v}$ in Theorem 3.2.1.

This example also reveals that Theorem 3.2.1 covers much broader class of functions than that of p.l.n. case. For this, we only have to verify that f_1 is not p.l.n. at $\bar{x} = 0$. Here we make use of a corresponding subgradient characterization available for p.l.n. functions.

Theorem 3.3.3. *(Levi, Poliquin and Thibault [19], Corollary 2.3) Let $f : X \to \overline{I\!\!R}$ be a l.s.c. function that is finite at $\bar{x}$. The following are equivalent:*

(a) f is primal-lower-nice at $\bar{x}$.

(b) There exist positive constants ε, c and R such that

$$\langle v_1 - v_2, x_1 - x_2 \rangle \geq -r|x_1 - x_2|^2$$

whenever $v_i \in \partial_p f(x_i)$, $|v_i| \leq cr$, $r \geq R$ and $|x_i - \bar{x}| \leq \varepsilon$, $i = 1, 2$.

If f_1 were p.l.n. at $\bar{x} = 0$ then there would be constants ε, c and R as in Theorem 3.3.3. Then for any $r > R$, consider the mapping T formed by adding r times the identity to the subgradient mapping of f_1,

$$T(x) := \frac{1}{2\sqrt{x}} + rx \quad \text{for} \quad x \in (0, \varepsilon).$$

The critical points of T are given by $T'(x) = -\frac{1}{4x^{\frac{3}{2}}} + r = 0$, and attained at $x_m := \frac{1}{(4r)^{\frac{2}{3}}}$. Since $T''(x) = \frac{3}{8x^{\frac{5}{2}}} > 0$, x_m is a local minimum for T. Now restrict the subgradients of f_1 such that $\frac{1}{2\sqrt{x}} \leq cr$, i.e, $x_0 := \frac{1}{4c^2r^2} \leq x$. Then T to be monotone on $[x_0, \varepsilon)$, x_m has to be less than or equal x_0. This requires that $r^2 \leq \frac{1}{2c^3}$. But, for the large values of r this is impossible and which contradicts the monotonicity of T required by Theorem 3.3.3. This confirms that f_1 is not p.l.n. at $\bar{x} = 0$.

CHAPTER 4

CALCULUS OF PROX-REGULAR FUNCTIONS

As noted in Chapter 2, Poliquin and Rockafellar, in their study of prox-regular functions, have obtained many functional properties, however calculus rules for these functions have not appeared yet. We fill this gap by developing basic calculus rules for prox-regular functions. A master key to our calculus is the following chain rule.

4.1. The Chain Rule

Here we establish the prox-regularity of a composite function obtained by composing a prox-regular function with a $\mathcal{C}^{1+}$ (differentiable with locally Lipschitz Jacobian) mapping under a natural constraint qualification.

Theorem 4.1.1. *(chain rule) Assume that $F : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable at $\bar{x}$ with the Jacobian mapping ∇F Lipschitz continuous near $\bar{x}$, $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ with $g(F(\bar{x}))$ finite, and that the following constraint qualification $(\mathcal{R})$ is satisfied at $F(\bar{x})$.*

$$(\mathcal{R}): \quad \textit{The only vector } \ y \in \partial^\infty g(F(\bar{x})) \ \textit{ with } \ \nabla F(\bar{x})^* y = 0 \ \textit{ is } \ y = 0.$$

Let $\bar{v} \in \partial(g \circ F)(\bar{x})$ and set

$$Y(\bar{x}, \bar{v}) := \{\, y \in \partial g(F(\bar{x}))\, ;\, \nabla F(\bar{x})^* y = \bar{v} \,\}.$$

Assume further that the outer function g is prox-regular at $F(\bar{x})$ for all $y \in Y(\bar{x}, \bar{v})$. Then the composite function $g \circ F$ is prox-regular at $\bar{x}$ for $\bar{v}$.

Proof. Let $v \in \partial(g \circ F)(x)$ and consider the set

$$Y(x,v) := \{\, y \in \partial g(F(x)) \,;\, \nabla F(x)^* y = v \,\}.$$

First we show that, for all (x, v) in an $(g \circ F)$-attentive neighborhood of $(\bar{x}, \bar{v})$ the subgradients y in $Y(x, v)$ are bounded.

Claim 1. *For $\bar{v} \in \partial(g \circ F)(\bar{x})$ there exists $\varepsilon > 0$ such that the set*

$$S := \{\, y \in Y(x,v) \,;\, |x - \bar{x}| < \varepsilon, |v - \bar{v}| < \varepsilon, \quad \textit{and} \quad |g(F(x)) - g(F(\bar{x}))| < \varepsilon \,\},$$

is bounded.

Proof of Claim 1. Suppose that the statement of the claim does not hold. Then there exist sequences $x_n \to \bar{x}$, $v_n \to \bar{v}$, and $y_n \in Y(x_n, v_n)$ with $|y_n| \to \infty$ and $g(F(x_n)) \to g(F(\bar{x}))$. Since $v_n = \nabla F(x_n)^* y_n$ with $y_n \in \partial g(F(x_n))$, by passing to the vectors

$$\frac{v_n}{|y_n|} = \nabla F(x_n)^* \frac{y_n}{|y_n|}, \tag{4.1.1}$$

and extracting a subsequence, we can suppose that $y_n/|y_n|$ converges to some y, with $|y| = 1$. Then $0 \neq y \in \partial^\infty g(F(\bar{x}))$, by the definition of singular limiting subgradients. At the same time we have $\nabla F(\bar{x})^* y = 0$ by (4.1.1) and the continuity of ∇F. This contradicts the constraint qualification $(\mathcal{R})$. $\square$

Thus, by Claim 1, and the closedness of the limiting proximal subdifferential set, in particular, we conclude that the set $Y(\bar{x}, \bar{v})$ is compact.

Now consider, for $\bar{v} \in \partial(g \circ F)(\bar{x})$, the set

$$Y(\bar{x},\bar{v}) = \{\, y \in \partial g(F(\bar{x})) \,;\, \nabla F(\bar{x})^* y = \bar{v} \,\}.$$

For each $y \in Y(\bar{x}, \bar{v})$, there exist parameters $\varepsilon_y > 0$ and $r_y > 0$ from the definition of prox-regularity of g at $F(\bar{x})$ for y, and hence we have a covering of $Y(\bar{x}, \bar{v})$ by open balls, i.e.,

$$Y(\bar{x}, \bar{v}) \subset \bigcup_{y \in Y(\bar{x}, \bar{v})} I\!\!B(y, \varepsilon_y).$$

The compactness of $Y(\bar{x}, \bar{v})$ allow us to find a finite subcovering:

$$Y(\bar{x}, \bar{v}) \subset \bigcup_{i=1}^{m} I\!\!B(y_i, \varepsilon_{y_i}) \quad \text{where} \quad y_i \in Y(\bar{x}, \bar{v}). \tag{4.1.2}$$

Now fix ε as in Claim 1.

Claim 2. *There exists $\tilde{\varepsilon} > 0$ such that $0 < \tilde{\varepsilon} < \varepsilon$ and*

$$\left.\begin{array}{r} |x - \bar{x}| < \tilde{\varepsilon} \\ |v - \bar{v}| < \tilde{\varepsilon} \\ |g(F(x)) - g(F(\bar{x}))| < \tilde{\varepsilon} \\ \textit{with} \quad v \in \partial(g \circ F)(x) \quad \textit{and} \quad y \in Y(x, v) \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} \exists\, y_i \in Y(\bar{x}, \bar{v}) \quad \textit{in} \quad (4.1.2) \\ \quad \textit{such that} \quad |y - y_i| < \varepsilon_{y_i} \\ \quad \textit{for some} \quad i = 1, \ldots, m. \end{array}\right.$$

Proof of Claim 2. Assume the contrary, i.e., there exist sequences $x_n \to \bar{x}$, $v_n \to \bar{v}$ with $y_n \in Y(x_n, v_n)$ and $g(F(x_n)) \to g(F(\bar{x}))$ such that for all $y_i \in Y(\bar{x}, \bar{v})$ one has

$$|y_n - y_i| \geq \varepsilon_{y_i} \quad i = 1, \ldots, m. \tag{4.1.3}$$

Then by Claim 1, y_n are bounded (eventually), and hence extracting a subsequence, we may suppose that y_n converges to some $\bar{y}$. Then $\bar{y} \in \partial g(F(\bar{x}))$ by the closedness of the graph of limiting subdifferentials. Since $y_n \in Y(x_n, v_n)$, i.e, $v_n = \nabla F(x_n)^* y_n$ with $y_n \in \partial g(F(x_n))$ and the continuity of ∇F, we also have $\bar{v} = \nabla F(\bar{x})^* \bar{y}$ with $\bar{y} \in \partial g(F(\bar{x}))$. Then $\bar{y} \in Y(\bar{x}, \bar{v})$ and hence by (4.1.2) there exists $y_i \in Y(\bar{x}, \bar{v})$ such that $|\bar{y} - y_i| < \varepsilon_{y_i}$ for some $i \in \{1, \ldots, m\}$. At the same time we have, by (4.1.3), $|\bar{y} - y_i| \geq \varepsilon_{y_i}$ for all $i \in \{1, \ldots, m\}$, which contradicts the preceding statement. $\square$

Recall that ε_i and r_i, $i = 1, \ldots, m$, are the parameters corresponding to the prox-regularity of g at $F(\bar{x})$ for $y_i \in Y(\bar{x}, \bar{v})$. Choose

$$\bar{\varepsilon} = \min\{\varepsilon_{y_i}\,;\, i = 1, \ldots, m\} \quad \text{and} \quad \bar{r} = \max\{r_i\,;\, i = 1, \ldots, m\}.$$

Then by Claim 2, and the continuity of F, there exists $\tilde{\varepsilon}$ such that $0 < \tilde{\varepsilon} < \min\{\varepsilon, \bar{\varepsilon}\}$ and

$$\left.\begin{array}{r} |x - \bar{x}| < \tilde{\varepsilon} \\ |v - \bar{v}| < \tilde{\varepsilon} \\ |g(F(x)) - g(F(\bar{x}))| < \tilde{\varepsilon} \\ \text{with} \quad v \in \partial(g \circ F)(x) \quad \text{and} \quad y \in Y(x, v) \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} \exists\, y_i \in Y(\bar{x}, \bar{v}) \quad \text{in} \quad (4.1.2) \\ \text{such that} \quad |y - y_i| < \varepsilon_{y_i} \\ \text{for some} \quad i = 1, \ldots, m, \\ \text{and} \quad |F(x) - F(\bar{x})| < \bar{\varepsilon}. \end{array}\right. \tag{4.1.4}$$

Consider a $(g \circ F)$-attentive $\tilde{\varepsilon}$-localization of $\partial(g \circ F)$ around $(\bar{x}, \bar{v})$ as in the left hand side of (4.1.4). We then have $|g(F(x)) - g(F(\bar{x}))| < \tilde{\varepsilon} < \varepsilon_{y_i}$, and by (4.1.4), $|F(x) - F(\bar{x})| < \bar{\varepsilon} < \varepsilon_{y_i}$, $|y - y_i| < \varepsilon_{y_i}$, $y_i \in Y(\bar{x}, \bar{v})$ for some $i = 1, \ldots, m$. Hence, invoking the prox-regularity of g at $F(\bar{x})$ for $y_i \in Y(\bar{x}, \bar{v})$ (with parameters ε_{y_i} and r_i) we get

$$\begin{aligned} g(F(x')) &\geq g(F(x)) + \langle y, F(x') - F(x) \rangle - \frac{r_i}{2}|F(x') - F(x)|^2 \\ &\geq g(F(x)) + \langle y, F(x') - F(x) \rangle - \frac{\bar{r}}{2}|F(x') - F(x)|^2 \end{aligned} \tag{4.1.5}$$

where $x' \in \mathbb{B}(\bar{x}, \tilde{\varepsilon})$ and $\bar{r} = \max\{r_i\,;\, i = 1, \ldots, m\}$.

Let k be the local Lipschitz constant for F and K be that of ∇F on the set $\mathbb{B}(\bar{x}, 2\tilde{\varepsilon})$ (we shrink $\tilde{\varepsilon}$ if necessary). Applying the local Lipschitzness of F to

(4.1.5) we obtain

$$g(F(x')) \geq g(F(x)) + \langle y, F(x') - F(x) \rangle - \frac{\bar{r}k^2}{2}|x' - x|^2. \tag{4.1.6}$$

To show that $g \circ F$ is prox-regular at $\bar{x}$ for $\bar{v} \in \partial(g \circ F)(\bar{x})$, we need $r' > 0$ large enough such that

$$g(F(x')) \geq g(F(x)) + \langle v, x' - x \rangle - \frac{r'}{2}|x' - x|^2 \quad \text{for all} \quad x' \in \mathbb{B}(\bar{x}; \bar{\varepsilon}) \tag{4.1.7}$$

whenever $v \in \partial(g \circ F)(x)$, $y \in Y(x, v)$, $|v - \bar{v}| < \bar{\varepsilon}$, $|x - \bar{x}| < \bar{\varepsilon}$, $|g(F(x)) - g(F(\bar{x}))| < \bar{\varepsilon}$.

Thus, by (4.1.6) we have (4.1.7) whenever the following inequality holds,

$$\langle v, x - x' \rangle + \frac{r'}{2}|x - x'|^2 \geq \langle y, F(x) - F(x') \rangle + \frac{\bar{r}k^2}{2}|x - x'|^2.$$

Or equivalently,

$$\langle v, x - x' \rangle + (\frac{r'}{2} - \frac{\bar{r}k^2}{2})|x - x'|^2 \geq \langle y, F(x) - F(x') \rangle. \tag{4.1.8}$$

Thus, we will be done if we can verify inequality (4.1.8). For that, choose r' large enough such that $M := (\frac{r'}{2} - \frac{\bar{r}k^2}{2}) > 0$ and $M \geq \eta K$, where η is the bound for $y \in S$ in Claim 1. Note that $y \in Y(x, v)$ in (4.1.7) same as in (4.1.8) are belong to the set S in Claim 1. This is because $\bar{\varepsilon} < \varepsilon$, by our choice.

Now consider the point y' defined by

$$y' := \frac{F(x') - F(x) - \nabla F(x)(x' - x)}{|x' - x|}$$

where we assume that $x' \neq x$, otherwise inequality (4.1.8) holds trivially. By the Mean Value Theorem, the norm of y' is bounded by $K|x' - x|$, and by Claim 1,

$y \in S$ are bounded by η. Utilizing these two bounds, we obtain the estimate

$$\begin{aligned}
-M|x'-x| &= -\frac{1}{K\eta} M\eta K|x'-x| \\
&\leq -\frac{1}{K\eta} M\eta|y'| \\
&\leq -|y||y'| \qquad (\text{ since } \; M \geq K\eta \; \text{ and } \; |y| \leq \eta) \\
&\leq \langle y, y' \rangle.
\end{aligned}$$

Hence, replacing y' by the defined expression gives

$$\langle \nabla F(x)(x-x'), y \rangle + M|x-x'|^2 \geq \langle y, F(x) - F(x') \rangle.$$

Now (4.1.8) follows since $v = \nabla F(x)^* y$ and $M = (\frac{r'}{2} - \frac{\bar{r}k^2}{2})$. ◻

In the framework of nonsmooth analysis, the chain rule for $f = g \circ F$ is the foundation for many other rules of calculus. For instance, it gives instant access to the following sum rule.

Corollary 4.1.2. *(sum rule) Suppose $f = f_1 + \cdots + f_m$, $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, $\bar{x} \in \operatorname{dom} f$, $\bar{v} \in \partial f(\bar{x})$ and the only combination of vectors $y_i \in \partial^\infty f_i(\bar{x})$ with $y_1 + \cdots + y_m = 0$ is $y_1 = \cdots = y_m = 0$. Assume also that, for $i = 1, \cdots, m$, f_i are prox-regular for all $v_i \in \partial f_i(\bar{x})$ such that $v_1 + \cdots + v_m = \bar{v}$. Then f is prox-regular at $\bar{x}$ for $\bar{v}$.*

Proof. Let $F : \mathbb{R}^n \to (\mathbb{R}^n)^m$ be the mapping that takes x to $(x, \ldots, x)$, and define the function $g : (\mathbb{R}^n)^m \to \overline{\mathbb{R}}$ by

$$g(x_1, \ldots, x_m) = f_1(x_1) + \cdots + f_m(x_m).$$

Then $f(x) = g(F(x))$, and the following subgradient formulas hold (cf. Proposition 2.1.14).

$$\begin{aligned}
\partial g(x_1, \ldots, x_m) &= \partial f_1(x_1) \times \cdots \times \partial f_m(x_m) \\
\partial^\infty g(x_1, \ldots, x_m) &\subseteq \partial^\infty f_1(x_1) \times \cdots \times \partial^\infty f_m(x_m).
\end{aligned}$$

Next we show that this composite function $g \circ F$ satisfies the constraint qualification $(\mathcal{R})$ at $F(\bar{x})$ of the Theorem 4.1.1. Indeed, for all $\tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_m) \in \partial^\infty g(\bar{x}, \ldots, \bar{x})$ with $\nabla F(\bar{x})^* \tilde{y} = 0$ imply

$$\Big(\tilde{y}_1, \ldots, \tilde{y}_m\Big) \begin{pmatrix} I_{n\times n} \\ \vdots \\ I_{n\times n} \end{pmatrix} = 0, \quad \text{where} \quad \tilde{y}_i \in \partial^\infty f_i(\bar{x}),\ i = 1, \ldots, m.$$

i.e., $\tilde{y}_1 + \cdots + \tilde{y}_m = 0$ with $\tilde{y}_i \in \partial^\infty f_i(\bar{x})$, $i = 1, \ldots, m$. Then by our assumption we have $\tilde{y}_1 = \cdots = \tilde{y}_m = 0$, as desired. Thus we have (cf. Theorem 2.1.12),

$$\partial f(x) \subseteq \nabla F(x)^* \partial g(F(x))$$

Since f_i are prox-regular for all $v_i \in \partial f_i(\bar{x})$ such that $v_1 + \cdots + v_m = \bar{v}$, it follows that g is prox-regular at $F(\bar{x})$ for all $y = (v_1, \cdots, v_m) \in \partial g(F(\bar{x}))$ such that $\nabla F(\bar{x})^* y = v_1 + \cdots + v_m = \bar{v}$, where $v_i \in \partial f_i(\bar{x})$, $i = 1, \ldots, m$. Hence, applying Theorem 4.1.1 for the composite function $g \circ F$ we conclude that f is prox-regular at $\bar{x}$ for $\bar{v}$. $\square$

4.2. Some Applications

Next, we record several applications of the Chain Rule (Theorem 4.1.1).

Corollary 4.2.1. *Assume that $F : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable at $\bar{x}$ with the Jacobian mapping ∇F Lipschitz continuous near $\bar{x}$, $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ with $g(F(\bar{x}))$ finite, and that the following constraint qualification $(\mathcal{R})$ is satisfied at $F(\bar{x})$.*

$$(\mathcal{R}): \quad \textit{The only vector} \quad y \in \partial^\infty g(F(\bar{x})) \quad \textit{with} \quad \nabla F(\bar{x})^* y = 0 \quad \textit{is} \quad y = 0.$$

Let $\bar{v} \in \partial(g \circ F)(\bar{x})$ and set

$$Y(\bar{x}, \bar{v}) = \big\{\, y \in \partial g\big(F(\bar{x})\big)\ ;\ \nabla F(\bar{x})^* y = \bar{v} \,\big\}.$$

Assume further that the outer function g is prox-regular at $F(\bar{x})$ for all $y \in Y(\bar{x}, \bar{v})$ and the composite function $g \circ F$ is subdifferentially continuous at $\bar{x}$ for $\bar{v}$. In this setting, $g \circ F$ is twice epi-differentiable at $\bar{x}$ for $\bar{v}$ if and only if $\partial(g \circ F)$ is proto-differentiable at $\bar{x}$ for $\bar{v}$ with

$$\big[\partial(g \circ F)\big]'_{\bar{x},\bar{v}}(\xi) \;=\; \partial\big[\tfrac{1}{2}(g \circ F)''_{\bar{x},\bar{v}}\big](\xi) \qquad \text{for all} \quad \xi.$$

Further, when $g \circ F$ is twice epi-differentiable at $\bar{x}$ for $\bar{v}$ with a finite second-order epi-derivative $(g \circ F)''_{\bar{x},\bar{v}}$ on a neighbourhood of 0, the composite function $g \circ F$ has a second-order expansion

$$(g \circ F)(x) \;=\; (g \circ F)(\bar{x}) + \langle \bar{v}, x - \bar{x} \rangle + (g \circ F)''_{\bar{x},\bar{v}}(x - \bar{x}) + o(|x - \bar{x}|^2).$$

Moreover, the composite function $g \circ F$ is integrable in the sense of the Theorem 3.2.1.

Proof. Since $g \circ F$ is prox-regular at $\bar{x}$ for $\bar{v}$ by Theorem 4.1.1, the stated results follow directly from Corollary 2.6.5, Theorem 2.6.9, and Theorem 3.2.1 (in the same order). □

The smoothness and convexity properties of Moreau envelopes of a prox-regular function can also be transformed into the above composite case.

Corollary 4.2.2. *Consider the composite function $g \circ F$ in the setting of Corollary 4.2.1. Then the Moreau envelope e_λ of $g \circ F$ is not only $\mathcal{C}^{1+}$ but also lower-$\mathcal{C}^2$ in a neighbourhood of $\bar{x}$ with*

$$De_\lambda \;=\; \big[\lambda I + [\partial(g \circ F)]^{-1}\big]^{-1}, \quad \text{and} \quad e_\lambda + \frac{r}{2(1 - \lambda r)} |\cdot|^2 \qquad \text{convex.}$$

Proof. Since $g \circ F$ is prox-regular at $\bar{x}$ for $\bar{v}$ by Theorem 4.1.1, the stated results follow directly from Proposition 2.4.6 and Theorem 2.5.2 (in the same order). $\square$

REFERENCES

1. H. Attouch, **Variational Convergence for Functions and Operators**, Pitman, Boston-London-Melbourne (1984).
2. J. P. Aubin and H. Frankowska, **Set-Valued Analysis**, Birkhäuser, 1990.
3. D. Azé and J. P. Penot, "Operations on convergent families of sets and functions," Optimization, **21** (1990), 521-534.
4. J. Benoist, "Intégration du sous-differentiel proximal : un contre example," Can. J. Math., **50** (1998), 242–265.
5. M. S. Berger, **Nonlinearity and Functional Analysis**, Lectures on Nonlinear Problems in Mathematical Analysis, Academic Press, New York, 1977.
6. S. K. Boralugoda, "Proto-differentiation and epi-differentiation," M. Sc. Thesis, University of Alberta, 1994.
7. J. M. Borwein and W. B. Moors, "Null sets and essentially smooth Lipschitz functions," SIAM J. Optimization, (to appear).
8. J. M. Borwein and D. Noll, "Second-order differentiability of convex functions in Banach spaces, Trans. Amer. Math. Soc., **342** (1994), 43–81.
9. J. M. Borwein and D. Preiss, "A smooth variational principle with applications to subdifferentiability and to differentiability of convex functions," Trans. Amer. Math. Soc., **303** (1987), 517-527.
10. J.V. Burke and R. A. Poliquin, "Optimality conditions for non-finite convex composite functions," Mathematical Programming **57** (1992), 103–120.

11. J. P. R. Christensen, **Topological and Borel Structure**, American Elsevier, New York, 1974.

12. F. H. Clarke, **Optimization and Nonsmooth Analysis**, Classics in Applied Math., vol. 5, SIAM Publications, Philadelphia, 1990 (originally published in 1983).

13. F. H. Clarke, **Methods of Dynamic and Nonsmooth Optimization**, CBMS-NSF Regional Conference Series in Applied Mathematics, **57**, SIAM Publications, Philadelphia, 1989.

14. F. H. Clarke, Yu. S. Ledyaev, R. J. Stern and P. R. Wolenski, **Nonsmooth Analysis and Control Theory**, Springer-Verlag, 1997.

15. C. Combari, A. Elhilali Alaoui, A. Levy, R. Poliquin and L. Thibault, "Convex composite functions in Banach spaces and primal-lower-nice property," Proc. Amer. Math. Soc., to appear.

16. R. Cominetti, "On Pseudo-differentiability," Trans. Amer. Math. Soc. **324** (1991), 843–865.

17. C. Do, "Generalized second derivatives of convex functions in reflexive Banach spaces," Trans. Amer. Math. Soc. **334** (1992), 281–301.

18. A. D. Ioffe, "Variational analysis of a composite function: a formula for the lower second order epi-derivative," J. Math. Anal. Appl. **160** (1991), 379–405.

19. A. Levy, R. A. Poliquin and L. Thibault, "Partial extension of Attouch's theorem with applications to proto-derivatives of subgradient mappings," Trans. Amer. Math. Soc., **347** (1995), 1269–1294.

20. A. Levy and R. T. Rockafellar, "Variational conditions and the proto-differentiation of partial subgradient mappings," Nonlinear Anal. Th. Meth. Appl.,

26 (1996), 1951–1964.

21. P. D. Loewen, **Optimal Control via Nonsmooth Analysis**, CRM Proceedings & Lecture Notes, American Mathematical Society, Providence, Rhode Island, 1993.

22. R. A. Poliquin, "Subgradient monotonicity and convex functions," Nonlinear Analysis, Theory, Methods & Applications, **14** (4) (1990), 305-317.

23. R. A. Poliquin, "Proto-differentiation of subgradient set-valued mappings," Canadian J. Math. **42** (1990), 520–532.

24. R. A. Poliquin, "Integration of subdifferentials of nonconvex functions," Nonlinear Anal. Th. Meth. Appl. **17** (1991), 385–398.

25. R. A. Poliquin, "An extension of Attouch's Theorem and its application to second-order epi-differentiation of convexly composite functions," Trans. Amer. Math. Soc. **332** (1992), 861–874.

26. R. A. Poliquin and R. T. Rockafellar, "Amenable functions in optimization," in *Nonsmooth Optimization Methods and Applications* (F. Giannessi, ed.), Gordon and Breach, Philadelphia, 1992, 338–353.

27. R. A. Poliquin and R. T. Rockafellar, "A calculus of epi-derivatives applicable to optimization," Canadian J. Math., **45** (4) (1993), 879–896.

28. R. A. Poliquin and R. T. Rockafellar, "Proto-derivative formulas for basic subgradient mappings in mathematical programming," Set-Valued Analysis, **2** (1994), 275–290.

29. R. A. Poliquin and R. T. Rockafellar, "Prox-regular functions in variational analysis," Trans. Amer. Math. Soc. **348** (1996), 1805–1838.

30. R. A. Poliquin and R. T. Rockafellar, "Generalized Hessian properties of

regularized nonsmooth functions," SIAM J. Optimization, **6** (1996), 1121–1137.

31. R. A. Poliquin and R. T. Rockafellar, "Second-order nonsmooth analysis in nonlinear programming," in *Recent Advances in Optimization* (D. Du, L. Qi and R. Womersley, eds.), World Scientific Publishers, 1995.

32. R. A. Poliquin, J. Vanderwerff and V. Zizler, "Convex composite representation of lower semicontinuous functions and renormings," C.R. Acad. Sci. Paris, t. 317 Série I (1993), 545–549.

33. R. T. Rockafellar, "Characterization of the subdifferentials of convex functions," Pacific J. Math. **17** (1966), 497-510.

34. R. T. Rockafellar, "Local boundedness of nonlinear monotone operators," Michigan Math. J. **16** (1969), 397–407

35. R. T. Rockafellar, **Convex Analysis**, Princeton University Press, Princeton, NJ, 1970.

36. R. T. Rockafellar, "Augmented Lagrangians and applications of the proximal point algorithm in convex programming," Math. of Op. Res. **1** (1976), 97–116.

37. R. T. Rockafellar, "Proximal subgradients, marginal values, and augmented Lagrangians in nonconvex optimization," Math. of Op. Res. **6** (1981), 424–436.

38. R. T. Rockafellar, "Favorable classes of Lipschitz continuous functions in subgradient optimization," *Progress in Nondifferentiable Optimization*, E. Nurminski (ed.), IIASA Collaborative Proceedings Series, International Institute of Applied Systems Analysis, Laxenburg, Austria 1982, 125-144.

39. R. T. Rockafellar, "Maximal monotone relations and the second derivatives of

nonsmooth functions," Ann. Inst. H. Poincaré: Analyse non linéaire **2** (1985), 167–184.

40. R. T. Rockafellar, "First- and second-order epi-differentiability in nonlinear programming," Trans. Amer. Math. Soc. **307** (1988), 75–107.

41. R. T. Rockafellar, "Proto-differentiability of set-valued mappings and its applications in optimization," in *Analyse Non Linéaire*, H. Attouch et al. (eds.), Gauthier-Villars, Paris (1989), 449–482.

42. R. T. Rockafellar, "Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives," Math. of Oper. Research **14** (1989), 462–484.

43. R. T. Rockafellar, "Generalized second derivatives of convex functions and saddle functions," Trans. Amer. Math. Soc. **320** (1990), 810–822.

44. R. T. Rockafellar, "Nonsmooth analysis and parametric optimization," in *Methods of Nonconvex Analysis* (A. Cellina, ed.) Springer-Verlag Lecture Notes in Math. No. 1446 (1990), 137–151.

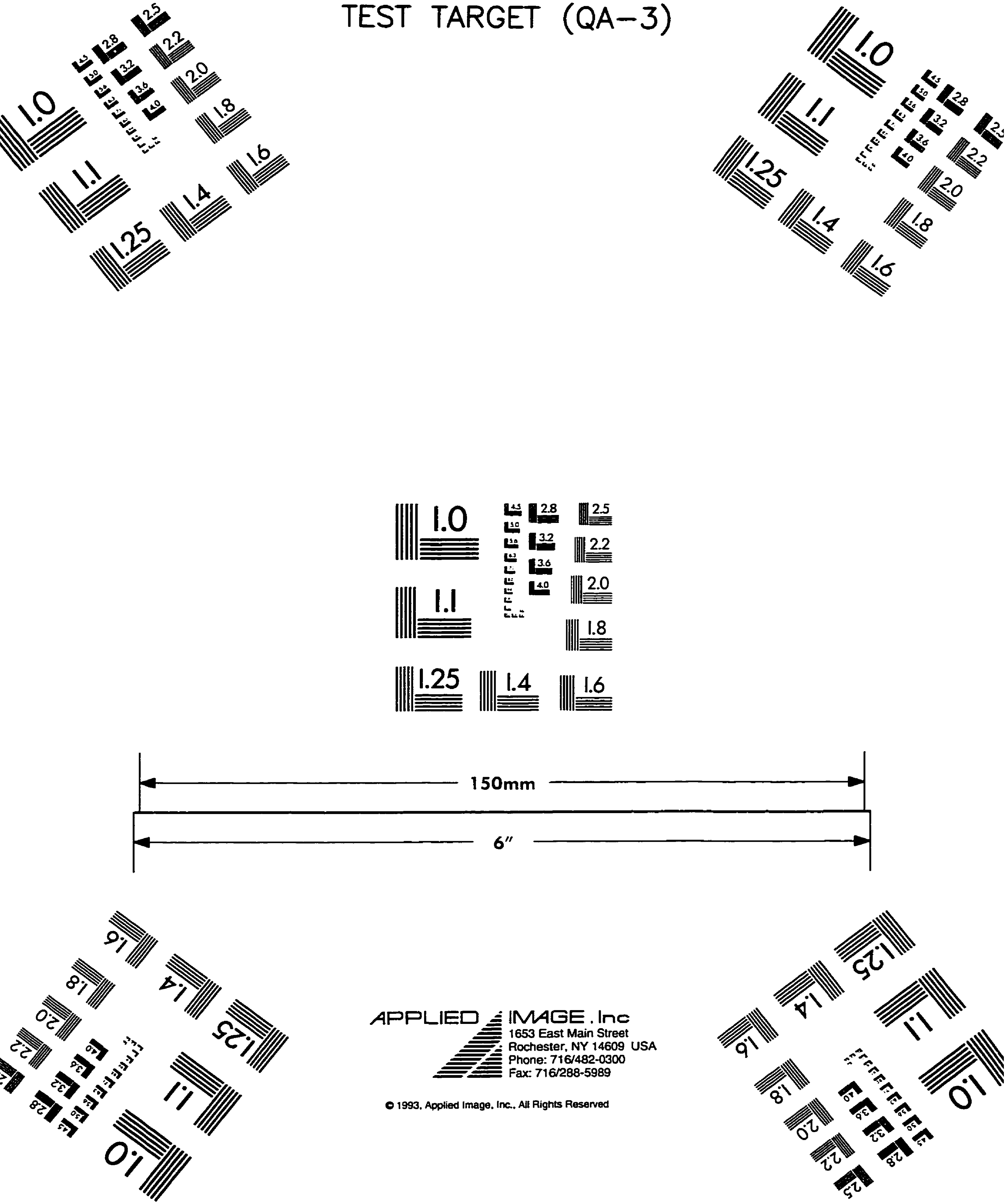
45. R. T. Rockafellar, "Lagrange multipliers and optimality," SIAM Review 35 (1993), 183–238.

46. R. T. Rockafellar and R. J-B Wets, **Variational Analysis**, Springer-Verlag, 1998.

47. T. Strömberg, "On regularization in Banach spaces," Ark. Mat. **34** (1996), 383-406.

48. L. Thibault and D. Zagrodny, "Integration of subdifferentials of lower semicontinuous functions," J. Math. Anal. Appl., **189** (1995), 33–58.

IMAGE EVALUATION TEST TARGET (QA-3)



APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989