Parametrically Prox-Regular Functions

by

Chayne Planiden

B.Sc. Hons., Universidad Autónoma de Baja California, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The College of Graduate Studies

(Mathematics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Okanagan)

April 2013

Abstract

Prox-regularity is a generalization of convexity that includes all lower-$\mathcal{C}^2$ functions. Therefore, the study of prox-regular functions provides insight on a broad spectrum of important functions. Parametrically prox-regular (para-prox-regular) functions are a further extension of this family, produced by adding a parameter. Such functions have been shown to play a key role in understanding the stability of minimizers in optimization problems. This thesis discusses para-prox-regular functions in $\mathbb{R}^n$.

We begin with some basic examples of para-prox-regular functions, and move on to the more complex examples of the convex and nonconvex proximal averages. We develop an alternate representation of para-prox-regular functions, related to the monotonicity of an f-attentive ϵ-localization as was done for prox-regular functions [25]. Levy in [18] provided proof of one implication of this relationship; we provide a characterization. We analyze two common forms of parametrized functions that appear in optimization: finite parametrized sum of functions, and finite parametrized max of functions. The example of strongly amenable functions by Poliquin and Rockafellar [27] is given, and a relaxation of its necessary conditions is presented. Some open questions and directions of further research are stated.

Statement of Co-Authorship

This thesis has been adapted from a manuscript co-authored with Dr. Warren Hare at the University of British Columbia, Okanagan campus. The manuscript has been submitted for publication to the Journal of Convex Analysis [12].

Table of Contents

List of Figures

Acknowledgements

Thank you to Dr. Warren Hare, my supervisor and co-author. You have brought out a love for research in me, and I look forward to continuing down this path you have helped me to discover.

Thank you to the committee members for agreeing to be a part of this very important moment in my career.

Thank you to NSERC and UBC for funding this research.

Dedication

To my wife:
As is the case with every other positive aspect of my life, this could not have happened without your love and support. Thank you so much for all that you do and all that you are for me.

I love you.

Chapter 1

Introduction

Optimization is a branch of mathematics that is of great practical importance in the world today. The very basic objective of an optimizer is to find the minimizers of a given function that is subject to a set of constraints. So that the reader may visualize this concept clearly, let us consider the following example.

Imagine that you are the CEO of a company that makes soft drinks. You want to design a plastic bottle that holds a certain volume of your product, and you want to use the minimum quantity of material possible in manufacturing the bottle. This is a logical objective, since you will be mass-producing the bottles and any savings in the amount of material used will affect the bottom line of the company. So the initial question is, what shape should the bottle be? This addresses the first part of the optimizer's objective: finding the minimizer. In the language of a mathematician, the goal is to find the shape that minimizes the surface area of a container while maintaining its volume unchanged. There are well-known formulas for the volume and surface area of any basic three-dimensional shape we would like to consider: a cylinder, a cube, etc. Calculus can be used to determine that the most efficient shape for our purposes is a sphere. If we want to contain a fixed volume of liquid in a container that uses a minimum of material, the container should be the shape of a ball.

Problem solved? Well, not really. How do you drink out of a spherical bottle? How do you keep it from rolling off the table? How do you stack them on the shelves in the grocery stores? We have solved one problem and created a multitude of others. So this is the moment to employ the other component to the optimizer's basic objective: the set of constraints. We will reconsider the same problem, but first we make a list of restrictions on the bottle, such as the following.

- It must stand up on its own on a flat surface,
- it must be a shape that can be held comfortably with one hand,

– it must be convenient to drink out of,

– it must fit in a standard cupholder of a vehicle,

and so on. One may list as many constraints as one wishes, and then ask the original question again. Now, subject to this set of constraints, what shape will use a minimum of material? So we continue with our calculations.

The next-best shape after a sphere is a cylinder with hemispheres on both ends. This one can be held comfortably with one hand, but it still does not meet the other criteria. So we continue the process until we have the shape in use today. Think about the plastic soda bottles we have in the stores. They are, in fact, cylinders with hemispheres on the top, and a spout added for drinking. The bottom is not a hemisphere of course, that would not stand up on its own, but it is not flat either. A flat bottom would be a waste of material, so the feet are little hemispheres. That is a more efficient use of the material, and that is why it is the shape we see everywhere soft drinks are sold.

This example is a great illustration of the importance of optimization. Every business is concerned with issues such as maximizing production, minimizing overhead expenses, maximizing efficiency, minimizing waste, maximizing profit, minimizing risk. So aspects of the research we do in this field are applicable to anyone and everyone in industry.

This thesis deals with a particular family of functions seen in optimization problems, called parametrically prox-regular (para-prox-regular) functions. In chapters that follow, we shall see the definition and particular uses of these functions, some examples, and a few methods of constructing them. For now, we will talk about the origin of these functions and why we are interested in learning about them. We begin with an important ancestor, the convex function.

In the late nineteenth century, mathematicians such as Minkowski [22], Brunn [8], and many others began to work with convex functions, as their desirable properties became apparent. Research on convexity continues today; convex analysis is a prominent area of research for optimizers. So what is a convex function? Graphically speaking, it is exactly what we might intuitively expect: a function that has a convex shape. Figure 1.1 illustrates three such functions.

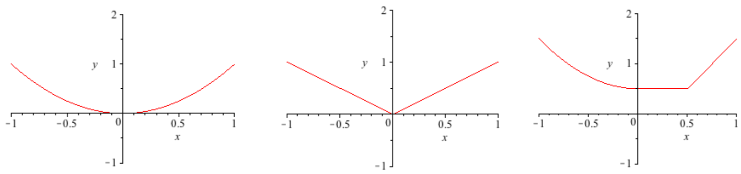

Figure 1.1: Three convex functions.

We will see the formal definition of convex functions in the next chapter, but for now let us just consider the graphs. If we want to verify that a function is convex, we can use a line-segment test. Every convex function has the property that, if you choose any two points that are on or above the graph of the function, the line segment that joins those two points is also entirely on or above the graph. For example, all three graphs in Figure 1.1 have that property. If we look more closely at one of those graphs, we can see more clearly that this is the case.

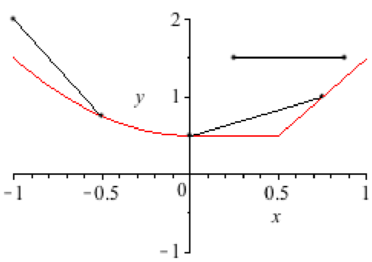

Figure 1.2: Line segments on or above the curve.

This gives us a way to prove graphically that a function is not convex. If we have the graph of a function and we can find two points that lie on or above the graph, but the line segment that joins them does not, then we know that we have a non-convex function. It has failed the line-segment test. For example, Figure 1.3 shows a non-convex function. Here we have chosen two points that are above the graph, but a section of the line segment that joins them falls below the graph. So this function is not convex.

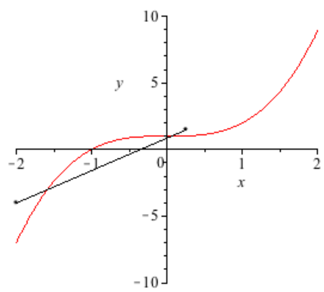

Figure 1.3: A non-convex function.

There are many benefits to working with convex functions, from an optimization point of view. However, there are other classes of functions that are not necessarily convex, but that also have desirable properties. Most notably among those is the class of differentiable functions, or so-called 'smooth' functions. Again considering them from a graphical standpoint, differentiable functions are functions whose graphs are smooth curves, with no gaps or sharp points in them. Taking a second look at Figure 1.1, we see that the graph on the left is that of a smooth function, while the other two are not, since they both have a kink in them. Figure 1.3 also represents a smooth function. So we can see that there is some overlap of the two classes, and that there are some differences between them. There exist functions that are both convex and differentiable, convex functions that are not differentiable, differentiable functions that are not convex, and functions that are neither. It would be beneficial to have a class of functions that covers both these types of functions, so that any theory developed in this new class would be applicable to both convex functions and smooth functions. This has already been done, and we call this new class prox-regular functions.

Prox-regularity was first introduced in the mid-nineties [10, 24, 25], so this branch of mathematics is less than twenty years old. The definition is quite technical and is better left to the next chapter, but it is a very useful property, as we shall see shortly. Many familiar classes of functions fall into this category, including convex functions and differentiable functions. Since its conception, a number of articles have been published on the topic [4, 6, 7, 11, 13–16, 20, 23, 26], furthering the theory and establishing practical, real-world applications for such functions.

Para-prox-regularity, which is the focus of this thesis, was first introduced by Levy, Poliquin and Rockafellar in 2000 [19]. There, it was used to analyze the stability of solutions to parametrized optimization problems. Para-prox-regular functions share the characteristics of prox-regular functions, but an extra level of complexity is added. The formal definition of parametric prox-regularity is presented in Definition 2.3.3. Little research has been done on para-prox-regular functions to date; the purpose of the present work is to take a step in that direction. We provide some examples of such functions, and develop an alternative characterization. We then examine several styles of constructing some commonly-used functions that are potentially para-prox-regular, and prove that indeed they are para-prox-regular under certain conditions.

The organization of this thesis is as follows. Chapter 2 contains basic definitions we will need, and the definitions of parametrized functions, prox-regular functions and para-prox-regular functions. It also provides some preliminary results for use in subsequent sections. Chapter 3 presents some examples of prox-regular and para-prox-regular functions. These include the proximal average [2, 3] and the NC-proximal average [14], which are parametrized functions designed to transform one function into another in a continuous manner. Chapter 4 contains an alternate definition of para-prox-regular functions. This result is similar to Poliquin's and Rockafellar's alternate definition of prox-regular functions presented in [25]. Chapter 5 presents three methods of constructing parametrized functions, and discusses when each method results in a para-prox-regular function. As a corollary, we answer an open question posed in [14]. The methods presented are: scalar multiplication of a function by a positive parameter (Lemma 5.2.3), the parametrized sum of a finite collection of functions (Theorem 5.2.5), and the finite parametrized max of continuous functions (Theorem 5.2.7). An expansion of [27, Theorem 3.2] is also seen in Chapter 5, extending the required two initial prox-regular functions to any finite number of prox-regular functions. Chapter 6 summarizes the work we have done and poses some suggestions for future research.

Chapter 2

Preliminary Definitions

This chapter includes the definitions of prox-regular and para-prox-regular functions. We will start with some more basic terms that we need in order to make the main definitions.

2.1 Basic Notation and Definitions

Throughout this thesis, we will use the following notation.

$\text{dom}\, f$	The domain of function f, i.e. all values of $x \in \mathbb{R}^n$ that deliver a real-valued $f(x)$. Example: if $f(x) := 2x + 1$ for $x \in \mathbb{R}$, then $\text{dom}\, f = \mathbb{R}$.
Proper function:	A function f is said to be *proper* if $f(x) > -\infty$ for all $x \in \mathbb{R}^n$, and there exists at least one $x \in \text{dom}\, f$ such that $f(x) < +\infty$.
$\mathcal{C}^0$	The set of all continuous functions.
$\mathcal{C}^1$	The set of all continuous, differentiable functions whose gradient is continuous.
$\mathcal{C}^2$	The set of all twice continuously differentiable functions whose gradient and Hessian are continuous.

For a vector $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, the *Euclidean norm* of x is denoted by $|x|$:

$$|x| := \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

For two vectors $x, y \in \mathbb{R}^n$, the *inner product* of x and y is denoted by $\langle x, y \rangle$:

$$\langle x, y \rangle := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

A set U is called *open* if for any $x \in U$ there exists $\epsilon > 0$ such that $V := \{y : |y - x| < \epsilon\} \subseteq U$.

A set U is called *closed* if fo any sequence $(x_n) \subseteq U$ that tends to some $\bar{x}$, $\bar{x} \in U$.

An *open ball* centered at $\bar{x}$ with radius δ, denoted $B_\delta(x)$, is defined as

$$B_\delta(\bar{x}) := \{x : |x - \bar{x}| < \delta\}.$$

A *closed ball* of the same type is denoted $\bar{B}_\delta(x)$, and is defined as

$$\bar{B}_\delta(x) := \{x : |x - \bar{x}| \leq \delta\}.$$

Definition 2.1.1. *(Derivative) A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be* Fréchet differentiable *(henceforth* differentiable*) if there exists a function $DT(x)$ such that*

$$\lim_{0 \neq \|y\| \to 0} \frac{\|f(x+y) - f(x) - DT(x)y\|}{\|y\|} = 0.$$

The function $DT(x)$ is called the Fréchet derivative *(henceforth* derivative*) of f at x, and $DT(x)$ defines the gradient of f at x, denoted $\nabla f(x)$:*

$$DT(x)y = \langle y, \nabla f(x) \rangle.$$

For a set of real numbers U, the *supremum* of U, denoted $\sup U$, is the least upper bound of U. That is, $\sup U$ is the smallest real number that is greater than or equal to every element of U. Similarly, the *infimum* of U, denoted $\inf U$, is the greatest lower bound of U; that is, the biggest real number that is less than or equal to every element of U.

The *supremum* of $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, denoted $\sup f$, is the smallest number that is greater than or equal to $f(x)$ for all $x \in \operatorname{dom} f$. In other words, it is the smallest upper bound of f. If no such number exists, the supremum is $+\infty$. If the supremum is equal to $f(x)$ for some $x \in \operatorname{dom} f$, then it is the *maximum* value of f and can also be denoted $\max f$.

Similarly, the *infimum* of f, denoted $\inf f$, is the greatest number that is less than or equal to $f(x)$ for all $x \in \operatorname{dom} f$. If no such number exists, the infimum is $-\infty$. If the infimum is equal to $f(x)$ for some $x \in \operatorname{dom} f$, then it is the *minimum* value of f and can also be denoted $\min f$.

If $\max f$ exists, then the set of all maximizers of f is denoted $\operatorname{argmax} f$:

$$\operatorname{argmax} f := \{x \in \mathbb{R}^n \text{ such that } f(x) = \max f\}.$$

If $\min f$ exists, then the set of all minimizers of f is denoted $\operatorname{argmin} f$:

$$\operatorname{argmin} f := \{x \in \mathbb{R}^n \text{ such that } f(x) = \min f\}.$$

Definition 2.1.2. *(Lower Semicontinuity) A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be* lower semicontinuous (lsc) *at $\bar{x}$ if for every $\epsilon > 0$ there exists a neighborhood U of $\bar{x}$ such that $f(x) \geq f(\bar{x} - \epsilon)$ for all $x \in U$. If f is lsc at x for all x such that $|x - \bar{x}| < \delta$ for some $\delta > 0$, then f is said to be* locally lsc *around $\bar{x}$.*

Definition 2.1.3. *(Lipschitz Continuity) A function $f : D \subseteq \mathbb{R}^m \to \mathbb{R}^n$ is called* Lipschitz continuous *if*

$$K := \sup_{x_1, x_2 \in D} \left\{ \frac{|f(x_1) - f(x_2)|}{|x_2 - x_1|} \right\}$$

is a real number. If so, then K is called the Lipschitz constant *of f, denoted* $\operatorname{Lip} f$.
If, rather than considering all $x_1, x_2 \in D$, we find that K is real for all x_1, x_2 such that $|x_1 - \bar{x}| < \epsilon$ and $|x_2 - \bar{x}| < \epsilon$ for some $\bar{x} \in D$ and some $\epsilon < 0$, then f is called locally Lipschitz continuous *at $\bar{x}$, with* local Lipschitz constant K *denoted* $\operatorname{Lip}(f, \bar{x})$.

Definition 2.1.4. *(Epigraph) The* epigraph *of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as*

$$\operatorname{epi}_f := \{(x, \alpha) : \alpha \geq f(x)\}.$$

It is the set of all points that lie on or above the curve of f.

Definition 2.1.5. *(Convex Set) A set $S \in \mathbb{R}^n$ is* convex *if for every $x_1, x_2 \in S$, the line segment $\alpha x_1 + (1 - \alpha) x_2$ for $\alpha \in [0, 1]$ is a subset of S.*

Definition 2.1.6. *(Convex Function) A function $f : \mathbb{R}^n \to \mathbb{R}$ is a* convex function *if* epi_f *is a convex set.*

Definition 2.1.7. *(Convex Hull) For any set $S \in \mathbb{R}^n$, the* convex hull *of S, denoted* $\operatorname{conv} S$*, is the smallest convex set that contains S.*

Definition 2.1.8. *(Little-o) For two real functions f and g, we say $f(x) = o(g(x))$ at x_0 if and only if $\lim\limits_{x \to x_0} \frac{f(x)}{g(x)} = 0$, with $x \neq x_0$ in limit and $g(x) \neq 0$ near x_0.*

2.2 Subgradients

The first definition we will see in detail is that of subgradients. For a continuously differentiable function, we know from calculus that its gradient exists everywhere on its domain, and that the gradient is calculated using the partial derivatives of the function [28, p. 47]. The gradient gives important information, such as the direction of steepest ascent of the function at any point. However, many of the functions we routinely use in the optimization field are not so well-behaved. One example is the norm function, $f(x) := |x|$. This function has a point of non-differentiability at the origin. For functions such as these, the definition of gradient fails us and we cannot apply the associated rules of calculus. So we would like to define a similar concept to the gradient that applies to non-smooth functions. That similar concept is what we call the subgradient. There are several types of subgradients; we first state their definitions, and follow with some examples and graphical interpretations.

Definition 2.2.1. *[28, Definition 8.3] Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper, $\bar{x} \in \operatorname{dom} f$. A vector $\bar{v} \in \mathbb{R}^n$ is*

a) *a* regular subgradient *of f at $\bar{x}$, written $\bar{v} \in \hat{\partial} f(\bar{x})$, if for all $x \in \operatorname{dom} f$*

$$f(x) \geq f(\bar{x}) + \langle \bar{v}, x - \bar{x} \rangle + o(|x - \bar{x}|). \qquad (2.2.1)$$

b) *a* (general) subgradient *of f at $\bar{x}$, written $\bar{v} \in \partial f(\bar{x})$, if there exist sequences $\{x^k\}_{k=1}^{\infty} \subseteq \operatorname{dom} f$ and $v^k \in \hat{\partial} f(x^k)$ for all $k \in \mathbb{N}$, such that $x^k \to \bar{x}$, $f(x_k) \to f(\bar{x})$, and $v^k \to \bar{v}$.*

c) *a* horizon subgradient *of f at $\bar{x}$, written $v \in \partial^{\infty} f(\bar{x})$, if the same holds as in b) except that instead of $v^k \to v$ one has $\lambda^k v^k \to v$ for some sequence $\lambda^k \searrow 0$.*

d) *a* proximal subgradient *of f at $\bar{x}$ if, when the error term $o(|x - \bar{x}|)$ of inequality (2.2.1) is $\frac{r}{2}|x - \bar{x}|^2$, the inequality holds for all x such that $|x - \bar{x}| < \delta$ for some $\delta > 0$.*

The set of all regular subgradients of f at $\bar{x}$ is denoted $\hat{\partial} f(\bar{x})$ and is called the *regular subdifferential* of f at $\bar{x}$. Similarly, the set of subgradients is denoted $\partial f(\bar{x})$ and is called the *subdifferential* and the set of horizon subgradients is denoted $\partial^\infty f(\bar{x})$ and called the *horizon subdifferential*. For a function $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, $f(x, \lambda)$, we shall also use $\partial_x f(\cdot, \bar{\lambda})$ to denote the subdifferential with respect to x.

To understand what subgradients are, we begin with the case of convex functions. Recall that a continuously differentiable function f is *convex* on an open set $U \subseteq \operatorname{dom} f$ if and only if

$$f(x) \geq f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle \tag{2.2.2}$$

for all $x, \bar{x} \in U$. For a convex non-differentiable function then, we would like to have a similar definition of convexity that employs subgradients at points of non-differentiability. Note that if we choose the zero function for the last term in inequality (2.2.1), we are left with

$$f(x) \geq f(\bar{x}) + \langle \bar{v}, x - \bar{x} \rangle \tag{2.2.3}$$

for all $x, \bar{x} \in U$. Any vector $\bar{v}$ that satisfies inequality (2.2.3) is referred to as a *convex subgradient*. In the case of a continuously differentiable function f, inequality (2.2.3) reverts to inequality (2.2.2), since the only subgradient of $f(\bar{x})$ is $\nabla f(\bar{x})$ [28, Proposition 8.5], as the next paragraph illustrates.

Graphically speaking, a differentiable function f is convex if and only if the function lies entirely above the linear function

$$f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle$$

of inequality (2.2.2). Here the gradient $\nabla f(\bar{x})$ describes the slopes of said linear functions at the point $(\bar{x}, f(\bar{x}))$. For example, the differentiable function $f : \mathbb{R} \to \mathbb{R}, f(x) = x^2$ is convex, since any tangent line to the function at any point lies completely below the graph of the function (see Figure 2.1). Now we can see why the gradient is the only subgradient of a continuously differentiable function, because a tangent line that contains $f(\bar{x})$ and has any slope other than $\nabla f(x)$ would not lie entirely below the graph of f [28, Exercise 8.8].

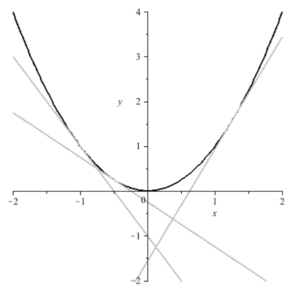

Figure 2.1: The function x^2 and some tangent lines.

For non-smooth convex functions, the convex subgradient has the same graphical interpretation. That is, a convex subgradient of f at $\bar{x}$ is any vector $\bar{v}$ such that the function $f(\bar{x}) + \langle \bar{v}, x - \bar{x} \rangle$ lies entirely below the graph of the function f. One of the differences here is that a subgradient at a point may not be unique. There is only one gradient at each point of a $\mathcal{C}^1$ function, but for a non-$\mathcal{C}^1$ function there might be multiple subgradients at the same point. If we look at the graph of the norm function (Figure 2.2), which is convex and non-smooth, then we see there are many lines that pass through the origin and lie below the graph of the function. Each of those lines is defined by a distinct subgradient.

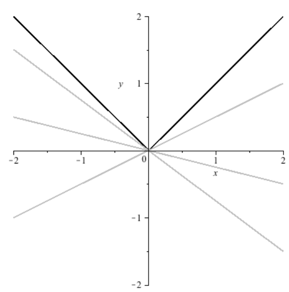

Figure 2.2: The norm function has multiple subgradients at the origin.

Now we move on to the non-convex case. On our way to visualizing (non-convex) subgradients, we define the epigraph, projection, and normal cone.

Definition 2.2.2. *The* epigraph *of a function* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, *denoted* epi f, *is the set of all points that lie on or above the graph of* *f. That is,*

$$\operatorname{epi} f := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} : \alpha \geq f(x)\}.$$

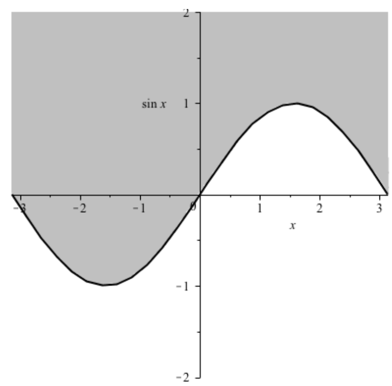

Figure 2.3: The epigraph of $\sin x$.

Definition 2.2.3. *Let $S = \operatorname{epi} f$. The* projection *of a point x onto S, denoted $P_S(x)$, is the set of all points in S that are closest to x. That is,*

$$P_S(x) = \{\bar{x} \in S : |x - \bar{x}| \leq |x - y| \text{ for all } y \in S\}.$$

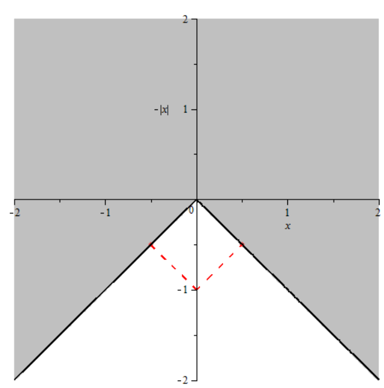

Figure 2.4: The projection of $(0, -1)$ onto $\operatorname{epi}_{-|x|}$.

Definition 2.2.4. *For a set $S \in \mathbb{R}^n$, the* proximal normal cone *to S at a point $y \in S$ is the cone*

$$N_S^P(y) = \{\lambda(x - y) : \lambda \geq 0, y \in P_S(x)\}.$$

In other words, the proximal normal cone at y consists of all vectors, called *proximal normals*, that point along the directions of projection back onto y.

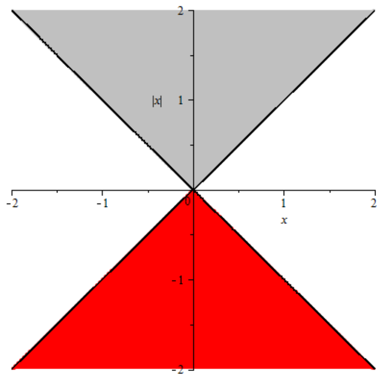

Figure 2.5: The proximal normal cone to $\text{epi}_{|x|}$ at the origin.

With this in mind, we may now give a graphical interpretation of proximal subgradients.

Proposition 2.2.5. *[28, Proposition 8.46] A vector v is a* proximal subgradient *to f at $\bar{x}$ if and only if $(v, -1)$ is a proximal normal to* $\text{epi}\, f$ *at $(\bar{x}, f(\bar{x}))$.*

For example, referring to Figure 2.5, we can see that the vector $(x, -1)$ lies in the red area if $x \in [-1, 1]$, and it does not otherwise. The interval $[-1, 1]$ then, is the set of proximal subgradients to $|x|$ at $x = 0$.

As we saw in Definition 2.2.1 b), if there exist sequences $\{x^k\}$ and $\{v^k\}$ such that $f(x^k) \to f(\bar{x})$, each v^k is a proximal subgradient of the corresponding $f(x^k)$, and $v^k \to \bar{v}$, then $\bar{v}$ is a subgradient of f at $\bar{x}$.

Finally, the horizon subgradients describe what happens at the outer edges of the normal cone. Consider the example of

$$f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}, f(x) = x^{\frac{1}{3}}.$$

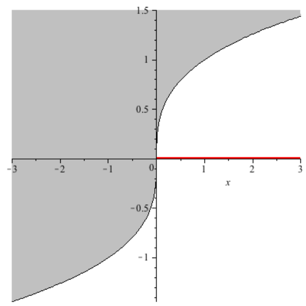

Figure 2.6: The graph of $f(x) = x^{\frac{1}{3}}$, and its normal cone at $x = 0$ in red.

In this case the proximal subdifferential is empty, since there does not exist any $x \in \mathbb{R}$ such that $(x, -1) \in N^P_{\operatorname{epi} f}(0, 0)$. However, the horizon subdifferential is not empty. To see this, let $v \in \mathbb{R}_+ = [0, +\infty)$. Consider any sequence $(x_k) \subseteq \mathbb{R}$ such that $x_k \to 0$ and $x_k \neq 0$. Then the corresponding sequence of proximal subgradients is (v_k) such that $v_k = \nabla f(x_k) = \frac{1}{3} x_k^{-\frac{2}{3}}$. By using the sequence of scalars $\lambda_k = 3v x_k^{\frac{2}{3}}$, we see that $\lambda_k \geq 0$ for all k, and that $\lambda_k \searrow 0$ since v is fixed and $x_k^{\frac{2}{3}} \searrow 0$. Then we have that $\lambda_k v_k = 3v x_k^{\frac{2}{3}} \frac{1}{3} x_k^{-\frac{2}{3}} = v$ for all k, so that $\lambda_k v_k \to v$. Hence, $\mathbb{R}_+ \subseteq \partial^\infty f(0)$. Now considering any $v < 0$, since $v_k \geq 0$ for all k, and since we must have $\lambda_k \geq 0$ for all k, then there do not exist sequences (v_k) and (λ_k) such that $\lambda_k v_k \to v$. Therefore, $\partial^\infty f(0) = \mathbb{R}_+$.

2.3 Prox-Regularity

Now we are ready for the definition of prox-regular functions.

Definition 2.3.1. *[28, Definition 13.27] A proper function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ that is finite at $\bar{x}$ is* prox-regular *at $\bar{x}$ for $\bar{v}$ with parameters $\epsilon > 0, r \geq 0$, where $\bar{v} \in \partial f(\bar{x})$, if and only if f is locally lsc at $\bar{x}$ and*

$$f(x') \geq f(x) + \langle v, x' - x \rangle - \frac{r}{2}|x' - x|^2 \tag{2.3.1}$$

whenever $|x' - \bar{x}| < \epsilon$, $|x - \bar{x}| < \epsilon$, $|f(x) - f(\bar{x})| < \epsilon$, $v \in \partial f(x)$ and $|v - \bar{v}| < \epsilon$. If this holds true for all $\bar{v} \in \partial f(\bar{x})$, then f is said to be prox-regular at $\bar{x}$ (without reference to any subgradient). If f is prox-regular for all $\bar{x} \in \operatorname{dom} f$, then it is said to be a prox-regular function (without reference to any point).

Prox-regularity implies for all x near enough to $\bar{x}$, v near enough to $\bar{v}$, and $f(x)$ near enough to $f(\bar{x})$, that v is a proximal subgradient of f at x, and that the radius or curvature $\frac{1}{r}$ of the quadratic term $-\frac{r}{2}|x-\bar{x}|^2$ is constant [28, Definition 8.45].

Let us consider a graphical interpretation of prox-regular functions. This is not a formal definition, but a visual aid so that the reader may better visualize what is happening. Graphically speaking, due to the last term on the right-hand side of (2.3.1), we can think of prox-regular functions as those functions which are locally bounded below by a tangent concave quadratic function with radius of curvature $-\frac{1}{r}$.

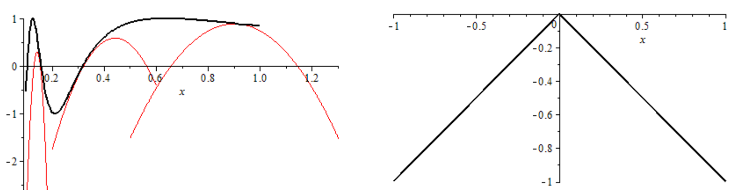

Figure 2.7: Left: a prox-regular function. Right: a function that is not prox-regular at $x = 0$.

In Figure 2.7, we have on the left a prox-regular function. At each point there exists a parabola that bounds the function below in a neighborhood of that point; three such parabolas are illustrated in red for three different points on the curve. Notice that for different points $(\bar{x}, f(\bar{x}))$, the values of ϵ and r may be different. Prox-regularity is a local property, and as such the ϵ-neighborhoods and the radii of curvature of the bounding parabolas are not necessarily the same for every point. Instead it only holds locally; for every point $\bar{x}$ at which f is prox-regular, there exist $\epsilon > 0$ and a tangent bounding parabola that is also tangent and bounding at every point in an ϵ-neighborhood of $\bar{x}$.

The function on the right of Figure 2.7, however, is not prox-regular at the origin. There is an interior sharp point there, hence there does not exist a parabola tangent to the curve at that point that lies entirely below the function.

An alternate representation of prox-regularity, which will be used in Chapter 5, is described in the following theorem.

Theorem 2.3.2. *[25, Theorem 3.2] A proper function* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *is prox-regular at* $\bar{x}$ *for* $\bar{v}$ *with parameters* $\epsilon > 0, r \geq 0$ *if and* *only if f is locally lsc at $\bar{x}$, $\bar{v} \in \partial f(\bar{x})$ and*

$$\langle v' - v, x' - x \rangle \geq -r|x' - x|^2$$

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

Theorem 2.3.2 gives us an expression that defines prox-regularity in terms of local pre-monotonicity of the subdifferential operator. This form is the inspiration for our main result in Chapter 4.

Para-prox-regularity is a natural extension of prox-regularity that allows for a parametrized function. The definition is the following.

Definition 2.3.3. *[19, Definition 2.1] The proper function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is* para-prox-regular *in x at $\bar{x}$ for $\bar{v}$ with compatible parametrization by λ at $\bar{\lambda}$ (also said to be para-prox-regular in x at $(\bar{x}, \bar{\lambda})$ for $\bar{v}$), with parameters $\epsilon > 0, r \geq 0$, if and only if f is locally lsc at $\bar{x}$, $\bar{v} \in \partial_x f(\bar{x}, \bar{\lambda})$ and*

$$f(x', \lambda) \geq f(x, \lambda) + \langle v, x' - x \rangle - \frac{r}{2}|x' - x|^2 \qquad (2.3.2)$$

whenever $|x' - \bar{x}| < \epsilon$, $|x - \bar{x}| < \epsilon$, $|f(x, \lambda) - f(\bar{x}, \bar{\lambda})| < \epsilon$, $v \in \partial_x f(x, \lambda)$, $|v - \bar{v}| < \epsilon$, and $|\lambda - \bar{\lambda}| < \epsilon$. The function f is continuously para-prox-regular in x at $(\bar{x}, \bar{\lambda})$ for $\bar{v}$ if, in addition, $f(x, \lambda)$ is continuous as a function of $(x, v, \lambda) \in \operatorname{gph} \partial_x f$ at $(\bar{x}, \bar{v}, \bar{\lambda})$. It is para-prox-regular in x at $(\bar{x}, \bar{\lambda})$ (with no mention of $\bar{v}$) if it is para-prox-regular in x at $(\bar{x}, \bar{\lambda})$ for all $\bar{v} \in \partial_x f(\bar{x}, \bar{\lambda})$, and it is a para-prox-regular function in x (with no mention of $\bar{x}$ or $\bar{\lambda}$) if it is para-prox-regular in x at $(\bar{x}, \bar{\lambda})$ for all $(\bar{x}, \bar{\lambda}) \in \operatorname{dom} f$.

A para-prox-regular function f of (x, λ) is a function that is prox-regular as a function of x at $(\bar{x}, \bar{\lambda})$, and continues to be so at $f(x, \lambda)$ for all x and for all λ in ϵ-neighborhoods of $\bar{x}$ and $\bar{\lambda}$, respectively. That is, if we do not wander too far away from $(\bar{x}, \bar{\lambda})$ in x or in λ, the prox-regularity property of f and its corresponding prox-regularity parameters are preserved.

Para-prox-regular functions were first presented by Levy, Poliquin and Rockafellar in [19], where they were used to analyze the stability of locally optimal solutions to minimization problems involving parametrized functions. They considered problems of the type

$$\text{minimize } f(x, \bar{\lambda}), \ x \in \mathbb{R}^n$$

and compared the solutions with those obtained by perturbing the function slightly by changing $\bar{\lambda}$ to a nearby λ. If a small change in λ results in only a small change in the solutions, the function is considered stable. Function stability is of vital importance in many real-world situations; consider the following example as an illustration.

Imagine it is your job to design a sport utility vehicle for retail sale. Such vehicles have a higher tendency to roll over compared to other vehicles, due to their higher center of gravity. So one of the concerns in designing an SUV is, given a constant highway speed of 80 kph, the vehicle must be able to navigate curves in the road of a certain radius, the curves that one typically finds on the highways. Let us suppose that you are successful in designing your vehicle to handle such curves at that speed. However, how many drivers drive at or below the posted speed limit? Since we know the rules of the road are not followed rigorously by everyone, it should be of great interest for you to know what will happen if a driver does not respect the 80 kph road sign and tries to take a curve at a higher rate of speed. Will the vehicle roll over at 81, 82, 85 kph? Can one drive as fast as 90 without danger? We would hope that a small increase in speed would result in only a small increase in the level of risk to the people in the vehicle.

In attemping to answer these types of questions, one may think of the problem as an analysis of the stability of the function 'level of risk' in terms of the parameter 'speed,' close to $\bar{\lambda} = 80$ kph. This examples serves to show that para-prox-regular functions, since they are used in such analyses, have a very important (sometimes even life-or-death important) role to play in the world. In the next chapter, we shall see some examples of specific functions that are para-prox-regular.

Chapter 3

Basic Examples, Proximal Average

3.1 Basic Examples

We begin with a few basic examples of para-prox-regular functions. The first example is the result of multiplying the function $|x|$ by the parameter λ.

Example 3.1.1. *For $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$, define*

$$f(x,\lambda) := \lambda|x|.$$

Then f is para-prox-regular at $(0,\bar{\lambda})$ for any $\bar{\lambda} > 0$, and it is not para-prox-regular at $(0,\bar{\lambda})$ at any $\bar{\lambda} \leq 0$, for any $\bar{v} \in \partial_x f(\bar{x},\bar{\lambda})$.

Proof: Let $\bar{x} = 0$ and $\bar{\lambda} > 0$. Then

$$f(x',\bar{\lambda}) \geq f(x,\bar{\lambda}) + \langle v, x' - x\rangle$$

for any x', x and for any $v \in \partial_x f(x,\bar{\lambda})$, as f is convex in x and $\bar{\lambda} > 0$. The same is true if we replace $\bar{\lambda}$ with any $\lambda > 0$, for the same reasons. Thus, for any $\epsilon \in (0,\bar{\lambda})$ (so that $|\lambda - \bar{\lambda}| < \epsilon$ implies $\lambda > 0$) and $r = 0$ we have

$$f(x',\lambda) \geq f(x,\lambda) + \langle v, x' - x\rangle - \frac{r}{2}|x' - x|^2$$

for $|x' - \bar{x}| < \epsilon$, $|x - \bar{x}| < \epsilon$, $|f(x,\lambda) - f(\bar{x},\bar{\lambda})| < \epsilon$, $v \in \partial_x f(x,\lambda)$, $|v - \bar{v}| < \epsilon$, and $|\lambda - \bar{\lambda}| < \epsilon$. That is, f is para-prox-regular at $(\bar{x},\bar{\lambda})$.

For $\bar{\lambda} < 0$, f ceases to be a prox-regular function at $\bar{x} = 0$, since $\partial_x f(\bar{x},\bar{\lambda}) = \emptyset$. So the conditions of para-prox-regularity immediately fail. For $\bar{\lambda} = 0$ para-prox-regularity fails as well, since we require f to be prox-regular in a neighborhood of $\bar{\lambda}$, but any neighborhood of $\bar{\lambda} = 0$ necessarily contains some $\lambda < 0$. $\square$

Notice that for $\bar{x} \neq 0$, f is para-prox-regular no matter the value of

$\bar{\lambda}$. This is because the function is continuously differentiable there, so all subgradients are gradients, and inequality (2.3.2) becomes trivially true.

This example hints at the next result: if a function is convex in terms of x no matter the value of λ, then it is para-prox-regular. This result will be used in Section 3.2 to prove that the proximal average is para-prox-regular.

Lemma 3.1.2. *Let* $(\bar{x}, \bar{\lambda}) \in \operatorname{dom} f$, *where* $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ *is convex as a function of* x *for all* λ *such that* $|\lambda - \bar{\lambda}| < \epsilon$, *where* $\epsilon > 0$. *Then* f *is para-prox-regular in* x *at* $(\bar{x}, \bar{\lambda})$, *with parameters* $r = 0$ *and* ϵ.

Proof: Since $f(x, \bar{\lambda})$ is convex in terms of x, for any $x \in \operatorname{dom} f(\cdot, \bar{\lambda})$ and any $v \in \partial_x f(x, \bar{\lambda})$ we have that

$$f(x', \bar{\lambda}) \geq f(x, \bar{\lambda}) + \langle v, x' - x \rangle \tag{3.1.1}$$

for all $x' \in \operatorname{dom} f(\cdot, \lambda)$. For every λ such that $|\lambda - \bar{\lambda}| < \epsilon$, since f remains convex, for any $x \in \operatorname{dom} f(\cdot, \lambda)$ and any $v_\lambda \in \partial f_x(x, \lambda)$ we have that

$$f(x', \lambda) \geq f(x, \lambda) + \langle v_\lambda, x' - x \rangle \tag{3.1.2}$$

for all $x' \in \operatorname{dom} f(\cdot, \lambda)$. In particular, inequality (3.1.2) is true for all x' and x such that $|x' - \bar{x}| < \epsilon$, $|x - \bar{x}| < \epsilon$ and $|f(x, \lambda) - f(\bar{x}, \bar{\lambda})| < \epsilon$ for any $\bar{x}$ fixed, and for all $v_\lambda \in \partial_x f(x, \lambda)$ such that $|v_\lambda - v| < \epsilon$. This demonstrates para-prox-regularity with parameters $r = 0$ and ϵ. □

The next two examples demonstrate that we can take a prox-regular function f and use the parameter λ in the argument of f to create para-prox-regular functions at particular $\bar{\lambda}$ values. Example 3.1.3 examines a linear shift of the argument by λ and Example 3.1.5 examines a scalar multiple of the argument by λ.

Example 3.1.3. *Let* $\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *be prox-regular at* $\bar{x}$ *for* $\bar{v} \in \partial \tilde{f}(\bar{x})$, *and let* $\lambda \in \mathbb{R}^n$. *Define*

$$f(x, \lambda) := \tilde{f}(x - \lambda).$$

Then f *is para-prox-regular at* $(\bar{x}, 0)$.

Proof: By assumption, there exist $\tilde{\epsilon} > 0$ and $\tilde{r} \geq 0$ such that

$$\tilde{f}(\tilde{x}') \geq \tilde{f}(\tilde{x}) + \langle v, \tilde{x}' - \tilde{x} \rangle - \frac{\tilde{r}}{2}|\tilde{x}' - \tilde{x}|^2 \tag{3.1.3}$$

whenever $\tilde{x}' \neq \tilde{x}$, $|\tilde{x}' - \bar{x}| < \tilde{\epsilon}$, $|\tilde{x} - \bar{x}| < \tilde{\epsilon}$, $|\tilde{f}(\tilde{x}) - \tilde{f}(\bar{x})| < \tilde{\epsilon}$, $v \in \partial \tilde{f}(\tilde{x})$, $|v - \bar{v}| < \tilde{\epsilon}$. We need to show that when $\bar{\lambda} = 0$, there exist $\epsilon > 0$ and $r \geq 0$ such that

$$f(x', \lambda) \geq f(x, \lambda) + \langle v, x' - x \rangle - \frac{r}{2}|x' - x|^2 \tag{3.1.4}$$

when $x' \neq x$, $|x' - \bar{x}| < \epsilon$, $|x - \bar{x}| < \epsilon$, $|\lambda - \bar{\lambda}| < \epsilon$, $|f(x, \lambda) - f(\bar{x}, \bar{\lambda})| < \epsilon$, $v \in \partial_x f(x, \lambda)$ and $|v - \bar{v}| < \epsilon$. Let $\epsilon = \frac{\tilde{\epsilon}}{2}$, $r = \tilde{r}$. Since $\epsilon < \tilde{\epsilon}$, inequality (3.1.3) holds when we replace $\tilde{r}$ with r and $\tilde{\epsilon}$ with ϵ everywhere. Select any λ such that $|\lambda - \bar{\lambda}| < \epsilon$, and set $\tilde{x}' = x' - \lambda$ and $\tilde{x} = x - \lambda$. Then

$$\begin{aligned} |\tilde{x} - \bar{x}| &= |x - \bar{x} - \lambda| \\ &\leq |x - \bar{x}| + |\lambda| \\ &< \epsilon + \epsilon \\ &= \tilde{\epsilon}, \end{aligned}$$

so we have that $|\tilde{x} - \bar{x}| < \tilde{\epsilon}$. Similarly, $|\tilde{x}' - \bar{x}| < \tilde{\epsilon}$. If in addition we restrict our function values using ϵ instead of $\tilde{\epsilon}$, then we have (remembering $\bar{\lambda} = 0$) that

$$|\tilde{f}(\tilde{x}) - \tilde{f}(\bar{x})| < \epsilon \Leftrightarrow |\tilde{f}(x - \lambda) - \tilde{f}(\bar{x} - \bar{\lambda})| < \epsilon \Leftrightarrow |f(x, \lambda) - f(\bar{x}, \bar{\lambda})| < \epsilon.$$

Finally we observe that $v \in \partial \tilde{f}(\tilde{x})$ if and only if $v \in \partial_x f(x, \lambda)$, so we may rewrite inequality (3.1.3) as

$$\tilde{f}(x' - \lambda) \geq \tilde{f}(x - \lambda) + \langle v, (x' - \lambda) - (x - \lambda) \rangle - \frac{r}{2}|(x' - \lambda) - (x - \lambda)|^2$$

when $x' \neq x$, $|x' - \bar{x}| < \epsilon$, $|x - \bar{x}| < \epsilon$, $|\lambda - \bar{\lambda}| < \epsilon$, $|f(x, \lambda) - f(\bar{x}, \bar{\lambda})| < \epsilon$, $v \in \partial_x f(x, \lambda)$, $|v - \bar{v}| < \epsilon$. That is, inequality (3.1.4) holds under the required conditions. □

For the next example, we first need a lemma that describes the subgradient chain rule.

Lemma 3.1.4. *[28, Example 10.7] Suppose $f(x) = g(F(x))$ for a proper, lsc function $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ and a smooth mapping $F : \mathbb{R}^n \to \mathbb{R}^m$, and let $\bar{x}$ be a point where f is finite and the gradient $\nabla F(\bar{x})$ has rank m. Then*

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

Example 3.1.5. *Let $\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be prox-regular at $\bar{x}$ for $\bar{v}$, and let $\lambda \in \mathbb{R}$. Define*

$$f(x, \lambda) := \tilde{f}(\lambda x).$$

Then f is para-prox-regular at $(\bar{x}, 1)$ for $\bar{v}$.

Proof: By assumption, there exist prox-regularity parameters $\tilde{\epsilon} > 0$ and $\tilde{r} \geq 0$ such that

$$\tilde{f}(\tilde{x}') \geq \tilde{f}(\tilde{x}) + \langle v, \tilde{x}' - \tilde{x} \rangle - \frac{\tilde{r}}{2}|\tilde{x}' - \tilde{x}|^2 \tag{3.1.5}$$

for $|\tilde{x}' - \bar{x}| < \tilde{\epsilon}$, $|\tilde{x} - \bar{x}| < \tilde{\epsilon}$, $|\tilde{f}(\tilde{x}) - \tilde{f}(\bar{x})| < \tilde{\epsilon}$, $v \in \partial \tilde{f}(\tilde{x})$, and $|v - \bar{v}| < \tilde{\epsilon}$. Substituting $\tilde{x}' = \lambda x'$ and $\tilde{x} = \lambda x$, inequality (3.1.5) becomes

$$\tilde{f}(\lambda x') \geq \tilde{f}(\lambda x) + \langle v, \lambda x' - \lambda x \rangle - \frac{\tilde{r}}{2}|\lambda x' - \lambda x|^2 \tag{3.1.6}$$

for $|\lambda x' - \bar{x}| < \tilde{\epsilon}$, $|\lambda x - \bar{x}| < \tilde{\epsilon}$, $|\tilde{f}(\lambda x) - \tilde{f}(\bar{x})| < \tilde{\epsilon}$, $v \in \partial \tilde{f}(\lambda x)$, and $|v - \bar{v}| < \tilde{\epsilon}$. The subgradient chain rule (Lemma 3.1.4 above) gives us $\lambda v \in \partial_x f(x, \lambda)$, so it is convenient to rewrite inequality (3.1.6) as

$$f(x', \lambda) \geq f(x, \lambda) + \langle \lambda v, x' - x \rangle - \frac{\tilde{r}\lambda^2}{2}|x' - x|^2 \tag{3.1.7}$$

for $|\lambda x' - \bar{x}| < \tilde{\epsilon}$, $|\lambda x - \bar{x}| < \tilde{\epsilon}$, $|f(x, \lambda) - f(\bar{x}, \bar{\lambda})| < \tilde{\epsilon}$, $\lambda v \in \partial_x f(x, \lambda)$, and $|v - \bar{v}| < \tilde{\epsilon}$. If we restrict our neighborhood of $\bar{x}$ further, say $|x' - \bar{x}| < \frac{\tilde{\epsilon}}{2}$ and $|x - \bar{x}| < \frac{\tilde{\epsilon}}{2}$, then there exists $\hat{\epsilon} > 0$ such that for all λ with $|\lambda - \bar{\lambda}| < \hat{\epsilon}$ and if $|x' - \bar{x}| < \frac{\tilde{\epsilon}}{2}$ and $|x - \bar{x}| < \frac{\tilde{\epsilon}}{2}$, then we necessarily have that $|\lambda x' - \bar{x}| < \tilde{\epsilon}$ and $|\lambda x - \bar{x}| < \tilde{\epsilon}$. Hence by taking $\epsilon = \min\{\frac{\tilde{\epsilon}}{2}, \hat{\epsilon}\}$, inequality (3.1.7) remains true for $|x' - \bar{x}| < \epsilon$, $|x - \bar{x}| < \epsilon$, $|f(x, \lambda) - f(\bar{x}, \bar{\lambda})| < \epsilon$, $\lambda v \in \partial_x f(x, \lambda)$, $|v - \bar{v}| < \epsilon$, and $|\lambda - \bar{\lambda}| < \epsilon$. Since $|\lambda - \bar{\lambda}| < \epsilon$, we may take $r = \max\{\tilde{r}, \tilde{r}|\bar{\lambda} + \epsilon|^2\}$. Renaming λv as $\tilde{v}$ we have that

$$f(x', \lambda) \geq f(x, \lambda) + \langle \tilde{v}, x' - x \rangle - \frac{r}{2}|x' - x|^2 \tag{3.1.8}$$

for $|x' - \bar{x}| < \epsilon$, $|x - \bar{x}| < \epsilon$, $|\lambda - \bar{\lambda}| < \epsilon$, $|f(x, \lambda) - f(\bar{x}, \bar{\lambda})| < \epsilon$, $\tilde{v} \in \partial_x f(x, \lambda)$, and $|\tilde{v} - \bar{v}| < \epsilon$. □

3.2 Proximal Average

The next examples are those of the *proximal average* and *NC-proximal average*, for which we will need some background information. Recall that for a proper function f, the *Fenchel conjugate* f^* is defined as

$$f^*(y) := \sup_y \{\langle x, y \rangle - f(y)\}.$$

For two convex functions f_0 and f_1, using the Fenchel conjugate, the proximal average is defined as follows.

Definition 3.2.1. *[2, Definition 4.1] The* proximal average PA_{f_0,f_1} *of two proper convex functions f_0 and f_1 is defined as*

$$PA_{f_0,f_1}(x, \lambda) := ((1-\lambda)(f_0(x) + \frac{1}{2}|x|^2)^* + \lambda(f_1(x) + \frac{1}{2}|x|^2)^*)^* - \frac{1}{2}|x|^2$$

for $\lambda \in [0, 1]$.

The proximal average has been shown to be an effective method of transforming one convex function into another in a continuous manner [2]. It has also been shown to be a convex function [3, Theorem 6.1], and as such it is prox-regular. In fact, it is para-prox-regular, as the following proposition demonstrates.

Proposition 3.2.2. *The proximal average PA_{f_0,f_1} of two proper convex functions f_0, f_1 is para-prox-regular for all $\lambda \in (0, 1)$.*

Proof: Bauschke, Matoušková and Reich [3, Theorem 6.1] proved that this function is convex in x for all $\lambda \in (0, 1)$, therefore it satisfies the conditions of Lemma 3.1.2. □

In [14], an alternate representation of the proximal average function was given using Moreau envelopes. The convexity of the proximal average is further studied in [17]. Recall that for a function f the *Moreau envelope* with parameter $r \geq 0$ is defined as

$$e_r f(x) := \inf_y \{f(y) + \frac{r}{2}|y - x|^2\},$$

and its associated *proximal point mapping* is defined as

$$P_r f(x) := \operatorname*{argmin}_y \{f(y) + \frac{r}{2}|y - x|^2\}.$$

If f is proper and convex, then $e_r f$ is proper for all $r > 0$. Properties of the Moreau envelope give rise to the equivalent definition of proximal average [14, equation (4)],

$$PA_{f_0,f_1}(x,\lambda) = -e_1(-(1-\lambda)e_1 f_0 - \lambda e_1 f_1)(x).$$

However, this formulation is only useful if f_0 and f_1 are convex functions. This is because the double envelope of a function is the original function only when it is a convex function. For two not necessarily convex functions f_0 and f_1, the *NC-proximal average* PA_r is similarly defined (see Definition 3.2.4 below). We need one more definition first, that of prox-boundedness.

Definition 3.2.3. *[28, Definition 1.23] A function f is* prox-bounded *if there exists $r > 0$ and a point $\bar{x}$ such that $e_r f(\bar{x}) > -\infty$. The infimum of the set of all such r is called the* threshold of prox-boundedness*.*

Definition 3.2.4. *[14, equation (5)] Let f_0 and f_1 be proper, prox-bounded functions, and let $r > 0$ be greater than the threshold of prox-boundedness for both f_0 and f_1. The* NC-proximal average *of f_0 and f_1 is defined as*

$$PA_r(x,\lambda) := -e_{r+\lambda(1-\lambda)}(-(1-\lambda)e_r f_0 - \lambda e_r f_1)(x).$$

Showing that the NC-proximal average is para-prox-regular requires a technical condition from [14, Theorem 4.6], specifically that the Lipschitz constant of $\lambda P_r f_0 + (1-\lambda)P_r f_1 - \mathrm{Id}$ is bounded above by 1. This condition ensures that the NC-proximal average is well-behaved. For the proof of the following example, we recall the definition of a lower-$\mathcal{C}^2$ function.

Definition 3.2.5. *[28, Definition 10.29](Lower-$\mathcal{C}^2$) A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be* lower-$\mathcal{C}^2$ *on $\mathbb{R}^n$ if on some neighborhood V of each $\bar{x} \in \mathbb{R}^n$ there is a representation*

$$f(x) = \max_{t \in T} f_t(x)$$

in which the functions f_t are of class $\mathcal{C}^2$ on V and the index set T is a compact space such that $f_t(x)$ and $\nabla f_t(x)$ depend continuously and jointly on $(t,x) \in T \times V$.

Proposition 3.2.6. *Any lower-$\mathcal{C}^2$ function is locally Lipschitz continuous.*

Proof:[28, Theorem 10.31 Proof] For each $t \in T$, the function f_t has Lipschitz constant $\operatorname{Lip} f_t(x) = |\nabla f_t(x)|$ by [28, Theorem 9.7]. Hence, $\operatorname{Lip} f(x) \leq \sup_{t\in T} |\nabla f_t(x)|$ by [28, Proposition 9.10]. This supremum is finite, since T is compact and ∇f is continuous in t. Therefore, f is Lipschitz continuous on V. $\square$

Example 3.2.7. *Consider two proper, prox-bounded, prox-regular functions f_0 and f_1. If the Lipschitz constant of $\lambda P_r f_0 + (1-\lambda)P_r f_1 - I$ is bounded above by 1, then for r sufficiently large, the NC-proximal average is para-prox-regular for all $\bar{\lambda} \in (0,1)$.*

Proof: Fix $\bar{\lambda} \in (0,1)$. Let $\bar{v} \in \partial_x PA_r(\bar{x},\bar{\lambda})$. Enlarging r if necessary, we know that $PA_r(x,\lambda)$ is locally Lipschitz continuous in λ [14, Theorem 4.6], and that it is a lower-$\mathcal{C}^2$ function in x [14, Proposition 2.5]. Since all lower-$\mathcal{C}^2$ functions are prox-regular [25], we have that $PA_r(x,\lambda)$ is prox-regular as a function of x at $\bar{x}$ for $\bar{v}$, say with parameters $\epsilon_1 > 0$ and $\rho \geq 0$. Therefore

$$PA_r(x',\bar{\lambda}) \geq PA_r(x,\bar{\lambda}) + \langle v, x'-x\rangle - \frac{\rho}{2}|x'-x|^2 \tag{3.2.1}$$

when $|x'-\bar{x}| < \epsilon_1$, $|x-\bar{x}| < \epsilon_1$, $|PA_r(x,\bar{\lambda}) - PA_r(\bar{x},\bar{\lambda})| < \epsilon_1$, $v \in \partial_x PA_r(x,\bar{\lambda})$, and $|v-\bar{v}| < \epsilon_1$. Notice that if $x' = x$, then inequality (3.2.1) is trivially true, and if $x' \neq x$, then we can replace ρ with $\rho+\delta$ for $\delta > 0$ so that we may change the inequality to a strict one. We will therefore (writing ρ to mean the new $\rho+\delta$ so as not to change notation) concern ourselves with the non-trivial case $x' \neq x$ and use

$$PA_r(x',\bar{\lambda}) > PA_r(x,\bar{\lambda}) + \langle v, x'-x\rangle - \frac{\rho}{2}|x'-x|^2 \tag{3.2.2}$$

instead of inequality (3.2.1). We have pointed out above that it is known that PA_r is continuous in λ, and therefore there exists $\epsilon_2 > 0$ such that inequality (3.2.2) remains true when $\bar{\lambda}$ is replaced by any λ such that $|\lambda-\bar{\lambda}| < \epsilon_2$. We can reduce ϵ_2 if necessary, so that $|\lambda-\bar{\lambda}| < \epsilon_2$ implies $\lambda \in (0,1)$. Taking $\epsilon_3 = \min\{\epsilon_1,\epsilon_2\}$, we have

$$PA_r(x',\lambda) > PA_r(x,\lambda) + \langle v, x'-x\rangle - \frac{\rho}{2}|x'-x|^2 \tag{3.2.3}$$

when $x \neq x'$, $|x'-\bar{x}| < \epsilon_3$, $|x-\bar{x}| < \epsilon_3$, $|PA_r(x,\lambda)-PA_r(\bar{x},\bar{\lambda})| < \epsilon_3$, $v \in \partial_x PA_r(x,\bar{\lambda})$, $|v-\bar{v}| < \epsilon_3$, and $|\lambda-\bar{\lambda}| < \epsilon_3$. If we use $\frac{\epsilon_3}{2}$-neighborhoods instead of ϵ_3-neighborhoods, then inequality (3.2.3) certainly remains true. Since PA_r is a $\mathcal{C}^1$ function with respect to x, we know that $\partial_x PA_r(x,\bar{\lambda}) = \{\nabla_x PA_r(x,\bar{\lambda})\}$, and by [14, Theorem 4.6], we know that $\nabla_x PA_r(x,\bar{\lambda})$ changes in a continuous manner with respect to λ. Hence there exists $\epsilon_4 > 0$ such that inequality (3.2.3) remains valid when $v \in \partial_x PA_r(x,\bar{\lambda})$ is replaced by $\hat{v} \in \partial_x PA_r(x,\lambda)$ for $|\hat{v}-v| < \epsilon_4$ and $|\lambda-\bar{\lambda}| < \epsilon_4$. Shrinking ϵ_4 if necessary, we can ensure that $\epsilon_4 < \frac{\epsilon_3}{2}$. Thus we have $|\hat{v}-v| < \frac{\epsilon_3}{2}$ and $|v-\bar{v}| < \frac{\epsilon_3}{2}$, which implies that $|\hat{v}-\bar{v}| < \frac{\epsilon_3}{2}+\frac{\epsilon_3}{2} = \epsilon_3$. Therefore, choosing $\epsilon = \min\{\epsilon_3,\epsilon_4\}$ we have that

$$PA_r(x',\lambda) > PA_r(x,\lambda) + \langle \hat{v}, x'-x\rangle - \frac{\rho}{2}|x'-x|^2 \tag{3.2.4}$$

when $|x'-\bar{x}| < \epsilon$, $|x-\bar{x}| < \epsilon$, $|PA_r(x,\lambda)-PA_r(\bar{x},\bar{\lambda})| < \epsilon$, $\hat{v} \in \partial_x PA_r(x,\lambda)$, $|\hat{v}-\bar{v}| < \epsilon$, and $|\lambda-\bar{\lambda}| < \epsilon$. □

Chapter 4

Relationship to Monotonicity

In this chapter we give a characterization of para-prox-regular functions in terms of monotonicity. As we shall see, it is possible to express the para-prox-regularity of a function f in terms of monotonicity of $T + rI$, where T is an f-attentive localization of ∂f as defined in Definition 4.1.1. This was done in [25] for prox-regular functions, and the same procedure works here in the para-prox-regular case. We shall present Theorem 4.2.1, which is the para-prox-regular analog of [25, Theorem 3.2], and Corollary 4.2.2, which is the characterization we seek.

4.1 Localization of the Subdifferential

We begin with the definition of an *f-attentive localization* of the subdifferential. An f-attentive localization of $\partial_x f$ around $(\bar{x}, \bar{\lambda}, \bar{v})$ for a fixed λ is a set-valued mapping $T_\lambda : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ whose graph in $\mathbb{R}^n \times \mathbb{R}^n$ is the intersection of $\operatorname{gph} \partial_x f$ with 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.) The following definition is adapted from [25, Definition 3.1].

Definition 4.1.1. *For an $\epsilon > 0$ and a fixed λ, the f-*attentive ϵ-localization of ∂f_x around $(\bar{x}, \bar{\lambda}, \bar{v})$ *is*

$$T_\lambda(x) = \begin{cases} \{v \in \partial_x f(x, \lambda) : |v - \bar{v}| < \epsilon\}, & |x - \bar{x}| < \epsilon, |f(x, \lambda) - f(\bar{x}, \bar{\lambda})| < \epsilon \\ \emptyset, & \textit{otherwise.} \end{cases} \tag{4.1.1}$$

In order to prove Theorem 4.2.1, we need the following well-known definition, propositions and lemma.

Definition 4.1.2. *Let $f = f_1 + \cdots + f_m$ for proper, lsc functions $f_i : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, and let $\bar{x} \in \operatorname{dom} f$. We say that f satisfies the* non-degeneracy condition (NDC) *at* $\bar{x}$ *if the only combination of vectors* $v_i \in \partial^\infty f_i(\bar{x})$ *with* $v_1 + \cdots + v_m = 0$ *is* $v_1 = \cdots = v_m = 0$.

Non-degeneracy combined with prox-regularity will allow us to conclude that the subdifferential of a sum of functions is equal to the sum of subdifferentials.

Proposition 4.1.3. *[28, Theorem 9.13] If* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *is* $\mathcal{C}^1$ *at* $\bar{x}$, *then* $\partial^\infty f(\bar{x}) = \{0\}$.

Proof: We know that $\nabla f(\bar{x})$ is the only subgradient of f at $\bar{x}$, and since f is continuously differentiable, the subgradients nearby $\bar{x}$ are gradients as well. So for a sequence (x_k) that tends to $\bar{x}$, the tail of the corresponding sequence (v_k) as in Definition 2.2.1 c) will be $(\nabla f(x_k))$ which tends to $\nabla f(\bar{x})$. Then for any sequence λ^k that tends to zero, $\lambda^k v^k$ necessarily tends to zero. Therefore, zero is the only horizon subgradient. □

Definition 4.1.2, coupled with Proposition 4.1.3, enables us to prove the following lemma. The lemma takes a para-prox-regular function at $\bar{x}$ for $\bar{v}$ and forms a new function that is para-prox-regular at 0 for 0, in effect shifting the original function. This is an extension of the prox-regular version found in [28, Exercise 13.35], and ir will allow us to make some assumptions in Theorem 4.2.1 without loss of generality in order to simplify the proof.

Lemma 4.1.4. *Let* f *be para-prox-regular at* $(\bar{x}, \bar{\lambda})$ *for* $\bar{v}$. *Define*

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

Then $\tilde{f}$ *is para-prox-regular at* $(0, \bar{\lambda})$ *for* 0.

Proof: By assumption, there exist $\epsilon > 0$ and $r \geq 0$ such that

$$f(x', \lambda) \geq f(x, \lambda) + \langle v, x' - x \rangle - \frac{r}{2}|x - \bar{x}|^2 \tag{4.1.2}$$

whenever $x' \neq x$, $|x' - \bar{x}| < \epsilon$, $|x - \bar{x}| < \epsilon$, $|f(x, \lambda) - f(\bar{x}, \bar{\lambda})| < \epsilon$, $v \in \partial_x f(x, \lambda)$, $|v - \bar{v}| < \epsilon$, and$|\lambda - \bar{\lambda}| < \epsilon$. For ease of discussion let $f_1(x, \lambda) = f(x + \bar{x}, \lambda)$ and $f_2(x, \lambda) = -\langle \bar{v}, x - \bar{x} \rangle$. Notice that $v \in \partial_x f(x, \lambda)$ implies $v - \bar{v} \in \partial_x f(x, \lambda) - \bar{v}$. Since f_2 is a continuously differentiable function, $\partial_x^\infty f_2(x, \lambda) = \{0\}$ for all (x, λ). Hence, the only solution to $v_1 + v_2 = 0$ for $v_i \in \partial_x^\infty f_i(\bar{x}, \lambda)$ is $v_1 = v_2 = 0$, and NDC holds. Since prox-regularity implies local regularity, Proposition **??** gives us that $\partial_x f(x, \lambda) - \bar{v} = \partial_x [f(x + \bar{x}, \lambda) - \langle \bar{v}, x - \bar{x} \rangle] = \partial_x \tilde{f}(x, \lambda)$. Therefore $v - \bar{v} \in \partial_x \tilde{f}(x, \lambda)$. Translating inequality (4.1.2) by $\bar{x}$, we may rewrite it as

$$f(x' + \bar{x}, \lambda) \geq f(x + \bar{x}, \lambda) + \langle v, (x' - \bar{x}) - (x - \bar{x}) \rangle - \frac{r}{2}|(x' - \bar{x}) - (x - \bar{x})|^2 \tag{4.1.3}$$

now with the conditions $x' \neq x$, $|x'| < \epsilon$, $|x| < \epsilon$, $|\tilde{f}(x,\lambda) - \tilde{f}(0,\bar{\lambda})| < \epsilon$, $v - \bar{v} \in \partial_x \tilde{f}(x,\lambda)$, $|v - \bar{v}| < \epsilon$, and $|\lambda - \bar{\lambda}| < \epsilon$. The $\bar{x}$ term is cancelled in two places, leaving

$$f(x' + \bar{x}, \lambda) \geq f(x + \bar{x}, \lambda) + \langle v, x' - x\rangle - \frac{r}{2}|x' - x|^2. \tag{4.1.4}$$

Subtracting $\langle \bar{v}, x' - \bar{x}\rangle$ and $\langle \bar{v}, x - \bar{x}\rangle$ from both sides of inequality (4.1.4), we get

$$\begin{aligned}&f(x' + \bar{x}, \lambda) - \langle \bar{v}, x' - \bar{x}\rangle - \langle \bar{v}, x - \bar{x}\rangle \geq \\ &f(x + \bar{x}, \lambda) - \langle \bar{v}, x - \bar{x}\rangle - \langle \bar{v}, x' - \bar{x}\rangle + \langle v, x' - x\rangle - \frac{r}{2}|x' - x|^2.\end{aligned}$$

Now the first two terms on each side of the above inequality are how we defined $\tilde{f}$, so we can rewrite

$$\begin{aligned}\tilde{f}(x', \lambda) - \langle \bar{v}, x - \bar{x}\rangle &\geq \tilde{f}(x, \lambda) - \langle \bar{v}, x' - \bar{x}\rangle + \langle v, x' - x\rangle - \frac{r}{2}|x' - x|^2 \\ &\Rightarrow \tilde{f}(x', \lambda) \geq \tilde{f}(x, \lambda) + \langle v, x' - x\rangle - \langle \bar{v}, x' - \bar{x} - x + \bar{x}\rangle - \frac{r}{2}|x' - x|^2 \\ &\Rightarrow \tilde{f}(x', \lambda) \geq \tilde{f}(x, \lambda) + \langle v, x' - x\rangle - \langle \bar{v}, x' - x\rangle - \frac{r}{2}|x' - x|^2 \\ &\Rightarrow \tilde{f}(x', \lambda) \geq \tilde{f}(x, \lambda) + \langle v - \bar{v}, x' - x\rangle - \frac{r}{2}|x' - x|^2.\end{aligned}$$

Therefore $\tilde{f}$ is para-prox-regular at $(0, \bar{\lambda})$ for 0. □

4.2 Characterization

Now we are ready to present the main result of this chapter: the relationship between para-prox-regularity and monotonicity.

Theorem 4.2.1. *Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be a proper function that is locally lsc near $(\bar{x}, \bar{\lambda})$. Then the following are equivalent.*

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

b) The vector $\bar{v}$ is a proximal subgradient of f with respect to x at $(\bar{x}, \bar{\lambda})$, and there exist $\epsilon > 0$ and $r \geq 0$ such that the f-attentive ϵ-localization T_λ of $\partial_x f$ at $(\bar{x}, \bar{\lambda}, \bar{v})$ has $T_\lambda + rI_x$ monotone for all λ such that $|\lambda - \bar{\lambda}| < \epsilon$. That is, there is a constant r such that

$$\langle v_1 - v_0, x_1 - x_0\rangle \geq -r|x_1 - x_0|^2 \tag{4.2.1}$$

when $v_i \in T_\lambda(x_i)$, $i \in \{0, 1\}$, and $|\lambda - \bar{\lambda}| < \epsilon$. Here I_x is the identity mapping with respect to x.

c) *There exist* $\epsilon > 0, r \geq 0$ *such that*

$$f(x, \bar{\lambda}) \geq f(\bar{x}, \bar{\lambda}) + \langle \bar{v}, x - \bar{x} \rangle - \frac{r}{2}|x - \bar{x}|^2 \qquad (4.2.2)$$

when $|x - \bar{x}| < \epsilon$, *and the mapping* $(\partial_x f + rI_x)^{-1}$ *has the following single-valuedness property near* $\bar{z} = \bar{v} + r\bar{x}$: *if* $|z - \bar{z}| < \epsilon$, *then for* $i \in \{0, 1\}$ *one has*

$$(x_i, \lambda) \in (\partial_x f + rI_x)^{-1}(z) \Rightarrow x_0 = x_1 \qquad (4.2.3)$$

when $|x_i - \bar{x}| < \epsilon$, $|f(x_i, \bar{\lambda}) - f(\bar{x}, \bar{\lambda})| < \epsilon$, *and* $|\lambda - \bar{\lambda}| < \epsilon$.

Proof:

a)⇒b): We have that f is para-prox-regular at $(\bar{x}, \bar{\lambda})$ for $\bar{v}$, so there exist $\epsilon > 0$ and $r \geq 0$ such that

$$\begin{array}{l} f(x_1, \lambda) \geq f(x_0, \lambda) + \langle v_0, x_1 - x_0 \rangle - \frac{r}{2}|x_1 - x_0|^2 \\ f(x_0, \lambda) \geq f(x_1, \lambda) + \langle v_1, x_0 - x_1 \rangle - \frac{r}{2}|x_0 - x_1|^2 \end{array} \qquad (4.2.4)$$

whenever $i \in \{0, 1\}$, $|x_i - \bar{x}| < \epsilon$, $|f(x_i, \lambda) - f(\bar{x}, \bar{\lambda})| < \epsilon$, $v_i \in \partial_x f(x_i, \lambda)$, $|v_i - \bar{v}| < \epsilon$, and $|\lambda - \bar{\lambda}| < \epsilon$. Adding the two inequalities in (4.2.4) together, we get

$$f(x_1, \lambda) + f(x_0, \lambda) \geq f(x_0, \lambda) + f(x_1, \lambda) + \langle v_0 - v_1, x_1 - x_0 \rangle - r|x_1 - x_0|^2,$$

which simplifies to inequality (4.2.1). Finally, since f is para-prox-regular at $(\bar{x}, \bar{\lambda})$ for $\bar{v}$, it is prox-regular in x at $\bar{x}$ as well, giving us that $\bar{v}$ is a proximal subgradient of f with respect to x at $(\bar{x}, \bar{\lambda})$.

b)⇒c): For this section of the proof and the next, without loss of generality we assume that $\bar{x} = 0$, $f(0, \bar{\lambda}) = 0$, and $\bar{v} = 0$ (by Lemma 4.1.4). Let $\bar{\epsilon} > 0$ and $\bar{r} \geq 0$ be the parameters assumed by condition b). Since $\bar{v}$ is a proximal subgradient of f with respect to x at $(\bar{x}, \bar{\lambda})$, we have prox-regularity of f with respect to x at $(\bar{x}, \bar{\lambda})$ and we can use Definition 2.3.1 to write

$$f(x, \bar{\lambda}) \geq f(\bar{x}, \bar{\lambda}) + \langle \bar{v}, x - \bar{x} \rangle - \frac{r}{2}|x - \bar{x}|^2 \qquad (4.2.5)$$

when $|x - \bar{x}| < \epsilon$ for some $\epsilon \in (0, \bar{\epsilon})$ and some $r \in (\bar{r}, +\infty)$.

To see the second statement in c), let $|\lambda - \bar{\lambda}| < \epsilon$. By decreasing ϵ and increasing r if necessary, we can arrange that $\epsilon < \frac{\bar{\epsilon}}{2r}$ with $r \geq 1$, hence $\epsilon < \frac{\bar{\epsilon}}{2}$. Suppose that for z such that $|z - (\bar{v} + r\bar{x})| < \epsilon$ we have $(x_i, \lambda) \in (\partial_x f + rI_x)^{-1}(z)$, $i \in \{0, 1\}$. By our assumptions that $\bar{x} = 0$,

$f(0,\bar\lambda)=0$, and $\bar v=0$, this simplifies to $|z|<\epsilon$. For $v_i=z-rx_i$ we have $v_i\in\partial_x f(x_i,\lambda)$, and by the triangle inequality $|v_i|\le|z|+r|x_i|<\frac{\bar\epsilon}{2}+\frac{\bar\epsilon}{2}=\bar\epsilon$. Also $|x_i|<\bar\epsilon$ and $|f(x_i,\lambda)|<\bar\epsilon$, thus, $v_i\in T_\lambda(x_i)$. So we have that

$$-\bar r|x_1-x_0|^2\le\langle v_1-v_0,x_1-x_0\rangle=\langle(z-rx_1)-(z-rx_0),x_1-x_0\rangle=-r|x_1-x_0|^2.$$

Since $r>\bar r$, we conclude that $x_1=x_0$.

c)⇒a): Remember that without loss of generality we assume here that $\bar x=0$, $f(0,\bar\lambda)=0$, and $\bar v=0$ (by Lemma 4.1.4). Let $\epsilon>0$ and $r\ge 0$ be the parameters assumed by condition c). It will suffice to show that there exists $\bar\epsilon\in(0,\epsilon)$ such that para-prox-regularity of f is satisfied by $\bar\epsilon$ and r at $\bar x=0$, $f(0,\bar\lambda)=0$ for $\bar v=0$. That is, we seek to show that whenever

$$v_0\in\partial_x f(x_0,\lambda),|v_0|<\bar\epsilon,|f(x_0,\lambda)|<\bar\epsilon, \tag{4.2.6}$$

and $|x|<\bar\epsilon,\ |\lambda-\bar\lambda|<\bar\epsilon$, we have

$$f(x,\lambda)\ge f(x_0,\lambda)+\langle v_0,x-x_0\rangle-\frac{r}{2}|x-x_0|^2. \tag{4.2.7}$$

In place of equation (4.2.7), we can aim at guaranteeing the stronger condition

$$f(x,\lambda)>f(x_0,\lambda)+\langle v_0,x-x_0\rangle-\frac{r}{2}|x-x_0|^2 \tag{4.2.8}$$

for $x\ne x_0$, $|x|\le\epsilon$, and $|\lambda-\bar\lambda|\le\epsilon$. Notice that

$$-\frac{r}{2}|x-x_0|^2=\langle rx_0,x-x_0\rangle-\frac{r}{2}|x|^2+\frac{r}{2}|x_0|^2,$$

so we can rewrite inequality (4.2.8) as

$$\begin{aligned}f(x,\lambda)&>f(x_0,\lambda)+\langle v_0,x-x_0\rangle+\langle rx_0,x-x_0\rangle-\frac{r}{2}|x|^2+\frac{r}{2}|x_0|^2\\&=f(x_0,\lambda)+\langle v_0+rx_0,x\rangle-\langle v_0+rx_0,x_0\rangle-\frac{r}{2}|x|^2+\frac{r}{2}|x_0|^2.\end{aligned}$$

Rearranging, we get

$$f(x,\lambda)-\langle v_0+rx_0,x\rangle+\frac{r}{2}|x|^2>f(x_0,\lambda)-\langle v_0+rx_0,x_0\rangle+\frac{r}{2}|x_0|^2.$$

Thus, we seek to show that for every fixed λ such that $|\lambda-\bar\lambda|\le\bar\epsilon$,

$$\operatorname*{argmin}_{|x|\le\bar\epsilon}\{f(x,\lambda)+\frac{r}{2}|x|^2-\langle z_0,x\rangle\}=\{(x_0,\lambda)\} \tag{4.2.9}$$

where $z_0 = v_0 + rx_0$. In summary, we need only demonstrate that conditions (4.2.6) imply equation (4.2.9) when $\bar{\epsilon}$ is small enough. Our assumption that f is locally lsc at $(0, \bar{\lambda})$ with $f(0, \bar{\lambda}) = 0$ ensures that f is lsc within a compact set of the form

$$C = \{(x, \lambda) : |x| \le \hat{\epsilon}, |\lambda - \bar{\lambda}| \le \hat{\epsilon}, f(x, \lambda) \le 2\hat{\epsilon}\}$$

for some $\hat{\epsilon} > 0$. In particular, f must be lsc at $(0, \bar{\lambda})$ itself, and since $f(0, \bar{\lambda}) = 0$, we have

$$\liminf_{x, |\lambda - \bar{\lambda}| \to 0} f(x, \lambda) = 0.$$

Shrinking ϵ if necessary, we can arrange that

$$|x| \le \epsilon, |\lambda - \bar{\lambda}| \le \epsilon \Rightarrow |x| \le \hat{\epsilon}, |\lambda - \bar{\lambda}| \le \hat{\epsilon}, \text{ and } f(x, \lambda) > -2\hat{\epsilon}. \tag{4.2.10}$$

Enlarging r if necessary, we can arrange that

$$|x|^2 < \frac{6\hat{\epsilon}}{r} \Rightarrow |x| < \epsilon, |x| < \frac{\epsilon^2}{2\hat{\epsilon}}, \text{ and } f(x, \lambda) > -\frac{\epsilon}{2}. \tag{4.2.11}$$

Define

$$g(z) := \inf_{|x|, |\lambda - \bar{\lambda}| \le \epsilon} \{f(x, \lambda) + \frac{r}{2}|x|^2 - \langle z, x \rangle\},$$

$$G(z) := \operatorname*{argmin}_{|x|, |\lambda - \bar{\lambda}| \le \epsilon} \{f(x, \lambda) + \frac{r}{2}|x|^2 - \langle z, x \rangle\}.$$

Since g is the pointwise infimum of a collection of affine functions of z, it is concave. Recall: $\bar{x} = 0$, $\bar{v} = 0$, and $f(\bar{x}, \bar{\lambda}) = 0$. These assumptions tell us that $g(0) = 0$ and $G(0) = \{0\}$, whereas $g(z) \le 0$ in general. We see this by noting that $f(0, \bar{\lambda}) = 0$ and hence $f(0, \bar{\lambda}) + \frac{r}{2}|0|^2 - \langle z, 0 \rangle = 0$, and since $g(z)$ is the infimum of all such expressions with $|x| < \epsilon$, then 0 is an upper bound for g. We claim that under these circumstances,

$$|z| < \frac{\hat{\epsilon}}{\epsilon} \Rightarrow \begin{cases} G(z) \ne \emptyset \text{ and} \\ |f(x, \lambda)| < 2\hat{\epsilon} \quad \text{ for all } (x, \lambda) \in G(z). \end{cases} \tag{4.2.12}$$

To see this, observe that because $g \le 0$, the minimization in the definition of $g(z)$ is unaffected if the attention is restricted to the points (x, λ) satisfying not only $|x| < \epsilon$ and $|\lambda - \bar{\lambda}| < \epsilon$, but also $f(x, \lambda) + \frac{r}{2}|x|^2 - \langle z, x \rangle \le \hat{\epsilon}$, in which case $f(x, \lambda) \le \hat{\epsilon} + \epsilon|z|$. Therefore, as long as $|z| < \frac{\hat{\epsilon}}{\epsilon}$, attention can be restricted to points (x,λ) satisfying $f(x,\lambda) < 2\hat{\epsilon}$. Recalling conditions (4.2.10) and the choice of the set C, we deduce that when $|z| \le \frac{\hat{\epsilon}}{\epsilon}$,

$$\begin{aligned} g(z) &= \inf_{(x,\lambda)\in C}\{f(x,\lambda) + \tfrac{r}{2}|x|^2 - \langle z, x\rangle\}, \\ G(z) &= \operatorname*{argmin}_{(x,\lambda)\in C}\{f(x,\lambda) + \tfrac{r}{2}|x|^2 - \langle z, x\rangle\}. \end{aligned} \tag{4.2.13}$$

Since f is lsc relative to C, which is compact, the infimum is sure to be attained. Thus, implication (4.2.12) is correct. Also we observe that g is finite on a neighborhood of $z = 0$, and hence continuous (by concavity) [28, Theorem 2.35]. Choose $\tilde{\epsilon} \in (0,\epsilon)$ small enough that $\tilde{\epsilon}(1+r) < \frac{\hat{\epsilon}}{\epsilon}$ and $g(z) > -\frac{\epsilon}{2}$ when $|z| < \tilde{\epsilon}(1+r)$. Under this choice we are ready to consider elements satisfying conditions (4.2.6) and show that we get equation (4.2.9). The vector $z_0 = v_0 + rx_0$ has $|z_0| \le |v_0| + r|x_0| < \tilde{\epsilon}(1+r)$, hence $|z_0| < \frac{\hat{\epsilon}}{\epsilon}$ and $g(z_0) > -\frac{\epsilon}{2}$. In particular, $G(z_0)$ must be nonempty. Consider any $(x_1,\lambda) \in G(z_0)$. We have to show on the basis of the single-valuedness property that $x_1 = x_0$. We have that $(x_0,\lambda) \in (\partial_x f + rI)^{-1}(z_0)$, due to the fact that $v_0 \in \partial_x f(x_0,\lambda)$ and hence $v_0 + rx_0 \in (\partial_x f + rI_x)(x_0,\lambda)$. If we can establish that

$$|x_1| < \epsilon, |\lambda| < \epsilon, |f(x_1,\lambda)| < \epsilon \text{ and } (x_1,\lambda) \in (\partial_f + rI)^{-1}(z_0), \tag{4.2.14}$$

then the single-valuedness property can be invoked and we will get $x_1 = x_0$ as required.

Because $|z_0| < \frac{\hat{\epsilon}}{\epsilon}$, we know from implication (4.2.12) that $|f(x_1,\lambda)| < 2\hat{\epsilon}$. Then from $\frac{r}{2}|x_1|^2 = \langle z_0, x_1\rangle - f(x_1,\lambda) + g(z_0)$, where $g(z_0) \le 0$, we know that $\frac{r}{2}|x_1|^2 \le |z_0||x_1| + |f(x_1,\lambda)| < \frac{\hat{\epsilon}}{\epsilon}\epsilon + 2\hat{\epsilon} = 3\hat{\epsilon}$, so $|x_1|^2 < \frac{6\hat{\epsilon}}{r}$. Through (4.2.11) this implies that $|x_1| < \epsilon$ and also that $|x_1| < \frac{\epsilon^2}{2\hat{\epsilon}}$ and $f(x_1,\lambda) > -\epsilon$. Then from $f(x_1,\lambda) = \langle z_0, x_1\rangle - \frac{r}{2}|x_1|^2 + g(z_0) \le |z_0||x_1| + |g(z_0)|$ and the fact that $g(z_0) < \frac{\epsilon}{2}$ we obtain $f(x_1) < \frac{\hat{\epsilon}}{\epsilon}\frac{\epsilon^2}{2\hat{\epsilon}} + \frac{\epsilon}{2} = \epsilon$. Hence $|x_1| < \epsilon$, $|\lambda - \bar{\lambda}| < \epsilon$, and $|f(x_1,\lambda)| < \epsilon$. The fact that the minimum for $g(z_0)$ is attained at x_1 gives us

$$f(x,\lambda) + \frac{r}{2}|x|^2 - \langle z_0, x\rangle \ge f(x_1,\lambda) + \frac{r}{2}|x_1|^2 - \langle z_0, x_1\rangle \tag{4.2.15}$$

when $|x| < \epsilon$ and $|\lambda - \bar{\lambda}| < \epsilon$. Hence, for all $|x| < \epsilon$ and $|\lambda - \bar{\lambda}| < \epsilon$,

$$\begin{aligned} f(x,\lambda) &\ge f(x_1,\lambda) + \langle z_0, x\rangle - \langle z_0, x_1\rangle - \frac{r}{2}|x|^2 + \frac{r}{2}|x_1|^2 \\ &= f(x_1,\lambda) + \langle z_0, x - x_1\rangle + r\langle x_1, x_1\rangle - \frac{r}{2}\langle x, x\rangle - \frac{r}{2}\langle x_1, x_1\rangle \\ &= f(x_1,\lambda) + \langle z_0 - rx_1, x - x_1\rangle - \frac{r}{2}|x - x_1|^2. \end{aligned}$$

Since $|x_1| < \epsilon$, this implies that $z_0 - rx_1 \in \partial_x f(x_1, \lambda)$, so $z_0 \in (\partial_x f + rI_x)(x_1, \lambda)$. The subgradient relation in (4.2.14) is therefore correct as well, and our single-valuedness assumption implies $x_0 = x_1$. Thus equation (4.2.9) holds and we are done. □

This result allows us to obtain the following characterization of para-prox-regularity. A similar result was shown in [25, Corollary 3.4]. Corollary 4.2.2 below is stronger than [25, Corollary 3.4] as the function defined here is para-prox-regular as opposed to prox-regular.

Corollary 4.2.2. *Let f be a proper, lsc function. Then f is para-prox-regular in x at $(\bar{x}, \bar{\lambda})$ for $\bar{v}$, with parameters $\epsilon > 0, r > 0$, if and only if $\bar{v}$ is a proximal subgradient of f with respect to x at $(\bar{x}, \bar{\lambda})$ and*

$$\langle v_1 - v_0, x_1 - x_0 \rangle \geq -r|x_1 - x_0|^2$$

when $i \in \{0, 1\}$, $|x_i - \bar{x}| < \epsilon$, $|f(x_i, \lambda) - f(\bar{x}, \bar{\lambda})| < \epsilon$, $v_i \in \partial_x f(x_i, \lambda)$, $|v_i - \bar{v}| < \epsilon$, and $|\lambda - \bar{\lambda}| < \epsilon$.

This result is an extension of [18, Proposition 3.1], which gives the proof of a) $\Rightarrow$ c) but with a stronger condition and stronger result. He assumes the NDC in his conditions, and arrives at the conclusion that $(\partial_x f + rI_x)^{-1}$ is not only single-valued, but also monotone and Lipschitz continuous.

Chapter 5

Constructing Para-prox-regular Functions

In this chapter we examine some common situations where parameters arise in optimization problems, and we explore when para-prox-regularity is present. In all cases we focus on identifying the prox-regularity parameters. In general we begin with prox-regular functions and build upon them. We already know that many important classes of functions are prox-regular, such as proper, lsc, convex functions, lower-$\mathcal{C}^2$ functions ($\mathcal{C}^2$ functions included), strongly amenable functions, and primal-lower-nice functions. Our goal here is to begin to analyze these functions to see what conditions are necessary for them to be para-prox-regular as well, or how we can combine or alter them to get para-prox-regular functions. In particular, we show that the scalar multiplication of a prox-regular function by a positive parameter results in a function that remains prox-regular. This leads to the construction of para-prox-regular functions via the weighted sum and weighted max of prox-regular functions.

5.1 Naming Parameters Explicitly

The following proposition and theorem are results first provided in [27], showing that strongly amenable functions, that is functions which can be locally represented as the composition of a $\mathcal{C}^2$ function with a proper, lsc, convex function, are prox-regular. In that work, Poliquin and Rockafellar simply conclude the function is prox-regular and they do not single out the prox-regularity parameters ϵ and r in the statement of the theorem. In the present work, the theorem has been restated with a proposition beforehand, in order to pull out the prox-regularity parameters from the proof into the statement of the theorem. In later results we require the ability to identify such parameters instead of merely stating that they exist. Other than clarifying the prox-regularity parameters, Proposition 5.1.1 and Theorem 5.1.2 and their proofs are essentially the same as [27, Theorem 3.1].

Proposition 5.1.1. *[27, Theorem 3.1] Let $f(x) = g(F(x))$, where $F : \mathbb{R}^n \to \mathbb{R}^m$ is of class $\mathcal{C}^2, g : \mathbb{R}^m \to \bar{\mathbb{R}}$ is lsc and proper, and suppose there is no vector $y \neq 0$ in $\partial^\infty g(F(\bar{x}))$ with $\nabla F(\bar{x})^* y = 0$, where $\bar{x} \in \operatorname{dom} f$. Then*

a) there exists $\epsilon_0 > 0$ such that

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

when $|x - \bar{x}| < \epsilon_0$ and $|f(x) - f(\bar{x})| < \epsilon_0$.

b) $S : (x, v) \to \{y \in \partial g(F(x)) : \nabla F(x)^* y = v\}$ has closed graph and is locally bounded at $(\bar{x}, \bar{v})$ for $\bar{v} \in \partial f(\bar{x})$, within the same neighborhoods $B_{\epsilon_0}(\bar{x})$ and $B_{\epsilon_0}(f(\bar{x}))$. In particular, $S(\bar{x}, \bar{v})$ is a compact set.

c) If $\bar{v} \in \partial f(\bar{x})$ is such that for every $y \in \partial g(F(\bar{x}))$ with $\nabla F(\bar{x})^* y = \bar{v}$ the function g is prox-regular at $F(\bar{x})$ with parameters $\epsilon_y > 0, r_y \geq 0$, then for the parameters $\epsilon = \min\{\epsilon_0, \min_y\{\epsilon_y\}\}$ and some $\bar{r} \geq 0$ we have

$$\langle y_1 - y_0, u_1 - u_0 \rangle \geq -\bar{r}|u_1 - u_0|^2 \tag{5.1.1}$$

whenever $y_i \in \partial g(u_i)$, $|u_i - F(\bar{x})| < \bar{\epsilon}$, $|g(u_i) - g(F(\bar{x}))| < \bar{\epsilon}$, and $\operatorname{dist}(y_i, S(\bar{x}, \bar{v})) < \bar{\epsilon}$.

Proof: See [27, Theorem 3.1] and the first half of its proof. □

The purpose of pulling out Proposition 5.1.1 from [27, Theorem 3.1] is that now we have an explicit expression for the ϵ that we will use in Theorem 5.1.2. In Theorem 5.1.2 we must provide greater detail, in order to fully describe the parameter r.

Theorem 5.1.2. *[27, Theorem 3.1] Let the conditions of Proposition 5.1.1 hold. Assume further that $\bar{v} \in \partial f(\bar{x})$ is a vector such that for every $y \in \partial g(F(\bar{x}))$ such that $\nabla F(\bar{x})^* y = \bar{v}$, the function g is prox-regular at $F(\bar{x})$ for y with parameters $\epsilon_y > 0, r_y \geq 0$. Then f is prox-regular at $\bar{x}$ for $\bar{v}$ with parameters $\epsilon > 0$ as in Proposition 5.1.1 c), and $r \geq 0$ as described by equations (5.1.2), (5.1.4), (5.1.6), and (5.1.8) below.*

Proof: By our assumptions, g is prox-regular at $F(\bar{x})$ for each $y \in S(\bar{x}, \bar{v})$, with constants ϵ_y and r_y. Since $S(\bar{x}, \bar{v})$ is compact, we know that there is a finite open cover of S using a finite number of the ϵ_y, say $\{\epsilon_1, \ldots, \epsilon_k\}$. Then using the corresponding $\{r_i, \ldots, r_k\}$ we can define

$$\bar{r} := \max_{i \in \{1, \ldots, k\}} \{r_i\} \geq 0. \tag{5.1.2}$$

We then have $\epsilon > 0$, $\bar{r} \geq 0$, S a compact set, and $y_i \in \partial g(F(x_i)) \cap S$ with $\nabla F(x_i)^* y_i = v_i$ such that

$$\langle y_1 - y_0, F(x_1) - F(x_0)\rangle \geq -\bar{r}|F(x_1) - F(x_0)|^2$$

when $|x_i - \bar{x}| < \epsilon$, $|f(x_i) - f(\bar{x})| < \epsilon$, $v_i \in \partial f(x_i)$, and $|v_i - \bar{v}| < \epsilon$. Note that when $v_i = \nabla F(x_i)^* y_i$ then

$$\begin{aligned}&\langle v_1 - v_0, x_1 - x_0\rangle = \\ &\langle \nabla F(x_1)^* y_1 - \nabla F(x_0)^* y_1 + \nabla F(x_0)^* y_1 - \nabla F(x_0)^* y_0, x_1 - x_0\rangle,\end{aligned}$$

which gives

$$\begin{aligned}&\langle v_1 - v_0, x_1 - x_0\rangle = \\ &\langle [\nabla F(x_1)^* - \nabla F(x_0)^*] y_1, x_1 - x_0\rangle + \langle \nabla F(x_0)^*(y_1 - y_0), x_1 - x_0\rangle.\end{aligned} \tag{5.1.3}$$

Because S is compact, we know that there exists L such that $|y| \leq L$ for all $y \in S$. Because $F \in \mathcal{C}^2$, we know that for $|x_0 - \bar{x}| < \epsilon$ and $|x_1 - \bar{x}| < \epsilon$ there exists K such that $|\nabla F(x_1) - \nabla F(x_0)| \leq K|x_1 - x_0|$. Hence, we can find

$$r_1 > 0 \text{ such that for all } y \in S, \quad |[\nabla F(x_1)^* - \nabla F(x_0)^*]y| \leq r_1|x_1 - x_0|. \tag{5.1.4}$$

Thus by the Cauchy-Schwarz Inequality, we have that for all $y \in S$,

$$\langle [\nabla F(x_1)^* - \nabla F(x_0)^*]y, x_1 - x_0\rangle \geq -r_1|x_1 - x_0|^2. \tag{5.1.5}$$

The final term in equation (5.1.3) can be written as $\langle \nabla\phi(x_0), x_1 - x_0\rangle$ for the $\mathcal{C}^2$ function $\phi(x) = \langle y_1 - y_0, F(x)\rangle$. Let

$$\begin{aligned}&r_2 \text{ be an upper bound for the eigenvalues of the Hessian of} \\ &x \to \langle \eta, F(x)\rangle,\end{aligned} \tag{5.1.6}$$

where x ranges over $B_\epsilon(\bar{x})$ and η ranges over the compact set $S - S$. Then in particular, by the Taylor expansion of ϕ about x_0, we have

$$\phi(x) \leq \phi(x_0) + \langle \nabla\phi(x_0), x - x_0\rangle + r_2|x - x_0|^2$$

when $|x - \bar{x}| \leq \epsilon$. It follows that

$$\begin{aligned}\langle \nabla\phi(x_0), x_1 - x_0\rangle &\geq -r_2|x_1 - x_0|^2 + \phi(x_1) - \phi(x_0) \\ &= -r_2|x_1 - x_0|^2 + \langle y_1 - y_0, F(x_1) - F(x_0)\rangle \\ &\geq -r_2|x_1 - x_0|^2 - \bar{r}|F(x_1) - F(x_0)|^2.\end{aligned} \tag{5.1.7}$$

But since $F \in \mathcal{C}^2$, there is also a constant k such that $|F(x_1) - F(x_0)| \leq k|x_1 - x_0|$ when $|x_i - \bar{x}| < \epsilon$. Using this fact with Inequalities (5.1.5) and (5.1.7) for the two terms at the end of equation (5.1.3) we obtain

$$\langle v_1 - v_0, x_1 - x_0 \rangle \geq -r_1|x_1 - x_0|^2 - r_2|x_1 - x_0|^2 - \bar{r}k^2|x_1 - x_0|^2.$$

Thus for

$$r := r_1 + r_2 + \bar{r}k^2 \tag{5.1.8}$$

we have that

$$\langle v_1 - v_0, x_1 - x_0 \rangle \geq -r|x_1 - x_0|^2$$

whenever $|x_i - \bar{x}| < \epsilon$, $|f(x_i) - f(\bar{x})| < \epsilon$, $v_i \in \partial f(x_i)$, and $|v_i - \bar{v}| < \epsilon$. □

5.2 Scalar Multiplication and Finite Sum

Now we are ready to begin with parametrized functions. Lemma 5.2.3 demonstrates that multiplying a prox-regular function by a positive parameter does not affect its prox-regularity. We shall use this property in the theorems that follow. First we require a pair of basic lemmas on computing subgradients after scaling a function. These are well-known results, but we provide short proofs for completeness.

Lemma 5.2.1. *[28, Proposition 10.19] Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\lambda > 0$. Then $\lambda \partial f = \partial \lambda f$.*

Proof: Let $\bar{x} \in \operatorname{dom} f$, $\bar{v} \in \partial f(\bar{x})$. Then there exists a sequence $\{x^k\} \subseteq \operatorname{dom} f$ that tends to $\bar{x}$, and a sequence $\{v^k\}$ such that $v^k \in \hat{\partial} f(x^k)$ and $v^k \to \bar{v}$. So for each (x^k, v^k) we have

$$f(x) \geq f(x^k) + \langle v^k, x - x^k \rangle + o(|x - x^k|). \tag{5.2.1}$$

Multiplying inequality (5.2.1) by λ yields

$$\lambda f(x) \geq \lambda f(x^k) + \langle \lambda v^k, x - x^k \rangle + o(|x - x^k|) \tag{5.2.2}$$

for each (x^k, v^k). Note that λ does not affect the little-o term. This gives us that each λv^k is a regular subgradient of $\lambda f(x^k)$. So we have a sequence $\lambda v^k \to \lambda \bar{v}$ such that $\lambda v^k \in \hat{\partial} \lambda f(x^k)$ and $x^k \to \bar{x}$, which says that $\lambda \bar{v} \in \partial \lambda f(\bar{x})$. Hence, $\lambda \partial f \subseteq \partial \lambda f$. Now by starting with $\lambda \bar{v} \in \partial \lambda f(\bar{x})$ and following the same procedure as above in reverse order, we see that $\partial \lambda f \subseteq \lambda \partial f$. Therefore, $\lambda \partial f = \partial \lambda f$. □

Lemma 5.2.2. *[28, Proposition 10.19] Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\lambda > 0$. Then $\partial^\infty \lambda f = \partial^\infty f$.*

Proof: Let $\bar{x} \in \operatorname{dom} f$, $\bar{v} \in \partial^\infty f(\bar{x})$. Then there exist sequences $\{x^k\}$, $\{\gamma^k\}$ and $\{v^k\}$ such that $x^k \to \bar{x}$, $\gamma^k \searrow 0$, $v^k \in \hat{\partial} f(x^k)$, and $\gamma^k v^k \to \bar{v}$. Since $\lambda > 0$, then the sequence $\frac{1}{\lambda}\gamma^k \searrow 0$, and we know from Lemma 5.2.1 that $\lambda v^k \in \hat{\partial} \lambda f(x^k)$. So we see that $\frac{1}{\lambda}\gamma^k \lambda v^k = \gamma^k v^k \to \bar{v}$, and we have the conditions needed to conclude that $\bar{v} \in \partial^\infty \lambda f(\bar{x})$. Hence, $\partial^\infty f(\bar{x}) \subseteq \partial^\infty \lambda f(\bar{x})$. By following the same procedure in reverse order, we see that $\partial^\infty \lambda f(\bar{x}) \subseteq \partial^\infty f(\bar{x})$, and we arrive at the conclusion that the two horizon subdifferentials are equal. □

Lemmas 5.2.1 and 5.2.2 will allow us to present our first minor result of this section, Lemma 5.2.3 below.

Lemma 5.2.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be prox-regular at $\bar{x}$ for $\bar{v}$ with parameters $\tilde{\epsilon} > 0$ and $\tilde{r} \geq 0$, and let $\lambda \in \mathbb{R}, \lambda > 0$. Then $\lambda f(x)$ is prox-regular at $\bar{x}$ for $\lambda \bar{v}$ with parameters $\epsilon = \min\{\tilde{\epsilon}, \lambda\tilde{\epsilon}\}$ and $r = \lambda \tilde{r}$.*

Proof: By prox-regularity of f, $\tilde{\epsilon}$ and $\tilde{r}$ are such that

$$f(x') \geq f(x) + \langle v, x' - x\rangle - \frac{\tilde{r}}{2}|x' - x|^2$$

whenever $|x' - \bar{x}| < \tilde{\epsilon}$, $|x - \bar{x}| < \tilde{\epsilon}$, $|f(x) - f(\bar{x})| < \tilde{\epsilon}$, $v \in \partial f(x)$, and $|v - \bar{v}| < \tilde{\epsilon}$. Multiplying by $\lambda > 0$ yields

$$\lambda f(x') \geq \lambda f(x) + \langle \lambda v, x' - x\rangle - \frac{\lambda\tilde{r}}{2}|x' - x|^2$$

whenever $|x' - \bar{x}| < \tilde{\epsilon}$, $|x - \bar{x}| < \tilde{\epsilon}$, $|f(x) - f(\bar{x})| < \tilde{\epsilon}$, $v \in \partial f(x)$, and $|v - \bar{v}| < \tilde{\epsilon}$. Since $v \in \partial f(x)$, we know by Lemma 5.2.1 that $\lambda v \in \partial \lambda f(x)$. When $|v - \bar{v}| < \tilde{\epsilon}$ and $|f(x) - f(\bar{x})| < \tilde{\epsilon}$ we have $|\lambda v - \lambda \bar{v}| < \lambda\tilde{\epsilon}$ and $|\lambda f(x) - \lambda f(\bar{x})| < \lambda\tilde{\epsilon}$. Let $\epsilon = \min\{\tilde{\epsilon}, \lambda\tilde{\epsilon}\}$, and let $r = \lambda\tilde{r}$. We conclude that

$$\lambda f(x') \geq \lambda f(x) + \langle \lambda v, x' - x\rangle - \frac{r}{2}|x' - x|^2$$

whenever $|x' - \bar{x}| < \epsilon$, $|x - \bar{x}| < \epsilon$, $|\lambda f(x) - \lambda f(\bar{x})| < \epsilon$, $\lambda v \in \partial f(x)$, and $|\lambda v - \lambda \bar{v}| < \epsilon$. □

A direct result of Lemma 5.2.3 is that the scalar multiplication of a prox-regular function with a positive parameter results in a para-prox-regular function.

Corollary 5.2.4. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be prox-regular at $\bar{x}$ for $\bar{v}$ with parameters $\tilde{\epsilon} > 0$ and $\tilde{r} \geq 0$, and let $\bar{\lambda} \in \mathbb{R}$, $\bar{\lambda} > 0$. Then $g(x, \lambda) := \lambda f(x)$ is para-prox-regular at $(\bar{x}, \bar{\lambda})$ for $\bar{\lambda}\bar{v}$, with parameters $\epsilon = \min\{\tilde{\epsilon}, \frac{\bar{\lambda}\tilde{\epsilon}}{2}, \frac{\bar{\lambda}}{2}\}$ and $r = \frac{3\bar{\lambda}\tilde{r}}{2}$.*

Proof: Let $\delta = \frac{\bar{\lambda}}{2}$ and consider $\lambda \in (\bar{\lambda} - \delta, \bar{\lambda} + \delta)$. Then for each such λ, using $\bar{\epsilon} = \min\{\tilde{\epsilon}, \delta\tilde{\epsilon}\}$ we have that

$$g(x', \lambda) \geq g(x, \lambda) + \langle \lambda v, x' - x \rangle - \frac{\lambda \tilde{r}}{2}|x' - x|^2$$

whenever $|x' - \bar{x}| < \bar{\epsilon}$, $|x - \bar{x}| < \bar{\epsilon}$, $|g(x, \bar{\lambda}) - g(\bar{x}, \bar{\lambda})| < \bar{\epsilon}$, $\lambda v \in \partial_x g(x, \lambda)$, and $|\lambda v - \bar{\lambda}\bar{v}| < \bar{\epsilon}$. Let $\epsilon = \min\{\delta, \bar{\epsilon}\}$ and $r = \sup_{\lambda}\{\lambda\tilde{r}\} = \frac{3\bar{\lambda}\tilde{r}}{2}$. Then we have

$$g(x', \lambda) \geq g(x, \lambda) + \langle \lambda v, x' - x \rangle - \frac{r}{2}|x' - x|^2$$

when $|x' - \bar{x}| < \epsilon$, $|x - \bar{x}| < \epsilon$, $|g(x, \lambda) - g(\bar{x}, \bar{\lambda})| < \epsilon$, $\lambda v \in \partial_x g(x, \lambda)$, $|\lambda v - \bar{\lambda} v| > \epsilon$, and $|\lambda - \bar{\lambda}| < \epsilon$. □

Next we have a minor extension of [27, Theorem 3.2], where Poliquin and Rockafellar showed that, under certain conditions, the sum of two prox-regular functions is again prox-regular. We extend this to the sum of any finite number of prox-regular functions. The result also extracts the prox-regularity parameters from the proof, thus allowing their use in later results.

Theorem 5.2.5. *For $i \in \{1, 2, \ldots, m\}$, let $f_i : \mathbb{R}^n \to \mathbb{R}$. Let $\bar{x} \in \bigcap_{i=1}^{m} \operatorname{dom} f_i$ and assume that NDC holds. Define $f(x) := \sum_{i=1}^{m} f_i(x)$ and let $\bar{v} \in \partial f(\bar{x})$. Suppose there exist $\epsilon_i > 0$ and $r_i \geq 0$ such that for any $v_i \in \partial f_i(\bar{x})$ with $\sum_{i=1}^{m} v_i = \bar{v}$ we have that f_i is prox-regular at $\bar{x}$ for v_i with parameters ϵ_i and r_i. Then f is prox-regular at $\bar{x}$ for $\bar{v}$ with parameters $\epsilon = \min_i\{\epsilon_i\}$ and $r = m \max_i\{r_i\}$.*

Proof: Define $g(u_1, u_2, \ldots, u_m) := \sum_{i=1}^{m} f_i(u_i)$ and $F(x) := (x, x, \ldots, x)$. Then g is lsc and proper, $F \in \mathcal{C}^2$, and $f(x) = g(F(x))$. Notice that $\nabla F(x) = (I, I, \ldots, I)^\top$, so the only possible $y \in \partial^\infty g(F(\bar{x}))$ with $\nabla F(\bar{x})^* y = 0$ is $y = 0$. Therefore the constraint qualification on y in Theorem 5.1.2 holds. We have that g is prox-regular at $F(\bar{x})$ for every $y \in \partial g(F(\bar{x}))$ with $\nabla F(\bar{x})^* y =$

$\bar{v}$, so all the conditions of Theorem 5.1.2 are satisfied. Let ϵ and r be as described in the proof of that theorem. In particular, $r_1 = 0$ (equation (5.1.4)) in this case since $\nabla F(x_1) - \nabla F(x_0) = 0$. Also $r_2 = 0$ (equation (5.1.6)), since $\langle \eta, F(x) \rangle = \langle \eta_1, x \rangle + \cdots + \langle \eta_n, x \rangle$, which implies $\nabla \langle \eta, F(x) \rangle = \eta_1 + \cdots + \eta_n$ and $\nabla^2 \langle \eta, F(x) \rangle = 0$. Finally,

$$\begin{aligned} |F(x_1) - F(x_0)| &= |(x_1 - x_0, x_1 - x_0, \ldots, x_1 - x_0)| \\ &= \sqrt{m|x_1 - x_0|^2} \\ &= \sqrt{m}|x_1 - x_0|, \end{aligned}$$

which gives $k = \sqrt{m}$ (equation (5.1.8)). Hence $r = m\bar{r} = m \max_i \{r_i\}$ (equation (5.1.2)). We conclude that f is prox-regular at $\bar{x}$ for $\bar{v}$ with these parameters. □

Corollary 5.2.6. *For $i \in \{1, 2, \ldots, m\}$, let $f_i : \mathbb{R}^n \to \mathbb{R}$ be prox-regular at $\bar{x}$ for $v_i \in \partial f_i(\bar{x})$ with parameters $\epsilon_i > 0$ and $r_i \geq 0$, and let NDC hold. Let $\lambda \in \mathbb{R}^m$, and fix $\bar{\lambda} = (\bar{\lambda}_1, \bar{\lambda}_2, \ldots, \bar{\lambda}_m) > 0$ (i.e. $\bar{\lambda}_i > 0$ for all i). Define*

$$f(x, \lambda) := \sum_{i=1}^{m} \lambda_i f_i(x).$$

Then $f(\cdot, \bar{\lambda})$ is prox-regular in x at $\bar{x}$ for $\bar{\lambda}\bar{v}$ with parameters $\epsilon = \min_i \{\epsilon_i, \bar{\lambda}_i \epsilon_i\}$ and $r = m \max_i \{\bar{\lambda}_i r_i\}$, where $\bar{v} = \sum_{i=1}^{m} \bar{\lambda}_i v_i \in \partial_x f(\bar{x}, \bar{\lambda})$.

Proof: Apply Lemma 5.2.3 to each f_i in Theorem 5.2.5. □

We now examine parametrized functions by multiplying each of the prox-regular functions f_i in Theorem 5.2.5 by a parameter λ_i, and we show that the weighted sum of prox-regular functions is a para-prox-regular function.

Theorem 5.2.7. *For $i \in \{1, 2, \ldots, m\}$, let $f_i : \mathbb{R}^n \to \mathbb{R}$. Let $\bar{x} \in \bigcap_{i=1}^{m} \operatorname{dom} f_i$ and $\bar{\lambda} = (\bar{\lambda}_1, \bar{\lambda}_2, \ldots, \bar{\lambda}_m) > 0$. Assume that NDC holds. Define*

$$f(x, \lambda) := \sum_{i=1}^{m} \lambda_i f_i(x)$$

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$, and let $\bar{v} \in \partial_x f(\bar{x}, \bar{\lambda})$. If f_i is prox-regular at $\bar{x}$ for all $\bar{v}_i \in \partial f_i(\bar{x})$ such that $\sum_{i=1}^{m} \bar{\lambda}_i \bar{v}_i = \bar{v}$, with parameters ϵ_i and r_i, then f is para-prox-regular at $(\bar{x}, \bar{\lambda})$ for $\bar{v}$, with parameters $\epsilon = \min_i \{\frac{\bar{\lambda}_i}{2}, \epsilon_i, \frac{\bar{\lambda}_i}{2}\epsilon_i\}$ and $r = m \max_i \{\frac{3\bar{\lambda}_i}{2} r_i\}$.

Proof: Since $\bar{\lambda}_i > 0$ and f_i is prox-regular at $\bar{x}$ for $\bar{v}_i$ for all i, Lemma 5.2.3 applies and we have that each $\bar{\lambda}_i f_i$ is prox-regular at $\bar{x}$ for $\bar{\lambda}_i \bar{v}_i$ with parameters $\tilde{\epsilon}_i = \min\{\epsilon_i, \bar{\lambda}_i \epsilon_i\}$ and $\tilde{r}_i = \bar{\lambda}_i r_i$. Since $\bar{\lambda}_i > 0$, we have $\partial^\infty \bar{\lambda}_i f_i = \partial^\infty f_i$ by Lemma 5.2.2, so NDC holds. Thus by Corollary 5.2.6, we have that $f(x, \bar{\lambda})$ is prox-regular in x at $\bar{x}$ for $\bar{v}$ with parameters $\bar{\epsilon} = \min_i\{\tilde{\epsilon}_i\}$ and $\bar{r} = m \max_i\{\tilde{r}_i\}$. (Note that here we are viewing f as a function of x, with $\bar{\lambda}$ fixed.) Define $\delta := \frac{1}{2}\min_i\{\bar{\lambda}_i\}$, and $\mathcal{B} := \{\lambda : \lambda_i \in [\bar{\lambda}_i - \delta, \bar{\lambda}_i + \delta] \text{ for all } i\}$. Since $\lambda > 0$ for all $\lambda \in \mathcal{B}$, by similar arguments as above we have that $f(x, \lambda)$ is prox-regular in x at $\bar{x}$ for $\bar{v}$ for every fixed $\lambda \in \mathcal{B}$, with parameters $\epsilon_\lambda = \min_i\{\epsilon_i, \lambda_i \epsilon_i\}$ and $r_\lambda = m \max_i\{\lambda_i r_i\}$. But $\min_{\lambda \in \mathcal{B}}\{\lambda_i\} = \delta$ and $\max_{\lambda \in \mathcal{B}}\{\lambda_i\} = \max_i\{\bar{\lambda}_i\} + \delta$, so for all $\lambda \in \mathcal{B}$ we have that $f(x, \lambda)$ is prox-regular in x at $\bar{x}$ for $\bar{v}$ with parameters $\hat{\epsilon} = \min\{\epsilon_i, \delta \epsilon_i\}$ and $\hat{r} = m \max_i\{(\bar{\lambda}_i + \delta) r_i\}$. Since $\bar{\lambda}_i + \delta \le \frac{3\bar{\lambda}_i}{2}$, we may overestimate and set $r = m \max_i\{\frac{3\bar{\lambda}_i}{2} r_i\}$. Therefore we may conclude that f is para-prox-regular at $(\bar{x}, \bar{\lambda})$ for $\bar{v}$ with parameters $\epsilon = \min\{\delta, \hat{\epsilon}\} = \min_i\{\frac{\bar{\lambda}_i}{2}, \epsilon_i, \frac{\bar{\lambda}_i}{2}\epsilon_i\}$ and $r = m \max_i\{\frac{3\bar{\lambda}_i}{2} r_i\}$. □

In the next example, for f_1 and f_2 lower-$\mathcal{C}^2$ functions this result has already been proved [14]. The lower-$\mathcal{C}^2$ requirement is generalized to prox-regular by using Theorem 5.2.7.

Example 5.2.8. *Let $f_1 : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $f_2 : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be prox-regular at $\bar{x} \in \bigcap_{i=1}^{2} \operatorname{dom} f_i$. Assume NDC holds. Define*

$$f(x, \lambda) := \lambda f_1(x) + (1 - \lambda) f_2(x).$$

Let $\bar{v} \in \partial_x f(\bar{x}, \bar{\lambda})$ such that $\bar{v} = \lambda v_1 + (1 - \lambda) v_2$ with $v_1 \in \partial f_1(\bar{x})$ and $v_2 \in \partial f_2(\bar{x})$. Then f is para-prox-regular at $(\bar{x}, \bar{\lambda})$ for $\bar{v} = \lambda v_1 + (1 - \lambda) v_2$, for any $\bar{\lambda} \in (0, 1)$.

5.3 Finite Maximum

The following proposition is a restatement of a result from an earlier work. We shall use it and the corollary that follows in the proof of Theorem 5.3.3, para-prox-regularity of a weighted finite max of $\mathcal{C}^0$ functions.

Proposition 5.3.1. *[19, Proposition 2.2] Suppose that $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is strongly amenable in x at $(\bar{x}, \bar{\lambda})$, in the sense that on some neighborhood of $(\bar{x}, \bar{\lambda})$ there is a composite representation $f(x, \lambda) = g(F(x, \lambda))$ in* *which* $F : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^m$ *is a* $\mathcal{C}^2$ *mapping and* $g : \mathbb{R}^m \to \mathbb{R}$ *is a convex, proper, lsc function for which* $F(\bar{x}, \bar{\lambda}) \in D := \operatorname{dom} g$ *and*

$$z \in N_D(F(\bar{x}, \bar{\lambda})), \nabla_x F(\bar{x}, \bar{\lambda})^* z = 0 \Rightarrow z = 0.$$

Then as long as $\bar{v} \in \partial_x f(\bar{x}, \bar{\lambda})$*, one has* f *continuously para-prox-regular in* x *at* $(\bar{x}, \bar{\lambda})$ *for* $\bar{v}$*.*

This allows us to state Corollary 5.3.2 below, which extends [25, Exercise 2.9], the similar result for the prox-regular case.

Corollary 5.3.2. *For* $i \in \{1, 2, \ldots, m\}$*, let* $f_i : \mathbb{R}^n \to \mathbb{R}$ *be* $\mathcal{C}^2$*. Let* $\bar{\lambda} = (\bar{\lambda}_1, \bar{\lambda}_2, \ldots, \bar{\lambda}_m) > 0$*. Let* $\bar{x} \in \bigcap_{i=1}^{m} \operatorname{dom} f_i$ *and define*

$$f(x, \lambda) := \max_i \{\lambda_i f_i(x)\}$$

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$*. Then* f *is continuously para-prox-regular in* x *at* $(\bar{x}, \bar{\lambda})$ *for* $\bar{v} \in \partial_x f(\bar{x}, \bar{\lambda})$*.*

Proof: Let

$$F(x, \lambda) = (\lambda_1 f_1(x), \lambda_2 f_2(x), \ldots, \lambda_m f_m(x))$$

and

$$g(u_1, u_2, \ldots, u_m) = \max\{u_1, u_2, \ldots, u_m\}.$$

Then F is $\mathcal{C}^2$, since each component is a product of two $\mathcal{C}^2$ functions, g is proper lsc convex, and $f(x, \lambda) = g(F(x))$.
Therefore f is strongly amenable. Since $D = \operatorname{dom} g$ is $\mathbb{R}^m$, $N_D(F(\bar{x}, \bar{\lambda})) = \{0\}$, so the constraint qualification of Proposition 5.3.1 holds. We have all the requirements of Proposition 5.3.1, and its conclusion is our desired result. □

Our next result is a relaxation of the $\mathcal{C}^2$ condition of Corollary 5.3.2. It is sufficient that the f_i be $\mathcal{C}^0$, prox-regular functions. In relaxing the result, we provide a direct proof and also extract the prox-regularity parameters from the proof.

Theorem 5.3.3. *For* $i \in \{1, 2, \ldots, m\}$*, let* $f_i : \mathbb{R}^n \to \mathbb{R}$ *be* $\mathcal{C}^0$*. Let* $\bar{\lambda} = (\bar{\lambda}_1, \bar{\lambda}_2, \ldots, \bar{\lambda}_m) > 0$*. Let* $\bar{x} \in \bigcap_{i=1}^{m} \operatorname{dom} f_i$*. Define*

$$f(x, \lambda) := \max_i \{\lambda_i f_i(x)\}$$

where $\lambda = (\lambda_1, \lambda_2 \ldots, \lambda_m)$, and let $\bar{v} \in \partial_x f(\bar{x}, \bar{\lambda})$. If the f_i are prox-regular at $\bar{x}$ for each $v_i \in \partial f_i(\bar{x})$ with parameters $\epsilon_i > 0$ and $r_i \geq 0$, and if f is Clarke regular at $\bar{x}$ (i.e. every normal vector to $\operatorname{epi} f$ at $\bar{x}$ is a regular normal vector), then f is para-prox-regular at $(\bar{x}, \bar{\lambda})$ for $\bar{v}$ with parameters $\epsilon = \min_i\{\epsilon_i, \frac{\bar{\lambda}_i \epsilon_i}{2}\}$ and $r = \max_i\{\frac{3\bar{\lambda}_i r_i}{2}\}$.

Proof: Let $A(\bar{x}, \bar{\lambda}) = \{i : f(\bar{x}, \bar{\lambda}) = \bar{\lambda}_i f_i(\bar{x})\}$ be the active set of indices. Then $\partial_x f(\bar{x}, \bar{\lambda}) = \operatorname*{conv}_{i \in A(\bar{x}, \bar{\lambda})} \{\bar{\lambda}_i \partial f_i(\bar{x})\}$ [9, Theorem 2.8.2], so every $\bar{v} \in \partial_x f(\bar{x}, \bar{\lambda})$ has the form $\sum_{i \in A(\bar{x}, \bar{\lambda})} \theta_i \bar{\lambda}_i v_i$ for some $v_i \in \partial f_i(\bar{x})$ and some set of non-negative $\{\theta_i\}_{i \in A(\bar{x}, \bar{\lambda})}$ such that $\sum_{i \in A(\bar{x}, \bar{\lambda})} \theta_i = 1$. Corollary 5.2.4 tells us that the $\lambda_i f_i$ are para-prox-regular at $(\bar{x}, \bar{\lambda}_i)$ for all $\bar{\lambda}_i v_i \in \partial \bar{\lambda}_i f_i(\bar{x})$. In particular, for each i we can take $\hat{\epsilon}_i = \min\{\epsilon_i, \frac{\bar{\lambda}_i \epsilon_i}{2}, \frac{\bar{\lambda}_i}{2}\}$ and $\hat{r}_i = \frac{3\bar{\lambda}_i r_i}{2}$ to get

$$\lambda_i f_i(x') \geq \lambda_i f_i(x) + \langle \lambda_i \hat{v}_i, x' - x \rangle - \frac{\hat{r}_i}{2}|x' - x|^2 \tag{5.3.1}$$

for $|x' - \bar{x}| < \hat{\epsilon}_i$, $|x - \bar{x}| < \hat{\epsilon}_i$, $|\lambda_i f_i(x) - \bar{\lambda}_i f(\bar{x})| < \hat{\epsilon}_i$, $\lambda_i \hat{v}_i \in \partial \lambda_i f_i(x)$, $|\lambda_i \hat{v}_i - \bar{\lambda}_i v_i| < \hat{\epsilon}_i$, and $|\lambda_i - \bar{\lambda}_i| < \hat{\epsilon}_i$. This is a system of m inequalities, so using $\epsilon = \min_i\{\hat{\epsilon}_i\}$ and $r = \max_i\{\hat{r}_i\}$ we can say that for all $i \in \{1, 2, \ldots, m\}$

$$\lambda_i f_i(x') \geq \lambda_i f_i(x) + \langle \lambda_i \hat{v}_i, x' - x \rangle - \frac{r}{2}|x' - x|^2 \tag{5.3.2}$$

for all $|x' - \bar{x}| < \epsilon$, $|x - \bar{x}| < \epsilon$, $|\lambda_i - \bar{\lambda}_i| < \epsilon$, $\lambda_i \hat{v}_i \in \partial \lambda_i f_i(x)$, $|\lambda_i \hat{v}_i - \bar{\lambda}_i v_i| < \epsilon$, $|\lambda_i f_i(x) - \bar{\lambda}_i f_i(\bar{x})| < \epsilon$ (with r and ϵ no longer dependent on i). Since $f(x', \lambda) \geq \lambda_i f_i(x')$ for all i, we can replace the left-hand side of inequality (5.3.2) with $f(x', \lambda)$ to get

$$f(x', \lambda) \geq \lambda_i f_i(x) + \langle \lambda_i \hat{v}_i, x' - x \rangle - \frac{r}{2}|x' - x|^2 \tag{5.3.3}$$

for all $|x' - \bar{x}| < \epsilon$, $|x - \bar{x}| < \epsilon$, $|\lambda_i - \bar{\lambda}_i| < \epsilon$, $\lambda_i \hat{v}_i \in \partial \lambda_i f_i(x)$, $|\lambda_i \hat{v}_i - \bar{\lambda}_i v_i| < \epsilon$, $|\lambda_i f_i(x) - \bar{\lambda}_i f_i(\bar{x})| < \epsilon$.

Let x', x, λ, v be such that $|x' - \bar{x}| < \epsilon$, $|x - \bar{x}| < \epsilon$, $v \in \partial_x f(x, \lambda)$, $|v - \bar{v}| < \epsilon$, $|\lambda - \bar{\lambda}| < \epsilon$, $|f(x, \lambda) - f(\bar{x}, \bar{\lambda})| < \epsilon$. Let $A(x, \lambda) = \{i : f(x, \lambda) = \lambda_i f_i(x)\}$. Then $\partial_x f(x, \lambda) = \operatorname*{conv}_{i \in A(x, \lambda)} \{\lambda_i \partial f_i(x)\}$. So $v \in \partial_x f(x, \lambda)$ has the form $\sum_{i \in A(x, \lambda)} \phi_i \lambda_i \hat{v}_i$ for some $\hat{v}_i \in \partial f_i(x)$ and some non-negative $\{\phi_i\}_{i \in A(x, \lambda)}$

such that $\sum_{i\in A(x,\lambda)} \phi_i = 1$. We can also substitute $f(x,\lambda)$ on the right-hand side of inequality (5.3.3) for those $i \in A(x,\lambda)$ to get

$$f(x',\lambda) \geq f(x,\lambda) + \langle \lambda_i \hat{v}_i, x' - x\rangle - \frac{r}{2}|x' - x|^2. \qquad (5.3.4)$$

Multiplying inequality (5.3.4) by ϕ_i and adding over $i \in A(x,\lambda)$ yields

$$\sum_{i\in A(x,\lambda)} \phi_i f(x',\lambda) \geq \sum_{i\in A(x,\lambda)} \phi_i f(x,\lambda) + \langle \sum_{i\in A(x,\lambda)} \phi_i \lambda_i \hat{v}_i, x' - x\rangle - \sum_{i\in A(x,\lambda)} \phi_i \frac{r}{2}|x' - x|^2$$

which simplifies to

$$f(x',\lambda) \geq f(x,\lambda) + \langle v, x' - x\rangle - \frac{r}{2}|x' - x|^2. \qquad (5.3.5)$$

Therefore f is para-prox-regular at $(\bar{x},\bar{\lambda})$ for $\bar{v}$, with parameters ϵ and r. □

Chapter 6

Conclusion

The goal of this thesis was to present para-prox-regular functions as one of the next logical areas of exploration in the study of prox-regularity, and to take a few steps in that direction. We understand the importance of prox-regular functions, as they cover convex functions and $\mathcal{C}^2$ functions among others, and we have seen that para-prox-regular functions are a very natural extension of prox-regular functions. We know that para-prox-regular functions have their place and their importance in the study of the field, since parametrized functions are commonly studied in optimization, in particular with problems that involve stability of minimizers [19, 26]. We have discussed and analyzed several examples of functions that are para-prox-regular, and seen how we can create such functions by manipulating a prox-regular function or a set of them. In each of those examples, we have defined explicitly the prox-regularity parameters that result. We have also seen an example of where para-prox-regularity of a prox-regular function can fail, namely $f(x, \lambda) = \lambda|x|$ at $\bar{\lambda} = 0$.

So we have made significant progress in further understanding how to create para-prox-regular functions, and in getting a feel for what they look like and how they behave. However, the results seen here are limited in at least two key aspects. First, our results concerning weighted parametrized sums and weighted parametrized max functions rely on starting with a finite number of prox-regular functions. A logical direction in which to continue with the development of these theorems would be to explore the case where the family of functions is possibly infinite. One of the difficulties here would be to be able to name, or even prove the existence of, the prox-regularity parameters. For a finite number of prox-regular functions f_i with respective parameters ϵ_i and r_i, we are guaranteed the existence of $\min\{\epsilon_i\}$ and $\max\{r_i\}$, which we used in Theorem 5.2.7 and Theorem 5.3.3. In general for an infinite number of such functions, we have no such guarantee.

Second, all the results found in this paper are for functions in $\mathbb{R}^n$, finite-dimensional space. A natural path to take in extending these results would be to reexamine them in the setting of a Hilbert or Banach space. This would be a particularly interesting challenge, as up to now we have relied on properties of finite-dimensional space in order to produce the results we have. For example, the proofs of Proposition 5.1.1 and Theorem 5.1.2 use the fact that a closed, bounded set is compact, a property which fails to be preserved when we make the change to infinite dimensions. Since 2005, there have been a number of papers published that extend particular properties of prox-regular functions from finite-dimensional spaces to Hilbert spaces [1, 5, 6, 21, 29], so it is likely the techniques and results seen in those works would be of use in progressing in this direction for para-prox-regular functions.

Bibliography

[1] M. Bačák, J. M. Borwein, A. Eberhard, and B. S. Mordukhovich. Infimal convolutions and Lipschitzian properties of subdifferentials for prox-regular functions in Hilbert spaces. *J. Convex Anal.*, 17(3-4):737–763, 2010. → pages 46

[2] H.H. Bauschke, Y. Lucet, and M. Trienis. How to transform one convex function continuously into another. *SIAM Rev.*, 50(1):115–132, 2008. → pages 5, 22

[3] H.H. Bauschke, E. Matoušková, and S. Reich. Projection and proximal point methods: convergence results and counterexamples. *Nonlinear Anal.*, 56(5):715–738, 2004. → pages 5, 22

[4] F. Bernard and L. Thibault. Prox-regularity of functions and sets in Banach spaces. *Set-Valued Anal.*, 12(1-2):25–47, 2004. → pages 4

[5] F. Bernard and L. Thibault. Prox-regular functions in Hilbert spaces. *J. Math. Anal. Appl.*, 303(1):1–14, 2005. → pages 46

[6] S. Boralugoda and R. A. Poliquin. Local integration of prox-regular functions in Hilbert spaces. *J. Convex Anal.*, 13(1):27–36, 2006. → pages 4, 46

[7] M. Bounkhel and A. Jofré. Subdifferential stability of the distance function and its applications to nonconvex economies and equilibrium. *J. Nonlinear Convex Anal.*, 5(3):331–347, 2004. → pages 4

[8] H. Brunn. Über Kerneigebiete. *Math. Ann.*, 73(3):436–440, 1913. → pages 2

[9] F. H. Clarke. *Optimization and nonsmooth analysis*, volume 5 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 1990. → pages 43

[10] F. H. Clarke, R. J. Stern, and P. R. Wolenski. Proximal smoothness and the lower-C^2 property. *J. Convex Anal.*, 2(1-2):117–144, 1995. → pages 4

[11] A. Eberhard and R. Wenczel. A study of tilt-stable optimality and sufficient conditions. *Nonlinear Anal.*, 75(3):1260–1281, 2012. → pages 4

[12] W. Hare and C. Planiden. Parametrically prox-regular functions, 2012. submitted to J. Convex Anal. → pages iii

[13] W. Hare and C. Sagastizábal. A redistributed proximal bundle method for nonconvex optimization. *SIAM J. Optim.*, 20(5):2442–2473, 2010. → pages 4

[14] W. L. Hare. A proximal average for nonconvex functions: a proximal stability perspective. *SIAM J. Optim.*, 20(2):650–666, 2009. → pages 5, 22, 23, 24, 25, 41

[15] W. L. Hare. A proximal method for identifying active manifolds. *Comput. Optim. Appl.*, 43(2):295–306, 2009. → pages

[16] W. L. Hare and R. A. Poliquin. Prox-regularity and stability of the proximal mapping. *J. Convex Anal.*, 14(3):589–606, 2007. → pages 4

[17] J. Johnstone, V. Koch, and Y. Lucet. Convexity of the proximal average. *J. Optim. Theory Appl.*, 148(1):107–124, 2011. → pages 22

[18] A. B. Levy. Calm minima in parameterized finite-dimensional optimization. *SIAM J. Optim.*, 11(1):160–178, 2000. → pages ii, 33

[19] A. B. Levy, R. A. Poliquin, and R. T. Rockafellar. Stability of locally optimal solutions. *SIAM J. Optim.*, 10(2):580–604, 2000. → pages 5, 16, 41, 45

[20] A. S. Lewis, D. R. Luke, and J. Malick. Local linear convergence for alternating and averaged nonconvex projections. *Found. Comput. Math.*, 9(4):485–513, 2009. → pages 4

[21] D. R. Luke. Finding best approximation pairs relative to a convex and prox-regular set in a Hilbert space. *SIAM J. Optim.*, 19(2):714–739, 2008. → pages 46

[22] H. Minkowski. *Diophantische Approximationen. Eine Einführung in die Zahlentheorie.* Chelsea Publishing Co., New York, 1957. → pages 2

[23] A. Moudafi. An algorithmic approach to prox-regular variational inequalities. *Appl. Math. Comput.*, 155(3):845–852, 2004. → pages 4

[24] R. A. Poliquin and R. T. Rockafellar. Generalized Hessian properties of regularized nonsmooth functions. *SIAM J. Optim.*, 6(4):1121–1137, 1996. → pages 4

[25] R. A. Poliquin and R. T. Rockafellar. Prox-regular functions in variational analysis. *Trans. Amer. Math. Soc.*, 348(5):1805–1838, 1996. → pages ii, 4, 5, 15, 24, 26, 33, 42

[26] R. A. Poliquin and R. T. Rockafellar. Tilt stability of a local minimum. *SIAM J. Optim.*, 8(2):287–299, 1998. → pages 4, 45

[27] R. A. Poliquin and R. T. Rockafellar. A calculus of prox-regularity. *J. Convex Anal.*, 17(1):203–210, 2010. → pages ii, 5, 34, 35, 39

[28] R.T. Rockafellar and R.J.B. Wets. *Variational analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998. → pages 9, 10, 13, 14, 15, 20, 23, 24, 27, 32, 37, 38

[29] M. Sebbah and L. Thibault. Metric projection and compatibly parameterized families of prox-regular sets in Hilbert space. *Nonlinear Anal.*, 75(3):1547–1562, 2012. → pages 46

Appendices

Appendix A: Glossary of notation

$\mathbb{R}$	Set of all real numbers.
$\mathbb{R}^n$	Set of all n-dimensional real vectors.
Σ	Summation.
$\lvert \cdot \rvert$	Euclidean norm.
$\langle \cdot, \cdot \rangle$	Vector inner product.
$\times$	Set product.
$\to$	Maps to.
∇	Function gradient.
argmin	Set of minimizers of a function.
inf	Infimum of a function.
sup	Supremum of a function.
$\mathcal{C}^0$	Set of continuous functions.
$\mathcal{C}^1$	Set of continuously differentiable functions.
$\mathcal{C}^{1+}$	Set of strictly continuously differentiable functions.
$\mathcal{C}^2$	Set of twice continuously differentiable functions.
dom	Domain of a function.
$:=$	Function definition.
$\hat{\partial}$	Set of regular subgradients.
∂	Set of subgradients.
∂^∞	Set of horizon subgradients.
f^*	Fenchel conjugate of function f.
PA_{f_0,f_1}	Proximal average of functions f_0 and f_1.
e_r	Moreau envelope.
P_r	Proximal mapping.
PA_r	NC proximal average.

Appendix B: Index