A DERIVATIVE-CODERIVATIVE INCLUSION IN SECOND-ORDER NONSMOOTH ANALYSIS

R. T. Rockafellar[1]
Department of Mathematics, University of Washington
Seattle, WA 98195-4350

and

D. Zagrodny[2]
Institute of Mathematics, Technical University of Łódź
Łódź, ul. Żwirki 36, Poland

Abstract. For twice smooth functions, the symmetry of the matrix of second partial derivatives is automatic and can be seen as the symmetry of the Jacobian matrix of the gradient mapping. For nonsmooth functions, possibly even extended-real-valued, the gradient mapping can be replaced by a subgradient mapping, and generalized second derivative objects can then be introduced through graphical differentiation of this mapping, but the question of what analog of symmetry might persist has remained open. An answer is provided here in terms of a derivative-coderivative inclusion.

September, 1996

[1] Supported by National Science Foundation Grant DMS 9500957.
[2] Supported by a Fulbright Grant.

1. Introduction

A familiar property of a C^2 function $f : \mathbb{R}^n \to \mathbb{R}$ at any point $\bar{x} \in \mathbb{R}^n$ is the symmetry of the Hessian matrix $\nabla^2 f(\bar{x})$, which is the Jacobian matrix $\nabla(\nabla f)(\bar{x})$ associated with the gradient mapping $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$. What analogous property might hold in the second-order nonsmooth analysis of functions $f : \mathbb{R}^n \to \overline{\mathbb{R}} = [-\infty, \infty]$ and their subgradient mappings $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$? This question has not previously been asked, but we pose it here in the framework of graphical derivatives and coderivatives of ∂f and provide an answer covering a major class of functions f.

Second-order nonsmooth analysis builds on the geometry of tangent and normal cones, just like first-order nonsmooth analysis. Recall that for a set $C \subset \mathbb{R}^n$, the *tangent cone* at a point $\bar{x} \in C$ is defined by

$$T_C(\bar{x}) = \limsup_{t \searrow 0} \frac{1}{t}[C - \bar{x}],$$

while the general (unconvexified) *normal cone* has the expression

$$N_C(\bar{x}) = \limsup_{x \underset{C}{\to} \bar{x}} T_C(x)^*$$

with "$x \underset{C}{\to} \bar{x}$" referring to convergence to $\bar{x}$ within C. The same cone $N_C(\bar{x})$ is generated if the polar cone $T_C(x)^*$ in this formula is replaced by the *proximal* normal cone

$$N_C^p(x) = \big\{ v \,\big|\, \exists\, \lambda > 0 : P_C(x + \lambda v) = x \big\},$$

where $P_C : \mathbb{R}^n \rightrightarrows C$ denotes the projection onto C. (We use " $\rightrightarrows$ " to signal the potential set-valuedness of a mapping.) A vector v belongs to $N_C^p(x)$ if and only if $x \in C$ and, for some $\lambda > 0$, one has

$$\langle v,\, x' - x \rangle \le \frac{1}{2\lambda}|x' - x|^2 \ \text{ for all } \ x' \in C.$$

The subgradient mapping ∂f associates with each point $\bar{x}$ with $f(\bar{x})$ finite the set

$$\partial f(\bar{x}) = \big\{ v \,\big|\, (v, -1) \in N_{\mathrm{epi}\, f}\big(\bar{x}, f(\bar{x})\big) \big\}.$$

Two notions of "second derivative" of f can be developed out of the tangent and normal cone geometry of the graph $\mathrm{gph}\, \partial f$ of ∂f in $\mathbb{R}^n \times \mathbb{R}^n$. For any choice of $\bar{v} \in \partial f(\bar{x})$ one can define the graphical *derivative* mapping $D(\partial f)(\bar{x} \,|\, \bar{v}) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ by

$$D(\partial f)(\bar{x} \,|\, \bar{v})(w) = \big\{ z \,\big|\, (w, z) \in T_{\mathrm{gph}\, \partial f}(\bar{x}, \bar{v}) \big\}$$

and also the *coderivative* mapping $D^*(\partial f)(\bar{x}\,|\,\bar{v}) : I\!R^n \rightrightarrows I\!R^n$ by

$$D^*(\partial f)(\bar{x}\,|\,\bar{v})(w) = \big\{ z \,\big|\, (z, -w) \in N_{\mathrm{gph}\,\partial f}(\bar{x}, \bar{v}) \big\}.$$

Such derivatives and coderivatives make sense with respect to any mapping (possibly set-valued), and they have already been studied extensively, in particular by Mordukhovich [1]. But are there special relationships between them in the subgradient setting?

A clue coming from the case of a C^2 function f is seen through the reduction of ∂f to the gradient mapping ∇f, which is single-valued and C^1. The notion for derivatives and coderivatives can be simplified then to $D(\nabla f)(\bar{x})(w)$ and $D^*(\nabla f)(\bar{x})(w)$, since $\bar{v} = \nabla f(\bar{x})$ automatically. It's easy to see that the mappings $D(\nabla f)(\bar{x})$ and $D^*(\nabla f)(\bar{x})$ both come out as the linear transformation corresponding to the Hessian matrix $\nabla^2 f(\bar{x})$. More generally for a C^1 mapping F with Jacobian $\nabla F(\bar{x})$, $DF(\bar{x})$ is the linear transformation with matrix $\nabla F(\bar{x})$, whereas $D^*F(\bar{x})$ is the linear transformation corresponding to $\nabla F(\bar{x})^*$, the transpose matrix, so this gives the idea of looking at graphical derivatives and coderivatives more generally for analogs of symmetry.

Beyond the C^2 case, one could begin by speculating that the mappings $D(\partial f)(\bar{x}\,|\,\bar{v})$ and $D^*(\partial f)(\bar{x}\,|\,\bar{v})$ might typically turn out to be the same despite the absence of Hessians. But simple examples puncture such hopes even for functions f of class C^{1+} (differentiable with ∇f locally lipschitzian). For instance, the function f on $I\!R^1$ with $f(x) = \frac{1}{2}x^2$ for $x \geq 0$ and $f(x) = -\frac{1}{2}x^2$ for $x \leq 0$, has $\partial f(x) = \nabla f(x) = f'(x) = |x|$ and consequently $D(\partial f)(0\,|\,0)(w) = |w|$ for all w. But $D^*(\partial f)(0\,|\,0)(w)$ is $\big\{ z \,\big|\, -|w| \leq z \leq |w| \big\}$ for $w \geq 0$ and $\{-w, w\}$ for $w < 0$. Note, however, that although $D(\partial f)(0\,|\,0)(w) \neq D^*(\partial f)(0\,|\,0)(w)$ in this example when $w \neq 0$, one does have $D(\partial f)(0\,|\,0)(w) \subset D^*(\partial f)(0\,|\,0)(w)$ for all w.

Our main goal in this paper is to prove that, in a range of important situations, the graph of the derivative mapping $D(\partial f)(\bar{x}\,|\,\bar{v})$ is always included within the graph of the coderivative mapping $D^*(\partial f)(\bar{x}\,|\,\bar{v})$.

Theorem 1.1. *Suppose that the function $f : I\!R^n \to \overline{I\!R}$ is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v}$, and that ∂f is proto-differentiable at $\bar{x}$ for $\bar{v}$ (these properties being present for any $\bar{v} \in \partial f(\bar{x})$ when f is fully amenable at $\bar{x}$). Then*

$$D(\partial f)(\bar{x}\,|\,\bar{v})(w) \subset D^*(\partial f)(\bar{x}\,|\,\bar{v})(w) \quad \text{for all } w. \tag{1.1}$$

The meaning of the terms in this statement is as follows. The function f is *subdifferentially continuous* at $\bar{x}$ for $\bar{v}$ if $\bar{v} \in \partial f(\bar{x})$ and, whenever $x^\nu \to \bar{x}$ and $v^\nu \to \bar{v}$ with $v^\nu \in \partial f(x^\nu)$, one automatically has $f(x^\nu) \to f(\bar{x})$. (Sequences in this paper are always indexed by superscript ν.) On the other hand, f is *prox-regular* at $\bar{x}$ for $\bar{v}$ if $\bar{v} \in \partial f(\bar{x})$, the set $\operatorname{epi} f$ is closed relative to a neighborhood of $\big(\bar{x}, f(\bar{x})\big)$, and there exist $\varepsilon > 0$ and $r > 0$ such that, whenever $v \in \partial f(x)$ with $|x - \bar{x}| < \varepsilon$ and $|v - \bar{v}| < \varepsilon$, one has

$$f(x') \geq f(x) + \langle v,\, x' - x\rangle - \frac{r}{2}|x' - x|^2 \text{ when } |x' - x| < \varepsilon.$$

These terms were introduced by Poliquin and Rockafellar [2], who noted that the properties are enjoyed not only by C^2 functions and convex functions, but more generally by strongly amenable functions.

As defined in [3], f is *amenable* at $\bar{x}$ if it has a local representation around $\bar{x}$ as $g \circ F$ for a C^1 mapping $F : V \to I\!R^m$ on a neighborhood V of $\bar{x}$ and a proper, lsc, convex function $g : I\!R^m \to \overline{I\!R}$ such that no nonzero vector $y \in N_{\operatorname{dom} g}\big(F(\bar{x})\big)$ has $\nabla F(\bar{x})^* y = 0$. It is *strongly amenable* when F is C^2 around $\bar{x}$ rather than just C^1, and it is *fully amenable* if, in addition, g is *piecewise linear-quadratic*, i.e., $\operatorname{dom} g$ is expressible as a union of finitely many polyhedral sets, relative to each of which g is given by a polynomial function of degree at most 2. These functions have a key role in second-order variational analysis and perturbation theory; see [3], [4], [5], [6].

Also included in the class of functions that are both subdifferentially continuous and prox-regular are all p.l.n. functions [7], [8] (cf. [2, Proposition 2.2]), and all lower-C^2 functions (cf. [2, Example 2.7]), hence in particular all C^{1+} functions. Indeed, these functions too are strongly amenable.

The concept of proto-differentiability of general set-valued mappings was introduced by Rockafellar [9]. The proto-differentiability of ∂f at $\bar{x}$ for $\bar{v}$, where $\bar{v} \in \partial f(\bar{x})$, means that the graphs of the difference quotient mappings

$$\Delta_t(\partial f)(\bar{x} \,|\, \bar{v}) : w \mapsto \frac{1}{t}\big[\partial f(\bar{x} + tw) - \bar{v}\big]$$

converge as $t \searrow 0$; in other words, they don't just have the graph of $D(\partial f)(\bar{x} \,|\, \bar{v})$ as their "lim sup" (as implied by the definition of that mapping), but as their limit in the sense of set convergence in $I\!R^n \times I\!R^n$. This property has been shown by Poliquin [10] to hold when f is fully amenable at $\bar{x}$. If f is a C^{1+} function, so that ∂f reduces to ∇f and $\bar{v} = \nabla f(\bar{x})$, proto-differentiability is equivalent to (one-sided) directional differentiability, and for that matter to the semidifferentiability or B-differentiability of ∇f at $\bar{x}$.

Corollary 1.2. *When $f : I\!R^n \to \overline{I\!R}$ is strongly amenable at $\bar{x}$, the inclusion (1.1) holds for any $\bar{v} \in \partial f(\bar{x})$ such that ∂f is proto-differentiable at $\bar{x}$ for $\bar{v}$. The proto-differentiability assumption is superfluous when f is fully amenable at $\bar{x}$, and in that case, moreover,*

$$\limsup_{\substack{(x,v) \to (\bar{x},\bar{v}) \\ v \in \partial f(x)}} \operatorname{gph} D(\partial f)(x \,|\, v) \;\subset\; \operatorname{gph} D^*(\partial f)(\bar{x} \,|\, \bar{v}). \tag{1.2}$$

Proof. This is clear from the foregoing citations and the fact that when f is fully amenable at $\bar{x}$ there exists $\varepsilon > 0$ such that f is fully amenable at every point x with $|x - \bar{x}| \leq \varepsilon$ and $f(x) \leq f(\bar{x}) + \varepsilon$. Then the inclusion (1.1) holds also at all points $(x, v) \in \operatorname{gph} \partial f$ sufficiently near to $(\bar{x}, \bar{v})$ (due to subdifferential continuity), and one gets (1.2), since

$$\limsup_{\substack{(x,v)\to(\bar{x},\bar{v})\\ v\in\partial f(x)}} \operatorname{gph} D^*(\partial f)(x \,|\, v) \;\subset\; \operatorname{gph} D^*(\partial f)(\bar{x} \,|\, \bar{v})$$

by the definition of these coderivative graphs via normal cones. □

Theorem 1.1 will be proved in Section 3, after we first lay a foundation in Section 2 in the easier setting of C^{1+} functions. The need for the proto-differentiability assumption in Theorem 1.1 will be demonstrated in Section 3 as well. The one-dimensional case, which goes through with weaker assumptions, will be treated in Section 4.

2. Smooth Functions with Lipschitzian Gradient

As a step along the way toward proving Theorem 1.1, we will demonstrate that the derivative-coderivative inclusion holds for the gradient mappings ∇f of functions f of class C^{1+} under assumptions of directional differentiability. Rather than working with the full graphical inclusion, we focus on establishing the inclusion $D(\nabla f)(\bar{x})(\bar{w}) \subset D(\nabla f)^*(\bar{x})(\bar{w})$ for a single vector $\bar{w}$. This impels us to assume an enhanced property of directional differentiability at $\bar{x}$ relative to $\bar{w}$.

Definition 2.1. *Let the single-valued mapping $F : I\!R^n \to I\!R^n$ be locally lipschitzian. We say that F is adequately directionally differentiable at $\bar{x}$ for $\bar{w}$ when*

(a) $\lim_{t\searrow 0}\big[F(\bar{x}+t\bar{w}) - F(\bar{x})\big]/t$ *exists and*

(b) *for some sequence $t^\nu \searrow 0$ one has*

$$\limsup_{s\to 0}\left\{\frac{1+s}{s}\limsup_{\nu\to\infty}\left|\frac{F(\bar{x}+t^\nu(1+s)w)-F(\bar{x})}{t^\nu(1+s)} - \frac{F(\bar{x}+t^\nu w)-F(\bar{x})}{t^\nu}\right|\right\} \leq O(w-\bar{w})$$

for a function $O : I\!R^n \to [0,\infty]$ that is continuous at 0 with $O(0) = 0$.

In what follows, we denote by $I\!B(a, r)$ the closed ball of radius r around a.

Proposition 2.2. *If there is a neighborhood $I\!B(\bar{w}, r)$ such that the directional derivative $\lim_{t\searrow 0}\big[F(\bar{x}+tw) - F(\bar{x})\big]/t$ exists for every $w \in I\!B(\bar{w}, r)$, then F is adequately directionally differentiable at $\bar{x}$ for $\bar{w}$. Thus in particular, if F is directionally differentiable at*

$\bar{x}$ *(in all directions), then, no matter what the choice of* $\bar{w}$*,* F *is adequately directionally differentiable at* $\bar{x}$ *for* $\bar{w}$*.*

Proof. The upper limit with respect to ν in (b) of the definition is 0 when $w \in I\!B(\bar{w}, r)$, regardless of the choice of the sequence $t^\nu \searrow 0$. $\square$

Corollary 2.3. *When* $n = 1$*, the existence of* $\lim_{t \searrow 0} \big[F(\bar{x} + t\bar{w}) - F(x)\big]/t$ *with* $\bar{w} \neq 0$ *already in itself ensures adequate directional differentiability at* $\bar{x}$ *for* $\bar{w}$*.*

Proof. The existence of the limit persists when $\bar{w}$ is replaced by multiples $\alpha\bar{w}$ for $\alpha > 0$, and because $n = 1$, such multiples constitute a neighborhood of $\bar{w}$. $\square$

The implication in Proposition 2.2 can't be reversed: there are cases where F is adequately directionally differentiable at $\bar{x}$ for $\bar{w}$, but there exists $r > 0$ such that F is not directionally differentiable at $\bar{x}$ for any $w \in I\!B(\bar{w}, r)$ that isn't a multiple of $\bar{w}$. For example, this is seen for $F : I\!R^2 \to I\!R^2$ with $\bar{x} = (0, 0)$, $\bar{w} = (1, 0)$, and $F(x_1, x_2) = \big(\varphi_1(x_1)\varphi_2(x_2), 0\big)$ when $\varphi_1 : I\!R \to I\!R$ is a C^1 function with $\varphi_1(0) = 1$, but $\varphi_2 : I\!R \to I\!R$ is a function with $\varphi_2(0) = 0$ whose right and left derivatives don't exist at 0.

Theorem 2.4. *Let* $f : I\!R^n \to I\!R$ *be a* C^{1+} *function such that* ∇f *is adequately directionally differentiable at* $\bar{x}$ *for* $\bar{w}$*. Then*

$$D(\nabla f)(\bar{x})(\bar{w}) \subset D(\nabla f)^*(\bar{x})(\bar{w}). \tag{2.1}$$

Corollary 2.5. *If* $f : I\!R^n \to I\!R$ *is a* C^{1+} *function such that* ∇f *is directionally differentiable at* $\bar{x}$*, then* $\operatorname{gph} D(\nabla f)(\bar{x}) \subset \operatorname{gph} D^*(\nabla f)(\bar{x})$*, i.e.,*

$$D(\nabla f)(\bar{x})(w) \subset D(\nabla f)^*(\bar{x})(w) \text{ for all } w \in I\!R^n.$$

If ∇f *is directionally differentiable at all points* x *in a neighborhood of* $\bar{x}$*, then also*

$$\limsup_{x \to \bar{x}} \operatorname{gph} D(\nabla f)(x) \subset \operatorname{gph} D^*(\nabla f)(\bar{x}).$$

Proof. This combines Theorem 2.4 with Proposition 2.2 and the fact that the "lim sup" inclusion holds automatically with $\operatorname{gph} D^*(\nabla f)(x)$ in place of $\operatorname{gph} D(\nabla f)(x)$. $\square$

The proof of Theorem 2.4 will require two auxiliary facts.

Lemma 2.6. *Let* C^ν *and* C *be nonempty, closed sets such that* $C^\nu \to C$ *as* $\nu \to \infty$*, and let* $\bar{v} \in N_C(\bar{x})$*. Then there exist* $x^\nu \in C^\nu$ *and* $v^\nu \in N^p_{C^\nu}(x^\nu)$ *with* $(x^\nu, v^\nu) \to (\bar{x}, \bar{v})$*.*

Proof. The pair $(\bar{x}, \bar{v})$ can be approximated arbitrarily closely by pairs (x, v) with $v \in N^p_C(x)$, so there is no harm is supposing that $\bar{v}$ is itself a proximal subgradient, i.e.,

$\bar v \in N^p_C(\bar x)$. By definition, then, there exists $\lambda > 0$ such that $\bar x$ is the unique point of C nearest to $\bar x + \lambda \bar v$. For each index ν, choose any

$$x^\nu \in P_{C^\nu}(\bar x + \lambda \bar v) = \operatorname*{argmin}_{x \in C^\nu} |x - (\bar x + \lambda \bar v)|.$$

The vector $(\bar x + \lambda \bar v) - x^\nu$ belongs then to $N^p_{C^\nu}(x^\nu)$. Let $v^\nu = \big[(\bar x + \lambda \bar v) - x^\nu\big]/\lambda$. Then v^ν belongs to $N_{C^\nu}(x^\nu)$ as well. By demonstrating that $x^\nu \to \bar x$, we will confirm at the same time that $v^\nu \to \bar v$.

Because $C^\nu \to C$, there exist points $\bar x^\nu \in C^\nu$ with $\bar x^\nu \to \bar x$. We have $|x^\nu - (\bar x + \lambda \bar v)| \le |\bar x^\nu - (\bar x + \lambda \bar v)|$, hence

$$\limsup_{\nu \to \infty} |x^\nu - (\bar x + \lambda \bar v)| \le \lim_{\nu \to \infty} |\bar x^\nu - (\bar x + \lambda \bar v)| = |\bar x - (\bar x + \lambda \bar v)| = \lambda |\bar v|.$$

Thus, the sequence of points x^ν is bounded, and every cluster point of this sequence lies in the closed ball of radius $\lambda|\bar v|$ around $\bar x + \lambda \bar v$. But this ball touches C only at $\bar x$. Therefore, the sequence in question has to converge to $\bar x$. □

Lemma 2.7. *When f is a function of class C^{1+}, there exists for any ball $I\!B(\bar x, r)$ a constant $\rho > 0$ such that the function $f_\rho(x) = f(x) + \frac{1}{2}\rho|x - \bar x|^2$ is convex on $I\!B(\bar x, r)$. One has $\partial f_\rho(\bar x) = \partial f(\bar x)$ and, for any vector $\bar v$ in this subgradient set, also*

$$D(\partial f_\rho)(\bar x \,|\, \bar v) = D(\partial f)(\bar x \,|\, \bar v) + \rho I \quad \text{and} \quad D^*(\partial f_\rho)(\bar x \,|\, \bar v) = D^*(\partial f)(\bar x \,|\, \bar v) + \rho I.$$

Proof. Since ∇f is locally lipschitzian, there exists $\rho > 0$ such that $|\nabla f(x') - \nabla f(x)| \le \rho|x' - x|$ when $x, x' \in I\!B(\bar x, \rho)$. Then f_ρ, having $\nabla f_\rho(x) = \nabla f(x) + \rho(x - \bar x)$, satisfies

$$\langle \nabla f(x') - \nabla f(x),\, x' - x\rangle \ge 0 \ \text{ for } \ x, x' \in I\!B(\bar x, \rho).$$

This guarantees that f_ρ is convex on $I\!B(\bar x, \rho)$.

We have $\partial f_\rho = \partial f + \rho \nabla h$ for the function $h(x) = \frac{1}{2}|x - \bar x|^2$. For any C^1 mapping $F : I\!R^n \to I\!R^n$ with $F(\bar x) = 0$, one has $D(\partial f + F)(\bar x \,|\, \bar v)(w) = D(\partial f)(\bar x \,|\, \bar v)(w) + \nabla F(x)w$ and also $D^*(\partial f + F)(\bar x \,|\, \bar v)(w) = D^*(\partial f)(\bar x \,|\, \bar v)(w) + \nabla F(x)^* w$ by the calculus of Mordukhovich in [1]. The formulas for $D(\partial f_\rho)(\bar x \,|\, \bar v)$ and $D^*(\partial f_\rho)(\bar x \,|\, \bar v)$ are then immediate. □

Proof of Theorem 2.4. On the basis of Lemma 2.7, there is no loss of generality in assuming that f is convex. Let $F = \nabla f$ and $\bar z \in DF(\bar x)(\bar w)$. This means that $(\bar w, \bar z) \in T_{\operatorname{gph} F}(\bar x, \bar v)$ for $\bar v = F(\bar x)$. In order to verify (2.1), we have to show that $\bar z \in D^* F(\bar x)(\bar w)$, or in other words that $(\bar z, -\bar w) \in N_{\operatorname{gph} F}(\bar x, \bar v)$.

By hypothesis, the mapping $F = \nabla f$ is locally lipschitzian, and conditions (a) and (b) of Definition 2.1 hold. The lipschitzian property guarantees through (a) that

$$\lim_{\substack{t \searrow 0 \\ w \to \bar{w}}} \frac{F(\bar{x}+tw) - F(\bar{x})}{t} = \lim_{t \searrow 0} \frac{F(\bar{x}+t\bar{w}) - F(\bar{x})}{t} = \bar{z}. \tag{2.2}$$

With respect to the sequence $t^\nu \searrow 0$ in (b), define the mappings $H^\nu : I\!\!R^n \to I\!\!R^n$ by $H^\nu(w) = \big[F(\bar{x}+t^\nu w) - F(\bar{x})\big]/t^\nu$, noting that then

$$H^\nu = \nabla h^\nu \quad \text{for} \quad h^\nu(w) = \frac{f(\bar{x}+t^\nu w) - f(\bar{x}) - t^\nu \langle \nabla f(\bar{x}), w\rangle}{t^{\nu\,2}},$$

these functions h^ν being convex and of class C^{1+}.

Relative to any compact, convex subset of $I\!\!R^n$, the mappings H^ν are equi-continuous, in fact equi-lipschitzian with a constant they inherit from F. It follows by application of the Arzela-Ascoli theorem that some subsequence converges uniformly on all compact subsets of $I\!\!R^n$ to a mapping $H : I\!\!R^n \to I\!\!R^n$, which itself must be locally lipschitzian. Passing to such a subsequence if necessary, we can suppose that the whole sequence has this convergence property. Then in particular the sets $G^\nu = \operatorname{gph} H^\nu$ converge to $G = \operatorname{gph} H$. Each H^ν, because it's the gradient mapping for some convex function, is maximal cyclically monotone; cf. [11]. The same just be true then of H, and this guarantees that $H = \nabla h$ for some convex function h (again from [11]); here h is C^{1+}. We have $H^\nu(0) = 0$ for all ν, hence also $H(0) = 0$.

By Rademacher's theorem, the mapping $H = \nabla h$ is differentiable almost everywhere. Hence we can find a sequence of vectors $w^\nu \to \bar{w}$ for which the Jacobian matrices $\nabla H(w^\nu)$ exist. These matrices must be symmetric; this property for a C^{1+} function h was proved by Rockafellar and Poliquin [12, Cor. 3.3] (any function of class C^{1+} is in particular a lower-C^2 function). The locally uniform convergence of H^ν to H implies that $H^\nu(w^\nu) \to H(\bar{w})$, and since $H^\nu(w^\nu) = \big[F(\bar{x}+t^\nu w^\nu) - F(\bar{x})\big]/t^\nu$ we deduce from (2.2) that $H(\bar{w}) = \bar{z}$.

Our assumption that F has the limit property in (b) of Definition 2.1 can now be put to use. In terms of the mappings H^ν we have

$$\Big(\frac{1+s}{s}\Big)\left|\frac{F(\bar{x}+t^\nu(1+s)w) - F(\bar{x})}{t^\nu(1+s)} - \frac{F(\bar{x}+t^\nu w) - F(\bar{x})}{t^\nu}\right|$$
$$= \Big(\frac{1+s}{s}\Big)\left|\frac{H^\nu\big([1+s]w\big)}{1+s} - H^\nu(w)\right| = \left|\frac{H^\nu(w+sw) - H^\nu(w)}{s} - H^\nu(w)\right|.$$

The convergence of H^ν to H yields from the limit property in (b) that

$$\limsup_{s \to 0} \left|\frac{H(w+sw) - H(w)}{s} - H(w)\right| \le O(w - \bar{w}).$$

Applying this at w^ν, where the expansion $H(w^\nu+z) = H(w^\nu)+\nabla H(w^\nu)z+o(|z|)$ is valid, we obtain

$$\limsup_{s\to 0}\left|\nabla H(w^\nu)w^\nu - \frac{o\big(s|w^\nu|\big)}{s} - H(w^\nu)\right| \le O(w^\nu - \bar{w})$$

and consequently $\big|\nabla H(w^\nu)w^\nu - H(w^\nu)\big| \le O(w^\nu - \bar{w})$, so that

$$\lim_{\nu\to\infty} \nabla H(w^\nu)w^\nu = H(\bar{w}) = \bar{z}. \tag{2.3}$$

Let $z^\nu = H(w^\nu)$, so that $(w^\nu, z^\nu) \in G = \operatorname{gph} H$ and $(w^\nu, z^\nu) \to (\bar{w}, \bar{z})$. The tangent cone $T_G(w^\nu, z^\nu)$ is the graph of the linear transformation with matrix $\nabla H(w^\nu)$; this is the subspace M of $I\!R^n \times I\!R^n$ consisting of all pairs $(\omega, \nabla H(w^\nu)\omega)$ as ω ranges over $I\!R^n$. The polar cone $T_G(w^\nu, z^\nu)^*$ is the orthogonal subspace $M^\perp$, which consists of all pairs $(-\nabla H(w^\nu)^*\zeta, \zeta)$ as ζ ranges over $I\!R^n$. But $\nabla H(w^\nu)^* = \nabla H(w^\nu)$. Hence $(\nabla H(w^\nu)w^\nu, -w^\nu) \in T_G(w^\nu, z^\nu)^*$. In passing to the limit as $\nu \to \infty$ and invoking (2.3) along with the definition of the normal cone $N_G(\bar{w}, \bar{z})$, we see that

$$\big(\bar{z}, -\bar{w}\big) \in N_G\big(\bar{w}, \bar{z}\big).$$

Lemma 2.6 now provides us with a sequence of points $(\bar{w}^\nu, \bar{z}^\nu) \in G^\nu = \operatorname{gph} H^\nu$ with $(\bar{w}^\nu, \bar{z}^\nu) \to (\bar{w}, \bar{z})$ and proximal normals $(\zeta^\nu, -\omega^\nu) \in N^p_{G^\nu}(\bar{w}^\nu, \bar{z}^\nu)$ such that $(\zeta^\nu, -\omega^\nu) \to (\bar{z}, -\bar{w})$. Here $\bar{z}^\nu = H^\nu(\bar{w}^\nu) = \big[F(\bar{x} + t^\nu\bar{w}^\nu) - F(\bar{x})\big]/t^\nu$ and moreover $G^\nu = \operatorname{gph} H^\nu = (1/t^\nu)\big[\operatorname{gph} F - (\bar{x}, \bar{v})\big]$ with $\bar{v} = F(\bar{x})$, so that

$$N^p_{G^\nu}(\bar{w}^\nu, \bar{z}^\nu) = N^p_{\operatorname{gph} F}(\bar{x} + t^\nu\bar{w}^\nu,\ \bar{v} + t^\nu\bar{z}^\nu) = N^p_{\operatorname{gph} F}\big(\bar{x} + t^\nu\bar{w}^\nu, F(\bar{x} + t^\nu\bar{w}^\nu)\big).$$

Because this cone contains $(\zeta^\nu, -\omega^\nu)$, we obtain in the limit as $(\zeta^\nu, -\omega^\nu) \to (\bar{z}, -\bar{w})$ and $\big(\bar{x} + t^\nu\bar{w}^\nu, F(\bar{x} + t^\nu\bar{w}^\nu)\big) \to \big(\bar{x}, F(\bar{x})\big)$ that $(\bar{z}, -\bar{w}) \in N_{\operatorname{gph} F}(\bar{x}, \bar{v})$. This is the conclusion that we needed. □

3. Extension to Nonsmooth Functions

Although Theorem 2.4 refers only to a special class of smooth functions, its applicability is actually much wider. This is best seen in the framework of Corollary 2.5. Everything there can be translated into geometry. The basic inclusion can be put in the form

$$J\Big[T_{\operatorname{gph}\nabla f}(\bar{x},\nabla f(\bar{x}))\Big] \subset N_{\operatorname{gph}\nabla f}(\bar{x},\nabla f(\bar{x})) \quad \text{for} \quad J:(w,z)\mapsto(z,-w). \tag{3.1}$$

The directional differentiability assumption, equivalent to proto-differentiability inasmuch as ∇f is locally lipschitzian, means that the sets $(1/t)\big[\operatorname{gph}\nabla f - (\bar{x},\nabla f(\bar{x}))\big]$ converge to $T_{\operatorname{gph}\nabla f}(\bar{x},\nabla f(\bar{x}))$ as $t\searrow 0$, rather than merely having this cone as their "lim sup." In principle, then, we can apply the result to the graph of any subgradient mapping that, by way of an orthogonal linear transformation L (a linear isometry), can be converted locally to a portion of $\operatorname{gph}\nabla f$. Some nonorthogonal transformations L can be utilized as well.

Lemma 3.1. *Let G and G' be subsets of $\mathbb{R}^n\times\mathbb{R}^n$ such that $G=L(G')$ for a nonsingular linear transformation L such that $L^*JL=J$, where $J(w,z)=(z,-w)$. Then*

$$J\big[T_G(0,0)\big]\subset N_G(0,0) \quad\Longleftrightarrow\quad J\big[T_{G'}(0,0)\big]\subset N_{G'}(0,0).$$

Proof. It is elementary from $G=L(G')$ that $T_G(0,0)=L\big[T_{G'}(0,0)\big]$ and $N_G(0,0)=(L^*)^{-1}\big[N_{G'}(0,0)\big]$. If $J\big[T_G(0,0)\big]\subset N_G(0,0)$, we get $L^*JL\big[T_{G'}(0,0)\big]\subset N_{G'}(0,0)$, hence $J\big[T_{G'}(0,0)\big]\subset N_{G'}(0,0)$ under our assumption about L. The converse implication must then be valid as well, since $G'=L^{-1}(G)$ and $J^{-1}=J$, so that $(L^{-1})^*JL^{-1}=J$. □

Remarkably, the graphs of many subgradient mappings, even for nonconvex functions with nonconvex domain, can be converted into the graphs of gradient mappings in just such a manner. This has emerged from the theory of Moreau envelopes as developed recently by Poliquin and Rockafellar [2].

Proposition 3.2 ([2], Thm. 4.4, Prop. 4.6)**.** *Let $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ be subdifferentially continuous and prox-regular at $\bar{x}=0$ for $\bar{v}=0$, and also bounded from below by a quadratic function on $\mathbb{R}^n$. Then for $\lambda>0$ sufficiently small the envelope function*

$$e_\lambda f(x)=\inf_{x'\in\mathbb{R}^n}\Big\{f(x')+\frac{1}{2\lambda}|x'-x|^2\Big\}$$

is C^{1+} on a neighborhood of 0 with $\nabla(e_\lambda f)(0)=0$ and, locally around $(0,0)$, one has

$$v=\nabla(e_\lambda f)(x) \quad\Longleftrightarrow\quad v\in\partial f(x-\lambda v).$$

This gives us precisely what we need in order to derive our main result. Note that although the results in Section 2 were formulated for finite functions of class C^{1+} everywhere on $I\!R^n$, all that really was involved was this property on a neighborhood of $\bar{x}$. Thus, the local character of Proposition 3.2 presents no handicap.

Proof of Theorem 1.1. No loss of generality is incurred in taking $\bar{x} = 0$ and $\bar{v} = 0$, since this merely amounts to a translation of the graph of ∂f. We can suppose f to be lower semicontinuous (since by closing $\operatorname{epi} f$, which in prox-regularity is already closed locally around $\big(\bar{x}, f(\bar{x})\big)$, there is no effect on the subgradients that we are dealing with). Furthermore, by adding to f the indicator of some compact neighborhood of $\bar{x} = 0$ if necessary, we can make f be bounded from below. Then we are in the framework of Proposition 3.2.

The last assertion of Proposition 3.2 tells us that the intersection of $\operatorname{gph} \partial f$ with some neighborhood of $(0,0)$ is mapped onto the intersection of $\operatorname{gph} \nabla(e_\lambda f)$ with some neighborhood of $(0,0)$ by the linear transformation L with inverse $L^{-1} : (x,v) \mapsto (x - \lambda v, v)$, namely $L : (x', v) \mapsto (x' + \lambda v, v)$. The assumed proto-differentiability of ∂f at 0 for 0 translates to the directional differentiability of $\nabla(e_\lambda f)$ at 0.

It remains only to check that $L^* J L = J$, because in that case the principle in Lemma 3.1 can be invoked to reach the desired conclusion. In matrix terms, the equation $L^* J L = J$ comes out as

$$\begin{bmatrix} I & 0 \\ \lambda I & I \end{bmatrix} \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} \begin{bmatrix} I & \lambda I \\ 0 & I \end{bmatrix} = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$

This matrix equation is correct, as verified by direct multiplication, and the proof is therefore complete. □

The proto-differentiability assumption in Theorem 1.1 can't be removed without endangering the derivative-coderivative inclusion. We establish this through an example.

Proposition 3.3. *There is a function f that meets all the assumptions of Theorem 1.1, except for the proto-differentiability of ∂f at $\bar{x}$ for $\bar{v}$, and for which the derivative-coderivative inclusion (1.1) fails. In fact, there is such a function that is finite, convex and of class C^{1+}.*

Proof. We construct an example in stages. First we define two auxiliary functions $p, q : I\!R \to I\!R$ in terms of indices $i = 1, 2, \ldots$ as follows:

$$p(t) := \begin{cases} 0 & \text{if } t \le 0 \text{ or } t \ge 2^{-1}, \\ 3(t - 2^{-i-1}) & \text{if } 2^{-i-1} \le t \le \frac{3}{4} 2^{-i}, \\ -3(t - 2^{-i}) & \text{if } \frac{3}{4} 2^{-i} \le t \le 2^{-i}, \end{cases}$$

$$q(t) := \begin{cases} 0 & \text{if } t \le 0 \text{ or } t \ge \frac{3}{4}2^{-4}, \\ 2(t - \frac{3}{4}2^{-4i-4}) & \text{if } \frac{3}{4}2^{-4i-4} \le t \le \frac{3}{4}2^{-4i-3}, \\ -\frac{1}{7}(t - \frac{3}{4}2^{-4i}) & \text{if } \frac{3}{4}2^{-4i-3} \le t \le \frac{3}{4}2^{-4i}. \end{cases}$$

To the right of the origin, these are nonnegative "sawtooth" functions with more and more teeth as the origin is approached; for p, the slopes alternate between -3 and 3, whereas for q they alternate between $-\frac{1}{7}$ and 2. Farther to the right of the origin, and everywhere to the left of the origin, both functions vanish. Now we define

$$f(x_1, x_2) := \int_0^{x_1} p(t)dt + \int_0^{x_2} q(t)dt + \delta_D(x_1, x_2) \quad \text{with}$$
$$D := \big\{(x_1, x_2) \in I\!R^2 \,\big|\, 0 \le 7x_1 \le 8x_2 \le 9x_1\big\}.$$

The convex set D is a "sector" bounded by two rays in the first quadrant of $I\!R^2$, the slopes of which are $\frac{7}{8}$ and $\frac{9}{8}$. Although f is not convex, the function f_0 that is defined in the same way, but with $p(t)$ and $q(t)$ replaced by the nondecreasing expressions $p_0(t) = p(t) + 3t$ and $q_0(t) = q(t) + t$, is convex; we have $f(x_1, x_2) = f_0(x_1, x_2) - \frac{3}{2}x_1^2 - \frac{1}{2}x_2^2$, so f is strongly amenable. (Any function obtained by subtracting a C^2 function from a proper, lsc, convex function is strongly amenable.) The subgradients of f are given by

$$\partial f(x_1, x_2) = \big(p(x_1), q(x_2)\big) + N_D(x_1, x_2), \tag{3.2}$$

where the normal cone is generated for points $(x_1, x_2) \in D$ by

$$N_D(x_1, x_2) = \begin{cases} \big\{s_1(7, -8) + s_2(-9, 8) \,\big|\, s_1 \ge 0,\ s_2 \ge 0\big\} & \text{if } (x_1, x_2) = (0, 0), \\ \big\{s_1(7, -8) \,\big|\, s_1 \ge 0\big\} & \text{if } 0 < 7x_1 = 8x_2, \\ \big\{s_2(-9, 8) \,\big|\, s_2 \ge 0\big\} & \text{if } 0 < 8x_2 = 9x_1, \\ \{(0, 0)\} & \text{if } 0 < 7x_1 < 8x_2 < 9x_1. \end{cases}$$

It is evident from these formulas that $(0, 0) \in \partial f(0, 0)$. We will work with $\bar{x} = (0, 0)$ and $\bar{v} = (0, 0)$ to get a counterexample to the inclusion in Theorem 1.1. We begin by observing that for $t^\nu = \frac{3}{4}2^{-4\nu-3}$ the points (t^ν, t^ν) are interior to D with

$$p(t^\nu) = -3\left(\tfrac{3}{4}2^{-4\nu-3} - 2^{-4\nu-3}\right) = t^\nu, \qquad q(t^\nu) = 2\left(\tfrac{3}{4}2^{-4\nu-3} - \tfrac{3}{4}2^{-4\nu-4}\right) = t^\nu,$$

so that $(t^\nu, t^\nu) \in \partial f(t^\nu, t^\nu)$ by (3.2). Thus we have

$$(1, 1) \in \frac{1}{t^\nu}\big[\partial f\big(t^\nu(1, 1)\big) - (0, 0)\big] \text{ as } t^\nu \searrow 0,$$

and consequently $(1,1) \in D(\partial f)\big((0,0) \,|\, (0,0)\big)(1,1)$. We will demonstrate, however, that $(1,1) \notin D(\partial f)^*\big((0,0) \,|\, (0,0)\big)(1,1)$, or in other words that

$$(1,1,-1,-1) \notin N_{\text{gph}\,\partial f}(0,0,0,0). \tag{3.3}$$

Suppose that (3.3) is false; it will be shown that this is impossible. At elements $(x_1^\nu, x_2^\nu, v_1^\nu, v_2^\nu) \to (0,0,0,0)$ in $\text{gph}\,\partial f$ there are proximal normals $(z_1^\nu, z_2^\nu, -w_1^\nu, -w_2^\nu) \to (1,1,-1,-1)$. For certain values $\lambda^\nu > 0$ we have

$$\begin{aligned}&\big\langle (z_1^\nu, z_2^\nu, -w_1^\nu, -w_2^\nu),\, (x_1, x_2, v_1, v_2) - (x_1^\nu, x_2^\nu, v_1^\nu, v_2^\nu) \big\rangle \\ &\quad \leq \frac{1}{2\lambda^\nu}\big|(x_1, x_2, v_1, v_2) - (x_1^\nu, x_2^\nu, v_1^\nu, v_2^\nu)\big|^2 \quad \text{for all } (x_1, x_2, v_1, v_2) \in \text{gph}\,\partial f.\end{aligned} \tag{3.4}$$

In particular, the points (x_1^ν, x_2^ν) belong to the set D.

Let us first of all eliminate the case of infinitely many of these points (x_1^ν, x_2^ν) lying on the upper edge of D, with $8x_2^\nu = 9x_1^\nu \geq 0$. If there were such a subsequence (we may as well suppose it's the whole sequence), we could consider any fixed ν and take $(x_1, x_2, v_1, v_2) = (x_1^\nu, x_2^\nu, v_1^\nu - 9s, v_2^\nu + 8s)$ for arbitrary $s > 0$; this would satisfy $(v_1, v_2) \in \partial f(x_1, x_2)$ by (3.2). Then in (3.4) we would have

$$-\big\langle (w_1^\nu, w_2^\nu), (-9s, 8s) \big\rangle \leq \frac{1}{2\lambda^\nu}\big|(-9s, 8s)\big|^2.$$

Dividing both sides by s and taking the limit as $s \searrow 0$, we would get $9w_1^\nu - 8w_2^\nu \leq 0$. In passing then to the limit as $\nu \to \infty$ with $(w_1^\nu, w_2^\nu) \to (1,1)$, we would obtain $1 \leq 0$, a contradiction.

In the case of a subsequence lying instead on the lower edge of D, with $7x_1^\nu = 8x_2^\nu \geq 0$, we could similarly take $(x_1, x_2, v_1, v_2) = (x_1^\nu, x_2^\nu, v_1^\nu + 7s, v_2^\nu - 8s)$ for fixed ν and arbitrary $s > 0$ to deduce from (3.4) that $-7w_1^\nu + 8w_2^\nu \leq 0$ and, in the limit, again obtain $1 \leq 0$.

So, let's assume that the points (x_1^ν, x_2^ν) are interior to D, and (harmlessly too) that they are near enough to $(0,0)$ to be in the "sawtooth" intervals for p and q; in other words, $x_1^\nu < 2^{-1}$ and $x_2^\nu < \frac{3}{4}2^{-4}$. We have $(v_1^\nu, v_2^\nu) = \big(p(x_1^\nu), q(x_2^\nu)\big)$ then by (3.2).

This time, for arbitrary fixed ν take $(x_1, x_2) = (x_1^\nu + s, x_2^\nu)$ in (3.4) for $s \neq 0$ small enough that this still belongs to D. Let $(v_1, v_2) = \big(p(x_1^\nu + s), q(x_2^\nu)\big)$, so that $(v_1, v_2) \in \partial f(x_1, x_2)$. Then (3.4) yields

$$z_1^\nu s - w_1^\nu\big[p(x_1^\nu + s) - p(x_1^\nu)\big] \leq \frac{1}{2\lambda^\nu}\big|\big(s, [p(x_1^\nu + s) - p(x_1^\nu)]\big)\big|^2.$$

In dividing by s and taking the limit first as $s \searrow 0$, and second as $s \nearrow 0$, we arrive at the pair of inequalities

$$z_1^\nu - w_1^\nu p'_+(x_1^\nu) \leq 0 \qquad \text{and} \qquad z_1^\nu - w_1^\nu p'_-(x_1^\nu) \geq 0,$$

where $p'_+(x_1^\nu)$ and $p'_-(x_1^\nu)$ denote the right and left derivatives of p at x_1^ν. In the limit as $\nu \to \infty$, we have $z_1^\nu \to 1$, $w_1^\nu \to 1$ and $p'_-(x_1^\nu) \le z_1^\nu/w_1^\nu \le p'_+(x_1^\nu)$, hence

$$\limsup_{\nu\to\infty} p'_-(x_1^\nu) \;\le\; 1 \;\le\; \liminf_{\nu\to\infty} p'_+(x_1^\nu).$$

But $p'_+(x_1^\nu)$ and $p'_-(x_1^\nu)$ can only take the values 3 or -3, so we must have $p'_+(x_1^\nu) = 3$ and $p'_-(x_1^\nu) = -3$ for all ν sufficiently large. This requires x_1^ν to correspond to a "lower tooth" of p; it has to have the form

$$x_1^\nu = 2^{-i^\nu} \;\text{ for some integer } i^\nu > 0 \text{ when } \nu \text{ is sufficiently large.} \tag{3.5}$$

Returning now to the inequality in (3.4), we temporarily fix ν and let $(x_1, x_2) = (x_1^\nu, x_2^\nu + s)$ for arbitrary $s \neq 0$ small enough that this point still belongs to D. We take $(v_1, v_2) = \big(p(x_1^\nu), q(x_2^\nu + s)\big)$, so that $(v_1, v_2) \in \partial f(x_1, x_2)$ by (3.2). Since $v_2^\nu = q(x_2^\nu)$, we obtain from this case of (3.4) that

$$z_2^\nu s - w_2^\nu\big[q(x_2^\nu + s) - q(x_2^\nu)\big] \le \frac{1}{2\lambda^\nu}\big|\big(s, [q(x_2^\nu + s) - q(x_2^\nu)]\big)\big|^2.$$

In dividing this by s and taking the limit first as $s \searrow 0$ and second as $s \nearrow 0$, we produce the inequalities

$$z_2^\nu - w_2^\nu q'_+(x_2^\nu) \le 0 \qquad\text{and}\qquad z_2^\nu - w_2^\nu q'_-(x_2^\nu) \ge 0.$$

In taking the limit then as $\nu \to \infty$ we have $z_2^\nu \to 1$, $w_2^\nu \to 1$ and $q'_-(x_2^\nu) \le z_2^\nu/w_2^\nu \le q'_+(x_2^\nu)$, so that

$$\limsup_{\nu\to\infty} q'_-(x_2^\nu) \;\le\; 1 \;\le\; \liminf_{\nu\to\infty} q'_+(x_2^\nu).$$

Here, the values of $q'_+(x_2^\nu)$ and $q'_-(x_2^\nu)$ can only be 2 or $-\frac{1}{7}$. Hence $q'_+(x_2^\nu) = 2$ and $q'_-(x_2^\nu) = -\frac{1}{7}$ for all ν sufficiently large. Then

$$x_2^\nu = \tfrac{3}{4}2^{-4j^\nu} \;\text{ for some integer } j^\nu > 0 \text{ when } \nu \text{ is sufficiently large.} \tag{3.6}$$

But since (x_1^ν, x_2^ν) belongs to the interior of D, we have $7/8 < x_2^\nu/x_1^\nu < 9/8$, which implies through (3.5) and (3.6) that $(7/8)(4/3) < 2^{-4j^\nu}/2^{-i^\nu} < (9/8)(4/3)$, or in other words that $7/6 < 2^{i^\nu - 4j^\nu} < 3/2$ for all ν sufficiently large. This isn't possible, because no power of 2 lies in the interval $[7/6, 3/2]$. The contradiction validates the claims about f.

Knowing that f gives a counterexample to the derivative-coderivative inclusion in Theorem 1.1 in the absence of proto-differentiability, we know also that the same is true for the convex function f_0 introduced early in the proof. Indeed, we have $\partial f_0(x_1,x_2) = \partial f(x_1,x_2) - (3x_1,x_2)$, so that

$$D(\partial f_0)\big((0,0)\,|\,(0,0)\big)(w_1,w_2) = D(\partial f)\big((0,0)\,|\,(0,0)\big)(w_1,w_2) + (3w_1,w_2),$$
$$D^*(\partial f_0)\big((0,0)\,|\,(0,0)\big)(w_1,w_2) = D^*(\partial f)\big((0,0)\,|\,(0,0)\big)(w_1,w_2) + (3w_1,w_2).$$

A counterexample is given then too by $f_1 = f_0 + \delta_B$ for any closed ball B centered at the origin, since such a function f_1 has $\operatorname{gph}\partial f_1$ agreeing with $\operatorname{gph}\partial f_0$ around $(0,0,0,0)$. Utilizing Lemma 3.1 once more, we see that the envelope function $e_\lambda f_1$ for any $\lambda > 0$ likewise provides a counterexample. Because f_1 is lsc, proper and convex with $\operatorname{dom} f$ bounded, the function $e_\lambda f_1$ is finite, convex and of class C^{1+}. □

4. The One-Dimensional Case.

The counterexample constructed in the proof of Proposition 3.3 is two-dimensional, and with good reason. For functions of a single real variable, there is no need for the assumption of proto-differentiability.

Theorem 4.1. *If a function $f : I\!R \to \overline{I\!R}$ is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v}$, then*

$$D(\partial f)(\bar{x}\,|\,\bar{v})(w) \subset D^*(\partial f)(\bar{x}\,|\,\bar{v})(w) \quad \text{for all } w.$$

In particular this holds for any function $f : I\!R \to \overline{I\!R}$ that is C^{1+} on a neighborhood of $\bar{x}$; then ∂f reduces to f', and $\bar{v} = f'(\bar{x})$, so the inclusion comes out as

$$D(f')(\bar{x})(w) \subset D^*(f')(\bar{x})(w) \quad \text{for all } w.$$

Proof. In drawing on the technique used to prove Theorem 1.1, we are able to reduce to the case where f is of class C^{1+} around $\bar{x}$. There is nothing to prove if $w = 0$. The cases where $w > 0$ or $w < 0$ are symmetric, so it's enough to deal with $w > 0$. But the mappings $D(f')(\bar{x})$ and $D^*(f')(\bar{x})$ are positively homogeneous, so we can even just take $w = 1$. We consider any $\bar{z} \in D(f')(\bar{x})(1)$ and work toward confirming that $\bar{z} \in D^*(f')(\bar{x})(1)$.

If the right derivative $(f')'_+(\bar{x})$ actually existed (in which case the set $D(f')(\bar{x})(1)$ would be the singleton $\{\bar{z}\}$), we would be able to reach our goal immediately by applying Theorem 2.4 on the basis of the fact in Corollary 2.3. We suppose therefore that

$$\bar{z}_- < \bar{z}_+ \quad \text{for} \quad \bar{z}_- := \liminf_{t \searrow 0} \frac{f'(\bar{x}+t) - f'(\bar{x})}{t}, \quad \bar{z}_+ := \limsup_{t \searrow 0} \frac{f'(\bar{x}+t) - f'(\bar{x})}{t}, \tag{4.1}$$

these values being finite because f' is locally lipschitzian around $\bar{x}$.

Since $\bar{z} \in D(f')(\bar{x})(1)$, there are $w^\nu \to 1$ and $t^\nu \searrow 0$ with $\big[f'(\bar{x}+t^\nu w^\nu)-f'(\bar{x})\big]/t^\nu \to \bar{z}$. The same limit then holds with w^ν replaced by 1. Thus,

$$\bar{z} = \lim_{\nu\to\infty} z^\nu \;\text{ for }\; z^\nu := \frac{f'(\bar{x}+t^\nu)-f'(\bar{x})}{t^\nu}.$$

Either $\bar{z} > \bar{z}_-$ or $\bar{z} = \bar{z}_-$, and our argument splits into these two cases.

Case 1: $\bar{z} > \bar{z}_-$. Without loss of generality, we can suppose also that $z^\nu > \bar{z}_-$ for all ν. Define the auxiliary functions $r^\nu : [0,t^\nu] \to I\!R$ by $r^\nu(t) = z^\nu t - \big[f'(\bar{x}+t)-f'(\bar{x})\big]$. Then $r^\nu(0) = 0 = r^\nu(t^\nu)$ and

$$\limsup_{t\searrow 0} \frac{r^\nu(t)}{t} = z^\nu - \bar{z}_- > 0,$$

so that $r^\nu(t)$ must be positive for some $t \in (0,t^\nu)$. Let $s^\nu \in \operatorname{argmin}\big\{r^\nu(t) \,\big|\, 0 \le t \le t^\nu\big\}$; then $s^\nu \to 0$. We have $r^\nu(t) \le r^\nu(s^\nu)$ for all $t \in [0,t^\nu]$, and therefore

$$z^\nu[t-s^\nu] - \big[f'(\bar{x}+t)-f'(\bar{x}+s^\nu)\big] \le 0 \;\text{ for all }\; t \in [0,t^\nu].$$

In terms of $x = \bar{x}+t$ and $x^\nu = \bar{x}+s^\nu$, this says that

$$\big\langle (z^\nu,-1),\, (x,f'(x)) - (x^\nu,f'(x^\nu))\big\rangle \le 0 \;\text{ for all } x \text{ sufficiently near } x^\nu.$$

Hence $(z^\nu,-1)$ belongs to the polar of the tangent cone $T_{\operatorname{gph} f'}\big(x^\nu,f(x^\nu)\big)$. As $\nu\to\infty$ we have $\big(x^\nu,f(x^\nu)\big) \to \big(\bar{x},f(\bar{x})\big)$ and $(z^\nu,-1) \to (\bar{z},-1)$ so $(\bar{z},-1) \in N_{\operatorname{gph} f'}\big(\bar{x},f(\bar{x})\big)$ by the definition of this normal cone. Thus, $\bar{z} \in D^*(f')(\bar{x})(1)$.

Case 2: $\bar{z} = \bar{z}_-$. On the basis of (4.1) there exists $\varepsilon > 0$ with $\varepsilon < \bar{z}_+ - \bar{z}_-$. Define the auxiliary functions $r^\nu_\varepsilon : [0,t^\nu] \to I\!R$ by $r^\nu_\varepsilon(t) = \big[z^\nu+\varepsilon\big]t - \big[f'(\bar{x}+t)-f'(\bar{x})\big]$. This gives us $r^\nu_\varepsilon(0) = 0$, $r_\varepsilon(t^\nu) = \varepsilon t^\nu > 0$, and

$$\liminf_{t\searrow 0} \frac{r^\nu_\varepsilon(t)}{t} = z^\nu - \bar{z}_+ + \varepsilon. \tag{4.2}$$

Since $z^\nu \to \bar{z}_-$ and $z_- - \bar{z}_+ + \varepsilon < 0$, the right side of (4.2) is negative for large ν; we can suppose this true for all ν. Then each function r^ν_ε has negative values in every neighborhood of 0 relative to $[0,t^\nu]$, in addition to being positive at t^ν, so the continuity of r^ν_ε ensures the existence of $t^\nu_0 \in (0,t^\nu)$ with $r^\nu_\varepsilon(t^\nu_0) = 0$. Select any $s^\nu \in \operatorname{argmin}\big\{r^\nu_\varepsilon(t) \,\big|\, 0 \le t \le t^\nu_0\big\}$; then $s^\nu \to 0$. Since $r^\nu_\varepsilon(t) \le r^\nu_\varepsilon(s^\nu)$ for all $t \in [0,t^\nu_0]$, we have

$$(z^\nu+\varepsilon)[t-s^\nu] - \big[f'(\bar{x}+t)-f'(\bar{x}+s^\nu)\big] \le 0 \;\text{ for all }\; t \in [0,t^\nu_0].$$

Once more setting $x = \bar{x}+t$ and $x^\nu = \bar{x}+s^\nu$, we see that

$$\big\langle (z^\nu+\varepsilon,-1),\, (x,f'(x)) - (x^\nu,f'(x^\nu))\big\rangle \le 0 \;\text{ for all } x \text{ sufficiently near } x^\nu.$$

Hence $(z^\nu + \varepsilon, -1)$ belongs to the polar cone $T_{\mathrm{gph}\, f'}\big(x^\nu, f(x^\nu)\big)^*$. The same must then be true for $(z^\nu, -1)$, because this cone is closed and our argument has worked for any $\varepsilon \in (0,\, \bar{z}_+ - \bar{z}_-)$. Since $\big(x^\nu, f(x^\nu)\big) \to \big(\bar{x}, f(\bar{x})\big)$ and $(z^\nu, -1) \to (\bar{z}, -1)$ as $\nu \to \infty$, we may conclude again that $(\bar{z}, -1) \in N_{\mathrm{gph}\, f'}\big(\bar{x}, f(\bar{x})\big)$, and thus that $\bar{z} \in D^*(f')(\bar{x})(1)$. □

References

1. B. S. Mordukhovich, "Generalized differential calculus for nonsmooth and set-valued mappings," J. Math. Anal. Appl. 183 (1994), 250–288.

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

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

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

5. R. T. Rockafellar, "Nonsmooth analysis and parametric optimization," in *Methods of Nonconvex Analysis* (A. Cellina, ed.), Springer-Verlag Lecture Notes in Math. 1446 (1990), 137–151.

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

7. R. A. Poliquin, "Integration of subdifferentials of nonconvex functions," Nonlinear Anal. Th. Meth. Appl. 17 (1991), 385–396.

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

9. R. T. Rockafellar, " Proto-differentiability of set-valued mappings and its applications in optimization," Analyse Non Linéaire (H. Attouch, J.-P. Aubin, F. Clarke and I. Ekeland, eds.), Gauthier-Villars, Paris, 1989, 449-482.

10. R. A. Poliquin, "Proto-differentiation of subgradient set-valued mappings," Canadian J. Math. 42 (1990), 510-532.

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

12. R. A. Poliquin and R. T. Rockafellar, "Generalized Hessian properties of regularized nonsmooth functions," SIAM J. Optimization, forthcoming.