

## PROX-REGULAR FUNCTIONS IN VARIATIONAL ANALYSIS

R. A. POLIQUIN AND R. T. ROCKAFELLAR

ABSTRACT. The class of prox-regular functions covers all l.s.c., proper, convex functions, lower- $\mathcal{C}^2$  functions and strongly amenable functions, hence a large core of functions of interest in variational analysis and optimization. The subgradient mappings associated with prox-regular functions have unusually rich properties, which are brought to light here through the study of the associated Moreau envelope functions and proximal mappings. Connections are made between second-order epi-derivatives of the functions and proto-derivatives of their subdifferentials. Conditions are identified under which the Moreau envelope functions are convex or strongly convex, even if the given functions are not.

### 1. INTRODUCTION

Fundamental insights into the properties of a function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$  in variational analysis come from the study of its *Moreau envelopes*  $e_\lambda$ , defined for  $\lambda > 0$  by

$$(1.1) \quad e_\lambda(x) = \min_{x'} \left\{ f(x') + \frac{1}{2\lambda} |x' - x|^2 \right\},$$

and the associated *proximal mappings*  $P_\lambda$ , defined by

$$(1.2) \quad P_\lambda(x) = \operatorname{argmin}_{x'} \left\{ f(x') + \frac{1}{2\lambda} |x' - x|^2 \right\}.$$

While the set-valued mappings  $P_\lambda$  relate to basic computational schemes in optimization, the functions  $e_\lambda$  provide a sort of *regularization* of  $f$ : as long as  $f$  is lower semicontinuous (l.s.c.) and minorized by some quadratic function, it is known that, for  $\lambda$  sufficiently small,  $e_\lambda$  is finite and locally Lipschitz continuous, and that  $e_\lambda$  increases pointwise to  $f$  as  $\lambda \downarrow 0$ . In terms of the operation “ $\square$ ” of infimal convolution in the space of all extended-real-valued functions on  $\mathbb{R}^n$ , one has  $e_\lambda = f \square \frac{1}{2\lambda} |\cdot|^2$ .

When  $f$  is convex,  $e_\lambda$  is convex for every  $\lambda > 0$ . This fact has long been important in variational analysis and optimization, but, rather surprisingly, the

---

Received by the editors December 21, 1994 and, in revised form, June 7, 1995.

1991 *Mathematics Subject Classification*. Primary 49A52, 58C06, 58C20; Secondary 90C30.

*Key words and phrases*. Prox-regularity, amenable functions, primal-lower-nice functions, proximal mappings, Moreau envelopes, regularization, subgradient mappings, nonsmooth analysis, variational analysis, proto-derivatives, second-order epi-derivatives, Attouch’s theorem.

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under grant OGP41983 for the first author and by the National Science Foundation under grant DMS-9200303 for the second author.

extent to which  $e_\lambda$  might be convex, even when  $f$  is not convex, has received little attention. Our aim is to address such imbalance here by systematically exploring the convexity-related properties of  $e_\lambda$  not only for their own sake but for the insights they afford in the study of the mapping  $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ , where  $\partial f(x)$  denotes the set of limiting proximal subgradients of  $f$  at  $x$ .

Motivation for our efforts comes from the strong connection in convex analysis between functions and their Moreau envelopes, both in subgradient theory and computation. There the proximal mappings  $P_\lambda$  can be used not only to parameterize the graph of  $\partial f$  but in support of solution techniques such as the proximal point algorithm. These proximal mappings appear actually as gradient mappings for the Moreau envelope functions associated with the Legendre-Fenchel transform  $f^*$  of  $f$ . The same cannot hold for nonconvex  $f$ , but there have been indications that much of importance can nevertheless be learned by examining that case closely.

Recently it was discovered by Thibault and Zagrodny [34] that if  $f$  belongs to the class of primal-lower-nice functions, defined by Poliquin [14] in connection with recovering a function from its subgradient mapping, then  $e_\lambda$  is lower- $\mathcal{C}^2$  locally when  $\lambda > 0$  is sufficiently small. A function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is *primal-lower-nice* (p.l.n.) at  $\bar{x}$ , a point where  $f$  is finite, if there exist  $R > 0$ ,  $c > 0$  and  $\epsilon > 0$  with the property that

$$(1.3) \quad f(x') > f(x) + \langle v, x' - x \rangle - \frac{r}{2} |x' - x|^2$$

whenever  $r > R$ ,  $|v| < cr$ ,  $v \in \partial f(x)$ ,  $|x' - \bar{x}| < \epsilon$  with  $x' \neq x$ , and  $|x - \bar{x}| < \epsilon$ . It is known that a function is lower- $\mathcal{C}^2$  if and only if it is expressible locally as the difference between a finite convex function and a positive multiple of  $\frac{1}{2} |\cdot|^2$  (cf. Rockafellar [26]), so this discovery opens doors to convex analysis and reveals a previously hidden route to deeper understanding of the behavior of Moreau envelopes. The significance of the lower- $\mathcal{C}^2$  property in regularization has been underscored by subsequent work of Levy, Poliquin and Thibault [10] in using it to obtain a generalization of Attouch's theorem on subgradient convergence beyond the context of convex functions  $f$ .

Here we work in a broader context than that of p.l.n. functions, demonstrating in particular that lower- $\mathcal{C}^2$  regularizations can exist under weaker assumptions than those in [34]. For this purpose we introduce the class of prox-regular functions  $f$ , which has interesting potential in several respects. Before stating the definition, we recall that  $f$  is *locally l.s.c.* at  $\bar{x}$  if  $f$  is l.s.c. relative to the set  $\{x \mid |x - \bar{x}| < \epsilon, f(x) < \alpha\}$  for some  $\epsilon > 0$  and  $\alpha > f(\bar{x})$ . This is equivalent to  $\text{epi } f$  being closed relative to a neighborhood of  $(\bar{x}, f(\bar{x}))$ . (Such a neighborhood is all that counts when the focus is on subgradients of  $f$  at  $\bar{x}$ .) When  $f$  is locally l.s.c. at  $\bar{x}$  it is in particular l.s.c. at  $\bar{x}$  itself in the usual sense: for every  $\beta < f(\bar{x})$  there exists  $\delta > 0$  such that  $f(x) > \beta$  when  $|x - \bar{x}| < \delta$ .

**Definition 1.1.** A function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  that is finite at  $\bar{x}$  is *prox-regular at  $\bar{x}$*  for  $\bar{v}$ , where  $\bar{v} \in \partial f(\bar{x})$ , if  $f$  is locally l.s.c. at  $\bar{x}$  and there exist  $\epsilon > 0$  and  $r > 0$  such that

$$(1.4) \quad f(x') > f(x) + \langle v, x' - x \rangle - \frac{r}{2} |x' - x|^2$$

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

Note that (1.4) requires  $v$  to be a proximal subgradient at  $x$ :  $v \in \partial_p f(x)$ . Beyond this, the definition calls for a kind of local uniformity in the parameter values involved in this proximal subgradient property. Also note that nothing would be changed in this definition if the subgradients  $v \in \partial f(x)$  in (1.4) were taken just to be proximal subgradients  $v \in \partial_p f(x)$ . Indeed if (1.4) holds for all proximal subgradients, then a similar inequality must hold, with  $r$  replaced by  $r + 1$  say, for all limiting proximal subgradients.

Obviously, any function that is p.l.n. is prox-regular. A big difference between a prox-regular function and a p.l.n. function, however, is that for a p.l.n. function condition (1.3) must hold for *all* subgradients and do so with a *linear growth condition*, whereas for a prox-regular function condition (1.4) only has to hold for subgradients close to a fixed  $\bar{v}$  and just in an  *$f$ -attentive* neighborhood of  $\bar{x}$ , i.e., a neighborhood making not only  $x$  close to  $\bar{x}$  but  $f(x)$  close to  $f(\bar{x})$ . (The  $f$ -attentive topology on  $\mathbb{R}^n$  is the coarsest topology that includes the usual topology and makes  $f$  continuous.) For instance, the function  $f$  on  $\mathbb{R}^1$  with  $f(x) = 1$  for  $x > 0$  but  $f(x) = 0$  for  $x \leq 0$  is easily seen to be prox-regular at  $\bar{x} = 0$  for  $\bar{v} = 0$ , yet it is not p.l.n. for this  $\bar{x}$ .

The scope and importance of the class of prox-regular functions is readily appreciated from the fact that it includes not only all l.s.c., proper, convex functions and all lower- $\mathcal{C}^2$  functions, but all strongly amenable functions, cf. Section 2. Strongly amenable functions are obtained by composing convex functions (extended-real-valued) with  $\mathcal{C}^2$  mappings under a constraint qualification. They have a basic role in convex and nonconvex optimization, both theoretically and computationally, cf. [11], [16], [27], [29].

A number of equivalent characterizations of prox-regular functions are worked out in Section 3; two in particular deserve special attention. We show that  $f$  is prox-regular if and only if an  $f$ -attentive localization of its subgradient mapping can be made monotone by the addition of a suitable multiple of the identity mapping. This “pre-monotonicity” property enables us to apply the theory of monotone mappings to the study of  $\partial f$ . We also give a geometric equivalence: we show that  $f$  is prox-regular at  $\bar{x}$  for  $\bar{v}$  if and only if  $\text{epi } f$  is *prox-regular* at  $(\bar{x}, f(\bar{x}))$  for  $(\bar{v}, -1)$ . This geometric equivalence is analogous to the equivalence between *Clarke regularity* of a function and Clarke regularity of its epigraph, cf [4]. Prox-regularity of sets is defined and studied in Section 2. A prox-regular set is quite close and in fact can be viewed as a “directionally local” version of a *proximally smooth* set. These sets were recently introduced in Clarke, Stern and Wolenski [6] in connection with the “paintability” of sets, the differentiability of distance functions, and the study of lower- $\mathcal{C}^2$  functions. Some related geometric ideas can be found also in an earlier paper of Vial [35].

Most questions about the prox-regularity of  $f$  at  $\bar{x}$  for a vector  $\bar{v} \in \partial f(\bar{x})$  can conveniently be normalized to the case where  $\bar{x} = 0$  and  $\bar{v} = 0$ , moreover with  $f(0) = 0$ . In Section 4 we explore the behavior of  $e_\lambda$  and  $P_\lambda$  in that setting. We show that for  $\lambda$  small enough,  $e_\lambda$  is  $\mathcal{C}^{1+}$  in a neighborhood of the origin, and we give a formula for the gradient mapping  $\nabla e_\lambda$  in terms of the proximal mapping  $P_\lambda$ . This development leads to the important conclusion that when  $f$  is prox-regular at  $\bar{x}$  for  $\bar{v}$ , the graph of  $\partial f$  in a localized sense around  $(\bar{x}, \bar{v})$  is a Lipschitz manifold of dimension  $n$  in  $\mathbb{R}^n \times \mathbb{R}^n$ , a property previously detected only for convex functions and their very close allies.

In Section 5 we show further that, in some neighborhood of 0, the function  $e_\lambda + s|\cdot|^2$  is convex for  $s = r/2(1 - \lambda r)$ ; here  $r$  is the constant that appears in (1.4). Further, we develop conditions under which  $e_\lambda$  itself is convex or strongly convex.

Section 6 is devoted to second-order aspects of prox-regular functions. We show that when  $f$  is prox-regular at  $\bar{x}$  for  $\bar{v}$ , it is twice epi-differentiable at  $\bar{x}$  for  $\bar{v}$  if and only if an  $f$ -attentive localization of the subgradient mapping  $\partial f$  is proto-differentiable at  $\bar{x}$  for  $\bar{v}$ . This has key significance for the sensitivity analysis of quasi-solutions to problems of optimization. Until now, such a result was available only for convex functions and strongly amenable functions; see [13], [15] and [29]. It has major implications for the second-order theory of Moreau envelopes, which we shall lay out in a separate paper [19]; see [20] for consequences in nonlinear programming. We also prove that when the second-order epi-derivative function associated with  $f$  at  $\bar{x}$  for  $\bar{v}$  is finite, it provides a second-order expansion in the traditional sense, although with a possibly nonquadratic second-order term.

## 2. EXAMPLES OF PROX-REGULARITY

The definition of prox-regularity centers on a pair  $(\bar{x}, \bar{v})$ , but the property in question then holds “locally”: if  $f$  is prox-regular at  $\bar{x}$  for a particular subgradient  $\bar{v} \in \partial f(\bar{x})$ , it is also prox-regular at  $\tilde{x}$  for  $\tilde{v} \in \partial f(\tilde{x})$  as long as  $(\tilde{x}, \tilde{v})$  is close enough to  $(\bar{x}, \bar{v})$  and such that  $f(\tilde{x})$  is close enough to  $f(\bar{x})$ . For many functions the latter provision is automatic, because closeness of subgradients already ensures closeness of function values, then the condition on function values in Definition 1.1 can be dropped as well. To facilitate the discussion of examples of prox-regularity, we give a term for this.

**Definition 2.1.** A function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is *subdifferentially continuous* at  $\bar{x}$  for  $\bar{v}$ , where  $\bar{v} \in \partial f(\bar{x})$ , if for every  $\delta > 0$  there exists  $\epsilon > 0$  such that  $|f(x) - f(\bar{x})| < \delta$  whenever  $|x - \bar{x}| < \epsilon$  and  $|v - \bar{v}| < \epsilon$  with  $v \in \partial f(x)$ .

Many important functions are subdifferentially continuous. In particular all p.l.n. functions enjoy this property.

**Proposition 2.2.** *If  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is p.l.n. at  $\bar{x}$ , then for all  $x$  in a neighborhood of  $\bar{x}$  it is subdifferentially continuous at  $x$  for any  $v \in \partial f(x)$ .*

*Proof.* Take  $R > 0$ ,  $c > 0$  and  $\epsilon > 0$  from the definition of p.l.n. given in Section 1 (see the discussion near (1.3)). Let  $|x - \bar{x}| < \epsilon$  and  $v \in \partial f(x)$ . Fix  $\delta > 0$  and  $\tilde{R} > R$ . We have for  $\epsilon_1 > 0$  small enough that

$$f(x') > f(x) + \langle v, x' - x \rangle - \frac{1}{2} \left( \frac{|v|}{c} + \tilde{R} \right) |x' - x|^2 > f(x) - \delta \text{ when } |x' - x| < \epsilon_1.$$

On the other hand, for  $v' \in \partial f(x')$  with  $|x' - \bar{x}| < \epsilon$  and  $|v' - v| < c\tilde{R}$  we have

$$f(x) > f(x') + \langle v', x - x' \rangle - \frac{1}{2} \left( \frac{|v|}{c} + \tilde{R} \right) |x - x'|^2.$$

We can now choose  $0 < \epsilon_2 < \epsilon$  with  $f(x) > f(x') - \delta$  whenever  $|x - x'| < \epsilon_2$  and  $|v' - v| < c\tilde{R}$ . Finally, for  $0 < \tilde{\epsilon} < \min\{\epsilon_1, \epsilon_2\}$  we have  $|f(x') - f(x)| < \delta$  whenever

$|x' - x| < \tilde{\epsilon}$  and  $|v' - v| < c\tilde{R}$ . This shows that  $f$  is subdifferentially continuous at  $x$  for  $v$ . □

Prox-regular functions are not in general subdifferentially continuous. For example the function mentioned in the introduction, namely the function  $f$  on  $\mathbb{R}^1$  with  $f(x) = 1$  for  $x > 0$  but  $f(x) = 0$  for  $x \leq 0$ , is prox-regular at  $\bar{x} = 0$  for  $\bar{v} = 0$ , yet it is not subdifferentially continuous at  $\bar{x}$  for  $\bar{v}$ . Nevertheless prox-regular functions enjoy a property that is close to subdifferential continuity.

**Proposition 2.3.** *If  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is prox-regular at  $\bar{x}$  for  $\bar{v}$ , then the following property close to subdifferential continuity is automatic: there exists  $\epsilon > 0$  such that*

$$\left. \begin{array}{l} (x', v') \rightarrow (x, v) \text{ with} \\ v' \in \partial f(x'), f(x') \leq f(\bar{x}) + \epsilon \end{array} \right\} \implies f(x') \rightarrow f(x)$$

whenever  $|x - \bar{x}| < \epsilon$  with  $|f(x) - f(\bar{x})| < \epsilon$  while  $|v - \bar{v}| < \epsilon$  with  $v \in \partial f(x)$ .

*Proof.* Take  $\epsilon > 0$  and  $r > 0$  from Definition 1.1 of prox-regularity. Because  $f$  is in particular l.s.c. at  $\bar{x}$ , we may assume that  $f(x') \geq f(\bar{x}) - \epsilon$  whenever  $|x' - \bar{x}| < \epsilon$  (which merely entails taking  $\epsilon$  smaller if necessary). Let  $|x - \bar{x}| < \epsilon$  with  $|f(x) - f(\bar{x})| < \epsilon$  and  $|v - \bar{v}| < \epsilon$  with  $v \in \partial f(x)$ . Now assume that  $x_k \rightarrow x$  and  $v_k \rightarrow v$  with  $v_k \in \partial f(x_k)$  and  $f(x_k) \leq f(\bar{x}) + \epsilon$ . We have eventually that  $|f(x_k) - f(\bar{x})| < \epsilon$ ,  $|v_k - \bar{v}| < \epsilon$  and  $|x_k - \bar{x}| < \epsilon$ , so that

$$\begin{aligned} f(x) &> f(x_k) + \langle v_k, x - x_k \rangle - \frac{r}{2}|x_k - x|^2, \\ f(x_k) &> f(x) + \langle v, x_k - x \rangle - \frac{r}{2}|x_k - x|^2. \end{aligned}$$

From this we conclude that  $f(x) \leq \liminf_k f(x_k) \leq \limsup_k f(x_k) \leq f(x)$ , as required. □

The following concept will provide a prime source of prox-regularity, showing that this property rules a much wider territory than might at first be thought.

**Definition 2.4.** A function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is *amenable* at  $\bar{x}$  if  $f(\bar{x})$  is finite and there is an open neighborhood  $U$  of  $\bar{x}$  on which  $f$  has a representation as  $g \circ F$  with  $F$  a  $\mathcal{C}^1$  mapping from  $U$  to a space  $\mathbb{R}^m$  and  $g$  a proper, lsc, convex function on  $\mathbb{R}^m$  such that

$$(2.1) \quad \text{there is no vector } y \neq 0 \text{ in } N_{\text{dom } g}(F(\bar{x})) \text{ with } \nabla F(\bar{x})^* y = 0.$$

If the mapping  $F$  is of class  $\mathcal{C}^2$  rather than just  $\mathcal{C}^1$ ,  $f$  is *strongly amenable* at  $\bar{x}$ .

Here  $N_{\text{dom } g}(F(\bar{x}))$  denotes the normal cone to the convex set  $\text{dom } g$  at  $F(\bar{x})$ , while  $\nabla F(\bar{x})^*$  is the adjoint of the Jacobian matrix  $\nabla F(\bar{x})$  for  $F$  at  $\bar{x}$ . The constraint qualification (2.1) is satisfied trivially, of course, if  $F(\bar{x}) \in \text{int}(\text{dom } g)$ .

The terminology of amenability was introduced in [16] and has been utilized further in [17] and [18]; see also [11]. Especially important among strongly amenable functions are the *fully amenable* functions, for which  $F$  is of class  $\mathcal{C}^2$  and  $g$  is not just convex but piecewise linear-quadratic. Such functions are remarkable for their second-order properties in variational analysis, developed in [28], [29] and

[30], which closely relate to the theory of perturbations of solutions to problems of optimization, cf. [31] and more recently [11]. For other perspectives on the study of functions expressible through composition of a smooth mapping with a convex function, see [3], [9], [10], [14], [15], [21].

Strongly amenable functions are p.l.n.; see [14]. This implies that they are prox-regular and subdifferentially continuous (see Proposition 2.3). We record this important fact in the following proposition, and provide a simpler proof (indeed for strongly amenable functions, it is much easier to show that they are prox-regular than to show that they are p.l.n.)

**Proposition 2.5.** *If  $f$  is strongly amenable at  $\bar{x}$ , then  $f$  is prox-regular and subdifferentially continuous at  $\bar{x}$  for every  $\bar{v} \in \partial f(\bar{x})$ .*

*Proof.* Consider a representation  $f = g \circ F$  on a neighborhood  $U$  of  $\bar{x}$  as provided by Definition 2.4. On the basis of this representation,  $f$  is l.s.c. relative to  $U$ , hence in particular locally l.s.c. at  $\bar{x}$ . Let  $\bar{v} \in \partial f(\bar{x})$ . For all  $x$  near  $\bar{x}$  we have  $\partial f(x) = \nabla F(x)^* \partial g(F(x))$ ; see [16]. Thus, for  $x \in U$  the vectors  $v \in \partial f(x)$  are the ones of the form  $v = \nabla F(x)^* y$  for some  $y \in \partial g(F(x))$ . Moreover, there exists  $\epsilon > 0$  such that the set of all  $y$  having this property with respect to  $x$  and  $v$  satisfying  $|x - \bar{x}| < \epsilon$  and  $|v - \bar{v}| < \epsilon$  is bounded in norm, say by  $\eta$  (for otherwise a contradiction to the constraint qualification (2.1) can be obtained); we suppose here that  $\epsilon$  is small enough that  $|x - \bar{x}| < \epsilon$  implies  $x \in U$ . Because  $F$  is of class  $\mathcal{C}^2$ , there exists  $r > 0$  such that

$$\begin{aligned} \langle y, F(x') - F(x) \rangle &\geq \langle \nabla F(x)^* y, x' - x \rangle - \frac{r}{2} |x' - x|^2 \\ &\text{when } |y| \leq \eta, |x - \bar{x}| < \epsilon, |x' - \bar{x}| < \epsilon. \end{aligned}$$

Then as long as  $|x - \bar{x}| < \epsilon$  and  $v \in \partial f(x)$  with  $|v - \bar{v}| < \epsilon$ , we have for any  $y \in \partial g(F(x))$  with  $\nabla F(x)^* y = v$  and any point  $x'$  with  $|x' - \bar{x}| < \epsilon$  that

$$\begin{aligned} f(x') - f(x) &= g(F(x')) - g(F(x)) \geq \langle y, F(x') - F(x) \rangle \\ &\geq \langle \nabla F(x)^* y, x' - x \rangle - \frac{r}{2} |x' - x|^2 = \langle v, x' - x \rangle - \frac{r}{2} |x' - x|^2. \end{aligned}$$

This not only tells us that  $f$  is prox-regular at  $\bar{x}$  but yields the bound  $f(x) - f(x') \leq \eta |F(x') - F(x)|$ . In the same way, if  $f$  has a subgradient  $v'$  at  $x'$  with  $|v' - \bar{v}| < \epsilon$ , we have  $f(x') - f(x) \leq \eta |F(x) - F(x')|$ . Hence, whenever  $(x, v)$  and  $(x', v')$  are sufficiently near to  $(\bar{x}, \bar{v})$  with  $v \in \partial f(x)$  and  $v' \in \partial f(x')$ , we have  $|f(x') - f(x)| \leq \eta |F(x') - F(x)|$ . In particular, we may conclude from the continuity of  $F$  that  $f$  is subdifferentially continuous at  $\bar{x}$ .  $\square$

The magnitude of the class of strongly amenable functions, and therefore that of the prox-regular functions, is evident from the following set of examples.

**Example 2.6.** A proper, l.s.c., convex function  $f$  on  $\mathbb{R}^n$  is strongly amenable at any  $\bar{x} \in \text{dom } f$  for any  $\bar{v} \in \partial f(\bar{x})$ , hence it is prox-regular and subdifferentially continuous for all such  $\bar{x}$  and  $\bar{v}$ .

*Details.* Here we have  $f = g \circ F$  with  $F$  the identity mapping and  $g = f$ .  $\square$

**Example 2.7.** If a function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is lower- $\mathcal{C}^2$  on an open set  $O$ , it is strongly amenable at any  $\bar{x} \in O$  for any  $\bar{v} \in \partial f(\bar{x})$ , hence it is prox-regular and subdifferentially continuous for all such  $\bar{x}$  and  $\bar{v}$ .

*Details.* On a neighborhood of  $\bar{x} \in O$  we can express  $f$  as  $f_0 - \mu | \cdot |^2$  for some finite, convex function  $f_0$  on  $\mathbb{R}^n$  and constant  $\mu$ , cf. [25]. Define  $g$  on  $\mathbb{R} \times \mathbb{R}^n$  by  $g(x_0, x) = x_0 + f_0(x)$  and take  $F$  to be the  $\mathcal{C}^2$  mapping  $x \mapsto (-\mu|x|^2, x)$ . Then  $f = g \circ F$ , and the definition of strong amenability is satisfied on  $O$ .  $\square$

**Example 2.8.** Suppose  $f = f_0 + \delta_C$  for a  $\mathcal{C}^2$  function  $f_0$  on  $\mathbb{R}^n$  and a set  $C \subset \mathbb{R}^n$  specified by the constraint system  $F(x) \in D$ , where  $F$  is a  $\mathcal{C}^2$  mapping from  $\mathbb{R}^n$  to  $\mathbb{R}^m$  and  $D$  is a closed, convex subset of  $\mathbb{R}^m$ . Let  $\bar{x}$  be a point of  $C$  where the following constraint qualification is satisfied:

$$(2.2) \quad \text{there is no vector } y \neq 0 \text{ in } N_D(F(\bar{x})) \text{ with } \nabla F(\bar{x})^* y = 0.$$

Then  $f$  is strongly amenable at  $\bar{x}$  for any  $\bar{v} \in \partial f(\bar{x})$ , and therefore it is prox-regular and subdifferentially continuous at  $\bar{x}$  for all such  $\bar{v}$ .

*Details.* We have  $f = g \circ F_0$  with  $F_0 : x \mapsto (f_0(x), F(x))$  and  $g(u_0, u) = u_0 + \delta_D$ , the latter being a l.s.c. convex function.  $\square$

In the standard case where  $D$  consists of the vectors  $u = (u_1, \dots, u_m)$  such that  $u_i \leq 0$  for  $i = 1, \dots, s$  but  $u_i = 0$  for  $i = s + 1, \dots, m$ , the constraint qualification (2.2) reduces to the Mangasarian-Fromovitz constraint qualification.

**Example 2.9.** Suppose  $f = \max\{f_1, \dots, f_m\}$  for a collection of  $\mathcal{C}^2$  functions  $f_i$  on  $\mathbb{R}^n$ . Then  $f$  is strongly amenable at every point  $\bar{x}$  for every subgradient  $\bar{v} \in \partial f(\bar{x})$ , hence it is everywhere prox-regular and subdifferentially continuous.

*Details.* Here we have  $f = g \circ F$  with  $F(x) = (f_1(x), \dots, f_m(x))$  and  $g(u_1, \dots, u_m) = \max\{u_1, \dots, u_m\}$ , the latter being a finite convex function.  $\square$

A “max function” as in Example 2.9 could also be substituted for the function  $f_0$  in Example 2.8. For further elaboration of the possibilities in producing prox-regular functions from known ones, see Section 6. In Example 2.9,  $f$  is in fact fully amenable, because  $g$  is not just convex but polyhedral convex. The same is true in Example 2.8 when  $D$  is a polyhedral set, as in the standard case that was mentioned.

These examples might give the impression that virtually every prox-regular function falls into the better category of functions that are strongly amenable. The function  $f$  on  $\mathbb{R}$  with  $f(x) = 0$  for  $x \leq 0$  but  $f(x) = 1$  for  $x > 0$ , used already to illustrate that prox-regular functions need not be subdifferentially continuous, not to speak of p.l.n., also shows that they need not be strongly amenable, but the epigraph of this function  $f$  agrees anyway around  $(0, f(0))$  with that of a certain strongly amenable function, the indicator of  $(-\infty, 0]$ . A more convincing illustration is furnished by  $f(x) = 0$  for  $x \leq 0$ ,  $f(x) = \sqrt{x}$  for  $x > 0$ , which is subdifferentially continuous and prox-regular at  $\bar{x} = 0$  for  $\bar{v} = 0$  but is not amenable at all at  $\bar{x}$ . Also illuminating is the similar example of  $f(x) = \sqrt{|x|}$ , which has these properties along with  $\partial f(0) = (-\infty, \infty)$ . Its epigraph has a cusp at  $(0, f(0))$ .

Proximal regularity can be defined for sets as well as for functions, and in this way the geometric side of the concept can be brought into view. For this purpose

we consider general sets  $C \subset \mathbb{R}^n$ , not necessarily convex, and denote by  $N_C(x)$  the set of limiting proximal normals to  $C$  at  $x$ . We speak of  $C$  being *locally closed* at  $\bar{x}$  if the intersection of  $C$  with some neighborhood of  $\bar{x}$  is closed.

**Definition 2.10.** A set  $C \subset \mathbb{R}^n$  is *prox-regular* at  $\bar{x}$  for  $\bar{v}$ , where  $\bar{x} \in C$  and  $\bar{v} \in N_C(\bar{x})$ , if  $C$  is locally closed at  $\bar{x}$  and there exist  $\epsilon > 0$  and  $r > 0$  such that whenever  $x \in C$  and  $v \in N_C(x)$  with  $|x - \bar{x}| < \epsilon$  and  $|v - \bar{v}| < \epsilon$ , then  $x$  is the unique nearest point of  $\{x' \in C \mid |x' - \bar{x}| < \epsilon\}$  to  $x + r^{-1}v$ .

In order to connect this with the prox-regularity of functions, we invoke the fact that  $N_C(x) = \partial\delta_C(x)$  for the indicator function  $\delta_C$  (which takes the value 0 on  $C$  but  $\infty$  everywhere else).

**Proposition 2.11.** A set  $C \subset \mathbb{R}^n$  is prox-regular at a point  $\bar{x} \in C$  for a vector  $\bar{v}$  if and only if its indicator function  $\delta_C$  is prox-regular at  $\bar{x}$  for  $\bar{v}$ .

*Proof.* We simply observe that the inequality (1.4) in the definition of prox-regularity in the case of  $f = \delta_C$  reduces to having

$$(2.3) \quad 0 > \langle v, x' - x \rangle - \frac{r}{2}|x' - x|^2$$

for all  $x' \in C$  with  $|x' - \bar{x}| < \epsilon$  and  $x' \neq x$ , when  $x \in C$  and  $v \in N_C(x)$  satisfy  $|x - \bar{x}| < \epsilon$  and  $|v - \bar{v}| < \epsilon$ . But (2.3) can equally well be written as

$$|x - (x + r^{-1}v)|^2 < |x' - (x + r^{-1}v)|^2,$$

so this condition means that  $x$  is the unique closest point of  $\{x' \in C \mid |x' - \bar{x}| < \epsilon\}$  to  $x + r^{-1}v$ .  $\square$

Convex sets, “weakly convex” sets (see [35]) and “proximally smooth” sets (see [6]) are all examples of prox-regular sets. Strong amenability provides further examples: a set  $C \subset \mathbb{R}^n$  is *strongly amenable* at one of its points  $\bar{x}$  if its indicator function  $\delta_C$  is strongly amenable at  $\bar{x}$ , or equivalently, there is an open neighborhood  $U$  of  $\bar{x}$  and a  $\mathcal{C}^2$  mapping  $F : U \rightarrow \mathbb{R}^m$  along with a closed, convex set  $D \subset \mathbb{R}^m$  such that

$$C \cap U = \{x \in U \mid F(x) \in D\},$$

and the constraint qualification is satisfied that no nonzero vector  $y \in N_D(F(\bar{x}))$  has  $\nabla F(\bar{x})^*y = 0$ . This definition parallels that given in [16] for the amenability and full amenability of sets, the first corresponding to the weaker assumption that  $F$  is only of class  $\mathcal{C}^1$  and the second insisting that  $D$  be polyhedral. It means that  $C$  can be represented locally as the set of points satisfying a nice kind of constraint system.

**Corollary 2.12.** If a set  $C$  is strongly amenable at  $\bar{x}$ , then  $C$  is prox-regular at  $\bar{x}$  for every vector  $\bar{v} \in N_C(\bar{x})$ .

*Proof.* This comes from Proposition 2.5 in light of the equivalence in Proposition 2.11.  $\square$

3. CHARACTERIZATIONS OF PROX-REGULARITY

We occupy ourselves next with alternative descriptions of prox-regularity which furnish useful insights in dealing with this property. Everything is local in reference to a pair  $(\bar{x}, \bar{v})$  with  $\bar{v} \in \partial f(\bar{x})$ , and it is essential that this be  $f$ -attentive localization in the following sense.

**Definition 3.1.** An  $f$ -attentive localization of  $\partial f$  around  $(\bar{x}, \bar{v})$ , is a (generally set-valued) mapping  $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  whose graph in  $\mathbb{R}^n \times \mathbb{R}^n$  is the intersection of  $\text{gph } \partial f$  with the product of an  $f$ -attentive neighborhood of  $\bar{x}$  and an ordinary neighborhood of  $\bar{v}$ ; this contrasts with an ordinary localization, in which the  $f$ -attentive neighborhood of  $\bar{x}$  is relaxed to an ordinary neighborhood. More specifically for an  $\epsilon > 0$ , the  $f$ -attentive  $\epsilon$ -localization of  $\partial f$  around  $(\bar{x}, \bar{v})$ , is the mapping  $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  defined by

$$(3.1) \quad T(x) = \begin{cases} \{v \in \partial f(x) \mid |v - \bar{v}| < \epsilon\} & \text{if } |x - \bar{x}| < \epsilon \text{ and } |f(x) - f(\bar{x})| < \epsilon, \\ \emptyset & \text{otherwise.} \end{cases}$$

In the examples seen above where  $f$  is subdifferentially continuous, we would not have to concern ourselves with  $f$ -attentive localization and could work with ordinary localization of  $\partial f$ . But in the absence of that continuity property some of the technical arguments on which our results will be based would fail unless we took the  $f$ -attentive approach. Anyway, this is a natural approach because it restricts the localization to subgradients  $v \in \partial f(x)$  associated with points  $(x, f(x))$  of  $\text{epi } f$  that are close to  $(\bar{x}, f(\bar{x}))$ , and it merely serves therefore to emphasize the epigraphical geometry on which the theory of subgradients is based.

**Theorem 3.2.** *When  $f$  is locally l.s.c. at  $\bar{x}$ , the following are equivalent.*

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

$$(3.2) \quad \langle v_1 - v_0, x_1 - x_0 \rangle \geq -r|x_1 - x_0|^2 \text{ when } v_i \in T(x_i), \ i = 0, 1.$$

- (c) *There exist  $\epsilon > 0$  and  $r > 0$ , with*

$$(3.3) \quad f(x) > f(\bar{x}) + \langle \bar{v}, x - \bar{x} \rangle - \frac{r}{2}|x - \bar{x}|^2 \text{ when } 0 < |x - \bar{x}| < \epsilon,$$

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

$$(3.4) \quad x_i \in (\partial f + rI)^{-1}(z) \text{ with } |x_i - \bar{x}| < \epsilon, \ |f(x_i) - f(\bar{x})| < \epsilon,$$

*then necessarily  $x_0 = x_1$ .*

*Proof.* In order to simplify some of the calculations that will be needed, we can assume without any loss of generality that  $\bar{x} = 0$ ,  $\bar{v} = 0$ , and  $f(0) = 0$ .

(a)  $\Rightarrow$  (b). Take  $\epsilon$  and  $r$  from Definition 1.1 of prox-regular, and for the same  $\epsilon$  let  $T$  be the  $f$ -attentive  $\epsilon$ -localization of  $\partial f$  as in (3.1). For  $i = 0, 1$ , let  $v_i \in T(x_i)$ . From prox-regularity we have

$$f(x_1) \geq f(x_0) + \langle v_0, x_1 - x_0 \rangle - \frac{r}{2}|x_1 - x_0|^2,$$

but by symmetry also

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

In adding these inequalities together, we get the inequality in (3.2).

(b)  $\Rightarrow$  (c). Let  $\bar{\epsilon}$  and  $\bar{r}$  be parameter values such that when  $T$  is the  $f$ -attentive  $\bar{\epsilon}$ -localization of  $\partial f$  the property in (b) holds for  $\bar{r}$ . For some  $\epsilon \in (0, \bar{\epsilon})$  and  $r \in (\bar{r}, \infty)$  we have (3.3), inasmuch as  $\bar{v} \in \partial_p f(\bar{x})$ . Decreasing  $\epsilon$  and increasing  $r$  further if necessary, we can arrange that  $\epsilon < \bar{\epsilon}/2r$  with  $r \geq 1$ , hence  $\epsilon < \bar{\epsilon}/2$ . Suppose now that (3.4) holds with  $|z - (\bar{v} + r\bar{x})| < \epsilon$ ; recall here that we have normalized our setting so that this simplifies to  $|z| < \epsilon$ . For  $v_i = z - rx_i$  we have  $v_i \in \partial f(x_i)$  and  $|v_i| \leq |z| + r|x_i| < \frac{1}{2}\bar{\epsilon} + \frac{1}{2}\bar{\epsilon} = \bar{\epsilon}$  and also  $|x_i| < \bar{\epsilon}$  and  $|f(x_i)| < \bar{\epsilon}$ . Thus,  $v_i \in T(x_i)$ . From the property in (b) we deduce that

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

Because  $r > \bar{r}$ , we conclude that  $x_1 = x_0$ , as required.

(c)  $\Rightarrow$  (a). This is the hardest part. Starting with  $\epsilon$  and  $r$  as in (c), it will suffice to demonstrate the existence of  $\bar{\epsilon} \in (0, \epsilon)$  such that the definition of prox-regularity of  $f$  is satisfied by  $\bar{\epsilon}$  and  $r$  at  $\bar{x} = 0$  for  $\bar{v} = 0$ : whenever

$$(3.5) \quad v_0 \in \partial f(x_0) \text{ with } |v_0| < \bar{\epsilon}, |x_0| < \bar{\epsilon}, |f(x_0)| < \bar{\epsilon},$$

one will have

$$f(x) \geq f(x_0) + \langle v_0, x - x_0 \rangle - \frac{r}{2}|x - x_0|^2 \text{ when } |x| < \bar{\epsilon}.$$

In place of the latter we can aim at guaranteeing the stronger condition

$$f(x) > f(x_0) + \langle v_0, x - x_0 \rangle - \frac{r}{2}|x - x_0|^2 \text{ when } 0 < |x| \leq \epsilon.$$

Because  $-(r/2)|x - x_0|^2 = \langle rx_0, x - x_0 \rangle - (r/2)|x|^2 - (r/2)|x_0|^2$ , we can write this as

$$f(x) - \langle v_0 + rx_0, x \rangle + \frac{r}{2}|x|^2 > f(x_0) - \langle v_0 + rx_0, x_0 \rangle + \frac{r}{2}|x_0|^2 \text{ when } 0 < |x| \leq \epsilon,$$

which means that

$$(3.6) \quad \operatorname{argmin}_{|x| \leq \epsilon} \left\{ f(x) + \frac{r}{2}|x|^2 - \langle z_0, x \rangle \right\} = \{x_0\} \text{ for } z_0 = v_0 + rx_0.$$

In summary, we need only demonstrate that when  $\bar{\epsilon}$  is small enough, (3.5) implies (3.6).

Our assumption that  $f$  is locally l.s.c. at 0 with  $f(0) = 0$  ensures that  $f$  is l.s.c. with respect to a compact set of the form

$$C = \{x \mid |x| \leq \hat{\epsilon}, f(x) \leq 2\hat{\epsilon}\}$$

for some  $\hat{\epsilon} > 0$ . In particular  $f$  must be l.s.c. at 0 itself:  $\liminf_{x \rightarrow 0} f(x) = 0$ . Replacing  $\epsilon$  by a smaller value and  $r$  by a larger value if necessary, which entails no loss of generality, we can arrange (for reasons yet to emerge) that

$$(3.7) \quad |x| \leq \epsilon \implies |x| \leq \hat{\epsilon}, f(x) > -2\hat{\epsilon},$$

$$(3.8) \quad |x|^2 < \frac{6\hat{\epsilon}}{r} \implies |x| < \epsilon, |x| < \frac{\epsilon^2}{2\hat{\epsilon}}, f(x) > -\frac{\epsilon}{2}.$$

With this background secured, we define

$$g(z) = \inf_{|x| \leq \epsilon} \left\{ f(x) + \frac{r}{2}|x|^2 - \langle z, x \rangle \right\},$$

$$G(z) = \operatorname{argmin}_{|x| \leq \epsilon} \left\{ f(x) + \frac{r}{2}|x|^2 - \langle z, x \rangle \right\}.$$

As the pointwise infimum of a collection of affine functions of  $z$ ,  $g$  is concave. Our assumptions tell us that  $g(0) = 0$  and  $G(0) = \{0\}$ , whereas  $g(z) \leq 0$  in general. Readers may recognize in  $-g$  the localized quadratic conjugate of  $f$  as developed in [12].

We claim that under the circumstances we have engineered,

$$(3.9) \quad |z| < \frac{\hat{\epsilon}}{\epsilon} \implies \begin{cases} G(z) \neq \emptyset, \\ |f(x)| < 2\hat{\epsilon} \text{ for all } x \in G(z). \end{cases}$$

To see this, observe that because  $g \leq 0$ , the maximization in the definition of  $g(z)$  is unaffected if attention is restricted to the points  $x$  satisfying not only  $|x| \leq \epsilon$  but also  $f(x) + (r/2)|x|^2 - \langle z, x \rangle \leq \hat{\epsilon}$ , in which case  $f(x) \leq \hat{\epsilon} + \epsilon|z|$ . Therefore, as long as  $|z| < \hat{\epsilon}/\epsilon$ , attention can be restricted to points  $x$  satisfying  $f(x) < 2\hat{\epsilon}$ . Recalling (3.7) and the choice of the set  $C$ , we deduce that

$$(3.10) \quad \left. \begin{aligned} g(z) &= \inf_{x \in C} \left\{ f(x) + \frac{r}{2}|x|^2 - \langle z, x \rangle \right\} \\ G(z) &= \operatorname{argmin}_{x \in C} \left\{ f(x) + \frac{r}{2}|x|^2 - \langle z, x \rangle \right\} \end{aligned} \right\} \text{ when } |z| \leq \frac{\hat{\epsilon}}{\epsilon}.$$

But  $f$  is l.s.c. relative to  $C$ , which is compact, so the infimum is sure to be attained. Thus, (3.9) is correct. Also we observe from (3.10) that, on a neighborhood of  $z = 0$ ,  $g$  is finite, hence continuous (by concavity).

Choose  $\bar{\epsilon} \in (0, \epsilon)$  small enough that  $\bar{\epsilon}(1+r) < \hat{\epsilon}/\epsilon$  and  $g(z) > -\epsilon/2$  when  $|z| < \bar{\epsilon}(1+r)$ . Under this choice we are ready to consider elements satisfying (3.5) and show that we get (3.6). The vector  $z_0 = v_0 + rx_0$  has  $|z_0| \leq |v_0| + r|x_0| < \bar{\epsilon}(1+r)$ , hence  $|z_0| < \hat{\epsilon}/\epsilon$  and  $g(z_0) > -\epsilon/2$ . In particular,  $G(z_0)$  must be nonempty.

Consider any  $x_1 \in G(z_0)$ . We have to show on the basis of the single-valuedness property in (c) that  $x_1$  must be  $x_0$ .

On the one hand we have  $x_0 \in (\partial f + rI)^{-1}(z_0)$  (due to the fact that  $v_0 \in \partial f(x_0)$ ), hence  $v_0 + rx_0 \in (\partial f + rI)(x_0)$ . If we can establish on the other hand that

$$(3.11) \quad x_1 \in (\partial f + rI)^{-1}(z_0) \text{ with } |x_1| < \epsilon, |f(x_1)| < \epsilon,$$

the property in (c) can be invoked and we will get  $x_1 = x_0$  as required.

Because  $|z_0| < \hat{\epsilon}/\epsilon$ , we know from (3.9) that  $|f(x_1)| < 2\hat{\epsilon}$ . Then from  $(r/2)|x_1|^2 = \langle z_0, x_1 \rangle - f(x_1) + g(z_0)$ , where  $g(z_0) \leq 0$ , we estimate that  $(r/2)|x_1|^2 \leq |z_0||x_1| + |f(x_1)| < (\hat{\epsilon}/\epsilon)\epsilon + 2\hat{\epsilon} = 3\hat{\epsilon}$ , so  $|x_1|^2 < 6\hat{\epsilon}/r$ . Through (3.8) this implies that  $|x_1| < \epsilon$  and also that  $|x_1| < \epsilon^2/2\hat{\epsilon}$  and  $f(x_1) > -\epsilon$ . Then from  $f(x_1) = \langle z_0, x_1 \rangle - (r/2)|x_1|^2 + g(z_0) \leq |z_0||x_1| + |g(z_0)|$  and the fact that  $g(z_0) < \epsilon/2$  we obtain  $f(x_1) < (\hat{\epsilon}/\epsilon)(\epsilon^2/2\hat{\epsilon}) + (\epsilon/2) = \epsilon$ . Hence the inequalities in (3.11) are true. The fact that the minimum for  $g(z_0)$  is attained at  $x_1$  gives us

$$f(x) + \frac{r}{2}|x|^2 - \langle z_0, x \rangle \geq f(x_1) + \frac{r}{2}|x_1|^2 - \langle z_0, x_1 \rangle \text{ when } |x| \leq \epsilon,$$

which translates to

$$f(x) \geq f(x_1) + \langle z_0 - rx_1, x - x_1 \rangle - \frac{r}{2}|x - x_1|^2 \text{ when } |x| \leq \epsilon.$$

Since  $|x_1| < \epsilon$ , this implies that  $z_0 - rx_1 \in \partial_p f(x_1)$ , so  $z_0 \in (\partial f + rI)(x_1)$ . The subgradient relation in (3.11) is therefore correct as well, and we are done.  $\square$

*Remark 3.3.* The proof of Theorem 3.2 shows specifically that when  $f$  is prox-regular at  $\bar{x}$  for  $\bar{v}$  with respect to the parameter values  $\epsilon > 0$  and  $r > 0$ , and  $T$  is the  $f$ -attentive  $\epsilon$ -localization of  $\partial f$  around  $(\bar{x}, \bar{v})$ , then  $T + rI$  is monotone.

It will be shown in Proposition 4.8 that the monotonicity of  $T + rI$  in condition (b) of Theorem 3.2 is in fact “locally maximal”.

The facts in Theorem 3.2 can be applied to the prox-regularity of sets by way of the identification of that property in Proposition 2.11 with the prox-regularity of indicator functions.

**Corollary 3.4.** *For  $C \subset \mathbb{R}^n$  and  $\bar{x} \in C$  with  $C$  locally closed at  $\bar{x}$ , the following are equivalent.*

- (a) *The set  $C$  is prox-regular at  $\bar{x}$  for  $\bar{v}$ .*
- (b) *The vector  $\bar{v}$  is a proximal normal to  $C$  at  $\bar{x}$ , and there exist  $\epsilon > 0$  and  $r > 0$  with*

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

*whenever for  $i = 0, 1$  one has  $x_i \in C$  and  $v_i \in N_C(x_i)$  with  $|x_i - \bar{x}| < \epsilon$  and  $|v_i - \bar{v}| < \epsilon$ .*

- (c) *There exist  $\epsilon > 0$  and  $r > 0$ , such that  $\bar{x}$  is the unique nearest point to  $\bar{x} + r^{-1}\bar{v}$  in the set  $\{x \in C \mid |x - \bar{x}| < \epsilon\}$ , and in addition, the following property holds: whenever  $z$  is a vector satisfying  $|z - (\bar{v} + r\bar{x})| < \epsilon$ , and for  $i = 0, 1$  one has*

$$x_i \in (N_C + rI)^{-1}(z), \quad |x_i - \bar{x}| < \epsilon, \quad x_i \in C,$$

*then necessarily  $x_0 = x_1$ .*

*Proof.* This is evident.  $\square$

To close the circle of geometry, we apply the notion of prox-regularity for sets back to functions by way of their epigraphs.

**Theorem 3.5.** *A function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is prox-regular at  $\bar{x}$  for  $\bar{v}$  if and only if its epigraph set  $\text{epi } f \subset \mathbb{R}^n \times \mathbb{R}$  is prox-regular at  $(\bar{x}, f(\bar{x}))$  for  $(\bar{v}, -1)$ .*

*Proof.* There is no loss of generality in focusing on the case where  $\bar{v} = 0$ , where calculations are simpler. This follows from the fact that the prox-regularity of  $f$  at  $\bar{x}$  for  $\bar{v}$  is equivalent to the prox-regularity of  $f - \langle \bar{v}, \cdot \rangle$  at  $\bar{x}$  for 0. At the same time, the epigraphs of  $f$  and  $f - \langle \bar{v}, \cdot \rangle$  differ only by an invertible affine transformation in  $\mathbb{R}^n \times \mathbb{R}$ , and such a transformation preserves the prox-regularity of sets. Instead of actually replacing  $f$  by  $f - \langle \bar{v}, \cdot \rangle$ , we can just consider  $\bar{v}$  to be the zero vector.

*Necessity.* Assume that  $f$  is prox-regular at  $\bar{x}$  for  $\bar{v} = 0$ , and take  $\epsilon > 0$  and  $r > 0$  from Definition 1.1. We first establish two claims.

*Claim 1.* Take  $R > r$ . Suppose that  $v \in \partial f(x)$  with  $|v| < \epsilon$ ,  $|f(x) - f(\bar{x})| < \epsilon$ , and  $|x - \bar{x}| < \epsilon'$ , where  $\epsilon' = \min \{\epsilon/2, 1/4R, 1/8R\epsilon\}$ . Then for any  $x' \neq x$  with  $|x' - \bar{x}| < \epsilon'$  and any  $\alpha \geq f(x')$ , we have

$$\begin{aligned} \text{dist}^2\left((x', \alpha), (x, f(x)) + \left(\frac{v}{2R}, \frac{-1}{2R}\right)\right) &= \left|x' - x - \frac{v}{2R}\right|^2 + \left(\alpha - f(x) + \frac{1}{2R}\right)^2 \\ &> \frac{|v|^2}{4R^2} + \frac{1}{4R^2} \\ &= \text{dist}^2\left((x, f(x)), (x, f(x)) + \left(\frac{v}{2R}, \frac{-1}{2R}\right)\right). \end{aligned}$$

*Proof of Claim 1.* First note that whenever  $x' \neq x$ ,  $|x' - \bar{x}| < \epsilon$ , one has

$$\alpha + \frac{1}{2R} \geq f(x') + \frac{1}{2R} > f(x) + \langle v, x' - x \rangle - \frac{R}{2}|x' - x|^2 + \frac{1}{2R},$$

hence,

$$\alpha - f(x) + \frac{1}{2R} > \langle v, x' - x \rangle - \frac{R}{2}|x' - x|^2 + \frac{1}{2R}.$$

It is easy to establish that the right side of this inequality is positive when  $|x' - x| < 1/4R$  and  $|x - \bar{x}||v| < 1/8R$ , as is true when  $|x' - x| < \epsilon'$  and  $|v| < \epsilon$ , in which case the inequality can be written equivalently as

$$\left(\alpha - f(x) + \frac{1}{2R}\right)^2 > \left[\langle v, x' - x \rangle - \frac{R}{2}|x' - x|^2 + \frac{1}{2R}\right]^2.$$

We calculate then that

$$\begin{aligned} &\left(\alpha - f(x) + \frac{1}{2R}\right)^2 + \left|x' - x - \frac{v}{2R}\right|^2 \\ &> \left[\langle v, x' - x \rangle - \frac{R}{2}|x' - x|^2 + \frac{1}{2R}\right]^2 + \left|x' - x - \frac{v}{2R}\right|^2 \\ &= \left[\langle v, x' - x \rangle - \frac{R}{2}|x' - x|^2\right]^2 + \frac{1}{4R^2} + \frac{1}{R} \left[\langle v, x' - x \rangle - \frac{R}{2}|x' - x|^2\right] \\ &\quad + |x' - x|^2 - \left\langle \frac{v}{R}, x' - x \right\rangle + \frac{|v|^2}{4R^2} \\ &= \left[\frac{|v|^2}{4R^2} + \frac{1}{4R^2}\right] + |x' - x|^2 + \left[\langle v, x' - x \rangle - \frac{R}{2}|x' - x|^2\right]^2 - \frac{1}{2}|x' - x|^2 \\ &= \left[\frac{|v|^2}{4R^2} + \frac{1}{4R^2}\right] + \frac{1}{2}|x' - x|^2 + \left[\langle v, x' - x \rangle - \frac{R}{2}|x' - x|^2\right]^2 \\ &\geq \frac{|v|^2}{4R^2} + \frac{1}{4R^2}, \end{aligned}$$

and this completes the proof of Claim 1.

*Claim 2.* If  $(v, b)$  is a proximal normal to  $\text{epi } f$  at  $(x, \alpha)$  with  $b \neq 0$ , then  $\alpha = f(x)$ .

*Proof of Claim 2.* By rescaling if necessary, we may assume that for  $c \geq f(x)$  small enough and  $c \neq \alpha$  we have

$$|(x, c) - [(x, \alpha) + (v, b)]|^2 > |(v, b)|^2.$$

This implies that  $(c - \alpha - b)^2 > b^2$  whenever  $c$  is close to  $\alpha$  with  $\alpha \neq c$ . From this we would arrive at a contradiction if  $\alpha \neq f(x)$ . This proves Claim 2.

*Remainder of the proof of necessity.*

Take  $R > 4r$  and  $\delta$  positive with  $\delta < \min \{1/2, \epsilon/4, 1/8R\epsilon, 1/4R\}$ . Take  $|x - \bar{x}| < \delta$  and  $|\beta - f(\bar{x})| < \delta$  with  $\beta \geq f(x)$ , and consider a proximal normal  $(v, b)$  to  $\text{epi } f$  at  $(x, \beta)$  with  $|(v, b) - (0, -1)| < \delta$ . Because  $\delta < 1/2$  we have  $b < 0$  (in fact  $-3/2 < b < -1/2$ ), and it follows then that  $\beta = f(x)$ ; cf. Claim 2. We have  $(v, b) = -b(-v/b, -1)$ , so  $-v/b$  is a proximal subgradient,  $-v/b \in \partial_p f(x)$ ; cf. [24]. We also have  $|-v/b| < \epsilon$ . Let  $\tilde{R} = R/(-2b)$ . Note that  $\tilde{R} > r$  and  $\tilde{R} < R$  (because  $1 < -2b < 3$ ). By Claim 1 we then have that

$$\text{dist}\left((x', \alpha), (x, f(x)) + \left(\frac{-v}{2b\tilde{R}}, \frac{-1}{2\tilde{R}}\right)\right) > \text{dist}\left((x, f(x)), (x, f(x)) + \left(\frac{-v}{2b\tilde{R}}, \frac{-1}{2\tilde{R}}\right)\right)$$

whenever  $|(x', \alpha) - (\bar{x}, f(\bar{x}))| < \delta$ . In other words

$$\text{dist}\left((x', \alpha), (x, f(x)) + (1/R)(v, b)\right) > \text{dist}\left((x, f(x)), (x, f(x)) + (1/R)(v, b)\right)$$

whenever  $|(x', \alpha) - (\bar{x}, f(\bar{x}))| < \delta$ . This shows that  $\text{epi } f$  is prox-regular at  $(\bar{x}, f(\bar{x}))$  for the vector  $(0, -1)$ .

*Sufficiency.* Assume that  $\text{epi } f$  is prox-regular at  $(\bar{x}, f(\bar{x}))$  for  $(0, -1)$ . This implies the existence of  $\epsilon > 0$  and  $r > 0$  such that whenever  $(v, -1) \in N_{\text{epi } f}(x, f(x))$  with  $|(x, f(x)) - (\bar{x}, f(\bar{x}))| < \epsilon$  and  $|(v, -1) - (0, -1)| < \epsilon$ , then  $(x, f(x))$  is the unique closest point of  $\text{epi } f$  to  $(x, f(x)) + r^{-1}(v, -1)$ . The relation  $(v, -1) \in N_{\text{epi } f}(x, f(x))$  is equivalent to  $v \in \partial f(x)$ . In terms of function  $g$  having as its graph the upper surface of the closed ball of radius

$$t = |r^{-1}(v, -1)| = r^{-1}\sqrt{|v|^2 + 1}$$

around  $(x, f(x)) + r^{-1}(v, -1)$  (this being a ball that touches  $\text{epi } f$  only at  $(x, f(x))$ ), the closest point condition means that  $f(x') > g(x')$  for all  $x' \neq x$  with  $|x' - (x + r^{-1}v)| \leq t$ . Since the surface of the ball consists of the points  $(x', \alpha')$  satisfying

$$|(x', \alpha') - [(x, f(x)) + r^{-1}(v, -1)]|^2 = |r^{-1}(v, -1)|^2,$$

the formula for  $g$  works out to

$$g(x') = f(x) - r^{-1} + \sqrt{|r^{-1}(v, -1)|^2 - |x' - (x + r^{-1}v)|^2}.$$

Thus we have

$$f(x') > f(x) - \frac{1}{r} + \sqrt{t^2 - \left|x' - \left(x + \frac{v}{r}\right)\right|^2} \text{ when } x' \neq x, |x' - (x + r^{-1}v)| \leq t,$$

where moreover

$$|x' - (x + r^{-1}v)| < t \iff |x' - x|^2 + 2\langle x' - x, r^{-1}v \rangle < r^{-2}.$$

Choose  $\epsilon' \in (0, \epsilon/2)$  small enough that the latter holds for all  $x', x, v$ , satisfying

$$(3.12) \quad |x' - \bar{x}| < \epsilon', |x - \bar{x}| < \epsilon', |f(x) - f(\bar{x})| < \epsilon', |v| < \epsilon'.$$

These conditions entail having  $|(v, -1) - (0, -1)| < \epsilon$  and  $|(x, f(x)) - (\bar{x}, f(\bar{x}))| < \epsilon$ , so they give us

$$(3.13) \quad f(x') > f(x) - \frac{1}{r} + \sqrt{t^2 - \left|x' - \left(x + \frac{v}{r}\right)\right|^2} \text{ under (3.12) when } x' \neq x, v \in \partial f(x).$$

*Claim 3.* As long as  $|x' - x| < 1/4r$  and  $|v| < 1/2$ , one has

$$(3.14) \quad -\frac{1}{r} + \sqrt{t^2 - \left|x' - \left(x + \frac{v}{r}\right)\right|^2} \geq \langle v, x' - x \rangle - \frac{19r}{2}|x' - x|^2.$$

Once this claim is established, we will know immediately from (3.13) that  $f$  is prox-regular at  $\bar{x}$  for  $\bar{v} = 0$  with respect to the parameter values  $\bar{r} = 19r$  and  $\bar{\epsilon} \in (0, \epsilon')$  small enough that  $\bar{\epsilon} < 1/4r$  and  $\bar{\epsilon} < 1/2$ .

*Proof of Claim 3.* Let  $u := x' - (x + r^{-1}v)$ . First note that

$$\begin{aligned} \frac{1}{r} + \langle v, x' - x \rangle - \frac{19r}{2}|x' - x|^2 &= \frac{1}{r} + \langle v, x' - x \rangle - \frac{r}{2}|x' - x|^2 - 9r|x' - x|^2 \\ &= \frac{1}{r} + \langle v, u + \frac{v}{r} \rangle - \frac{r}{2}\left|u + \frac{v}{r}\right|^2 - 9r|x' - x|^2 \\ &= \frac{1}{r} + \langle v, u \rangle + \frac{|v|^2}{r} - \frac{r|u|^2}{2} - \langle u, v \rangle - \frac{|v|^2}{2r} - 9r|x' - x|^2 \\ &= \frac{1}{r} + \frac{|v|^2}{2r} - \frac{r|u|^2}{2} - 9r|x' - x|^2. \end{aligned}$$

Hence

$$\begin{aligned} r \left[ \frac{1}{r} + \langle v, x' - x \rangle - \frac{19r}{2}|x' - x|^2 \right] &= 1 + \frac{|v|^2}{2} - \frac{r^2|u|^2}{2} - 9r^2|x' - x|^2 \\ &= \frac{1}{2}[1 + s] - 9r^2|x' - x|^2, \end{aligned}$$

where

$$\begin{aligned} s &:= 1 + |v|^2 - r^2|u|^2 = 1 + |v|^2 - |ur|^2 \\ &= 1 + |v|^2 - |r(x' - x) - v|^2 = 1 - r^2|x' - x|^2 + 2r\langle v, x' - x \rangle, \end{aligned}$$

and consequently

$$s \geq 1 - r^2|x' - x|^2 - 2r|v||x' - x| > 1 - r^2\left(\frac{1}{16r^2}\right) - 2r|v|\left(\frac{1}{4r}\right) > 1 - \frac{1}{16} - \frac{|v|}{2} > 0$$

under our stipulation that  $|x' - x| < 1/4r$  and  $|v| < 1/2$ . Note further then that

$$r\sqrt{t^2 - \left|x' - \left(x + \frac{v}{r}\right)\right|^2} = r\sqrt{\frac{1}{r^2} + \frac{|v|^2}{r^2} - |u|^2} = \sqrt{1 + |v|^2 - r^2|u|^2} = \sqrt{s}.$$

We can therefore translate the desired inequality (3.14) into

$$(3.15) \quad \sqrt{s} \geq \frac{1}{2}[1 + s] - 9r^2|x' - x|^2.$$

All we have to do is verify that, by insisting not only on  $|v| < 1/2$  but  $|x| < 1/4r$ , we ensure that (3.15) holds. We can write (3.15) equivalently as

$$0 \geq \frac{1}{2}[1 - 2\sqrt{s} + s] - 9r^2|x' - x|^2 = \frac{1}{2}[1 - \sqrt{s}]^2 - 9r^2|x' - x|^2$$

and thus pose the goal in terms of having

$$(3.16) \quad 9r^2|x' - x|^2 \geq [1 - \sqrt{s}]^2.$$

The cases where  $[1 - \sqrt{s}] > 0$  or  $[1 - \sqrt{s}] \leq 0$  can then be analyzed separately.

If  $[1 - \sqrt{s}] > 0$  in (3.16), what we need is  $[1 - \sqrt{s}] \leq 3r|x' - x|$ . This means that  $1 - 3r|x' - x| \leq \sqrt{s}$ . Since we know  $|x' - x| \leq 1/4r$ , all we have to check is whether

$$1 - 6r|x' - x| + 9r^2|x' - x|^2 \leq s = 1 - r^2|x' - x|^2 + 2r\langle v, x' - x \rangle.$$

This comes down to  $10r^2|x' - x|^2 - 6r|x' - x| - 2r\langle v, x' - x \rangle \leq 0$ , i.e.

$$|x' - x| \left[ 10r^2|x' - x| - 6r - 2r\left\langle \frac{x' - x}{|x' - x|}, v \right\rangle \right] \leq 0.$$

Because  $|v| < 1/2$  we have

$$10r^2|x' - x| - 6r - 2r\left\langle \frac{x' - x}{|x' - x|}, v \right\rangle \leq 10r^2|x' - x| - 6r + r.$$

This last expression can be made nonpositive when  $|x' - x| \leq 1/2r$ , which is true here.

If  $[1 - \sqrt{s}] \leq 0$  in (3.16), what we need is  $[\sqrt{s} - 1] \leq 3r|x' - x|$ . This means that  $1 + 3r|x' - x| \geq \sqrt{s}$ . All we have to check is whether

$$1 + 6r|x' - x| + 9r^2|x' - x|^2 \geq s = 1 - r^2|x' - x|^2 + 2r\langle v, x' - x \rangle.$$

This comes down to  $10r^2|x' - x|^2 + 6r|x' - x| - 2r\langle v, x' - x \rangle \geq 0$ , i.e.

$$|x' - x| \left[ 10r^2|x' - x| + 6r - 2r\left\langle \frac{x' - x}{|x' - x|}, v \right\rangle \right] \geq 0.$$

Because  $|v| < 1/2$  we have

$$10r^2|x' - x| + 6r - 2r\left\langle \frac{x' - x}{|x' - x|}, v \right\rangle \geq 10r^2|x' - x| + 6r - r.$$

This last expression is obviously nonnegative. This completes the proof of Claim 3 and therefore that of the theorem.  $\square$

4. REGULARITY PROPERTIES OF MOREAU ENVELOPES

In this section we study the envelope functions  $e_\lambda$  and proximal mappings  $P_\lambda$  associated with a function  $f$  in (1.1) and (1.2) for the purpose of learning more about the behavior of  $f$  around a point  $\bar{x}$  where  $f$  is prox-regular for a vector  $\bar{v} \in \partial f(\bar{x})$ . In such a situation  $\bar{v}$  is actually a proximal subgradient of  $f$  at  $\bar{x}$ , as we know from Theorem 3.2. The analysis is greatly simplified by normalizing to the case where  $\bar{x} = 0$  and  $\bar{v} = 0$ , along with  $f(\bar{x}) = 0$ , and taking the quadratic inequality in the proximal subgradient condition for  $\bar{v}$  to be global. We work therefore under the baseline assumptions that

$$(4.1) \quad \begin{cases} f \text{ is locally l.s.c. at } 0 \text{ with } f(0) = 0, \text{ and} \\ r > 0 \text{ is such that } f(x) > -\frac{r}{2}|x|^2 \text{ for all } x \neq 0, \end{cases}$$

which imply that

$$(4.2) \quad e_\lambda(0) = 0 \text{ and } P_\lambda(0) = \{0\} \text{ when } \lambda \in (0, 1/r).$$

These assumptions entail no real loss of generality for the task we have set before us. The shift to  $\bar{x} = 0$ ,  $\bar{v} = 0$ , and  $f(\bar{x}) = 0$  is nothing more than a change of variables, followed by subtracting a constant from  $f$ . We are interested primarily in properties of  $f$  and  $\partial f$  that depend only on the local geometry of  $\text{epi } f$  around  $(\bar{x}, f(\bar{x}))$ , and for this we can even, if necessary, add to  $f$  the indicator of some compact neighborhood of  $\bar{x}$  so as to make  $\text{dom } f$  be bounded. By taking that neighborhood small enough we can get the quadratic inequality for  $\bar{v}$  to hold for all  $x \neq \bar{x}$ . Alternatively, under the assumption that  $f$  is bounded below by some quadratic the latter could be arranged simply by taking  $r$  sufficiently large.

We begin with some estimates that depend only on these assumptions and do not yet actually call for  $f$  to be prox-regular at  $\bar{x} = 0$  for  $\bar{v} = 0$ .

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

$$(4.3) \quad f(x') + \frac{1}{2\lambda}|x' - x|^2 \leq e_\lambda(x) + \rho \implies \begin{cases} |x'| \leq 2\mu|x| + \sqrt{2\lambda\mu\rho}, \\ f(x') \leq \frac{1}{2\lambda}|x|^2 + \rho, \\ f(x') \geq -\frac{r}{2}\left(2\mu|x| + \sqrt{2\lambda\mu\rho}\right)^2. \end{cases}$$

*Proof.* From the possibility of taking  $x' = 0$  in formula (1.1), we observe the general inequality that

$$(4.4) \quad e_\lambda(x) \leq \frac{1}{2\lambda}|x|^2 \text{ for all } x.$$

Using this and the inequality in (4.1), we obtain for any  $x'$  satisfying the condition on the left of (4.3) that

$$-\frac{r}{2}|x'|^2 + \frac{1}{2\lambda}|x' - x|^2 \leq \frac{1}{2\lambda}|x|^2 + \rho,$$

which simplifies to  $(1 - \lambda r)|x'|^2 \leq 2\langle x', x \rangle + 2\lambda\rho$ . In multiplying this through by  $\mu$ , we get  $|x'|^2 \leq 2\mu|x'||x| + 2\lambda\mu\rho$ , which can be written as  $(|x'| - \mu|x|)^2 \leq \mu^2|x|^2 + 2\lambda\mu\rho$  and implies that

$$|x'| \leq \mu|x| + \sqrt{\mu^2|x|^2 + 2\lambda\mu\rho} \leq 2\mu|x| + \sqrt{2\lambda\mu\rho}.$$

This is the first inequality on the right side of (4.3), and it yields the third inequality through the lower bound in (4.1). The middle inequality in (4.3) is obvious from (4.4). □

**Proposition 4.2.** *Under assumptions (4.1), consider any  $\lambda \in (0, 1/r)$ . For any  $\epsilon > 0$  there is a neighborhood  $X$  of  $\bar{x} = 0$  such that*

- (a)  $e_\lambda$  is Lipschitz continuous on  $X$  with constant  $\epsilon$ ,
- (b)  $P_\lambda(x)$  is nonempty when  $x \in X$ ,
- (c)  $|x'| < \epsilon$ ,  $|f(x')| < \epsilon$  and  $\lambda^{-1}|x - x'| < \epsilon$  for all  $x' \in P_\lambda(x)$  when  $x \in X$ ,
- (d)  $\{(x, x') \mid x \in X, x' \in P_\lambda(x)\}$  is compact.

*Proof.* Let  $\mu = (1 - \lambda r)^{-1}$ . From Lemma 4.1 we know that  $|x'| \leq 2\mu|x|$  whenever  $x' \in P_\lambda(x)$ , because the implication can be invoked for every  $\rho > 0$  when  $x' \in P_\lambda(x)$ .

The local lower semicontinuity of  $f$  in (4.1) guarantees the existence of  $\epsilon' \in (0, \epsilon)$  such that the set  $C := \{x' \mid |x'| \leq \epsilon', f(x') \leq \epsilon'\}$  is compact, and  $f$  is l.s.c. relative to  $C$ . Choose  $\delta > 0$  and  $\rho > 0$  small enough that  $(\epsilon' + \delta)/\lambda \leq \epsilon$  and

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

and let  $X = \{x \mid |x| \leq \delta\}$ . By Lemma 4.1,

$$x \in X, \quad f(x') + \frac{1}{2\lambda}|x' - x|^2 \leq e_\lambda(x) + \rho \implies x' \in C.$$

Then as long as  $x \in X$ , the minimization over  $x \in \mathbb{R}^n$  in the definitions of  $e_\lambda(x)$  and  $P_\lambda(x)$  can be restricted without loss to  $x' \in C$ . In that case one is minimizing an l.s.c. function over a compact set, so the minimum is attained. Therefore,  $\emptyset \neq P_\lambda(x) \subset C$ , which gives (b) and the first two inequalities in (c). Because  $|x'| \leq 2\mu|x|$  when  $x' \in P_\lambda(x)$ , implying that  $|\lambda^{-1}(x - x')| \leq [(1 + 2\mu)/\lambda]|x|$ , the third inequality in (c) follows from these and the assumption that  $\delta(1 + 2\mu)/\lambda \leq \epsilon'$ .

Our expression of  $e_\lambda$  in terms of minimizing over  $C$  represents  $e_\lambda$  relative to  $X$  as the pointwise minimum of the collection of functions  $\phi_{x'} : x \mapsto f(x') + (1/2\lambda)|x' - x|^2$  indexed by  $x' \in C$ . These functions are continuously differentiable with gradient  $\nabla\phi_{x'} = (x' - x)/\lambda$ , so that

$$|\nabla\phi_{x'}(x)| \leq \frac{|x'| + |x|}{\lambda} \leq \frac{\epsilon' + \delta}{\lambda} \leq \epsilon$$

when  $x \in X$ , and consequently they are Lipschitz continuous on  $X$  with constant  $\epsilon$ . Then  $e_\lambda$  must have the same property; thus, (a) is true. The set in (d) is the same as

$$\left\{ (x, x') \in X \times C \mid f(x') + \frac{1}{2\lambda}|x' - x|^2 - e_\lambda(x) \leq 0 \right\},$$

where  $X$  and  $C$  are compact and the function in the inequality is l.s.c. on  $X \times C$ . Therefore, (d) is true too. □

**Proposition 4.3.** *Under assumptions (4.1), there exists for each  $\lambda \in (0, 1/r)$  a neighborhood of  $\bar{x} = 0$  on which*

- (a)  $\partial e_\lambda(x) \subset \{ \lambda^{-1}(x - x') \mid x' \in P_\lambda(x) \}$ , where
- (b)  $x' \in P_\lambda(x) \implies \lambda^{-1}(x - x') \in \partial f(x')$ , i.e.,  $x' \in (I + \lambda \partial f)^{-1}(x)$ .

*Proof.* First we verify that (b) holds for all  $x$ . When  $x' \in P_\lambda(x)$  we have

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

so that  $f(x'') - f(x') \geq q(x'')$  for the quadratic function

$$q(x'') = (|x' - x|^2 - |x'' - x|^2)/2\lambda.$$

We have  $q(x') = 0$  and  $\nabla q(x') = \lambda^{-1}(x - x')$ , so the latter must be a proximal subgradient of  $f$  at  $x'$ : we have  $\lambda^{-1}(x - x') \in \partial_p f(x')$ . In particular, we have (b).

Moving on to (a), we fix  $\lambda \in (0, 1/r)$  and choose a neighborhood  $X$  of 0 with the properties in Proposition 4.2. Consider any point  $x \in X$  and any proximal subgradient  $v \in \partial_p e_\lambda(x)$ : there exist  $s > 0$  and  $\delta > 0$  such that

$$(4.5) \quad e_\lambda(w) \geq e_\lambda(x) + \langle v, w - x \rangle - \frac{s}{2}|w - x|^2 \text{ when } |w - x| < \delta.$$

By our choice of  $x$  we have  $P_\lambda(x) \neq \emptyset$ . Consider any  $x' \in P_\lambda(x)$ . We have

$$\begin{aligned} e_\lambda(x) &= f(x') + \frac{1}{2\lambda}|x' - x|^2, \\ e_\lambda(w) &\leq f(w') + \frac{1}{2\lambda}|w' - w|^2 \text{ for all } w', w, \end{aligned}$$

and therefore by (4.5) that

$$f(w') + \frac{1}{2\lambda}|w' - w|^2 \geq f(x') + \frac{1}{2\lambda}|x' - x|^2 + \langle v, w - x \rangle - \frac{s}{2}|w - x|^2 \text{ when } |w - x| < \delta.$$

Specializing this to  $w' = x'$ , we obtain in terms of  $u = w - x$  that

$$\frac{1}{2\lambda}|u - (x' - x)|^2 \geq \frac{1}{2\lambda}|x' - x|^2 + \langle v, u \rangle - \frac{s}{2}|u|^2 \text{ when } |u| < \delta,$$

which we can expand on the left and then rewrite as

$$\partial \langle v - \lambda^{-1}(x - x'), u \rangle \leq (s + \lambda^{-1})|u|^2 \text{ when } |u| < \delta.$$

This requires  $v - \lambda^{-1}(x - x') = 0$ , i.e.,  $v = \lambda^{-1}(x - x')$ . Therefore, the inclusion in (a) is valid for  $\partial_p e_\lambda$ . But then it must hold for all limiting subgradients, i.e., for  $\partial e_\lambda$ , because of the compactness in Proposition 4.2(d).  $\square$

These observations reveal a major fact about the Moreau envelopes of prox-regular functions. For the purpose of stating it, we recall that a function is of class  $\mathcal{C}^{1+}$  if it is differentiable and its gradient mapping is Lipschitz continuous.

**Theorem 4.4.** *Suppose that  $f$  is prox-regular at  $\bar{x} = 0$  for  $\bar{v} = 0$  with respect to  $\epsilon$  and  $r$ , in particular with (4.1) holding. Let  $T$  be the  $f$ -attentive  $\epsilon$ -localization  $T$  of  $\partial f$  around  $(0, 0)$ . Then for each  $\lambda \in (0, 1/r)$  there is a neighborhood  $X$  of  $\bar{x} = 0$  such that, on  $X$ , the mapping  $P_\lambda$  is single-valued and Lipschitz continuous with constant  $1/(1 - \lambda r)$  and*

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

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

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

*Proof.* Choose  $X$  open and small enough that the properties in Propositions 4.2 and 4.3 hold on  $X$ . Then for  $x \in X$  we have  $P_\lambda(x)$  nonempty by 4.2(b), while  $\partial e_\lambda(x)$  is nonempty by 4.2(a) and satisfies the inclusion in 4.3(a). In this inclusion and the one in 4.3(b) we can replace  $\partial f$  by  $T$  because of 4.2(c). To obtain the formulas claimed here for  $P_\lambda(x)$  and  $\partial e_\lambda(x)$ , we need only demonstrate that  $(I + \lambda T)^{-1}$  cannot be multivalued on  $X$ , since a Lipschitz function with at most one subgradient at each point of an open set must be a  $C^1$  function.

Suppose that  $x_i \in (I + \lambda T)^{-1}(x) \cap X$ ,  $i = 0, 1$ . Then  $(x - x_i)/\lambda \in T(x_i)$ . Invoking the prox-regularity of  $f$ , we have the monotonicity of  $T + rI$  by Theorem 3.2 (Remark 3.3) and therefore

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

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

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

$$\left\langle \left[ \frac{x_1 - x'_1}{\lambda} \right] - \left[ \frac{x_0 - x'_0}{\lambda} \right], x'_1 - x'_0 \right\rangle \geq -r|x'_1 - x'_0|^2,$$

so  $(1 - \lambda r)|x'_1 - x'_0|^2 \leq \langle x_1 - x_0, x'_1 - x'_0 \rangle \leq |x_1 - x_0| |x'_1 - x'_0|$ . This can be written in the form  $|x'_1 - x'_0| \leq (1 - \lambda r)^{-1} |x_1 - x_0|$ .  $\square$

A variant of the gradient formula in Theorem 4.4 can be obtained as follows.

**Lemma 4.5.** *For any mapping  $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  and any  $\lambda > 0$ , one has the identity*

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

*Proof.* By direct manipulation we have

$$\begin{aligned} z \in \lambda^{-1} [I - (I + \lambda T)^{-1}](w) &\iff \lambda z \in w - (I + \lambda T)^{-1}(w) \\ &\iff w - \lambda z \in (I + \lambda T)^{-1}(w) \\ &\iff w \in (w - \lambda z) + \lambda T(w - \lambda z) \\ &\iff z \in T(w - \lambda z) \iff w - \lambda z \in T^{-1}(z) \\ &\iff w \in (\lambda I + T^{-1})(z) \iff z \in (\lambda I + T^{-1})^{-1}(w), \end{aligned}$$

and this is what was required.  $\square$

Incidentally, the identity in this lemma is also valid for  $\lambda < 0$  with the factor  $\lambda^{-1}$  on the left converted to a factor  $\lambda$  on the right. In that mode the identity is valid too for  $\lambda = 0$  by separate argument.

**Proposition 4.6.** *In Theorem 4.4, the gradient formula can be expressed equivalently as:*

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

*Proof.* Simply combine Theorem 4.4 with Lemma 4.5. □

For the next result, we recall from [26] that a set  $M \subset \mathbb{R}^N$  with  $N > n$  is a Lipschitz manifold of dimension  $n$  around one of its points if there is an open neighborhood  $U$  of that point and a one-to-one mapping between  $U$  and an open subset  $O$  of  $\mathbb{R}^n \times \mathbb{R}^{N-n}$ , continuously differentiable in both directions, under which  $U \cap M$  is identified with  $O \cap \text{gph } F$  for some Lipschitz continuous mapping  $F$  from an open subset of  $\mathbb{R}^n$  into  $\mathbb{R}^{N-n}$ .

**Theorem 4.7.** *If the function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is prox-regular at  $\bar{x}$  for a vector  $\bar{v} \in \partial f(\bar{x})$ , then for any  $\epsilon > 0$  sufficiently small, the graph of the  $f$ -attentive  $\epsilon$ -localization of  $\partial f$  at  $(\bar{x}, \bar{v})$  is a Lipschitz manifold of dimension  $n$  around  $(\bar{x}, \bar{v})$  in  $\mathbb{R}^n \times \mathbb{R}^n$ . When  $f$  is subdifferentially continuous, this can be said of the graph of  $\partial f$  itself.*

*Proof.* There is no loss in taking  $\epsilon$  small enough that the definition of proximal regularity is satisfied by  $\epsilon$  and some  $r$ . Reducing to the case of  $\bar{x} = 0$  and  $\bar{v} = 0$ , we observe from Theorem 4.4 that, for any  $\lambda \in (0, 1/r)$ , the pairs  $(x, v) \in \text{gph } T$  correspond locally to the pairs  $(w, w') \in \text{gph } P_\lambda$  under the linear transformation  $(x, v) \mapsto (x + \lambda v, x)$ , having inverse  $(w, w') \mapsto (w', (w - w')/\lambda)$ . Since  $P_\lambda$  is Lipschitz continuous around  $\bar{x} = 0$  as a mapping from  $\mathbb{R}^n$  into  $\mathbb{R}^n$ ,  $\text{gph } T$  then fits the definition of being a Lipschitz manifold of dimension  $n$  around  $(\bar{x}, \bar{v})$  in the  $2n$ -dimensional space  $\mathbb{R}^n \times \mathbb{R}^n$ . When  $f$  is subdifferentially continuous, there is no need for  $f$ -attentiveness, and the property can be asserted directly for  $\text{gph } \partial f$ . □

Building on this, we can add to the characterization of prox-regularity in Theorem 3.2. Let us speak of a mapping  $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  as *locally maximal monotone* relative to  $(\bar{x}, \bar{v}) \in \text{gph } S$  if there is a neighborhood  $U$  of  $(\bar{x}, \bar{v})$  in  $\mathbb{R}^n \times \mathbb{R}^n$  such that for every monotone mapping  $S' : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  with  $\text{gph } S' \supset \text{gph } S$ , one has  $U \cap \text{gph } S' = U \cap \text{gph } S$ .

**Proposition 4.8.** *If the function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is prox-regular at  $\bar{x}$  for  $\bar{v} \in \partial f(\bar{x})$  with parameter values  $\epsilon > 0$  and  $r > 0$ , the  $f$ -attentive  $\epsilon$ -localization  $T$  of  $\partial f$  at  $(\bar{x}, \bar{v})$  has the property that  $T + rI$  is not just monotone but locally maximal monotone relative to  $(\bar{x}, \bar{v} + r\bar{x})$ . When  $f$  is subdifferentially continuous, this can be said of  $\partial f + rI$ .*

*Proof.* We can suppose  $(\bar{x}, \bar{v}) = (0, 0)$ . The elements  $(x, v) \in \text{gph } T$  correspond one-to-one to those of  $\text{gph } S$  for  $S = T + rI$  under  $(x, v) \longleftrightarrow (x, x + rv)$ , this being affine in both directions. Hence by Theorem 4.7,  $\text{gph } S$  is a Lipschitz manifold of dimension  $n$  around  $(0, 0)$ . The same is then true for the graph of the mapping  $P = (I + S)^{-1}$ ; the correspondence between  $\text{gph } S$  and  $\text{gph } P$  is given by  $(x, y) \longleftrightarrow (x + y, x)$ . The monotonicity of  $S$  implies that  $P$  is nonexpansive (hence Lipschitz continuous) relative to its domain  $D$  in  $\mathbb{R}^n$ . Some neighborhood of  $(0, 0)$  in  $\text{gph } P$  thus corresponds one-to-one to a subset of  $D$  containing  $0$  under a mapping that is Lipschitz continuous in both directions. Since  $\text{gph } P$  is a Lipschitz manifold of dimension  $n$  around  $(0, 0)$ , it follows that a subset of  $D$  containing  $0$  corresponds

in such a way to an open subset of  $\mathbb{R}^n$ , and therefore that  $D$  is a neighborhood of 0. For any monotone mapping  $S'$  with  $\text{gph } S' \supset \text{gph } S$ , the mapping  $P' = (I + S')^{-1}$ , whose graph corresponds one-to-one with that of  $S'$ , is nonexpansive too, and  $\text{gph } P' \supset \text{gph } P$ . Therefore,  $P'$  can do no more than coincide with  $P$  on a neighborhood of 0. This means that the graph of  $S'$  must agree with that of  $S$  on a neighborhood of  $(0, 0)$ , and hence that  $S$  is locally maximal monotone with respect to  $(0, 0)$ .  $\square$

## 5. CONVEXITY OF MOREAU ENVELOPES

Continuing in the same setting as laid down at the beginning of Section 3, we investigate now the local properties of convexity of the functions  $e_\lambda$ .

**Lemma 5.1.** *Let  $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  be any set-valued mapping. Suppose that  $T = \sigma I + M$  where  $M$  is monotone and  $\sigma$  is any value in  $\mathbb{R}$  (positive, negative, zero). Let  $\lambda > 0$  be small enough that  $1 + \lambda\sigma > 0$ . Then the mapping  $S_\lambda$  given by either side of the identity in Lemma 4.5 can be expressed by*

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

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

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

*Proof.* Let  $\kappa = \lambda/(1 + \lambda\sigma)$ . We have

$$\begin{aligned} S_\lambda &= \lambda^{-1} [I - (I + \lambda(\sigma I + M))^{-1}] \\ &= \lambda^{-1} [I - [(1 + \lambda\sigma)I + \lambda M]^{-1}] \\ &= \lambda^{-1} [I - [(1 + \lambda\sigma)I \circ (I + \kappa M)]^{-1}] \\ &= \lambda^{-1} [I - (I + \kappa M)^{-1} \circ (1 + \lambda\sigma)^{-1} I] \\ &= \lambda^{-1} [(1 + \lambda\sigma)I - (I + \kappa M)^{-1}] \circ (1 + \lambda\sigma)^{-1} I \\ &= \lambda^{-1} [\lambda\sigma I + \kappa(\kappa^{-1}[I - (I + \kappa M)^{-1}])] \circ (1 + \lambda\sigma)^{-1} I. \end{aligned}$$

Applying Lemma 4.5 to write  $\kappa^{-1}[I - (I + \kappa M)^{-1}]$  as  $(\kappa I + M^{-1})^{-1}$ , we get

$$\begin{aligned} S_\lambda &= [\sigma I + \lambda^{-1}\kappa(\kappa I + M^{-1})^{-1}] \circ (1 + \lambda\sigma)^{-1} I \\ &= (1 + \lambda\sigma)^{-1} \sigma I + (1 + \lambda\sigma)^{-1} (\kappa I + M^{-1})^{-1} \circ (1 + \lambda\sigma)^{-1} I \\ &= (1 + \lambda\sigma)^{-1} \sigma I + M', \end{aligned}$$

as claimed. Because  $M$  is monotone and  $\kappa > 0$ , we have  $\kappa I + M^{-1}$  monotone and consequently  $M'$  monotone.  $\square$

Note that the estimate in this lemma is the sharpest possible under such general assumptions, since in the case where  $T = \sigma I$  ( $M = 0$ ) one has  $S_\lambda = \sigma(1 + \lambda\sigma)^{-1} I$ .

**Theorem 5.2.** *Suppose that  $f$  is prox-regular at  $\bar{x} = 0$  for  $\bar{v} = 0$  with respect to  $\epsilon$  and  $r$ , in particular with (4.1) holding, and let  $\lambda \in (0, 1/r)$ . Then on some neighborhood of 0 the function*

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

*is convex.*

*Proof.* Recall from Remark 3.3 that  $T + rI$  is monotone. Take  $\sigma = -r$  in the preceding lemma. Then for  $S_\lambda$ , the mapping given by the identity in Lemma 4.5, we have  $S_\lambda + r(1 - \lambda r)^{-1}I$  monotone. But this mapping is the gradient of the (Lipschitz) function in question. Hence, this function is convex.  $\square$

**Corollary 5.3.** *If  $f$  is prox-regular at  $\bar{x} = 0$  for  $\bar{v} = 0$ , and  $\lambda$  is sufficiently small, then on some neighborhood of the origin  $e_\lambda$  is a lower- $\mathcal{C}^2$  function, hence in particular prox-regular itself.*

*Proof.* This holds because a finite function, which we know  $e_\lambda$  to be, is lower- $\mathcal{C}^2$  if and only if it can be expressed as the difference of a convex function and a  $\mathcal{C}^2$  function, cf. [25]. The prox-regularity of lower- $\mathcal{C}^2$  functions has been observed in Example 2.7.  $\square$

The convexity of  $e_\lambda$  itself has a full characterization. To state it, we recall that set-valued mapping  $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  is *proto-differentiable* at a point  $x$  for an element  $v \in T(x)$  if the set-valued mappings

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

regarded as a family indexed by  $t > 0$ , graph-converge as  $t \downarrow 0$  (i.e., one has set convergence of the graphs). If so, the limit mapping is denoted by  $T'_{x,v}$  and called the *proto-derivative* of  $T$  at  $x$  for  $v$ ; see [18], [28], [30]. This proto-derivative mapping assigns to each  $\xi \in \mathbb{R}^n$  a subset  $T'_{x,v}(\xi)$  of  $\mathbb{R}^n$ , which could be empty for some choices of  $\xi$ . Proto-differentiability is *semi-differentiability* when instead of just graph convergence of the mappings  $\Delta_{x,v,t}T$  one has

$$[T(x + t\xi') - v]/t \rightarrow T'_{x,v}(\xi) \text{ as } \xi' \rightarrow \xi, t \downarrow 0.$$

**Proposition 5.4.** *Suppose that  $f$  is prox-regular at  $\bar{x} = 0$  for  $\bar{v} = 0$  with respect to  $\epsilon$  and  $r$ , in particular with (4.1) holding, and let  $\lambda \in (0, 1/r)$ . Let  $T$  be the  $f$ -attentive  $\epsilon$ -localization  $T$  of  $\partial f$  around  $(0, 0)$ . Then the following conditions are equivalent:*

- (a) *The function  $e_\lambda$  is convex on a neighborhood of 0.*
- (b) *There is a neighborhood  $U$  of  $(0, 0)$  such that if  $T_0$  is the localization of  $T$  obtained by intersecting the graph of  $T$  with  $U$ , then  $T_0^{-1} + \lambda I$  is monotone.*
- (c) *There is a neighborhood  $U$  of  $(0, 0)$  such that at all points  $(x, v) \in U \cap \text{gph } T$  where  $T$  is proto-differentiable, the proto-derivative mapping  $T'_{x,v} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  is such that  $(T'_{x,v})^{-1} + \lambda I$  is monotone.*
- (d) *Same as (c) but with restriction to the points  $(x, v)$  where in addition the graph of  $T'_{x,v}$  is an  $n$ -dimensional subspace of  $\mathbb{R}^n \times \mathbb{R}^n$ .*

*Proof.* We have  $e_\lambda$  convex locally if and only if  $\nabla e_\lambda = S_\lambda$  is monotone locally around the point  $(0, 0)$  in its graph, or equivalently,  $S_\lambda^{-1}$  has such local monotonicity.

Recall that  $S_\lambda = (\lambda I + T^{-1})^{-1}$ , which means that  $S_\lambda^{-1} = \lambda I + T^{-1}$ . This gives the equivalence between (a) and (b). The local monotonicity of  $S_\lambda$  implies that of its proto-derivative mappings where they exist. Proto-derivative mappings for  $S_\lambda$  have the form  $\lambda I + (T'_{x,v})^{-1}$  in terms of proto-derivative mappings for  $T$ , and their monotonicity thus corresponds to the mappings  $(T'_{x,v})^{-1} + \lambda I$  being monotone. Because the mapping  $S_\lambda = \nabla e_\lambda$  is Lipschitz continuous around 0, it is differentiable a.e., hence in particular proto-differentiable a.e., (proto-differentiability being the same as semi-differentiability in this case; see [28]). Through integration (using Fubini's theorem) one sees that  $\nabla e_\lambda$  is monotone around 0 if and only if, on some neighborhood of 0, the Jacobians  $\nabla(\nabla e_\lambda)$  are monotone almost wherever they exist:

$$\langle \nabla e_\lambda(x_1) - \nabla e_\lambda(x_0), x_1 - x_0 \rangle = \int_0^1 \nabla(\nabla e_\lambda)(x_0 + t(x_1 - x_0))(x_1 - x_0) dt.$$

From this we get the equivalence between (b) and (d). Condition (c) is sandwiched between these, so it is equivalent to them too. (The graph of  $T'_{x,v}$  is a linear subspace of dimension  $n$  precisely at the points of the Lipschitz manifold  $\text{gph } T$  where a linear tangent space exists; see [26].)  $\square$

For strong monotonicity of  $e_\lambda$ , we have the following sufficient condition.

**Proposition 5.5.** *Suppose that  $f$  is prox-regular at  $\bar{x} = 0$  for  $\bar{v} = 0$  with respect to  $\epsilon$  and  $r$ , and let  $\lambda \in (0, 1/r)$ . Let  $T$  be the  $f$ -attentive  $\epsilon$ -localization  $T$  of  $\partial f$  around  $(0, 0)$ . Suppose  $T$  is strongly monotone with modulus  $\mu > 0$ , i.e.,  $T - \mu I$  is monotone. Then, on some neighborhood of 0, one has the strong convexity of  $e_\lambda$  with modulus  $\mu/(1 + \lambda\mu)$ , i.e., the convexity of*

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

*Proof.* This comes out of Lemma 5.1 for  $\sigma = \mu$ , because the gradient mapping of the function in question is  $S_\lambda - [\mu/(1 + \lambda\mu)]I$  with  $S_\lambda$  the mapping given by the identity in Lemma 4.5.  $\square$

We now work on conditions in terms of proto-derivatives of  $T$  for  $T - \mu I$  to be monotone for a given  $\mu$ . The following lemma will enter the development.

**Lemma 5.6.** *Suppose  $P$  is a Lipschitz continuous mapping from an open convex set  $O \subset \mathbb{R}^n$  into  $\mathbb{R}^n$ . Then the following conditions are equivalent for any  $\alpha > 0$ .*

- (a)  $P^{-1} - \alpha I$  is monotone.
- (b) For all  $y \in O$  where  $P$  is proto-differentiable (hence semi-differentiable), the proto-derivative mapping  $P'_y$  mapping is such that  $(P'_y)^{-1} - \alpha I$  is monotone.
- (c) For all  $y \in O$  where  $P$  is differentiable, the Jacobian matrix  $\nabla P(y)$  satisfies

$$\langle \eta, \nabla P(y)\eta \rangle \geq \alpha |\nabla P(y)\eta|^2 \text{ for all } \eta \in \mathbb{R}^n.$$

*Proof.* Condition (a) implies condition (b) through the fact that the proto-derivative of a monotone mapping, if it exists, is another monotone mapping. Condition (b) implies condition (c) as the specialization to the case where the proto-derivative  $P'_y$  happens to exist as a linear mapping, which we know by Rademacher's theorem

is true a.e.  $y \in O$ . We must show now that (c) implies (a). Condition (a) means that

$$(5.1) \quad \langle y_1 - y_0, P(y_1) - P(y_0) \rangle \geq \alpha |P(y_1) - P(y_0)|^2 \text{ for all } y_0, y_1 \in O.$$

Because  $P$  is continuous, we need only verify this for a dense set of pairs  $(y_0, y_1) \in O \times O$ , hence (by Fubini's Theorem) only for a choice of  $y_0$  and  $y_1$  such that  $P$  is differentiable a.e. on the line segment  $[y_0, y_1]$ . Let  $\eta = y_1 - y_0$ . In terms of  $\phi(t) := \langle \eta, P(y_0 + t\eta) \rangle$ , the instance of (5.1) we want to check translates to

$$(5.2) \quad \phi(1) - \phi(0) \geq \alpha |P(y_0 + \eta) - P(y_0)|^2.$$

We have  $\phi(t)$  Lipschitz continuous with

$$\phi'(t) = \langle \eta, \nabla P(y_0 + t\eta)\eta \rangle \geq \alpha |P(y_0 + t\eta)\eta|^2 \text{ for a.e. } t \in [0, 1],$$

where the inequality is based on the assumption in (c). Therefore

$$\phi(1) - \phi(0) = \int_0^1 \phi'(t)dt \geq \alpha \int_0^1 |P(y_0 + t\eta)\eta|^2 dt.$$

Using the fact that for any vector function  $p(t)$  we have  $\int_0^1 |p(t)|^2 dt \geq |\int_0^1 p(t)dt|^2$  (Jensen's inequality relative to the probability space  $[0, 1]$  under the uniform distribution), we get

$$\int_0^1 |P(y_0 + t\eta)\eta|^2 dt \geq \left| \int_0^1 P(y_0 + t\eta)\eta dt \right|^2 = |P(y_0 + \eta) - P(y_0)|^2.$$

This yields (5.2), as desired. □

**Proposition 5.7.** *Suppose that  $f$  is prox-regular at  $\bar{x}$  for  $\bar{v}$  with respect to  $\epsilon$  and  $r$ . Let  $T$  be the  $f$ -attentive  $\epsilon$ -localization  $T$  of  $\partial f$  around  $(\bar{x}, \bar{v})$ . Then the following conditions on  $T$  and a value  $\mu > 0$  are equivalent:*

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

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

(c) Same as (b) but with restriction to the points  $(x, v)$  where in addition the graph of  $T'_{x,v}$  is an  $n$ -dimensional subspace of  $\mathbb{R}^n \times \mathbb{R}^n$ .

*Proof.* We may assume that  $\bar{x} = 0 = \bar{v}$ , in particular with (4.1) holding. Consider any  $\rho > r$  such that  $\rho + \mu > 0$ , where  $r$  is a local constant from the definition of prox-regularity. Let  $\alpha = \rho + \mu$ . We have  $T - \mu I$  monotone if and only if  $(T + \rho I) - \alpha I$  is monotone. Let  $P = (T + \rho I)^{-1}$ , so that the question of whether  $(T + \rho I) - \alpha I$  is monotone locally is that of whether  $P^{-1} - \alpha I$  is monotone locally. Since  $T + rI$  is a monotone mapping, in fact it is locally maximal monotone in a graph around  $(0, 0)$ , and  $P = (M + (\rho - r)I)^{-1}$  with  $\rho - r > 0$ , we know that  $P$  is Lipschitz continuous on some neighborhood of 0. Applying Lemma 5.6, we reduce the monotonicity

question for  $P^{-1} - \alpha I$  to that of  $(P'_v)^{-1} - \alpha I$  at points  $v$  near 0 where the proto-derivative exists, or the actual derivative exists (equivalent to the proto-derivative being linear, since proto-differentiability implies semi-differentiability when  $P$  is Lipschitz continuous). But proto-derivatives of  $P$  correspond through graph geometry to proto-derivatives of  $T$ . The conditions translate thereby into ones of monotonicity of  $[T'_{x,v} + \rho I] - \alpha I = T'_{x,v} - \mu I$  at points  $(x, v) \in \text{gph} T$  near  $(0, 0)$ . They come out then as (b) and (c).  $\square$

**Corollary 5.8.** *Suppose that  $f$  is prox-regular at  $\bar{x}$  for  $\bar{v}$  with respect to  $\epsilon$  and  $r$ . Let  $T$  be the  $f$ -attentive  $\epsilon$ -localization  $T$  of  $\partial f$  around  $(\bar{x}, \bar{v})$ . Then the following conditions on  $T$  are equivalent:*

- (a)  $T$  is monotone locally around the point  $(\bar{x}, \bar{v}) \in \text{gph} T$ .
- (b) There is a neighborhood  $U$  of  $(\bar{x}, \bar{v})$  such that at all points  $(x, v) \in U \cap \text{gph} T$  where  $T$  is proto-differentiable, the proto-derivative mapping  $T'_{x,v} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  is monotone.
- (c) Same as (b) but with restriction to the points  $(x, v)$  where in addition the graph of  $T'_{x,v}$  is an  $n$ -dimensional subspace of  $\mathbb{R}^n \times \mathbb{R}^n$ .

*Proof.* Apply Proposition 5.7 to  $T_\mu = T + \mu I$  for all  $\mu > 0$ .  $\square$

In Section 6, we establish the connection between the second-order epi-derivatives of  $f$  and the proto-derivatives of an  $f$ -attentive localization of  $\partial f$ . Once we have this connection we will be able to add more conditions in Proposition 5.7 and Corollary 5.8; see Corollaries 6.3 and 6.4.

*Remark 5.9.* If the function  $f$  is also subdifferentially continuous, then all results in this section concerning  $T$  as an  $f$ -attentive localization of  $\partial f$  at  $(\bar{x}, \bar{v})$  can be restated in terms of  $T$  being an ordinary localization.

### 6. SECOND-ORDER THEORY

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

$$(6.1) \quad \Delta^2_{\bar{x}, \bar{v}, t} f(\xi) = [f(\bar{x} + t\xi) - f(\bar{x}) - t\langle \bar{v}, \xi \rangle] / \frac{1}{2}t^2 \text{ for } t > 0,$$

epi-converge to a proper function as  $t \downarrow 0$ . The epi-limit is then the *second epi-derivative* function  $f''_{\bar{x}, \bar{v}} : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ . see [17], [27] and [29]. This function, when it exists, is l.s.c., proper and positively homogeneous of degree 2. We say that  $f$  is *strictly twice epi-differentiable* at  $\bar{x}$  for  $\bar{v} \in \partial f(\bar{x})$  if more generally the functions

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

epi-converge to a proper function as  $t \downarrow 0$ ,  $x \rightarrow \bar{x}$  with  $f(x) \rightarrow f(\bar{x})$  and  $v \rightarrow \bar{v}$  with  $v \in \partial f(x)$ . The limit function is then the second epi-derivative function  $f''_{\bar{x}, \bar{v}}$ ; in particular the function  $f$  is twice epi-differentiable. Similarly the set-valued mapping  $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  is *strictly proto-differentiable* at a point  $\bar{x}$  for an element  $\bar{v} \in T(\bar{x})$  if the set-valued mappings

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

graph-converge as  $t \downarrow 0$ ,  $x \rightarrow \bar{x}$  and  $v \rightarrow \bar{v}$  with  $v \in T(x)$ . The limit is then the proto-derivative mapping  $T'_{\bar{x}, \bar{v}}$ ; in particular the mapping  $T$  is proto-differentiable.

**Theorem 6.1.** *Assume that  $f$  is prox-regular at  $\bar{x}$  for  $\bar{v}$  with constants  $\epsilon$  and  $r$ . Let  $T$  be the  $f$ -attentive  $\epsilon$ -localization of  $\partial f$  around  $(\bar{x}, \bar{v})$ .*

(a) *The function  $f$  is twice epi-differentiable at  $\bar{x}$  for  $\bar{v}$  if and only if  $T$  is proto-differentiable at  $\bar{x}$  for  $\bar{v}$ .*

(b) *The function  $f$  is strictly twice epi-differentiable at  $\bar{x}$  for  $\bar{v}$  if and only if  $T$  is strictly proto-differentiable at  $\bar{x}$  for  $\bar{v}$ .*

Moreover, in either (a) or (b), one has

$$T'_{\bar{x}, \bar{v}}(\xi) = \partial\left[\frac{1}{2}f''_{\bar{x}, \bar{v}}\right](\xi) \text{ for all } \xi.$$

*Proof.* We concentrate on (b), since the argument for that will cover the one for (a) as a simple specialization. Without loss of generality we can suppose that  $\bar{x} = 0$ ,  $\bar{v} = 0$ ,  $f(0) = 0$ , and, as explained at the beginning of Section 4, work under the assumption that (4.1) holds. In addition we may assume, without any loss of generality, that the domain of  $f$  is included in the closed ball of radius  $\epsilon$ . Consider any  $\lambda \in (0, 1/r)$  and the function

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

There is a neighborhood of 0 on which this function is  $\mathcal{C}^{1+}$  by Theorem 4.4 and convex by Theorem 5.2, the gradient mapping being

$$\nabla \hat{e}_\lambda = \nabla e_\lambda + \frac{r}{1 - \lambda r}I.$$

We begin with the proof of (a). The convexity ensures through an easy modification of [30 (Thm. 2.2)] that  $\hat{e}_\lambda$  is strictly twice epi-differentiable at  $\bar{x} = 0$  for  $\bar{v} = 0$  if and only if  $\nabla \hat{e}_\lambda$  is strictly proto-differentiable at 0 for 0 (the latter actually being the same as the semi-differentiability of  $\nabla \hat{e}_\lambda$  at 0 because  $\nabla \hat{e}_\lambda$  is Lipschitz continuous around 0), with

$$(6.2) \quad \partial\left[\frac{1}{2}(\hat{e}_\lambda)''_{0,0}\right](\xi) = (\nabla \hat{e}_\lambda)'_{0,0}(\xi) \text{ for all } \xi.$$

Here we have

$$(6.3) \quad (\hat{e}_\lambda)''_{0,0}(\xi) = (e_\lambda)''_{0,0}(\xi) + \frac{r}{(1 - \lambda r)}|\xi|^2, \quad (\nabla \hat{e}_\lambda)'_{0,0} = (\nabla e_\lambda)'_{0,0}(\xi) + \frac{r}{1 - \lambda r}I,$$

hence also

$$(6.4) \quad \partial\left[\frac{1}{2}(e_\lambda)''_{0,0}\right](\xi) = (\nabla e_\lambda)'_{0,0}(\xi) \text{ for all } \xi;$$

But  $\nabla e_\lambda$  has been identified locally with  $[\lambda I + T^{-1}]^{-1}$  in Proposition 4.6. The graph of the latter mapping is the image of the graph of  $T$  under the invertible linear transformation  $(x, v) \mapsto (x + \lambda v, v)$  from  $\mathbb{R}^n \times \mathbb{R}^n$  onto itself. Since strict proto-differentiability at 0 for 0 is a geometric property of graphs at  $(0, 0)$  that is maintained when graphs are subjected to an invertible linear transformation, and the proto-derivative mappings themselves then correspond under the same

transformation, we deduce that  $\nabla e_\lambda$  is strictly proto-differentiable at 0 for 0 if and only if  $T$  has this property, in which event there is the formula

$$(6.5) \quad (\nabla e_\lambda)'_{0,0} = [\lambda I + D^{-1}]^{-1} \text{ with } D = T'_{0,0}.$$

It follows now that  $e_\lambda$  is strictly twice epi-differentiable at 0 for 0 if and only if  $T$  is strictly proto-differentiable at 0 for 0, in which event one has

$$(6.6) \quad \partial[\frac{1}{2}(e_\lambda)''_{0,0}] = [\lambda I + D^{-1}]^{-1} \text{ with } D = T'_{0,0}.$$

The same then holds for  $\hat{e}_\lambda$  in place of  $e_\lambda$ , the formula for that case being

$$(6.7) \quad \partial[\frac{1}{2}(\hat{e}_\lambda)''_{0,0}] = [\lambda I + D^{-1}]^{-1} + \frac{r}{1 - \lambda r} I \text{ with } D = T'_{0,0}.$$

A technical advantage of (6.7) over (6.6), which will be put to use presently, is that the functions  $(\hat{e}_\lambda)''_{0,0}$  are convex by virtue of the fact that each of the functions  $\hat{e}_\lambda$  is convex in some neighborhood of 0.

Next we observe that for each  $\lambda$  small enough, there exists  $\epsilon_\lambda > 0$  such that if  $|x| < \epsilon_\lambda$  with  $|f(x)| < \epsilon_\lambda$  and  $|v| < \epsilon_\lambda$  with  $v \in \partial f(x)$ , the functions

$$f_{x,v,t}(\xi) := \frac{f(x + t\xi) - f(x) - t\langle v, \xi \rangle}{t^2},$$

$$e_{\lambda,x+\lambda v,t}(\xi) := \frac{e_\lambda(x + \lambda v + t\xi) - e_\lambda(x + \lambda v) - t\langle \nabla e_\lambda(x + \lambda v), \xi \rangle}{t^2},$$

are related like  $f$  and  $e_\lambda$  themselves, i.e., the Moreau  $\lambda$ -envelope of  $f_{x,v,t}$  is  $e_{\lambda,x+\lambda v,t}$ :

$$(6.8) \quad \begin{aligned} & \min_{\xi'} \left\{ f_{x,v,t}(\xi') + \frac{1}{2\lambda} |\xi' - \xi|^2 \right\} \\ &= \min_{\xi'} \left\{ \frac{f(x + t\xi') - f(x) - t\langle v, \xi' \rangle}{t^2} + \frac{1}{2\lambda t^2} |t\xi' - t\xi|^2 \right\} \\ &= \frac{1}{t^2} \left[ \min_{\xi'} \left\{ f(x + t\xi') - t\langle v, \xi' \rangle + \frac{1}{2\lambda} |t\xi' - t\xi|^2 \right\} - f(x) \right] \\ &= \frac{1}{t^2} \left[ \min_{\xi'} \left\{ f(x + t\xi') + \frac{1}{2\lambda} |t\xi' - (\lambda v + t\xi)|^2 \right\} \right. \\ & \quad \left. - f(x) - \frac{\lambda}{2} |v|^2 - t\langle \xi, v \rangle \right] \\ &= \frac{1}{t^2} \left[ \min_{\xi'} \left\{ f(x + t\xi') + \frac{1}{2\lambda} |(x + t\xi') - (x + \lambda v + t\xi)|^2 \right\} \right. \\ & \quad \left. - f(x) - \frac{\lambda}{2} |v|^2 - t\langle \xi, v \rangle \right] \\ &= \frac{1}{t^2} \left[ e_\lambda(x + \lambda v + t\xi) - f(x) - \frac{\lambda}{2} |v|^2 - t\langle \xi, v \rangle \right]. \end{aligned}$$

By Theorem 4.4 and Proposition 4.6 there exists  $\epsilon_\lambda > 0$  such that  $e_\lambda(x + \lambda v) = f(x) + (\lambda/2)|v|^2$  and  $\nabla e_\lambda(x + \lambda v) = v$  whenever  $v \in \partial f(x)$  with  $|x| < \epsilon_\lambda$ ,  $|f(x)| < \epsilon_\lambda$  and  $|v| < \epsilon_\lambda$ . Hence, for such  $x$  and  $v$ , the Moreau  $\lambda$ -envelope of  $f_{x,v,t}$  at  $\xi$  is equal

to  $(1/t^2)[e_\lambda(x + \lambda v + t\xi) - e_\lambda(x + \lambda v) - t\langle \nabla e_\lambda(x + \lambda v), \xi \rangle]$ , which by definition is  $e_{\lambda, x+\lambda v, t}(\xi)$ . The strict twice epi-differentiability of  $f$  at 0 for 0 corresponds to  $2f_{x,v,t}$  epi-converging as  $t \downarrow 0$  and  $x \rightarrow 0$  with  $f(x) \rightarrow f(0)$ , while  $v \rightarrow 0$  with  $v \in \partial f(x)$ . On the other hand the strict twice epi-differentiability of  $e_\lambda$  corresponds to  $2e_{\lambda, x', t}$  epi-converging as  $t \downarrow 0$  and  $x' \rightarrow 0$ . The latter is the same as having  $2e_{\lambda, x+\lambda v, t}(\xi)$  epi-converge as  $t \downarrow 0$  and  $x \rightarrow 0$  with  $f(x) \rightarrow 0$ , while  $v \rightarrow 0$  with  $v \in T(x)$ . This is due to the one to one correspondence between the points  $x'$  close to 0 and the points  $x + \lambda v$  for  $v \in T(x)$ . Because  $f$  is prox-regular at 0 for 0, and  $\text{dom } f \subset \epsilon \mathcal{B}$ , ( $\mathcal{B}$  the closed unit ball), we have  $f(x + t\xi) - f(x) - t\langle v, \xi \rangle > -(r/2)|t\xi|^2$  whenever  $|x| < \epsilon$  with  $|f(x)| < \epsilon$  and  $|v| < \epsilon$  with  $v \in \partial f(x)$ . In other words, the functions  $f_{x,v,t}$  are uniformly bounded below by the quadratic  $-(r/2)|\cdot|^2$  whenever  $|x| < \epsilon$  with  $|f(x)| < \epsilon$  and  $|v| < \epsilon$  with  $v \in \partial f(x)$ . From [15 (Prop. 2.1)] we know then that in order to have  $f_{x,v,t}$  epi-converge to some function  $g$ , as  $x \rightarrow 0$  with  $f(x) \rightarrow 0$  while  $v \rightarrow 0$  with  $v \in \partial f(x)$ , it is necessary and sufficient that, for all  $\lambda$  sufficiently small, the  $\lambda$ -envelope of  $f_{x,v,t}$  converges pointwise on  $\mathbb{R}^n$  as  $x \rightarrow 0$  with  $f(x) \rightarrow 0$  while  $v \rightarrow 0$  with  $v \in \partial f(x)$ , the limit being the  $\lambda$ -envelope of  $g$ . We conclude that  $f$  is strictly twice epi-differentiable at 0 for 0 if and only if, for all  $\lambda$  sufficiently small,  $e_\lambda$  is strictly twice epi-differentiable at 0 for 0, in which case  $\frac{1}{2}(e_\lambda)''_{0,0}$  is the Moreau envelope  $\tilde{e}_\lambda$  for the function  $\tilde{f} := \frac{1}{2}f''_{0,0}$ .

The equivalence just established, along with the earlier one about the strict twice epi-differentiability of  $e_\lambda$ , proves the equivalence claimed in part (b) of the theorem. As already mentioned, the proof of part (a) is an elementary specialization of this proof.

From now on, our efforts are devoted to verifying the derivative formula in the theorem, and for this purpose we assume that  $f$  is indeed twice epi-differentiable at 0 for 0, i.e., that the function  $\tilde{f} = \frac{1}{2}f''_{0,0}$  exists. Then the Moreau envelope  $\tilde{e}_\lambda$  of  $\tilde{f}$  increases to  $\tilde{f}$  as  $\lambda \downarrow 0$ , hence in particular epi-converges to  $\tilde{f}$ . Since  $\tilde{e}_\lambda = \frac{1}{2}(e_\lambda)''_{0,0}$ , we see from the first equation in (6.3) that  $\frac{1}{2}(\tilde{e}_\lambda)''_{0,0}$  epi-converges to  $\tilde{f} + \frac{1}{2}r|\cdot|^2$  as  $\lambda \downarrow 0$ . But  $(\tilde{e}_\lambda)''_{0,0}$  is convex for  $\lambda \in (0, 1/r)$ . It follows by Attouch's theorem that the mappings  $\partial[\frac{1}{2}(\tilde{e}_\lambda)''_{0,0}]$  converge graphically to  $\partial\tilde{f}$  as  $\lambda \downarrow 0$ . But by (6.7) these mappings converge graphically to  $D$ . We conclude that  $D = \partial\tilde{f} = \partial[\frac{1}{2}f''_{0,0}]$ , as required.  $\square$

**Corollary 6.2.** *Suppose  $f$  is prox-regular and subdifferentially continuous at  $\bar{x}$  for  $\bar{v}$ .*

(a) *The function  $f$  is twice epi-differentiable at  $\bar{x}$  for  $\bar{v}$  if and only if  $\partial f$  is proto-differentiable at  $\bar{x}$  for  $\bar{v}$ .*

(b) *The function  $f$  is strictly twice epi-differentiable at  $\bar{x}$  for  $\bar{v}$  if and only if  $\partial f$  is strictly proto-differentiable at  $\bar{x}$  for  $\bar{v}$ .*

Moreover, in either (a) or (b), one has  $(\partial f)'_{\bar{x}, \bar{v}}(\xi) = \partial(\frac{1}{2}f''_{\bar{x}, \bar{v}})(\xi)$  for all  $\xi$ .

*Proof.* In this case the  $f$ -attentiveness in the localization of  $\partial f$  to  $T$  is superfluous.  $\square$

For the next statement we recall that a function  $q : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is a *generalized quadratic function* when it can be expressed as

$$q(\xi) = \begin{cases} \frac{1}{2}\langle x, Qx \rangle & \text{if } x \in N, \\ \infty & \text{if } x \notin N, \end{cases}$$

where  $Q$  is a symmetric  $n \times n$  matrix and  $N$  is a linear subspace of  $\mathbb{R}^n$ . A proper, l.s.c., function  $q$  has this property if and only if the graph of  $\partial q : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  is a linear subspace of  $\mathbb{R}^n \times \mathbb{R}^n$  (the subspace necessarily being of dimension  $n$ ), cf. [26].

**Corollary 6.3.** *Suppose that  $f$  is prox-regular at  $\bar{x}$  for  $\bar{v}$  with constants  $\epsilon$  and  $r$ , and let  $T$  be the  $f$ -attentive  $\epsilon$ -localization of  $\partial f$  around  $(\bar{x}, \bar{v})$ . Then the equivalent conditions in Proposition 5.7 are equivalent also to the following conditions:*

(a) *There is a neighborhood  $U$  of  $(\bar{x}, \bar{v})$  such that at all points  $(x, v) \in U \cap \text{gph } T$  where  $f$  is twice epi-differentiable, the second-order epi-derivative  $f''_{x,v}$  is strongly convex with modulus  $\mu$ .*

(b) *Same as (a) but with restriction to the points  $(x, v)$  where in addition the second-order epi-derivative is a generalized quadratic function.*

*Proof.* Conditions (b) and (c) of Proposition 5.7 can then be stated as (a) and (b) here using Theorem 6.1.  $\square$

**Corollary 6.4.** *Suppose that  $f$  is prox-regular at  $\bar{x}$  for  $\bar{v}$  with constants  $\epsilon$  and  $r$ , and let  $T$  be the  $f$ -attentive  $\epsilon$ -localization of  $\partial f$  around  $(\bar{x}, \bar{v})$ . Then the equivalent conditions in Corollary 5.8 are equivalent also to the following conditions:*

(a) *There is a neighborhood  $U$  of  $(\bar{x}, \bar{v})$  such that at all points  $(x, v) \in U \cap \text{gph } T$  where  $f$  is twice epi-differentiable, the second-order epi-derivative  $f''_{x,v}$  is convex.*

(b) *Same as (a) but with restriction to the points  $(x, v)$  where in addition the second-order epi-derivative is a generalized quadratic function.*

*Proof.* Conditions (b) and (c) of Corollary 5.8 can then be stated as (a) and (b) here using Theorem 6.1.  $\square$

The proof of Theorem 6.1 has revealed additional facts concerning  $f$  and the second-order properties of its Moreau envelopes  $e_\lambda$ , which we record next.

**Theorem 6.5.** *Suppose that  $f$  is prox-regular at  $\bar{x} = 0$  for  $\bar{v} = 0$  with respect to  $\epsilon$  and  $r$ , in particular with (4.1) holding, and let  $\lambda \in (0, 1/r)$ .*

(a) *The function  $f$  is twice epi-differentiable at 0 for 0 if and only if  $e_\lambda$  has this property.*

(b) *The function  $f$  is strictly twice epi-differentiable at 0 for 0 if and only if  $e_\lambda$  has this property.*

*In either (a) or (b),  $(e_\lambda)''_{0,0}$  is the Moreau  $\lambda$ -envelope of  $f''_{0,0}$ . Then the function  $f''_{0,0} + r|\cdot|^2$  is nonnegative and convex with*

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

*Proof.* See the proof of Theorem 6.1.  $\square$

**Corollary 6.6.** *Suppose that  $f$  is prox-regular at  $\bar{x} \in \text{argmin } f$  for  $\bar{v} = 0$  with respect to  $\epsilon$  and  $r$ . Let  $T$  be the  $f$ -attentive  $\epsilon$ -localization of  $\partial f$  around  $(\bar{x}, \bar{v})$ . There is a neighborhood  $U$  of  $(\bar{x}, \bar{v})$  and  $\bar{\lambda} > 0$  such that at all points  $(x, v) \in U \cap \text{gph } T$ , and for all  $0 < \lambda < \bar{\lambda}$  we have:*

(a) *The function  $f$  is twice epi-differentiable at  $x$  for  $v$  if and only if  $e_\lambda$  has this property at  $x + \lambda v$  for  $v$ .*

(b) *The function  $f$  is strictly twice epi-differentiable at  $x$  for  $v$  if and only if  $e_\lambda$  has this property at  $x + \lambda v$  for  $v$ ,*

In either (a) or (b),  $(e_\lambda)''_{x+\lambda v, v}$  is the Moreau  $\lambda$ -envelope of  $f''_{x, v}$ . Then the function  $f''_{x, v}$  plus  $(1/\lambda)|\cdot|^2$  is nonnegative and convex with

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

*Proof.* Assume that  $\bar{x} = 0$  with  $f(0) = 0$ . Consider  $(\tilde{x}, \tilde{v}) \in \text{gph}T$  and the function  $\tilde{f}(x) := f(x + \tilde{x}) - f(\tilde{x}) - \langle \tilde{v}, x \rangle$ . There is a neighborhood  $U$  of  $(0, 0)$  and  $R > r$  such that for all points  $(\tilde{x}, \tilde{v}) \in U \cap \text{gph}T$ , we have  $\tilde{f}(x) \geq -(R/2)|x|^2$  for all  $x$ . To see this first assume that  $|\tilde{x}| < (\epsilon/4)$ ,  $|f(\tilde{x})| < \epsilon$  and that  $|\tilde{v}| < \epsilon$ , then

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

whenever  $|x + \tilde{x}| < \epsilon$ . On the other hand when  $|x + \tilde{x}| \geq \epsilon$ , in particular this means that  $|x| \geq (3/4)\epsilon$ , notice that

$$\tilde{f}(x) \geq f(\bar{x}) - f(\tilde{x}) - \langle \tilde{v}, x \rangle \geq -\epsilon - \langle \tilde{v}, x \rangle \geq -\epsilon - \epsilon|x|.$$

This last quantity is greater than  $-(R/2)|x|^2$  provided  $|x| \geq (1/R)[\epsilon + \sqrt{\epsilon^2 + 2\epsilon R}]$ . Finally this last quantity can be made smaller than  $(3/4)\epsilon$  by choosing  $R$  large enough. So  $\tilde{f}(x) \geq -(R/2)|x|^2$  for all  $x$ .

As in (6.8), it is very easy to verify that for  $0 < \lambda < (1/R)$

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

here  $\tilde{e}_\lambda$  is the Moreau  $\lambda$ -envelope of  $\tilde{f}$ . From this we conclude that  $(\tilde{e}_\lambda)''_{0, 0} = (e_\lambda)''_{\tilde{x}+\lambda\tilde{v}, \tilde{v}}$ . Finally note that  $\nabla e_\lambda(\tilde{x} + \lambda\tilde{v}) = \tilde{v}$ . (because  $\tilde{v} \in \partial f(\tilde{x})$ ),  $\tilde{f}''_{0, 0} = f''_{x, v}$ , and that  $\tilde{f}$  is prox-regular at 0 for 0 with respect to  $\epsilon$  and  $R$ , in particular with (4.1) holding ( $\tilde{f}$  in place of  $f$  and  $R$  in place of  $r$ ). Now simply apply Theorem 6.5 to the function  $\tilde{f}$ . □

We finish off by deriving from the theory that has been developed a powerful property of the second-order epi-derivatives of prox-regular functions.

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

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

*In the case of  $f$  strictly twice epi-differentiable at  $\bar{x}$  for  $\bar{v}$ , the second-order difference quotient functions  $\Delta^2_{x, v, t} f$  not only epi-converge to  $f''_{\bar{x}, \bar{v}}$  as  $t \downarrow 0$  and  $(x, v) \rightarrow (\bar{x}, \bar{v})$  with  $f(x) \rightarrow f(\bar{x})$  and  $v \in \partial f(x)$ , but converge uniformly on all bounded sets.*

*Proof.* Let the parameter values for the prox-regularity be  $\epsilon$  and  $r$ . Without loss of generality we can put ourselves in the setting where  $\bar{x} = 0, \bar{v} = 0$ , and (4.1)

holds. Then by Theorem 3.2, an  $f$ -attentive localization of  $\partial f + rI$  around  $(0, 0)$  is monotone, in fact by Proposition 4.8 it is locally maximal monotone relative to  $(0, 0)$ . To simplify matters, we can add the term  $(r/2)|\cdot|^2$  to  $f$ , which has the effect of adding  $rI$  to  $\partial f$  and  $r|\cdot|^2$  to  $f''_{0,0}$ . In this way we can reduce to the case where  $f$  is nonnegative and a certain  $f$ -attentive localization  $T$  of  $\partial f$  around  $(0, 0)$  is locally maximal monotone.

Let  $f_0(\xi) = \frac{1}{2}f''_{0,0}(\xi)$ , and for  $t > 0$ ,  $f_t(\xi) = \frac{1}{2}\Delta_{0,0,t}^2 f(\xi) = f(t\xi)/t^2$ , these functions, like  $f$ , being nonnegative and having the value 0 at the origin. Our assumption of second-order epi-differentiability means that  $f_t$  epi-converges to  $f_0$  as  $t \downarrow 0$ . Theorem 6.1(a) tells us that this is equivalent to  $T$  being proto-differentiable at 0 for 0, i.e., the mappings  $T_t(\xi) = T(t\xi)/t$  converging in graph to  $T_0 = T'_{0,0}$  as  $t \downarrow 0$ , and that  $T_0 = \partial f_0$ . These mappings, like  $T$  are monotone. Hence  $f_0$ , being l.s.c. and proper with  $\partial f_0$  monotone, must be convex [12], hence in particular lower- $\mathcal{C}^2$  (cf. [25]). Because the convex function  $f_0$  is finite everywhere, the mapping  $\partial f_0 = T_0$  is everywhere nonempty-valued.

Let  $\bar{T}$  be any maximal monotone extension of  $T$ . Because  $T$  is locally maximal around  $(0, 0)$ , we know there is a neighborhood  $U$  of 0 in  $\mathbb{R}^n$  such that  $\bar{T}(x) \cap U = T(x) \cap U$  when  $x \in U$ . Since the proto-differentiability of  $T$  at 0 for 0 depends only on the nature of the graph of  $T$  in a neighborhood of  $(0, 0)$ , this property carries over to  $\bar{T}$ . Thus, the mappings  $\bar{T}_t(\xi) = \bar{T}(t\xi)/t$ , which are maximal monotone globally, converge in graph to  $T_0$  too.

Consider any simplex in  $\mathbb{R}^n$  having the origin in its interior; this simplex is the convex hull of its vertex set  $\{\xi_0, \xi_1, \dots, \xi_n\}$ . For certain  $\delta > 0$  sufficiently small, it will be true that whenever  $|\xi'_i - \xi_i| \leq \delta$  for  $i = 0, 1, \dots, n$ , the convex hull of  $\{\xi'_0, \xi'_1, \dots, \xi'_n\}$  is likewise a simplex having the origin in its interior. Because  $\bar{T}_t$  converges in graph to  $T_0$  and  $T_0(\xi_i) \neq \emptyset$  for all  $i$ , we know that when  $t > 0$  is sufficiently small there exists for each  $i$  a point  $\xi'_i$  with  $|\xi'_i - \xi_i| \leq \delta$  and  $\bar{T}_t(\xi'_i) \neq \emptyset$ . The set  $\text{dom } \bar{T}_t := \{\xi \mid T_t(\xi) \neq \emptyset\}$  then includes  $\{\xi'_0, \xi'_1, \dots, \xi'_n\}$ , so the interior of its convex hull contains the origin. But  $\bar{T}_t$  is maximal monotone, so the interior of the convex hull of  $\text{dom } \bar{T}_t$  is the same as the interior of  $\text{dom } \bar{T}_t$ , with  $\bar{T}_t$  locally bounded on this interior; cf. [21, Thm. 1]. It follows that  $\bar{T}_t$  is locally bounded at 0, hence  $\bar{T}$  has this property. Then  $T$  itself must be locally bounded at 0.

The local boundedness of  $T$  at 0 says in particular that whenever  $v_k$  is a proximal subgradient of  $f$  at  $x_k$ , and  $x_k \rightarrow 0$  with  $f(x_k) \rightarrow f(0)$ , the sequence  $\{v_k\}$  is bounded. This condition is known to imply that  $f$  is Lipschitz continuous around 0, cf. [24]. Then actually  $T = \partial f$  on a neighborhood of 0. The monotonicity of  $T$  implies that  $f$  is convex on such a neighborhood, say  $U$ ; cf. [12]. But then the functions  $f_t$ , which epi-converge to  $f_0$ , are convex on  $t^{-1}U$ . Since convex functions epi-converge on an open set  $O$  to a finite function if and only if they converge pointwise uniformly on every bounded subset of  $O$  (see Salinetti and Wets [33]), our claim about the second-order expansion of  $f$  is justified.

Now that we know  $f$  is convex on a neighborhood of 0, we also know that  $f$  is subdifferentially continuous at 0 and, for  $(x, v)$  sufficiently near to  $(0, 0)$  in  $\text{gph } \partial f$ , the function  $\Delta_{x,v,t} f$  is convex. Under the assumption that  $f$  is not just twice epi-differentiable at  $\bar{x} = 0$  for  $\bar{v} = 0$ , but strictly twice epi-differentiable, the convex functions  $\Delta_{x,v,t} f$  epi-converge to  $f_0$ . Again, appealing to the same fact about epi-convergence of convex functions to a finite function, we conclude that the difference quotients  $\Delta_{x,v,t}$  must converge pointwise to  $f_0$  uniformly on all bounded sets.  $\square$

## REFERENCES

1. H. Attouch, *Variational convergence for functions and operators*, Pitman, 1984. MR **86f**:49002
2. H. Attouch, and R. J-B Wets, *Epigraphical analysis*, *Analyse non linéaire* **6** (1989), 73–100. MR **91g**:90161
3. J.V. Burke and R. A. Poliquin, *Optimality conditions for non-finite convex composite functions*, *Mathematical Programming* **57** (1992), 103–120. MR **93g**:90066
4. F. H. Clarke, *Optimization and Nonsmooth Analysis*, *Classics in Applied Math.*, vol. 5, SIAM Publications, Philadelphia, 1990 (originally published in 1983). MR **85m**:49002
5. ———, *Methods of Dynamic and Nonsmooth Optimization*, CBMS-NSF Regional Conference Series in Applied Mathematics 57, SIAM Publications, Philadelphia, 1989. MR **91j**:49001
6. F. H. Clarke, R. J. Stern and P. R. Wolenski, *Proximal smoothness and the lower- $C^2$  property*, preprint (1994).
7. R. Cominetti, *On Pseudo-differentiability*, *Trans. Amer. Math. Soc.* **324** (1991), 843–865. MR **91h**:26009
8. C. Do, *Generalized second derivatives of convex functions in reflexive Banach spaces*, *Trans. Amer. Math. Soc.* **334** (1992), 281–301. MR **93a**:49011
9. A. D. Ioffe, *Variational analysis of a composite function: a formula for the lower second order epi-derivative*, *J. Math. Anal. Appl.* **160** (1991), 379–405. MR **92m**:46061
10. A. Levy, R. A. Poliquin and L. Thibault, *Partial extension of Attouch's theorem with applications to proto-derivatives of subgradient mappings*, *Trans. Amer. Math. Soc.* **347** (1995), 1269–1294. MR **95**:07
11. A. Levy and R. T. Rockafellar, *Variational conditions and the proto-differentiation of partial subgradient mappings*, *Nonlinear Anal. Th. Meth. Appl.* (to appear).
12. R. A. Poliquin, *Subgradient monotonicity and convex functions*, *Nonlinear Analysis, Theory, Methods & Applications*, **14** (1990), 305–317. MR **91b**:90155
13. ———, *Proto-differentiation of subgradient set-valued mappings*, *Canadian J. Math.* **42** (1990), 520–532. MR **91g**:49007
14. ———, *Integration of subdifferentials of nonconvex functions*, *Nonlinear Anal. Th. Meth. Appl.* **17** (1991), 385–398. MR **92j**:49008
15. ———, *An extension of Attouch's Theorem and its application to second-order epi-differentiation of convexly composite functions*, *Trans. Amer. Math. Soc.* **332** (1992), 861–874. MR **93a**:49013
16. R. A. Poliquin and R. T. Rockafellar, *Amenable functions in optimization*, *Nonsmooth Optimization Methods and Applications* (F. Giannessi, ed.), Gordon and Breach, Philadelphia, 1992, pp. 338–353. MR **95d**:49033
17. ———, *A calculus of epi-derivatives applicable to optimization*, *Canadian J. Math.* **45** (1993), 879–896. MR **94d**:49023
18. ———, *Proto-derivative formulas for basic subgradient mappings in mathematical programming*, *Set-Valued Analysis*, **2** (1994), 275–290. MR **95c**:49024
19. ———, *Generalized Hessian properties of regularized nonsmooth functions*, *SIAM J. Optimization* (to appear).
20. ———, *Second-order nonsmooth analysis in nonlinear programming*, *Recent Advances in Nonsmooth Optimization* (D. Du, L. Qi and R. Womersley, eds.), World Scientific Publishers, 1995, pp. 322–350.
21. R. A. Poliquin, J. Vanderwerff and V. Zizler, *Convex composite representation of lower semicontinuous functions and renormings*, *C.R. Acad. Sci. Paris* **317 Série I** (1993), 545–549. MR **94i**:46102
22. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970. MR **43**:445
23. ———, *Local boundedness of nonlinear monotone operators*, *Michigan Math.* **16** (1969), 397–407. MR **40**:6229
24. ———, *Proximal subgradients, marginal values, and augmented Lagrangians in nonconvex optimization*, *Math. of Op. Res.* **6** (1981), 424–436. MR **83m**:90088
25. ———, *Favorable classes of Lipschitz continuous functions in subgradient optimization*, *Progress in Nondifferentiable Optimization* (E. Nurminski, ed.), IIASA Collaborative Pro-

- ceedings Series, International Institute of Applied Systems Analysis, Laxenburg, Austria, 1982, pp. 125-144. MR **85c**:90069
26. ———, *Maximal monotone relations and the second derivatives of nonsmooth functions*, Ann. Inst. H. Poincaré: Analyse non linéaire **2** (1985), 167–184. MR **87c**:49021
  27. ———, *First- and second-order epi-differentiability in nonlinear programming*, Trans. Amer. Math. Soc. **307** (1988, 75–107). MR **90a**:90216
  28. ———, *Proto-differentiability of set-valued mappings and its applications in optimization*, Analyse Non Linéaire (H. Attouch et al., eds.), Gauthier-Villars, Paris, 1989, pp. 449–482. MR **90k**:90140
  29. ———, *Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives*, Math. of Oper. Research **14** (1989), 462–484. MR **91b**:49022
  30. ———, *Generalized second derivatives of convex functions and saddle functions*, Trans. Amer. Math. Soc. **320** (1990), 810–822. MR **91b**:90190
  31. ———, *Nonsmooth analysis and parametric optimization*, Methods of Nonconvex Analysis (A. Cellina, ed.), vol. 1446, Springer-Verlag Lecture Notes in Math, 1990, pp. 137–151. MR **91i**:49016
  32. ———, *Lagrange multipliers and optimality*, SIAM Review **35** (1993), 183–238. MR **94h**:49004
  33. G. Salinetti and R. J-B Wets, *On the relation between two types of convergence for convex functions*, J. Math. Anal. Appl. **60** (1977), 211–226. MR **57**:18828
  34. L. Thibault and D. Zagrodny, *Integration of subdifferentials of lower semicontinuous functions*, J. Math. Anal. Appl. **189** (1995), 33–58.
  35. J. P. Vial, *Strong and weak convexity of sets and functions*, Math of Oper. Res. **8** (1983), 231–259. MR **84m**:90107

DEPARTMENT OF MATHEMATICAL SCIENCES, UNIVERSITY OF ALBERTA, EDMONTON, ALBERTA,  
CANADA T6G 2G1

*E-mail address:* [rene@fenchel.math.ualberta.ca](mailto:rene@fenchel.math.ualberta.ca)

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WASHINGTON, SEATTLE, WASHINGTON 98195

*E-mail address:* [rtr@math.washington.edu](mailto:rtr@math.washington.edu)