Proximal Operators: Semismoothness and Variational Properties

MATH 470 Honours Research Project
McGill University

Ariel Goodwin

December 15, 2020

1 Introduction

> *"For since the fabric of the universe is most perfect, and is the work of a most wise Creator, nothing whatsoever takes place in the universe in which some relation of maximum and minimum does not appear."*
>
> — Leonhard Euler

Mathematical optimization is a well-motivated subject of study that generally focuses on the maximization or minimization of a function, subject to certain constraints. Without loss of generality we speak of minimization, and for the purposes of this writeup we consider only the problem of continuous optimization where the function we seek to minimize is continuous (or at least *lower semicontinuous*).

Applications of continuous optimization include problems in the areas of image recovery, machine learning, and operations research. In machine learning, for example, we may seek to minimize the error associated with a certain prediction model. Such work has become increasingly popular over the last few decades and continues to develop rapidly. In the 20th century, much work was done on the special case of linear optimization, known as linear programming. Linear programming problems are characterized by affine constraints and objective functions. These problems arise frequently in economics and business management, and the theory of linear programming is generally well-studied.

Convex analysis is an analytic and geometric body of knowledge that lends itself to convex optimization, in which the objective function and constraints are convex. This subject has been developed over the last few centuries, with a few notable figures including Rockafellar, Moreau, Fenchel, and Minkowski. It has given us a rich set of tools for solving modern optimization problems, and a wealth of theory to be expanded upon.

In the work that follows we concern ourselves with convex functions and a particular tool due to Moreau that is known as the *proximal operator* of a closed proper convex function. The proximal operator is a tool employed by many convex optimization algorithms, see [2] for examples. A computation of interest that often show up as sub-problems in convex constrained optimization is determining the projection of points onto the level sets or epigraph of a closed proper convex function. This is a problem where the proximal operator arises naturally and its smoothness properties give us information about how efficiently we can compute these projections. The contents of this writeup pertain to some results and examples related to the semismoothness and semismoothness* of the proximal operator via the semismoothness* of the subdifferential of the associated function.

I would like to thank Tim Hoheisel for providing the opportunity to work on this project and for his insights and guidance over the duration of the semester.

2 Preliminaries

Let $\mathbb{E}$ be a finite-dimensional real vector space, and $\overline{\mathbb{R}}$ the extended real numbers i.e., $\mathbb{R} \cup \{\pm\infty\}$. A set $C \subseteq \mathbb{E}$ is *convex* if

$$\lambda x + (1-\lambda)y \in C \ \ \forall x, y \in C, \ \lambda \in [0,1]$$

A set $K \subseteq \mathbb{E}$ is a *cone* if $\lambda x \in K$ for all $\lambda \geq 0$. Given a cone K, the *polar cone* K^* is

$$K^* = \{v \mid \langle v, u\rangle \leq 0 \, \forall u \in K\}$$

We define the *epigraph* of a function $f\colon \mathbb{E} \to \overline{\mathbb{R}}$ to be the set

$$\operatorname{epi} f = \{(x, \alpha) \in \mathbb{E} \times \mathbb{R} \mid f(x) \leq \alpha\}$$

We say that f is *proper* if f never takes on the value $-\infty$ and its domain

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

is nonempty. A function $f\colon \mathbb{E} \to \overline{\mathbb{R}}$ is said to be convex if its epigraph is a convex set, and it is easy to verify that this agrees with the other common definition of convexity characterized by $\operatorname{dom} f$ being convex and

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \ \ \forall x, y \in \operatorname{dom} f, \ \lambda \in [0,1]$$

For a function f its *level sets* are given by

$$\operatorname{Lev}(f, \alpha) = \{x \in \mathbb{E} \mid f(x) \leq \alpha\}$$

We say that a function f as above is *closed* if its epigraph is a closed set. An equivalent property is *lower semicontinuity*, meaning that

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

for all $x \in \mathbb{E}$. In fact, the famous result that a continuous function attains its minimum over a nonempty compact set holds also for functions that are only lower semicontinuous. Finally, for a function f we define the *subdifferential* of f at $x \in \operatorname{dom} f$ to be

$$\partial f(x) = \{u \in \mathbb{E} \mid f(y) \geq f(x) + \langle u, y - x \rangle \ \forall y \in \mathbb{E}\}$$

We define the class Γ_0 as the set of closed, proper, and convex functions on a finite-dimensional vector space $\mathbb{E}$ with values in $\mathbb{R} \cup \{+\infty\}$. In particular, we explore various properties of the set-valued proximal operator, defined by

$$P_\lambda f(\bar{x}) = \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \, \lambda f(u) + \frac{1}{2} \|u - \bar{x}\|, \text{ for } \lambda > 0, \ \bar{x} \in \mathbb{E}$$

A powerful generalization of the usual Euclidean projection operator, the proximal operator emerges in many applications of optimization.

3 Motivation

The motivating example for this research project is the problem of projecting onto the epigraph and level sets of a function $f \in \Gamma_0$. The recent work of Hoheisel, Friedlander, and Goodwin [6] presents a framework in which these projections can be computed by solving a scalar convex optimization problem using a generalized Newton method. The generalization in this method comes from the notion of a semismooth gradient of the objective function, and it is this semismoothness property that provides theoretical guarantees of convergence rates (see [9]). The classical notion of semismoothness applies generally to functions from $\mathbb{R}^n \to \mathbb{R}^m$. A locally Lipschitz continuous function $\varphi \colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$, Ω open, is *semismooth* at $\bar{z} \in \Omega$ if φ is directionally differentiable at z and there exists $\varepsilon > 0$ such that

$$\|\varphi'(z; z - \bar{z}) - \varphi(\bar{z}; z - \bar{z})\| = o(\|z - \bar{z}\|), \ \ \forall z \in B_\varepsilon(\bar{z})$$

It turns out that the semismoothness properties of the gradient in this problem are intimately related to the semismoothness properties of the proximal operator, and these properties can be traced back to the semismoothness of the subdifferential of f. This is slight abuse of language; the subdifferential and prox operator are set-valued mappings so the notion of semismoothness is not defined for these functions. We will introduce the proper terminology after a brief discussion of the prerequisite tools of variational analysis.

4 Variational Analysis

We introduce some tools from variational analysis to aid in our discussion. The subject matter, as described by Rockafellar and Wets in their monograph *Variational Analysis* encompasses "problems of optimization, equilibrium, control, and stability of linear and nonlinear systems" ([10], page i). These ideas are useful for analyzing set-valued mappings such as the prox operator.

Given a set-valued mapping $S\colon \mathbb{E}_1 \rightrightarrows \mathbb{E}_2$, we define the *outer limit* of S at $\bar{x}$ to be

$$\operatorname*{Lim\,sup}_{x\to\bar{x}} S(x) = \{y \mid \exists\{x_k\} \to \bar{x}, y_k \in S(x_k) \;:\; y_k \to y\}$$

For a given subset $A \subseteq \mathbb{E}$, we have the *tangent cone* and *regular normal cone*, respectively, of A at $\bar{x} \in A$

$$T_A(\bar{x}) = \operatorname*{Lim\,sup}_{t\downarrow 0} \frac{A-\bar{x}}{t} = \left\{w \mid \exists\,\{x_k\}, \{t_k\},\; x_k \underset{A}{\to} \bar{x},\; t_k \downarrow 0,\; (x_k-\bar{x})/t_k \to w\right\}$$

$$\hat{N}_A(\bar{x}) = \{v \mid \langle v,\, y\rangle \le 0\; \forall y \in T_A(\bar{x})\}$$

Notice that the regular normal cone is the polar cone of the tangent cone. It is also worthwhile to note that they are both closed cones; the closedness of the tangent cone can be seen by considering a diagonal argument on the sequences associated to a convergent sequence in $T_A(\bar{x})$. The normal cone is closed because the polar cone is always closed. The fact that they are cones follows directly from the definition because we can easily rescale the sequence in the tangent cone definition and the inequality in the regular normal cone definition will be preserved by nonnegative scalar multiples.

As an example to help us understand the behaviour of the tangent cone we prove the following well-known result about the tangent cone to a closed cone at the origin.

Proposition 4.1. Let K be a closed cone. Then $T_K(0) = K$.

Proof. First suppose $v \in K$. Then for all $n \in \mathbb{N}$, $(1/n)v \in K$ because K is a cone. Also, the sequence $(1/n)v$ converges in K to the vector 0 as $n \to \infty$. It follows that

$$v = \frac{(1/n)v - 0}{(1/n)} \in T_K(0)$$

by definition of the tangent cone. On the other hand, suppose $v \in T_K(0)$. Then there exist sequences $\{x_k\}, \{t_k\}$ such that $x_k \underset{K}{\to} 0$, $t_k \downarrow 0$ such that

$$\frac{x_k}{t_k} \to v$$

as $k \to \infty$. The fact that $x_k \in K$ for all $k \in \mathbb{N}$ combined with the cone property ensures $x_k/t_k \in K$ for all k. Finally the closedness of K tells us that the limit vector v lies in K, proving the equality. $\square$

Remark 4.1. The tangent and regular normal cones are inherently local structures. More precisely, the tangent cone to A at $\bar{x}$ is the same as the tangent cone to $U \cap A$ at $\bar{x}$ where U is a neighbourhood of $\bar{x}$. This is due to the nature of the convergent sequences defined in the definition of the tangent cone. The definition of the regular normal cone in terms of the tangent cone carries over this property.

We extend the definition of the regular normal cone to the *limiting normal cone* defined by

$$N_A(\bar{x}) = \operatorname*{Lim\,sup}_{x\to\bar{x}} \hat{N}_A(x)$$

A set A is *Clarke regular* at $\bar{x}$ if A is *locally closed* at $\bar{x}$ and $N_A(\bar{x}) = \hat{N}_A(\bar{x})$. Local closedness is the property that $A \cap V_\varepsilon$ is closed for some closed neighbourhood V_ε of $\bar{x}$.

The *coderivative* of S at $(\bar{x}, \bar{u}) \in \operatorname{gph} S$ is $D^*S(\bar{x}|\bar{u}) \colon \mathbb{E}_2 \rightrightarrows \mathbb{E}_1$ characterized by

$$v \in D^*S(\bar{x}|\bar{u})(y) \iff (v, -y) \in N_{\operatorname{gph} S}(\bar{x}, \bar{u})$$

The *graphical derivative* of S at $(\bar{u}, \bar{v})$ is $DS(\bar{u}|\bar{v}) \colon \mathbb{E}_1 \rightrightarrows \mathbb{E}_2$ characterized by

$$z \in DS(\bar{u}|\bar{v})(w) \iff (w, z) \in T_{\operatorname{gph} S}(\bar{u}, \bar{v})$$

Remark 4.2. We often study convex sets and functions so it is useful to understand how these cones behave when the set in question is convex.

This next result is given in Rockafellar-Wets and proven here in more detail.

Theorem 4.3. [10, Theorem 6.9] Let $C \subseteq \mathbb{R}^n$ be convex, and $\bar{x} \in C$. Then

$$N_C(\bar{x}) = \hat{N}_C(\bar{x}) = \{v \mid \langle v,\, x - \bar{x}\rangle \leq 0 \,\forall x \in C\}$$

$$T_C(\bar{x}) = \operatorname{cl}\{w \mid \exists \lambda > 0 \text{ such that } \bar{x} + \lambda w \in C\}$$

Proof. Let $T = \{w \mid \exists \lambda > 0 \text{ such that } \bar{x} + \lambda w \in C\}$. If $w \in T$, then we can use the convexity of C to see that

$$\alpha(\bar{x} + \lambda w) + (1 - \alpha)\bar{x} = \bar{x} + \alpha\lambda w \in C \;\; \forall \alpha \in (0, 1)$$

Then as $\alpha \downarrow 0$, we have $\alpha\lambda \downarrow 0$ and we find that $\bar{x} + \alpha\lambda w \underset{C}{\to} \bar{x}$ while $(\bar{x} + \alpha\lambda w - \bar{x})/(\alpha\lambda w) = w$ so this converges to w trivially. Thus $w \in T_C(\bar{x})$ by definition of the tangent cone. This proves $T \subseteq T_C(\bar{x})$. Taking the closure along with the fact that the tangent cone is closed implies $\operatorname{cl} T \subseteq T_C(\bar{x})$. On the other hand, if $w \in T_C(\bar{x})$ then there exist sequences $\{x_k\}, x_k \underset{C}{\to} \bar{x}, \{t_k\}, t_k \downarrow 0$. We have

$$\bar{x} + t_k \left(\frac{x_k - \bar{x}}{t_k} \right) \in C$$

so w is a limit of points belonging to T i.e., $w \in \operatorname{cl} T$ which proves the desired equality.

Let $N = \{v \mid \langle v,\, x - \bar{x}\rangle \leq 0 \,\forall x \in C\}$. Suppose $v \in \hat{N}_C(\bar{x})$. Let $x \in C$. Using the convexity of C we can take any sequence $\{t_k\}, t_k \downarrow 0$ and for sufficiently large k we have $t_k x + (1 - t_k)\bar{x} \in C$. Then

$$\frac{t_k x + (1 - t_k)\bar{x} - \bar{x}}{t_k} = x - \bar{x}$$

so in fact $x - \bar{x} \in T_C(\bar{x})$. Thus $\langle v,\, x - \bar{x}\rangle = 0$ so $\hat{N}_C(\bar{x}) \subseteq N$. For the reverse inclusion, we take $v \in N$ and use our newfound characterization of the tangent cone to write any $y \in T_C(\bar{x})$ as the limit of points $\{y_k\}$ satisfying

$$\bar{x} + \lambda_k y_k = x_k \iff y_k = (x_k - \bar{x})/\lambda_k$$

Consequently,

$$\langle v,\, y_k\rangle = \frac{1}{\lambda_k}\,\langle v,\, x_k - \bar{x}\rangle \le 0$$

By continuity of the inner product $\langle v,\, y\rangle \le 0$. This proves the first equality. It remains to prove that the limiting normal cone is contained in the regular normal cone. If $v \in N_C(\bar{x})$ there exist sequences $\{x_k\}$, $\{v_k\}$, $x_k \underset{C}{\to} \bar{x}$ and $v_k \to v$ with $v_k \in \hat{N}_C(x_k)$ for each k. Using our new characterization of the regular normal cone we have

$$\langle v,\, x - x_k\rangle \le 0$$

for any $x \in C$. Letting $k \to \infty$ proves that $v \in \hat{N}_C(\bar{x})$, completing the proof. $\square$

Computing tangent cones is a usual first step in the examples that follow so we require some properties on how these cones behave under familiar set operations. The text of Aubin and Frankowska, *Set-Valued Analysis* [1], deals with such results extensively. An important subtlety is that their definition of the tangent cone $T_C(\bar{x})$, which is referred to as the *contingent cone* in the text, admits the possibility that the point $\bar{x}$ belongs to the closure of C. This of course coincides with the definition given earlier if $\bar{x}$ already belongs to the set but we must be aware of when it lies in the closure. To distinguish these cones when they arise we use a superscript $\mathcal{A}$. The following theorem is given in Aubin and Frankowska's text.

Theorem 4.4. [1, Table 4.1] If $K_i \subseteq \mathbb{R}^n$ $(i = 1, \dots, m)$ and $x \in \mathrm{cl}\,(\bigcup_i K_i)$ then $T^{\mathcal{A}}_{\bigcup_{i=1}^m K_i}(x) = \bigcup_{i \in I(x)} T^{\mathcal{A}}_{K_i}(x)$ where $I(x) = \{i \mid x \in \mathrm{cl}\, K_i\}$

5 Semismoothness

Ultimately we wish to understand the semismoothness of a scalar function to compute epigraphical projections, and our analysis requires us to study the semismoothness of the proximal operator. Semismoothness is a property of functions between vector spaces, while the proximal operator is a set-valued mapping. This is where the related notion of *semismooth** sets and functions, due to Gfrerer and Outrata [8], becomes important. The definition makes use of the *directional normal cone* at $\bar{x}$ in direction $\bar{u}$

$$N_A(\bar{x}; \bar{u}) = \operatorname*{Lim\,sup}_{\substack{u \to \bar{u} \\ t \downarrow 0}} \hat{N}_A(\bar{x} + tu)$$

First we present a simple lemma about the directional normal cone.

Lemma 5.1. Let $A \subseteq \mathbb{E}$. Then for all $\bar{x}, \bar{u} \in \mathbb{E}$, $N_A(\bar{x}; \bar{u}) \subseteq N_A(\bar{x})$ and if $\bar{u} \notin T_A(\bar{x})$ then $N_A(\bar{x}; \bar{u}) = \emptyset$.

Proof. Let $y \in N_A(\bar{x}; \bar{u})$. Then $y \in \operatorname{Lim\,sup}_{\substack{u \to \bar{u} \\ t \downarrow 0}} \hat{N}_A(\bar{x} + tu)$ so there exist sequences $\{u_k\} \to \bar{u}$, $\{t_k\} \downarrow 0$, $y_k \in \hat{N}_A(\bar{x} + t_k u_k)$, $y_k \to y$. Then we may apply

Proposition 6.6 from Rockafellar-Wets to see that the limit vector y belongs to $N_A(\bar{x})$.

For the second part we prove the contrapositive. If $N_A(\bar{x};\bar{u}) \neq \emptyset$, choose $v \in N_A(\bar{x};\bar{u})$. Then by definition of the directional normal cone, there exist sequences $\{u_k\} \to \bar{u}$, $\{t_k\} \downarrow 0$, $\{v_k\} \to v$ such that $v_k \in \hat{N}_A(\bar{x}+t_k u_k)\ \forall k$. Then the sequence $\{\bar{x}+t_k u_k\} \underset{A}{\to} \bar{x}$ and $(\bar{x}+t_k u_k-\bar{x})/t_k \to \bar{u}$ which implies $\bar{u} \in T_A(\bar{x})$ by definition. $\square$

We now introduce the companion notion of semismoothness*. A set $C \subseteq \mathbb{E}$ is said to be *semismooth** at $\bar{x} \in C$ if we have

$$\langle z,\, u\rangle = 0 \ \ \forall u \in \mathbb{E},\ z \in N_C(\bar{x};u)$$

Moreover, a set-valued mapping $S\colon \mathbb{E}_1 \rightrightarrows \mathbb{E}_2$ is said to be semismooth* at $(\bar{x},\bar{y}) \in \operatorname{gph} S$ if $\operatorname{gph} S$ is semismooth* at $(\bar{x},\bar{y})$.

Of particular interest for us is the semismoothness* of the proximal map as a function of λ and of $\bar{x}$. Here we present a special case of a more general resut in [6] that demonstrates the relationship between semismoothness* of the prox and the subdifferential.

Theorem 5.2. [6, Corollary 3] Let $f \in \Gamma_0$, and $(\bar{x},\bar{\lambda}) \in \mathbb{E}\times\mathbb{R}_{++}$. Then the map $P_\lambda f(x) = \operatorname{argmin}_{u\in\mathbb{E}} f(u) + \frac{1}{2\lambda}\|u-x\|^2$ is semismooth* at $\big((\bar{x},\bar{\lambda}), P_{\bar{\lambda}} f(\bar{x})\big)$ if ∂f is semismooth* at $\left(P_{\bar{\lambda}} f(\bar{x}), \frac{\bar{x}-P_{\bar{\lambda}} f(\bar{x})}{\bar{\lambda}}\right)$.

Proof. Since $\mathbb{E}$ is a finite-dimensional real vector space, we may assume $\mathbb{E} = \mathbb{R}^n$. Define the function $F\colon \mathbb{R}^n \times \mathbb{R}_{++} \times \mathbb{R}^n \to \mathbb{R}^{2n}$ by $F(x,\lambda,z) = \begin{pmatrix} z \\ \frac{x-z}{\lambda}\end{pmatrix}$. Then for any (x,λ,z) in the domain of F, we compute the Jacobian to find

$$F'(x,\lambda,z) = \begin{pmatrix} 0 & 0 & I \\ \frac{1}{\lambda}I & -\frac{1}{\lambda^2}I(x-z) & -\frac{1}{\lambda}I \end{pmatrix}$$

A straightforward calculation then reveals that $\ker F'(x,\lambda,z)^* = \{0\}$, where $*$ denotes the adjoint operator. We now use a result from Dontchev and Rockafellar [5, p. 247] which tells us that this condition implies F is *metrically regular* at (x,λ,z) (this is a condition for the proposition we wish to employ but the exact details are not important so we omit the definition). Then the map $(x,\lambda,z) \mapsto F(x,\lambda,z) - \operatorname{gph}\partial f$ is metrically regular. We remark that

$$\begin{aligned}
(\bar{x},\bar{\lambda},v) \in F^{-1}(\operatorname{gph}\partial f) &\iff \left(v, \frac{\bar{x}-v}{\bar{\lambda}}\right) \in \operatorname{gph}\partial f \\
&\iff \frac{\bar{x}-v}{\bar{\lambda}} \in \partial f(v) \\
&\iff 0 \in \partial f(v) + \frac{1}{\bar{\lambda}}(v-\bar{x}) \\
&\iff v \in \operatorname*{argmin}_{u\in\mathbb{R}^n} f(u) + \frac{1}{2\lambda}\|u-\bar{x}\|^2 \\
&\iff v \in P_{\bar{\lambda}} f(\bar{x}) \\
&\iff \big((\bar{x},\bar{\lambda}),v\big) \in \operatorname{gph} P_\lambda f(x)
\end{aligned}$$

The first equivalence is directly from the definition of F, just note that we have slightly abused the column vector notation and written it instead as a row vector. The fourth equivalence is Fermat's optimality condition (see e.g. [2]), and the remaining equivalencies are immediate from the definition of the graph, prox, and subdifferential.

Thus $F^{-1}(\operatorname{gph}\partial f) = \operatorname{gph} P_{(\cdot)}f(\cdot)$. Now from [6, Proposition 2] we use the semismoothness* of $\operatorname{gph}\partial f$ at $\left(P_{\bar\lambda}f(\bar x), \frac{\bar x - P_{\bar\lambda}f(\bar x)}{\bar\lambda}\right) = F(\bar x, \bar\lambda, P_{\bar\lambda}f(\bar x))$ along with the metric regularity of F to conclude that $F^{-1}(\operatorname{gph}\partial f)$ is semismooth* at $(\bar x, \bar\lambda, P_{\bar\lambda}f(\bar x))$ or equivalently $\operatorname{gph} P_{(\cdot)}f(\cdot)$ is semismooth* at $((\bar x, \bar\lambda), P_{\bar\lambda}f(\bar x))$ i.e., the prox is semismooth* at $((\bar x, \bar\lambda), P_{\bar\lambda}f(\bar x))$. □

With this in mind, it is useful to have some conditions guaranteeing the semismoothness* of ∂f, which we provide in the following theorem.

Theorem 5.3. [6, Proposition 3] Let $f \in \Gamma_0(\mathbb{E})$. Then ∂f is semismooth* at $(\bar x, \bar y) \in \operatorname{gph}\partial f$ if either of the following hold:

i. f is twice continuously differentiable at $\bar x$

ii. f is piecewise linear-quadratic (in which case ∂f is semismooth* on $\mathbb{E}$)

Proof. i) Since $f \in C^2$, we may apply [10, Example 8.34] to see that

$$D(\partial f)(\bar x|\bar y) = \nabla(\partial f)(\bar x) = \nabla^2 f(\bar x) = \nabla^2 f(\bar x)^* = D^*(\partial f)(\bar x|\bar y)$$

The third equality follows because $f \in C^2$ implies the Hessian is symmetric.

To show that $\operatorname{gph}\partial f$ is semismooth* at $(\bar x, \bar y)$, we turn to the definition. We must show that for all $(u, v) \in \mathbb{E} \times \mathbb{E}$ and $(x^*, y^*) \in N_{\operatorname{gph}\partial f}((\bar x, \bar y), (u, v))$ that $\langle (x^*, y^*), (u, v)\rangle = 0$. Using Lemma 3.1, we see that the property is satisfied vacuously for $(u, v) \notin T_{\operatorname{gph}\partial f}((\bar x, \bar y))$ so we only need to consider $(u, v) \in T_{\operatorname{gph}\partial f}((\bar x, \bar y))$.

By definition, we have $v \in DS(\bar x|\bar y)(u) = \nabla^2 f(\bar x)u$. Then for any $(x^*, y^*) \in N_{\operatorname{gph}\partial f}((\bar x, \bar y), (u, v)) \subseteq N_{\operatorname{gph}\partial f}((\bar x, \bar y))$ (the inclusion follows from Lemma 3.1) we again have by definition that $x^* \in D^*(\partial f)(\bar x|\bar y)(-y^*) = -\nabla^2 f(\bar x)y^*$. We find

$$\langle (x^*, y^*), (u, v)\rangle = \langle x^*, u\rangle + \langle y^*, v\rangle = \left\langle -\nabla^2 f(\bar x)y^*, u\right\rangle + \left\langle y^*, \nabla^2 f(\bar x)u\right\rangle = 0$$

which proves the first statement.

ii) Under the second condition, we apply [10, Proposition 12.30] to see that ∂f is piecewise polyhedral, i.e., $\operatorname{gph}\partial f$ is a finite union of polyhedral sets. Then as a corollary of [8, Propositions 3.4/3.5] we obtain the desired result that ∂f is semismooth* at all points in $\operatorname{gph}\partial f$. □

5.1 Examples

5.1.1 Euclidean Norm on $\mathbb{R}$

We focus on analyzing the function $f \in \Gamma_0$, $f\colon \mathbb{R}^n \to \mathbb{R}$ defined by $f(x) = \|x\|_2$. This function is differentiable everywhere but the origin, and it is a well-known result (see [2, Example 3.3]) that the subdifferential of f at the origin is the closed unit ball. To summarize:

$$\partial f(x) = \begin{cases} \frac{x}{\|x\|_2} & x \neq 0 \\ \operatorname{cl} B_1(0) & x = 0 \end{cases}$$

If we then examine the graph of this mapping, we get the formula

$$\operatorname{gph} \partial f = \left\{ \left(x, \frac{x}{\|x\|_2} \right) \mid x \neq 0 \right\} \cup \{(0, u) \mid u \in \operatorname{cl} B_1(0)\} \tag{1}$$

The simplest case is when $n = 1$ and we find that this set is equal to $\mathbb{R}_{--} \times \{-1\} \cup \mathbb{R}_{++} \times \{1\} \cup \{0\} \times [-1, 1]$. In the one-dimensional case f reduces to the absolute value function which is actually piecewise linear-quadratic and so this falls into the realm of Theorem 5.3. However, it is worthwhile to work through the details of this more trivial example from the definitions to build up our intuition for the higher dimensional analogues that will follow. The graph lies in a product space so we are at a loss for straightforward imagery once we go beyond $n = 1$.

We can visualize this set as a subset of $\mathbb{R}^2$ as in the following figure. We remark that this set is not convex, and this holds in general for $n > 1$ by considering a point x belonging to the lefthand set in (1), and y a point belonging to the righthand set in (1).

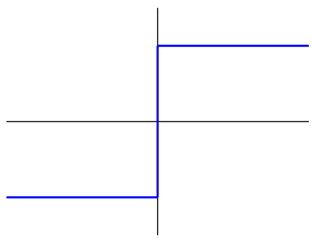

Figure 1: $\operatorname{gph} \partial f$ viewed as subset of $\mathbb{R}^2$ when $n = 1$

The tangent cone: The set $T_{\operatorname{gph} \partial f}((\bar{x}, \bar{y}))$ can be thought of as the set of directions from which we can approach the point $(\bar{x}, \bar{y}) \in \operatorname{gph} \partial f$ from within $\operatorname{gph} \partial f$. It is clear that for a point in $\mathbb{R}_{--} \times \{-1\}$ or $\mathbb{R}_{++} \times \{1\}$ that the only directions along which that point can be approached within $\operatorname{gph} \partial f$ are the directions $(\pm 1, 0)$. Similarly for a point in $\{0\} \times (-1, 1)$, the directions are $(0, \pm 1)$.

The slightly more interesting points are $(0, \pm 1)$ where the possible directions of approach are along $(0, -1)$ and $(1, 0)$ for the point with positive y coordinate, and along $(0, 1)$ and $(-1, 0)$ for the point with negative y coordinate. Of course these properties hold for any scaling of the direction vector chosen (this is part of the verification that the tangent cone is indeed a cone), so we find that for example the tangent cone at $(1,1)$ is $T_{\operatorname{gph}\partial f}((1,1)) = \{(\lambda,0) \mid \lambda \in \mathbb{R}\}$.

The regular normal cone: Consider a point $(x,1)$ in $\mathbb{R}_{++} \times \{1\}$. From our discussion above, we determined the tangent cone at this point to be

$$T_{\operatorname{gph}\partial f}((x,1)) = \{(\lambda,0) \mid \lambda \in \mathbb{R}\}$$

Then $v = (v_1, v_2) \in \hat{N}_{\operatorname{gph}\partial f}((x,1))$ if and only if $v_1 \leq 0$ and $v_1 \geq 0$ so we find that

$$N_{\operatorname{gph}\partial f}((x,1)) = \{(0,\lambda) \mid \lambda \in \mathbb{R}\}$$

We find that the result holds for points $(x,-1)$ in $\mathbb{R}_{--} \times \{-1\}$, and that the regular normal cone for points in $\{0\} \times (-1,1)$ is precisely the tangent cone of points in $\mathbb{R}_{--} \times \{-1\} \cup \mathbb{R}_{++} \times \{1\}$.

Now consider the point $(0,1)$. Using our results on the tangent cone at this point, we find that

$$\hat{N}_{\operatorname{gph}\partial f}((0,1)) = \{v = (v_1,v_2) \mid v_1 \leq 0, v_2 \geq 0\}$$

The directional normal cone: Checking semismoothness of the graph at a point $\bar{x}$ requires computing the directional normal cone $N_{\operatorname{gph}\partial f}(\bar{x};\bar{u})$ for a direction vector $\bar{u} \in \mathbb{R} \times \mathbb{R}$. Using Lemma 3.1, we can restrict our choice of $\bar{u}$ to vectors in the tangent cone at $\bar{x}$. The only points of interest in the graph are $(0,\pm 1)$ because it is certainly smooth elsewhere. We focus on $\bar{x} = (0,1)$. From Lemma 3.1, we only need to consider the directions contained in $T_{\operatorname{gph}\partial f}(\bar{x})$. We will consider the direction $\bar{u} = (1,0)$; the other direction is analogous.

$$\begin{aligned} N_{\operatorname{gph}\partial f}(\bar{x};\bar{u}) &= \operatorname*{Lim\,sup}_{\substack{u \to \bar{u} \\ t \downarrow 0}} \hat{N}_{\operatorname{gph}\partial f}(\bar{x} + tu) \\ &\supseteq \left\{ \begin{pmatrix} 0 \\ \lambda \end{pmatrix} \,\middle|\, \lambda \in \mathbb{R} \right\} \end{aligned}$$

The inclusion follows taking the sequence $u_k = (1,0)$ and any sequence $t_k \downarrow 0$. Then every set inside of the Lim sup falls into the case of the regular normal cone for points in $\mathbb{R}_{++} \times \{1\}$. On the other hand, if $y \in N_{\operatorname{gph}\partial f}(\bar{x};\bar{u})$ then there exist sequences $\{t_k\} \downarrow 0$, $\{u_k\} \to \bar{u}$ such that $\bar{x} + t_k u_k \in \operatorname{gph}\partial f$, $y_k \in \hat{N}_{\operatorname{gph}\partial f}(\bar{x} + t_k u_k)$ for all k with $y_k \to y$. For the sequence of points $\bar{x} + t_k u_k$ to remain in $\operatorname{gph}\partial f$ we must have that eventually the point u_k (and thus $\bar{x} + t_k u_k$) is within the half-line extending from $\bar{x}$ to infinity (and in particular is not $\bar{x}$). It follows that $y \in \left\{ \begin{pmatrix} 0 \\ \lambda \end{pmatrix} \,\middle|\, \lambda \in \mathbb{R} \right\}$ so in fact

$$N_{\operatorname{gph}\partial f}(\bar{x};\bar{u}) = \left\{ \begin{pmatrix} 0 \\ \lambda \end{pmatrix} \,\middle|\, \lambda \in \mathbb{R} \right\}$$

Using this argument, it is easy to show that for the other direction vector $\tilde{u} = (0,-1)$ that

$$N_{\operatorname{gph}\partial f}(\bar{x};\tilde{u}) = \left\{ \begin{pmatrix} \lambda \\ 0 \end{pmatrix} \,\middle|\, \lambda \in \mathbb{R} \right\}$$

Thus for any $z \in N_{\mathrm{gph}\,\partial f}(\bar{x}; \bar{u})$ it holds that $\langle z, \bar{u}\rangle = 0 \cdot 1 + \lambda \cdot 0 = 0$ and similarly for the direction $\tilde{u}$. This verifies that indeed $\mathrm{gph}\,\partial f$ is semismooth* at $(0, 1)$, which we already knew via Theorem 5.3 because the 2-norm is piecewise linear-quadratic in one dimension.

5.2 The Tangent Cone to the Euclidean Norm on $\mathbb{R}^n$

Recall our formula for the graph of the subdifferential to the Euclidean norm on $\mathbb{R}^n$.

$$\mathrm{gph}\,\partial f = \left\{ \left(x, \frac{x}{\|x\|_2} \right) \mid x \neq 0 \right\} \cup \{(0, u) \mid u \in \mathrm{cl}\, B_1(0)\}$$

We can be quite general in our analysis of certain parts of this set. For example, consider the subset $B = \{0\} \times B_1(0)$ (note that we have omitted the boundary). We illustrate the local essence of the tangent and normal cones in the following lemma.

Lemma 5.4. The tangent and normal cones at $\bar{x} \in B \subseteq \mathrm{gph}\,\partial f$ to $\mathrm{gph}\,\partial f$ satisfy

$$T_{\mathrm{gph}\,\partial f}(\bar{x}) = \mathrm{cl}\,\{w \mid \exists \lambda > 0 \text{ such that } \bar{x} + \lambda w \in B\}$$

$$N_{\mathrm{gph}\,\partial f}(\bar{x}) = \hat{N}_{\mathrm{gph}\,\partial f}(\bar{x}) = \{v \mid \langle v,\, x - \bar{x}\rangle \leq 0\, \forall x \in B\}$$

Proof. Let $\bar{x} = (0, \bar{y}) \in B$. If $w \in T_{\mathrm{gph}\,\partial f}(\bar{x})$ then there exist sequences $\{(x_k, y_k)\}$ converging in $\mathrm{gph}\,\partial f$ to $\bar{x}$, $\{t_k\} \downarrow 0$ for which $(x_k - \bar{x})/t_k \to w$. Since $||\bar{y}|| < 1$, any sequence (x_k, y_k) converging in $\mathrm{gph}\,\partial f$ to $\bar{x}$ will eventually satisfy $\|y_k\| < 1$. This immediately forces x_k to zero using the definition of $\mathrm{gph}\,\partial f$. Thus there is no loss of generality in assuming that such a sequence $\{(x_k, y_k)\}$ originates in the set B. As a result this reduces to the tangent cone of B at $\bar{x}$, and we can use the convexity of B to apply Theorem 4.3 and arrive at the given characterization of the tangent and normal cones to the points of B. ☐

Although we will not use this lemma to prove that $\mathrm{gph}\,\partial f$ is semismooth* over B, it is interesting to see how the tangent cone is a local structure in the sense that it only depends on the neighbourhoods around a given point. This idea will be used in the next result but instead for the directional normal cone.

Theorem 5.5. Let $f = \|\cdot\|_2$ be the Euclidean norm on $\mathbb{R}^n$ and let $(0, \bar{y}) \in B$. Then ∂f is semismooth* at $(0, \bar{y}) \in B$.

Proof. We must show that $\mathrm{gph}\,\partial f$ is semismooth* at $(0, \bar{y})$. Since $B_1(0)$ is open there exists a closed neighbourhood V_ϵ around $\bar{y}$ for which V_ϵ is contained in $B_1(0)$. Thus any sequence of points converging to $(0, \bar{y})$ will eventually lie in the set $V = \{0\} \times V_\epsilon$. For any direction $\bar{u}$ and any $z \in N_{\mathrm{gph}\,\partial f}\left((0, \bar{y}); \bar{u}\right)$ there exist sequences $\{u_k\}, \{t_k\}$ with $u_k \to \bar{u}$, $t_k \downarrow 0$ so that the sequence $z_k \to z$ satisfies $z_k \in \hat{N}_{\mathrm{gph}\,\partial f}\left((0, \bar{y}) + t_k u_k\right)$ for all k.

Eventually the points $(0, \bar{y}) + t_k u_k$ must belong to V, so there is no loss of generality in assuming that any sequences associated with $z \in N_{\mathrm{gph}\,\partial f}\left((0, \bar{y}); \bar{u}\right)$

originate from V. The set V is closed and convex so it follows from [7, Lemma 2.1] that

$$\begin{aligned} N_{\operatorname{gph}\partial f}\left((0,\bar{y});\bar{u}\right) &= N_V\left((0,\bar{y});\bar{u}\right) \\ &= \{z \in N_V\left((0,\bar{y})\right) \mid \langle z,\, \bar{u}\rangle = 0\} \\ &= \{z \in N_{\operatorname{gph}\partial f}\left((0,\bar{y})\right) \mid \langle z,\, \bar{u}\rangle = 0\} \end{aligned}$$

The first equality follows from our local approximation of $(0,\bar{y})$, the second equality by the cited lemma, and the third equality again by the locality of the normal cone. Of course this is precisely the definition of $\operatorname{gph}\partial f$ being semismooth* at $(0,\bar{y})$ so we are done. ☐

Based on this lemma, consider the following decomposition of $\operatorname{gph}\partial f$ into three disjoint sets.

$$\operatorname{gph}\partial f = \underbrace{\{0\}\times B_1(0)}_{=:B} \sqcup \underbrace{\{0\}\times S^{n-1}}_{=:S} \sqcup \underbrace{\left\{\left(x, \frac{x}{\|x\|_2}\right) \mid x \in \mathbb{R}^n \setminus \{0\}\right\}}_{=:L}$$

The choice for this decomposition is based on the one-dimensional example. The set B is relatively well-behaved, and the set L was well-behaved when $n = 1$ but it is less clear that this is the case in higher dimensions. In the one-dimensional example, those cusps when B met with L were the more difficult points—these points are encapsulated by the set S. Using this decomposition we now derive a formula for the tangent cone to $\operatorname{gph}\partial f$ at points in S in terms of the Aubin-Frankowska tangent cones to these other sets.

Theorem 5.6. The tangent cone to $\operatorname{gph}\partial f$ at $\bar{x} \in S$ is given by

$$T_{\operatorname{gph}\partial f}(\bar{x}) = T_B^{\mathcal{A}}(\bar{x}) \cup T_S(\bar{x}) \cup T_L^{\mathcal{A}}(\bar{x})$$

Proof. It is not difficult to see that the point $\bar{x}$ belongs to the closures of B, S, and L. The formula then follows from Theorem 4.4, and remarking that $\bar{x} \in S$ so the $\mathcal{A}$-tangent cone is just the usual tangent cone to that set. ☐

The next logical step is to develop an understanding of these $\mathcal{A}$-tangent cones and the tangent cone to S at $\bar{x}$. The latter is easy enough to do with some existing theory.

Theorem 5.7. The tangent cone to S at $\bar{x} = (0,\bar{v})$ is given by

$$T_S(\bar{x}) = \{0\} \times \left\{ w = (w_1,\dots,w_n) \in \mathbb{R}^n \mid \sum_{i=1}^{n} w_i \bar{v}_i = 0 \right\}$$

Proof. The set S is the Cartesian product of a singleton and the unit $(n-1)$-sphere S^{n-1}. The singleton is closed and has normal cone equal to $\{0\}$. The sphere S^{n-1} is closed and is in fact a smooth manifold:

$$S^{n-1} = \{x \in \mathbb{R}^n \mid F(x) := \|x\|_2^2 - 1 = 0\}$$

Note that [10, Example 6.8] deals with this situation and allows us to determine that S^{n-1} is Clarke regular and that the tangent cone admits the following formula

$$T_{S^{n-1}}(\bar{v}) = \{w \in \mathbb{R}^n \mid \nabla F(\bar{v})w = 0\} = \left\{w \in \mathbb{R}^n \mid \sum_{i=1}^{n} w_i \bar{v}_i = 0\right\}$$

Knowing that both of the sets in the Cartesian product are Clarke regular and closed, we use [10, Proposition 6.41] to write the tangent cone as a product of tangent cones

$$T_S(\bar{x}) = T_{\{0\}}(0) \times T_{S^{n-1}}(\bar{v}) = \{0\} \times \left\{w \in \mathbb{R}^n \mid \sum_{i=1}^{n} w_i \bar{v}_i = 0\right\}$$

□

5.2.1 Euclidean Norm in $\mathbb{R}^2$

Building off of the first example, we consider the Euclidean norm in a slightly higher dimension. Using our intuition from the one-dimensional example, we recall that the points of difficulty were on the boundary of the subset $\{0\} \times \operatorname{cl} B_1(0)$ so we consider some test points along this boundary. Suppose $\bar{x} = (0, (1, 0))$ (note that in the first coordinate 0 denotes the 0 vector in $\mathbb{R}^2$). In order to compute the directional normal cone we must first understand the tangent cone.

Theorem 5.8. Let $f = \|\cdot\|_2$ be the Euclidean norm on $\mathbb{R}^2$. Then the tangent cone to $\operatorname{gph} \partial f$ at $\bar{x} = (0, (1, 0))$ satisfies

$$T_{\operatorname{gph} \partial f}(\bar{x}) \supseteq \{(0, (v_1, v_2)) \mid v_1 \leq 0\} \cup \{((\lambda, 0), 0) \mid \lambda \geq 0\}$$

Proof. First we show that the set

$$T_1 := \left\{(0, (\cos\theta, \sin\theta)) \mid \theta \in \left(\frac{\pi}{2}, \frac{3\pi}{2}\right)\right\}$$

is contained in the given tangent cone. Given a vector $w = (0, (\cos\theta, \sin\theta))$ where θ lies in the appropriate interval, we construct a sequence of vectors $\{x_k\}$ converging in $\operatorname{gph} \partial f$ to $\bar{x}$ and a sequence $t_k \downarrow 0$ for which the quotient $(x_k - \bar{x})/t_k$ converges to w. Let $\{t_k\}$ be any nonnegative sequence converging to 0. Define

$$\tilde{x}_k = (0, (t_k \cos\theta + 1, t_k \sin\theta))$$

It is clear that as $k \to \infty$ that $\tilde{x}_k \to \bar{x}$ and that $(\tilde{x}_k - \bar{x})/t_k \to w$. We have to make some modifications to ensure that this sequence is contained in $\operatorname{gph} \partial f$. Since the first coordinate is zero we want the second coordinate to have norm less than or equal to 1.

$$\sqrt{(t_k \cos\theta + 1)^2 + (t_k \sin\theta)^2} = \sqrt{t_k^2 + 2t_k \cos\theta + 1} \leq 1 \iff t_k \leq -2\cos\theta$$

Since θ belongs to the open interval $\left(\frac{\pi}{2}, \frac{3\pi}{2}\right)$ it holds that $-2\cos\theta > 0$ so this inequality will hold for k sufficiently large because $t_k \downarrow 0$. From here we can modify the sequence $\{t_k\}$ to consist of only those points for which the inequality holds (say $\{t_n\} := \{t_{k_n}\}$), and define the sequence $\{x_n\}$ as $\{\tilde{x}_{k_n}\}$ to obtain two sequences with all of the desired properties. Thus $w \in T_{\mathrm{gph}\,\partial f}(\bar{x})$ and so $T_1 \subseteq T_{\mathrm{gph}\,\partial f}(\bar{x})$. To deal with the extreme angles $\frac{\pi}{2}, \frac{3\pi}{2}$, consider the sequences

$$x_k = \left(0, \left(\cos\frac{\pi}{2^k}, \sin\frac{\pi}{2^k}\right)\right),\ t_k = \frac{\pi}{2^k}$$

The sequence $\{x_k\}$ converges in $\mathrm{gph}\,\partial f$ to $\bar{x}$ and the relevant difference quotient evaluates to

$$\lim_{k\to\infty} \frac{\left(0, \left(\cos\frac{\pi}{2^k} - 1, \sin\frac{\pi}{2^k}\right)\right)}{\frac{\pi}{2^k}} = (0, (0, 1))$$

Similarly we find that $(0, (0, -1))$ belongs to the tangent cone by taking the same sequence but negating the sine coordinate. Because the tangent cone is a cone we obtain

$$\left\{(0, (\lambda\cos\theta, \lambda\sin\theta)) \mid \theta \in \left[\frac{\pi}{2}, \frac{3\pi}{2}\right], \lambda \geq 0\right\} \subseteq T_{\mathrm{gph}\,\partial f}(\bar{x})$$

Noting that this set is precisely

$$\{(0, (v_1, v_2)) \mid v_1 \leq 0\}$$

we have proven part of the first inclusion.

Next, notice that $((1, 0), 0) \in T_{\mathrm{gph}\,\partial f}(\bar{x})$ by taking the sequence

$$\mathrm{gph}\,\partial f \ni x_k = \left(t_k(1, 0), \frac{t_k(1, 0)}{\|t_k(1, 0)\|}\right) = ((t_k, 0), (1, 0)) \to \bar{x}$$

for any sequence $\{t_k\} \downarrow 0$. The quotient $(x_k - \bar{x})/t_k = ((1, 0), 0)$ so we are done.

□

To show semismoothness* from the definition we need a full characterization of the tangent cone. While trying to show that the converse inclusion holds in Theorem 5.8 an error was discovered in the case analysis that could not be resolved in a timely manner. In fact, the tangent cone is larger so it remains the subject of future work to provide a a more complete formula. However, we present here the general sketch of the work done so far and illustrate how it might be completed.

5.2.2 Roadmap for completing Theorem 5.8

If we suppose $w \in T_{\mathrm{gph}\,\partial f}(\bar{x})$ then there exist sequences $\{x_k\}$, $\{t_k\}$ as prescribed by the tangent cone definition. The sequence $\{x_k\}$ converges within $\mathrm{gph}\,\partial f$ so we consider two cases.

Case 1: Infinitely many terms of $\{x_k\}$ belong to $B = \{0\} \times \mathrm{cl}\,B_1(0)$. In this case we may pass to a subsequence $\{x_{k_n}\}$ that lies in B. We take the sequence $\{t_{k_n}\}$ accordingly. The original sequence $(x_k - \bar{x})/t_k$ is convergent so we may pass to this subsequence and still converge to w. The points x_{k_n} are of the form

$$x_{k_n} = (0, (x^{(1)}_{k_n}, x^{(2)}_{k_n}))$$

which implies

$$w = \left(0, \lim_{n\to\infty}\left(\frac{x_{k_n}^{(1)}-1}{t_{k_n}}, \frac{x_{k_n}^{(2)}}{t_{k_n}}\right)\right)$$

The second coordinate in the inner limit will be finite because the sequence is convergent, and the first coordinate satisfies nonpositivity because $(x_{k_n}^{(1)}, x_{k_n}^{(2)}) \in \operatorname{cl} B_1(0)$ so $x_{k_n}^{(1)} \leq 1$. This property holds in the limit so w belongs to the first set in the union formula provided.

Case 2: Infinitely many terms of $\{x_k\}$ belong to the set $V = \left\{\left(x, \frac{x}{\|x\|}\right) \mid x \in \mathbb{R}^2 \setminus \{0\}\right\}$. As before we pass to the appropriate subsequences $\{x_{k_n}\}$, $\{t_{k_n}\}$. The points x_{k_n} have the form

$$x_{k_n} = \left((a_n, b_n), \frac{(a_n, b_n)}{\sqrt{a_n^2+b_n^2}}\right)$$

We know that $x_{k_n} \to 0$ so $a_n, b_n \to 0$ and also at least one of a_n or b_n is always nonzero. Examining the vector w we find

$$w = \lim_{n\to\infty}\left(\frac{(a_n, b_n)}{t_{k_n}}, \left(\frac{a_n - \sqrt{a_n^2+b_n^2}}{t_{k_n}\sqrt{a_n^2+b_n^2}}, \frac{b_n}{t_{k_n}\sqrt{a_n^2+b_n^2}}\right)\right)$$

We have supposed that the sequence $(x_{k_n} - \bar{x})/t_{k_n}$ converges to $w \in \mathbb{R}^4$ so we know that $(a_n, b_n)/t_{k_n}$ converges coordinate-wise to a vector in $\mathbb{R}^2$.

From here there are a few possibilities. If $a_n \sim t_{k_n} \sim b_n$, we consider the sequence in the last coordinate of w. The proportionality of b_n to t_{k_n} and the fact that $a_n, b_n \to 0$ imply that this coordinate will blow up to $\pm\infty$ unless b_n is eventually 0. We deduce that b_n is 0 for n sufficiently large. Then the formula reduces to

$$w = \lim_{n\to\infty}\left(\frac{(a_n, 0)}{t_{k_n}}, \left(\frac{a_n - |a_n|}{t_{k_n}|a_n|}, 0\right)\right)$$

There is no issue of division by zero here because a_n and b_n cannot be zero simultaneously. If $a_n < 0$ then the formula in the third coordinate reduces to $-2/t_{k_n}$ which would diverge to $-\infty$ so this case is impossible. We conclude that $a_n > 0$ so

$$w = \lim_{n\to\infty}\left(\frac{(a_n, 0)}{t_{k_n}}, 0\right) = ((c, 0), 0)$$

for some $c > 0$ which proves that w belongs to the second set in the union formula provided.

Now, if $b_n = o(t_{k_n})$ and $a_n = o(t_{k_n})$ then the first two coordinates go to zero. It suffices to show that the third coordinate is nonpositive to conclude that this belongs to the first set in the given union formula. If $a_n \leq 0$ then $a_n - \sqrt{a_n^2+b_n^2} \leq 0$ as we are subtracting a nonnegative value from a nonpositive value. If $a_n > 0$ then the inequality also holds by squaring by both sides and subtracting a_n^2. Thus we are done with this possibility.

The cases that are unresolved are when $b_n = o(t_{k_n})$ and $a_n \sim t_{k_n}$, and vice versa. From these cases one can show that there are actually more elements in the tangent cone than are presented in the formula of Theorem 5.8.

As a consequence of the weakened version of Theorem 5.8 we can only obtain a partial result on the regular normal cone to the graph at this point.

Corollary 5.8.1. Let $f = \|\cdot\|_2$ be the Euclidean norm on $\mathbb{R}^2$. Then the normal cone to $\operatorname{gph}\partial f$ at $\bar{x} = (0,(1,0))$ is

$$\hat{N}_{\operatorname{gph}\partial f}(\bar{x}) \subseteq \{v = ((x,y),(z,0)) \mid x \leq 0,\ z \geq 0,\ y \in \mathbb{R}\}$$

Proof. If v belongs to the normal cone at $\bar{x}$ then $\langle v,\, y\rangle \leq 0$ for all $y \in T_{\operatorname{gph}\partial f}(\bar{x})$. If $v = ((x,t),(z,w))$ and $y = ((y_1,y_2),(y_3,y_4))$ then this is equivalent to $xy_1 + ty_2 + zy_3 + wy_4 \leq 0$.

Now we can substitute y for some of the vectors in the tangent cone. Letting $y = ((1,0),(0,0))$ implies $x \leq 0$. Taking $y = ((0,0),(-1,0))$ implies $z \geq 0$. Finally let $y = ((0,0),(0,\pm 1))$ to deduce that $w = 0$. This proves v belongs to the set on the right-hand side giving the desired inclusion.

□

6 Directional Differentiability

Before we conclude we briefly discuss directional differentiability. We have spent significant time focusing on issues of semismoothness* and now we provide a theorem relating the notions of semismoothness, semismoothness*, and directional differentiability, due to Gfrerer and Outrata [8, Corollary 3.8].

Proposition 6.1. Let $G\colon D \subseteq \mathbb{R} \to \mathbb{R}\cup\{+\infty\}$ be locally Lipschitz at $\bar{x} \in \operatorname{int} D$. Then the following are equivalent:

1. G is semismooth at $\bar{x}$
2. G is semismooth* and directionally differentiable at $\bar{x}$

We conclude from the proposition that to pass from semismoothness* of the prox operator to semismoothness we require directional differentiability.

Recall that the prox operator can be viewed as a function of $\lambda > 0$ and $\bar{x} \in \mathbb{E}$ simultaneously. In our applications, $\bar{x}$ remains fixed so we mainly concern ourselves with directional differentiability with respect to λ. Nonetheless, it is interesting to consider whether the prox operator is necessarily directionally differentiable as a function of $\bar{x}$. The answer is negative, and we will discuss a counterexample due to Shapiro [11].

6.1 Counterexample via Projection Operator

A specific case of the prox is when $f = \delta_C\colon \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is the indicator function for a nonempty closed convex set $C \subseteq \mathbb{R}^n$.

$$\delta_C(x) = \begin{cases} 0 & x \in C, \\ +\infty & x \notin C \end{cases}$$

The prox then amounts to the familiar Euclidean projection operator $P_C\colon \mathbb{R}^n \to \mathbb{R}^n$. In the counterexample we construct a set such that the projection operator onto this set is not directionally differentiable at a certain point.

To this end, let $\{x_n\}$ be the following sequence in $\mathbb{R}$: $x_n = \frac{\pi}{2^n}$ for all $n \in \mathbb{N}$. Note that this sequence is monotonically decreasing to 0. Define $C \subseteq \mathbb{R}^2$ to be the convex hull of the set $S = \{(\cos x_n, \sin x_n) \mid n \in \mathbb{N}\} \cup \{(0,0),(1,0)\}$. Notice that S is closed due to the addition of its only limit point, and it is clearly contained in $B_1(0)$ so it is bounded. Thus S is compact, and it follows from Carathéodory's theorem that the convex hull of S is compact. In particular, C is closed, nonempty, and convex.

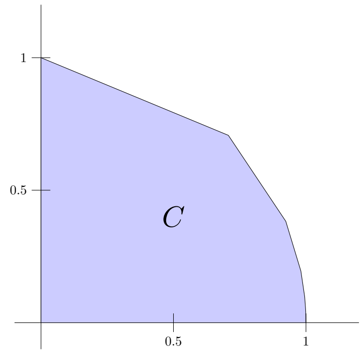

Figure 2: The set C we are projecting on.

Define two sequences in $\mathbb{R}$ as follows:

$$s_n = \sin x_n + (2 - \cos x_n)\tan\left(\frac{x_n + x_{n+1}}{2}\right)$$

$$t_n = \sin x_n + (2 - \cos x_n)\tan\left(\frac{x_n + x_{n-1}}{2}\right)$$

Observe that

$$\begin{aligned}
\lim_{n\to\infty} \frac{\sin x_n}{s_n} &= \lim_{n\to\infty} \frac{\sin\frac{\pi}{2^n}}{\sin\frac{\pi}{2^n} + (2-\cos\frac{\pi}{2^n})\tan\left(\frac{\frac{\pi}{2^n}+\frac{\pi}{2^{n+1}}}{2}\right)} \\
&= \lim_{n\to\infty} \frac{\sin\frac{\pi}{2^n}}{\sin\frac{\pi}{2^n} + (2-\cos\frac{\pi}{2^n})\tan\left(\frac{1}{2}\frac{\pi}{2^n}(1+\frac{1}{2})\right)} \\
&= \lim_{x\downarrow 0} \frac{\sin x}{\sin x + (2-\cos x)\tan\left(\frac{x}{2}(1+\frac{1}{2})\right)} \\
&= \lim_{x\downarrow 0} \frac{\cos x}{\cos x + \sin x\tan\left(\frac{x}{2}(1+\frac{1}{2})\right) + (2-\cos x)\frac{1}{2}(1+\frac{1}{2})\sec^2\left(\frac{x}{2}(1+\frac{1}{2})\right)} \\
&= \frac{2}{3+\frac{1}{2}}
\end{aligned}$$

Here the fourth equality follows from L'Hospital's rule. Using similar manipulations, we also obtain that $\lim_{n\to\infty}\frac{\sin x_n}{t_n} = \frac{2}{3+2}$. In particular, the values of these limits are distinct. Let $\bar{x} = (2,0)$ and $d = (0,1)$. It is easy to check that $P_C(\bar{x}) = (1, 0)$, and using geometric considerations one can show that $P_C(\bar{x}+t_n d) = P_C(\bar{x}+s_n d) = (\cos x_n, \sin x_n)$ for all $n \geq 2$. If we let $P_{C,2}$ denote the second coordinate of the projection onto C, it follows that

$$\begin{aligned}
\lim_{n\to\infty} \frac{P_{C,2}(\bar{x}+s_n d) - P_{C,2}(\bar{x})}{s_n} &= \lim_{n\to\infty} \frac{\sin x_n}{s_n} \\
&\neq \lim_{n\to\infty} \frac{\sin x_n}{t_n} \\
&= \lim_{n\to\infty} \frac{P_{C,2}(\bar{x}+t_n d) - P_{C,2}(\bar{x})}{t_n}
\end{aligned}$$

From this we conclude that the directional derivative at $\bar{x}$ in direction d does not exist. This shows that the projection operator, and hence the prox operator, need not be directionally differentiable as a function of $\bar{x} \in \mathbb{E}$.

6.2 Another Counterexample

For this second counterexample we make use of the following result on proximal operators due to Combettes and Pesquet [4].

Theorem 6.1. [4, Proposition 2.4] Let ρ be a function defined from $\mathbb{R}$ to $\mathbb{R}$. Then ρ is the proximal operator of a function in Γ_0 if and only if it is nonexpansive and increasing.

Nonexpansivity is just the Lipschitz condition with Lipschitz constant $L \leq 1$ so that $|\rho(x) - \rho(y)| \leq |x - y|$ for all $x, y \in \mathbb{R}$. Consider the function $\rho\colon \mathbb{R} \to \mathbb{R}$ defined by

$$\rho(x) = \begin{cases} \frac{1}{6}x\sin(-\log|x|) + \frac{1}{2}x & x \neq 0 \\ 0 & x = 0 \end{cases}$$

We analyze the derivative to determine the various properties of this function. Notice that for $x > 0$

$$\rho'(x) = \frac{1}{6}\left(\sin(-\log x) - \cos(-\log x)\right) + \frac{1}{2} \geq -\frac{\sqrt{2}}{6} + \frac{1}{2} > 0$$

and similarly for $x < 0$

$$\rho'(x) = \frac{1}{6}\left(\sin(-\log(-x)) - \cos(-\log(-x))\right) + \frac{1}{2} \geq -\frac{\sqrt{2}}{6} + \frac{1}{2} > 0$$

so this function is increasing on $\mathbb{R} \setminus \{0\}$ and the verification at 0 is straightforward.

To verify nonexpansivity on $\mathbb{R}_{++}$, we take the derivative and find that for $x > 0$

$$\rho'(x) = \frac{1}{6}\left(\sin(-\log x) - \cos(-\log x)\right) + \frac{1}{2} \leq \frac{\sqrt{2}}{6} + \frac{1}{2} \leq 1$$

As before, the same argument works for $x < 0$. It remains to show that it is nonexpansive around $x = 0$. We remark that if $y < 0 < x$ then

$$\begin{aligned}
|\rho(x) - \rho(y)| &= \Big|\frac{1}{6}x\sin(-\log x) + \frac{1}{2}x - \frac{1}{6}y\sin(-\log(-y)) - \frac{1}{2}y\Big| \\
&\leq \frac{1}{6}|x\sin(-\log x) - y\sin(-\log(-y))| + \frac{1}{2}|x-y| \\
&= \frac{1}{6}\left(|x(\sin(-\log x) - \sin(-\log(-y))) + (x-y)\sin(-\log(-y))|\right) + \frac{1}{2}|x-y| \\
&\leq \frac{1}{6}|x||\sin(-\log x) - \sin(-\log(-y))| + \frac{1}{6}|x-y||\sin(-\log(-y))| + \frac{1}{2}|x-y| \\
&\leq \frac{2}{6}|x| + \frac{1}{6}|x-y| + \frac{1}{2}|x-y| \\
&\leq \frac{2}{6}|x-y| + \frac{4}{6}|x-y| \\
&= |x-y|
\end{aligned}$$

as desired. The cases where one of x or y is zero are easier and symmetrical. By Theorem 6.1, we deduce that ρ is indeed the proximal operator for some function in Γ_0. Now we demonstrate that ρ is not directionally differentiable at 0 in any direction (left or right). We prove that it has no right derivative, and the case of left derivative follows by symmetry of the function.

We write
$$\lim_{h\downarrow 0}\frac{\rho(0+h)-\rho(0)}{h} = \lim_{h\downarrow 0}\frac{1}{6}\sin(-\log h) + \frac{1}{2}$$

If we consider the sequence $\{h_n\}$ defined by $h_n = e^{-\frac{\pi}{2}n}$ then $h_n \downarrow 0$ and

$$\frac{1}{6}\sin(-\log h_n) = \frac{1}{6}\sin(\frac{\pi}{2}n)$$

which oscillates so the limit on the righthand side above does not exist. Thus we have furnished another example of a prox operator that is not directionally differentiable as a function of its base point.

7 Final Remarks

We conclude our study having considered some examples and counterexamples related to our results on the proximal operator. Rectifying the Euclidean norm example will be the subject of future work, ultimately working towards a total classification of the semismoothness* (or nonsemismoothness*) properties of this function's (subdifferential) graph. The results discussed in this paper emphasize the implications of semismoothness* of the subdifferential but there is still work to be done on how necessary such conditions are.

We also hope to gain further insight on the directional differentiability of the prox as a function of $\lambda > 0$ while the parameter $\bar{x}$ is fixed. This is important for our epigraphical projection framework and we have already shown that there is no such regularity when λ is fixed and $\bar{x}$ varies.

References

[1] J.-P. Aubin and Hélène Frankowska. *Set-Valued Analysis.* Modern Birkhäuser Classics, Birkhäuser Basel, 2009.

[2] A. Beck. *First-Order Methods in Optimization.* MOS-SIAM Series on Optimization, 2017.

[3] M. Benko, H. Gfrerer, and J.V. Outrata. *Calculus for Directional Limiting Normal Cones and Subdifferentials.* Set-Valued Var. Anal 27, 713–745, 2019.

[4] P.L. Combettes and J.-C. Pesquet. *Proximal Thresholding Algorithm for Minimization over Orthonormal Bases.* SIAM Journal on Optimization, 18(4), 1351–1376, 2008.

[5] A.L. Dontchev and R.T. Rockafellar. *Implicit Functions and Solution Mappings. A View from Variational Analysis.* Springer Series in Operations Research and Financial Engineering, Springer-Verlag New York, 2014.

[6] M.P. Friedlander, A. Goodwin, and T. Hoheisel. *From perspective maps to epigraphical projections.* arXiv:2102.06809 [math.OC]. Submitted to Mathematics of Operations Research, February 2021.

[7] H. Gfrerer. *Optimality conditions for disjunctive programs based on generalized differentiation with application to mathematical programs with equilibrium constraints.* arXiv:1611.08257, 2016.

[8] H. Gfrerer and J.V. Outrata. *On a semismooth* Newton method for solving generalized equations.* arXiv:1904.09167, 2019.

[9] J.S. Pang and L. Qi. *A globally convergent Newton method for convex SC1 minimization problems.* Journal of Optimization Theory and Applications 85(3), 1995, pp. 633–648.

[10] R.T. Rockafellar and R.J.-B. Wets. *Variational Analysis.* Grundlehren der Mathematischen Wis- senschaften, Vol. 317, Springer-Verlag, Berlin, 1998.

[11] A. Shapiro. *Directionally nondifferentiable metric projection.* Journal of Optimization Theory and Applications 81(1), 1994, pp. 203–204.