

# Subgradient Projectors: Extensions, Theory, and Characterizations

Heinz H. Bauschke\*, Caifang Wang<sup>†</sup>, Xianfu Wang,<sup>‡</sup> and Jia Xu<sup>§</sup>

April 13, 2017

## Abstract

Subgradient projectors play an important role in optimization and for solving convex feasibility problems. For every locally Lipschitz function, we can define a subgradient projector via generalized subgradients even if the function is not convex. The paper consists of three parts. In the first part, we study basic properties of subgradient projectors and give characterizations when a subgradient projector is a cutter, a local cutter, or a quasi-nonexpansive mapping. We present global and local convergence analyses of subgradient projectors. Many examples are provided to illustrate the theory. In this second part, we investigate the relationship between the subgradient projector of a prox-regular function and the subgradient projector of its Moreau envelope. We also characterize when a mapping is the subgradient projector of a convex function. In the third part, we focus on linearity properties of subgradient projectors. We show that, under appropriate conditions, a linear operator is a subgradient projector of a convex function if and only if it is a convex combination of the identity operator and a projection operator onto a subspace. In general, neither a convex combination nor a composition of subgradient projectors of convex functions is a subgradient projector of a convex function.

**2010 Mathematics Subject Classification:** Primary 49J52; Secondary 49J53, 47H04, 47H05, 47H09.

**Keywords:** Approximately convex function, averaged mapping, cutter, essentially strictly differentiable function, fixed point, limiting subgradient, local cutter, local quasi-firmly nonexpansive mapping, local quasi-nonexpansive mapping, local Lipschitz function, linear cutter, linear firmly nonexpansive mapping, linear subgradient projection operator, Moreau envelope, projection, prox-bounded, proximal mapping, prox-regular function, quasi-firmly nonexpansive mapping, quasi-nonexpansive mapping,  $(C, \varepsilon)$ -firmly nonexpansive mapping, subdifferentiable function, subgradient projection operator.

## 1 Introduction

Studies of optimization problems and convex feasibility problems have led in recent years to the development of a theory of subgradient projectors, which is a projection to a certain half-space.

---

\*Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. E-mail: heinz.bauschke@ubc.ca.

<sup>†</sup>Department of Mathematics, Shanghai Maritime University, Shanghai, China. Email: cfwang@shmtu.edu.cn.

<sup>‡</sup>Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. E-mail: shawn.wang@ubc.ca.

<sup>§</sup>Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. Email: jia.xu@ubc.ca.

Rather than finding projections on level sets of original functions, the iterative algorithms find projections on half spaces which include the 0-level set of the function. Polyak developed subgradient projector iteration for convex functions [46, 47, 48], and they are further developed by Censor, Combettes, Fukushima, Kiwiel, Yamada and others, and applied to many kinds of optimization problems [22, 21, 9, 24, 25, 29, 34, 35, 44, 19, 55]. In [12], we give a systematic study for subgradient projectors of convex functions. Convexity is often a too strong assumption for the needs of applications. In a recent work [43], Pang studied finitely convergent algorithms for nonconvex inequality problems involving approximately convex functions. The subgradient projector by Pang used the Clarke subdifferential instead of the Mordukhovich limiting subdifferential. To this day, however, there is a lack of systematic theory on the subgradient projector when a function is possibly nonconvex.

The goal of this paper is to carry out the basic theory of subgradient projectors for possibly nonconvex functions on a finite dimensional space, which is thus aimed ultimately at applications to diverse problems of nonconvex optimization. Non-differentiable and nonconvex functions arise in many optimization problems. As far as nonconvex functions are concerned, the cutter theory or  $\mathcal{T}$ -class developed by Cegielski [20], Bauschke, Borwein and Combettes [8], and Bauschke, Wang, Wang and Xu [13] furnish the new approach to subgradient projectors, without appealing to the existence theory on subgradient projectors for convex functions. Our study shows that subgradient projectors for nonconvex functions have many attractive analytical properties. Among all results presented here, we discover that while cutters and quasi-nonexpansive mappings on  $\mathbb{R}^n$  are global, cutters and quasi-nonexpansive mappings on a neighborhood are more useful for functions which are locally convex around the desired point, say a critical point or a feasible point. This paper not only includes some results from [54], but also many refinements and new advances.

Since definitions and proofs are much simpler in the finite dimensional space, and many technical complications do not even appear, we shall work in the finite dimensional space only. For the convenience of readers, our main results are presented in three parts. In the first part, we study extensions and theory of subgradient projectors. In the second part, we consider subgradient projectors of Moreau envelopes and conditions under which a mapping is the subgradient projector of a convex function. The third part is devoted to linear subgradient projectors.

The remainder of this paper is organized as follows.

Part I consists of Sections 2–6. Section 2 provides an extension of subgradient projectors from convex functions to possibly nonconvex functions; Section 3 is devoted to calculus of subgradient projectors; Section 4 deals with whether one can recover a function from its subgradient projector, and fixed point closed property of a subgradient projector; conditions on functions under which their subgradient projectors are cutters or local cutters are presented in Section 5. Section 6 is devoted to convergence analysis of subgradient projectors by using theory from cutters, local cutters, quasinonexpansive mapping, and local quasinonexpansive mappings. Under appropriate assumptions, we show that subgradient projectors are  $(C, \varepsilon)$ -firmly nonexpansive, a very useful concept introduced by Hesse and Luke for studying local linear convergence of a variety of algorithms.

Part II consists of Sections 7–8. For prox-bounded and prox-regular functions, their Moreau envelopes are differentiable. Section 7 studies the subgradient projectors of Moreau envelopes of prox-bounded and prox-regular functions, and their connections to subgradient projectors of original functions. We show that if  $f$  is proper, lsc, prox-bounded, and prox-regular on  $\mathbb{R}^n$ , then  $f$  is a difference of convex functions (Corollary 7.9); and that if  $f$  is  $C^2$ ,  $\min f = 0$ , and  $\nabla f(x) \neq 0$  for every  $x \in \mathbb{R}^n \setminus \operatorname{argmin} f$ , then the subgradient projector  $G_f$  of  $f$  is a cutter if and only if the

subgradient projector  $G_{e_\lambda f}$  of the envelope  $e_\lambda f$  is a cutter for every  $\lambda > 0$  (Propositions 7.22 and 7.24). Section 8 characterizes when a mapping is actually a subgradient projector of a convex function.

Part III consists of Section 9–11. It is interesting to ask when a subgradient projector is linear, and what special properties a linear subgradient projector possesses. To the best of our knowledge, this question has not been explored in the literature. Section 9 studies linear subgradient projectors and their distinguished features. In particular, we give a nonlinear cutter which is nonexpansive but not firmly nonexpansive, and the example is much simpler than the one given by Cegielski [20]. In Section 10, using results from Section 9, we show that in general neither a convex combination nor a composition of subgradient projectors of convex functions is a subgradient projector of a convex function. Finally, in Section 11, we completely characterize linear subgradient projectors on  $\mathbb{R}^2$ , and give explicit formulae for the corresponding functions.

The notation that we employ is for the most part standard; however, a partial list is provided for the reader's convenience. Throughout this paper,  $\mathbb{R}^n$  is the  $n$ -dimensional Euclidean space with inner product  $\langle \cdot, \cdot \rangle$  and induced norm  $\| \cdot \|$ , i.e.,  $(\forall x \in \mathbb{R}^n) \|x\| := \sqrt{\langle x, x \rangle}$ . The identity operator on  $\mathbb{R}^n$  is  $\text{Id}$ . For a mapping  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , its fixed point set is denoted by  $\text{Fix } T := \{x \in \mathbb{R}^n \mid Tx = x\}$ ; its kernel is  $\ker T := \{x \in \mathbb{R}^n \mid Tx = 0\}$ ; its range is  $\text{ran } T := \{y \in \mathbb{R}^n \mid y = Tx \text{ for some } x \in \mathbb{R}^n\}$ .

For a function  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ , its  $\alpha$ -level set is denoted by  $\text{lev}_\alpha f := \{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}$ ; its effective domain is  $\text{dom } f := \{x \in \mathbb{R}^n \mid f(x) < +\infty\}$ . For a set-valued mapping  $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ , the domain, range and fixed point set of  $F$  are given by  $\text{dom } F := \{x \in \mathbb{R}^n \mid F(x) \neq \emptyset\}$ ,  $\text{ran } F := \cup_{x \in \mathbb{R}^n} F(x)$ , and  $\text{Fix } F := \{x \in \mathbb{R}^n \mid x \in F(x)\}$  respectively. We use  $\mathbf{B}(x, \delta)$  for the closed ball centered at  $x \in \mathbb{R}^n$  with radius  $\delta > 0$ .  $\mathbb{R}_+$  denotes the set of non-negative real numbers, and  $\mathbb{N}$  denotes the set of non-negative integers  $\{0, 1, 2, \dots\}$ .

For a set  $C \subseteq \mathbb{R}^n$ , its distance function is

$$d_C : \mathbb{R}^n \rightarrow [0, +\infty) : x \mapsto \inf \{\|x - y\| \mid y \in C\},$$

and the projection operator onto  $C$  is

$$P_C : \mathbb{R}^n \rightrightarrows C : x \mapsto \{p \in C \mid \|x - p\| = d_C(x)\}.$$

The indicator function of  $C$  is  $\iota_C : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  is defined by  $\iota_C(x) := 0$  if  $x \in C$  and  $\iota_C(x) := +\infty$  if  $x \notin C$ . We write  $\text{int } C$  for the interior, and  $\text{bdry}(C) := C \setminus \text{int } C$  for the boundary of  $C$ , respectively. For a subspace  $L \subseteq \mathbb{R}^n$ , its orthogonal complement is defined to be

$$L^\perp := \{y \in \mathbb{R}^n \mid \langle y, x \rangle = 0, x \in L\}.$$

When  $x, y \in \mathbb{R}^n$ , the line segment between  $x, y$  is given by  $[x, y] := \{(1 - \lambda)x + \lambda y \mid 0 \leq \lambda \leq 1\}$ .

## Part I

# Extensions to possibly nonconvex functions and basic theory

## 2 An extension of subgradient projector via limiting subgradients

To introduce subgradient projectors for possible non-convex functions, we need the following generalized subgradients [51, 40, 39, 31].

**Definition 2.1** Consider a function  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  and a point  $\bar{x} \in \mathbb{R}^n$  with  $f(\bar{x})$  finite. For a vector  $v \in \mathbb{R}^n$ , one says that

(i)  $v$  is a regular (or Fréchet) subgradient of  $f$  at  $\bar{x}$ , written  $v \in \hat{\partial}f(\bar{x})$ , if

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

(ii)  $v$  is a limiting (or Mordukhovich) subgradient of  $f$  at  $\bar{x}$ , written  $v \in \partial f(\bar{x})$ , if there are sequences  $x_\nu \rightarrow \bar{x}$ ,  $f(x_\nu) \rightarrow f(\bar{x})$  and  $v_\nu \in \hat{\partial}f(x_\nu)$  with  $v_\nu \rightarrow v$ .

A locally Lipschitz function is subdifferentially regular at  $\bar{x}$  with  $f(\bar{x})$  finite if  $\partial f(\bar{x}) = \hat{\partial}f(\bar{x})$ , see [51, Corollary 8.11], [39]. It is well-known that when  $f$  is locally Lipschitz,  $\partial f$  is nonempty-valued everywhere; when  $f$  is lower semicontinuous (lsc), the set of points at which  $\partial f$  is nonempty-valued is at least dense in the domain of  $f$ , [51, Corollary 8.10]. Furthermore,  $\partial f$  is the usual Fenchel subdifferential when  $f$  is convex. All of these results can be found in [18, 17, 40, 51].

A function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is called subdifferentiable if  $\partial f(x) \neq \emptyset$  for every  $x \in \mathbb{R}^n$ . While every locally Lipschitz functions on  $\mathbb{R}^n$  is a subdifferentiable function, a subdifferentiable function might not be locally Lipschitz, e.g.,

$$f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} 1 & \text{if } x \leq 0, \\ 1 - \sqrt{x} & \text{if } x > 0, \end{cases}$$

see also [51, page 359]. The key concept we shall study is the subgradient projection operator.

**Definition 2.2** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be lsc and subdifferentiable, and let  $s : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a selection of  $\partial f$ . The subgradient projector of  $f$  is defined by

$$(1) \quad G_{f,s} : \mathbb{R}^n \rightarrow \mathbb{R}^n : x \mapsto \begin{cases} x - \frac{f(x)}{\|s(x)\|^2} s(x) & \text{if } f(x) > 0 \text{ and } 0 \notin \partial f(x), \\ x & \text{otherwise.} \end{cases}$$

When it is not necessary to emphasize the selection  $s$ , we will write  $G_f$ . It is also convenient to introduce the set-valued mapping associated with  $f$  by

$$(2) \quad \mathcal{G}_f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n : x \mapsto \{G_{f,s}(x) \mid s \text{ is a selection of } \partial f\}$$

with  $G_{f,s}$  being given in (1).

Although subgradient projectors have been well studied for convex functions [12, 21, 24, 20, 44, 46, 48, 55, 42], the extension to possibly nonconvex functions is new.

When  $f$  is convex and  $\inf_{\mathbb{R}^n} f \leq 0$ ,  $G_{f,s}$  reduces to

$$G_{f,s} : \mathbb{R}^n \rightarrow \mathbb{R}^n : x \mapsto \begin{cases} x - \frac{f(x)}{\|s(x)\|^2} s(x) & \text{if } f(x) > 0, \\ x & \text{otherwise,} \end{cases}$$

where  $s : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a selection of  $\partial f$  with  $s(x) \in \partial f(x)$ . When  $f$  is continuously differentiable on  $\mathbb{R}^n \setminus \text{lev}_0 f$ ,  $G_f$  reduces to

$$G_f : \mathbb{R}^n \rightarrow \mathbb{R}^n : x \mapsto \begin{cases} x - \frac{f(x)}{\|\nabla f(x)\|^2} \nabla f(x) & \text{if } f(x) > 0 \text{ and } \nabla f(x) \neq 0, \\ x & \text{otherwise.} \end{cases}$$

The geometric interpretation and motivation of the subgradient projector come from the following:

**Proposition 2.3** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be lsc and subdifferentiable, and let  $s$  be a selection of  $\partial f$ .*

(i) *Whenever  $f(x) > 0$ ,  $0 \notin \partial f(x)$ , we have*

$$G_{f,s}(x) = P_{H_-(s(x),x)}(x)$$

*where the half space*

$$(3) \quad H_-(s(x), x) := \{z \in \mathbb{R}^n \mid f(x) + \langle s(x), z - x \rangle \leq 0\}.$$

(ii) *The fixed point set of  $G_{f,s}$  is*

$$\text{Fix } G_{f,s} = \{x \in \mathbb{R}^n \mid 0 \in \partial f(x)\} \cup \text{lev}_0 f = \text{Fix } \mathcal{G}_f.$$

*If  $f$  is locally Lipschitz, then  $\text{Fix } G_{f,s}$  is closed.*

(iii) *If  $f$  is convex and  $\inf_{\mathbb{R}^n} f \leq 0$ , then  $\text{Fix } G_{f,s} = \text{lev}_0 f$ .*

*Proof.* (i). According to [7] or [20, page 133], for the half space

$$H_-(a, \beta) := \{z \in \mathbb{R}^n \mid \langle a, z \rangle \leq \beta\},$$

where  $a \in \mathbb{R}^n$ ,  $a \neq 0$  and  $\beta \in \mathbb{R}$ , its metric projection is given by

$$(4) \quad P_{H_-(a,\beta)} x = \begin{cases} x - \frac{\langle a, x \rangle - \beta}{\|a\|^2} a & \text{if } \langle a, x \rangle > \beta, \\ x & \text{if } \langle a, x \rangle \leq \beta. \end{cases}$$

Apply (4) with  $a := s(x)$ ,  $\beta := \beta(x) = \langle s(x), x \rangle - f(x)$ .

(ii). This follows from the definition of  $G_f$ . When  $f$  is locally Lipschitz,  $\partial f$  is upper-semicontinuous [51, Proposition 8.7], so

$$\{x \in \mathbb{R}^n \mid 0 \in \partial f(x)\}$$

is closed. Being a union of two closed sets,  $\text{Fix } G$  is closed.

(iii). When  $f$  is convex,  $0 \in \partial f(x)$  gives  $f(x) = \min_X f$ , so  $f(x) \leq 0$ . Then

$$\{x \in \mathbb{R}^n \mid 0 \in \partial f(x)\} \subseteq \text{lev}_0 f.$$

Thus (iii) follows from (ii). ■

**Remark 2.4** (i) Proposition 2.3(i) uses the Euclidean distance. Following [36, 8, 11, 10], one may define Bregman subgradient projectors for lsc and subdifferentiable functions. This will be explored in future work.

(ii) Proposition 2.3(ii) shows that for subdifferential functions, the fixed point of  $G_f$  gives  $x \in \mathbb{R}^n$  such that

$$0 \in \partial f(x) \text{ or } f(x) \leq 0.$$

Proposition 2.3(iii) shows that for convex functions, the fixed point of  $G_f$  gives  $x \in \mathbb{R}^n$  such that

$$f(x) \leq 0.$$

We give two simple examples to illustrate the difference of subgradient projectors between convex and nonconvex functions.

**Example 2.5** Consider  $f : \mathbb{R}^n \rightarrow \mathbb{R} : x \mapsto \sqrt[k]{\|x\|} = \|x\|^{1/k}$  where  $k > 0$ . Then

(i) When  $k \leq 1$ ,  $f$  is convex,  $G_f = (1 - k) \text{Id}$  is firmly nonexpansive.

(ii) When  $k > 1$ ,  $f$  is not convex,  $G_f = (1 - k) \text{Id}$  is not monotone and need not be nonexpansive, e.g,  $k = 3$ .

*Proof.* For  $x \neq 0$ , we have

$$\nabla f(x) = \frac{1}{k} \|x\|^{1/k-1} \frac{x}{\|x\|},$$

and the result follows from the definition of  $G_f$ . ■

Let  $\mathbf{B}$  denote the closed unit ball of  $\mathbb{R}^n$ . According to [51, Exercise 8.14], for a nonempty  $C \subseteq \mathbb{R}^n$  the normal cone and regular normal cone mapping are respectively defined  $N_C := \partial \iota_C$  and  $\hat{N}_C := \hat{\partial} \iota_C$ . Recall

**Fact 2.6** ([51, Example 8.53]) For  $f := d_C$  in the case of a closed set  $C \neq \emptyset$  in  $\mathbb{R}^n$ , one has at any point  $\bar{x} \in C$  that

$$\partial f(\bar{x}) = N_C(\bar{x}) \cap \mathbf{B}, \quad \hat{\partial} f(\bar{x}) = \hat{N}_C(\bar{x}) \cap \mathbf{B}.$$

On the other hand, for any  $\bar{x} \notin C$ , one has

$$\partial f(\bar{x}) = \frac{\bar{x} - P_C(\bar{x})}{d_C(\bar{x})}, \quad \hat{\partial} f(\bar{x}) = \begin{cases} \left\{ \frac{\bar{x} - \bar{x}}{d_C(\bar{x})} \right\} & \text{if } P_C(\bar{x}) = \{\bar{x}\}, \\ \emptyset & \text{otherwise.} \end{cases}$$

**Example 2.7 (subgradient projectors of distance functions)** Let  $C \neq \emptyset$  be a closed set in  $\mathbb{R}^n$ . Then  $\mathcal{G}_{d_C} = P_C$ .

*Proof.* Let  $s$  be a selection of  $\partial d_C$ . By Fact 2.6,

$$(\forall x \notin C) s(x) = \frac{x - p(x)}{d_C(x)}$$

where  $p(x) \in P_C(x)$ . We show  $G_{d_C, s} = p$ .

When  $x \in C$ , we have  $P_C(x) = \{x\}$  and  $p(x) = x$ . Because  $d_C(x) = 0$  for  $x \in C$ , the definition of  $G_{d_C, s}$  gives  $G_{d_C, s}(x) = x$ . Thus  $G_{d_C, s}(x) = x = p(x)$  for  $x \in C$ .

When  $x \notin C$ ,  $d_C(x) > 0$  and  $0 \notin \partial d_C(x)$  because every  $x^* \in \partial d_C(x)$  has  $\|x^*\| = 1$  by Fact 2.6. Then for  $x \notin C$ ,

$$G_{d_C,s}(x) = x - d_C(x) \frac{x - p(x)}{d_C(x)} = p(x).$$

Altogether,  $G_{d_C,s}(x) = p(x)$  for every  $x \in \mathbb{R}^n$ . ■

When  $C$  is nonempty, closed and convex, the projection mapping  $P_C$  is single-valued, and  $d_C$  is continuously differentiable on  $\mathbb{R}^n \setminus C$ . Example 2.7 implies:

**Fact 2.8** ([11], [25]) *Let  $C \subseteq \mathbb{R}^n$  be nonempty, closed and convex. Then  $\mathcal{G}_{d_C} = P_C$ .*

What happens if we take the subgradient projector of a distance function to a set where the distance is taken with respect to another norm? The following example illustrates that using the Euclidean norm for  $d_C$  in Example 2.7 is essential.

**Example 2.9** Define

$$f : \mathbb{R}^2 \rightarrow \mathbb{R} : (x_1, x_2) \mapsto |x_1| + |x_2|,$$

the distance function to  $C := \{(0,0)\}$  in  $\ell_1$ -norm. When  $x_1 > 0, x_2 > 0, x_1 \neq x_2$ , we have  $G_f(x_1, x_2) = ((x_1 - x_2)/2, (x_2 - x_1)/2) \neq (0,0) = P_C(x_1, x_2)$ . Even using  $\|\cdot\|_\infty$  the dual norm of  $\|\cdot\|_1$  for  $s(x_1, x_2)$ , we have  $G_f(x_1, x_2) = (-x_2, -x_1) \neq (0,0) = P_C(x_1, x_2)$ .

**Remark 2.10** Example 2.7 might lead the reader to believe that  $\mathcal{G}_f$  is a monotone operator. This holds for any twice differentiable convex function  $f : \mathbb{R} \rightarrow \mathbb{R}$ ; see [12, Proposition 8.2]. However, this fails for  $f : \mathbb{R}^2 \rightarrow \mathbb{R} : (x_1, x_2) \mapsto |x_1|^p + |x_2|^p$  when  $1 < p < 2$ ; see [12, Proposition 10.1(iii)].

The following example shows that the assumption that  $f$  being subdifferentiable is important in Definition 2.2.

**Example 2.11** The function defined by

$$f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} 1/|x| & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

has  $G_f = 2 \text{Id}$  on  $\mathbb{R}$  so that  $\text{Fix } G_f = \text{Fix } 2 \text{Id} = \{0\}$ . However, the function defined by

$$g : \mathbb{R} \rightarrow (-\infty, +\infty] : x \mapsto \begin{cases} 1/|x| & \text{if } x \neq 0, \\ +\infty & \text{if } x = 0, \end{cases}$$

has  $G_g = 2 \text{Id}$  on  $\mathbb{R} \setminus \{0\}$ . Because  $G_g$  is not defined at  $x = 0$ , we have  $\text{Fix } G_g = \emptyset$  but  $\text{Fix } 2 \text{Id} = \{0\}$ .

### 3 Calculus for subgradient projectors

In this section we obtain calculus results for subgradient projectors defined in Section 2 related to representations of subgradient projectors for max functions, compositions of functions with a linear operator, and positive powers of nonnegative functions. Subdifferential calculus is the main tool for proving these results.

A mapping  $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^k$  is called strictly differentiable at  $\bar{x}$  if the Fréchet derivative  $\Phi'(\bar{x})$  exists and

$$\lim_{\substack{x, y \rightarrow \bar{x} \\ y \neq x}} \frac{\|\Phi(x) - \Phi(y) - \Phi'(\bar{x})(x - y)\|}{\|x - y\|} = 0.$$

The following facts on subdifferentials are crucial to study the calculus of subgradient projectors.

**Fact 3.1** ([39, Theorem 6.5], [40, Theorem 1.110(ii)]) *Assume that  $F : \mathbb{R}^n \rightarrow \mathbb{R}^k$  is locally Lipschitz at  $\bar{x} \in \mathbb{R}^n$ , and  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  is strictly differentiable at  $F(\bar{x})$ . Then for  $f(x) = g(F(x))$ , one has*

$$\partial f(\bar{x}) = \partial \langle g'(F(\bar{x})), F \rangle(\bar{x}) \text{ with } \bar{y} = F(\bar{x}).$$

For a matrix  $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , let  $A^\top$  denote its transpose.

**Fact 3.2** ([39, Theorem 6.7(i)], [40, Proposition 1.112(i)], or [51, Exercise 10.7]) *Let  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be strictly differentiable at  $\bar{x}$  with  $F'(\bar{x})$  being invertible, and suppose that  $f(x) = g(F(x))$  with  $g : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  being lsc around  $\bar{y} = F(\bar{x})$  and  $f$  being finite at  $\bar{x}$ . Then*

$$\partial f(\bar{x}) = (F'(\bar{x}))^\top \partial g(\bar{y}) \text{ with } \bar{y} = F(\bar{x}).$$

**Fact 3.3** ([39, Theorem 7.5(ii)], [40, Theorem 3.46(ii)]) *Let  $f_1, f_2 : \mathbb{R}^n \rightarrow \mathbb{R}$  be locally Lipschitz at  $\bar{x}$  and  $J(\bar{x}) := \{j \mid f_j(\bar{x}) = \max\{f_1, f_2\}(\bar{x})\}$ . Then*

$$\partial \max\{f_1, f_2\}(\bar{x}) \subseteq \text{conv} \bigcup \{\partial f_j(\bar{x}) \mid j \in J(\bar{x})\},$$

where the equality holds and  $\max\{f_1, f_2\}$  is subdifferentially regular at  $\bar{x}$  if the function  $f_j$  is subdifferentially regular at  $\bar{x}$  for  $j \in J(\bar{x})$ .

**Proposition 3.4** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be lsc and subdifferentiable.*

(i) *If  $k > 0$  then  $\mathcal{G}_{kf} = \mathcal{G}_f$ .*

(ii) *Let  $\alpha \in \mathbb{R}$  and  $s$  be a selection of  $\partial f$ . Define*

$$G_{f,\alpha} : \mathbb{R}^n \rightarrow \mathbb{R}^n : x \mapsto \begin{cases} x - \frac{f(x) - \alpha}{\|s(x)\|^2} s(x) & \text{if } f(x) > \alpha \text{ and } 0 \notin \partial f(x), \\ x & \text{otherwise.} \end{cases}$$

*Then  $G_{f,\alpha} = G_{f-\alpha,s}$ .*

(iii) *Let  $\alpha > 0$  and  $s$  be a selection of  $\partial f$ . Then*

$$G_{f-\alpha,s}(x) = \begin{cases} G_{f,s}(x) + \frac{\alpha s(x)}{\|s(x)\|^2} & \text{if } f(x) > \alpha \text{ and } 0 \notin \partial f(x), \\ x & \text{otherwise.} \end{cases}$$

*Proof.* (i). By Fact 3.1,  $\partial(kf) = k\partial f$ . Note that  $kf(x) > 0$  if and only if  $f(x) > 0$ , and  $0 \notin \partial(kf)(x)$  if and only if  $0 \notin \partial f(x)$ . When  $kf(x) > 0$  and  $0 \notin \partial(kf)(x)$ , for  $s(x) \in \partial f(x)$ , we have  $ks(x) \in \partial(kf)(x)$  so that

$$G_{kf,ks}(x) = x - \frac{kf(x)}{\|ks(x)\|^2} ks(x) = x - \frac{f(x)}{\|s(x)\|^2} s(x) = G_{f,s}(x).$$



When  $kf(x) \leq 0$  or  $0 \in \partial(kf)(x)$ , we have  $f(x) \leq 0$  or  $0 \in \partial f(x)$ , so  $G_{kf,ks}(x) = x = G_{f,s}(x)$ .

(ii). It suffices to note that  $\partial(f - \alpha) = \partial f$ .

(iii). When  $f(x) > \alpha$  and  $0 \notin \partial f(x)$ , we have  $f(x) > 0$  and  $0 \notin \partial f(x)$ . Then

$$G_{f-\alpha,s}(x) = x - \frac{f(x) - \alpha}{\|s(x)\|^2} s(x) = x - \frac{f(x)}{\|s(x)\|^2} s(x) + \frac{\alpha}{\|s(x)\|^2} s(x) = G_{f,s}(x) + \frac{\alpha}{\|s(x)\|^2} s(x).$$

When  $f(x) \leq \alpha$  or  $0 \in \partial f(x)$ ,  $G_{f-\alpha,s}(x) = x$  by the definition. ■

**Proposition 3.5** Assume that  $f_1, f_2 : \mathbb{R}^n \rightarrow \mathbb{R}$  are locally Lipschitz and subdifferentially regular. For the maximum function  $g := \max\{f_1, f_2\}$ , one has

$$\mathcal{G}_g(x) = \begin{cases} \mathcal{G}_{f_1}(x) & \text{if } g(x) > \max\{f_2(x), 0\}, 0 \notin \partial f_1(x), \\ \mathcal{G}_{f_2}(x) & \text{if } g(x) > \max\{f_1(x), 0\}, 0 \notin \partial f_2(x), \\ V(x) & \text{if } g(x) = f_1(x) = f_2(x) > 0, 0 \notin \text{conv}(\partial f_1(x) \cup \partial f_2(x)), \\ x & \text{if } g(x) \leq 0, \text{ or } 0 \in \text{conv}(\partial f_1(x) \cup \partial f_2(x)), \end{cases}$$

where

$$V(x) := \left\{ x - \frac{f_i(x)}{\|s(x)\|^2} s(x) \mid s(x) \in \text{conv}(\partial f_1(x) \cup \partial f_2(x)) \right\}.$$

*Proof.* When  $g(x) > 0$ , we consider three cases: (i).  $g(x) > f_2(x)$ ; (ii).  $g(x) > f_1(x)$ ; (iii).  $g(x) \leq \min\{f_2(x), f_1(x)\}$  which is  $g(x) = f_1(x) = f_2(x)$ . Also note that

$$\partial g(x) = \text{conv}(\partial f_1(x) \cup \partial f_2(x))$$

when  $f_1(x) = f_2(x)$  by Fact 3.3. ■

**Proposition 3.6** Assume that  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is lsc and subdifferentiable, and that  $g(x) := f(kx)$  with  $0 \neq k \in \mathbb{R}$ . Then

$$\mathcal{G}_g(x) = \frac{1}{k} \mathcal{G}_f(kx)$$

for every  $x \in \mathbb{R}^n$ . Moreover,  $\text{Fix } \mathcal{G}_g = \frac{1}{k} \text{Fix } \mathcal{G}_f$ .

*Proof.* By Fact 3.2,  $\partial g(x) = k\partial f(y)$  where  $y = kx$ , so  $0 \notin \partial g(x)$  if and only if  $0 \notin \partial f(y)$  with  $y = kx$ . Let  $s$  be a selection of  $\partial f$ . When  $g(x) > 0$  and  $0 \notin \partial g(x)$ , we have  $f(kx) > 0$  and  $0 \notin \partial f(kx)$ , therefore

$$(5) \quad G_{g,ks(k\cdot)}(x) = x - \frac{f(kx)}{\|ks(y)\|^2} ks(y) = x - \frac{1}{k} \frac{f(kx)}{\|s(y)\|^2} s(y)$$

$$(6) \quad = \frac{1}{k} \left( kx - \frac{f(kx)}{\|s(y)\|^2} s(y) \right) = \frac{1}{k} G_{f,s}(kx)$$

where  $s(y) \in \partial f(y)$  with  $y = kx$ .

When  $g(x) \leq 0$  or  $0 \in \partial g(x)$ , we have  $f(y) \leq 0$  or  $0 \in \partial f(y)$  with  $y = kx$ , thus

$$G_{g,ks(k\cdot)}(x) = x = \frac{1}{k} kx = \frac{1}{k} G_{f,s}(kx).$$

This establishes the result. ■

**Proposition 3.7** Let  $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be unitary and  $b \in \mathbb{R}^n$ , and let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be lsc and subdifferentiable. Define  $g : \mathbb{R}^n \rightarrow \mathbb{R} : x \mapsto f(Ax + b)$ . Then

$$(7) \quad \mathcal{G}_g(x) = A^\top(\mathcal{G}_f(Ax + b) - b)$$

for every  $x \in \mathbb{R}^n$ . Furthermore,

$$(8) \quad \text{Fix } \mathcal{G}_g = A^\top(\text{Fix } \mathcal{G}_f - b).$$

*Proof.* Let  $s$  be a selection of  $\partial f$ . By Fact 3.2,  $\partial g(x) = A^\top \partial f(y)$  where  $y = Ax + b$ . As  $A$  is unitary,  $\|A^\top s(y)\| = \|s(y)\|$  for every  $s(y) \in \partial f(y)$ . When  $g(x) > 0$  and  $0 \notin \partial g(x)$ , we have  $f(Ax + b) > 0$  and  $0 \notin \partial f(y)$  with  $y = Ax + b$ , therefore

$$(9) \quad \mathcal{G}_{g, A^\top s(A \cdot + b)}(x) = x - \frac{f(Ax + b)}{\|A^\top s(y)\|^2} A^\top s(y) = A^\top \left( Ax + b - \frac{f(Ax + b)}{\|s(y)\|^2} s(y) - b \right)$$

$$(10) \quad = A^\top(\mathcal{G}_{f,s}(Ax + b) - b).$$

When  $g(x) \leq 0$  or  $0 \in \partial g(x)$ , we have  $f(y) \leq 0$  or  $0 \in \partial f(y)$  with  $y = Ax + b$ , thus

$$\mathcal{G}_{g, A^\top s(A \cdot + b)}(x) = x = A^\top(Ax + b - b) = A^\top(\mathcal{G}_{f,s}(Ax + b) - b).$$

Hence (7) holds. Finally, (8) follows from (7). ■

**Corollary 3.8** Let  $a \in \mathbb{R}^n$ ,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be lsc and subdifferentiable, and  $g(x) := f(x - a)$ . Then  $\mathcal{G}_g(x) = \mathcal{G}_f(x - a) + a$  for every  $x \in \mathbb{R}^n$ . Moreover,  $\text{Fix } \mathcal{G}_g = a + \text{Fix } \mathcal{G}_f$ .

**Theorem 3.9** Assume that  $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$  is locally Lipschitz, and  $g := f^k$  with  $k > 0$ . Then

$$\mathcal{G}_g = \left(1 - \frac{1}{k}\right) \text{Id} + \frac{1}{k} \mathcal{G}_f.$$

*Proof.* By Fact 3.1,  $\partial g(x) = kf(x)^{k-1} \partial f(x)$  when  $f(x) > 0$ . Let  $s$  be a selection of  $\partial f$ . When  $g(x) > 0$  and  $0 \notin \partial g(x)$ , we have  $f(x) > 0$  and  $0 \notin \partial f(x)$ , therefore

$$(11) \quad (\text{Id} - \mathcal{G}_{g, kf^{k-1}s})(x) = \frac{f(x)^k}{\|kf(x)^{k-1}s(x)\|^2} kf(x)^{k-1}s(x) = \frac{1}{k} \frac{f(x)}{\|s(x)\|^2} s(x)$$

$$(12) \quad = \frac{1}{k} (\text{Id} - \mathcal{G}_{f,s})(x).$$

When  $g(x) = 0$  or  $0 \in \partial g(x)$ , we have  $f(x) = 0$  or  $0 \in \partial f(x)$ , thus

$$(\text{Id} - \mathcal{G}_{g, kf^{k-1}s})(x) = 0 = (\text{Id} - \mathcal{G}_{f,s})(x).$$

Therefore,  $\text{Id} - \mathcal{G}_{g, kf^{k-1}s} = \frac{1}{k} (\text{Id} - \mathcal{G}_{f,s})$  which gives  $\mathcal{G}_{g, kf^{k-1}s} = \left(1 - \frac{1}{k}\right) \text{Id} + \frac{1}{k} \mathcal{G}_{f,s}$ . ■

**Remark 3.10** While Theorem 3.9 says that the convex combination of  $\text{Id}$  and  $\mathcal{G}_{f,s}$  is a subgradient projector, the set of subgradient projectors is not a convex set; see Theorems 10.1, 10.3 in Section 10. Note that if  $U_i : \mathcal{H} \rightarrow \mathcal{H}$  is a cutter (see [20] or Definition 5.1) with a common fixed point,  $i \in I := \{1, 2, \dots, m\}$ , and  $w : \mathcal{H} \rightarrow \Delta_m$  is an appropriate weight function, then the operator  $U := \sum_{i \in I} w_i U_i$  is a cutter, cf. [20, Corollary 2.1.49].

**Corollary 3.11** For  $f := d_C^2$  in the case of a closed set  $C \neq \emptyset$  in  $\mathbb{R}^n$ , one has

$$\mathcal{G}_f = \frac{\text{Id} + P_C}{2}.$$

*Proof.* Combine Theorem 3.9 and Example 2.7. ■

**Example 3.12 (penalty function)** Assume that  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  is locally Lipschitz. In optimization, for a direct constraint given by

$$C := \{x \in \mathbb{R}^n \mid g(x) \leq 0\},$$

one can define penalty substitutes. Two popular penalty functions associated with  $g$  are: the linear penalty  $\theta_1 \circ g(x) = t_+(g(x))$  and quadratic penalty  $\theta_2 \circ g(x) = t_+^2(g(x))$  where  $t_+ := \max\{0, t\}$ , cf. [51, page 4]. We have

$$(\forall x \in \mathbb{R}^n) \mathcal{G}_{\theta_1 \circ g}(x) = \mathcal{G}_g(x).$$

Because  $\theta_2 \circ g = (\theta_1 \circ g)^2$ , by Theorem 3.9 we obtain

$$\mathcal{G}_{\theta_2 \circ g} = \frac{\text{Id} + \mathcal{G}_{\theta_1 \circ g}}{2}.$$

The following is immediate from the definition of subgradient projectors.

**Proposition 3.13** Let  $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$  be lsc and subdifferentiable such that  $f \equiv g$  on an open set  $O \subseteq \mathbb{R}^n$ . Then  $\mathcal{G}_f = \mathcal{G}_g$  on  $O$ .

**Remark 3.14** For calculus of subgradient projectors of convex functions, see [12, 44].

## 4 Basic properties of subgradient projectors

In this section under appropriate conditions we show that the subgradient projector can determine a function uniquely up to a positive scalar multiplication, that the subgradient projector enjoys the fixed point closedness property, and that the subgradient projector is continuous if the function is strictly differentiable.

We start with some elementary properties of subgradient projectors.

**Theorem 4.1** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be lsc and subdifferentiable, and  $G_{f,s}$  be given by (1). Then the following hold:

(i) We have

$$(13) \quad \|x - G_{f,s}(x)\| = \frac{f(x)}{\|s(x)\|} \quad \text{and}$$

$$(14) \quad \frac{x - G_{f,s}(x)}{\|x - G_{f,s}(x)\|^2} = \frac{1}{f(x)} s(x)$$

for every  $x$  satisfying  $f(x) > 0$  and  $0 \notin \partial f(x)$ . In particular, when  $f$  is locally Lipschitz, one has

$$(15) \quad \frac{x - G_{f,s}(x)}{\|x - G_{f,s}(x)\|^2} = \frac{s(x)}{f(x)} \in \partial(\ln f)(x);$$

when  $f$  is continuously differentiable, one has

$$(16) \quad \frac{x - G_f(x)}{\|x - G_f(x)\|^2} = \nabla(\ln f(x)).$$

(ii) Set  $g := \ln f$  when  $f > 0$ . Then whenever  $f(x) > 0$  and  $0 \notin \partial f(x)$  we have

$$(17) \quad G_{f,s}(x) = x - \frac{c(x)}{\|c(x)\|^2}$$

where  $c(x) = \frac{s(x)}{f(x)} \in \partial g(x)$ . If  $f$  is continuously differentiable on  $\mathbb{R}^n \setminus \text{lev}_0 f$ , then

$$(18) \quad G_f(x) = x - \frac{\nabla g(x)}{\|\nabla g(x)\|^2},$$

whenever  $f(x) > 0$  and  $\nabla f(x) \neq 0$ .

*Proof.* (i). By the definition of  $G_{f,s}$ , when  $f(x) > 0$  and  $0 \notin \partial f(x)$ ,  $x - G_{f,s}(x) = \frac{f(x)}{\|s(x)\|^2} s(x)$ . Therefore,

$$\|x - G_{f,s}(x)\| = \frac{f(x)}{\|s(x)\|}.$$

It follows that

$$(19) \quad x - G_{f,s}(x) = \frac{f(x)^2}{\|s(x)\|^2} \frac{s(x)}{f(x)} = \|x - G_{f,s}(x)\|^2 \frac{s(x)}{f(x)},$$

equivalently,  $\frac{x - G_{f,s}(x)}{\|x - G_{f,s}(x)\|^2} = \frac{s(x)}{f(x)}$ . When  $f$  is locally Lipschitz, (15) holds because Fact 3.1 gives  $\partial(\ln f)(x) = \frac{\partial f(x)}{f(x)}$  when  $f(x) > 0$ . When  $f$  is continuously differentiable,  $s(x) = \nabla f(x)$ , hence (16) follows from  $\nabla(\ln f(x)) = \nabla f(x)/f(x)$  when  $f(x) > 0$ .

(ii). By Fact 3.1, we have  $\partial g(x) = \frac{\partial f(x)}{f(x)}$  when  $f(x) > 0$ . Then  $c(x) = \frac{s(x)}{f(x)}$  where  $s(x) \in \partial f(x)$ . (17) follows since  $\frac{1}{\|c(x)\|^2} = \frac{f(x)^2}{\|s(x)\|^2}$  when  $f(x) > 0$  and  $0 \notin \partial f(x)$ .

When  $f$  is continuously differentiable on  $\mathbb{R}^n \setminus \text{lev}_0 f$ , the same holds for  $g$ , so (18) follows. ■

#### 4.1 When is a mapping $T$ a subgradient projector?

**Theorem 4.2** *Given a mapping  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ . The following are equivalent:*

- (i)  $T$  is the subgradient projector of a locally Lipschitz function.
- (ii) There exists a locally Lipschitz function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  such that

$$(20) \quad \frac{x - T(x)}{\|x - T(x)\|^2} \in \partial(\ln f(x)) \quad \text{whenever } f(x) > 0 \text{ and } 0 \notin \partial f(x),$$

$$Tx = x \quad \text{whenever } f(x) \leq 0 \text{ or } 0 \in \partial f(x).$$

*Proof.* (i) $\Rightarrow$ (ii). Suppose that  $T = G_{f,s}$  with  $f$  being locally Lipschitz. Apply Theorem 4.1(i) to obtain (20).

(ii) $\Rightarrow$ (i). Assume that (20) holds. By Fact 3.1,  $\partial \ln f = \frac{\partial f}{f}$  when  $f > 0$ . When  $f(x) > 0$  and  $0 \notin \partial f(x)$ , (20) gives

$$\frac{1}{\|x - Tx\|} = \frac{\|s(x)\|}{f(x)} \quad \text{i.e., } \|x - Tx\| = \frac{f(x)}{\|s(x)\|}$$

where  $s(x) \in \partial f(x)$ . Then using (20) again we obtain  $x - Tx = \|x - Tx\|^2 \frac{s(x)}{f(x)}$  so that

$$Tx = x - \|x - Tx\|^2 \frac{s(x)}{f(x)} = x - \left( \frac{f(x)}{\|s(x)\|} \right)^2 \frac{s(x)}{f(x)} = x - \frac{f(x)}{\|s(x)\|^2} s(x)$$

as required. ■

Can the functions Theorem 4.2(i) and (ii) be different? This is answered in the next subsection.

## 4.2 Recovering $f$ from its subgradient projector $G_f$

Can one determine the function  $f$  if  $G_f$  is known? To this end, we recall the concept of essentially strictly differentiable functions by Borwein and Moors [15, Section 4].

**Definition 4.3** *A locally Lipschitz function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is called essentially strictly differentiable on an open set  $O \subseteq \mathbb{R}^n$  if  $f$  is strictly differentiable everywhere on  $O$  except possibly on a Lebesgue null set.*

This class of functions has been extensively studied by Borwein and Moors [15]. This class of functions includes finite-valued convex functions, Clarke regular locally Lipschitz functions, semismooth locally Lipschitz functions,  $C^1$  functions and others, [15, pages 323-328]. If a locally Lipschitz function is essentially strictly differentiable, then  $\partial f$  is single-valued almost everywhere. Moreover, the Clarke subdifferential  $\partial_c f$ , which can be written as  $\text{conv } \partial f$  (the convex hull of  $\partial f$ ) when  $f$  is locally Lipschitz [40, Theorem 3.57], can be recovered by every densely defined selection  $s \in \partial f$ ; see, e.g., [15]. We refer the reader to [23] and [51] for details on the Clarke subdifferential.

**Fact 4.4** *Let  $f, g$  be locally Lipschitz on a polygonally connected and open subset  $O$  of  $\mathbb{R}^n$ . If  $\nabla f = \nabla g$  almost everywhere on  $O$ , then  $h := f - g$  is a constant on  $O$ .*

*Proof.* We prove this by contradiction. Rademacher's Theorem says that a locally Lipschitz function is differentiable almost everywhere, see, e.g., [28, page 81]. By the assumption,  $h$  is locally Lipschitz, so  $\nabla h = 0$  almost everywhere. Suppose that  $x, y \in O$  and  $h(x) \neq h(y)$ . As  $O$  is polygonally connected, there exists  $z \in O$  such that either  $[x, z] \subseteq O$  with  $h(x) \neq h(z)$  or  $[z, y] \subseteq O$  with  $h(z) \neq h(y)$ . Without loss of generality, assume  $[z, y] \subseteq O$  and  $h(z) \neq h(y)$ . As  $h$  is differentiable almost everywhere, by Fubini's Theorem [49, Theorem 6.2.2, page 110], we can choose  $\tilde{z}$  nearby  $z$  and  $\tilde{y}$  nearby  $y$  so that both  $h$  is differentiable and  $\nabla h = 0$  almost everywhere on  $[\tilde{z}, \tilde{y}] \subseteq O$ , and  $h(\tilde{z}) \neq h(\tilde{y})$ . Then

$$h(\tilde{y}) - h(\tilde{z}) = \int_0^1 \langle \nabla h(\tilde{z} + t(\tilde{y} - \tilde{z})), \tilde{y} - \tilde{z} \rangle dt = \int_0^1 0 dt = 0$$

which contradicts  $h(\tilde{z}) \neq h(\tilde{y})$ . ■

**Theorem 4.5** Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a subgradient projector. Suppose that there exist two essentially strictly differentiable functions  $f, f_1 : \mathbb{R}^n \rightarrow \mathbb{R}$  such that  $G_{f,s} = T = G_{f_1,s_1}$  with  $s$  being a selection of  $\partial f$  and  $s_1$  being a selection of  $\partial f_1$ . Then on each polygonally connected component of  $\mathbb{R}^n \setminus \text{Fix } T$  there exists  $k > 0$  such that  $f = kf_1$ .

*Proof.* Assume that there exist two essentially strictly differentiable and locally Lipschitz functions  $f, f_1$  such that  $T = G_{f,s} = G_{f_1,s_1}$ . Since  $T$  has a full domain, we have  $\text{dom } f = \text{dom } f_1 = \mathbb{R}^n$ . By Theorem 4.2, we have

$$\begin{aligned} \frac{x - T(x)}{\|x - T(x)\|^2} &\in \partial(\ln f(x)) \quad \text{whenever } x \in \mathbb{R}^n \setminus \text{Fix } T, \\ \frac{x - T(x)}{\|x - T(x)\|^2} &\in \partial(\ln f_1(x)) \quad \text{whenever } x \in \mathbb{R}^n \setminus \text{Fix } T. \end{aligned}$$

As  $f, f_1$  are locally Lipschitz, both  $\ln f, \ln f_1$  are locally Lipschitz on  $\mathbb{R}^n \setminus \text{Fix } T$ . Then

$$\partial \ln f = \frac{1}{f} \partial f, \quad \partial \ln f_1 = \frac{1}{f_1} \partial f_1$$

by Fact 3.1 or [23, Theorem 2.3.9(ii)]. Because  $f, f_1$  are essentially strictly differentiable and locally Lipschitz,  $\partial f, \partial f_1$  are single-valued almost everywhere [15], thus

$$\partial(\ln f_1(x)) = \frac{x - T(x)}{\|x - T(x)\|^2} = \partial(\ln f(x)) \quad \text{almost everywhere on } \mathbb{R}^n \setminus \text{Fix } T.$$

By Fact 4.4, on each polygonally connected component of  $\mathbb{R}^n \setminus \text{Fix } T$ , there exists  $c \in \mathbb{R}$  such that  $\ln f - \ln f_1 = c$ , which implies that  $f_1 = kf$  for  $k = e^c > 0$ . ■

**Example 4.6** Define

$$f_1 : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} 2x & \text{if } x > 0, \\ 0 & \text{if } -1 \leq x \leq 0, \\ -3(x+1) & \text{if } x < -1, \end{cases}$$

and

$$f_2 : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} x & \text{if } x > 0, \\ 0 & \text{if } -1 \leq x \leq 0, \\ -1(x+1) & \text{if } x < -1. \end{cases}$$

Then

$$(\forall x \in \mathbb{R}) \quad G_{f_1}(x) = G_{f_2}(x) = \begin{cases} 0 & \text{if } x \geq 0, \\ x & \text{if } -1 \leq x \leq 0, \\ -1 & \text{if } x < -1. \end{cases}$$

The set  $\mathbb{R} \setminus [-1, 0]$  has two connected components  $(-\infty, -1)$  and  $(0, +\infty)$ . We have  $f_1 = 3f_2$  on  $(-\infty, -1)$ , and  $f_1 = 2f_2$  on  $(0, +\infty)$ .

The following example shows that Theorem 4.5 fails if one removes the assumption of essentially strictly differentiability.

**Example 4.7** In [16], Borwein, Moors and Wang showed that generically nonexpansive Lipschitz functions have their limiting subdifferentials identically equal to the unit ball; see also [53]. Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a locally Lipschitz function such that  $\partial f(x) = \mathbf{B}$  for every  $x \in \mathbb{R}^n$ . Since  $0 \in \partial f(x)$  for every  $x \in \mathbb{R}^n$ , in view of Definition 2.2 we have  $G_f = \text{Id}$ . As such, generically nonexpansive Lipschitz functions have a subgradient projector equal to the identity mapping.

### 4.3 Fixed point closed property and continuity

**Definition 4.8** We say that an operator  $T : D \rightarrow \mathbb{R}^n$  is fixed point closed at  $x \in D$  if for every sequence  $x_k \rightarrow x$  with  $x_k - Tx_k \rightarrow 0$  one has  $x = Tx$ . If this holds for every  $x \in D$ , we say that  $T$  has the fixed point closed property on  $D$ .

In [20], Cegielski calls the fixed point closed property of  $T$  as  $\text{Id} - T$  being closed at 0.

**Theorem 4.9 (fixed-point closed property)** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be locally Lipschitz and  $G_{f,s}$  be given by Definition 2.2. Then  $G_{f,s}$  is fixed-point closed at every  $x \in \mathbb{R}^n$ , i.e.,

$$(21) \quad \|y - G_{f,s}(y)\| \rightarrow 0 \text{ and } y \rightarrow x \quad \Rightarrow \quad x = G_{f,s}(x).$$

*Proof.* Assume that a sequence  $(y_n)_{n \in \mathbb{N}}$  in  $\mathbb{R}^n$  satisfies

$$(22) \quad \|y_n - G_{f,s}(y_n)\| \rightarrow 0 \text{ and } y_n \rightarrow x.$$

Consider three cases.

Case 1. If there exists infinitely many  $y_n$ 's, say  $(y_{n_k})_{k \in \mathbb{N}}$ , such that  $0 \in \partial f(y_{n_k})$ . Since  $\partial f$  is upper semicontinuous, taking limit when  $k \rightarrow \infty$  gives  $0 \in \partial f(x)$ . Hence  $x = G_f(x)$ .

Case 2. If there exists infinitely many  $y_n$ 's, say  $(y_{n_k})_{k \in \mathbb{N}}$ , such that  $f(y_{n_k}) \leq 0$ . Taking limit when  $k \rightarrow \infty$  and using the continuity of  $f$  at  $x$  gives

$$f(x) = \lim_{k \rightarrow \infty} f(y_{n_k}) \leq 0.$$

Hence  $x = G_f(x)$ .

Case 3. There exists  $N \in \mathbb{N}$  such that  $f(y_n) > 0$  and  $0 \notin \partial f(y_n)$  when  $n > N$ . Then by (13),

$$(23) \quad f(y_n) = \|y_n - G_{f,s}(y_n)\| \|s(y_n)\|.$$

As  $f$  is continuous at  $x$ ,  $f$  is locally Lipschitz around  $x$ , so  $\partial f$  is locally bounded around  $x$ . Therefore,

$$f(x) = \lim_{n \rightarrow \infty} f(y_n) = \lim_{n \rightarrow \infty} (\|y_n - G_{f,s}(y_n)\| \|s(y_n)\|) = 0$$

since  $\|y_n - G_{f,s}(y_n)\| \rightarrow 0$ . Hence  $x = G_{f,s}(x)$ .

Altogether,  $x \in \text{Fix } G_{f,s}$ . This establishes (21) because  $(y_n)_{n \in \mathbb{N}}$  was an arbitrary sequence satisfying (22). ■

The following result generalizes [12, Theorem 5.6].

**Theorem 4.10** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a locally Lipschitz function and essentially strictly differentiable, and let  $G_{f,s}$  be given by Definition 2.2. Suppose that  $x \in \mathbb{R}^n \setminus \text{Fix } G_{f,s}$ . Then the following statements are equivalent:

- (i)  $G_{f,s}$  is continuous at  $x$ .
- (ii)  $f$  is strictly differentiable at  $x$ .

Consequently,  $G_{f,s}$  is continuous on  $\mathbb{R}^n \setminus \text{Fix } G_{f,s}$  if and only if  $f$  is continuously differentiable on  $\mathbb{R}^n \setminus \text{Fix } G_{f,s}$ .

*Proof.* (ii) $\Rightarrow$ (i). Assume that  $f$  is strictly differentiable at  $x \in \mathbb{R}^n \setminus \text{Fix } G_{f,s}$ . Under the assumption,  $s : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is continuous at  $x$  and  $s(x) \neq 0$ . The result follows from the definition  $G_{f,s}$ :

$$y \mapsto y - \frac{f(y)}{\|s(y)\|^2} s(y).$$

(i) $\Rightarrow$ (ii). Assume that  $G_{f,s}$  is continuous at  $x \in \mathbb{R}^n \setminus \text{Fix } G_{f,s}$ . By (14),

$$s(y) = f(y) \frac{y - G_{f,s}(y)}{\|y - G_{f,s}(y)\|^2}$$

so  $s$  is continuous at  $x$ . Because  $s$  is a selection of  $\partial f$  and  $f$  is essentially strictly differentiable, we conclude that  $f$  is strictly differentiable at  $x$ .

Note that  $\text{Fix } G_{f,s}$  is closed by Proposition 2.3(ii). The remaining result follows from the fact that on an open set on which a function is finite, the function is continuously differentiable if and only if the function is strictly differentiable; cf. [51, Corollary 9.19]. ■

We illustrate Theorem 4.10 by two examples.

**Example 4.11** Define

$$f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} |x| & \text{if } x \leq 1, \\ 2x - 1 & \text{if } x > 1. \end{cases}$$

Then

$$G_{f,s}(x) = \begin{cases} 0 & \text{if } x < 1, \\ 1/2 & \text{if } x > 1, \\ 1 - \frac{1}{s(x)} \text{ where } s(x) \in [1, 2] & \text{if } x = 1, \end{cases}$$

is discontinuous at  $x = 1$ , because  $f$  is not differentiable at  $x = 1$ .

*Proof.* When  $x < 0$ ,  $G_{f,s}(x) = x - \frac{-x}{(-1)^2}(-1) = 0$ ; When  $x = 0$ ,  $f(0) = 0$ , so  $G_{f,s}(0) = 0$ ; When  $0 < x < 1$ ,  $G_{f,s}(x) = x - \frac{x}{1^2}(1) = 0$ ; When  $x > 1$ ,  $G_{f,s}(x) = x - \frac{2x-1}{2^2}2 = 1/2$ ; When  $x = 1$ ,  $\partial f(1) = [1, 2]$ , so

$$G_{f,s}(x) = x - \frac{1}{s(x)^2}(s(x)) = x - \frac{1}{s(x)}$$

where  $s(x) \in [1, 2]$ . ■

The next example gives a function that is differentiable but not strictly differentiable at 0, and that its subgradient projector is not continuous at 0.

**Example 4.12** Define

$$f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} x^2 \sin \frac{1}{x} + x + 1 & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Then  $f$  is differentiable everywhere, but not strictly differentiable at 0. The subgradient projector

$$G_f(x) = \begin{cases} x - \frac{x^2 \sin(1/x) + x + 1}{2x \sin(1/x) - \cos(1/x^2) + 1} & \text{if } f(x) > 0 \text{ and } f'(x) \neq 0, \\ x & \text{otherwise,} \end{cases}$$

is not continuous at 0.



*Proof.* At  $x = 0$ ,  $f(0) = 1$  and  $f'(0) = 1$ . The function  $f$  is not strictly differentiable at 0 because  $f'$  is not continuous at 0. Since  $\lim_{x \rightarrow 0} G_f(x)$  does not exist, the subgradient projector is not continuous at 0. ■

How about the continuity of  $G_{f,s}$  on  $\text{Fix } G_{f,s}$ ? Since  $G_{f,s} = \text{Id}$  on  $\text{Fix } G_{f,s}$ , it is always continuous at  $x \in \text{int}(\text{Fix } G_{f,s})$ . The following result deals with the case of  $x \in \text{bdry}(\text{Fix } G_{f,s})$ .

**Theorem 4.13** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be locally Lipschitz,  $G_{f,s}$  be given by Definition 2.2, and  $x \in \text{bdry}(\text{Fix } G_{f,s})$ .*

(i) *Assume that  $f(x) > 0$  and  $0 \in \partial f(x)$ . Then  $G_{f,s}$  is discontinuous at  $x$ .*

(ii) *Assume that  $f(x) \leq 0$ . Suppose that one of the following holds:*

(a)  *$\exists \alpha > 0$  such that*

$$(24) \quad (\forall y : f(y) > 0, 0 \notin \partial f(y)) \quad \alpha f(y) + \langle s(y), x - y \rangle \leq 0.$$

*In particular, this is true when  $f$  is convex.*

(b)

$$(25) \quad 0 \notin \partial f(x).$$

(c)

$$(26) \quad \liminf_{\substack{y \rightarrow x \\ f(y) > 0, 0 \notin \partial f(y)}} \|s(y)\| > 0.$$

*Then  $G_{f,s}$  is continuous at  $x$ .*

*Proof.* (i). As  $x \in \text{bdry}(\text{Fix } G_f)$ , there exists a sequence  $(y_k)_{k \in \mathbb{N}}$  such that  $y_k \rightarrow x$ ,  $f(y_k) > 0$  and  $0 \notin \partial f(y_k)$ . Because  $f$  is locally Lipschitz and  $s(y_k) \in \partial f(y_k)$ ,  $(s(y_k))_{k \in \mathbb{N}}$  is locally bounded. By taking a subsequence if necessary, we can assume that  $\|s(y_k)\| \rightarrow l \in \mathbb{R}_+$ . Taking limit when  $k \rightarrow \infty$  yields that

$$\|y_k - G_{f,s}(y_k)\| = \frac{f(y_k)}{\|s(y_k)\|} \rightarrow \frac{f(x)}{l}$$

which is  $+\infty$  if  $l = 0$  or a positive number if  $l > 0$ . Because  $G_{f,s}(x) = x$ , this shows that  $G_{f,s}$  is not continuous at  $x$ .

(ii). To show that  $G_{f,s}$  is continuous at  $x$ , it suffices to show that

$$(27) \quad \lim_{\substack{y \rightarrow x \\ f(y) > 0, 0 \notin \partial f(y)}} \frac{f(y)}{\|s(y)\|} = 0.$$

Indeed, by Theorem 4.1, when  $f(y) > 0$  and  $0 \notin \partial f(y)$ , we have

$$\|y - G_{f,s}(y)\| = \frac{f(y)}{\|s(y)\|}.$$

Then (27) gives that  $\lim_{y \rightarrow x} G_{f,s}(y) = \lim_{y \rightarrow x} (G_{f,s}(y) - y) + \lim_{y \rightarrow x} y = x$ . When  $y \in \text{Fix } G_{f,s}$ ,  $G_{f,s}(y) = y$ , clearly  $\lim_{y \rightarrow x} G_{f,s}(y) = x$ . Hence  $G_{f,s}$  is continuous at  $x$ .

Now (24) gives

$$f(y) \leq \frac{\langle s(y), y - x \rangle}{\alpha} \leq \|s(y)\| \frac{\|y - x\|}{\alpha}$$

so that

$$\frac{f(y)}{\|s(y)\|} \leq \frac{\|y - x\|}{\alpha}$$

which implies (27).

Next, we show (25) implies (26). Note that (25) gives  $d_{\partial f(x)}(0) > 0$  since  $\partial f(x)$  is closed by [51, Theorem 8.6]. Because  $f$  is locally Lipschitz, in view of [51, Proposition 8.7], we have that  $\limsup_{y \rightarrow x} \partial f(y) \subseteq \partial f(x)$ , hence  $0 \notin \partial f(y)$  for  $y$  sufficiently nearby  $x$ . Invoking [51, Corollary 4.7(b)], we obtain

$$\liminf_{y \rightarrow x} d_{\partial f(y)}(0) \geq d_{\partial f(x)}(0),$$

from which it follows that

$$(28) \quad \liminf_{\substack{y \rightarrow x \\ f(y) > 0, 0 \notin \partial f(y)}} \|s(y)\| \geq \liminf_{\substack{y \rightarrow x \\ f(y) > 0, 0 \notin \partial f(y)}} d_{\partial f(y)}(0)$$

$$(29) \quad \geq \liminf_{y \rightarrow x} d_{\partial f(y)}(0) \geq d_{\partial f(x)}(0) > 0,$$

and this gives (26).

Finally, (26) gives (27) because  $\lim_{y \rightarrow x} f(y) = 0$  and

$$0 \leq \liminf_{\substack{y \rightarrow x \\ f(y) > 0, 0 \notin \partial f(y)}} \frac{f(y)}{\|s(y)\|} \leq \limsup_{\substack{y \rightarrow x \\ f(y) > 0, 0 \notin \partial f(y)}} \frac{f(y)}{\|s(y)\|} \leq \frac{\lim_{y \rightarrow x, f(y) > 0} f(y)}{\liminf_{\substack{y \rightarrow x \\ f(y) > 0, 0 \notin \partial f(y)}} \|s(y)\|} = 0.$$

■

Here is an example showing the result of Theorem 4.13(i).

**Example 4.14** (1). Define  $f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto x^3 + 1$ . Then

$$G_f(x) = \begin{cases} \frac{2x}{3} - \frac{1}{3x^2} & \text{if } x \neq 0 \text{ and } x > -1, \\ x & \text{if } x = 0 \text{ or } x \leq -1, \end{cases}$$

has  $\text{Fix } G_f = (-\infty, -1] \cup \{0\}$ , and  $G_f$  is not continuous at  $x = 0$ .

(2). Define

$$f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} x + 1 & \text{if } x \leq 0, \\ 1 & \text{if } x \geq 0. \end{cases}$$

Then

$$G_f(x) = \begin{cases} -1 & \text{if } -1 < x < 0, \\ x & \text{if } x \leq -1 \text{ or } x \geq 0, \end{cases}$$

has  $\text{Fix } G_f = (-\infty, -1] \cup [0, +\infty)$ , and  $G_f$  is not continuous at  $x = 0$ .

## 4.4 The family of subgradient projectors

**Theorem 4.15** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be locally Lipschitz. Then the following are equivalent:*

- (i)  $\mathcal{G}_f$  is a single-valued.
- (ii)  $f$  is strictly differentiable on  $\mathbb{R}^n \setminus \text{Fix } \mathcal{G}_f$ .

*Proof.* (i) $\Rightarrow$ (ii). By (14) in Theorem 4.1, when  $x \in \mathbb{R}^n \setminus \text{Fix } \mathcal{G}_f$ , we have

$$s(x) = f(x) \frac{x - G_{f,s}(x)}{\|x - G_{f,s}(x)\|^2}$$

where  $s(x) \in \partial f(x)$ . By the assumption  $\mathcal{G}_f = T$  for an everywhere single-valued  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , so

$$s(x) = f(x) \frac{x - Tx}{\|x - Tx\|^2}.$$

It follows that  $\partial f(x)$  is a singleton, so  $f$  is strictly differentiable at  $x$  by [51, Theorem 9.18]. Therefore,  $f$  is strictly differentiable on  $\mathbb{R}^n \setminus \text{Fix } \mathcal{G}_f$ .

(ii) $\Rightarrow$ (i). Clear. ■

**Theorem 4.16** *Let  $C \subseteq \mathbb{R}^n$  be a nonempty closed set. Then the following are equivalent:*

- (i)  $\mathcal{G}_{d_C}$  is a single-valued.
- (ii)  $C$  is convex.

*Proof.* According to Fact 2.6, we have  $\text{Fix } \mathcal{G}_{d_C} = C$ .

(i) $\Rightarrow$ (ii). By Theorem 4.15,  $d_C$  is strictly differentiable on  $\mathbb{R}^n \setminus C$ . Fact 2.6 shows that  $P_C$  is single-valued for every  $x \in \mathbb{R}^n \setminus C$ . Hence,  $C$  is convex; cf. [27, Theorem 12.7].

(ii) $\Rightarrow$ (i). Apply Fact 2.8. ■

## 5 When is the subgradient projector $G_f$ a cutter or local cutter?

In this section we provide conditions for a subgradient projector to be a cutter or local cutter, and an explicit nonconvex function with a cutter subgradient projector. Along the way some calculus on cutter subgradient projectors are also developed.

### 5.1 Cutters, quasi-firmly nonexpansive mappings, and local cutters

Recall the following well-known algorithmic operators.

**Definition 5.1** ([20, page 53]) *Let  $D$  be a nonempty subset of  $\mathbb{R}^n$  and  $T : D \rightarrow \mathbb{R}^n$ . We say that  $T$  is a cutter if  $\text{Fix } T \neq \emptyset$  and*

$$(30) \quad (\forall x \in D)(\forall u \in \text{Fix } T) \quad \langle x - Tx, u - Tx \rangle \leq 0.$$

**Definition 5.2** ([20, page 56]) Let  $D$  be a nonempty subset of  $\mathbb{R}^n$  and  $T : D \rightarrow \mathbb{R}^n$ . We say that  $T$  is quasi-firmly nonexpansive (quasi-fne) if  $\text{Fix } T \neq \emptyset$  and

$$(\forall x \in D)(\forall u \in \text{Fix } T) \quad \|Tx - u\|^2 + \|x - Tx\|^2 \leq \|x - u\|^2.$$

In [20, page 56], quasi-fne mappings are called strongly quasicononexpansive mappings.

The following fact says that a cutter is strongly Fejer monotone with respect to the set of its fixed points, and that cutters and quasi-fne mappings are the same, see [20, page 108].

**Fact 5.3** ([20, Theorem 2.1.39, Lemma 2.1.36])

- (i) A mapping  $T : D \rightarrow \mathbb{R}^n$  is a cutter if and only if  $T$  is quasi-fne.
- (ii) Let  $T : D \rightarrow \mathbb{R}^n$  be a cutter. Then  $T$  is always continuous on  $\text{Fix } T$ .
- (iii) Let  $T : D \rightarrow \mathbb{R}^n$  be a cutter. Then  $\text{Fix } T$  is closed and convex.

In Definitions 5.1 and (i), they require that  $T$  satisfies the inequalities for all  $x \in D$  and  $u \in \text{Fix } T$ . In practice, the sets  $D$  and  $\text{Fix } T$  might be too large to verify those inequalities. We now introduce local cutters and locally quasi-firmly nonexpansive mappings.

**Definition 5.4** A mapping  $T : D \rightarrow \mathbb{R}^n$  is a local cutter at  $\bar{x} \in \text{Fix } T$  if  $\text{Fix } T \neq \emptyset$  and there exists  $\delta > 0$  such that

$$(31) \quad (\forall x \in \mathbf{B}(\bar{x}, \delta) \cap D)(\forall u \in \mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T) \quad \langle x - Tx, u - Tx \rangle \leq 0.$$

**Definition 5.5** A mapping  $T : D \rightarrow \mathbb{R}^n$  is locally quasi-firmly nonexpansive (locally quasi-fne) at  $\bar{x} \in \text{Fix } T$  if there exists  $\delta > 0$  such that

$$(\forall x \in \mathbf{B}(\bar{x}, \delta) \cap D)(\forall u \in \mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T) \quad \|Tx - u\|^2 + \|Tx - x\|^2 \leq \|x - u\|^2.$$

A localized version of Fact 5.3(i) comes next.

**Proposition 5.6** A mapping  $T : D \rightarrow \mathbb{R}^n$  is a local cutter at  $\bar{x} \in \text{Fix } T$  if and only if  $T$  is locally quasi-fne at  $\bar{x} \in \text{Fix } T$ .

*Proof.* This follows from  $\|x - u\|^2 = \|Tx - u\|^2 + \|x - Tx\|^2 + 2 \langle x - Tx, Tx - u \rangle$ . ■

**Proposition 5.7** Assume that  $T : \mathbb{R} \rightarrow \mathbb{R}$  and  $\text{Fix } T \neq \emptyset$ . Then  $T$  is a cutter on  $\mathbb{R}$  if and only if

$$(32) \quad (\forall x \in \mathbb{R}) \quad Tx \in [x, P_{\text{Fix } T}x].$$

*Proof.* The sufficiency is clear. Conversely, when  $x \in \text{Fix } T$ , (32) clearly holds. Assume  $x \notin \text{Fix } T$  and  $c \in \text{Fix } T$ . Because  $T$  is from  $\mathbb{R}$  to  $\mathbb{R}$ , there exists  $\lambda \in \mathbb{R}$  such that

$$Tx = (1 - \lambda)x + \lambda c.$$

As  $T$  is a cutter, we have

$$(x - Tx)(c - Tx) = -\lambda(1 - \lambda)(x - c)^2 \leq 0,$$

which gives  $0 \leq \lambda \leq 1$ , so that  $Tx \in [x, c]$ . Since  $c \in \text{Fix } T$  was arbitrary, it follows that  $x \in [x, P_{\text{Fix } T}x]$ . ■

**Remark 5.8** Compare Proposition 5.7 to Corollary 8.4, which characterizes the subgradient projector of a convex function on  $\mathbb{R}$ .

For a nonempty convex set  $C \subseteq \mathbb{R}^n$ , the recession cone of  $C$  is  $\text{rec } C := \{x \in \mathbb{R}^n \mid x + C \subseteq C\}$ . The negative polar of  $K \subseteq \mathbb{R}^n$  is  $K^- := \{y \in \mathbb{R}^n \mid \langle y, x \rangle \leq 0, \forall x \in K\}$ .

**Proposition 5.9** *Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a cutter. Then*

$$\text{ran}(\text{Id} - T) \subseteq (\text{rec}(\text{Fix } T))^-.$$

*Consequently, when  $\text{Fix } T$  is a linear subspace,  $\text{ran}(\text{Id} - T) \subseteq (\text{Fix } T)^\perp$ . In other words,  $\text{ran}(\text{Id} - T) \subseteq (\ker(\text{Id} - T))^\perp$ .*

*Proof.* Let  $x - Tx \in \text{ran}(\text{Id} - T)$  and  $v \in \text{rec}(\text{Fix } T)$ . Then for every  $k > 0$  and  $u \in \text{Fix } T$ , we have  $u + kv \in \text{Fix } T$ . The assumption of  $T$  being a cutter implies

$$\langle x - Tx, u + kv - Tx \rangle \leq 0 \quad \Rightarrow \quad \langle x - Tx, u/k + v - Tx/k \rangle \leq 0.$$

When  $k \rightarrow \infty$  this gives  $\langle x - Tx, v \rangle \leq 0$ . Since  $v \in \text{rec}(\text{Fix } T)$  was arbitrary, we have  $x - Tx \in (\text{rec}(\text{Fix } T))^-$ .

When  $\text{Fix } T$  is a linear subspace,  $\text{Fix } T = \text{rec}(\text{Fix } T)$  and  $(\text{rec}(\text{Fix } T))^- = (\text{rec}(\text{Fix } T))^\perp$ . ■

## 5.2 Characterizations of $G_f$ being a cutter or local cutter

Our first result characterizes the class of functions  $f$  for which its  $G_{f,s}$  is a cutter.

**Lemma 5.10** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be lsc and subdifferentiable, and let  $G_{f,s}$  be given by Definition 2.2. Suppose that  $f(x) > 0, 0 \notin \partial f(x)$ . Then*

$$\langle x - G_{f,s}(x), u - G_{f,s}(x) \rangle = \frac{f(x)}{\|s(x)\|^2} (f(x) + \langle s(x), u - x \rangle).$$

*Proof.* Let  $f(x) > 0, 0 \notin \partial f(x)$ . The definition of  $G_{f,s}$  gives

$$\begin{aligned} & \langle x - G_{f,s}(x), u - G_{f,s}(x) \rangle \\ &= \left\langle \frac{f(x)s(x)}{\|s(x)\|^2}, u - x + \frac{f(x)}{\|s(x)\|^2}s(x) \right\rangle \\ &= \frac{f(x)}{\|s(x)\|^2} \langle s(x), u - x \rangle + \frac{f^2(x)}{\|s(x)\|^2} \\ &= \frac{f(x)}{\|s(x)\|^2} (f(x) + \langle s(x), u - x \rangle). \end{aligned}$$

■

**Theorem 5.11 (level sets of tangent planes including the target set)** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be lsc and subdifferentiable, let  $G_{f,s}$  be given by Definition 2.2, and*

$$S := \{u \in \mathbb{R}^n \mid f(u) \leq 0 \text{ or } 0 \in \partial f(u)\}.$$

*Then the following hold:*

(i)  $G_{f,s}$  is a cutter if and only if whenever  $x \notin S$  and  $u \in S$  one has

$$f(x) + \langle s(x), u - x \rangle \leq 0.$$

(ii) Let  $\bar{x} \in S$  and  $\delta > 0$ .  $G_{f,s}$  is a cutter on  $\mathbf{B}(\bar{x}, \delta)$  if and only if for all  $x \in \mathbf{B}(\bar{x}, \delta) \setminus S$  and  $u \in S \cap \mathbf{B}(\bar{x}, \delta)$  one has

$$f(x) + \langle s(x), u - x \rangle \leq 0.$$

*Proof.* (i). When  $f(x) \leq 0$  or  $0 \in \partial f(x)$ ,  $x = G_{f,s}(x)$ , (31) holds for  $T = G_{f,s}$ . Assume that  $f(x) > 0$ ,  $0 \notin \partial f(x)$  and  $s(x) \in \partial f(x)$ . By Lemma 5.10,

$$\langle x - G_{f,s}(x), u - G_{f,s}(x) \rangle = \frac{f(x)}{\|s(x)\|^2} (f(x) + \langle s(x), u - x \rangle).$$

Since  $f(x) > 0$ , we deduce that

$$\langle x - G_{f,s}(x), u - G_{f,s}(x) \rangle \leq 0 \quad \Leftrightarrow \quad f(x) + \langle s(x), u - x \rangle \leq 0.$$

Hence, the result follows from Definition 5.1.

(ii). Apply the same arguments as in above with  $x \in \mathbf{B}(\bar{x}, \delta)$  and  $u \in S \cap \mathbf{B}(\bar{x}, \delta)$ . ■

One immediately obtains the following:

**Fact 5.12** ([20, page 146]) *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be convex, let  $G_{f,s}$  be given by Definition 2.2, and  $\text{lev}_0 f \neq \emptyset$ . Then  $G_{f,s}$  is a cutter. Consequently,  $G_{f,s}$  is continuous at every  $x \in \text{lev}_0 f$ .*

*Proof.* As  $\text{lev}_0 f \neq \emptyset$ ,  $\text{Fix } G_{f,s} = \text{lev}_0 f$ . Assume that  $f(x) > 0$ . For  $u \in G_{f,s}$ ,  $f(u) \leq 0$ . By the convexity of  $f$  we have

$$f(x) + \langle s(x), u - x \rangle \leq f(u) \leq 0.$$

Theorem 5.11 shows that  $G_{f,s}$  is a cutter. The remaining result follows from Fact 5.3(ii). ■

In Fact 5.12,  $\text{lev}_0 f \neq \emptyset$  is required, as the following example shows.

**Example 5.13** (1). Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be defined by  $(\forall x \in \mathbb{R}) \quad f(x) := \exp |x|$ . Then  $\text{lev}_0 f = \emptyset$  and

$$G_f(x) = \begin{cases} x - 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ x + 1 & \text{if } x < 0. \end{cases}$$

In particular, this  $G_f$  is discontinuous at  $x = 0$  and not a cutter. Moreover,  $G_f$  is not monotone. (2). Consider

$$f : \mathbb{R}^n \rightarrow \mathbb{R} : x \mapsto \exp(\|x\|^2/2).$$

We have  $\text{lev}_0 f = \emptyset$  and

$$G_f(x) = \begin{cases} x - \frac{x}{\|x\|^2} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

In particular,  $G_f$  is not continuous at 0, so not a cutter.

**Example 5.14** The nonconvex function

$$f : \mathbb{R}^n \rightarrow \mathbb{R} : x \mapsto \begin{cases} \|x\|^2 & \text{if } \|x\| \leq 1, \\ 1 & \text{if } \|x\| > 1, \end{cases}$$

has  $G_f$  being a cutter on a neighborhood of 0, but not a cutter on  $\mathbb{R}^n$ .

It is instructive to consider  $d_C$  where  $C \subseteq \mathbb{R}^n$  is closed and nonempty.

**Proposition 5.15** *Let  $C \subseteq \mathbb{R}^n$  be closed and nonempty, and  $s$  be a selection of  $\partial d_C$ . Then  $G_{d_C, s}$  is a cutter if and only if the set  $C$  is convex.*

*Proof.* By Fact 2.6,  $0 \notin \partial d_C(\bar{x})$  whenever  $\bar{x} \notin C$ , because  $\|v\| = 1$  for every  $v \in \partial d_C(\bar{x})$ . This implies that  $\text{Fix } G_{d_C, s} = C$ .

Assume that  $G_{d_C, s}$  is a cutter. Then  $\text{Fix } G_{d_C, s} = C$  is convex. Conversely, assume that  $C$  is convex. We have  $d_C$  is convex, consequently,  $G_{d_C, s}$  is a cutter. ■

**Theorem 5.16** *Let  $k \geq 1$ ,  $f : \mathbb{R}^n \rightarrow [0, +\infty)$  be locally Lipschitz, and let  $G_{f, s}$  be given by Definition 2.2. Suppose that  $G_{f, s}$  is a cutter, and  $\text{Fix } G_{f, s} \neq \emptyset$ . If  $g = f^k$ , then  $G_{g, k f^{k-1} s}$  is a cutter.*

*Proof.* By Theorem 3.9,  $G_{g, k f^{k-1} s} = (1 - 1/k) \text{Id} + 1/k G_{f, s}$ . As  $\text{Id}$  and  $G_{f, s}$  are both cutters, and  $\text{Fix } G_{f, s} \cap \text{Fix } \text{Id} = \text{Fix } G_{f, s} \neq \emptyset$ , being a convex combination of cutters,  $G_{g, k f^{k-1} s}$  is a cutter by [20, Corollary 2.1.49, page 62]. ■

In Corollary 11.6, we give an example showing even though  $G_{f^2, 2f_s}$  is a cutter,  $G_{f, s}$  might not be a cutter; so the converse of Theorem 5.16 is not true.

**Theorem 5.17** *Let  $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be unitary,  $b \in \mathbb{R}^n$ , let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be lsc and subdifferentiable, and let  $G_{f, s}$  be given by Definition 2.2. Define  $g : \mathbb{R}^n \rightarrow \mathbb{R} : x \mapsto f(Ax + b)$ . If  $G_{f, s}$  is a cutter, then  $G_{g, A^T s(A \cdot + b)}$  is a cutter.*

*Proof.* Let  $x \in \mathbb{R}^n$ ,  $u \in \text{Fix } G_{g, A^T s(A \cdot + b)}$ . Proposition 3.7 gives

$$G_{g, A^T s(A \cdot + b)}(x) = A^T (G_{f, s}(Ax + b) - b), \quad Au + b \in \text{Fix } G_{f, s}.$$

Since  $A$  is unitary and  $G_{f, s}$  is a cutter, we have

$$(33) \quad \|x - G_{g, A^T s(A \cdot + b)}(x)\|^2 = \|x - A^T (G_{f, s}(Ax + b) - b)\|^2 = \|Ax + b - G_{f, s}(Ax + b)\|^2$$

$$(34) \quad \leq \|Ax + b - (Au + b)\|^2 - \|G_{f, s}(Ax + b) - (Au + b)\|^2$$

$$(35) \quad = \|x - u\|^2 - \|G_{f, s}(Ax + b) - Au - b\|^2$$

$$(36) \quad = \|x - u\|^2 - \|A^T (G_{f, s}(Ax + b) - b) - u\|^2$$

$$(37) \quad = \|x - u\|^2 - \|G_{g, A^T s(A \cdot + b)}(x) - u\|^2.$$

Hence  $G_{g, A^T s(A \cdot + b)}$  is a cutter by Fact 5.3(i). ■

**Corollary 5.18** *Let  $B$  be an  $n \times n$  symmetric matrix. Define*

$$f : \mathbb{R}^n \rightarrow \mathbb{R} : x \mapsto \frac{1}{2} x^T B x.$$

Then

$$(38) \quad G_f(x) = \begin{cases} x - \frac{x^\top Bx}{2\|Bx\|^2} Bx & \text{if } x^\top Bx > 0 \text{ and } Bx \neq 0, \\ x & \text{otherwise.} \end{cases}$$

Moreover, the following are equivalent:

- (i)  $G_f$  is a cutter.
- (ii)  $B$  is positive semidefinite or negative semidefinite.

*Proof.* (38) follows from Definition 2.2.

Because  $B$  is symmetric, there exists an orthogonal matrix  $Q$  such that  $Q^\top BQ = D$ , where  $D$  is an  $n \times n$  diagonal matrix whose diagonal entries are eigenvalues of  $B$ . Using  $x = Qy$ , Theorem 5.17 shows that  $G_f$  is a cutter if and only if  $G_g$  is a cutter, where  $g : \mathbb{R}^n \rightarrow \mathbb{R} : y \mapsto f(Qy) = \frac{1}{2}y^\top Dy$ .

(i) $\Rightarrow$ (ii). (i) implies that  $G_g$  is a cutter. This means that

$$(39) \quad (\forall y \in \mathbb{R}^n : y^\top Dy > 0)(\forall u \in \mathbb{R}^n : u^\top Du \leq 0) \quad y^\top Du \leq \frac{1}{2}y^\top Dy.$$

We will show that all nonzero diagonal entries of  $D$  have the same sign. Suppose to the contrary that there exist diagonal entries of  $D$  such that  $\lambda_i > 0, \lambda_j < 0$ . Put  $y_k = 0, u_k = 0$  for  $k = 1, \dots, n, k \neq i, j$ . Then (39) reduces to that whenever  $\lambda_i y_i^2 + \lambda_j y_j^2 > 0$  and

$$(40) \quad \lambda_i u_i^2 + \lambda_j u_j^2 \leq 0,$$

we have

$$(41) \quad \lambda_i y_i u_i + \lambda_j y_j u_j \leq \frac{1}{2}(\lambda_i y_i^2 + \lambda_j y_j^2).$$

Fix  $(y_i, y_j)$  such that  $\lambda_i y_i^2 + \lambda_j y_j^2 > 0$  and  $y_j < 0$ . When  $u_i = 0, u_j \rightarrow +\infty$ , (40) is verified but (41) fails to hold. This contradicts that  $G_g$  is a cutter. Hence all nonzero diagonal entries of  $D$  must have the same sign, which implies that  $B$  is positive semidefinite if positive sign, and  $B$  is negative semidefinite if negative sign.

(ii) $\Rightarrow$ (i). When  $B$  is positive semidefinite,  $f$  is convex, we apply Fact 5.12. When  $B$  is negative semidefinite,  $G_g = \text{Id}$  is a cutter. ■

**Theorem 5.19** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be lsc and subdifferentiable, and let  $G_{f,s}$  be given by Definition 2.2. Assume that  $\mathbb{R} \ni k \neq 0$  and  $g(x) = f(kx)$ . If  $G_{f,s}$  is a cutter, then  $G_{g,ks(k\cdot)}$  is a cutter.

*Proof.* Proposition 3.6 gives  $G_{g,ks(k\cdot)}(x) = \frac{1}{k}G_{f,s}(kx)$ , and  $\text{Fix } G_{g,ks(k\cdot)} = \frac{1}{k}\text{Fix } G_{f,s}$ . Let  $x \in X$  and  $u \in \text{Fix } G_{g,ks(k\cdot)}$ . We have

$$(42) \quad \langle x - G_{g,ks(k\cdot)}(x), u - G_{g,ks(k\cdot)}(x) \rangle = \langle x - 1/kG_{f,s}(kx), u - 1/kG_{f,s}(kx) \rangle$$

$$(43) \quad = 1/k^2 \langle kx - G_{f,s}(kx), ku - G_{f,s}(kx) \rangle \leq 0$$

since  $G_{f,s}$  is a cutter. Therefore,  $G_{g,ks(k\cdot)}$  is a cutter. ■

One might ask: If each function  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$  has  $G_{f_i}$  being a cutter, must the maximum  $g := \max\{f_1, f_2\}$  have  $G_g$  being a cutter? The answer is negative as the following example shows.



**Example 5.20** Let  $f_1, f_2 : \mathbb{R}^n \rightarrow \mathbb{R}$  be defined by  $f_1(x) := 1 + x$  and  $f_2(x) := 1 - x$  on  $\mathbb{R}$ . Each  $G_{f_i}$  is a cutter by Fact 5.12. The function  $g(x) := \max\{f_1(x), f_2(x)\}$  has

$$G_g(x) = \begin{cases} -1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x < 0, \end{cases}$$

which is not continuous at  $x = 0$ , so  $G_g$  is not a cutter.

### 5.3 A nonconvex function whose $G_f$ is a cutter

**Example 5.21** If  $f$  is not convex,  $G_{f,s}$  need not be a cutter. Consider

$$f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto 1 - \exp(-x^2).$$

Then the subgradient projector of  $f$  is

$$G_{f,s}(x) = \begin{cases} x - \left(\frac{1}{2x \exp(-x^2)} - \frac{1}{2x}\right) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

and  $\text{Fix } G_f = \{0\}$ . However  $G_f$  is not a cutter. Indeed, when  $|x| > \sqrt{2}$  we have

$$\begin{aligned} f(x) + s(x)(0 - x) &= 1 - \exp(-x^2) + (2x \exp(-x^2))(0 - x) \\ &= \frac{\exp(x^2) - (1 + 2x^2)}{\exp(x^2)} \geq \frac{1 + x^2 + \frac{x^4}{2} - (1 + 2x^2)}{\exp(x^2)} \\ &= \frac{x^2(x^2 - 2)}{2 \exp(x^2)} > 0. \end{aligned}$$

By Theorem 5.11,  $G_f$  is not a cutter.

**Example 5.22** Even though  $f$  is not convex,  $G_{f,s}$  may still be a cutter. Define

$$f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} 0 & \text{if } x \leq 0, \\ x & \text{if } 0 \leq x \leq 20/7, \\ 8(x - 2.5) & \text{if } 20/7 \leq x \leq 3, \\ 2(x - 1) & \text{if } x > 3. \end{cases}$$

Then  $f$  is not convex since  $f'(x)$  is not monotone on  $[20/7, +\infty)$ . However, its subgradient projector

$$G_{f,s}(x) = \begin{cases} x & \text{if } x \leq 0, \\ 0 & \text{if } 0 < x < 20/7, \\ \frac{20}{7} - \frac{20}{7} \frac{1}{s(x)} & \text{if } x = 20/7, \text{ where } s(x) \in [1, 8], \\ 2.5 & \text{if } 20/7 < x < 3, \\ 3 - \frac{4}{s(x)} & \text{if } x = 3, \text{ where } s(x) \in \{2, 8\}, \\ 1 & \text{if } x > 3. \end{cases}$$

is a cutter. To see this, by Theorem 5.11, it suffices to consider zero level sets of tangent planes. Indeed,

Let  $f(u) \leq 0$ , i.e.,  $u \leq 0$ . When  $x_0 > 3$ ,

$$f(x_0) + s(x_0)(u - x_0) = 2(u - 1) \leq 0;$$

when  $x_0 = 3$ ,

$$f(x_0) + s(x_0)(u - x_0) = 4 + s(3)(u - 3) \leq 4 + 2(u - 3) \leq 0;$$

where  $2 \leq s(3) \leq 8$ ; when  $20/7 < x_0 < 3$ ,

$$f(x_0) + s(x_0)(u - x_0) = 8(u - 2.5) \leq 0;$$

when  $x_0 = 20/7$ ,

$$f(x_0) + s(x_0)(u - x_0) = 70/2 + s(20/7)(u - 20/7) \leq u \leq 0;$$

where  $1 \leq s(20/7) \leq 8$ ; when  $0 < x_0 < 20/7$ ,

$$f(x_0) + s(x_0)(u - x_0) = u \leq 0.$$

See Corollary 11.6(ii) for an example on  $\mathbb{R}^2$ . Note that even if  $G_f$  is continuous, it does not mean that  $G_f$  is a cutter, e.g., see Example 2.5(ii). In [20], Cegielski developed a systematic theory for cutters. The theory of cutters can be used to study the class of functions (Theorem 5.11) whose subgradient projectors are cutters.

One might also ask: If  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  has  $G_{f,s}$  being a cutter, does  $g := f + r$  have  $G_{g,s}$  being a cutter for every  $r \in \mathbb{R}$ ? In general, the answer is negative. When  $f$  is convex and  $\text{lev}_0 f \neq \emptyset$ , it follows from Fact 5.12 that  $G_{f-r,s}$  is a cutter whenever  $r > 0$ . This might fail for  $r < 0$  as the following example shows.

**Example 5.23** Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be defined by  $f(x) := e^{|x|} - 1$ . Then

$$G_{f,s}(x) = \begin{cases} x - 1 + e^{-x} & \text{if } x > 0, \\ x + 1 - e^x & \text{if } x < 0, \\ 0 & \text{if } x = 0, \end{cases}$$

is a cutter by Fact 5.12. However, for  $g : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto e^{|x|}$ , we have  $g = f + 1$  and but  $G_{g,s}$  is not a cutter by Example 5.13(1).

For a nonconvex function function, although  $G_{f,s}$  is a cutter,  $G_{f-r,s}$  might not be a cutter even when  $r > 0$ .

**Example 5.24** Let  $f$  be given by Example 5.22, and  $g := f - 20/7$ . Then

$$G_{g,s}(x) = \begin{cases} x & \text{if } x \leq 20/7, \\ \frac{20}{7} & \text{if } 20/7 < x < 3, \\ s \in \{17/7, 20/7\} & \text{if } x = 3, \\ 17/7 & \text{if } x > 3. \end{cases}$$

As shown in Example 5.22,  $G_{f,s}$  is a cutter. However,  $G_{g,s}$  is not a cutter by using Proposition 5.7 or by direct calculations using Definition 5.1.

## 6 Convergence analysis of subgradient projectors

In this section, we study the convergence of the sequence generated by the subgradient projector. When the function is convex, the convergence analysis has been fairly well known; see, e.g., [47, Section 5.3], [46], [9], and [20]. For nonconvex functions, we demonstrate that the convergence results on cutters, local cutters, quasi-ne mappings, and local quasi-ne mappings can be effectively used. It turns out that local cutters and local quasi-ne mappings are more appropriate for nonconvex functions.

In addition to cutters and local cutters, see Definitions 5.1 and 5.4, quasi-nonexpansive mappings and local quasi-nonexpansive mappings are also useful for the convergence analysis.

### 6.1 Quasi-nonexpansive mappings and local quasi-nonexpansive mappings

According to [7, page 59], and [20, page 47], we define:

**Definition 6.1** *Let  $D$  be a nonempty subset of  $\mathbb{R}^n$  and  $T : D \rightarrow \mathbb{R}^n$ . We say that*

(i)  *$T$  is quasinonexpansive (quasi-ne) if*

$$(\forall x \in D)(\forall y \in \text{Fix } T) \quad \|Tx - y\| \leq \|x - y\|.$$

(ii) *A mapping  $T : D \rightarrow D$  is said to be asymptotically regular at  $x \in D$  if  $\|T^{k+1}x - T^kx\| \rightarrow 0$  as  $k \rightarrow \infty$ ; it is said to be asymptotic regular on  $D$  if it is so at every  $x \in D$ .*

Definition 6.1(ii) requires that  $T$  satisfy the inequalities for all  $x \in D$  and  $y \in \text{Fix } T$ . In practice, the sets  $D$  and  $\text{Fix } T$  might be too large to verify those inequalities. We now introduce locally quasinonexpansive mappings.

**Definition 6.2** *A mapping  $T : D \rightarrow \mathbb{R}^n$  is locally quasinonexpansive (locally quasi-ne) at  $\bar{x} \in \text{Fix } T$  if there exists  $\delta > 0$  such that*

$$(\forall x \in \mathbf{B}(\bar{x}, \delta) \cap D) (\forall y \in \mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T) \quad \|Tx - y\| \leq \|x - y\|.$$

The connection between quasi-ne mappings and quasi-fne mappings is given by the following fact.

**Fact 6.3** ([9, Proposition 2.3(v)  $\Leftrightarrow$  (vi)], [20, Corollary 2.1.33]) *Let  $D$  be a nonempty subset of  $\mathbb{R}^n$ , and  $T : D \rightarrow \mathbb{R}^n$  with  $\text{Fix } T \neq \emptyset$ . Then the following are equivalent:*

(i)  *$T$  is quasi-fne.*

(ii)  *$2T - \text{Id}$  is quasi-ne.*

The following result says that quasi-ne, nonexpansiveness, and local quasi-ne are the same for linear mappings. Although the equivalence of quasi-ne and nonexpansiveness for linear mappings has been given in [7, Exercise 4.4], the equivalence to local quasi-ne is new.

**Proposition 6.4** *Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a linear operator. Then the following are equivalent:*

(i)  $T$  is quasi-ne.

(ii)  $T$  is nonexpansive.

(iii) There exists  $\delta > 0$  and  $\bar{x} \in \text{Fix } T$  such that  $T$  is quasi-ne on  $\mathbf{B}(\bar{x}, \delta)$ .

*Proof.* (i) $\Rightarrow$ (ii). Since  $0 \in \text{Fix } T$ , we have  $\|Tx\| \leq \|x\|$  for every  $x \in \mathbb{R}^n$ . Hence  $T$  is nonexpansive.

(ii) $\Rightarrow$ (i). Clear.

(ii) $\Rightarrow$ (iii). Clear.

(iii) $\Rightarrow$ (ii). The assumption means that there exists  $\bar{x} \in \text{Fix } T$  and  $\delta > 0$  such that

$$(44) \quad (\forall x \in \mathbf{B}(\bar{x}, \delta))(\forall y \in \mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T) \quad \|Tx - y\| \leq \|x - y\|.$$

Let  $v \in \mathbf{B}(0, \delta)$ . Using  $T\bar{x} = \bar{x}$ ,  $y = \bar{x}$ , and  $T$  being linear, from (44) we obtain  $\|Tv\| = \|T(\bar{x} + v) - T\bar{x}\| = \|T(\bar{x} + v) - \bar{x}\| \leq \|(\bar{x} + v) - \bar{x}\| = \|v\|$ . Since  $v \in \mathbf{B}(0, \delta)$  was arbitrary and  $T$  is linear, we have  $\|Tv\| \leq \|v\|$  for every  $v \in \mathbb{R}^n$ . Hence  $T$  is nonexpansive.  $\blacksquare$

**Remark 6.5** Fact 6.3 and Proposition 6.4 hold in Hilbert spaces. We formulate them only in  $\mathbb{R}^n$ .

The following example illustrates that for nonlinear  $T$ , quasineonexpansiveness and nonexpansiveness are different.

**Example 6.6** Define

$$T : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} \frac{|x|}{2} \sin \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Then  $T$  is quasi-ne but not nonexpansive.

*Proof.*  $T$  is quasi-ne because that  $\text{Fix } T = \{0\}$  and

$$(\forall x \in \mathbb{R}) \quad |T(x) - 0| = \left| \frac{|x|}{2} \sin \frac{1}{x} \right| \leq \frac{|x|}{2} \leq |x|.$$

$T$  is not nonexpansive because for  $x > 0$ , we have

$$T'(x) = \frac{1}{2} \sin \frac{1}{x} - \frac{1}{2x} \cos \frac{1}{x}$$

and  $|T'(1/(2n\pi))| = n\pi > 1$ .  $\blacksquare$

For analogous results on linear cutters, see Proposition 9.1 in Section 9.

Although we have developed calculus for  $G_f$  being cutters in Section 5, most results also hold for quasi-ne mappings. We single out two most important ones.

**Theorem 6.7** Let  $k \geq 1$  and  $f : \mathbb{R}^n \rightarrow [0, +\infty)$  be locally Lipschitz, and let  $G_{f,s}$  be given by Definition 2.2. Suppose that  $G_{f,s}$  is quasi-ne and  $\text{Fix } G_{f,s} \neq \emptyset$ . If  $g = f^k$ , then  $G_{g, kf^{k-1}s}$  is quasi-ne.

*Proof.* Apply Theorem 3.9 and [7, Exercise 4.11].  $\blacksquare$

**Theorem 6.8** Let  $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be unitary and  $b \in \mathbb{R}^n$ , let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be lsc and subdifferentiable, and let  $G_{f,s}$  be given by Definition 2.2. Define  $g : \mathbb{R}^n \rightarrow \mathbb{R} : x \mapsto f(Ax + b)$ . If  $G_{f,s}$  is quasi-ne, then  $G_{g, A^T s(A \cdot + b)}$  is quasi-ne.

*Proof.* Apply Proposition 3.7 and Definition 6.1. ■

**Corollary 6.9** *Let  $k \geq 1$  and  $f : \mathbb{R}^n \rightarrow [0, +\infty)$  be locally Lipschitz with  $G_{f,s}$  be given by Definition 2.2. Suppose that  $g = f^2$  and  $\text{Fix } G_{f,s} \neq \emptyset$ . Then  $G_{f,s}$  is quasi-ne if and only if  $G_{g,2fs}$  is quasi-fne.*

*Proof.* By Theorem 3.9,  $G_{g,2fs} = \frac{G_{f,s} + \text{Id}}{2}$ . The result then follows from Fact 6.3. ■

## 6.2 Convergences of cutters, local cutters, quasi-ne mappings, and local quasi-ne mappings

**Proposition 6.10 (convergence of iterates of a cutter)** *Let  $D \subseteq \mathbb{R}^n$  be a nonempty closed convex set,  $T : D \rightarrow D$  be an operator with a fixed point, and that  $T$  has the fixed point closed property on  $D$ . If  $T$  is a cutter, then for every  $x \in D$ , the sequence  $(T^k x)_{k \in \mathbb{N}}$  converges to a point  $z \in \text{Fix } T$ .*

*Proof.* Since  $T$  is a cutter,  $T$  is quasi-fne by Fact 5.3(i), so quasi-ne. Moreover,  $T$  is asymptotically regular by [20, Theorem 3.4.3]. The result now follows from [20, Theorem 3.5.2]. ■

**Proposition 6.11 (convergence of iterates of a locally quasi-fne mapping)** *Let  $D$  be nonempty closed convex subset of  $\mathbb{R}^n$ , let  $T : D \rightarrow D$  and  $\text{Fix } T \neq \emptyset$ . Assume that*

- (i) *There exists  $\bar{x} \in \text{Fix } T$  and  $\delta > 0$  such that  $T$  is locally quasi-fne (see Definition 5.5);*
- (ii)  *$T$  has the fixed-point closed property.*

Let  $x_0 \in D \cap \mathbf{B}(\bar{x}, \delta)$ . Set

$$(\forall k \in \mathbb{N}) \quad x_{k+1} = Tx_k.$$

Then  $(x_k)_{k \in \mathbb{N}}$  converges to a point  $z \in \mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T$ .

*Proof.* By assumption (i),

$$(45) \quad (\forall x \in \mathbf{B}(\bar{x}, \delta) \cap D) (\forall y \in \mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T) \quad \|Tx - y\|^2 + \|Tx - x\|^2 \leq \|x - y\|^2.$$

With  $x_0 \in D \cap \mathbf{B}(\bar{x}, \delta)$ , equation (45) gives

$$(\forall y \in \mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T) \quad \|Tx_0 - y\|^2 + \|Tx_0 - x_0\|^2 \leq \|x_0 - y\|^2,$$

so  $\|x_1 - \bar{x}\| \leq \|x_0 - \bar{x}\| \leq \delta$ . By induction, we have that

$$(46) \quad (x_k)_{k \in \mathbb{N}} \text{ is a sequence in } \mathbf{B}(\bar{x}, \delta).$$

Moreover, equation (45) gives

$$(\forall k \in \mathbb{N}) (\forall y \in \mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T) \quad \|Tx_k - y\|^2 + \|Tx_k - x_k\|^2 \leq \|x_k - y\|^2,$$

so  $(x_k)_{k \in \mathbb{N}}$  is Fejér monotone with respect to  $C := \mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T$ , and  $\|x_{k+1} - x_k\| \rightarrow 0$  as  $k \rightarrow \infty$ .

Let  $x$  be a cluster point of  $(x_k)_{k \in \mathbb{N}}$ , say  $x_{k_l} \rightarrow x$ . Since  $Tx_{k_l} - x_{k_l} \rightarrow 0$ , and  $T$  is fixed-point closed, we have  $Tx - x = 0$ . Moreover,  $\|x - \bar{x}\| \leq \delta$  because of (46). Thus,  $x \in C$ . Applying [7, Theorem 5.5], we conclude that  $x_k \rightarrow z \in C$ . ■

**Proposition 6.12 (convergence of iterates of a locally quasi-ne mapping)** *Let  $D$  be nonempty closed convex subset of  $\mathbb{R}^n$ , let  $T : D \rightarrow D$  and  $\text{int}(\text{Fix } T) \neq \emptyset$ . Assume that*

- (i) *There exists  $\bar{x} \in \text{Fix } T$  and  $\delta > 0$  such that  $T$  is locally quasi-ne (see Definition 6.2);*
- (ii)  *$\text{int}(\mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T) \neq \emptyset$ ;*
- (iii)  *$T$  has the fixed-point closed property.*

Let  $x_0 \in D \cap \mathbf{B}(\bar{x}, \delta)$ . Set

$$(\forall k \in \mathbb{N}) \quad x_{k+1} = Tx_k.$$

Then  $(x_k)_{k \in \mathbb{N}}$  converges to a point  $z \in \mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T$ .

*Proof.* By assumption (i), there exists  $\delta > 0$  such that

$$(47) \quad (\forall x \in \mathbf{B}(\bar{x}, \delta) \cap D) (\forall y \in \mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T) \quad \|Tx - y\| \leq \|x - y\|;$$

With  $x_0 \in D \cap \mathbf{B}(\bar{x}, \delta)$ , equation (47) gives

$$(\forall y \in \mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T) \quad \|Tx_0 - y\| \leq \|x_0 - y\|,$$

so  $\|x_1 - \bar{x}\| \leq \|x_0 - \bar{x}\| \leq \delta$ . By induction, we have that

$$(48) \quad (x_k)_{k \in \mathbb{N}} \text{ is a sequence in } \mathbf{B}(\bar{x}, \delta).$$

Moreover, equation (47) gives

$$(\forall k \in \mathbb{N})(\forall y \in \mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T) \quad \|Tx_k - y\| \leq \|x_k - y\|,$$

so  $(x_k)_{k \in \mathbb{N}}$  is Fejér monotone with respect to  $C := \mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T$ . As  $\text{int } C \neq \emptyset$ , we have that  $x_k \rightarrow z \in \mathbb{R}^n$  by [7, Proposition 5.10]. This implies that  $\|Tx_k - x_k\| = \|x_{k+1} - x_k\| \rightarrow 0$  and  $x_k \rightarrow z$  as  $k \rightarrow \infty$ .

Since  $T$  is fixed-point closed, we have  $Tz - z = 0$ . Moreover,  $\|z - \bar{x}\| \leq \delta$  because of (48). Hence  $x_k \rightarrow z \in \mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T$ . ■

### 6.3 Applications to subgradient projectors

**Theorem 6.13** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be locally Lipschitz,  $G_{f,s}$  be given by Definition 2.2, and*

$$S := \{x \in \mathbb{R}^n \mid 0 \in \partial f(x)\} \cup \text{lev}_0 f \neq \emptyset.$$

*If the subgradient projector  $G_{f,s}$  is a cutter, then for every  $x \in \mathbb{R}^n$ , the sequence  $(G_{f,s}^k x)_{k \in \mathbb{N}}$  converges to a point  $z$  such that either  $0 \in \partial f(z)$  or  $f(z) \leq 0$ .*

*Proof.* Combine Theorem 4.9 and Proposition 6.10. ■

To proceed, it will be convenient to single out:

**Lemma 6.14** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be lsc and subdifferentiable, and  $G_{f,s}$  be given by Definition 2.2. When  $f(x) > 0$ ,  $0 \notin \partial f(x)$ , and  $y \in \mathbb{R}^n$ , we have*

$$(49) \quad \|G_{f,s}(x) - y\|^2 = \|x - y\|^2 + \frac{f(x)}{\|s(x)\|^2} (f(x) + 2 \langle y - x, s(x) \rangle).$$

*Proof.* This follows from

$$\begin{aligned}\|G_{f,s}(x) - y\|^2 &= \left\| x - y - \frac{f(x)}{\|s(x)\|^2} s(x) \right\|^2 \\ &= \|x - y\|^2 + \frac{f^2(x)}{\|s(x)\|^2} - 2 \left\langle x - y, \frac{f(x)}{\|s(x)\|^2} s(x) \right\rangle.\end{aligned}$$

■

**Theorem 6.15** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be locally Lipschitz,  $G_{f,s}$  be given by Definition 2.2, and

$$S := \{x \in \mathbb{R}^n \mid 0 \in \partial f(x)\} \cup \text{lev}_0 f.$$

Then the following hold:

(i)  $G_{f,s}$  is quasi-ne if and only if

$$(50) \quad (\forall x \notin S) (\forall y \in S) \quad f(x) + 2 \langle y - x, s(x) \rangle \leq 0.$$

(ii) Assume that  $\text{int } S \neq \emptyset$ , and (50) holds. Then for every  $x \in \mathbb{R}^n$ , the sequence  $(G_{f,s}^k x)_{k \in \mathbb{N}}$  converges to a point  $z \in S$ .

*Proof.* (i). By Lemma 6.14, when  $x \notin S$ , and  $y \in S$ , we have

$$(51) \quad \|G_{f,s}(x) - y\|^2 = \|x - y\|^2 + \frac{f(x)}{\|s(x)\|^2} (f(x) + 2 \langle y - x, s(x) \rangle).$$

In view of Definition 6.1, assumption (50) is equivalent to  $G_{f,s}$  being quasi-ne.

(ii). By (i), the sequence  $(G_{f,s}^k x)_{k \in \mathbb{N}}$  is Fejér monotone with respect to  $S$ . Since  $\text{int } S \neq \emptyset$ , by [7, Proposition 5.10], the sequence  $(G_{f,s}^k x)_{k \in \mathbb{N}}$  converges to a point  $z \in \mathbb{R}^n$ . Write  $x_k = G_{f,s}^k x$ . Then  $(\forall k \in \mathbb{N}) x_{k+1} = G_{f,s}(x_k)$ . As  $f$  is locally Lipschitz at  $z$  and  $x_k \rightarrow z$ , the sequence  $(s(x_k))_{k \in \mathbb{N}}$  is bounded. Since

$$x_{k+1} - x_k = -\frac{f(x_k)}{\|s(x_k)\|^2} s(x_k)$$

and  $\lim_{k \rightarrow \infty} x_k = z$ , we have

$$|f(z)| = \lim_{k \rightarrow \infty} |f(x_k)| = \lim_{k \rightarrow \infty} \|x_{k+1} - x_k\| \|s(x_k)\| = 0.$$

Hence  $z \in S$ .

■

**Theorem 6.16** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be locally Lipschitz, and  $G_{f,s}$  be given by Definition 2.2, and

$$S := \{x \in \mathbb{R}^n \mid 0 \in \partial f(x)\} \cup \text{lev}_0 f \neq \emptyset.$$

Assume that the subgradient projector  $G_{f,s}$  is locally quasi-fne at  $\bar{x} \in S$ , i.e., there exists  $\delta > 0$  such that

$$(52) \quad (\forall x \in \mathbf{B}(\bar{x}, \delta) \setminus S) (\forall y \in \mathbf{B}(\bar{x}, \delta) \cap S) \quad f(x) + \langle y - x, s(x) \rangle \leq 0.$$

Then for every  $x_0 \in \mathbf{B}(\bar{x}, \delta)$ , the sequence  $(x_k)_{k \in \mathbb{N}}$  defined by

$$(\forall k \in \mathbb{N}) \quad x_{k+1} = G_{f,s}(x_k)$$

converges to a point  $z \in \mathbf{B}(\bar{x}, \delta) \cap S$ .

*Proof.* (52) guarantees that  $G_{f,s}$  is locally quasi-fne at  $\bar{x} \in S$ . Indeed, when  $x \notin S$  and  $y \in S \cap \mathbf{B}(\bar{x}, \delta)$ , using Lemma 6.14, (52) and (13), we have

$$(53) \quad \|G_{f,s}(x) - y\|^2 = \|x - y\|^2 + \frac{f(x)}{\|s(x)\|^2} (f(x) + 2 \langle y - x, s(x) \rangle)$$

$$(54) \quad = \|x - y\|^2 + \frac{f^2(x)}{\|s(x)\|^2} \frac{2(f(x) + \langle y - x, s(x) \rangle) - f(x)}{f(x)}$$

$$(55) \quad \leq \|x - y\|^2 + \|G_f(x) - x\|^2 \left( \frac{-f(x)}{f(x)} \right)$$

$$(56) \quad \leq \|x - y\|^2 - \|G_{f,s}(x) - x\|^2.$$

In view of Theorem 4.9, it suffices to apply Proposition 6.11. ■

**Theorem 6.17** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be locally Lipschitz,  $G_{f,s}$  be given by Definition 2.2, and*

$$S := \{x \in \mathbb{R}^n \mid 0 \in \partial f(x)\} \cup \text{lev}_0 f$$

*with  $\text{int } S \neq \emptyset$ . Assume that there exist  $\bar{x} \in S$  and  $\delta > 0$  such that*

$$(57) \quad (\forall x \in \mathbf{B}(\bar{x}, \delta) \setminus S) (\forall y \in \mathbf{B}(\bar{x}, \delta) \cap S) f(x) + 2 \langle y - x, s(x) \rangle \leq 0.$$

*Assume further that  $\text{int}(\mathbf{B}(\bar{x}, \delta) \cap S) \neq \emptyset$ . Then for every  $x_0 \in \mathbf{B}(\bar{x}, \delta)$ , the sequence  $(x_k)_{k \in \mathbb{N}}$  defined by*

$$(\forall k \in \mathbb{N}) x_{k+1} = G_{f,s}(x_k)$$

*converges to a point  $z \in \mathbf{B}(\bar{x}, \delta) \cap S$ .*

*Proof.* (57) guarantees that  $G_{f,s}$  is locally quasi-ne at  $\bar{x}$ . Indeed, for every  $x \in \mathbf{B}(\bar{x}, \delta) \setminus S$  and  $y \in \mathbf{B}(\bar{x}, \delta) \cap S$ , using Lemma 6.14 and (57) we have

$$(58) \quad \|G_{f,s}(x) - y\|^2 = \|x - y\|^2 + \frac{f(x)}{\|s(x)\|^2} (f(x) + 2 \langle y - x, s(x) \rangle)$$

$$(59) \quad \leq \|x - y\|^2.$$

By Theorem 4.9,  $G_{f,s}$  has the fixed point closed property. Therefore, Proposition 6.12 applies. ■

**Example 6.18** (1). Define

$$f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} 1 - |x| & \text{if } |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Because  $\text{Fix } G_{f,s} = \{x \in \mathbb{R} \mid |x| \geq 1\}$  is not convex, we have that  $G_{f,s}$  is not a cutter. However,  $f$  satisfies the assumptions of both Theorems 6.16 and 6.17 so that the local convergence theory applies.

(2). Define

$$f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} 0 & \text{if } x \leq 0, \\ x & \text{if } 0 \leq x \leq 1, \\ 1 & \text{if } 1 \leq x \leq 2, \\ x - 1 & \text{if } x \geq 2. \end{cases}$$

As  $\text{Fix } G_{f,s} = (-\infty, 0] \cup [1, 2]$ ,  $G_{f,s}$  is not a cutter. However, both Theorems 6.16 and 6.17 apply.



## 6.4 Finite convergence and $(C, \varepsilon)$ -firmly nonexpansiveness

Finite termination algorithms for subgradient projectors of convex functions have been studied in [46, 29, 13]. Recently, in [43] Pang studied finite convergent algorithms of subgradient projectors of locally Lipschitz functions defined in terms of the Clarke subdifferential. Naturally, one asks what his result implies about the subgradient projector defined by us. To this end, let us recall lower- $C^k$  functions defined by Rockafellar and Wets [51, Definition 10.29], and approximate convex functions by Nghai, Luc, and Théra [41], respectively.

**Definition 6.19** A function  $f : O \rightarrow \mathbb{R}$ , where  $O$  is an open subset in  $\mathbb{R}^n$ , is said to be lower  $C^k$  on  $O$  if on some neighborhood  $V$  of each  $\bar{x} \in O$  there is a representation  $f(x) := \max_{t \in T} f_t(x)$  in which  $f_t$  is of class  $C^k$  on  $V$  and the index set  $T$  is compact such that  $f_t(x)$  and all its partial derivatives through order  $k$  depend continuously not just on  $x \in V$  but jointly on  $(t, x) \in T \times V$ .

**Definition 6.20** A function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is approximately convex at  $\bar{x} \in \mathbb{R}^n$  if for every  $\varepsilon > 0$  there exists  $\delta > 0$  such that

$$(\forall x, y \in \mathbf{B}(\bar{x}, \delta)) (\forall \lambda \in (0, 1)) \quad f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) + \varepsilon \lambda(1 - \lambda) \|x - y\|.$$

**Fact 6.21** (See [2, Theorem 4.5], [26, Corollary 3]) Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be locally Lipschitz at  $\bar{x}$ . Then the following are equivalent:

- (i)  $f$  is lower- $C^1$  around  $\bar{x}$ .
- (ii)  $f$  is approximately convex at  $\bar{x}$ .
- (iii) for every  $\varepsilon > 0$  there exists  $\delta > 0$  such that

$$(\forall x, y \in \mathbf{B}(\bar{x}, \delta)) (x^* \in \partial_c f(x)) \quad f(y) \geq f(x) + \langle x^*, y - x \rangle - \varepsilon \|x - y\|.$$

**Theorem 6.22 (finite convergence for accelerated subgradient projectors)** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be locally Lipschitz, and let  $\bar{x} \in \mathbb{R}^n$  satisfy

- (i)  $f(\bar{x}) = 0$ ;
- (ii)  $0 \notin \partial f(\bar{x})$ ;
- (iii)  $f$  is lower- $C^1$  around  $\bar{x}$ .

Suppose that the strictly decreasing sequence  $(\varepsilon_k)_{k \in \mathbb{N}}$  converges to 0 at a sublinear rate. Then there exist  $\delta > 0$  and  $\bar{\varepsilon} > 0$  such that for every  $x_0 \in \mathbf{B}(\bar{x}, \delta)$  and  $\varepsilon_0 < \bar{\varepsilon}$ , the sequence  $(x_k)_{k \in \mathbb{N}}$  defined by

$$(60) \quad (\forall k \in \mathbb{N}) \quad x_{k+1} = x_k - \frac{\varepsilon_k + f(x_k)}{\|s_k\|^2} s_k, \quad \text{where } s_k \in \partial f(x_k),$$

converges in finitely many iterations, i.e.,  $f(x_k) \leq 0$  for some  $k \in \mathbb{N}$ .

*Proof.* Since  $f$  is lower- $C^1$  around  $\bar{x}$ ,  $f$  is Clarke regular around  $\bar{x}$ ; see [51, Theorem 10.31]. Thus, the Clarke subdifferential and the limiting subdifferential of  $f$  are the same around  $\bar{x}$ . Because  $\partial f$  is upper semicontinuous, when  $\delta$  is sufficiently small, (ii) guarantees that  $0 \notin \partial f(x)$  for every  $x \in \mathbf{B}(\bar{x}, \delta)$ , which implies that (60) is well defined. By Fact 6.21, (iii) is equivalent to  $f$  being approximately convex at  $\bar{x}$ . The result then follows from [43, Theorem 3].  $\blacksquare$

**Remark 6.23** See [2, 41, 52] for more characterizations on lower- $C^1$  functions and approximately convex functions.

Let  $\varepsilon \geq 0$  and  $C \subseteq \mathbb{R}^n$ . In [30], Hesse and Luke studied  $(C, \varepsilon)$ -firmly nonexpansive mappings; see also [37].

**Definition 6.24** Let  $C, D$  be nonempty subsets of  $\mathbb{R}^n$  and  $T : D \rightarrow \mathbb{R}^n$ .  $T$  is called  $(C, \varepsilon)$ -firmly nonexpansive if

$$(\forall x \in D)(\forall y \in C) \|Tx - Ty\|^2 + \|(x - Tx) - (y - Ty)\|^2 \leq (1 + \varepsilon)\|x - y\|^2.$$

**Theorem 6.25** ( $(C, \varepsilon)$ -firmly nonexpansiveness of  $G_{f,s}$ ) Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be locally Lipschitz,  $G_{f,s}$  be given by Definition 2.2, and

$$S := \{x \in \mathbb{R}^n \mid 0 \in \partial f(x)\} \cup \text{lev}_0 f.$$

Suppose that  $\bar{x} \in \mathbb{R}^n$  satisfies

- (i)  $f(\bar{x}) = 0$ ;
- (ii)  $0 \notin \partial f(\bar{x})$ ;
- (iii)  $f$  is lower- $C^1$  around  $\bar{x}$ .

Then for every  $\varepsilon > 0$  there exists  $\delta > 0$  such that on  $\mathbf{B}(\bar{x}, \delta)$  the subgradient projector  $G_{f,s}$  is  $(S \cap \mathbf{B}(\bar{x}, \delta), \tilde{\varepsilon})$ -firmly nonexpansive, in which  $\tilde{\varepsilon} = 1 + 8L\varepsilon/d_{\partial f(\bar{x})}(0)^2$  and  $L$  being the Lipschitz modulus of  $f$  around  $\bar{x}$ .

*Proof.* Let  $\alpha := d_{\partial f(\bar{x})}(0)/2$ . Then  $\alpha > 0$  by (ii). For every  $\varepsilon > 0$ , we can find  $\delta > 0$  such that

•

$$(61) \quad f(y) \geq f(x) + \langle x^*, y - x \rangle - \varepsilon\|x - y\|, \text{ when } x, y \in \mathbf{B}(\bar{x}, \delta), x^* \in \partial f(x).$$

This follows from (iii) and Fact 6.21.

- $\|s(x)\| \geq \alpha$  whenever  $s(x) \in \partial f(x)$  and  $x \in \mathbf{B}(\bar{x}, \delta)$ . This is because that  $\partial f(\bar{x})$  is compact,  $\partial f$  is upper semicontinuous, and (ii).
- $|f(x) - f(y)| \leq L\|x - y\|$  whenever  $x, y \in \mathbf{B}(\bar{x}, \delta)$ . This is possible because  $f$  is locally Lipschitz around  $\bar{x}$ .

Since  $0 \notin \partial f(y)$  for  $y \in \mathbf{B}(\bar{x}, \delta)$ , we must have  $f(y) \leq 0$  if  $y \in S \cap \mathbf{B}(\bar{x}, \delta)$ . Thus, (61) gives

$$(62) \quad (\forall x \in \mathbf{B}(\bar{x}, \delta))(\forall y \in S \cap \mathbf{B}(\bar{x}, \delta))(\forall x^* \in \partial f(x)) \quad f(x) + \langle x^*, y - x \rangle \leq \varepsilon\|x - y\|.$$

Put  $C := S \cap \mathbf{B}(\bar{x}, \delta)$ . When  $f(x) > 0$ ,  $0 \notin \partial f(x)$ , and  $y \in S \cap \mathbf{B}(\bar{x}, \delta)$ , using Lemma 6.14, (62), and (13), we have

$$(63) \quad \|G_{f,s}(x) - y\|^2 = \|x - y\|^2 + \frac{f(x)}{\|s(x)\|^2} (f(x) + 2\langle y - x, s(x) \rangle)$$

$$(64) \quad = \|x - y\|^2 + \frac{f^2(x)}{\|s(x)\|^2} \frac{2(f(x) + \langle y - x, s(x) \rangle) - f(x)}{f(x)}$$

$$\begin{aligned}
(65) \quad & \leq \|x - y\|^2 + \frac{2f(x)}{\|s(x)\|^2} \varepsilon \|y - x\| + \|G_{f,s}(x) - x\|^2 \left( \frac{-f(x)}{f(x)} \right) \\
(66) \quad & \leq \|x - y\|^2 + \frac{2(f(x) - f(y))}{\alpha^2} \varepsilon \|x - y\| - \|G_{f,s}(x) - x\|^2 \\
(67) \quad & \leq \|x - y\|^2 + \frac{2L\|x - y\|}{\alpha^2} \varepsilon \|x - y\| - \|G_{f,s}(x) - x\|^2 \\
(68) \quad & = \|x - y\|^2 + \frac{2L\varepsilon}{\alpha^2} \|x - y\|^2 - \|G_{f,s}(x) - x\|^2 \\
(69) \quad & \leq (1 + \varepsilon) \|x - y\|^2 - \|G_{f,s}(x) - x\|^2.
\end{aligned}$$

This completes the proof. ■

**Remark 6.26** Observe that both Theorems 6.22 and 6.25 aim for solving nonconvex inequality problems, e.g., finding a point  $x$  such that  $f(x) \leq 0$  with  $f(x) = -e^{-x^2} + 1/2$  and  $f$  satisfying the assumptions at  $\bar{x} = \sqrt{\ln 2}$ . However, they do not apply to  $f = d_C$ .

This completes Part I. In Part II, we will study subgradient projectors of Moreau envelopes, and their connections to subgradient projectors of original functions.

## Part II

# Subgradient projectors of Moreau envelopes and characterizations

## 7 Subgradient projectors of Moreau envelopes

When  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  is lsc,  $\partial f(x)$  might be empty for some  $x \in \mathbb{R}^n$ . However,  $e_\lambda f$  has much better properties when  $f$  is prox-bounded, see, e.g., Fact 7.5 below; and this is the motivation for us to study subgradient projectors of Moreau envelopes below. To do this, we need to study the relationship between  $G_{e_\lambda f}$  and  $G_{f,s}$ .

Recall that for a proper, lsc function  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  and parameter value  $\lambda > 0$ , the Moreau envelope  $e_\lambda f$  and proximal mapping  $P_\lambda f$  are defined respectively by

$$\begin{aligned}
e_\lambda f : \mathbb{R}^n \rightarrow (-\infty, +\infty] : x \mapsto \inf_w \left\{ f(w) + \frac{1}{2\lambda} \|x - w\|^2 \right\}, \text{ and} \\
P_\lambda f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n : x \mapsto \operatorname{argmin}_w \left\{ f(w) + \frac{1}{2\lambda} \|x - w\|^2 \right\}.
\end{aligned}$$

When  $f$  is proper, lsc, and convex, we refer the reader to [7, Chapter 12] and [50] for the properties  $e_\lambda f$  and  $P_\lambda f$ . When  $f$  is a proper and lsc function, not necessarily convex, in [45] Poliquin and Rockafellar coined the notions of prox-boundedness and prox-regularity of functions; see also [51, page 610].

**Definition 7.1** (i) A function  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  is prox-bounded if there exists  $\lambda > 0$  such that  $e_\lambda f(x) > -\infty$  for some  $x \in \mathbb{R}^n$ . The supremum of the set of all such  $\lambda$  is the threshold  $\lambda_f$  of prox-boundedness for  $f$ .

(ii) A function  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  is prox-regular at  $\bar{x}$  for  $\bar{v}$  if  $f$  is finite and locally lsc at  $\bar{x}$  with  $\bar{v} \in \partial f(\bar{x})$ , and there exists  $\varepsilon > 0$  and  $\rho \geq 0$  such that

$$f(x') \geq f(x) + \langle v, x' - x \rangle - \frac{\rho}{2} \|x' - x\|^2 \quad \text{for all } \|x' - \bar{x}\| \leq \varepsilon$$

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

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

We give a simple example to illustrate the concepts of prox-regularity and prox-bounded of functions.

**Example 7.2** (1). The function  $f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto -|x|$  is prox-bounded with  $\lambda_f = +\infty$ . However,  $f$  is not prox-regular at  $x = 0$ .

(2). The function  $f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto x^3$  is prox-regular on  $\mathbb{R}$ . However,  $f$  is not prox-bounded.

(3). The function  $f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto -x^2/2$  is prox-regular on  $\mathbb{R}$ , and prox-bounded with  $\lambda_f = 1$ .

(4). The function  $f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto x^3 - |x|$  not prox-regular at  $x = 0$ , and not prox-bounded.

In the sequel, we shall also need the following key concepts.

**Definition 7.3** ([51, page 614]) Let  $O$  be a nonempty open subset of  $\mathbb{R}^n$ . We say that  $f : O \rightarrow \mathbb{R}$  is  $C^{1+}$  if  $f$  is differentiable with  $\nabla f$  Lipschitz continuous.

Set  $q : \mathbb{R}^n \rightarrow \mathbb{R} : x \mapsto \|x\|^2/2$ .

**Definition 7.4** ([51, page 567]) A proper, lsc function  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  is  $\mu$ -hypoconvex for some  $\mu > 0$  if  $f + \mu^{-1}q$  is convex.

## 7.1 Fine properties of prox-regular functions

Two major facts about the Moreau envelopes of prox-bounded functions and prox-regular functions are:

**Fact 7.5** ([51, Example 10.32]) Let  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  be proper, lsc, and prox-bounded with threshold  $\lambda_f$ . Then for every  $\lambda \in (0, \lambda_f)$ , the function  $-e_\lambda f$  is lower  $C^2$ , hence semidifferentiable, locally Lipschitz, Clarke regular, and

$$\begin{aligned} \partial[-e_\lambda f](x) &= \lambda^{-1}[\text{conv } P_\lambda f(x) - x], \\ \emptyset \neq \partial[e_\lambda f](x) &\subseteq \lambda^{-1}[x - P_\lambda f(x)]. \end{aligned}$$

**Fact 7.6** ([5, Proposition 5.3], [51, Proposition 13.37]) Let  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  be lsc, proper, and prox-bounded with threshold  $\lambda_f$ . Suppose that  $f$  is prox-regular at  $\bar{x}$  for  $\bar{v} \in \partial f(\bar{x})$ . Then for all  $\lambda \in (0, \lambda_f)$  there is a neighborhood  $U_\lambda$  of  $\bar{x} + \lambda\bar{v}$  for which the following equivalent properties hold:

(i)  $e_\lambda f$  is  $C^{1+}$  on  $U_\lambda$ .

(ii)  $P_\lambda f$  is nonempty, single-valued, monotone and Lipschitz continuous on  $U_\lambda$ . Further,

$$\nabla e_\lambda f = (\text{Id} - P_\lambda f) / \lambda \quad \text{on } U_\lambda.$$

**Proposition 7.7** *Let  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  be proper, lsc, and prox-bounded with threshold  $\lambda_f$ . Then for every  $\lambda \in (0, \lambda_f)$ , one has  $\text{dom } P_\lambda f = \mathbb{R}^n$ . Consequently,  $\text{ran}(\text{Id} + \lambda \partial f) = \mathbb{R}^n$ .*

*Proof.* As  $0 < \lambda < \lambda_f$ , we have  $\text{dom } P_\lambda f = \mathbb{R}^n$ . To complete the proof, it suffices to apply [51, Example 10.2]:  $P_\lambda f \subseteq (\text{Id} + \lambda \partial f)^{-1}$ . ■

**Proposition 7.8 (global prox-regularity implies hypoconvexity)** *Let  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  be proper, lsc, and prox-bounded with threshold  $\lambda_f$ . Suppose that  $f$  is prox-regular on  $\mathbb{R}^n$ . Then for every  $\lambda \in (0, \lambda_f)$ , the following hold:*

- (i) *The function  $f + \lambda^{-1} \mathfrak{q}$  is convex.*
- (ii)  *$P_\lambda f = (\text{Id} + \lambda \partial f)^{-1}$  is single-valued and Lipschitz continuous on  $\mathbb{R}^n$ .*
- (iii)  *$\nabla e_\lambda f = (\text{Id} - P_\lambda f) / \lambda$ .*

*Proof.* When  $\lambda \in (0, \lambda_f)$ , we have  $\text{dom } P_\lambda f = \mathbb{R}^n$ . Since  $f$  is prox-regular, by Fact 7.6, for  $v \in \partial f(x)$  there exists an open neighborhood  $U_\lambda$  of  $x + \lambda v$  such that  $P_\lambda f$  is single-valued and locally Lipschitz. Proposition 7.7 implies that  $P_\lambda f$  is single-valued and locally Lipschitz on  $\mathbb{R}^n$ . As  $P_\lambda f$  is always monotone, cf. [51, Proposition 12.19],  $P_\lambda f$  is maximally monotone by [51, Example 12.7]. Then (i) and (ii) follow from [51, Proposition 12.19]. To obtain (iii), one can apply Fact 7.6(ii). ■

Proposition 7.8(i) immediately implies:

**Corollary 7.9** *Let  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  be proper, lsc, and prox-bounded. Suppose that  $f$  is prox-regular on  $\mathbb{R}^n$ . Then the function  $f$  is a difference of two convex functions.*

Characterizations of prox-regularity on an open subset is given by

**Fact 7.10** ([51, Theorem 10.33], [51, Proposition 13.33]) *Let  $f : O \rightarrow \mathbb{R}$ , where  $O$  is a nonempty open set in  $\mathbb{R}^n$ . The following are equivalent:*

- (i) *The function  $f$  is lower  $C^2$  on  $O$ .*
- (ii) *Relative to some neighborhood of each point of  $O$ , there is an expression  $f = g - \rho \mathfrak{q}$  in which  $g$  is finite, convex function, and  $\rho > 0$ .*
- (iii)  *$f$  is prox-regular and locally Lipschitz on  $O$ .*

**Corollary 7.11** *Let  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  and let  $O$  be a nonempty open subset of  $\mathbb{R}^n$ . Suppose that  $f$  is prox-regular and locally Lipschitz on  $O$ . Then for every compact convex subset  $S$  of  $O$ , there exists  $\rho > 0$  such that  $f + \rho \mathfrak{q}$  is convex on  $S$ .*

*Proof.* Let  $x \in S$ . By Fact 7.10, there exists an open ball  $\mathbf{B}(x, \delta_x) \subseteq O$  and  $\rho_x$  such that  $f + \rho_x \mathfrak{q}$  is convex on  $\mathbf{B}(x, \delta_x)$ . Select from the covering of  $S$  by various balls  $\mathbf{B}(x, \delta_x)$  a finite covering, say  $\mathbf{B}(x_i, \delta_{x_i})$  with  $i = 1, \dots, m$ . Let  $\rho := \max\{\rho_{x_1}, \dots, \rho_{x_m}\}$ . As  $f + \rho \mathfrak{q}$  is convex on each  $\mathbf{B}(x_i, \delta_{x_i})$ , and  $S \subseteq \bigcup \mathbf{B}(x_i, \delta_{x_i})$ , we obtain that  $f + \rho \mathfrak{q}$  is convex on  $S$ . ■

## 7.2 Relationship among $(\partial e_\lambda f)^{-1}(0)$ , $\text{Fix } P_\lambda f$ and $(\partial f)^{-1}(0)$

**Proposition 7.12** *Let  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  be proper, lsc, and prox-bounded with threshold  $\lambda_f$ . Then for every  $\lambda \in (0, \lambda_f)$ , the following hold:*

- (i) *For every  $\alpha \in \mathbb{R}$ , the level set  $\text{lev}_\alpha f \neq \emptyset$  if and only if  $\text{lev}_\alpha(e_\lambda f) \neq \emptyset$ . Moreover,  $\text{lev}_\alpha(e_\lambda f) \supseteq \text{lev}_\alpha f$ .*
- (ii)  $0 \in \partial e_\lambda f(x) \Rightarrow x \in P_\lambda f(x) \Rightarrow 0 \in \partial f(x)$ .
- (iii) *If, in addition,  $f$  is prox-regular at  $\bar{x}$  for  $\bar{v} \in \partial f(\bar{x})$ , then on a neighborhood  $U_\lambda$  of  $\bar{x} + \lambda \bar{v}$  one has*

$$0 = \nabla e_\lambda f(x) \Leftrightarrow x = P_\lambda f(x).$$

*When  $f$  is prox-regular on  $\mathbb{R}^n$ , one has*

$$(\forall x \in \mathbb{R}^n) \quad 0 = \nabla e_\lambda f(x) \Leftrightarrow x = P_\lambda f(x) \Leftrightarrow 0 \in \partial f(x).$$

*Proof.*

(i). Since  $\inf f = \inf e_\lambda f$  and  $\text{argmin } f = \text{argmin } e_\lambda f$  by [51, Example 1.46],  $\text{lev}_\alpha f \neq \emptyset$  if and only if  $\text{lev}_\alpha(e_\lambda f) \neq \emptyset$  for every  $\alpha \in \mathbb{R}$ . The inclusion follows from  $e_\lambda f \leq f$ .

(ii). By Fact 7.5, we have  $\partial e_\lambda f(x) \subseteq \lambda^{-1}[x - P_\lambda f(x)]$ . This gives the first implication. By [51, Example 10.2],  $P_\lambda f(x) \subseteq (\text{Id} + \lambda \partial f)^{-1}(x)$  for all  $x \in \mathbb{R}^n$ . The second implication follows.

(iii). By Fact 7.6 or [45, Theorem 4.4], the Moreau envelope  $e_\lambda f$  is  $C^{1+}$  on a neighborhood  $U_\lambda$  of  $\bar{x} + \lambda \bar{v}$  with  $\nabla e_\lambda f = \lambda^{-1}[\text{Id} - P_\lambda f]$  on  $U_\lambda$ . When  $f$  is prox-regular on  $\mathbb{R}^n$ , one has  $\nabla e_\lambda f = \lambda^{-1}[\text{Id} - P_\lambda f]$ , and  $P_\lambda f = (\text{Id} + \lambda \partial f)^{-1}$  is single-valued on  $\mathbb{R}^n$  by Proposition 7.8. ■

**Fact 7.13** [51, Proposition 12.19] *For a proper, lsc function  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ , assume that  $f$  is  $\mu$ -hypoconvex for some  $\mu > 0$ . Then  $P_\mu f = (\text{Id} + \mu \partial f)^{-1}$ , and for all  $\lambda \in (0, \mu)$  the mapping  $P_\lambda f = (\text{Id} + \lambda \partial f)^{-1}$  is Lipschitz continuous with constant  $\mu / [\mu - \lambda]$ .*

Under the assumption of  $f$  being  $\mu$ -hypoconvex for some  $\mu > 0$ , when  $\lambda > 0$  is sufficiently small  $e_\lambda f$  gives rise to a smooth regularization of  $f$ .

**Proposition 7.14** *For a proper, lsc function  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ , assume that  $f$  is  $\mu$ -hypoconvex for some  $\mu > 0$ . Then for every  $\lambda \in (0, \mu)$ , the following hold:*

- (i)  $e_\lambda f$  is  $C^{1+}$  and  $\nabla e_\lambda f = \lambda^{-1}(\text{Id} - P_\lambda f)$  on  $\mathbb{R}^n$ .
- (ii)  $\nabla e_\lambda f(x) = 0 \Leftrightarrow 0 \in \partial f(x)$ .

*Proof.* As  $f$  is  $\mu$ -hypoconvex,  $f$  is prox-regular and prox-bounded. By Fact 7.13 and Fact 7.6,

$$\nabla e_\lambda f = \lambda^{-1}[\text{Id} - P_\lambda f] = \lambda^{-1}[\text{Id} - (\text{Id} + \lambda \partial f)^{-1}].$$

■

**Remark 7.15** Proposition 7.14(ii) can also be obtained from [33, Theorem 4.4], in which the authors study the Bregman envelope and proximal mapping of proper, lsc, and prox-bounded functions.

**Proposition 7.16** For a proper, lsc function  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ , assume that  $f := \max\{f_1, \dots, f_m\}$  with  $f_i$  being  $C^2$  and that  $f$  is prox-bounded below. Then for every  $\lambda > 0$  sufficiently small, one has

$$0 = \nabla e_\lambda f(x) \Leftrightarrow x = P_\lambda f(x) \Leftrightarrow 0 \in \partial f(x).$$

*Proof.* By [51, Proposition 13.33] or [45, Example 2.9],  $f$  is prox-regular everywhere on  $\mathbb{R}^n$ . By Proposition 7.8,  $P_\lambda f = (\text{Id} + \lambda \partial f)^{-1}$ . By Fact 7.6, we have  $\nabla e_\lambda f = \lambda^{-1}[\text{Id} - P_\lambda f]$ . It remains to apply Proposition 7.12(iii), and  $P_\lambda f = (\text{Id} + \lambda \partial f)^{-1}$  being single-valued. ■

For a sequence  $\{C_k\}_{k \in \mathbb{N}}$  of subsets of  $\mathbb{R}^n$ , its limit and outer limit are denoted respectively by  $\lim_{k \rightarrow \infty} C_k$  and  $\limsup_{k \rightarrow \infty} C_k$ ; see [51, page 109].

**Proposition 7.17** Let  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  be proper, lsc, and prox-bounded with threshold  $\lambda_f > 0$ . Assume that  $C \subseteq \mathbb{R}^n$  is nonempty and closed. Then for every  $\alpha \in \mathbb{R}$  one has  $\lim_{\lambda \downarrow 0} \text{lev}_\alpha(e_\lambda f + \iota_C) = \text{lev}_\alpha(f + \iota_C)$ .

*Proof.* In view of [51, Theorem 7.4(d)],  $e_\lambda f + \iota_C$  converges epigraphically to  $f + \iota_C$  when  $\lambda \downarrow 0$ . By [51, Proposition 7.7], for every  $\alpha \in \mathbb{R}$  there exists  $\alpha_\lambda \downarrow \alpha$  such that  $\lim_{\lambda \downarrow 0} \text{lev}_{\alpha_\lambda}(e_\lambda f + \iota_C) = \text{lev}_\alpha(f + \iota_C)$ . Since

$$\text{lev}_\alpha(f + \iota_C) \subseteq \text{lev}_\alpha(e_\lambda f + \iota_C) \subseteq \text{lev}_{\alpha_\lambda}(e_\lambda f + \iota_C),$$

we obtain  $\lim_{\lambda \downarrow 0} \text{lev}_\alpha(e_\lambda f + \iota_C) = \text{lev}_\alpha(f + \iota_C)$ . ■

### 7.3 The subgradient projector of $e_\lambda f$

The following result extends [12, Proposition 3.1(viii)] and [11, Example 4.9(ii)] from convex functions to possibly nonconvex functions.

**Theorem 7.18 (subgradient projector of Moreau envelopes of a prox-regular function)** Suppose that  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  is proper, lsc, and prox-bounded with threshold  $\lambda_f$ , and that  $f$  is prox-regular. Then for every  $\lambda \in (0, \lambda_f)$ , the subgradient projector of  $e_\lambda f$  is given by

$$G_{e_\lambda f} : \mathbb{R}^n \rightarrow \mathbb{R}^n : x \mapsto \begin{cases} x - \lambda \frac{e_\lambda f(x)}{\|x - P_\lambda f(x)\|^2} (x - P_\lambda f(x)) & \text{if } e_\lambda f(x) > 0 \text{ and } x \neq P_\lambda f(x), \\ x & \text{otherwise,} \end{cases}$$

and  $\text{Fix } G_{e_\lambda f} = \text{lev}_0(e_\lambda f) \cup \{x \in \mathbb{R}^n \mid x = P_\lambda f(x)\}$ . When  $x = P_\lambda f(x)$ , we have  $0 \in \partial f(x)$ . Moreover,  $\lim_{\lambda \downarrow 0} \text{lev}_0(e_\lambda f) = \text{lev}_0 f$ .

*Proof.* Apply Propositions 7.12, 7.8, and 7.17 with  $C = \mathbb{R}^n$  and  $\alpha = 0$ . ■

The restriction of  $G_f$  to a subset  $D \subseteq \mathbb{R}^n$  is denoted by  $G_f|_D$  and is the operator defined by

$$G_f|_D : D \rightarrow \mathbb{R}^n, \quad G_f|_D(x) = G_f(x) \text{ for every } x \in D.$$

**Theorem 7.19 (functions being prox-regular at the critical point)** Suppose that  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  is proper, lsc, and prox-bounded with threshold  $\lambda_f$ , and that  $f$  is prox-regular at  $\bar{x}$  for  $0 \in \partial f(\bar{x})$ . Then for every  $\lambda \in (0, \lambda_f)$ , there exists a closed neighborhood  $U_\lambda$  of  $\bar{x}$  for which

$$(70) \quad (\forall x \in U_\lambda) \quad G_{e_\lambda f}|_{U_\lambda}(x) = \begin{cases} x - \lambda \frac{e_\lambda f(x)}{\|x - P_\lambda f(x)\|^2} (x - P_\lambda f(x)) & \text{if } e_\lambda f(x) > 0 \text{ and } x \neq P_\lambda f(x), \\ x & \text{otherwise,} \end{cases}$$

$$(71) \quad \text{Fix } G_{e_\lambda f}|_{U_\lambda} = (\text{lev}_0(e_\lambda f) \cup \{x \in \mathbb{R}^n \mid x = P_\lambda f(x)\}) \cap U_\lambda, \text{ and}$$

$$(72) \quad x = P_\lambda f(x) \quad \Rightarrow \quad 0 \in \partial f(x).$$

Moreover,  $\limsup_{\lambda \downarrow 0} (\text{lev}_0(e_\lambda f) \cap U_\lambda) \subseteq \text{lev}_0 f$ .

*Proof.* Apply Fact 7.6 and Proposition 7.12(iii) to obtain (70)–(72). Since

$$(\text{lev}_0(e_\lambda f) \cap U_\lambda) \subseteq \text{lev}_0(e_\lambda f), \text{ and}$$

$$\lim_{\lambda \downarrow 0} \text{lev}_0(e_\lambda f) = \bigcap_{k \geq 1} \text{lev}_0(e_{1/k} f)$$

by [51, Exercise 4.3(b)], it suffices to use Proposition 7.17 with  $C = \mathbb{R}^n$  and  $\alpha = 0$ . ■

**Theorem 7.20 (subgradient projector of Moreau envelopes of a hypoconvex function)** *Suppose that  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  is proper and lsc, and that  $f$  is  $\mu$ -hypoconvex for some  $\mu > 0$ . Then for every  $\lambda \in (0, \mu)$ , the subgradient projector of  $e_\lambda f$  is given by*

$$G_{e_\lambda f} : \mathbb{R}^n \rightarrow \mathbb{R}^n : x \mapsto \begin{cases} x - \lambda \frac{e_\lambda f(x)}{\|x - P_\lambda f(x)\|^2} (x - P_\lambda f(x)) & \text{if } e_\lambda f(x) > 0 \text{ and } 0 \notin \partial f(x), \\ x & \text{otherwise,} \end{cases}$$

$$\text{Fix } G_{e_\lambda f} = \text{lev}_0(e_\lambda f) \cup \{x \in \mathbb{R}^n \mid 0 \in \partial f(x)\}, \text{ and}$$

$$\{x \in \mathbb{R}^n \mid 0 = \nabla e_\lambda f(x)\} = \{x \in \mathbb{R}^n \mid 0 \in \partial f(x)\}.$$

Moreover,  $\lim_{\lambda \downarrow 0} \text{lev}_0(e_\lambda f) = \text{lev}_0 f$ .

*Proof.* Apply Propositions 7.14 and 7.17 with  $C = \mathbb{R}^n$  and  $\alpha = 0$ . ■

Theorems 7.18 and 7.20 imply that if one can solve  $x_\lambda \in \text{Fix}(G_{e_\lambda f})$ , then either  $0 \in \partial f(x_\lambda)$  for some  $\lambda > 0$  or the subsequential limits of  $(x_\lambda)$  will lie in  $\text{lev}_0 f$  when  $\lambda \downarrow 0$ .

**Remark 7.21** Moreau envelopes of nonconvex functions in infinite dimensional spaces have been intensively studied; see, e.g., [5, 6, 32, 3]. Thus, it is possible to have analogues of Theorems 7.18, 7.19, 7.20 in infinite dimensional spaces. However, this is beyond the scope of this paper.

Cutters are important for studying convergence of iterative methods; see, e.g., [9, 20, 13]. It is natural to ask whether  $G_{e_\lambda f}$  is a cutter in the case that  $G_f$  is a cutter. Although we cannot answer this in general, the following special case is true.

**Proposition 7.22** *Let  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  be proper, lsc, and prox-regular. Suppose that  $\min f = 0$ ,  $f$  is strictly differentiable at every  $x \in \text{argmin } f$ , and that  $0 \notin \partial f(x)$  for every  $x \in \mathbb{R}^n \setminus \text{argmin } f$ . Then for every  $\lambda > 0$  the following hold:*

(i)  $\text{Fix } G_{e_\lambda f} = \text{Fix } G_f$ .

(ii) *If  $G_{f,s}$  is a cutter for every selection  $s$  of  $\partial f$ , then  $G_{e_\lambda f}$  is a cutter.*



*Proof.* As  $\min f > -\infty$ , the function  $f$  is prox-bounded with threshold  $r_f = +\infty$ .

(i). Note that  $\min f = \min e_\lambda f$ ,  $\operatorname{argmin} f = \operatorname{argmin} e_\lambda f$ . The assumption  $\min f = 0$  implies  $\operatorname{lev}_0 f = \operatorname{lev}_0 e_\lambda f = \operatorname{argmin} f$ . Because  $f$  is prox-regular on  $\mathbb{R}^n$  and  $r_f = +\infty$ , for every  $\lambda > 0$  we have  $\nabla e_\lambda f = \lambda^{-1}(\operatorname{Id} - \operatorname{Prox}_\lambda f)$  and  $\operatorname{Prox}_\lambda f = (\operatorname{Id} + \lambda \partial f)^{-1}$  being single-valued by Proposition 7.8. This gives

$$\{x \in \mathbb{R}^n \mid \nabla e_\lambda f(x) = 0\} = \{x \in \mathbb{R}^n \mid 0 \in \partial f(x)\}.$$

Then

$$\operatorname{Fix} G_{e_\lambda f} = \operatorname{lev}_0 e_\lambda f \cup \{x \in \mathbb{R}^n \mid \nabla e_\lambda f(x) = 0\} = \operatorname{lev}_0 f \cup \{x \in \mathbb{R}^n \mid 0 \in \partial f(x)\} = \operatorname{Fix} \mathcal{G}_f.$$

(ii). Assume that  $e_\lambda f(x) > 0$  and  $0 \neq \nabla e_\lambda f(x)$ . Since  $f$  is prox-regular,  $\nabla e_\lambda f(x) = \lambda^{-1}(x - \operatorname{Prox}_\lambda f(x)) \neq 0$ . By the definition of  $\operatorname{Prox}_\lambda f$ , we have

$$(73) \quad 0 \neq \lambda^{-1}(x - \operatorname{Prox}_\lambda f(x)) \in \partial f(\operatorname{Prox}_\lambda f(x)),$$

which implies that  $\operatorname{Prox}_\lambda f(x) \notin \operatorname{argmin} f$ . Indeed, if  $\operatorname{Prox}_\lambda f(x) \in \operatorname{argmin} f$ , the assumption gives  $\partial f(\operatorname{Prox}_\lambda f(x)) = \{0\}$  which contradicts (73). Thus,  $f(\operatorname{Prox}_\lambda f(x)) > 0$ . Because  $\operatorname{Prox}_\lambda f(x) \notin \operatorname{argmin} f$ , the assumption also gives  $0 \notin \partial f(\operatorname{Prox}_\lambda f(x))$ . These arguments, (i), Theorem 5.11, and  $G_{f,s}$  being a cutter altogether imply that

$$(74) \quad f(\operatorname{Prox}_\lambda f(x)) + \left\langle \lambda^{-1}(x - \operatorname{Prox}_\lambda f(x)), u - \operatorname{Prox}_\lambda f(x) \right\rangle \leq 0$$

if  $u \in \operatorname{Fix} G_{f,s}$ ,  $e_\lambda f(x) > 0$  and  $\nabla e_\lambda f(x) \neq 0$ .

Now we show that  $G_{e_\lambda f}$  is a cutter. Let  $u \in \operatorname{Fix} G_{e_\lambda f}$ ,  $e_\lambda f(x) > 0$ , and  $\nabla e_\lambda f(x) \neq 0$ . In view of (74) and (i), we calculate

$$\begin{aligned} & e_\lambda f(x) + \left\langle \lambda^{-1}(x - \operatorname{Prox}_\lambda f(x)), u - x \right\rangle \\ &= e_\lambda f(x) + \left\langle \lambda^{-1}(x - \operatorname{Prox}_\lambda f(x)), u - \operatorname{Prox}_\lambda f(x) \right\rangle + \left\langle \lambda^{-1}(x - \operatorname{Prox}_\lambda f(x)), \operatorname{Prox}_\lambda f(x) - x \right\rangle \\ &= f(\operatorname{Prox}_\lambda f(x)) + \frac{1}{2\lambda} \|x - \operatorname{Prox}_\lambda f(x)\|^2 + \left\langle \lambda^{-1}(x - \operatorname{Prox}_\lambda f(x)), u - \operatorname{Prox}_\lambda f(x) \right\rangle - \\ & \quad \lambda^{-1} \|x - \operatorname{Prox}_\lambda f(x)\|^2 \\ &= f(\operatorname{Prox}_\lambda f(x)) + \left\langle \lambda^{-1}(x - \operatorname{Prox}_\lambda f(x)), u - \operatorname{Prox}_\lambda f(x) \right\rangle - \frac{1}{2\lambda} \|x - \operatorname{Prox}_\lambda f(x)\|^2 \\ &\leq -\frac{1}{2\lambda} \|x - \operatorname{Prox}_\lambda f(x)\|^2 \leq 0. \end{aligned}$$

Theorem 5.11(i) concludes the proof. ■

A local version of Theorem 7.22 comes as follows.

**Proposition 7.23** *Let  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  be proper, lsc, and prox-regular at  $\bar{x}$  for  $\bar{v} = 0$ , and let*

$$S := \{x \in \mathbb{R}^n \mid 0 \in \partial f(x)\} \cup \operatorname{lev}_0 f.$$

*Suppose that  $\min f = 0$ , and there exists  $\delta > 0$  such that*

(i) *For every selection  $s$  of  $\partial f$ ,  $G_{f,s}$  is a cutter on  $\mathbf{B}(\bar{x}, \delta)$ , i.e.,*

$$(\forall x \in \mathbf{B}(\bar{x}, \delta) \setminus S)(\forall u \in S \cap \mathbf{B}(\bar{x}, \delta)) \quad f(x) + \langle s(x), u - x \rangle \leq 0.$$

- (ii)  $f$  is strictly differentiable at every  $u \in \operatorname{argmin} f \cap \mathbf{B}(\bar{x}, \delta)$ , and that  $0 \notin \partial f(x)$  for every  $x \in (\mathbb{R}^n \setminus \operatorname{argmin} f) \cap \mathbf{B}(\bar{x}, \delta)$ .

Then for every  $\lambda > 0$  there is a neighborhood of  $\bar{x}$  on which  $G_{e_\lambda f}$  is a cutter.

*Proof.* Because  $\min f > -\infty$ , the function  $f$  is prox-bounded with threshold  $r_f = +\infty$ . Since  $\min f = \min e_\lambda f$  and  $\operatorname{argmin} f = \operatorname{argmin} e_\lambda f$ , the assumption  $\min f = 0$  implies  $\operatorname{lev}_0 f = \operatorname{lev}_0 e_\lambda f = \operatorname{argmin} f$ . Because  $f$  is prox-regular at  $\bar{x}$  for  $\bar{v} = 0$ , and  $r_f = +\infty$ , by Proposition 7.6 for every  $\lambda > 0$  there exists  $\delta > \delta_1 > 0$  such that on  $\mathbf{B}(\bar{x}, \delta_1)$  the proximal mappings

$$(75) \quad P_\lambda f \text{ is Lipschitz continuous, } P_\lambda f(\bar{x}) = \bar{x},$$

and

$$(76) \quad \nabla e_\lambda f = \lambda^{-1}(\operatorname{Id} - \operatorname{Prox}_\lambda f).$$

By (75) there exists  $\delta_1 > \delta_2 > 0$  such that

$$(77) \quad P_\lambda f(x) \in \mathbf{B}(\bar{x}, \delta_1) \text{ when } x \in \mathbf{B}(\bar{x}, \delta_2).$$

*Claim 1.* For every  $u \in \operatorname{Fix} G_{f,s} \cap \mathbf{B}(\bar{x}, \delta_2)$  we have

$$(78) \quad f(\operatorname{Prox}_\lambda f(x)) + \left\langle \lambda^{-1}(x - \operatorname{Prox}_\lambda f(x)), u - \operatorname{Prox}_\lambda f(x) \right\rangle \leq 0$$

if  $e_\lambda f(x) > 0$ ,  $\nabla e_\lambda f(x) \neq 0$  and  $x \in \mathbf{B}(\bar{x}, \delta_2)$ .

Indeed, let  $e_\lambda f(x) > 0$  and  $0 \neq \nabla e_\lambda f(x)$  and  $x \in \mathbf{B}(\bar{x}, \delta_2)$ . In view of (76),  $\nabla e_\lambda f(x) = \lambda^{-1}(x - \operatorname{Prox}_\lambda f(x)) \neq 0$ . By the definition of  $\operatorname{Prox}_\lambda f$  or [45, Proposition 4.3(b)], we have

$$(79) \quad 0 \neq \lambda^{-1}(x - \operatorname{Prox}_\lambda f(x)) \in \partial f(\operatorname{Prox}_\lambda f(x)).$$

This implies that  $\operatorname{Prox}_\lambda f(x) \notin \operatorname{argmin} f$ . Suppose to the contrary that  $\operatorname{Prox}_\lambda f(x) \in \operatorname{argmin} f$ . Then the assumption (ii) and (77) give  $\partial f(\operatorname{Prox}_\lambda f(x)) = \{0\}$  which contradicts (166). Thus,

$$(80) \quad f(\operatorname{Prox}_\lambda f(x)) > 0.$$

Because  $\operatorname{Prox}_\lambda f(x) \notin \operatorname{argmin} f$  and (77), the assumption (ii) also ensures

$$(81) \quad 0 \notin \partial f(\operatorname{Prox}_\lambda f(x)).$$

Therefore, (78) follows from assumptions (i) and (ii).

*Claim 2.*  $G_{e_\lambda f}$  is a cutter on  $\mathbf{B}(\bar{x}, \delta_2)$ .

To this end, let  $u \in \operatorname{Fix} G_{e_\lambda f} \cap \mathbf{B}(\bar{x}, \delta_2)$ ,  $x \in \mathbf{B}(\bar{x}, \delta_2)$ ,  $e_\lambda f(x) > 0$ , and  $\nabla e_\lambda f(x) \neq 0$ . Then  $u \in \operatorname{Fix} G_f \cap \mathbf{B}(\bar{x}, \delta_2)$ ,  $f(\operatorname{Prox}_\lambda f(x)) > 0$ ,  $0 \notin \partial f(\operatorname{Prox}_\lambda f(x))$  by (80), (81). Using (78) we calculate

$$\begin{aligned} & e_\lambda f(x) + \left\langle \lambda^{-1}(x - \operatorname{Prox}_\lambda f(x)), u - x \right\rangle \\ &= e_\lambda f(x) + \left\langle \lambda^{-1}(x - \operatorname{Prox}_\lambda f(x)), u - \operatorname{Prox}_\lambda f(x) \right\rangle + \left\langle \lambda^{-1}(x - \operatorname{Prox}_\lambda f(x)), \operatorname{Prox}_\lambda f(x) - x \right\rangle \\ &= f(\operatorname{Prox}_\lambda f(x)) + \frac{1}{2\lambda} \|x - \operatorname{Prox}_\lambda f(x)\|^2 + \left\langle \lambda^{-1}(x - \operatorname{Prox}_\lambda f(x)), u - \operatorname{Prox}_\lambda f(x) \right\rangle - \end{aligned}$$

$$\begin{aligned}
& \lambda^{-1} \|x - \text{Prox}_\lambda f(x)\|^2 \\
&= f(\text{Prox}_\lambda f(x)) + \left\langle \lambda^{-1}(x - \text{Prox}_\lambda f(x)), u - \text{Prox}_\lambda f(x) \right\rangle - \frac{1}{2\lambda} \|x - \text{Prox}_\lambda f(x)\|^2 \\
&\leq -\frac{1}{2\lambda} \|x - \text{Prox}_\lambda f(x)\|^2 \leq 0.
\end{aligned}$$

Hence,  $G_{e_\lambda f}$  is a cutter on  $\mathbf{B}(\bar{x}, \delta_2)$  by Theorem 5.11(ii). ■

Is it possible that  $G_{e_\lambda f}$  is a cutter for every  $\lambda > 0$  but  $G_f$  is not a cutter? This is partially answered by the following result.

**Proposition 7.24** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be  $C^2$  and prox-bounded below. If  $G_{e_\lambda f}$  is a cutter for all sufficiently small  $\lambda > 0$ , then  $G_f$  is a cutter.*

*Proof.* Let  $\lambda > 0$  be sufficiently small. Proposition 7.8 yields that  $e_\lambda f$  is  $C^{1+}$ . Write

$$S := \{x \in \mathbb{R}^n \mid 0 \in \nabla f(x)\} \cup \text{lev}_0 f,$$

$$S_\lambda := \{x \in \mathbb{R}^n \mid \nabla e_\lambda f(x) = 0\} \cup \text{lev}_0 e_\lambda f.$$

Using that  $\nabla e_\lambda f(x) = 0 \Leftrightarrow 0 \in \partial f(x)$  by Proposition 7.16 and that  $e_\lambda f \leq f$ , we have  $S \subseteq S_\lambda$ . Since  $G_{e_\lambda f}$  is a cutter, by Theorem 5.11 we obtain

$$S_\lambda \subseteq \{u \in \mathbb{R}^n \mid e_\lambda f(x) + \langle \nabla e_\lambda f(x), u - x \rangle \leq 0\}$$

whenever  $e_\lambda f(x) > 0$  and  $\nabla e_\lambda f(x) \neq 0$ . It follows that

$$(82) \quad S \subseteq \{u \in \mathbb{R}^n \mid e_\lambda f(x) + \langle \nabla e_\lambda f(x), u - x \rangle \leq 0\}$$

whenever  $e_\lambda f(x) > 0$  and  $\nabla e_\lambda f(x) \neq 0$ . By [3, Theorem 3.10] or [32, Theorem 5.1],

$$(83) \quad \nabla f(x) = \limsup_{m \rightarrow \infty} \nabla e_{\lambda_m} f(x_m)$$

in which  $x_m \rightarrow x$ ,  $e_{\lambda_m} f(x_m) \rightarrow f(x)$ ,  $\lambda_m \downarrow 0$ . Whenever  $f(x) > 0$ ,  $\nabla f(x) \neq 0$ , (83) implies that for sufficiently large  $m$ , it holds that  $e_{\lambda_m} f(x_m) > 0$  and  $\nabla e_{\lambda_m} f(x_m) \neq 0$ . Then by (82),

$$(\forall u \in S) \quad e_{\lambda_m} f(x_m) + \langle \nabla e_{\lambda_m} f(x_m), u - x_m \rangle \leq 0.$$

Passing to the limit when  $m \rightarrow \infty$ , we have

$$(\forall u \in S) \quad f(x) + \langle \nabla f(x), u - x \rangle \leq 0.$$

Hence,  $G_f$  is a cutter by using Theorem 5.11 again. ■

## 7.4 The subgradient projector of $d_C$ when $C$ is prox-regular at a point

In this subsection, instead of functions we shall consider sets which are prox-regular at some points. Recall that a set  $C \subseteq \mathbb{R}^n$  is prox-regular at  $\bar{x} \in C$  for  $\bar{v} \in N_C(\bar{x})$  when  $\iota_C$  is prox-regular at  $\bar{x}$  for  $\bar{v}$ ; see [51, Exercise 13.31].

**Example 7.25** Let  $C \subseteq \mathbb{R}^n$  be closed and  $\bar{x} \in C$ . If  $C$  is prox-regular at  $\bar{x}$  for  $\bar{v} = 0$ , then there exists a neighborhood  $U$  of  $\bar{x}$  on which

- (i)  $P_C$  is single-valued and Lipschitz;
- (ii)  $P_C = (\text{Id} + T)^{-1}$  for some localization  $T$  of  $N_C$  around  $(\bar{x}, 0)$ ;
- (iii)  $d_C$  is strictly differentiable on  $U \setminus C$  with  $\nabla d_C = \frac{\text{Id} - P_C}{d_C}$ ;
- (iv)  $G_{d_C} = P_C$ ;
- (v)  $G_{d_C^2} = \frac{\text{Id} + P_C}{2}$ .

*Proof.* (i), (ii), and (iii) are given in [51, page 618].

To see (iv), let  $x \in U \setminus C$ . Since  $d_C(x) > 0$  and  $\nabla d_C(x) = \frac{x - P_C(x)}{d_C(x)} \neq 0$ , we have

$$\begin{aligned} G_{d_C}(x) &= x - \frac{d_C(x)}{\|\nabla d_C(x)\|^2} \nabla d_C(x) \\ &= x - (x - P_C(x)) = P_C(x). \end{aligned}$$

When  $x \in U \cap C$ ,  $G_{d_C}(x) = x = P_C(x)$ .

(v) follows from (iv) and Theorem 3.9. ■

**Remark 7.26** Sets which satisfy the assumption on  $C$  in Theorem 7.25 include convex sets, strongly amenable sets, etc; see, e.g., [51, page 442]. See also [1] for recent advances on prox-regular sets and uniformly prox-regular sets.

According to Example 7.25(iv), when  $C$  is prox-regular at  $\bar{x}$  for  $\bar{v} = 0$ , we have  $G_{d_C} = P_C$  around a neighborhood of  $\bar{x}$ . What happens if, in addition,  $G_{d_C}$  is a cutter or quasi nonexpansive on the neighborhood?

**Proposition 7.27** *Let  $C \subseteq \mathbb{R}^n$  be closed and  $\bar{x} \in C$ , and let  $C$  be prox-regular at  $\bar{x}$  for  $\bar{v} = 0$ . Suppose that there exists  $\delta > 0$  such that one of the following holds:*

- (i)  $P_C$  is a cutter on  $\mathbf{B}(\bar{x}, \delta)$ , i.e.,

$$(84) \quad (\forall x \in \mathbf{B}(\bar{x}, \delta)) (\forall u \in C \cap \mathbf{B}(\bar{x}, \delta)) \quad \langle x - P_C(x), u - P_C(x) \rangle \leq 0.$$

- (ii)  $P_C$  is quasi-ne on  $\mathbf{B}(\bar{x}, \delta)$ , i.e.,

$$(85) \quad (\forall x \in \mathbf{B}(\bar{x}, \delta)) (\forall u \in C \cap \mathbf{B}(\bar{x}, \delta)) \quad \|P_C(x) - u\| \leq \|x - u\|.$$

Then  $C \cap \mathbf{B}(\bar{x}, \delta)$  is convex.

*Proof.* By Example 7.25, there exists  $\delta > 0$  such that  $P_C$  is single-valued and Lipschitz on the closed ball  $\mathbf{B}(\bar{x}, \delta)$ .

- (i). Assume that (84) holds. On the one hand, (84) gives that

$$C \cap \mathbf{B}(\bar{x}, \delta) \subseteq \bigcap_{x \in \mathbf{B}(\bar{x}, \delta)} \{u \in \mathbf{B}(\bar{x}, \delta) \mid \langle x - P_C(x), u - P_C(x) \rangle \leq 0\}.$$

On the other hand, let

$$y \in \bigcap_{x \in \mathbf{B}(\bar{x}, \delta)} \{u \in \mathbf{B}(\bar{x}, \delta) \mid \langle x - P_C(x), u - P_C(x) \rangle \leq 0\}.$$

Then  $y \in \mathbf{B}(\bar{x}, \delta)$  and  $\langle x - P_C(x), y - P_C(x) \rangle \leq 0$  for every  $x \in \mathbf{B}(\bar{x}, \delta)$ . Taking  $x = y$  we have  $\langle y - P_C(y), y - P_C(y) \rangle \leq 0$ , which implies  $y = P_C(y)$ , so  $y \in C$ . Therefore,  $y \in \mathbf{B}(\bar{x}, \delta) \cap C$ . Hence

$$C \cap \mathbf{B}(\bar{x}, \delta) = \bigcap_{x \in \mathbf{B}(\bar{x}, \delta)} \{u \in \mathbf{B}(\bar{x}, \delta) \mid \langle x - P_C(x), u - P_C(x) \rangle \leq 0\},$$

and consequently  $C \cap \mathbf{B}(\bar{x}, \delta)$  is a convex set.

(ii). Similar arguments as (i) show that

$$C \cap \mathbf{B}(\bar{x}, \delta) = \bigcap_{x \in \mathbf{B}(\bar{x}, \delta)} \{u \in \mathbf{B}(\bar{x}, \delta) \mid \|P_C x - u\| \leq \|x - u\|\}.$$

To finish the proof, it suffices to observe that in the Euclidean space  $\mathbb{R}^n$ , for every  $x, y \in \mathbb{R}^n$  the set  $\{u \in \mathbb{R}^n \mid \|y - u\| \leq \|x - u\|\}$  is a half space when  $x \neq y$ , and the whole space  $\mathbb{R}^n$  if  $x = y$ . ■

**Proposition 7.28** *Let  $C \subseteq \mathbb{R}^n$  be closed and  $\bar{x} \in C$ . If there exists  $\delta > 0$  such that  $C \cap \mathbf{B}(\bar{x}, \delta)$  is convex, then there exists  $\delta_1 > 0$  such that  $P_C$  is a cutter on  $\mathbf{B}(\bar{x}, \delta_1)$ . Consequently,  $P_C$  is a quasi-fne on  $\mathbf{B}(\bar{x}, \delta_1)$ .*

*Proof.* Observe that for  $p \in P_C(x)$ ,

$$\|p - \bar{x}\| \leq \|p - x\| + \|x - \bar{x}\| \leq 2\|x - \bar{x}\|.$$

Thus, we can choose  $0 < \delta_1 < \delta$  sufficiently small, e.g.,  $\delta_1 < \delta/2$ , such that

$$\|x - \bar{x}\| < \delta_1 \text{ implies } (\forall p \in P_C(x)) \quad \|p - \bar{x}\| < \delta.$$

This implies that whenever  $x \in \mathbf{B}(\bar{x}, \delta_1)$ , we have  $P_C(x) \subseteq C \cap \mathbf{B}(\bar{x}, \delta)$ . Then

$$(86) \quad (\forall x \in \mathbf{B}(\bar{x}, \delta_1)) \quad P_C(x) = P_{C \cap \mathbf{B}(\bar{x}, \delta)}(x).$$

Because  $C \cap \mathbf{B}(\bar{x}, \delta)$  is closed and convex,  $P_{C \cap \mathbf{B}(\bar{x}, \delta)}$  is firmly nonexpansive on  $\mathbb{R}^n$ . From (86), we have that  $P_C$  is firmly nonexpansive on  $\mathbf{B}(\bar{x}, \delta_1)$ , that is,

$$(\forall x, y \in \mathbf{B}(\bar{x}, \delta_1)) \quad \|P_C(x) - P_C(y)\|^2 + \|(\text{Id} - P_C)(x) - (\text{Id} - P_C)(y)\|^2 \leq \|x - y\|^2.$$

It follows that

$$(\forall x \in \mathbf{B}(\bar{x}, \delta_1))(\forall y \in C \cap \mathbf{B}(\bar{x}, \delta_1)) \quad \|P_C(x) - y\|^2 + \|x - P_C(x)\|^2 \leq \|x - y\|^2.$$

Hence  $P_C$  is a cutter on  $\mathbf{B}(\bar{x}, \delta_1)$ . ■

## 8 Characterization of subgradient projectors of convex functions

Subgradient projectors of convex functions are quasi-fne, so algorithms developed in [20] or [7] can be applied; see also Theorem 6.13. Therefore, in practice, it is useful to have available some results on whether a mapping is a subgradient projector of a convex function. This is the goal of this section. The results in this section provide some checkable conditions for convergence of iterated subgradient projectors in Section 6.

The following result is of independent interest.

**Proposition 8.1** Let  $C \subseteq \mathbb{R}^n$  be closed and convex. Assume that the function  $f : \mathbb{R}^n \setminus C \rightarrow \mathbb{R}$  satisfies

- (i)  $f \geq 0$  on  $\mathbb{R}^n \setminus C$ ;
- (ii)  $f$  is convex on every convex subsets of  $\mathbb{R}^n \setminus C$ ;
- (iii) Whenever  $x \in \text{bdry}(\mathbb{R}^n \setminus C)$ , one has  $\lim_{\substack{y \rightarrow x \\ y \in \mathbb{R}^n \setminus C}} f(y) = 0$ . That is,  $\lim_{i \rightarrow \infty} f(y_i) = 0$  whenever  $(y_i)_{i \in \mathbb{N}}$  is a sequence in  $\mathbb{R}^n \setminus C$  converging to a boundary point  $x$  of  $\mathbb{R}^n \setminus C$ .

Define

$$g : \mathbb{R}^n \rightarrow \mathbb{R} : x \mapsto \begin{cases} f(x) & \text{if } x \notin C, \\ 0 & \text{if } x \in C. \end{cases}$$

Then  $g$  is convex on  $\mathbb{R}^n$ .

*Proof.* Let  $x, y \in \mathbb{R}^n$ ,  $0 \leq \lambda \leq 1$ . We need to show

$$(87) \quad g(\lambda x + (1 - \lambda)y) \leq \lambda g(x) + (1 - \lambda)g(y).$$

We consider three cases.

- (i). If  $[x, y] \subseteq \mathbb{R}^n \setminus C$ ,  $g = f$  is convex on  $[x, y]$  by the assumption.
- (ii). If  $\lambda x + (1 - \lambda)y \in C$ , then

$$g(\lambda x + (1 - \lambda)y) = 0 \leq \lambda g(x) + (1 - \lambda)g(y)$$

since  $g(x), g(y) \geq 0$ .

- (iii).  $\lambda x + (1 - \lambda)y \notin C$  and  $[x, y] \cap C \neq \emptyset$ . In particular,  $x, y$  cannot both be in  $C$ . We consider two subcases.

Subcase 1.  $x \in C$  and  $y \notin C$ . As  $y \notin C$ , there exists  $z \in \text{bdry}(C)$  such that

$$\lambda x + (1 - \lambda)y \in [z, y] \subseteq \mathbb{R}^n \setminus C \quad \text{and} \quad f(z) = 0.$$

Because

$$\lambda x + (1 - \lambda)y = \alpha z + (1 - \alpha)y \quad \text{for some } 0 \leq \alpha \leq 1,$$

and  $f$  is convex on  $[z, y]$ , we have

$$(88) \quad f(\lambda x + (1 - \lambda)y) = f(\alpha z + (1 - \alpha)y) \leq \alpha f(z) + (1 - \alpha)f(y) = (1 - \alpha)f(y).$$

Now

$$z = \beta x + (1 - \beta)y \quad \text{for some } 0 \leq \beta \leq 1, \text{ and}$$

$$\begin{aligned} \lambda x + (1 - \lambda)y &= \alpha z + (1 - \alpha)y = \alpha(\beta x + (1 - \beta)y) + (1 - \alpha)y \\ &= (\alpha\beta)x + (1 - \alpha\beta)y \end{aligned}$$

give  $\lambda = \alpha\beta$ . Therefore, by (88),  $g(x) = 0$  and  $g(y) = f(y) \geq 0$ ,

$$(89) \quad g(\lambda x + (1 - \lambda)y) = f(\lambda x + (1 - \lambda)y)$$

$$(90) \quad \leq (1 - \alpha\beta)f(y) = (1 - \lambda)g(y) + \lambda g(x),$$

which is (87).

Subcase 2.  $x \notin C$  and  $y \notin C$ . By the assumption, there exists  $z \in \text{bdry}(C)$  such that

$$\lambda x + (1 - \lambda)y \in [z, y] \quad \text{or} \quad \lambda x + (1 - \lambda)y \in [x, z],$$

say  $\lambda x + (1 - \lambda)y \in [z, y]$ . Then

$$\lambda x + (1 - \lambda)y = \alpha z + (1 - \alpha)y \quad \text{for some } 0 \leq \alpha \leq 1.$$

As  $f$  is convex on  $[z, y]$ ,  $f(z) = 0$ ,

$$(91) \quad g(\lambda x + (1 - \lambda)y) = f(\alpha z + (1 - \alpha)y) \leq \alpha f(z) + (1 - \alpha)f(y) = (1 - \alpha)f(y).$$

Now

$$z = \beta x + (1 - \beta)y \quad \text{for some } 0 \leq \beta \leq 1, \text{ and}$$

$$\begin{aligned} \lambda x + (1 - \lambda)y &= \alpha z + (1 - \alpha)y = \alpha(\beta x + (1 - \beta)y) + (1 - \alpha)y \\ &= (\alpha\beta)x + (1 - \alpha\beta)y \end{aligned}$$

give  $\lambda = \alpha\beta$ . Then by (91), using  $g(x) = f(x) \geq 0$ ,  $g(y) = f(y) \geq 0$ , we obtain

$$(92) \quad g(\lambda x + (1 - \lambda)y) \leq (1 - \alpha\beta)f(y) = (1 - \lambda)f(y)$$

$$(93) \quad \leq (1 - \lambda)f(y) + \lambda f(x)$$

$$(94) \quad = (1 - \lambda)g(y) + \lambda g(x),$$

which is (87).

Combining (i)–(iii), we conclude that  $g$  is convex on  $\mathbb{R}^n$ . ■

**Theorem 8.2** Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $C := \{x \in \mathbb{R}^n \mid Tx = x\}$  be closed convex. Then  $T$  is a subgradient projector of a convex function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  with  $\text{lev}_0 f = C$  if and only if there exists  $g : \mathbb{R}^n \rightarrow [-\infty, +\infty)$  such that  $g : \mathbb{R}^n \setminus C \rightarrow \mathbb{R}$  is locally Lipschitz,  $g(x) = -\infty$  for every  $x \in C$ , and

(i) for every  $x \in \mathbb{R}^n \setminus C$ ,  $\frac{x - Tx}{\|x - Tx\|^2} \in \partial g(x)$ ;

(ii) the function defined by

$$f(x) := \begin{cases} \exp(g(x)) & \text{if } x \notin C, \\ 0 & \text{if } x \in C, \end{cases}$$

is convex.

In this case,  $T = G_f$ .

*Proof.* “ $\Rightarrow$ ”: Assume that  $T$  is a subgradient projector, say  $T = G_{f_1}$  with  $f_1 : \mathbb{R}^n \rightarrow \mathbb{R}$  being convex and  $\text{lev}_0 f_1 = C$ . Then  $f = \max\{0, f_1\}$  is convex and  $G_f = G_{f_1}$ . Put  $g = \ln f$  and  $C = \text{lev}_0 f$ . Since  $f$  is locally Lipschitz,  $g$  is locally Lipschitz on  $\mathbb{R}^n \setminus C$ . Note that  $\partial g(x) = (\partial f(x))/f(x)$  when  $f(x) > 0$ . Apply Theorem 4.1(i) to obtain (i).

“ $\Leftarrow$ ”: Assume that (i), (ii) hold. When  $x \notin C$ , (i) and (ii) give

$$\|x - Tx\| = \frac{1}{\|c(x)\|}, \quad \partial g(x) = \frac{\partial f(x)}{f(x)}$$

where  $c(x) \in \partial g(x)$ . Using (i) again, we have

$$(95) \quad Tx = x - \frac{\|x - Tx\|^2 c(x)}{\|c(x)\|^2} = G_f(x)$$

by Theorem 4.1(ii). Moreover, when  $x \in C$ ,  $Tx = x = G_f(x)$ . Hence  $T = G_f$ . ■

For an  $n \times n$  symmetric matrix  $A$ , by  $A \succeq 0$  we mean that  $A$  is positive semidefinite.

**Theorem 8.3** *Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $C := \{x \in \mathbb{R}^n \mid Tx = x\}$ . Suppose that  $C$  is closed and convex, and  $T$  is continuously differentiable on  $\mathbb{R}^n \setminus C$ . Define*

$$T_1 : \mathbb{R}^n \setminus C \rightarrow \mathbb{R}^n : x \mapsto \frac{x - Tx}{\|x - Tx\|^2}.$$

*Then  $T$  is a subgradient projector of a convex function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  with  $\text{lev}_0 f = C$  and being differentiable on  $\mathbb{R}^n \setminus C$  if and only if*

(i) *For every  $x \in \mathbb{R}^n \setminus C$ , the matrix  $T_1(x)(T_1(x))^* + \nabla T_1(x) \succeq 0$ ;*

(ii) *There exists a function  $g : \mathbb{R}^n \rightarrow [-\infty, +\infty)$  such that*

$$\begin{aligned} & (\forall x \in \mathbb{R}^n \setminus C) \quad \nabla g(x) = T_1(x), \\ & (\forall x \in \text{bdry}(C)) \quad \lim_{\substack{y \rightarrow x \\ y \in \mathbb{R}^n \setminus C}} g(y) = -\infty, \quad \text{and } (\forall x \in C) \quad g(x) = -\infty. \end{aligned}$$

*Proof. “ $\Rightarrow$ ”:* Assume that  $T = G_f$  with  $f$  being convex and  $\text{lev}_0 f = C$ . Theorem 4.10 shows that  $f$  is continuously differentiable on  $\mathbb{R}^n \setminus C$ . By Theorem 4.1(i), we can put  $g = \ln f$  to obtain (ii). Moreover, as  $f = \exp(g)$ , thanks to (16) in Theorem 4.1, for every  $x \notin C$  we have

$$\begin{aligned} \nabla f(x) &= e^{g(x)} \nabla g(x) = e^{g(x)} T_1(x), \\ \nabla^2 f(x) &= e^{g(x)} T_1(x)(T_1(x))^* + e^{g(x)} \nabla T_1(x) = e^{g(x)} (T_1(x)(T_1(x))^* + \nabla T_1(x)). \end{aligned}$$

Since  $f$  is convex,  $\nabla^2 f(x) \succeq 0$ , and this is equivalent to

$$T_1(x)(T_1(x))^* + \nabla T_1(x) \succeq 0$$

which is (i).

*“ $\Leftarrow$ ”:* Assume that (i) and (ii) hold. Put  $f = \exp(g)$ . Then  $\text{lev}_0 f = C$ , and for  $x \in \mathbb{R}^n \setminus C$ ,

$$\begin{aligned} \nabla f(x) &= e^{g(x)} \nabla g(x) = e^{g(x)} T_1(x), \\ \nabla^2 f(x) &= e^{g(x)} T_1(x)(T_1(x))^* + e^{g(x)} \nabla T_1(x) = e^{g(x)} (T_1(x)(T_1(x))^* + \nabla T_1(x)). \end{aligned}$$

(i) and (ii) imply that  $f$  is differentiable and convex on convex subsets of  $\mathbb{R}^n \setminus C$ , and  $f \equiv 0$  on  $C$ . By Proposition 8.1,  $f$  is convex on  $\mathbb{R}^n$ .

Moreover, when  $x \neq Tx$  we have

$$(96) \quad G_f(x) = x - \left( \frac{f(x)}{\|\nabla f(x)\|} \right)^2 \frac{\nabla f(x)}{f(x)} = x - \frac{T_1(x)}{\|T_1(x)\|^2}$$

$$(97) \quad = x - \frac{\frac{x-Tx}{\|x-Tx\|^2}}{\left( \frac{1}{\|x-Tx\|} \right)^2} = x - (x - Tx) = Tx.$$

■



**Corollary 8.4** Let  $T : \mathbb{R} \rightarrow \mathbb{R}$  and  $C := \{x \in \mathbb{R} \mid Tx = x\}$ . Suppose that  $C$  is a closed interval, and  $T$  is continuously differentiable on  $\mathbb{R} \setminus C$ . Then  $T$  is a subgradient projector of a convex function  $f : \mathbb{R} \rightarrow \mathbb{R}$  with  $\text{lev}_0 f = C$  and being differentiable on  $\mathbb{R} \setminus C$  if and only if

(i)  $T$  is monotonically increasing on convex subsets of  $\mathbb{R} \setminus C$ ;

(ii) The function

$$g(x) = \int_a^x \frac{1}{s - Ts} ds$$

satisfies  $\lim_{x \downarrow \sup(C)} g(x) = -\infty$  for some  $a > \sup(C)$ ; and  $\lim_{x \uparrow \inf(C)} g(x) = -\infty$  for some  $a < \inf(C)$ .

*Proof.* Define  $n : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto x - Tx$ . Then for every  $x \notin C$ ,  $T_1(x) = \frac{1}{n(x)}$ . Theorem 8.3(i) is equivalent to

$$\frac{1}{n^2(x)} - \frac{n'(x)}{n^2(x)} \geq 0.$$

This is the same as  $n'(x) \leq 1$ , which transpires to  $T'(x) \geq 0$ . ■

**Remark 8.5** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be continuously differentiable,  $\text{lev}_0 f = \text{Fix } T$ , and  $G_f = T$ . Can one use  $T$  to decide whether  $f$  is convex? The proof of Theorem 8.3 implies that  $\nabla f = f \cdot T_1$  where

$$T_1 : \mathbb{R}^n \setminus \text{lev}_0 f \rightarrow \mathbb{R}^n : x \mapsto \frac{x - Tx}{\|x - Tx\|^2}.$$

If

$$(98) \quad f \cdot T_1 \text{ is monotone on convex subsets of } \mathbb{R}^n \setminus \text{lev}_0 f,$$

then  $f$  is convex on convex subsets of  $\mathbb{R}^n \setminus \text{lev}_0 f$ . Using Proposition 8.1, we conclude that  $\max\{0, f\}$  is convex on  $\mathbb{R}^n$ . When  $T$  is continuously differentiable, (98) is equivalent to

$$(99) \quad (\forall x \in \mathbb{R}^n \setminus C) \quad T_1(x)(T_1(x))^* + \nabla T_1(x) \succeq 0.$$

On  $\mathbb{R}$ , (99) is equivalent to

$$(100) \quad T \text{ is monotonically increasing on convex subsets of } \mathbb{R} \setminus C.$$

**Corollary 8.6** Let  $T : \mathbb{R} \rightarrow \mathbb{R}$  and  $C := \{x \in \mathbb{R} \mid Tx = x\}$ . Let  $C$  be a closed interval, and  $T$  be continuously differentiable on  $\mathbb{R} \setminus C$ . Define

$$N : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto x - Tx.$$

Suppose that

(i)  $N$  is nonexpansive;

(ii) The function

$$g(x) = \int_a^x \frac{1}{s - Ts} ds$$

satisfies  $\lim_{x \downarrow \sup(C)} g(x) = -\infty$  for some  $a > \sup(C)$ ; and  $\lim_{x \uparrow \inf(C)} g(x) = -\infty$  for some  $a < \inf(C)$ .

Then  $T$  is a subgradient projector of a convex function  $f : \mathbb{R} \rightarrow \mathbb{R}$  with  $\text{lev}_0 f = C$  and being differentiable on  $\mathbb{R} \setminus C$ . In particular, the assumption (i) holds when  $T$  is firmly nonexpansive.

*Proof.* It suffices to observe that  $T = \text{Id} - N$ . Since  $N$  is nonexpansive,  $T$  is monotone. Also note that  $T$  is firmly nonexpansive if and only if  $N$  is. ■

We illustrate Corollary 8.4 with three examples. They demonstrate that both conditions (i) and (ii) in Corollary 8.4 are needed. More precisely, (i) is for the convexity of  $f$ ; (ii) is for  $\text{lev}_0 f = C$ .

**Example 8.7** Define  $T : \mathbb{R} \rightarrow \mathbb{R}$  by

$$T(x) := \begin{cases} x - \sqrt{x} + \sqrt{x}e^{-2\sqrt{x}} & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

Then  $T$  is a subgradient projector of the nonconvex function

$$f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} e^{2\sqrt{x}} - 1 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

In this case,  $T$  fails to be monotone, but  $T$  verifies condition (ii) of Corollary 8.4.

*Proof.* When  $x > 0$ ,  $f'(x) = e^{2\sqrt{x}}x^{-1/2}$ , so that

$$f''(x) = \frac{e^{2\sqrt{x}}(1 - 1/(2\sqrt{x}))}{x}.$$

Since  $f''(x) < 0$  when  $x < 1/4$ ,  $f$  is not convex on  $\mathbb{R}$ . Now we show that

(i).  $T$  fails to be monotone. This is equivalent to verify that for some  $x$  we have  $N'(x) > 1$  where  $N(x) = x - Tx$ . Indeed,

$$N'(x) = \left( \sqrt{x} - \frac{\sqrt{x}}{e^{2\sqrt{x}}} \right)' = \frac{1}{2} \frac{e^{2\sqrt{x}} - 1}{e^{2\sqrt{x}}\sqrt{x}} + e^{-2\sqrt{x}}.$$

L'Hospital's rule gives

$$\lim_{x \rightarrow 0^+} \frac{e^{2\sqrt{x}} - 1}{e^{2\sqrt{x}}\sqrt{x}} = 2,$$

so  $\lim_{x \rightarrow 0^+} N'(x) = 2$ . Therefore,  $T$  is not monotone.

(ii).  $T$  satisfies condition (ii) of Corollary 8.4. For  $x > 0$ ,

$$N(x) = \frac{e^{2\sqrt{x}} - 1}{e^{2\sqrt{x}}x^{-1/2}}.$$

With  $a > 0$ , we have

$$g(x) = \int_a^x \frac{1}{N(s)} ds = \int_a^x \frac{e^{2\sqrt{s}}s^{-1/2}}{e^{2\sqrt{s}} - 1} dx = \ln(e^{2\sqrt{x}} - 1) - \ln(e^{2\sqrt{a}} - 1).$$

Clearly,  $\lim_{x \rightarrow 0^+} g(x) = -\infty$ . Hence (ii) holds. ■

**Example 8.8** Define

$$T : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} x - \frac{1}{2x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Then  $T = G_f$  where  $f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto e^{x^2}$ . However,  $\text{lev}_0 f = \emptyset$  but  $\text{Fix}(T) = \{0\}$ . In this case, in Corollary 8.4 condition (i) holds but condition (ii) fails.

*Proof.* We have

$$N(x) = x - T(x) = \frac{1}{2x}$$

and  $N'(x) = -\frac{1}{2x^2}$ . Therefore,  $T$  is monotone on  $(0, +\infty)$  and  $(-\infty, 0)$ . This says that condition (i) of Corollary 8.4 holds. However, when  $a > 0$ , for  $x > 0$  we have

$$g(x) = \int_a^x \frac{1}{N(x)} dx = \int_a^x 2x dx = x^2 - a^2.$$

Then  $\lim_{x \rightarrow 0^+} g(x) = -a^2$ , so condition (ii) of Corollary 8.4 fails. ■

**Example 8.9** Define  $T : \mathbb{R} \rightarrow \mathbb{R}$  by

$$x \mapsto \begin{cases} x - \sqrt{x} & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ x - \sqrt{-x} & \text{if } x < 0. \end{cases}$$

Then  $T = G_f$  where the nonconvex function

$$f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} e^{2\sqrt{x}} & \text{if } x \geq 0, \\ e^{-2\sqrt{-x}} & \text{if } x < 0. \end{cases}$$

However,  $\text{lev}_0 f = \emptyset$  but  $\text{Fix } T = \{0\}$ . In this case, both conditions (i) and (ii) in Corollary 8.4 fail.

*Proof.* The function  $f(x) = e^{2\sqrt{x}}$  is nonconvex on  $[0, +\infty)$ , see Example 8.7.  $G_f = T$  follows by direct calculations.

Condition (i) of Corollary 8.4 fails:  $T$  is not monotonically increasing on  $[0, +\infty)$  since  $T'(x) = 1 - \frac{1}{2\sqrt{x}} < 0$  when  $x > 0$  is sufficiently near 0.

Condition (ii) of Corollary 8.4 fails. Indeed,  $N(x) = x - T(x) = \sqrt{x}$  when  $x \geq 0$ . When  $a > 0$ , for  $x > 0$  we have

$$g(x) = \int_a^x \frac{1}{\sqrt{s}} ds = 2\sqrt{x} - 2\sqrt{a},$$

so that  $\lim_{x \rightarrow 0^+} g(x) = -2\sqrt{a}$ . ■

For further properties of subgradient projectors of convex functions, we refer the reader to [44, 54, 12].

This completes Part II. We will investigate conditions under which a subgradient projector is linear in part III.

## Part III

# Linear subgradient projectors

## 9 Characterizations of $G_{f,s}$ when $G_{f,s}$ is linear

We shall see in this section that under appropriate conditions a linear operator is a subgradient projector of a convex function if and only if it is a convex combination of the identity operator and a projection operator on a subspace (Theorems 9.6 and 9.11). For subgradient projectors of convex functions, see [12, 44, 9, 46, 47, 48]. We begin with

### 9.1 Linear cutters are precisely linear firmly nonexpansive mappings

**Proposition 9.1** *Let  $\mathcal{H}$  be a Hilbert space, and  $T : \mathcal{H} \rightarrow \mathcal{H}$  be a linear operator. Then the following are equivalent:*

- (i)  *$T$  is a cutter, i.e., quasi-firmly nonexpansive.*
- (ii)  *$T$  is firmly nonexpansive.*
- (iii) *There exists  $\delta > 0$  and  $\bar{x} \in \text{Fix } T$  such that  $T$  is a cutter on  $\mathbf{B}(\bar{x}, \delta)$ , i.e., a local cutter.*

*Proof.* (i) $\Rightarrow$ (ii). Assume that  $T$  is a cutter. Then for every  $x \in X$  and  $u \in \text{Fix } T$ ,

$$\langle x - Tx, u - Tx \rangle = \langle Tx - x, Tx - u \rangle \leq 0.$$

Put  $u = 0$ . We have

$$\langle Tx - x, Tx - 0 \rangle \leq 0 \quad \Rightarrow \quad \|Tx\|^2 \leq \langle x, Tx \rangle.$$

Hence  $T$  is firmly nonexpansive, see [7, Corollary 4.3].

(ii) $\Rightarrow$ (i). Assume that  $T$  is firmly nonexpansive. Let  $u \in \text{Fix } T$ . Then  $Tu = u$  and

$$(101) \quad \langle Tx - x, Tx - u \rangle = \langle Tx - x, Tx - Tu \rangle$$

$$(102) \quad = \langle Tx - Tu + Tu - x, Tx - Tu \rangle$$

$$(103) \quad = \|Tx - Tu\|^2 + \langle Tu - x, Tx - Tu \rangle$$

$$(104) \quad = \|Tx - Tu\|^2 - \langle x - u, Tx - Tu \rangle \leq 0.$$

Hence  $T$  is a cutter.

(iii) $\Rightarrow$ (ii). By the assumption

$$(\forall x \in \mathbf{B}(\bar{x}, \delta)) (\forall u \in \mathbf{B}(\bar{x}, \delta) \cap \text{Fix } T) \quad \langle x - Tx, u - Tx \rangle \leq 0.$$

As  $T\bar{x} = \bar{x}$ , and  $T$  is linear, for  $x = \bar{x} + v$  with  $\|v\| \leq \delta$ , we have

$$\begin{aligned} 0 &\geq \langle x - Tx, \bar{x} - Tx \rangle = \langle Tx - x, Tx - \bar{x} \rangle \\ &= \langle T(x - \bar{x}) - (x - \bar{x}), T(x - \bar{x}) \rangle = \|Tv\|^2 - \langle v, Tv \rangle. \end{aligned}$$

Since  $T$  is linear, we have  $\|Tx\|^2 \leq \langle x, Tx \rangle$  for every  $x \in X$ , so  $T$  is firmly nonexpansive, see [7, Corollary 4.3].

Since (i)  $\Leftrightarrow$  (ii), and (i) implies (iii), the proof is done. ■

The following example says that Proposition 9.1 fails if  $T$  is not linear.

**Example 9.2** Define the continuous nonlinear mapping

$$T : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} x/2 & \text{if } -2 \leq x \leq 2, \\ 3 - x & \text{if } 2 \leq x \leq 3, \\ -(3 + x) & \text{if } -3 \leq x \leq -2, \\ 0 & \text{otherwise.} \end{cases}$$

Then  $T$  is a cutter, nonexpansive, but not firmly nonexpansive as  $T$  is not monotone; cf. [7, Proposition 4.2(iv)].

Indeed,  $\text{Fix } T = \{0\}$ . This means that  $T$  is a cutter if and only if  $(T(x))^2 \leq xT(x)$ . When  $-2 \leq x \leq 2$ , we have

$$(T(x))^2 = \frac{x^2}{4} \leq x \frac{x}{2} = xT(x);$$

When  $2 \leq x \leq 3$ ,

$$(T(x))^2 = (3 - x)^2 = (3 - x)(3 - x) \leq x(3 - x);$$

when  $-3 \leq x \leq -2$ ,

$$(T(x))^2 = [-(x + 3)]^2 = [-(x + 3)][-(x + 3)] \leq x[-(3 + x)];$$

when  $|x| > 3$ ,

$$(T(x))^2 = 0 = xT(x).$$

Hence  $T$  is a cutter. Clearly,  $T$  is nonexpansive. As  $T$  is not monotone, we conclude that  $T$  is not firmly nonexpansive.

**Remark 9.3** Observe that Example 9.2 is much simpler than the example on  $\mathbb{R}^2$  constructed by Cegielski [20, Example 2.2.8, page 68].

## 9.2 Subgradient projector of powers of a quadratic function

It is natural to investigate subgradient projectors of quadratic functions or their variants first. In the following result, we assume  $B \neq 0$  because that  $B = 0$  gives  $G_f = \text{Id}$  with  $f \equiv 0$ .

**Theorem 9.4** *Let  $a > 0$  and  $B \neq 0$  being an  $n \times n$  symmetric and positive semidefinite matrix. Consider the function*

$$f : \mathbb{R}^n \rightarrow \mathbb{R} : x \mapsto (x^* B x)^{1/(2a)}.$$

*Then the following hold:*

(i)  $\text{lev}_0 f = \{x \in \mathbb{R}^n \mid Bx = 0\}$ .

(ii) *We have*

$$G_f(x) = \begin{cases} x - a \frac{x^* B x}{\|Bx\|^2} Bx & \text{if } Bx \neq 0, \\ x & \text{if } Bx = 0. \end{cases}$$

(iii)  $G_f$  is linear if and only if  $B = \lambda P_L$  where  $\lambda > 0$  and  $L \subseteq X$  is a subspace. In this case

$$\ker B = L^\perp,$$

$$f(x) = \lambda^{1/(2a)} (d_{L^\perp}(x))^{1/a} \text{ and}$$

$$G_f = \text{Id} - aP_L = (1 - a) \text{Id} + aP_{L^\perp}.$$

(iv) Assume that  $G_f$  is linear. Then  $G_f$  is a cutter if and only if  $0 < a \leq 1$ .

*Proof.* (i). Since  $B$  is symmetric and positive semidefinite, there exists a matrix  $A$  such that  $B = A^*A$ ; see, e.g., [38, page 558]. Then  $Ax = 0 \Leftrightarrow Bx = 0$ . The result follows because  $f(x) = \|Ax\|^{1/a}$ .

(ii).  $G_f$  follows from direct calculations.

(iii). " $\Rightarrow$ ": Assume that  $G_f$  is linear. The mapping

$$x \mapsto T_1(x) := a^{-1}(x - G_f(x)) = \begin{cases} \frac{x^*Bx}{\|Bx\|^2} Bx & \text{if } Bx \neq 0, \\ 0 & \text{if } Bx = 0, \end{cases}$$

is linear. Let  $\lambda_1, \lambda_2 > 0$  be any two eigenvalues of  $B$ . We show that  $\lambda_1 = \lambda_2$ . Suppose that  $\lambda_1 \neq \lambda_2$ . Take unit length eigenvector  $v_i$  associated with  $\lambda_i$ . Note that  $\langle v_1, v_2 \rangle = 0$ ,  $Bv_i \neq 0$  and  $B(v_1 + v_2) = \lambda_1 v_1 + \lambda_2 v_2 \neq 0$ . As  $T_1$  is linear, we have  $T_1(v_1 + v_2) = T_1 v_1 + T_1 v_2$ . Now

$$(105) \quad T_1(v_1 + v_2) = \frac{(v_1 + v_2)^* B(v_1 + v_2)}{\|B(v_1 + v_2)\|^2} B(v_1 + v_2)$$

$$(106) \quad = \frac{(v_1 + v_2)^* (\lambda_1 v_1 + \lambda_2 v_2)}{\|\lambda_1 v_1 + \lambda_2 v_2\|^2} (\lambda_1 v_1 + \lambda_2 v_2)$$

$$(107) \quad = \frac{\lambda_1 + \lambda_2}{\lambda_1^2 + \lambda_2^2} (\lambda_1 v_1 + \lambda_2 v_2),$$

$$(108) \quad T_1 v_1 + T_1 v_2 = \frac{v_1^* B v_1}{\|B v_1\|^2} B v_1 + \frac{v_2^* B v_2}{\|B v_2\|^2} B v_2$$

$$(109) \quad = \frac{\lambda_1 \|v_1\|^2}{\|\lambda_1 v_1\|^2} \lambda_1 v_1 + \frac{\lambda_2 \|v_2\|^2}{\|\lambda_2 v_2\|^2} \lambda_2 v_2$$

$$(110) \quad = v_1 + v_2.$$

As  $\{v_1, v_2\}$  are linearly independent, the above gives  $\lambda_1 = \lambda_2$  which contradicts  $\lambda_1 \neq \lambda_2$ . Therefore, all positive eigenvalues of  $B$  have to be equal. Hence, we have

$$B = \lambda U^* \begin{pmatrix} \text{Id} & 0 \\ 0 & 0 \end{pmatrix} U$$

where  $U$  is an orthogonal matrix,  $\lambda > 0$ ,  $\text{Id}$  is an  $m \times m$  identity matrix with  $m = \text{rank } B$ . The matrix

$$U^* \begin{pmatrix} \text{Id} & 0 \\ 0 & 0 \end{pmatrix} U$$

is idempotent and symmetric, so it is a matrix associated with an orthogonal projection onto a closed subspace, say  $P_L$ , [38, page 430, page 433]. Hence

$$B = \lambda P_L$$

which implies that  $Bx = 0$  if and only if  $P_Lx = 0$ , i.e.,  $\ker B = L^\perp$ . Then when  $P_Lx \neq 0$ ,

$$T_1(x) = \frac{x^*Bx}{\|Bx\|^2}Bx = \frac{\lambda x^*P_Lx}{x^*\lambda P_L\lambda P_Lx}\lambda P_Lx = \frac{\lambda x^*P_Lx}{\lambda^2 x^*P_Lx}\lambda P_Lx = P_Lx;$$

when  $P_Lx = 0$ ,  $T_1x = 0 = P_Lx$ . Hence  $T_1 = P_L$ . It follows that

$$G_f = \text{Id} - aT_1 = \text{Id} - aP_L = (1 - a)\text{Id} + a(\text{Id} - P_L) = (1 - a)\text{Id} + aP_{L^\perp}.$$

We proceed to find the expression for  $f(x)$ :

$$(111) \quad f(x) = (x^*Bx)^{1/(2a)} = (x^*\lambda P_Lx)^{1/(2a)}$$

$$(112) \quad = \lambda^{1/(2a)}(x^*P_LP_Lx)^{1/(2a)} = \lambda^{1/(2a)}(\|P_Lx\|^2)^{1/(2a)}$$

$$(113) \quad = \lambda^{1/(2a)}(\|x - P_{L^\perp}x\|^2)^{1/(2a)} = \lambda^{1/(2a)}(d_{L^\perp}(x)^2)^{1/(2a)}$$

$$(114) \quad = \lambda^{1/(2a)}(d_{L^\perp}(x))^{1/a}.$$

“ $\Leftarrow$ ”: Assume that  $B = \lambda P_L$  for  $\lambda > 0$  and some subspace  $L \subseteq \mathbb{R}^n$ . The assumption gives

$$f(x) = \lambda^{1/(2a)}(d_{L^\perp}(x))^{1/a}.$$

By Proposition 3.4(i),  $G_f = G_{(d_{L^\perp})^{1/a}}$ . By Theorem 3.9,  $G_f = (1 - a)\text{Id} + aG_{d_{L^\perp}}$ . By Fact 2.8,

$$G_f = (1 - a)\text{Id} + aP_{L^\perp}.$$

Hence  $G_f$  is linear.

(iv). “ $\Rightarrow$ ”: Assume that  $G_f$  is linear and a cutter. By Fact 9.1,  $G_f$  is firmly nonexpansive, so is  $\text{Id} - G_f$ . By (ii)  $\text{Id} - G_f = aP_L$ ,  $aP_L$  has to be nonexpansive. Because  $B \neq 0$ , we have  $L \neq \{0\}$ . Take  $0 \neq x \in L$ . The nonexpansiveness requires

$$\|aP_Lx - aP_L0\| = \|ax\| \leq \|x\|$$

so that  $a \leq 1$ .

“ $\Leftarrow$ ”: Assume that  $0 < a \leq 1$ . Since  $x \mapsto (x^*Bx)^{1/2}$  is convex, and the function  $[0, +\infty) \ni t \mapsto t^{1/a}$  is convex and increasing when  $0 < a \leq 1$ , we have that  $x \mapsto f(x) = ((x^*Bx)^{1/2})^{1/a}$  is convex. Then  $G_f$  is a cutter by Fact 5.12.  $\blacksquare$

We illustrate Theorem 9.4(iv) with the following example.

**Example 9.5** Let  $a > 1$ . Consider

$$f : \mathbb{R}^n \rightarrow \mathbb{R} : x \mapsto (x^*x)^{1/(2a)} = \|x\|^{1/a}.$$

Then  $f$  is not convex, and  $G_f(x) = (1 - a)x$  for every  $x \in \mathbb{R}^n$ . Although  $G_f$  is linear, it is not a cutter since it is not monotone; see, e.g., Proposition 9.1.

### 9.3 Symmetric and linear subgradient projectors

The following result completely characterizes symmetric and linear subgradient projectors.

**Theorem 9.6** *Assume that  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is linear and symmetric. Then the following are equivalent:*

- (i)  $T$  is a subgradient projector of a convex function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  with  $\text{lev}_0 f \neq \emptyset$ .
- (ii)  $T = G_f$  where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is given by

$$(115) \quad f(x) = K(x^* P_L x)^{1/(2\lambda)} = K(d_{L^\perp}(x))^{1/\lambda}$$

where  $0 < \lambda \leq 1$ ,  $K > 0$ , and  $L \subseteq \mathbb{R}^n$  is a subspace such that  $L^\perp = \text{Fix } T$ . In this case,  $G_f = (1 - \lambda) \text{Id} + \lambda P_{L^\perp}$ .

*Proof.* (i) $\Rightarrow$ (ii). Assume that  $T = G_f$  for some convex function. Since  $T$  is linear and a cutter,  $T$  is firmly nonexpansive by Proposition 9.1. Then  $T_1 = \text{Id} - T$  is firmly nonexpansive by [7, Proposition 4.2].

We consider two cases.

Case 1.  $\text{int lev}_0 f \neq \emptyset$ . We have  $T_1 \equiv 0$  on an open set  $\mathbf{B}(x_0, \varepsilon) \subseteq \text{lev}_0 f$ , i.e.,

$$T_1(x_0 + b) = 0 \text{ for every } \|b\| < \varepsilon.$$

As  $T_1$  is linear,  $T_1(b) = T_1(x_0 + b) - T_1(x_0) = 0 - 0 = 0$  when  $\|b\| < \varepsilon$ , so  $T_1 \equiv 0$  on  $\mathbb{R}^n$ . Thus,  $T = \text{Id}$  on  $\mathbb{R}^n$ . Then  $T = G_f$  with  $f \equiv 0$ . This means that (ii) holds with  $L = \{0\}$ ,  $\lambda = 1$  and  $K > 0$ .

Case 2.  $\text{int lev}_0 f = \emptyset$ . Since  $\text{lev}_0 f$  is a proper subspace, it is an intersection of a finite collection of hyper-planes [50, Corollary 1.4.1], so  $\mathbb{R}^n \setminus \text{lev}_0 f$  is union of a finite collection of open half spaces. As  $T_1$  is continuous, we only need to consider

$$T_1(x) = \frac{f(x)}{\|\nabla f(x)\|^2} \nabla f(x) \quad \text{when } f(x) > 0.$$

Then

$$\|T_1(x)\| = \frac{f(x)}{\|\nabla f(x)\|} \quad \text{and} \quad \frac{\nabla f(x)}{f(x)} = \frac{T_1(x)}{\|T_1(x)\|^2}.$$

Since  $T$  is symmetric,  $T_1$  is symmetric, so there exists an orthogonal matrix  $Q$  such that  $Q^* T_1 Q = D$  where  $D$  is a diagonal matrix and  $Q^*$  denotes the transpose of  $Q$ . Put  $g = \ln f$  and  $x = Qy$ . When  $y \notin Q^*(\text{Fix } T)$ , we have

$$(\nabla g)(Qy) = \frac{T_1 Qy}{\|T_1 Qy\|^2}.$$

Multiplying both sides by  $Q^*$  and using  $Q^*$  being an isometry (i.e.,  $\|Q^*z\| = \|z\|$  for every  $z \in \mathbb{R}^n$ ) give

$$Q^*(\nabla g)(Qy) = \frac{Q^* T_1 Qy}{\|T_1 Qy\|^2} = \frac{Dy}{\|Q^* T_1 Qy\|^2} = \frac{Dy}{\|Dy\|^2}.$$

If we put  $h = g \circ Q$ , then  $\nabla h(y) = Q^* \nabla g(Qy)$  for every  $y \in \mathbb{R}^n \setminus (Q^* \text{Fix } T)$ , so

$$(\forall y \in \mathbb{R}^n \setminus (Q^* \text{Fix } T)) \quad \nabla h(y) = \frac{Dy}{\|Dy\|^2}.$$



Moreover,  $\mathbb{R}^n \setminus (Q^* \text{Fix } T)$  is a finite union of open half spaces, because  $Q^* \text{Fix } T$  is a proper subspace of  $\mathbb{R}^n$ . Write

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

When  $\lambda_1 = \cdots = \lambda_n = 0$ , this is covered in Case 1. We thus assume that  $T_1 \neq 0$ . As  $T_1$  is monotone, we can and do assume that  $\lambda_1, \dots, \lambda_m > 0$  and  $\lambda_{m+1} = \cdots = \lambda_n = 0$ . Then

$$\nabla h(y) = \left( \frac{\lambda_1 y_1}{\sum_{k=1}^m \lambda_k^2 y_k^2}, \dots, \frac{\lambda_m y_m}{\sum_{k=1}^m \lambda_k^2 y_k^2}, 0, \dots, 0 \right).$$

Since  $h$  has continuous second order derivatives on the nonempty open  $\mathbb{R}^n \setminus (Q^* \text{Fix } T)$ , it must hold that

$$\frac{\partial^2 h}{\partial y_i \partial y_j} = \frac{\partial^2 h}{\partial y_j \partial y_i}$$

which gives

$$(116) \quad \frac{-2\lambda_j \lambda_i^2 y_i y_j}{\sum_{k=1}^m \lambda_k^2 y_k^2} = \frac{-2\lambda_i \lambda_j^2 y_i y_j}{\sum_{k=1}^m \lambda_k^2 y_k^2}$$

when  $1 \leq i, j \leq m, i \neq j$ . As  $\text{int lev}_0 f = \text{int Fix } T = \emptyset$ , (116) holds on the nonempty open  $\mathbb{R}^n \setminus (Q^* \text{Fix } T)$ , so we have  $\lambda_i = \lambda_j$ . Because  $1 \leq i, j \leq m$  were arbitrary, we obtain that  $\lambda_1 = \cdots = \lambda_m$ . Hence

$$T_1 = Q \begin{pmatrix} \lambda \text{Id}_m & 0 \\ 0 & 0 \end{pmatrix} Q^* = \lambda Q \begin{pmatrix} \text{Id}_m & 0 \\ 0 & 0 \end{pmatrix} Q^* = \lambda P_L$$

where  $L \subseteq \mathbb{R}^n$  is a linear subspace; see [38, page 430]. More precisely,  $T_1$  is a positive multiple of an orthogonal projector with

$$(117) \quad \text{Fix } T = \ker T_1 = L^\perp.$$

Now  $T_1$  is firmly nonexpansive and  $T_1^* = T_1$ , this implies that

$$\frac{T_1 + T_1^*}{2} - T_1^* T_1 = T_1 - T_1^2 = Q \begin{pmatrix} (\lambda - \lambda^2) \text{Id} & 0 \\ 0 & 0 \end{pmatrix} Q^*$$

is positive semidefinite, so  $0 \leq \lambda \leq 1$ . Because  $T_1 \neq 0$  in this case, we obtain  $0 < \lambda \leq 1$ .

Therefore, when  $x \notin \text{Fix } T$ ,

$$\nabla \ln f(x) = \frac{T_1 x}{\|T_1 x\|^2} = \frac{\lambda P_L x}{\|\lambda P_L x\|^2} = \frac{1}{\lambda} \frac{P_L x}{\|P_L x\|^2}.$$

Note that  $P_L^* = P_L, P_L^2 = P_L$ ,

$$\nabla \ln \|P_L x\| = \frac{1}{\|P_L x\|} \nabla \|P_L x\| = \frac{1}{\|P_L x\|} P_L^* \frac{P_L x}{\|P_L x\|} = \frac{P_L x}{\|P_L x\|^2}.$$

It follows that

$$\nabla \ln f(x) = \frac{1}{\lambda} \nabla \ln \|P_L x\| = \nabla \ln \|P_L x\|^{1/\lambda}.$$

On each connected and open component of  $\mathbb{R}^n \setminus \text{Fix } T$ , this is equivalent to

$$\ln f(x) = \ln \|P_L x\|^{1/\lambda} + c$$

for some constant  $c \in \mathbb{R}$ . Taking exp both sides gives

$$(118) \quad f(x) = K \|P_L x\|^{1/\lambda} = K (\|P_L x\|^2)^{1/(2\lambda)} = K (x^* P_L x)^{1/(2\lambda)}$$

where  $K = \exp(c) > 0$ . As  $P_L = \text{Id} - P_{L^\perp}$ , we obtain

$$f(x) = K \|x - P_{L^\perp} x\|^{1/\lambda} = K (d_{L^\perp}(x))^{1/\lambda}.$$

Moreover,

$$T = G_f = \text{Id} - T_1 = \text{Id} - \lambda P_L = \text{Id} - \lambda (\text{Id} - P_{L^\perp}) = (1 - \lambda) \text{Id} + \lambda P_{L^\perp}$$

where  $L^\perp = \text{Fix } T$  by (117). One can apply the same argument on each connected and open component of  $\mathbb{R}^n \setminus \text{Fix } T$ , while one might have different constant  $K$ 's in (118), but  $\lambda$  will be the same. Indeed, suppose that there exist  $0 < \lambda, \lambda_1 \leq 1, \lambda \neq \lambda_1$  such that

$$(1 - \lambda) \text{Id} + \lambda P_{\text{Fix } T} = (1 - \lambda_1) \text{Id} + \lambda_1 P_{\text{Fix } T}.$$

Then  $P_{\text{Fix } T} = \text{Id}$  so that  $\text{Fix } T = \mathbb{R}^n$ , which contradicts that  $\text{int } \text{Fix } T = \emptyset$ . Using the same  $K > 0$  for all connected and open component of  $\mathbb{R}^n \setminus \text{Fix } T$ , one obtains (115).

(ii) $\Rightarrow$ (i). Clear. ■

Theorem 9.6 is proved under the assumption that the linear subgradient projector of a convex function is symmetric. We think that the assumption of symmetry is superfluous; cf. Theorem 11.2.

**Conjecture 9.7** *If  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is convex and its subgradient projector  $G_{f,s}$  is linear, then  $G_{f,s}$  must be symmetric.*

Note that when  $f$  is not convex,  $G_{f,s}$  can be nonsymmetric; see Corollary 11.6(ii).

## 9.4 Characterization of linear subgradient projectors

In subsection 9.3, we assume that the linear operator is symmetric. What happens if the linear operator is not symmetric? For this purpose we need the following result.

**Proposition 9.8** *Let  $M : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be linear, monotone and*

$$(\forall x \in \mathbb{R}^n \setminus \ker M) \quad \nabla h(x) = \frac{Mx}{\|Mx\|^2}$$

*where the function  $h : \mathbb{R}^n \setminus \ker M \rightarrow \mathbb{R}$ . If  $\dim \text{ran } M \neq 2$ , then  $M$  is symmetric.*

*Proof.* If  $\dim \text{ran } M = 0$ , then  $M = 0$ , so it is symmetric. Let us assume that  $\dim \text{ran } M > 0$  and  $\dim \text{ran } M \neq 2$ . Since  $h$  has continuous mixed second order derivatives at  $x$  whenever  $Mx \neq 0$ , the Hessian matrix  $\nabla^2 h(x)$  is symmetric. As

$$\nabla^2 h(x) = \frac{\|Mx\|^2 M - Mx(\nabla \|Mx\|^2)^*}{\|Mx\|^4} = \frac{\|Mx\|^2 M - 2Mx x^* M^* M}{\|Mx\|^4},$$

the symmetric property means that

$$\|Mx\|^2 M - 2Mxx^*M^*M = (\|Mx\|^2 M - 2Mxx^*M^*M)^* = M^*\|Mx\|^2 - 2M^*Mxx^*M^*$$

whenever  $Mx \neq 0$ . Put  $y = Mx$ . The above is simplified to

$$\frac{M - M^*}{2} = \frac{yy^*}{\|y\|^2} M - M^* \frac{yy^*}{\|y\|^2}.$$

Denote the projection operator on the line spanned by  $\{y\}$ ,  $\text{span}(y)$ , by  $P_y := \frac{yy^*}{\|y\|^2}$ . We have

$$(119) \quad (\forall y \in \text{ran } M) \quad \frac{M - M^*}{2} = P_y M - M^* P_y.$$

Since  $M$  is monotone,  $\text{ran } M = \text{ran } M^*$ ; see, e.g., [14, Theorem 3.2]. Let  $\{e_i \mid i = 1, \dots, m\}$  be an orthonormal basis of  $\text{ran } M$ . Then

$$(120) \quad P_{\text{ran } M} = \sum_{i=1}^m e_i e_i^*.$$

Note that

$$(121) \quad M^* P_{\text{ran } M} = M^* P_{\text{ran } M^*} = M^* (P_{\text{ran } M^*} + P_{(\text{ran } M^*)^\perp}) = M^*$$

because  $M^* P_{(\text{ran } M^*)^\perp} = 0$ . To see this, let  $y \in (\text{ran } M^*)^\perp$ . For every  $z \in \mathbb{R}^n$ ,

$$\langle M^* y, z \rangle = \langle y, Mz \rangle = 0$$

because  $Mz \in \text{ran } M = \text{ran } M^*$ . Because  $z \in \mathbb{R}^n$  was arbitrary, we must have  $M^* y = 0$ . Since

$$(122) \quad \frac{M - M^*}{2} = P_{e_i} M - M^* P_{e_i}$$

by (119), summing up (122) from  $i = 1$  to  $i = m$ , followed by using (120) and (121), we obtain

$$\frac{m}{2} (M - M^*) = \left( \sum_{i=1}^m P_{e_i} \right) M - M^* \left( \sum_{i=1}^m P_{e_i} \right) = P_{\text{ran } M} M - M^* P_{\text{ran } M} = M - M^*,$$

that is,

$$\left( \frac{m}{2} - 1 \right) (M - M^*) = 0.$$

Hence  $M - M^* = 0$  because  $m \neq 2$ , and so  $M$  is symmetric. ■

The proof of Proposition 9.8 requiring  $\dim \text{ran } M \neq 2$  seems bizarre. However, the following examples show that Proposition 9.8 fails when  $\dim \text{ran } M = 2$ .

**Example 9.9** When  $\dim \text{ran } M = 2$ , although  $M : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is linear, monotone and

$$(\forall x \in \mathbb{R}^n \setminus \ker M) \quad \nabla h(x) = \frac{Mx}{\|Mx\|^2},$$

one cannot guarantee that  $M$  is symmetric.

To see this, let  $x = (x, y)^* \in \mathbb{R}^2$ .

(1). Define

$$M := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Then  $M$  is linear, monotone,  $\dim \operatorname{ran} M = 2$  and

$$\nabla \arctan(y/x) = \frac{\begin{pmatrix} -y \\ x \end{pmatrix}}{x^2 + y^2} = \frac{M\mathbf{x}}{\|M\mathbf{x}\|^2}$$

whenever  $\mathbf{x} \neq 0$ . However,  $M$  is not symmetric.

(2). Define

$$M := \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

Then  $M$  is linear, firmly nonexpansive and

$$\nabla \left( \frac{\ln(x^2 + y^2)}{2} + \arctan(y/x) \right) = \frac{\begin{pmatrix} x - y \\ y + x \end{pmatrix}}{x^2 + y^2} = \frac{M\mathbf{x}}{\|M\mathbf{x}\|^2}$$

whenever  $\mathbf{x} \neq 0$ . However,  $\dim \operatorname{ran} M = 2$  and  $M$  is not symmetric.

**Conjecture 9.10** *Let  $M : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be linear, monotone and*

$$(\forall x \in \mathbb{R}^n \setminus \ker M) \quad \nabla h(x) = \frac{Mx}{\|Mx\|^2}$$

*where the function  $h : \mathbb{R}^n \setminus \ker M \rightarrow \mathbb{R}$ . If  $\dim \operatorname{ran} M = 2$  and  $\exp(h)$  is convex on convex subsets of  $\mathbb{R}^n \setminus \ker M$ , then  $M$  is symmetric.*

Combining Theorem 9.6 and Proposition 9.8, we obtain the following characterization of linear subgradient projectors.

**Theorem 9.11** *Assume that  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is linear and  $\dim \operatorname{ran}(\operatorname{Id} - T) \neq 2$ . Then the following are equivalent:*

- (i)  $T$  is a subgradient projector of a convex function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  with  $\operatorname{lev}_0 f \neq \emptyset$ .
- (ii)  $T = G_f$  where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is given by

$$f(x) := K(x^* P_L x)^{1/(2\lambda)} = K(d_{L^\perp}(x))^{1/\lambda}$$

*where  $0 < \lambda \leq 1$ ,  $K > 0$ , and  $L \subseteq \mathbb{R}^n$  is a subspace such that  $L^\perp = \operatorname{Fix} T$ . In this case,  $G_f = (1 - \lambda) \operatorname{Id} + \lambda P_{L^\perp}$ .*

*Proof.* (i) $\Rightarrow$ (ii). Assume that  $T = G_f$  for some convex function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . Then  $T$  is a cutter by Fact 5.12. As  $T$  is linear, in view of Proposition 9.1,  $T$  is firmly nonexpansive, so  $M := \operatorname{Id} - T$  is firmly nonexpansive, in particular, monotone. By Theorem 4.1(i),

$$(123) \quad \nabla h(x) = \frac{Mx}{\|Mx\|^2}$$

where  $h(x) = \ln f(x)$ ,  $f(x) > 0$ . Since  $\text{Fix } T = \text{lev}_0 f = \ker M$ , (123) is equivalent to

$$\nabla h(x) = \frac{Mx}{\|Mx\|^2}$$

when  $Mx \neq 0$ . Proposition 9.8 shows that  $M$  is symmetric, so is  $T = \text{Id} - M$ . It suffices to apply Theorem 9.6 to obtain (ii).

(ii) $\Rightarrow$ (i). Clear. ■

## 10 Subgradient projectors of convex functions are not closed under convex combinations and compositions

A convex combination of cutters is a cutter, see [20, Corollary 2.1.49] or [7, Proposition 4.34]. Convex combinations of a finite family of cutters with a common fixed point are effectively used in simultaneous cutter methods; see [20, Section 5.8], [7, Corollary 5.18].

A question that naturally arises is whether the set of subgradient projectors of convex functions is convex. Theorem 9.6 allows us to show that the answer is negative. While Theorem 10.1 works only in  $\mathbb{R}^2$ , Theorem 10.3 works in  $\mathbb{R}^n$  with  $n \geq 2$ .

**Theorem 10.1** *In  $\mathbb{R}^2$ , a convex combination of subgradient projectors of convex functions need not be a subgradient projector of a convex function.*

*Proof.* Let  $L := \{0\} \times \mathbb{R} \subseteq \mathbb{R}^2$  and  $M := \{x = (x_1, x_2) \in \mathbb{R}^2 \mid x_1 + x_2 = 0\}$ . Both  $L, M$  are proper linear subspaces of  $\mathbb{R}^2$ . Define  $f, g : \mathbb{R}^2 \rightarrow \mathbb{R}$  by

$$(124) \quad (\forall x \in \mathbb{R}^2) \quad f(x) := K_1(d_{L^\perp}(x))^{1/\lambda_1}, \quad g(x) := K_2(d_{M^\perp}(x))^{1/\lambda_2}$$

where  $0 < \lambda_1 \neq \lambda_2 < 1$ ,  $K_1, K_2 > 0$ . By Theorem 9.6, we have

$$(125) \quad G_f = (1 - \lambda_1)\text{Id} + \lambda_1 P_{L^\perp}, \quad \text{and } G_g = (1 - \lambda_2)\text{Id} + \lambda_2 P_{M^\perp}.$$

Now consider  $\lambda_3 G_f + (1 - \lambda_3)G_g$  where  $0 < \lambda_3 < 1$ . Then

$$(126) \quad \lambda_3 G_f + (1 - \lambda_3)G_g$$

$$(127) \quad = \lambda_3((1 - \lambda_1)\text{Id} + \lambda_1 P_{L^\perp}) + (1 - \lambda_3)((1 - \lambda_2)\text{Id} + \lambda_2 P_{M^\perp})$$

$$(128) \quad = (1 - \lambda_2 + \lambda_2 \lambda_3 - \lambda_1 \lambda_3)\text{Id} + \lambda_1 \lambda_3 P_{L^\perp} + \lambda_2(1 - \lambda_3)P_{M^\perp}.$$

We show that  $\lambda_3 G_f + (1 - \lambda_3)G_g$  is not a subgradient projector of a convex function by contradiction. Suppose that  $\lambda_3 G_f + (1 - \lambda_3)G_g$  is a subgradient projector of a convex function. By Theorem 9.6, there are  $0 < \lambda < 1$  and  $S$  which is a subspace of  $\mathbb{R}^2$  such that

$$(129) \quad \lambda_3 G_f + (1 - \lambda_3)G_g = (1 - \lambda)\text{Id} + \lambda P_{S^\perp}.$$

Therefore, we have

$$(130) \quad (1 - \lambda_2 + \lambda_2 \lambda_3 - \lambda_1 \lambda_3)\text{Id} + \lambda_1 \lambda_3 P_{L^\perp} + \lambda_2(1 - \lambda_3)P_{M^\perp} = (1 - \lambda)\text{Id} + \lambda P_{S^\perp}.$$

Naturally, the set of fixed points of left-hand side is equal to the set of fixed points of right-hand side. Thus we have

$$(131) \quad \text{Fix}((1 - \lambda) \text{Id} + \lambda P_{S^\perp})$$

$$(132) \quad = \text{Fix}((1 - \lambda_2 + \lambda_2 \lambda_3 - \lambda_1 \lambda_3) \text{Id} + \lambda_1 \lambda_3 P_{L^\perp} + \lambda_2(1 - \lambda_3) P_{M^\perp}).$$

By [7, Proposition 4.34], we have

$$(133) \quad \text{Fix}((1 - \lambda_2 + \lambda_2 \lambda_3 - \lambda_1 \lambda_3) \text{Id} + \lambda_1 \lambda_3 P_{L^\perp} + \lambda_2(1 - \lambda_3) P_{M^\perp}) = L^\perp \cap M^\perp.$$

Also,

$$(134) \quad \text{Fix}((1 - \lambda) \text{Id} + \lambda P_{S^\perp}) = S^\perp.$$

Hence, using definitions of  $L$ ,  $M$ , and (131)-(134), it follows that

$$(135) \quad \{(0, 0)\} = L^\perp \cap M^\perp$$

$$(136) \quad = \text{Fix}((1 - \lambda_2 + \lambda_2 \lambda_3 - \lambda_1 \lambda_3) \text{Id} + \lambda_1 \lambda_3 P_{L^\perp} + \lambda_2(1 - \lambda_3) P_{M^\perp})$$

$$(137) \quad = \text{Fix}((1 - \lambda) \text{Id} + \lambda P_{S^\perp})$$

$$(138) \quad = S^\perp.$$

Therefore  $S^\perp = \{(0, 0)\}$ , which implies  $S = \mathbb{R}^2$ . In terms of matrices, we have

$$(139) \quad P_{L^\perp} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad P_{M^\perp} = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}, \quad \text{and } P_{S^\perp} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

In particular,  $P_{L^\perp}, P_{S^\perp}$  are diagonal matrices, but  $P_{M^\perp}$  is not. Hence, equation (130) is not true. Therefore,  $\lambda_3 G_f + (1 - \lambda_3) G_g$  is not a subgradient projector of a convex function.  $\blacksquare$

Our next result needs averaged mappings.

**Definition 10.2** (See [4], [7, Definition 4.23]) *Let  $\lambda \in (0, 1)$ . An operator  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is  $\lambda$ -averaged if there exists a nonexpansive operator  $N : \mathbb{R}^n \rightarrow \mathbb{R}^n$  such that  $T = (1 - \lambda) \text{Id} + \lambda N$ .*

**Theorem 10.3** *Let  $n \geq 2, 0 < \lambda_1 < 1, 0 < \lambda_2 < 1, 0 < \lambda < 1$ . Suppose that  $L, M$  are linear subspaces of  $\mathbb{R}^n$  satisfying  $L = M^\perp, M = L^\perp$ , and that both  $L$  and  $M$  are proper linear subspaces of  $\mathbb{R}^n$ . Define*

$$f : \mathbb{R}^n \rightarrow \mathbb{R} : x \mapsto (d_{L^\perp}(x))^{1/\lambda_1}, \text{ and}$$

$$g : \mathbb{R}^n \rightarrow \mathbb{R} : x \mapsto (d_{M^\perp}(x))^{1/\lambda_2}.$$

*If  $\frac{1-\lambda}{\lambda} \neq \frac{\lambda_2}{\lambda_1}$ , then  $(1 - \lambda)G_f + \lambda G_g$  is not a subgradient projector of a convex function.*

*Proof.* By Theorem 9.6, we have

$$(140) \quad G_f = (1 - \lambda_1) \text{Id} + \lambda_1 P_{L^\perp}, \text{ and } G_g = (1 - \lambda_2) \text{Id} + \lambda_2 P_{M^\perp}.$$

Then

$$(141) \quad (1 - \lambda)G_f + \lambda G_g$$

$$(142) \quad = (1 - \lambda) ((1 - \lambda_1) \text{Id} + \lambda_1 P_{L^\perp}) + \lambda ((1 - \lambda_2) \text{Id} + \lambda_2 P_{M^\perp})$$

$$(143) \quad = [(1-\lambda)(1-\lambda_1) + \lambda(1-\lambda_2)] \text{Id} + \lambda_1(1-\lambda)P_{L^\perp} + \lambda\lambda_2P_{M^\perp}$$

$$(144) \quad = \beta \text{Id} + (1-\beta)(\gamma P_{L^\perp} + (1-\gamma)P_{M^\perp}),$$

where  $\beta := (1-\lambda)(1-\lambda_1) + \lambda(1-\lambda_2)$  and  $\gamma := \frac{\lambda_1(1-\lambda)}{1-(1-\lambda)(1-\lambda_1)-\lambda(1-\lambda_2)}$ . We observe that  $0 < \beta < 1$  and  $\gamma \neq \frac{1}{2}$ . Indeed,

$$(145) \quad 0 < \beta = (1-\lambda)(1-\lambda_1) + \lambda(1-\lambda_2) < (1-\lambda) + \lambda = 1.$$

Also,

$$(146) \quad \gamma = \frac{1}{2}$$

$$(147) \quad \iff \frac{\lambda_1(1-\lambda)}{1-(1-\lambda)(1-\lambda_1)-\lambda(1-\lambda_2)} = \frac{\lambda\lambda_2}{1-(1-\lambda)(1-\lambda_1)-\lambda(1-\lambda_2)}$$

$$(148) \quad \iff \lambda_1(1-\lambda) = \lambda\lambda_2$$

$$(149) \quad \iff \frac{1-\lambda}{\lambda} = \frac{\lambda_2}{\lambda_1}.$$

By the assumption:  $\frac{1-\lambda}{\lambda} \neq \frac{\lambda_2}{\lambda_1}$ , so  $\gamma \neq \frac{1}{2}$ . Since  $G_f, G_g$  are linear and symmetric, so is  $(1-\lambda)G_f + \lambda G_g$ . We show that  $(1-\lambda)G_f + \lambda G_g$  is not a subgradient projector of a convex function by contradiction. If  $(1-\lambda)G_f + \lambda G_g$  is a subgradient projector of a convex function, by Theorem 9.6, we have

$$(150) \quad (1-\lambda)G_f + \lambda G_g = (1-\alpha) \text{Id} + \alpha P_{S^\perp}$$

where  $0 < \alpha < 1$  and  $S$  is a subspace of  $X$ . Note that  $G_f, G_g$  are averaged mappings, so is  $(1-\lambda)G_f + \lambda G_g$ . Because

$$(151) \quad \text{Fix } G_f = L^\perp, \text{ and } \text{Fix } G_g = M^\perp,$$

by [7, Proposition 4.34], we obtain

$$(152) \quad \text{Fix}((1-\lambda)G_f + \lambda G_g) = \text{Fix } G_f \cap \text{Fix } G_g = L^\perp \cap M^\perp = \{0\}.$$

Because  $\text{Fix}((1-\alpha) \text{Id} + \alpha P_{S^\perp}) = S^\perp$ , using (150) and (152) we obtain  $S^\perp = \{0\}$ . Therefore, in view of equation (150), we have

$$(153) \quad (1-\lambda)G_f + \lambda G_g = (1-\alpha) \text{Id}.$$

Combing (144) and (153) gives

$$(154) \quad \beta \text{Id} + (1-\beta)(\gamma P_{L^\perp} + (1-\gamma)P_{M^\perp}) = (1-\alpha) \text{Id}.$$

We proceed to analyze  $\alpha, \beta$ . Take  $x \in M \setminus \{0\}$ , which is possible since  $M \neq \{0\}$ . Then  $P_{M^\perp}x = 0, P_{L^\perp}x = P_Mx = x$ . Equation (154) gives

$$(155) \quad \beta x + (1-\beta)(\gamma x) = (1-\alpha)x,$$

which implies

$$(156) \quad \beta + (1-\beta)\gamma = 1-\alpha.$$

Take  $x \in L \setminus \{0\}$ , which is possible since  $L \neq \{0\}$ . Then  $P_{L^\perp}x = 0, P_{M^\perp}x = P_Lx = x$ . Equation (154) gives

$$(157) \quad \beta x + (1 - \beta)(1 - \gamma)x = (1 - \alpha)x,$$

which implies

$$(158) \quad \beta + (1 - \beta)(1 - \gamma) = 1 - \alpha.$$

Subtracting equation (156) from equation (158), we have

$$(159) \quad (1 - \beta)(1 - 2\gamma) = 0$$

which implies  $\beta = 1$  or  $\gamma = \frac{1}{2}$ . This contradicts the choices of  $\lambda, \lambda_1, \lambda_2$ . ■

If two nearest point projectors onto subspaces commute, then their composition is the projection onto the intersection of the subspaces; see [27, Lemma 9.2]. One referee asks whether there is an analogue when two linear subgradient projectors commute. The answer is negative. To this end, we need an auxiliary result.

**Lemma 10.4** *Let  $L, M \subseteq \mathbb{R}^n$  be two subspaces, and  $\lambda_i \in [0, 1)$  with  $i = 1, 2$ . Then the following are equivalent:*

- (i)  $(\lambda_1 \text{Id} + (1 - \lambda_1)P_{L^\perp})(\lambda_2 \text{Id} + (1 - \lambda_2)P_{M^\perp}) = (\lambda_2 \text{Id} + (1 - \lambda_2)P_{M^\perp})(\lambda_1 \text{Id} + (1 - \lambda_1)P_{L^\perp})$ .
- (ii)  $P_{L^\perp}P_{M^\perp} = P_{M^\perp}P_{L^\perp}$ .
- (iii)  $P_LP_M = P_MP_L$ .

*Proof.* (i) $\Leftrightarrow$ (ii): This follows from

$$(160) \quad (\lambda_1 \text{Id} + (1 - \lambda_1)P_{L^\perp})(\lambda_2 \text{Id} + (1 - \lambda_2)P_{M^\perp})$$

$$(161) \quad = \lambda_1\lambda_2 \text{Id} + \lambda_1(1 - \lambda_2)P_{M^\perp} + (1 - \lambda_1)\lambda_2P_{L^\perp} + (1 - \lambda_1)(1 - \lambda_2)P_{L^\perp}P_{M^\perp},$$

and

$$(162) \quad (\lambda_2 \text{Id} + (1 - \lambda_2)P_{M^\perp})(\lambda_1 \text{Id} + (1 - \lambda_1)P_{L^\perp})$$

$$(163) \quad = \lambda_1\lambda_2 \text{Id} + \lambda_2(1 - \lambda_1)P_{L^\perp} + (1 - \lambda_2)\lambda_1P_{M^\perp} + (1 - \lambda_1)(1 - \lambda_2)P_{M^\perp}P_{L^\perp}.$$

(ii) $\Leftrightarrow$ (iii): Since  $P_{L^\perp} = \text{Id} - P_L, P_{M^\perp} = \text{Id} - P_M$ , (ii) is equivalent to

$$(\text{Id} - P_L)(\text{Id} - P_M) = (\text{Id} - P_M)(\text{Id} - P_L),$$

which is (iii) after simplifications. ■

**Theorem 10.5** *In  $\mathbb{R}^2$ , even though two linear subgradient projectors of convex functions commute, its composition need not be a subgradient projector of a convex function.*

*Proof.* Let  $0 < \lambda_1 < \lambda_2 < 1$ . Because

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = P_{\mathbb{R} \times \{0\}}, \text{ and } \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = P_{\{0\} \times \mathbb{R}},$$



by Theorem 9.6, there exist two convex functions  $f, g : \mathbb{R}^2 \rightarrow \mathbb{R}$  such that

$$G_f = \lambda_1 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + (1 - \lambda_1) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda_1 \end{pmatrix}, \text{ and}$$

$$G_g = \lambda_2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + (1 - \lambda_2) \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \lambda_2 & 0 \\ 0 & 1 \end{pmatrix}.$$

These two subgradient projectors are commutative by Lemma 10.4 or a direct calculation:

$$G_f G_g = G_g G_f = \begin{pmatrix} \lambda_2 & 0 \\ 0 & \lambda_1 \end{pmatrix}.$$

We claim that  $T := G_f G_g$  is not a subgradient projector of a convex function. We prove this by contradiction. Suppose that  $T$  is a subgradient projector. Since  $T$  is symmetric and linear, by Theorem 9.6, there exists  $0 \leq \lambda \leq 1$  such that

$$(164) \quad T = \lambda \text{Id} + (1 - \lambda)P$$

where  $P$  is a projector onto a subspace of  $\mathbb{R}^2$ . We consider five cases.

Case 1.  $\lambda = 0$ . This gives  $P = T$ . Because  $P$  is a projector, its eigenvalues are 0 or 1. This is impossible, since  $0 < \lambda_i < 1$  and  $\lambda_1 \neq \lambda_2$ .

Case 2.  $\lambda = 1$ . This gives  $T = \text{Id}$ . This is impossible, since  $\lambda_1 \neq \lambda_2$ .

Case 1 and Case 2 implies that  $0 < \lambda < 1$ . This gives

$$P = \begin{pmatrix} \frac{\lambda_2 - \lambda}{1 - \lambda} & 0 \\ 0 & \frac{\lambda_1 - \lambda}{1 - \lambda} \end{pmatrix}.$$

Case 3.  $\lambda > \lambda_1$ . Then  $\lambda_1 - \lambda < 0$ . This is impossible, since the eigenvalues of  $P$  have to be nonnegative.

Case 4.  $\lambda = \lambda_1$ . Since

$$\frac{\lambda_2 - \lambda}{1 - \lambda} > 0,$$

and  $P$  has eigenvalues only of 0 or 1, we have

$$\frac{\lambda_2 - \lambda}{1 - \lambda} = 1.$$

It follows that  $\lambda_2 = 1$ , which is impossible.

Case 5.  $0 < \lambda < \lambda_1$ . Then

$$\frac{\lambda_1 - \lambda}{1 - \lambda} > 0, \quad \frac{\lambda_2 - \lambda}{1 - \lambda} > 0.$$

Since  $P$  has eigenvalues only of 0 or 1, we must have

$$\frac{\lambda_2 - \lambda}{1 - \lambda} = \frac{\lambda_1 - \lambda}{1 - \lambda} = 1$$

from which  $\lambda_1 = \lambda_2$ . This is impossible.

Altogether, (164) does not hold. Using Theorem 9.6 again, we conclude that  $T$  is not a subgradient projector of a convex function. ■

## 11 A complete analysis of linear subgradient projectors on $\mathbb{R}^2$

In this section we turn our attention to linear operators on  $\mathbb{R}^2$ . One nice feature is that we are able to not only characterize when the linear operator is a subgradient projector but also give explicit formulae for the corresponding functions.

Is every linear mapping from  $\mathbb{R}^2$  to  $\mathbb{R}^2$  a subgradient projector of an essentially strictly differentiable function (convex or nonconvex) on  $\mathbb{R}^2$ ? The answer is no by Theorem 11.2 below. Theorem 11.2(iii) also shows that Theorem 9.11 fails if  $\dim \operatorname{ran}(\operatorname{Id} - T) \neq 2$  is removed.

We start with a simple result about essentially strictly differentiable functions, see Definition 4.3.

**Lemma 11.1** *Let  $O \subseteq \mathbb{R}^n$  be a nonempty open set and  $f : O \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  be an essentially strictly differentiable function. If there exists a continuous selection  $s : O \rightarrow \mathbb{R}^n$  with  $s(x) \in \partial f(x)$  for every  $x \in O$ , then  $f$  is strictly differentiable on  $O$ . Consequently,  $f$  is continuously differentiable on  $O$ .*

*Proof.* By [15, Theorem 2.4, Corollary 4.2],  $f$  has a minimal Clarke subdifferential  $\partial_c f$ , and  $\partial_c f$  can be recovered by every dense selection of  $\partial_c f$ . Since  $s(x) \in \partial f(x) \subseteq \partial_c f(x)$ , and  $s$  is continuous on  $O$ , we have  $\partial_c f(x) = \partial f(x) = \{s(x)\}$  for every  $x \in O$ , which implies that  $f$  is strictly differentiable at  $x$ ; see, e.g., [51, page 362, Theorem 9.18] or [40, Theorem 3.54]. Hence  $f$  is strictly differentiable on  $O$ . ■

We consider the linear operator  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  defined by

$$(165) \quad T := \begin{pmatrix} 1 - a & -b \\ -c & 1 - d \end{pmatrix}$$

where

$$(166) \quad a^2 + b^2 + c^2 + d^2 \neq 0 \quad (\text{i.e., } (a, b, c, d) \neq (0, 0, 0, 0)).$$

Note that when  $a = b = c = d = 0$ , we have  $T = \operatorname{Id} = G_f$  with  $f \equiv 0$ .

**Theorem 11.2** *Let  $T$  be given by (165). Then  $T$  is a subgradient projector of an essentially strictly differentiable function on  $\mathbb{R}^2 \setminus \operatorname{Fix} T$  if and only if one of the following holds:*

- (i)  $a = b = c = 0, d \neq 0$ :  $T = G_f$  where  $f(x_1, x_2) := K|x_2|^{1/d}$  for some  $K > 0$ ;  $b = c = d = 0, a \neq 0$ :  $T = G_f$  where  $f(x_1, x_2) := K|x_1|^{1/a}$  for some  $K > 0$ .
- (ii)  $a \neq 0, d \neq 0, b = c \neq 0, ad = c^2$ :  $T = G_f$  where  $f(x_1, x_2) := K|ax_1 + cx_2|^{a/(a^2+c^2)}$  for some  $K > 0$ .
- (iii)  $a = d, b = -c$ , and  $a^2 + c^2 \neq 0$ :  $T = G_f$  where

$$(167) \quad f(x_1, x_2) := \begin{cases} K(x_1^2 + x_2^2)^{\frac{a}{2(a^2+c^2)}} \exp\left(-\frac{c}{a^2+c^2} \arctan\left(\frac{x_1}{x_2}\right)\right) & \text{if } x_2 \neq 0, \\ 0 & \text{if } (x_1, x_2) = (0, 0), \\ K|x_1|^{\frac{a}{(a^2+c^2)}} \exp\left(-\frac{|c|}{a^2+c^2} \frac{\pi}{2}\right) & \text{if } x_1 \neq 0, x_2 = 0, \end{cases}$$

for some  $K > 0$ , and  $f$  is lsc. In particular, when  $c \neq 0$ ,  $f$  is not convex.

*Proof.* Observe that (166) implies that  $\text{Fix } T$  is a proper subspace of  $\mathbb{R}^2$ . Assume that  $T$  is a subgradient projector. By Theorem 4.1 and Lemma 11.1, we can find a differentiable function  $g : (\mathbb{R}^2 \setminus \text{Fix } T) \rightarrow \mathbb{R}$  such that

$$\text{for every } x \in \mathbb{R}^2 \setminus \text{Fix } T, \quad \frac{x - Tx}{\|x - Tx\|^2} = \nabla g(x).$$

Because

$$x - Tx = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix},$$

we have

$$\begin{aligned} \frac{\partial g}{\partial x_1} &= \frac{ax_1 + bx_2}{(ax_1 + bx_2)^2 + (cx_1 + dx_2)^2}, \\ \frac{\partial g}{\partial x_2} &= \frac{cx_1 + dx_2}{(ax_1 + bx_2)^2 + (cx_1 + dx_2)^2}. \end{aligned}$$

Since

$$\begin{aligned} \frac{\partial^2}{\partial x_1 \partial x_2} g(x_1, x_2) &= -\frac{(a^2b - bc^2 + 2acd)x_1^2 + (b^3 + bd^2)x_2^2 + 2(ab^2 + ad^2)x_1x_2}{((ax_1 + bx_2)^2 + (cx_1 + dx_2)^2)^2}, \\ \frac{\partial^2}{\partial x_2 \partial x_1} g(x_1, x_2) &= -\frac{(a^2c + c^3)x_1^2 + (cd^2 - b^2c + 2abd)x_2^2 + 2(c^2d + a^2d)x_1x_2}{((ax_1 + bx_2)^2 + (cx_1 + dx_2)^2)^2}, \end{aligned}$$

on the nonempty open set of  $\mathbb{R}^2 \setminus \text{Fix } T$ , we have

$$\frac{\partial^2}{\partial x_1 \partial x_2} g(x_1, x_2) = \frac{\partial^2}{\partial x_2 \partial x_1} g(x_1, x_2).$$

This leads to

$$\begin{cases} a^2b - bc^2 + 2acd = a^2c + c^3, & (1) \\ b^3 + bd^2 = cd^2 - b^2c + 2abd, & (2) \\ ab^2 + ad^2 = c^2d + a^2d. & (3) \end{cases}$$

Now multiplying (2) by  $a$ , followed by subtracting it with (3) multiplied by  $b$ , gives  $(ad - bc)(ab + cd) = 0$ . It suffices to consider two cases:

- Case  $ad = bc$ . (1) implies  $(b - c)(a^2 + c^2) = 0$ . Observe that  $b - c \neq 0$  is impossible by (2). Then the following two subcases could happen.

- i.  $b = c = 0$ . Then (3)  $\Rightarrow ad(a - d) = 0$ . This means

$$(168) \quad a = b = c = 0, d \neq 0,$$

or

$$(169) \quad b = c = d = 0, a \neq 0.$$

- ii.  $b = c \neq 0$ , which implies  $a \neq 0, d \neq 0$ , and  $ad = c^2$ .

- Case  $ab + cd = 0$ . (1) implies  $(b + c)(a^2 + c^2) = 0$ . When  $b = -c = 0$ , (3) gives  $ad(a - d) = 0$ , which leads to (168), (169), or  $a = d \neq 0$ . It remains to consider the case  $b = -c \neq 0$ . Then (2) and (3) imply  $a = d$ . Moreover, we can and do assume  $a^2 + c^2 \neq 0$  since  $a = c = 0$  gives (168) by (2).

In summary, we only have the following three cases.

Case 1.  $a = b = c = 0, d \neq 0$ . Then we get

$$g(x_1, x_2) = \frac{\ln|x_2|}{d} + C_1, \text{ if } x_2 \neq 0.$$

Or  $b = c = d = 0, a \neq 0$ . Then we get

$$g(x_1, x_2) = \frac{\ln|x_1|}{a} + C_1, \text{ if } x_1 \neq 0.$$

Case 2.  $a \neq 0, d \neq 0, b = c \neq 0, ad = c^2$ . Then we get

$$g(x_1, x_2) = \frac{a}{a^2 + c^2} \ln|ax_1 + cx_2| + C_2, \text{ if } ax_1 + cx_2 \neq 0.$$

Case 3.  $a = d, b = -c$ , and  $a^2 + c^2 \neq 0$ . Then we get

$$g(x_1, x_2) = \frac{a}{2(a^2 + c^2)} \ln(x_1^2 + x_2^2) - \frac{c}{a^2 + c^2} \arctan\left(\frac{x_1}{x_2}\right) + C_3, \text{ if } x_2 \neq 0.$$

Since  $g = \ln f$ , we obtain  $f = \exp(g)$  by using Case 1-Case 2. For Case 3, we obtain

$$f(x_1, x_2) = K(x_1^2 + x_2^2)^{\frac{a}{2(a^2 + c^2)}} \exp\left(-\frac{c}{a^2 + c^2} \arctan\left(\frac{x_1}{x_2}\right)\right) \text{ if } x_2 \neq 0$$

for some  $K > 0$ . However, when  $c \neq 0$ ,  $f$  is not continuous at  $(\bar{x}_1, 0)$  with  $\bar{x}_1 \neq 0$  since

$$\lim_{x_1 \rightarrow \bar{x}_1, x_2 \downarrow 0} \arctan \frac{x_1}{x_2} \neq \lim_{x_1 \rightarrow \bar{x}_1, x_2 \uparrow 0} \arctan \frac{x_1}{x_2} = \pm \frac{\pi}{2}.$$

The function given by (167) is lsc but not continuous at every  $(\bar{x}_1, 0)$ . Moreover,  $f$  is not convex on  $\mathbb{R}^2$  since a finite-valued convex function on a finite dimensional space is continuous; see, e.g., [7, Corollary 8.31].

It is interesting to ask for what selection  $s \in \partial f$ , we have  $G_f = T$  on  $\mathbb{R}^2$ . On  $\mathbb{R}^2 \setminus \{(x_1, x_2) \mid x_2 \neq 0\}$ , one clearly chooses  $s = \nabla f$ . It remains to determine the subgradient of  $f$  at  $(\bar{x}_1, 0)$ . Indeed, when  $x_2 \neq 0$ ,  $f(x_1, x_2) = \exp(g(x_1, x_2))$ , so that  $\nabla f(x_1, x_2) = f(x_1, x_2) \nabla g(x_1, x_2)$ , i.e.

$$\nabla f(x_1, x_2) = K(x_1^2 + x_2^2)^{\frac{a}{2(a^2 + c^2)}} \exp\left(-\frac{c}{a^2 + c^2} \arctan\left(\frac{x_1}{x_2}\right)\right) \frac{1}{a^2 + c^2} \frac{(ax_1 - cx_2, ax_2 + cx_1)}{x_1^2 + x_2^2}.$$

When  $(x_1, x_2) \rightarrow (\bar{x}_1, 0)$ ,  $cx_1/x_2 > 0$ , we have

$$f(x_1, x_2) \rightarrow K|\bar{x}_1|^{\frac{a}{(a^2 + c^2)}} \exp\left(-\frac{|c|}{a^2 + c^2} \frac{\pi}{2}\right) = f(\bar{x}_1, 0),$$

and

$$\nabla f(x_1, x_2) \rightarrow K|\bar{x}_1|^{\frac{a}{(a^2+c^2)}} \exp\left(-\frac{|c|}{a^2+c^2} \frac{\pi}{2}\right) \frac{1}{a^2+c^2} \left(\frac{a}{\bar{x}_1}, \frac{c}{\bar{x}_1}\right).$$

Therefore, by the definition of limiting subdifferentials (see Definition 2.1),

$$(170) \quad K|\bar{x}_1|^{\frac{a}{(a^2+c^2)}} \exp\left(-\frac{|c|}{a^2+c^2} \frac{\pi}{2}\right) \frac{1}{a^2+c^2} \left(\frac{a}{\bar{x}_1}, \frac{c}{\bar{x}_1}\right) \in \partial f(\bar{x}_1, 0).$$

Hence, we can choose  $s(\bar{x}_1, 0)$  to be the limiting subgradient given by (170). ■

**Remark 11.3** Note that  $\partial f(\bar{x}_1, 0)$  is not a singleton when  $\bar{x}_1 \neq 0$  and  $c \neq 0$  in Theorem 11.2(iii). Thus, in Theorem 11.2(iii), we only have  $T \in \mathcal{G}_f$  when  $c \neq 0$ . In order to make  $f$  continuous on  $\mathbb{R}^n$ , we need  $c = 0$ , in which case (167) reduces to

$$f(x_1, x_2) = K\|(x_1, x_2)\|^{1/a},$$

and  $G_f = (1-a)\text{Id}$ . Clearly,  $f$  is not convex when  $a > 1$ . This has been discussed in Example 2.5.

**Corollary 11.4** *Let  $T$  be given by (165). Suppose that one of the following holds:*

- (i)  $|b| \neq |c|$ .
- (ii)  $b = c = 0, a \neq 0, d \neq 0$ , and  $a \neq d$ .

*Then there exists no  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  being essentially strictly differentiable such that  $T = G_f$ .*

**Corollary 11.5** *Let  $T$  be given by (165). Suppose that  $b = c = 0, 0 < a < 1, 0 < d < 1$ , and  $a \neq d$ . Then  $T$  is firmly nonexpansive, and there exists no  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  being essentially strictly differentiable such that  $T = G_f$ .*

**Corollary 11.6** (i) *The skew linear mapping*

$$T := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

*is not firmly nonexpansive, so not a cutter. However,  $T$  is a subgradient projector of a nonconvex, discontinuous but lsc function  $f_1$  given by*

$$(171) \quad f_1(x, y) := \begin{cases} (x^2 + y^2)^{1/4} \exp(- (1/2) \arctan(x/y)) & \text{if } y \neq 0, \\ 0 & \text{if } (x, y) = (0, 0), \\ |x|^{1/2} \exp(-\pi/4) & \text{if } x \neq 0, y = 0. \end{cases}$$

(ii) *The linear mapping*

$$T := \begin{pmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix}$$

*is firmly nonexpansive and a cutter. However,  $T$  is a subgradient projector of a nonconvex, discontinuous but lsc function  $f_2$  given by*

$$(172) \quad f_2(x, y) := \begin{cases} (x^2 + y^2)^{1/2} \exp(- \arctan(x/y)) & \text{if } y \neq 0, \\ 0 & \text{if } (x, y) = (0, 0), \\ |x| \exp(-\pi/2) & \text{if } x \neq 0, y = 0. \end{cases}$$

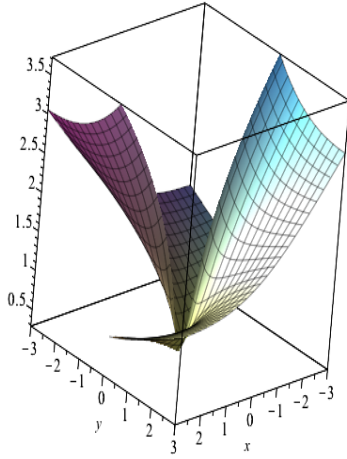


Figure 1: Plot of function given by (171)

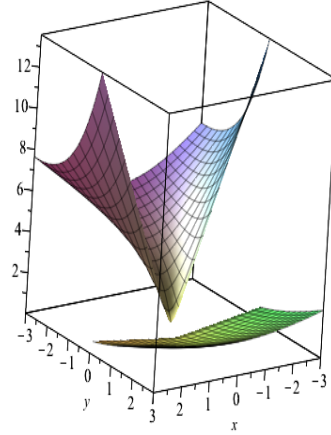


Figure 2: Plot of function given by (172)

Note that  $f_2 = f_1^2$  in Corollary 11.6 and  $G_{f_2} = (\text{Id} + G_{f_1})/2$ .

**Remark 11.7** Corollary 11.5 and Corollary 11.6 together show that although the set of cutters and the set of subgradient projectors have a nonempty intersection, they are different because neither one contains the other.

By Theorem 4.5, there exists no continuous convex function  $f$  such that  $G_f = T$  in either case of Corollary 11.6. Corollary 11.6 says that  $T = G_f$  being linear and firmly nonexpansive does not imply that  $f$  is convex. A key point below is that if  $T = G_f$  is linear and  $f$  is convex on  $\mathbb{R}^2$ , then Theorem 11.2 implies that  $T$  has to be firmly nonexpansive and symmetric.

**Corollary 11.8** *Let  $T$  be given by (165). Then  $T$  is a subgradient projector of a convex function if and only if one of the following holds:*

- (i)  $a = b = c = 0, d \neq 0, 0 < d \leq 1$ :  $T = G_f$  where  $f(x_1, x_2) = K|x_2|^{1/d}$  for some  $K > 0$ ; or  $b = c = d = 0, a \neq 0, 0 < a \leq 1$ :  $T = G_f$  where  $f(x_1, x_2) = K|x_1|^{1/a}$  for some  $K > 0$ .
- (ii)  $a \neq 0, d \neq 0, b = c \neq 0, ad = c^2, a \geq a^2 + c^2$ :  $T = G_f$  where  $f(x_1, x_2) = K|ax_1 + cx_2|^{a/(a^2+c^2)}$  for some  $K > 0$ .
- (iii)  $a = d, b = c = 0, 0 < a \leq 1$ :  $T = G_f$  where  $f(x_1, x_2) = K(x_1^2 + x_2^2)^{\frac{1}{2a}}$  for some  $K > 0$ .

## Acknowledgments

The authors thank two anonymous referees for careful reading and constructive suggestions on the paper. HHB was partially supported by a Discovery Grant of the Natural Sciences and Engineering Research Council of Canada (NSERC) and by the Canada Research Chair Program. CW was partially supported by National Natural Science Foundation of China (11401372). XW was partially supported by a Discovery Grant of the Natural Sciences and Engineering Research Council of Canada (NSERC). JX was supported by by NSERC grants of HHB and XW.

## References

- [1] S. Adly, F. Nacry, and L. Thibault, Preservation of prox-regularity of sets with applications to constrained optimization, *SIAM J. Optim.* 26 (2016), no. 1, 448–473.
- [2] D. Aussel, A. Daniilidis, and L. Thibault, Subsmooth sets: functional characterizations and related concepts, *Trans. Amer. Math. Soc.* 357 (2005), no. 4, 1275–1301.
- [3] M. Bacák, 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 (2010), no. 3-4, 737–763.
- [4] J.B. Baillon, R.R. Bruck, and S. Reich, On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces, *Houston J. Math.* 4 (1978), 1–9.
- [5] F. Bernard and L. Thibault, Prox-regular functions in Hilbert spaces, *J. Math. Anal. Appl.* 303 (2005), no. 1, 1–14.
- [6] F. Bernard and L. Thibault, Prox-regularity of functions and sets in Banach spaces, *Set-Valued Anal.* 12 (2004), no. 1-2, 25–47.
- [7] H.H. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011.
- [8] H.H. Bauschke, J.M. Borwein, and P.L. Combettes, Bregman monotone optimization algorithms, *SIAM J. Control Optim.* 42 (2003), 596–636.
- [9] H.H. Bauschke and P.L. Combettes, A weak-to-strong convergence principle for Fejermotone methods in Hilbert spaces, *Math. Oper. Res.* 26 (2001), 248–264.
- [10] H.H. Bauschke, P.L. Combettes, and D. Noll, Joint minimization with alternating Bregman proximity operators, *Pac. J. Optim.* 2 (2006), no. 3, 401–424.
- [11] H.H. Bauschke, J. Chen, and X. Wang, A Bregman projection method for approximating fixed points of quasi-Bregman nonexpansive mappings, *Appl. Anal.* 94 (2015), 75–84.
- [12] H.H. Bauschke, C. Wang, X. Wang, and J. Xu, On subgradient projectors, *SIAM J. Optim.* 25 (2015), no. 2, 1064–1082.
- [13] H.H. Bauschke, C. Wang, X. Wang, and J. Xu, On the finite convergence of a projected cutter method, *J. Optim. Theory Appl.* 165 (2015), no. 3, 901–916.
- [14] H.H. Bauschke, X. Wang, and L. Yao, Monotone linear relations: maximality and Fitzpatrick functions, *J. Convex Anal.* 16 (2009), 673–686.
- [15] J.M. Borwein and W.B. Moors, Essentially smooth Lipschitz functions, *J. Funct. Anal.* 149 (1997), no. 2, 305–351.
- [16] J.M. Borwein, W.B. Moors, and X. Wang, Generalized subdifferentials: a Baire categorical approach, *Trans. Amer. Math. Soc.* 353 (2001), no. 10, 3875–3893.
- [17] J.M. Borwein and A.S. Lewis, *Convex Analysis and Nonlinear Optimization: Theory and Examples*, Springer, New York, second edition, 2006.

- [18] J.M. Borwein and Q.J. Zhu, *Techniques of Variational Analysis*, Springer-Verlag, New York, 2005.
- [19] T.D. Capricelli, *Algorithmes de Projections Convexes Généralisées et Applications en Imagerie Médicale*, Ph.D. dissertation, University Pierre & Marie Curie (Paris 6), Paris, June 2008.
- [20] A. Cegielski, *Iterative Methods for Fixed Point Problems in Hilbert Spaces*, Lecture Notes in Mathematics, 2057, Springer, Heidelberg, 2012.
- [21] Y. Censor and A. Lent, Cyclic subgradient projections, *Math. Programming* 24 (1982), no. 1, 233–235.
- [22] Y. Censor, W. Chen, and H. Pajoohesh, Finite convergence of a subgradient projections method with expanding controls, *Appl. Math. Optim.* 64 (2011), no. 2, 273–285.
- [23] F.H. Clarke, *Optimization and Nonsmooth Analysis*, second edition, Classics in Applied Mathematics, 5. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.
- [24] P.L. Combettes, Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections, *IEEE Trans. Image Process.* 6 (1997), no. 4, 493–506.
- [25] P.L. Combettes and J. Luo, An adaptive level set method for nondifferentiable constrained image recovery, *IEEE Trans. Image Process.* 11 (2002), no. 11, 1295–1304.
- [26] A. Daniilidis and P. Georgiev, Approximate convexity and submonotonicity, *J. Math. Anal. Appl.* 291 (2004), no. 1, 292–301.
- [27] F. Deutsch, *Best Approximation in Inner Product Spaces*, Springer, 2001.
- [28] L.C. Evans and R.F. Gariepy, *Measure Theory and Fine Properties of Functions*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1992.
- [29] M. Fukushima, A finitely convergent algorithm for convex inequalities, *IEEE Trans. Automat. Control* 27 (1982), no. 5, 1126–1127.
- [30] R. Hesse and D.R. Luke, Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems, *SIAM J. Optim.* 23 (2013), no. 4, 2397–2419.
- [31] A.D. Ioffe, Approximate subdifferentials and applications, I: the finite-dimensional theory, *Trans. Amer. Math. Soc.* 281 (1984), no. 1, 389–416.
- [32] A. Jourani, L. Thibault, and D. Zagrodny, Differential properties of the Moreau envelope, *J. Funct. Anal.* 266 (2014), no. 3, 1185–1237.
- [33] C. Kan and W. Song, The Moreau envelope function and proximal mapping in the sense of the Bregman distance, *Nonlinear Anal.* 75 (2012), no. 3, 1385–1399.
- [34] K.C. Kiwiel, The efficiency of subgradient projection methods for convex optimization. I. General level methods, *SIAM J. Control Optim.* 34 (1996), no. 2, 660–676.
- [35] K.C. Kiwiel, The efficiency of subgradient projection methods for convex optimization. II. Implementations and extensions, *SIAM J. Control Optim.* 34 (1996), no. 2, 677–697.



- [36] K.C. Kiwiel, A Bregman-projected subgradient method for convex constrained nondifferentiable minimization, *Operations Research Proceedings 1996 (Braunschweig)*, 26–30, Springer, Berlin, 1997.
- [37] D.R. Luke, N.H. Thao, and M.K. Tam, Quantitative convergence analysis of iterated expansive, set-valued mappings, <http://arxiv.org/abs/1605.05725>, May 18, 2016.
- [38] C. Meyer, *Matrix Analysis and Applied Linear Algebra*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.
- [39] B.S. Mordukhovich and Y. Shao, Nonsmooth sequential analysis in Asplund spaces, *Trans. Amer. Math. Soc.* 348 (1996), 1235–1280.
- [40] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation I*, Springer-Verlag, 2006.
- [41] H. Van Ngai, D.T. Luc, and M. Théra, Approximate convex functions, *J. Nonlinear Convex Anal.* 1 (2000), no. 2, 155–176.
- [42] N. Ogura and I. Yamada, A deep outer approximating half space of the level set of certain quadratic functions, *J. Nonlinear Convex Anal.* 6 (2005), no. 1, 187–201.
- [43] C.H. Jeffery Pang, Finitely convergent algorithm for nonconvex inequality problems, <http://arxiv.org/abs/1405.7280>, July 31, 2014.
- [44] B. Pauwels, Subgradient projection operators, <http://arxiv.org/abs/1403.7237v1>, March 2014.
- [45] R.A. Poliquin and R.T. Rockafellar, Prox-regular functions in variational analysis, *Trans. Amer. Math. Soc.* 348 (1996), no. 5, 1805–1838.
- [46] B.T. Polyak, Minimization of unsmooth functionals, *U.S.S.R. Computational Mathematics and Mathematical Physics* 9 (1969), 14–29. (The original version appeared in *Akademija Nauk SSSR. Žurnal Vychislitel' noř Matematiki i Matematičeskoř Fiziki* 9 (1969), 509–521.)
- [47] B.T. Polyak, *Introduction to Optimization*, Optimization Software, 1987.
- [48] B.T. Polyak, Random algorithms for solving convex inequalities, in *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, D. Butnariu, Y. Censor, and S. Reich (editors), pages 409–422, Elsevier, 2001.
- [49] L.F. Richardson, *Measure and Integration : A Concise Introduction to Real Analysis*, Wiley, 2009.
- [50] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.
- [51] R.T. Rockafellar and R. J-B Wets, *Variational Analysis*, Springer, corrected 3rd printing, 2009.
- [52] J.E. Spingarn, Submonotone subdifferentials of Lipschitz functions, *Trans. Amer. Math. Soc.* 264 (1981), no. 1, 77–89.
- [53] X. Wang, Subdifferentiability of real functions, *Real Anal. Exchange* 30 (2004/05), no. 1, 137–171.
- [54] J. Xu, *Subgradient Projectors: Theory, Extensions, and Algorithms*, Ph.D. dissertation, University of British Columbia, Kelowna, April 2016.

- [55] I. Yamada and N. Ogura, Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions, *Numer. Funct. Anal. Optim.* 25 (2004), no. 7-8, 593–617.