A CALCULUS OF EPI-DERIVATIVES APPLICABLE TO OPTIMIZATION

R. A. Poliquin and R. T. Rockafellar*

Abstract. When an optimization problem is represented by its essential objective function, which incorporates constraints through infinite penalties, first- and second-order conditions for optimality can be stated in terms of the first- and second-order epi-derivatives of that function. Such derivatives also are the key to the formulation of subproblems determining the response of a problem's solution when the data values on which the problem depends are perturbed. It is vital for such reasons to have available a calculus of epi-derivatives. This paper builds on a central case already understood, where the essential objective function is the composite of a convex function and a smooth mapping with certain qualifications, in order to develop differentiation rules covering operations such as addition of functions and a more general form of composition. Classes of "amenable" functions are introduced to mark out territory in which this sharper form of nonsmooth analysis can be carried out.

Keywords. Epi-derivatives, generalized second derivatives, amenable sets and functions, nonsmooth analysis, composite optimization, parametric optimization, sensitivity analysis.

Preprint, July 1991
Revised slightly, April 1992

---

* 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–8819586 for the second author.

1. Introduction.

Any optimization problem in $I\!R^n$ can be formulated in terms of minimizing an extended real-valued function f over all of $I\!R^n$. For instance if the given task is to minimize a function $f_0 : I\!R^n \to I\!R$ over a set $C \subset I\!R^n$, one can take $f(x) = f_0(x)$ for $x \in C$ but $f(x) = \infty$ for $x \notin C$. Then f is called the *essential* objective function for the problem. In general, when minimizing a function $f : I\!R^n \to \overline{I\!R} := I\!R \cup \{\pm\infty\}$ the effective domain $\operatorname{dom} f := \{x \mid f(x) < \infty\}$ represents the feasible solutions under consideration.

A central case is that of *composite optimization*, where f can be expressed as $f = g \circ F$ for a smooth mapping $F : I\!R^n \to I\!R^m$ and a lower semicontinuous, proper, convex function $g : I\!R^m \to \overline{I\!R}$. Then $\operatorname{dom} f = F^{-1}(\operatorname{dom} g)$. A vast class of problems can be perceived as having this form, and results about generalized derivatives of f in a context of nonsmooth analysis can accordingly be applied to characterize optimal solutions. The study of perturbations of optimal solutions benefits from such an approach as well, since the notion of an optimization problem in $x \in I\!R^n$ dependent on a parameter vector $u \in I\!R^d$ can be identified with that of an extended-real-valued function of $(u, x) \in I\!R^d \times I\!R^n$.

The goal of this paper is the derivation of some calculus rules for working in this context. These rules concern first- and second-order epi-derivatives, as introduced in Rockafellar [1] and developed further in Rockafellar [2],[3],[4], Cominetti [5], Do [6], Poliquin [7],[8], and Poliquin and Rockafellar [9]. A lower semicontinuous function $f : I\!R^n \to \overline{I\!R}$ is said to be *epi-differentiable* at a point x where $f(x)$ is finite if the first-order difference quotient functions $\Delta_{x,t} f : I\!R^n \to \overline{I\!R}$ defined by

$$\Delta_{x,t} f(\xi) = \big[f(x + t\xi) - f(x)\big]/t \ \text{ for } \ t > 0$$

epi-converge as $t \searrow 0$, the limit being a proper function (somewhere finite, nowhere $-\infty$). This limit is then the *epi-derivative* function f'_x. Epi-convergence refers to the convergence of the epigraphs of the functions in question as subsets of $I\!R^n \times I\!R$.

Similarly, f is *twice epi-differentiable* at x relative to a vector $v \in I\!R^n$ if it is epi-differentiable at x and the second-order difference quotient functions $\Delta^2_{x,v,t} f : I\!R^n \to \overline{I\!R}$ defined by

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

epi-converge to a proper function as $t \searrow 0$. The limit function is then the second-order epi-derivative, denoted by $f''_{x,v}(\xi)$.

Optimality conditions that mimic the classical ones for a smooth function can readily be stated for a twice epi-differentiable function f, as observed in Rockafellar [2, Thm. 2.2].

Necessary conditions: If $\bar{x}$ furnishes a local minimum of f, then $f'_x(\xi) \geq 0$ for all ξ and $f''_{x,0}(\xi) \geq 0$ for all ξ.

Sufficient conditions: If $\bar{x}$ is a point where $f'_x(\xi) \geq 0$ for all ξ and $f''_{x,0}(\xi) > 0$ for all $\xi \neq 0$, then $\bar{x}$ furnishes a local minimum of f in the strong sense.

These conditions are quite simple in nature but broad in applications. Although similar conditions can be brought to fruition under weaker restrictions on f than twice epi-differentiability, as developed recently by Ioffe [10] with only semiconvergence in the epi-graphical sense, most of the functions typically arising as essential objectives in finite-dimensional optimization actually do happen to be twice epi-differentiable. This has been demonstrated in Rockafellar [2] along with the fact that the standard kinds of optimality conditions, and many more properties as well, then follow from the specific form taken by the epi-derivatives in such cases. Likewise, epi-differentiation leads to a strong and versatile framework for the sensitivity analysis of solutions to problems of optimization [4],[9].

It is important therefore to ascertain as far as possible whether a function is once or twice epi-differentiable, and if so, what the derivatives are. The chief tool so far has been a chain rule established in Rockafellar [1] and supplemented by duality relations in Rockafellar [3] (for generalizations see Cominetti [5] and Do [6]). The effectiveness of a chain rule approach, as evidenced already in the papers cited, leads us to define two classes of functions according to the availability of local composite representations. We then work out a calculus within these classes, showing at the same time how the classes are preserved under various operations.

2. Amenable Functions

The idea of specifying a class of functions through the existence of certain composite representations is new to nonsmooth analysis but long familiar in other areas of mathematics, such as the theory of differentiable manifolds. In employing it here, our aim is to capture local aspects of convexity and smoothness which activate a sharper form of subdifferential calculus.

Definition 2.1. *A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ will be called* amenable *at $\bar{x}$, a point where $f(\bar{x})$ is finite, if on some open neighborhood V of $\bar{x}$ there is a $\mathcal{C}^1$ mapping $F : V \to \mathbb{R}^m$ and a proper, lower semicontinuous, convex function $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ such that $f(x) = g\big(F(x)\big)$ for $x \in V$ and*

$$\text{there is no } y \neq 0 \text{ in } N\big(F(\bar{x})|\operatorname{dom} g\big) \text{ with } \nabla F(\bar{x})^* y = 0. \tag{2.1}$$

Here $\nabla F(\bar{x})$ denotes the $m \times n$ Jacobian matrix of F at $\bar{x}$, and $\nabla F(\bar{x})^*$ is its transpose. Further, $N\big(F(\bar{x})|\operatorname{dom} g\big)$ is the normal cone to the nonempty convex set $\operatorname{dom} g$ at the point $F(\bar{x})$. It is appropriate to view (2.1) as a local *constraint qualification* for the condition $F(x) \in \operatorname{dom} g$, which locally around $\bar{x}$ describes the elements of $\operatorname{dom} f$, cf. [1],[2]. In terms of the tangent cone $T\big(F(\bar{x})|\operatorname{dom} g\big)$ to $\operatorname{dom} g$ at $F(\bar{x})$, which is polar to the normal cone

$N\big(F(\bar{x})|\operatorname{dom} g\big)$, the constraint qualification (2.1) can be written equivalently as

$$\nabla F(\bar{x})\, I\!R^n + T\big(F(\bar{x})|\operatorname{dom} g\big) = I\!R^m,$$

where $\nabla F(\bar{x})\, I\!R^n$ denotes the set of all vectors of the form $\nabla F(\bar{x})w$ with $w \in I\!R^n$. (This is because the vectors y belonging to the convex cone polar to $\nabla F(\bar{x})\, I\!R^n + T\big(F(\bar{x})|\operatorname{dom} g\big)$ are precisely the ones in $N\big(F(\bar{x})|\operatorname{dom} g\big)$ satisfying $\nabla F(\bar{x})^* y = 0$. A convex cone in $I\!R^m$ is equal to all of $I\!R^m$ if and only if its polar consists of just the zero vector.)

A special case of amenability is encountered when $m = n$ and F is a smooth mapping with nonsingular Jacobian, giving a local change of coordinates. The constraint qualification (2.1) holds trivially in that case. The realm of amenable functions thus includes all functions that would be lower semicontinuous, proper, convex functions "except for a poor choice of coordinates," or in other words, all curvilinear distortions of convex functions (and their effective domains). That notion falls short of conveying the essence of the class, however, because many functions that exhibit amenability do not appear to fit this picture (cf. the examples given below).

For the study of second-order properties, a refinement of amenability is useful.

Definition 2.2. *A function $f : I\!R^n \to \overline{I\!R}$ will be called fully amenable at $\bar{x}$ if the conditions in the preceding definition can be satisfied with the extra stipulation that F is a $\mathcal{C}^2$ mapping and g is piecewise linear-quadratic (convex). The latter means that* $\operatorname{dom} g$ *can be expressed as the union of a finite collection of polyhedral (convex) sets, on each of which g is given by a polynomial function with no terms higher than degree two.*

Examples of piecewise linear-quadratic (convex) functions g are polyhedral functions (having polyhedral epigraph), such as the indicator function δ_C and support function σ_C of a polyhedral set C. The convex function $\frac{1}{2}d_C^2$, where d_C gives the distance to such a set, is piecewise linear-quadratic although not polyhedral. A convex function is piecewise linear-quadratic if and only if its subdifferential mapping is polyhedral in the sense of Robinson [11], i.e., has a union of finitely many polyhedral sets as its graph, cf. Sun [12]. Therefore, the conjugate of a convex, piecewise linear-quadratic function is again piecewise linear-quadratic.

To appreciate the breadth of the classes specified in Definitions 2.1 and 2.2, it is important to understand that a given function f need not come already supplied with a composite representation of one of the types indicated, in order to be eligible for consideration. We only have to know that such a representation can be devised, at least locally.

Example 2.3. *Any lower semicontinuous, proper, convex function f is amenable at all points in* $\operatorname{dom} f$. *Any convex, piecewise linear-quadratic function f is fully amenable at all points in* $\operatorname{dom} f$.

Here the mapping F in Definitions 2.1 and 2.2 can be taken to be the identity.

Example 2.4. *Any $\mathcal{C}^1$ function f is everywhere amenable, whereas any $\mathcal{C}^2$ function f is everywhere fully amenable.*

This is the case where $m = 1$ in Definitions 2.1 and 2.2, and $g(w) = w$.

Example 2.5. *If $f = \max\{f_1, \dots, f_m\}$ for a family of C^1 functions $f_i : I\!R^n \to I\!R$, then f is everywhere amenable. If each f_i is $\mathcal{C}^2$, f is everywhere fully amenable.*

Obtain this example by taking $F(x) = \big(f_1(x), \dots, f_m(x)\big)$ along with $g(w_1, \dots, w_m) = \max\{w_1, \dots, w_m\}$. The function g is polyhedral.

A geometric side to amenability is reflected in a specialization to indicator functions, which provides further examples to which our calculus will be directed.

Definition 2.6. *A set $C \subset I\!R$ will be called amenable at a point $\bar{x} \in C$, if its indicator function δ_C is amenable at $\bar{x}$, or in other words, if for some open neighborhood V of $\bar{x}$ there is a $\mathcal{C}^1$ mapping $F : V \to I\!R^m$ and a closed, convex set $D \subset I\!R^m$ such that $V \cap C = \big\{x \in V \,\big|\, F(x) \in D\big\}$ and*

$$\text{there is no } y \neq 0 \text{ in } N\big(F(\bar{x})|D\big) \text{ with } \nabla F(\bar{x})^* y = 0. \tag{2.2}$$

Similarly, C is fully amenable at $\bar{x}$ if δ_C is fully amenable at $\bar{x}$, which means that the condition on F and D can be satisfied with F a $\mathcal{C}^2$ mapping and D a polyhedral set.

Again, the constraint qualification can be written in terms of tangents instead of normals: (2.2) is equivalent to

$$\nabla F(\bar{x})\, I\!R^n + T\big(F(\bar{x})|C\big) = I\!R^m.$$

Example 2.7. *Any closed, convex set C is amenable at all of its points. Any polyhedral set C is fully amenable at all of its points. (More generally, C is fully amenable at $\bar{x}$ if there is a polyhedral neighborhood V of $\bar{x}$ such that $C \cap V$ is polyhedral.)*

Example 2.8. *Let the set $C \subset I\!R^n$ be given by a system of finitely many constraints*

$$f_i(x) \leq 0 \text{ for } i = 1, \dots, s, \quad f_i(x) = 0 \text{ for } i = s+1, \dots, m, \tag{2.3}$$

involving $\mathcal{C}^1$ functions $f_i : I\!R^n \to I\!R$. For C to be amenable at a point $\bar{x} \in C$, it is necessary and sufficient that the Mangasarian-Fromovitz constraint qualification be satisfied at $\bar{x}$. In the case of $\mathcal{C}^2$ functions f_i, the same criterion gives full amenability.

Here let $F(x) = \big(f_1(x), \dots, f_m(x)\big)$ and let D be the polyhedral set in $I\!R^m$ consisting of all $w = (w_1, \dots, w_m)$ such that $w_i \leq 0$ for $i = 1, \dots, s$ but $w_i = 0$ for $i = s+1, \dots, m$.

One has $\bar{x} \in C$ if and only if $F(\bar{x}) \in D$, and then the cone $N\big(F(\bar{x})|D\big)$ consists of the vectors $y = (y_1, \ldots, y_m)$ such that $y_i \geq 0$ for $i \in [1, s]$ with $f_i(\bar{x}) = 0$, but $y_i = 0$ for $i \in [1, s]$ with $f_i(\bar{x}) < 0$; y_i can be anything for $i \in [s+1, m]$. Condition (2.2) requires that there be no vector of this type such that $\sum_{i=1}^m y_i \nabla f_i(\bar{x}) = 0$, except for $y = (0, \ldots, 0)$. This is well known as the equivalent dual form of the Mangasarian-Fromovitz constraint qualification.

The calculus rules in Section 3 will show how these primitive examples of amenability can be combined into others through addition, composition and further operations. In the present section the aim is to record the consequences of amenability which inspire such calculus.

We shall need to refer to subgradients not only of convex functions but nonconvex functions. There are several routes that can be taken in defining subgradients in the nonconvex case, but they all arrive at the same place as far as amenable functions are concerned, as will be seen. For the purpose at hand we rely on the formulation of Clarke [13], [14], which we now review to the basic extent needed.

The Clarke normal cone to a set $C \subset \mathbb{R}^n$ at a point $\bar{x} \in C$ is the closed convex hull of the cone consisting of the zero vector and all the vectors v for which there exists a sequence of points $x^\nu \notin \operatorname{cl} C$ (with $\nu = 1, 2, \ldots$) having nearest point projections $\bar{x}^\nu \in \operatorname{cl} C$ with $\bar{x}^\nu \to \bar{x}$, such that $\lambda^\nu(x^\nu - \bar{x}^\nu) \to v$ for some choice of scalars $\lambda^\nu > 0$ (cf. [13, section 2.4]). This cone is denoted here by $N(\bar{x}|C)$ rather than $N_C(\bar{x})$ to facilitate the treatment of sets with complicated labels like $\operatorname{dom} f$. When C is convex, $N(\bar{x}|C)$ agrees with the normal cone in the sense of convex analysis to which we have already referred.

The set C is *Clarke regular* (tangentially regular) at $\bar{x}$ if C is closed relative to some neighborhood of $\bar{x}$ and the cone polar to $N(\bar{x}|C)$, which is the Clarke tangent cone $T(\bar{x}|C)$, coincides with contingent cone (the Bouligand contingent cone) to C at $\bar{x}$ (cf. [13, p. 55]). This means that the vectors in $N(\bar{x}|C)$ are precisely the vectors v such that

$$\langle v,\, x - \bar{x} \rangle \leq o(x - \bar{x}) \;\text{ for }\; x \in C.$$

Many common types of sets are known to be Clarke regular, for instance convex sets and smooth manifolds, as well as sets defined by nice constraints as in Example 2.8 (cf. [13, pp. 55–59]).

For a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ these geometric notions are applied to the epigraph $\operatorname{epi} f = \big\{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \,\big|\, \alpha \geq f(x)\big\}$. A vector $v \in \mathbb{R}^n$ is a *subgradient* (generalized gradient) of f at $\bar{x}$, if $f(\bar{x})$ is finite and $(v, -1)$ belongs to the normal cone $N\big(\bar{x}, f(\bar{x}) \,\big|\, \operatorname{epi} f\big)$. It is a *horizon subgradient* (singular subgradient) if instead $(v, 0)$ belongs to this normal cone. These conditions are denoted by $v \in \partial f(\bar{x})$ and $v \in \partial^\infty f(\bar{x})$, respectively. Again, the general concept reduces to the familiar one of convex analysis when f is convex. When $f = \delta_C$ one has $\partial f(\bar{x}) = \partial^\infty f(\bar{x}) = N(\bar{x}|C)$.

The function f is *Clarke regular* at $\bar{x}$ if the set $\operatorname{epi} f$ is Clarke regular at $\big(\bar{x}, f(\bar{x})\big)$. In that event the vectors $v \in \partial f(\bar{x})$ are the vectors satisfying

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

Convex functions and smooth functions, in particular, are Clarke regular. This property is of strong interest in nonsmooth analysis because of its simplifying effect on various formulas for subgradients, and many other examples of Clarke regular functions are known in consequence of the theory of such formulas, cf. Clarke [13, pp. 59–61 and section 2.9].

The following theorem extracts from the results in Rockafellar [1][15] and Poliquin [7] the main implications for amenability as well as Clarke regularity. In this we recall from Rockafellar [16] that the set-valued mapping $\partial f : I\!R^n \rightrightarrows I\!R^n$ is *proto-differentiable* at x relative to the element $v \in \partial f(x)$ if the (set-valued) difference quotient mappings

$$\big(\Delta_{x,v,t}\partial f\big)(\xi) = \big[\partial f(x + t\xi) - v\big]/t$$

graph-converge as $t \searrow 0$. If so, the limit mapping is denoted by $(\partial f)'_{x,v}$ and called the *proto-derivative*. (It assigns to each $\xi \in I\!R^n$ a subset $(\partial f)'_{x,v}(\xi)$ of $I\!R^n$, which could be empty.)

Theorem 2.9. *If f is amenable at $\bar{x}$, then f is both epi-differentiable and Clarke regular at $\bar{x}$ with*

$$\begin{aligned} &\partial f(\bar{x}) = \big\{\, v \,\big|\, \langle v, \xi\rangle \leq f'_{\bar{x}}(\xi) \ \text{ for all } \xi \,\big\}, && \partial^\infty f(\bar{x}) = N(\bar{x}|\operatorname{dom} f),\\ &f'_{\bar{x}}(\xi) = \sup\big\{\, \langle v, \xi\rangle \,\big|\, v \in \partial f(\bar{x}) \,\big\}, && \operatorname{dom} f'_{\bar{x}} = T(\bar{x}|\operatorname{dom} f). \end{aligned} \tag{2.4}$$

If f is fully amenable at $\bar{x}$, it is in fact twice epi-differentiable there relative to every $v \in \partial f(\bar{x})$ (but not relative to any $v \notin \partial f(\bar{x})$). Moreover, the subgradient mapping ∂f is then proto-differentiable at $\bar{x}$ relative to every $v \in \partial f(\bar{x})$, with

$$(\partial f)'_{\bar{x},v}(\xi) = \partial\big(\tfrac{1}{2} f''_{\bar{x},v}\big)(\xi) \quad \text{for all } \xi. \tag{2.5}$$

Proof. The first-order properties are based on the composite representation in Definitions 2.1 and 2.2 along with the chain rule in [15, Prop. 2.2]. The properties of second-order epi-differentiability are based similarly on the chain rule in Rockafellar [1, Theorem 4.5]. The proto-differentiability of ∂f was established by Poliquin [7]. □

Formula (2.5) relating the proto-derivative of the subgradient mapping to the subgradients of the second-order epi-derivative was first established in the convex case by Rockafellar [3]. The formula was later extended to the setting of Theorem 2.9 by Poliquin [7], and recently, by Poliquin [8], to the setting of the composition of an arbitrary lower semicontinuous convex function and a $\mathcal{C}^2$ mapping with the constraint qualification (2.1). Formula (2.5) has tremendous applications to the study of perturbations of optimal solutions and associated multipliers in parametric optimization. In the setting of parametric optimization the formula is used to show that the proto-derivatives of the solution mapping can be obtained as primal and dual pairs for an auxiliary derivative problem; see Rockafellar [4] and Poliquin and Rockafellar [9].

In the case of an indicator function $f = \delta_C$, the first- and second-order epi-derivatives in Theorem 2.9 provide information about the local structure of C at $\bar{x}$. But while the first-derivative function is itself an indicator function (namely, for the tangent cone to C at $\bar{x}$), the second-derivative function is *not* an indicator, except in special circumstances such as C being polyhedral. Instead it provides a functional description of the "curvature" properties of C at $\bar{x}$.

Relative to a specific representation $f = g \circ F$, the epi-derivatives, normal vectors and subgradients in Theorem 2.9 come out according to [1, Theorem 4.5] as given by

$$\begin{aligned}
\partial f(\bar{x}) &= \nabla F(\bar{x})^* \partial g\big(F(\bar{x})\big), \\
N(\bar{x} | \operatorname{dom} f) &= \nabla F(\bar{x})^* N\big(F(\bar{x}) | \operatorname{dom} g\big), \\
f'_{\bar{x}}(\xi) &= g'_{F(\bar{x})}(\nabla F(\bar{x})\xi), \\
f''_{\bar{x},v}(\xi) &= \max_{y \in Y(\bar{x},v)} \Big\{ g''_{F(\bar{x}),y}\big(\nabla F(\bar{x})\xi\big) + \Big\langle \xi, \nabla^2 \langle y, F\rangle(\bar{x})\xi \Big\rangle \Big\}.
\end{aligned} \tag{2.6}$$

Here we refer to

$$Y(\bar{x}, v) = \big\{\, y \,\big|\, y \in \partial g\big(F(\bar{x})\big) \quad \text{with} \quad \nabla F(\bar{x})^* y = v \,\big\},$$

and to the function

$$\langle y, F\rangle : I\!R^n \to I\!R \text{ with } \langle y, F\rangle(x) := \big\langle y, F(x)\big\rangle \qquad (\text{where } y \in I\!R^m).$$

Actually the maximum in the second-order formula in (2.6) may be taken over $\operatorname{ext} Y(\bar{x}, v)$ i.e., the set of extreme points of $Y(\bar{x}, v)$.

An immediate consequence of the first-order formula in (2.6) is that in the case of a fully amenable function f, the set of subgradients of f at $\bar{x}$ is a *polyhedral* set, and the epi-derivative f'_x is a *piecewise linear positively homogeneous (of degree 1) convex function*. An important feature of the second-order formula in (2.6) is that the maximum is over a *finite* set (because $Y(\bar{x}, v)$ is a polyhedral set); other features of the second-order formula are identified below.

In the second-derivative formula in (2.6) the function g is of course piecewise linear-quadratic. Then, according to [1, Theorem 3.1], whenever $y \in \partial g(u)$ one actually has

$$\begin{aligned}
g''_{u,y}(\omega) &= \lim_{t \searrow 0} \big[g(u + t\omega) - g(u) - t\langle \omega, y\rangle\big] / \tfrac{1}{2}t^2 \\
&= \begin{cases} \lim_{t \searrow 0} \big[g(u + t\omega) - g(u) - t g'_u(\omega)\big] / \tfrac{1}{2}t^2 & \text{if } \langle \omega, y\rangle = g'_u(\omega), \\ \infty & \text{if } \langle \omega, y\rangle < g'_u(\omega). \end{cases}
\end{aligned} \tag{2.7}$$

It follows from (2.7) that

$$\operatorname{dom} f''_{\bar{x},v} = \big\{\, \xi \in \operatorname{dom} f'_x \;\big|\; f'_x(\xi) = \langle v, \xi\rangle \,\big\}.$$

The second-order formula in (2.6) is written differently than the one in [1, Theorem 4.5]; the reason for presenting it in this form is apparent from the chain rule formula in Theorem 3.5. The formulas are of course equal because for any $y \in Y(\bar{x}, v)$

$$\langle \nabla F(\bar{x})\xi, y\rangle = \langle \xi, \nabla F(\bar{x})^* y\rangle = \langle \xi, v\rangle,$$

It follows (by combining the second-order formula in (2.6) with (2.7)) that for any $\bar{y} \in Y(\bar{x}, v)$

$$f''_{\bar{x},v}(\xi) = g''_{F(\bar{x}),\bar{y}}\big(\nabla F(\bar{x})\xi\big) + \max_{y\in \operatorname{ext} Y(\bar{x},v)} \Big\{ \Big\langle \xi, \nabla^2\langle y, F\rangle(\bar{x})\xi \Big\rangle \Big\}. \tag{2.8}$$

By adding and subtracting $\lambda\|\xi\|^2$ to $f''_{\bar{x},v}(\xi)$, where λ is chosen so that for any $y \in \operatorname{ext} Y(\bar{x}, v)$ the function $\Big\langle \xi, \nabla^2\langle y, F\rangle(\bar{x})\xi \Big\rangle + \lambda\|\xi\|^2$ is convex, and because the maximum of finitely many purely quadratic functions is piecewise linear-quadratic, we have the following characterization: *The second-order epi-derivative of a fully amenable function is the sum of a piecewise linear-quadratic convex function homogeneous of degree 2 and a quadratic function.* By using formula (2.5) we have the following subgradient version: *The proto-derivative of the subgradient mapping of a fully amenable function is the sum of a polyhedral (in the sense of Robinson) homogeneous piecewise linear maximal monotone set-valued mapping and a symmetric linear transformation.*

Proposition 2.10. *If f is fully amenable as in Definition 2.2 and $f = g\circ F$ is a local representation around $\bar{x}$ in the sense required in that definition, then*

$$\begin{aligned}(\partial f)'_{\bar{x},v}(\xi) &= \operatorname{co}\Big\{\, \nabla F(\bar{x})^*(\partial g)'_{F(\bar{x}),y}\big(\nabla F(\bar{x})\xi\big) + \nabla^2\langle y, F\rangle(\bar{x})\xi \;\big|\; y \in \operatorname{ext} M(\bar{x}, v, \xi) \,\Big\} \\ &= \bigcup_{y\in M(\bar{x},v,\xi)} \Big\{\, \nabla F(\bar{x})^*(\partial g)'_{F(\bar{x}),y}\big(\nabla F(\bar{x})\xi\big) + \nabla^2\langle y, F\rangle(\bar{x})\xi \,\Big\},\end{aligned}$$

where $M(\bar{x}, v, \xi)$ denotes the set of vectors y furnishing the maximum in the second-derivative formula in (2.6).

Proof. To obtain the proto-derivative of the subgradient mapping all we need to do, according to (2.5), is evaluate the subgradient of the second-order epi-derivative. According to formula (2.8) and the calculus for the maximum over a compact set of quadratic functions (see Clarke [13]), we need only show that for any $y \in Y(\bar{x}, v)$ we have

$$\partial\big(g''_{F(\bar{x}),y} \circ \nabla F(\bar{x})\big)(\xi) = \nabla F(\bar{x})^* \partial g''_{F(\bar{x}),y}(\nabla F(\bar{x})\xi). \tag{2.9}$$

To show (2.9) first notice that (trivially)

$$\big(g''_{F(\bar{x}),y} \circ \nabla F(\bar{x})\big)'_{\xi}(\xi') = \big(g''_{F(\bar{x}),y}\big)'_{\nabla F(\bar{x})\xi}(\nabla F(\bar{x})\xi'). \tag{2.10}$$

Because the expression on the right of (2.10) is a lower semicontinuous function of ξ', the same can be said of the expression on the left of (2.10). This last remark enables us to show (2.9) because the closure of the directional derivative of a convex function is the support function of its subdifferential; see Rockafellar [17]. □

Although amenability may seem to be a condition focused on a single point at a time, it is truly a local condition *around* a point, as established by the next theorem. Amenability is therefore a much stronger condition than Clarke regularity, since a function—even one that is Lipschitz continuous—can be Clarke regular almost everywhere and yet fail to be Clarke regular on a dense set of points. This fact adds further motivation to the search for criteria for verifying amenability.

It also deserves to be noted that for functions f that are Clarke regular at $\bar{x}$, but not amenable there, the horizon subgradient set $\partial^\infty f(\bar{x})$ need not reduce to the normal cone $N(\bar{x}|\,\mathrm{dom}\, f)$ as it does in Theorem 2.9. A simple example is the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 1$ when $x > 0$, $f(x) = 0$ with $x \le 0$. At $\bar{x} = 0$ one has $\partial^\infty f(\bar{x}) = \mathbb{R}_+$ but $N(\bar{x}|\,\mathrm{dom}\, f) = \{0\}$.

Theorem 2.11. *If a function f is amenable at $\bar{x}$, there is a neighborhood U of $\bar{x}$ such that f is lower semicontinuous relative to U and amenable at all points $x \in U \cap \mathrm{dom}\, f$. In addition,*

$$\mathrm{gph}\,\partial f \ \textit{and}\ \mathrm{gph}\, N\big(\cdot\,\big|\,\mathrm{dom}\, f\big) \ \textit{are closed relative to}\ (U \cap \mathrm{dom}\, f) \times \mathbb{R}^n.$$

Likewise, if a set C is amenable at $\bar{x}$, there is a neighborhood U of $\bar{x}$ such that C is closed relative to U and amenable at all points of $U \cap C$. In addition,

$$\mathrm{gph}\, N\big(\cdot\,\big|C\big) \ \textit{is closed relative to}\ (U \cap C) \times \mathbb{R}^n.$$

All these assertions are valid also for full amenability.

Proof. If $f = g \circ F$ on a neighborhood of $\bar{x}$ in the pattern of Definition 2.1, it is clear that f is lower semicontinuous on some neighborhood and therefore bounded away from $-\infty$ on some neighborhood, since $f(\bar{x})$ is finite. The issue is whether condition (2.1) must carry over to all points of $\mathrm{dom}\, f$ sufficiently near to $\bar{x}$. If not, there would be a sequence $x^\nu \to \bar{x}$ along with nonzero vectors $y^\nu \in N\big(F(x^\nu)\big|\,\mathrm{dom}\, g\big)$ such that $\nabla F(x^\nu)^* y^\nu = 0$. By passing to the vectors $y^\nu/|y^\nu|$ (which still satisfy the same condition) and extracting a subsequence, we can suppose that y^ν converges to some y, where $|y| = 1$. Then $0 \ne y \in \partial g(\bar{x})$, because the graph of the subdifferential mapping associated with a lower semicontinuous, proper, convex function is closed [17,Theorem 24.4]. At the same time we have $\nabla F(\bar{x})^* y = 0$ by the continuity of the first derivatives of F. This situation would contradict the amenability of f at $\bar{x}$.

The fact that $\operatorname{gph} \partial f$ is closed relative to $(U \cap \operatorname{dom} f) \times \mathbb{R}^n$ follows from the first-order formula in (2.6) and the constraint qualification (2.1), with appeal again to the closedness of $\operatorname{gph} g$. Likewise one obtains the closedness of $\operatorname{gph} N\big(\cdot \big| \operatorname{dom} f\big)$ relative to $(U \cap \operatorname{dom} f) \times \mathbb{R}^n$: although $\operatorname{dom} f$ might not itself be closed, because the convex set $\operatorname{dom} g$ might not be closed, one can rely on the fact that (by convexity) $N\big(u \big| \operatorname{dom} g\big) = N\big(u \big| \operatorname{cl} \operatorname{dom} g\big)$ when $u \in \operatorname{dom} g$, where the set $\operatorname{gph} N\big(\cdot \,|\, \operatorname{cl} \operatorname{dom} g\big)$ is closed.

The claims in the case of a set C can be established similarly, or simply by specializing f to δ_C. For full amenability, no additional arguments are needed. □

3. Calculus Rules

Criteria for the preservation of Clarke regularity under various constructions applied to sets and functions have long been known and can be found in Clarke [13] and Borwein and Ward [18] as well as earlier work of Clarke [19] and Rockafellar [20]. Although amenability is a distinctly stronger property than Clarke regularity, the criteria for its preservation follow a similar pattern. From this standpoint the reader should see the *first-order* results in the following theorems essentially as observations that known theory has systematically sharper consequences than understood before, when applied in a more select yet very common situation.

The *second-order* results, on the other hand, have a different scope than anything previously offered through the strong properties in Theorem 2.9. For results on the calculus of other kinds of generalized second derivatives in nonsmooth analysis, we refer to Hiriart-Urruty [21], Hiriart-Urruty and Seeger [22], Cominetti and Correa [23], and Ioffe [10][24].

It is well to note at the outset that rules one might think would be easy to establish directly from the definitions of epi-derivatives actually present serious technical hurdles. This is due to the reliance of the amenability definitions on epi-convergence instead of pointwise convergence of functions. For instance, when two function sequences $\{f_1^\nu\}$ and $\{f_2^\nu\}$ epi-converge to f_1 and f_2, respectively, it does not immediately follow that $\{f_1^\nu + f_2^\nu\}$ epi-converges to $f_1 + f_2$. Conditions implying this are known for convex functions, cf. McLinden and Bergstrom [25], but not in any simple way for nonconvex functions, apart from some cases where epi-convergence can be seen to reduce to pointwise convergence.

Theorem 3.1 (addition rule). *Assume the functions $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ for $i = 1, \ldots, m$ are amenable at $\bar{x}$ and such that*

$$\text{if } v_1 + \ldots + v_m = 0 \text{ with } v_i \in N(\bar{x} | \operatorname{dom} f_i), \text{ then } v_1 = \cdots = v_m = 0. \tag{3.1}$$

Then the function $f = f_1 + \cdots + f_m$ is amenable at all points x in some neighborhood of $\bar{x}$ relative to $\operatorname{dom} f = \operatorname{dom} f_1 \cap \cdots \cap \operatorname{dom} f_m$*, with*

$$\begin{aligned} f'_x(\xi) &= (f'_1)_x(\xi) + \cdots + (f'_m)_x(\xi), \\ \partial f(x) &= \partial f_1(x) + \cdots + \partial f_m(x), \\ N(x|\operatorname{dom} f) &= N(x|\operatorname{dom} f_1) + \cdots + N(x|\operatorname{dom} f_m). \end{aligned} \tag{3.2}$$

If each f_i is fully amenable at $\bar{x}$, there is the additional conclusion that f is fully amenable at such neighboring points x, with

$$f''_{x,v}(\xi) = \max_{\substack{v_1+\cdots+v_m=v \\ v_i \in \partial f_i(x)}} \Big\{ (f_1)''_{x,v_1}(\xi) + \cdots + (f_m)''_{x,v_m}(\xi) \Big\} \text{ for all } v \in \partial f(x), \tag{3.3}$$

and in terms of the set $V(x, v, \xi)$ giving the elements $(v_1, \ldots, v_m)$ for which the maximum in this formula is achieved, also

$$(\partial f)'_{x,v}(\xi) = \bigcup_{(v_1,\ldots,v_m) \in V(x,v,\xi)} \Big\{ (\partial f_1)'_{x,v_1}(\xi) + \cdots + (\partial f_m)'_{x,v_m}(\xi) \Big\} \tag{3.4}$$

Proof. By assumption, for $i = 1, \ldots, m$ there exists on a neighborhood V_i of $\bar{x}$ a $\mathcal{C}^1$ mapping $F_i : V_i \to I\!R^{d_i}$ and a lsc, proper, convex function $g_i : I\!R^{d_i} \to \overline{I\!R}$ such that

$$\text{there is no } y_i \neq 0 \text{ in } N\big(F(\bar{x})| \operatorname{dom} g_i\big) \text{ with } \nabla F_i(\bar{x})^* y_i = 0. \tag{3.5}$$

On the neighborhood $V = V_1 \cap \cdots \cap V_m$ of $\bar{x}$ let $F : V \to I\!R^{d_1} \times \cdots \times I\!R^{d_m}$ be given by $F(x) = \big(F_1(x), \ldots, F_m(x)\big)$, and let $g : I\!R^{d_1} \times \cdots \times I\!R^{d_m} \to \overline{I\!R}$ be given by $g(w_1, \ldots, w_m) = g_1(w_1) + \cdots + g_m(w_m)$. Then $g\big(F(x)\big) = f_1(x) + \cdots + f_m(x) = f(x)$. Moreover, F is of class $\mathcal{C}^1$ and g is lower semicontinuous, proper, convex with

$$\begin{aligned} \operatorname{dom} g &= \operatorname{dom} g_1 \times \cdots \times \operatorname{dom} g_m, \\ \partial g(w) &= \partial g(w_1) \times \cdots \times \partial g_m(w_m), \\ N(w|\operatorname{dom} g) &= N(w_1|\operatorname{dom} g_1) \times \cdots \times N(w_m|\operatorname{dom} g_m) \end{aligned}$$

(these expressions for normal cones and subgradients being immediate in the context of convex analysis). Due to the product form of $N(w|\operatorname{dom} g)$ and the block-diagonal structure of the Jacobian $\nabla F(\bar{x})$, the fact that (3.5) holds for every i translates into the constraint qualification (3.1). Thus, f is amenable.

The same reasoning when the f_i's are fully amenable establishes that f is fully amenable. In that case the mapping F is $\mathcal{C}^2$ because each F_i is $\mathcal{C}^2$, and the function g is piecewise linear-quadratic because each g_i is piecewise linear-quadratic.

Applying the formulas in (2.6) relative to our local representation, we obtain on the first-order level that

$$\begin{aligned} \partial f(\bar{x}) &= \nabla F_1(\bar{x})^* \partial g_1\big(F_1(\bar{x})\big) + \cdots + \nabla F_m(\bar{x})^* \partial g_m\big(F_m(\bar{x})\big), \\ N(\bar{x}|\operatorname{dom} f) &= \nabla F_1(\bar{x})^* N\big(F_1(\bar{x})|\operatorname{dom} g_1\big) + \cdots + \nabla F_m(\bar{x})^* N\big(F_m(\bar{x})|\operatorname{dom} g_m\big), \\ f'_{\bar{x}}(\xi) &= (g_1)'_{F_1(\bar{x})}\big(\nabla F_1(\bar{x})\xi\big) + \cdots + (g_m)'_{F_m(\bar{x})}\big(\nabla F_m(\bar{x})\xi\big), \end{aligned}$$

where by these same formulas (2.6) as applied to the individual f_i's we have

$$\begin{aligned} \partial f_i(\bar{x}) &= \nabla F_i(\bar{x})^* \partial g_i\big(F_i(\bar{x})\big), \\ N(\bar{x}|\operatorname{dom} f_i) &= \nabla F_i(\bar{x})^* N\big(F_i(\bar{x})|\operatorname{dom} g_i\big), \\ (f_i)'_{\bar{x}}(\xi) &= (g_i)'_{F_i(\bar{x})}\big(\nabla F_i(\bar{x})\xi\big). \end{aligned} \tag{3.6}$$

If the convex functions g_i are piecewise linear-quadratic, their second-order epi-derivatives are expressed by (2.7), from which it is evident that for $u = (u_1, \cdots, u_m)$

$$g''_u(\omega) = (g_1)''_{u_1}(\omega_1) + \cdots + (g_m)''_{u_m}(\omega_m).$$

Because $\langle y, F\rangle = \langle y_1, F_1\rangle + \cdots + \langle y_m, F_m\rangle$ for $y = (y_1, \ldots, y_m) \in I\!R^{d_1} \times \cdots \times I\!R^{d_m}$, it is also clear that

$$\Big\langle \xi, \nabla^2\langle y, F\rangle(\bar{x})\xi \Big\rangle = \Big\langle \xi, \nabla^2\langle y_1, F_1\rangle(\bar{x})\xi \Big\rangle + \cdots + \Big\langle \xi, \nabla^2\langle y_m, F_m\rangle(\bar{x})\xi \Big\rangle.$$

We therefore deduce from (2.6) on the second-order level, relative to full amenability, that

$$f''_{\bar{x},v}(\xi) = \max_{y \in U(\bar{x},v)} \sum_{i=1}^{m} \Big\{ (g_i)''_{F_i(\bar{x}),y_i}\big(\nabla F_i(\bar{x})\xi\big) + \Big\langle \xi, \nabla^2\langle y_i, F_i\rangle(\bar{x})\xi \Big\rangle \Big\}, \tag{3.7}$$

where $y \in U(\bar{x}, v)$ if and only if $y_i \in \partial g_i\big(F_i(\bar{x})\big)$ for $i = 1, \ldots, m$ and $\nabla F_1(\bar{x})^* y_1 + \cdots + \nabla F_m(\bar{x})^* y_m = v$. At the same time we have

$$(f_i)''_{\bar{x},v_i}(\xi) = (g_i)''_{F_i(\bar{x}),y_i}\big(\nabla F_i(\bar{x})\xi\big) + \Big\langle \xi, \nabla^2\langle y_i, F_i\rangle(\bar{x})\xi \Big\rangle$$

by (2.6). Thus, (3.7) agrees with (3.3).

To prove (3.4), recall that for a fully amenable function the set of subgradients is polyhedral. Therefore for some finite index set J and $v_i^j \in \partial f_i(\bar{x})$ we have

$$f''_{\bar{x},v}(\xi) = \max_{j \in J} \Big\{ (f_1)''_{\bar{x},v_1^j}(\xi) + \cdots + (f_m)''_{\bar{x},v_m^j}(\xi) \Big\}. \tag{3.8}$$

Recall further that the second-order epi-derivative of an amenable function is the sum of a piecewise linear-quadratic convex function with a quadratic. Therefore we easily have that

$$\big(f''_{\bar{x},v}\big)'_{\xi}(\xi') = \max_{j\in J(\xi)} \Big\{ \big((f_1)''_{\bar{x},v_1^j}\big)'_{\xi}(\xi') + \cdots + \big((f_m)''_{\bar{x},v_m^j}\big)'_{\xi}(\xi') \Big\}, \tag{3.9}$$

where $J(\xi)$ is the set of indices where the maximum is attained. Because the directional derivatives appearing on the right hand side of (3.9) are lower semicontinuous (as functions of ξ') we deduce that the left hand side of (3.9) is also lower semicontinuous. From this is follows that $\big(f''_{\bar{x},v}\big)'_{\xi}(\cdot)$ is the support function of $\partial f''_{\bar{x},v}(\xi)$, and that

$$\partial f''_{\bar{x},v}(\xi) = \bigcup_{j\in J(\xi)} \Big\{ \partial(f_1)''_{\bar{x},v_1^j}(\xi) + \cdots + \partial(f_m)''_{\bar{x},v_m^j}(\xi) \Big\}.$$

To complete the proof of (3.4) simply invoke formula (2.5).

To complete the proof, it is necessary only to demonstrate, from our assumptions, that condition (3.1) holds not just for $\bar{x}$, but for all x in some neighborhood of $\bar{x}$ relative to $\operatorname{dom} f$. Then not only will f be amenable (or, as the case may be, fully amenable) at such neighboring points, which we already could conclude from Theorem 2.11, but the differentiation formulas we have established at $\bar{x}$ will be valid at those points x as well.

Consider a sequence of points $x^\nu \in \operatorname{dom} f$ with $x^\nu \to \bar{x}$, and suppose that (3.1) (with $\bar{x}$ replaced by x^ν) is not satisfied at any of these points. It must be verified that this hypothesis leads to a contradiction with our knowledge that (3.1) holds at $\bar{x}$. For each index ν we have the existence of vectors $v_i^\nu \in N(x^\nu|\operatorname{dom} f_i)$, not all 0, such that $v_1^\nu + \cdots + v_m^\nu = 0$. This property of the vectors v_i^ν is retained if they are rescaled by a common factor λ^ν for each ν. Without loss of generality, therefore, we can assume that $|v_1^\nu| + \cdots + |v_m^\nu| = 1$ for all ν. Then, by passing to subsequences if necessary, we can suppose that v_i^ν converges for each i to a certain v_i. Clearly $|v_1| + \cdots + |v_m| = 1$, so the vectors v_i are not all 0. We must prove that $v_i \in N(\bar{x}|\operatorname{dom} f_i)$, however.

Fixing i and returning to the representation of $f_i = g_i \circ F_i$ that we utilized earlier, we invoke Theorem 2.11 in recalling that f_i is amenable also at points near to $\bar{x}$ within $\operatorname{dom} f_i$, hence at the points x^ν for ν sufficiently large. The relations in (3.6) therefore hold at such points x^ν as well as at $\bar{x}$. This gives us vectors $y_i^\nu \in N\big(F_i(x^\nu)|\operatorname{dom} g_i\big)$ such that $\nabla F_i(x^\nu)^* y_i^\nu = v_i^\nu$. If these vectors y_i^ν formed an unbounded sequence, we could obtain by passing to a subsequence corresponding to ν in a certain index set N that $0 < |y_i^\nu| \xrightarrow[\nu\in N]{} \infty$ and $\bar{y}_i^\nu := y_i^\nu/|y_i^\nu| \xrightarrow[\nu\in N]{} \bar{y}_i \neq 0$. Since $\nabla F_i(x^\nu)^* \bar{y}_i^\nu = v_i^\nu/|y_i^\nu| \xrightarrow[\nu\in N]{} 0$ we would get $\nabla F_i(\bar{x})^* \bar{y}_i = 0$, yet from the fact that $\bar{y}_i^\nu \in N\big(F_i(x^\nu)|\operatorname{dom} g_i\big)$ with $F_i(x^\nu) \to F_i(\bar{x}) \in \operatorname{dom} g_i$, we would have $\bar{y}_i \in N\big(F_i(\bar{x})|\operatorname{dom} g_i\big)$, inasmuch as g_i is convex. The existence of such a vector $\bar{y}_i$ would be contrary to the constraint qualification assumed for the representation $f_i = g_i \circ F_i$ at $\bar{x}$. It follows that the sequence of vectors y_i^ν must be bounded. A subsequence must converge then to some y_i. By parallel reasoning we are able to conclude that $y_i \in N\big(F(\bar{x})|\operatorname{dom} g_i\big)$ and $\nabla F_i(\bar{x})^* \bar{y}_i = v_i$. This proves by way of (3.6) that $v_i \in N(\bar{x}|\operatorname{dom} f_i)$, as required. □

Corollary 3.2. *Let $\bar{x} \in C = \bigcap_{i=1}^m C_i$, where the sets C_i are all amenable at $\bar{x}$, and the constraint qualification is satisfied that if $\sum_{i=1}^m v_i = 0$ with $v_i \in N(\bar{x}|C_i)$, then $v_i = 0$ for all i. Then C is amenable at $\bar{x}$ with*

$$N(\bar{x}|C) = \sum\nolimits_{i=1}^m N(\bar{x}|C_i), \qquad T(\bar{x}|C) = \bigcap\nolimits_{i=1}^m T(\bar{x}|C_i).$$

If each C_i is fully amenable at $\bar{x}$, then C is fully amenable at $\bar{x}$ as well.

The domain condition (3.1) in Theorem 3.1 reduces in the case of $f = f_1 + f_2$ to

$$N(x|\operatorname{dom} f_1) \cap -N(x|\operatorname{dom} f_2) = \{0\}. \tag{3.10}$$

Inasmuch as the cones $N(x|\operatorname{dom} f_i)$ are closed and convex, this relation can be written in dual form as

$$T(x|\operatorname{dom} f_1) - T(x|\operatorname{dom} f_2) = I\!R^n, \tag{3.11}$$

where $T(x|\operatorname{dom} f_i) = N(x|\operatorname{dom} f_i)^*$ (polar cone). The tangent cone condition in (3.11) is the kind of condition that has been used in the study of Clarke regularity by Ward and Borwein [18]. For convex functions f_i (or more generally, functions f_i for which $\operatorname{dom} f_i$ is a convex set), (3.10) and (3.11) are equivalent to the relative interior condition

$$\operatorname{ri}(\operatorname{dom} f_1) \cap \operatorname{ri}(\operatorname{dom} f_2) \neq \emptyset, \ \text{ with } \ \dim(\operatorname{dom} f_1) + \dim(\operatorname{dom} f_2) = n.$$

This is the condition commonly invoked when calculating subgradients in convex analysis (cf. [17, Theorem 23.8]), except that the dimensionality restriction is superfluous in that context.

Condition (3.10) is trivially satisfied, for instance, when $\bar{x}$ belongs to the interior of either $\operatorname{dom} f_1$ or $\operatorname{dom} f_2$. We record some common instances.

Corollary 3.3. *Suppose $f = f_1 + f_2$ for a function $f_1 : I\!R^n \to \overline{I\!R}$ that happens to be amenable at $\bar{x}$ and a $\mathcal{C}^1$ function $f_2 : I\!R^n \to I\!R$. Then f is amenable at all points x in some neighborhood of $\bar{x}$ relative to $\operatorname{dom} f = \operatorname{dom} f_1$, with*

$$\partial f(x) = \partial f_1(x) + \nabla f_2(x), \qquad f'_x(\xi) = (f_1)'_x(\xi) + \big\langle \nabla f_2(x), \xi \big\rangle.$$

If f_1 is fully amenable at $\bar{x}$ and f_2 is $\mathcal{C}^2$, then f is fully amenable at all such neighboring points x, and for each $v \in \partial f(x)$ one has

$$\begin{aligned} f''_{x,v}(\xi) &= (f_1)''_{x,v_1}(\xi) + \big\langle \xi, \nabla^2 f_2(x)\xi \big\rangle, \\ (\partial f)'_{x,v}(\xi) &= (\partial f_1)'_{x,v_1}(\xi) + \nabla^2 f_2(x)\xi, \quad \textit{where } v_1 = v - \nabla f_2(x). \end{aligned}$$

The second derivative formula in Corollary 3.2 is covered also by a result in Rockafellar [1, Proposition 2.10] which does not assume full amenability but merely the twice epi-differentiability of f_1 at x relative to each $v_1 \in \partial f_1(x)$.

Corollary 3.4. *Suppose $f = f_0 + \delta_C$ for a finite function $f_0 : I\!R^n \to I\!R$ and a set $C \subset I\!R^n$. If f_0 and C are amenable at a point $\bar{x} \in C$, then f is amenable at every point x in some neighborhood of $\bar{x}$ relative to C, with*

$$\partial f(x) = \partial f_0(x) + N(x|C), \qquad f'_x(\xi) = (f_0)'_x(\xi) + \delta_{T(x|C)}(\xi). \tag{3.12}$$

If f_0 and C are fully amenable at $\bar{x}$, there is the additional conclusion that f is fully amenable at all such neighboring points x with

$$f''_{x,v}(\xi) = \max_{\substack{v_0+v_1=v \\ v_0\in\partial f_0(x),\, v_1\in N(x|C)}} \Big\{ (f_0)''_{x,v_0}(\xi) + (\delta_C)''_{x,v_1}(\xi) \Big\} \text{ for all } v \in \partial f(x). \tag{3.13}$$

When f_0 happens to be differentiable at x, this reduces to

$$f''_{x,v}(\xi) = (f_0)''_{x,v_0}(\xi) + (\delta_C)''_{x,v_1}(\xi) \text{ with } v_0 = \nabla f_0(x),\ v_1 = v - v_0 \in N(x|C).$$

We move on now to a general chain rule. Although amenability has the operation of composition already built into its definition, such a rule still has significant content because it avoids the necessity in every application of having to revert to a composite representation in which the "outer" function g is convex.

Theorem 3.5 (chain rule)**.** *Suppose $f(x) = g\big(F(x)\big)$ for a C^1 mapping $F : I\!R^n \to I\!R^d$ and a function $g : I\!R^d \to \overline{I\!R}$. Let $\bar{x}$ be a point such that g is amenable at $F(\bar{x})$ and*

$$\text{there is no } y \neq 0 \text{ in } N\big(F(\bar{x})|\operatorname{dom} g\big) \text{ with } \nabla F(\bar{x})^* y = 0. \tag{3.14}$$

Then f is amenable at all points x in some neighborhood of $\bar{x}$ relative to $\operatorname{dom} f$, with

$$\begin{aligned} \partial f(x) &= \nabla F(x)^* \partial g\big(F(x)\big), \\ N(x|\operatorname{dom} f) &= \nabla F(x)^* N\big(F(x)|\operatorname{dom} g\big), \\ f'_x(\xi) &= g'_{F(x)}\big(\nabla F(x)\xi\big). \end{aligned}$$

If g is fully amenable at $F(\bar{x})$ and F is a $\mathcal{C}^2$ mapping, f is fully amenable at all such neighboring points x, with

$$f''_{x,v}(\xi) = \max_{\substack{y\in\partial g(F(x)) \\ \nabla F(x)^* y = v}} \Big\{ g''_{F(x),y}\big(\nabla F(x)\xi\big) + \Big\langle \xi, \nabla^2\langle y, F\rangle(x)\xi \Big\rangle \Big\} \tag{3.15}$$

and, in terms of the set $M(x,v,\xi)$ of vectors y achieving the maximum in this formula, also

$$(\partial f)'_{x,v}(\xi) = \bigcup_{y\in M(x,v,\xi)} \Big\{ \nabla F(x)^*(\partial g)'_{F(x),y}\big(\nabla F(x)\xi\big) + \nabla^2\langle y, F\rangle(x)\xi \Big\}. \tag{3.16}$$

Proof. From the hypothesis there is a local representation $g(w) = h\big(G(w)\big)$ in a neighborhood of $\bar{w} := F(\bar{x})$, where G is a $\mathcal{C}^1$ mapping, h is a lower semicontinuous, proper, convex function, and

$$\text{there is no } z \neq 0 \text{ in } N\big(G(\bar{w})|\,\mathrm{dom}\, h\big) \text{ with } \nabla G(\bar{w})^* z = 0. \tag{3.17}$$

For this we have through specialization of (2.6) the formulas

$$\begin{aligned} \partial g(\bar{w}) &= \nabla G(\bar{w})^* \partial h\big(G(\bar{w})\big), \\ N(\bar{w}|\,\mathrm{dom}\, g) &= \nabla G(\bar{w})^* N\big(G(\bar{w})|\,\mathrm{dom}\, h\big), \\ g'_{\bar{w}}(\omega) &= h'_{G(\bar{w})}\big(\nabla G(\bar{w})\omega\big). \end{aligned} \tag{3.18}$$

The key is to consider the local representation $f(x) = h\big(H(x)\big)$ of f, where $H = G \circ F$ and $\nabla H(\bar{x}) = \nabla G(\bar{w}) \nabla F(\bar{x})$. We must check that this representation satisfies the constraint qualification associated with amenability. Suppose $z \in N\big(H(\bar{x})|\,\mathrm{dom}\, h\big)$ and $\nabla H(\bar{x})^* z = 0$. Let $y = \nabla G(\bar{w})^* z$. We have $\nabla F(\bar{x})^* y = 0$ by the product form of $\nabla H(\bar{x})$, but also $y \in N(\bar{w}|\,\mathrm{dom}\, g)$ by the middle formula in (3.18). Our assumption (3.14) implies that $y = 0$. But then $z = 0$ by (3.17).

The representation $f = h \circ H$ fits the original pattern in the definition of amenability and confirms that property for f at $\bar{x}$. It further allows us to invoke the first-order formulas in (2.6) with the appropriate shift of notation:

$$\begin{aligned} \partial f(\bar{x}) &= \nabla H(\bar{x})^* \partial h\big(H(\bar{x})\big), \\ N(\bar{x}|\,\mathrm{dom}\, f) &= \nabla H(\bar{x})^* N\big(H(\bar{x})|\,\mathrm{dom}\, h\big), \\ f'_{\bar{x}}(\xi) &= h'_{H(\bar{x})}\big(\nabla H(\bar{x})\xi\big). \end{aligned} \tag{3.19}$$

These formulas, in combination with the ones in (3.18), immediately yield the first-order formulas asserted in the theorem, at least at the point $\bar{x}$.

When F is $\mathcal{C}^2$ and g is fully amenable at $\bar{w} = F(\bar{x})$, we can choose G to be $\mathcal{C}^2$ and h to be piecewise linear-quadratic, verifying from the representation $f = h \circ H$ that f is fully amenable at $\bar{x}$. In very much the same way we then obtain the second-order formula in the theorem at $\bar{x}$. We have

$$g''_{\bar{w},y}(\omega) = \max_{\substack{z \in \partial h\big(G(\bar{w})\big) \\ \nabla G(\bar{w})^* z = y}} \Big\{ h''_{G(\bar{w}),z}\big(\nabla G(\bar{w})\omega\big) + \Big\langle \omega, \nabla^2 \langle z, G\rangle(\bar{w})\omega \Big\rangle \Big\}, \tag{3.20}$$

but on the other hand

$$f''_{\bar{x},v}(\xi) = \max_{\substack{z \in \partial h\big(H(\bar{x})\big) \\ \nabla H(\bar{x})^* z = v}} \Big\{ h''_{H(\bar{x}),z}\big(\nabla H(\bar{x})\xi\big) + \Big\langle \xi, \nabla^2 \langle z, H\rangle(\bar{x})\xi \Big\rangle \Big\}. \tag{3.21}$$

One calculates by the classical chain rule for smooth functions that

$$\nabla^2\langle z, H\rangle(\bar{x}) = \nabla F(\bar{x})^*\nabla^2\langle z, G\rangle(\bar{w})\nabla F(\bar{x}) + \nabla^2\big\langle \nabla G(\bar{w})^* z, F\big\rangle(\bar{x}).$$

In placing this expression in (3.21) and using (3.20) together with the first-order relations already available, we get the claimed formula (3.15).

We omit the proof of (3.16) because the proof follows an already well established pattern (see the proof of the addition rule) i.e., write (3.15) as the maximum over finitely many points, obtain the directional derivative, take subgradients and finally (in this case) invoke the formula in Proposition 2.10. □

Corollary 3.6. *Suppose $f(x) = g(Ax + a)$ for a linear transformation $A : I\!\!R^n \to I\!\!R^m$, a vector $a \in I\!\!R^m$ and a function $g : I\!\!R^m \to \overline{I\!\!R}$. Let $\bar{x}$ be a point such that g is amenable at $A\bar{x} + a$ and*

$$\text{there is no } y \neq 0 \text{ in } N(A\bar{x} + a \,|\, \operatorname{dom} g) \text{ with } A^* y = 0. \tag{3.20}$$

Then f is amenable at all points x in some neighborhood of $\bar{x}$ relative to $\operatorname{dom} f$, with

$$\begin{aligned} \partial f(x) &= A^*\partial g(Ax + a), \\ N(x|\operatorname{dom} f) &= A^* N(Ax + a|\operatorname{dom} g), \\ f'_x(\xi) &= g'_{Ax+a}(A\xi). \end{aligned}$$

If g is fully amenable at $A\bar{x}$, f is fully amenable at all such neighboring points x, with

$$f''_{x,v}(\xi) = \max_{\substack{y \in \partial g(Ax+a) \\ A^* y = v}} g''_{Ax+a,\, y}(A\xi) \tag{3.23}$$

and, in terms of the set $M(x, v, \xi)$ of all vectors y achieving the maximum in this formula, also

$$(\partial f)'_{x,v}(\xi) = \bigcup_{y \in M(x,v,\xi)} A^*(\partial g)'_{Ax+a,\, y}(A\xi). \tag{3.24}$$

Another direct consequence of Theorem 3.5 is a rule for "partial epi-differentiation."

Corollary 3.7 (partial epi-differentiation)**.** *For a function $f : I\!\!R^d \times I\!\!R^n \to \overline{I\!\!R}$, consider for each $u \in I\!\!R^d$ the function $f_u := f(u, \cdot)$ on $I\!\!R^n$. Suppose f is amenable at $(\bar{u}, \bar{x})$ and*

$$\text{there is no } y \neq 0 \text{ with } (y, 0) \in N\big(\bar{u}, \bar{x} \,\big|\, \operatorname{dom} f\big). \tag{3.25}$$

Then for all pairs (u, x) in some neighborhood of $(\bar{u}, \bar{x})$ relative to $\operatorname{dom} f$, the function f_u is amenable at x with

$$\begin{aligned} (f_u)'_x(\xi) &= f'_{(u,x)}(0, \xi), \\ \partial f_u(x) &= \big\{\, y \,\big|\, \exists v \text{ with } (y, v) \in \partial f(u, x) \,\big\}, \\ N(x|\operatorname{dom} f_u) &= \big\{\, y \,\big|\, \exists v \text{ with } (y, v) \in N(u, x \,|\, \operatorname{dom} f) \,\big\}. \end{aligned} \tag{3.26}$$

If f is fully amenable at $(\bar{u}, \bar{x})$, then f_u is fully amenable at x, with

$$(f_u)''_{x,v}(\xi) = \max_{(y,v)\in\partial f(u,x)} f''_{(u,x),(y,v)}(0,\xi). \tag{3.27}$$

and, in terms of the set $M(x, v, \xi)$ of vectors y achieving the maximum in this formula, also

$$(\partial f_u)'_{x,v}(\xi) = \bigcup_{y\in M(x,v,\xi)} (\partial f)'_{(u,x),(y,v)}(0,\xi). \tag{3.28}$$

Proof. Focusing first on $u = \bar{u}$, consider $f_{\bar{u}}$ to be the composition $f\circ F$ where $F(x) = (\bar{u}, x)$. Since F is affine, its second derivatives all vanish, and Theorem 3.5 gives the desired results. Observe now through Theorem 2.11 (first part) that the condition on $(\bar{u}, \bar{x})$ is inherited by all points (u, x) in some neighborhood of $(\bar{u}, \bar{x})$ relative to $\operatorname{dom} f$. For such (u, x), therefore, the same argument can be applied, and Theorem 3.5 once more gives the formulas claimed. □

Remark: The maximum in the second-order formulas 3.3, 3.13, 3.15, 3.23, and 3.27 may be taken over the set of corresponding extreme points. By doing, as in Proposition 2.10, we then have alternative versions of the proto-derivative formulas 3.4, 3.16, 3.24, and 3.28.

References

1. R. T. Rockafellar, "First- and second-order epi-differentiability in nonlinear programming," *Tarns. Amer. Math. Soc.* **307** (1988), 75–107.
2. R. T. Rockafellar, "Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives," *Math. of Oper. Research* **14** (1989), 462–484.
3. R. T. Rockafellar, "Generalized second derivatives of convex functions and saddle functions," Transactions Amer. Math. Soc. **320** (1990), 810–822.
4. R. T. Rockafellar, "Nonsmooth analysis and parametric optimization," in *Methods of Nonconvex Analysis* (A. Cellina, ed.) Springer-Verlag Lecture Notes in Math. No. 1446 (1990), 137–151.
5. R. Cominetti, "On Pseudo-differentiability," Transactions Amer. Math. Soc., to appear.
6. C. N. Do, "Generalized second derivatives of convex functions in reflexive Banach spaces," Transactions Amer. Math. Soc., to appear.
7. R. A. Poliquin, "Proto-differentiation of subgradient set-valued mappings," Canadian J. Math **42** (1990), 520–532.

8. R. A. Poliquin, "An extension of Attouch's Theorem and its application to second-order epi-differentiation of convex composite functions," preprint.

9. R. A. Poliquin and R. T. Rockafellar, "Differentiability of solution mappings and sensitivity analysis in parametric optimization," preprint.

10. A. D. Ioffe, "Variational analysis of a composite function: a formula for the lower second order directional derivative," J. Math. Anal. Appl., forthcoming.

11. S. M. Robinson, "Some continuity properties of polyhedral multifunctions," Math. Programming Study **14** (1981), 206–214.

12. J. Sun, *On Monotropic Piecewise quadratic Programming*, dissertation, Department of Applied Math., University of Washington, August, 1986.

13. F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, 1983. Republished by the Centre de Recherches Mathématiques, Université de Montréal (C.P. 6128 "A," Montréal, Québec, Canada, H3C 3J7).

14. F. H. Clarke, *Methods of Dynamic and Nonsmooth Optimization*, CBMS-NSF Regional Conference Series in Applied Mathematics, **57**, SIAM Publications, Philadelphia, PA, 1989.

15. R. T. Rockafellar, "Dualization of subgradient conditions for optimality."

16. R. T. Rockafellar, "Proto-differentiability of set-valued mappings and its applications in optimization," in *Analyse Non Linéaire*, H. Attouch et al. (eds.), Gauthier-Villars, Paris (1989), 449–482.

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

18. D. E. Ward and J. M. Borwein, "Nonsmooth calculus in finite dimensions," SIAM J. Control Opt. **25** (1987), 1312–1340.

19. F. H. Clarke, "Generalized gradients and applications," Transactions Amer. Math. Math. Soc. **205** (1975), 247–262.

20. R. T. Rockafellar, "Directionally Lipschitzian functions and subdifferential calculus," Proc. London Math. Soc. **39** (1979), 331–355.

21. J.-B. Hiriart-Urruty, "Calculus rules on the approximate second order directional derivative of a convex function," SIAM J. Control Opt. **22** (1984), 381–404.

22. J.-B. Hiriart-Urruty and A. Seeger, "Calculus rules on a new set-valued second order derivative for convex functions," Nonlinear Analysis Th. Meth. Appl. **13** (1989), 721–738.

23. R. Cominetti and R. Correa, "A generalized second order derivative in nonsmooth optimization," SIAM J. Control Opt. **28** (1990), 789–809.

24. A. D. Ioffe, “Variational analysis of a composite function: a formula for the lower second order epi-derivative.”

25. L. McLinden and R.C. Bergstrom, “Preservation of convergence of convex sets and functions,” Transactions Amer. Math. Soc. **268** (1981), 127–142.