SECOND-ORDER CONVEX ANALYSIS

R. Tyrrell Rockafellar*

Abstract. The classical theorem of Alexandrov asserts that a finite convex function has a second-order Taylor expansion almost everywhere, even though its first partial derivatives may only exist almost everywhere. A theorem of Mignot provides a generic linearization of the subgradient mapping associated with such a function but leaves open the question of symmetry of the matrix that appears in this linearization. This paper clears up the gap between these results and goes on to a broader theory of second-order semi-derivatives of a convex function in relation to first-order semi-derivatives of its subgradient mapping. Connections with generalized derivatives based on approximations utilizing variational convergence are illuminated as well.

Keywords: Convex analysis, generalized second-derivatives, monotone mappings

AMS Subject Classification: 26B25, 49J52, 58C20

* Department of Mathematics, Box 354350, University of Washington, Seattle, WA 98195-4350. This work was supported in part by the National Science Foundation under grant DMS–9803089.

1. Introduction

In first-order convex analysis, a central notion is that of "subgradient." Associated with any closed, proper, convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is the set-valued mapping $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, which gives for each x the subgradient set $\partial f(x) = \{ v \mid f(x') \geq f(x) + \langle v, x' - x \rangle \}$. This mapping is known to be maximal monotone and to determine f uniquely up to an additive constant. Its effective domain $\operatorname{dom} \partial f$, although not necessarily convex, has the same relative interior and closure as $\operatorname{dom} f = \{ x \mid f(x) < \infty \}$.

For purposes of second-order convex analysis, both f and ∂f offer possibilities. On the one hand, second-order approximations of f can be explored, but first-order approximations of ∂f also deserve consideration in accordance with the notion of second derivatives being obtainable by differentiating first derivatives. A basic issue, however, is what should be meant by "approximation." That term should refer of course to some concept of nearness, but nearness in the traditional sense of locally uniform pointwise convergence of functions isn't appropriate when the functions can be discontinuous and take on ∞.

In this paper, we look at approximations of both f and ∂f. We start with the reconciliation of some early results on Taylor-like expansions, which are based on uniform convergence, and proceed to extensions in which the approximating expressions come from directional derivatives that are merely one-sided and correspond to semi-differentiation. Then we go on to approximations based instead on set convergence, applied graphically and epi-graphically, in order to gain further insights and connections with duality. These approximations, in terms of epi-derivatives and proto-derivatives, provide criteria for semi-derivative expansions in particular.

A general reference for second-order nonsmooth analysis of possibly nonconvex functions is the recent book [1], Chapter 13. Our aim here is to bring out special properties relevant to that theory that hold under convexity. Some of these properties can be extracted from the broader picture, but others follow a separate track or, in taking advantage of convexity, can rely on much simpler arguments. We also aim at using this setting to trace the motivation for some of the ideas that have come to dominate second-order nonsmooth analysis. Much of that motivation came from convexity, even though the convex analysis book [2] developed no second-order theory at all, and indeed with only the exception of Alexandrov's theorem on quadratic expansions, little was known in that direction when [2] was written.

2. Second derivatives based on Taylor-like expansions

A well known theorem of Rademacher asserts that a locally Lipschitz continuous mapping from an open subset O of $\mathbb{R}^n$ to $\mathbb{R}^d$ for some $d \geq 1$ is differentiable almost everywhere. This can be applied to convex functions because they are locally Lipschitz continuous on sets where they are finite. In our context of a closed, proper, convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ (which is adopted throughout this paper) we obtain the fact that at almost every point $x \in \operatorname{int} \operatorname{dom} f$ there is a first-order linear (i.e., affine) expansion

$$f(x+w) = f(x) + \langle v, w \rangle + o(|w|), \tag{2.1}$$

where $|w|$ denotes the Euclidean norm and $o(t)$ is the notation for an error term such that $o(t)/t \to 0$ as $t \to 0$.

The existence of the expansion (2.1) is the very definition of f being differentiable at x and describes the circumstances in which the gradient $\nabla f(x)$ exists, this being the vector v. Such points x thus form the set $\operatorname{dom} \nabla f$, which lies within $\operatorname{int} \operatorname{dom} f$ and differs from it only by a negligible set.

Convexity allows us to go further than these generic first-order expansions. The following theorem of Alexandrov [3] from 1939 stands as the primary classical fact in second-order convex analysis. A corresponding geometric result in the language of convex surfaces was obtained by Busemann and Feller [4] in 1936.

Theorem 2.1 (Alexandrov)**.** *At almost every point $x \in \operatorname{int} \operatorname{dom} f$ there is a second-order quadratic expansion in the form*

$$f(x+w) = f(x) + \langle v, w \rangle + \tfrac{1}{2}\langle Aw, w \rangle + o(|w|^2). \tag{2.2}$$

In particular (2.2) implies (2.1) and ensures that $v = \nabla f(x)$, but the status of the matrix A is less clear. The quadratic form $\langle Aw, w \rangle$ depends only on the symmetric part of A, i.e. the matrix $\frac{1}{2}(A + A^*)$ (with * denoting transpose), so there's no loss of generality in taking A itself symmetric in (2.2). Whether the entries of A can be interpreted as second partial derivatives of f is nevertheless not so easy to answer.

Ordinarily, any discussion of second partial derivatives of f with respect to the components of $x = (x_1, \ldots, x_n)$ presupposes that the first partial derivatives $(\partial f/\partial x_j)(x_1, \ldots, x_n)$ exist locally. Here we can be sure of their existence almost everywhere in $\operatorname{int} \operatorname{dom} f$, namely at the points $x \in \operatorname{dom} \nabla f$, but not everywhere. We do everywhere have subgradients, i.e., the nonemptiness of $\partial f(x)$, and we know from convex analysis [2; Theorem 25.1] that $\partial f(x)$ reduces to a single vector v if and only if f is differentiable at x with $\nabla f(x) = v$.

A theorem of Mignot [5] from 1976 provides in this respect an interesting parallel to Alexandrov's theorem and also a challenge. Mignot's result is valid for maximal monotone mappings $T : I\!\!R^n \rightrightarrows I\!\!R^n$ in general, but we state it now only for $T = \partial f$. In doing so, we use $I\!\!B$ to denote the closed unit ball in $I\!\!R^n$ (with respect to the Euclidean norm).

Theorem 2.2 (Mignot)**.** *At almost every point $x \in \operatorname{int}\operatorname{dom}\partial f$ there is a first-order linear expansion in the form*

$$\partial f(x+w) \;\subset\; v + Aw + o(|w|)I\!\!B. \tag{2.3}$$

The inclusion in (2.3), along with the fact that $\partial f(x+w) \neq \emptyset$ when w is sufficiently small, implies that $x \in \operatorname{dom}\nabla f$ with $v = \nabla f(x)$ and allows us to think of (2.3) as defining the *differentiability* of ∂f at x, even though ∂f is a generally set-valued mapping. The matrix A is in this sense the *Jacobian* of ∂f at x. The approximation that is afforded fits the usual notion of differentiability when restricted to the set of points where ∂f is single-valued, i.e., to $\operatorname{dom}\nabla f$.

Theorem 2.3. *The points $x \in \operatorname{int}\operatorname{dom} f$ at which the expansion (2.3) holds are the points $x \in \operatorname{dom}\nabla f$ at which the expansion*

$$\nabla f(x+w) \;=\; \nabla f(x) + Aw + o(|w|) \tag{2.4}$$

holds with respect to $\{w \,|\, x+w \in \operatorname{dom}\nabla f\}$, this being a neighborhood of $w=0$ except for the possible omission of a negligible set of points $w \neq 0$.

Proof. We get (2.4) from (2.3) simply by restricting to the points $x+w$ where ∂f is single-valued. For the converse derivation of (2.3) from (2.4), a characterization of ∂f in terms of ∇f comes into play. According to [2; Theorem 25.6], we have at any point $x' \in \operatorname{int}\operatorname{dom} f$ that $\partial f(x')$ is the convex hull of the compact set $\overline{\nabla} f(x')$, consisting of all cluster points of sequences $\{\nabla f(x^\nu)\}_{\nu=1}^\infty$ at points $x^\nu \in \operatorname{dom}\nabla f$ such that $x^\nu \to x'$. Through this we get from (2.4), written in the form $\nabla f(x+w) \in \nabla f(x) + Aw + o(|w|)I\!\!B$, that the inclusion $\overline{\nabla} f(x+w) \subset \nabla f(x) + Aw + o(|w|)I\!\!B$ without any need for restricting $x+w$ to $\operatorname{dom}\nabla f$. Since the right side of this inclusion is a convex set, it remains valid when the convex hull is taken on the left, an operation which yields (2.3). $\square$

It's natural in the presence of the expansion property of Theorem 2.3 to define ∇f to be *differentiable at x in the extended sense* and f to be *twice differentiable at x in the extended sense*. The entries of the matrix A in (2.4) can legitimately be regarded then as the *second partial derivatives* of f at x, even though first partial derivatives might fail to exist at a certain points $x+w$ near x. We'll refer to A therefore as the *Hessian* of f at x

and employ for it the notation

$$\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, \ldots, x_n)\right]_{i,j=1}^{n,n}$$

just as in the case of f being twice differentiable at x in the classical sense, where $\operatorname{dom} \nabla f$ is fully a neighborhood of x. In such terms we obtain statements about second-order differentiability, first from Theorem 2.3 and then from Theorem 2.2.

Corollary 2.4. *The subgradient mapping ∂f is differentiable at a point $x \in \operatorname{int} \operatorname{dom} f$ if and only if the gradient mapping ∇f is differentiable at x in the extended sense, or in other words, f is twice differentiable at x in the extended sense. The Hessian matrix $\nabla^2 f(x)$ serves then as the Jacobian matrix for ∂f at x, with*

$$\partial f(x+w) \subset \nabla f(x) + \nabla^2 f(x) w + o(|w|) I\!\!B. \tag{2.5}$$

Corollary 2.5. *The function f is twice differentiable in the extended sense at almost every point $x \in \operatorname{int} \operatorname{dom} f$, those points forming a subset of $\operatorname{dom} \nabla f$.*

Corollary 2.5, as a reincarnation of Theorem 2.2, is tantalizing in its similarity to Theorem 2.1. In comparing Theorem 2.2 to Theorem 2.1, we observe that the open convex sets $\operatorname{int} \operatorname{dom} f$ and $\operatorname{int} \operatorname{dom} \partial f$ are identical and that in both expansions we have $v = \nabla f(x)$. Could the points x at which the two expansions exist likewise be the same, and could it be true that the matrix A is the same in both cases, thus equaling the $\nabla^2 f(x)$? A serious hurdle is that we have no assurance in (2.3) and (2.4) of the matrix being symmetric. Classical examples remind us that a twice differentiable function f not of class $\mathcal{C}^2$ can have $\partial^2 f / \partial x_i \partial x_j \neq \partial^2 f / \partial x_j \partial x_i$. How do we know that can't happen even for convex f?

We'll prove, though, that the conjecture is true. For this purpose we'll have to look more closely at difference quotients and their convergence, which will be instructive for subsequent developments as well. To begin with, let's recall that in writing the expansion (2.1) in the equivalent form $f(x + \tau w) = f(x) + \tau \langle v, w \rangle + o(\tau |w|)$ we can interpret it as saying that, as $\tau \searrow 0$ the difference quotient functions

$$\Delta_\tau f(x) : \, w \mapsto \frac{f(x+\tau w) - f(x)}{\tau} \tag{2.6}$$

converge uniformly on bounded subsets of $I\!\!R^n$ to the linear function $w \mapsto \langle v, w \rangle$. Through a similar notational maneuver, the expansion (2.2) in Alexandrov's theorem can be interpreted as saying that, as $\tau \searrow 0$, the second-order difference quotient functions

$$\Delta_\tau^2 f(x) : \, w \mapsto \frac{f(x+\tau w) - f(x) - \tau \langle v, w \rangle}{\frac{1}{2}\tau^2}, \quad \text{where } v = \nabla f(x), \tag{2.7}$$

converge uniformly on bounded subsets of $I\!R^n$ to the quadratic function $w \mapsto \langle Aw, w\rangle$.

Also important in this context for their relationship with the expansion (2.3) in Mignot's theorem are the first-order difference quotient mappings

$$\Delta_\tau[\partial f](x) : w \mapsto \frac{\partial f(x+\tau w) - v}{\tau}, \quad \text{where } v = \nabla f(x). \tag{2.8}$$

These are set-valued, so the issue of their behavior as $\tau \searrow 0$ is a bit more subtle, but still it comes down to uniform convergence, properly construed.

Proposition 2.6. *The expansion (2.3) holds at x if and only if, for every bounded set $W \subset I\!R^n$ and every $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\emptyset \neq \Delta_\tau[\partial f](x)(w) - Aw \subset \varepsilon I\!B \;\; \textit{for all } w \in W \;\textit{ when } \tau \in (0, \delta). \tag{2.9}$$

Proof. This is no more than a careful restatement of (2.3) with w replaced by τw, making explicit the fact that, since $x \in \operatorname{int}\operatorname{dom}\partial f$, we have $\partial f(x+\tau w) \neq$ for all $w \in W$ when τ is sufficiently small. □

Without the nonemptiness on the left in (2.9), it wouldn't be right to speak of uniform convergence being expressed by the expansion in (2.3).

Proposition 2.7. *For any $\tau > 0$, the function $\Delta^2_\tau f(x) : I\!R^n \to \overline{I\!R}$ is closed, proper and convex and nonnegative, while the mapping $\Delta_\tau[\partial f](x) : I\!R^n \rightrightarrows I\!R^n$ is maximal monotone. Moreover*

$$\partial[\tfrac{1}{2}\Delta^2_\tau f(x)] \;=\; \Delta_\tau[\partial f](x). \tag{2.10}$$

Proof. Let $\varphi_\tau = \frac{1}{2}\Delta^2_\tau f(x)$. We have $\varphi_\tau(w) \geq 0$ by virtue of the convexity inequality $f(x+\tau w) \geq f(x) + \tau\langle v, w\rangle$, since $v = \nabla f(x)$. Also, we can write $\varphi_\tau(w) = \tau^{-1}[f_\tau(w) - \langle v, w\rangle]$ for $f_\tau(w) = \tau^{-1}[f(x+\tau w) - f(x)]$. It's obvious that f_τ is again closed, proper and convex, with $\partial f_\tau(w) = \partial f(x+\tau w)$ and consequently that φ_τ is closed, proper and convex with $\partial\varphi_\tau(w) = \tau^{-1}[\partial f(x+\tau w) - v] = \Delta_\tau[\partial f](x)(w)$. As the subgradient mapping $\partial\varphi_\tau$, $\Delta_\tau[\partial f](x)$ is maximal monotone. (The latter also follows from that property of ∂f itself and the defining formula for $\Delta_\tau[\partial f](x)$.) □

The difference quotient relationship in Proposition 2.7 provides the bridge we are seeking between Theorems 2.1 and 2.2.

Theorem 2.8. *The points x for which the expansion (2.2) holds in Alexandrov's theorem are the same as the ones for which the expansion (2.3) holds in Mignot's theorem. Moreover* *the matrix A in (2.3) is always symmetric when it exists and can be identified with the symmetric matrix A in (2.2). In addition, this matrix is positive semidefinite.*

Proof. Suppose first that the expansion (2.2) holds with A symmetric, and let $\varphi(w) = \frac{1}{2}\langle Aw, w\rangle$, so that φ is differentiable with $\nabla\varphi(w) = Aw$. As already observed, (2.2) means that the functions $\varphi_\tau = \frac{1}{2}\Delta_\tau^2 f(x)$, which by Proposition 2.7 are convex, converge pointwise to φ as $\tau \searrow 0$ and do so uniformly on bounded sets. Convexity is preserved under pointwise convergence, so φ must be convex as well; hence A is positive semidefinite.

Uniform convergence of convex functions entails a kind of uniform convergence of their subgradient mappings; on the basis of [2; Theorem 24.5], one has for every bounded set $W \subset \mathbb{R}^n$ and every $\varepsilon > 0$ the existence of $\delta > 0$ such that

$$\partial\varphi_\tau(w) \subset \partial\varphi(w) + \varepsilon\mathbb{B} \text{ when } \tau \in (0,\delta),\ w \in W,$$

where in addition δ can be chosen small enough that $\partial\varphi_\tau(w) \neq 0$ in these circumstances. Since $\partial\varphi_\tau(w) = \Delta_\tau[\partial f](w)$ by Proposition 2.7, whereas $\partial\varphi(w)$ reduces to Aw, we see we have the property set forth in Proposition 2.6 as describing the expansion (2.3). Thus, (2.2) implies (2.3) with the same symmetric matrix A.

To establish the converse implication, suppose now that the expansion (2.3) holds. Fix any $\rho > 0$ and any $\varepsilon > 0$. We wish to demonstrate the existence of $\delta > 0$ such that

$$\left|\Delta_\tau^2 f(x)(w) - \langle Aw, w\rangle\right| \leq \varepsilon \text{ when } \tau \in (0,\delta),\ |w| \leq \rho, \tag{2.11}$$

in order to confirm that, as $\delta \searrow 0$, the functions $\Delta_\tau^2 f(x)$ converge , uniformly on all bounded sets to the function $w \mapsto \langle Aw, w\rangle$.

By applying Proposition 2.6 to $W = \rho\mathbb{B}$ and $\varepsilon' = \varepsilon/\rho$ and restricting ∂f to $\operatorname{dom}\nabla f$, we get the existence of $\delta > 0$ such that

$$\begin{aligned} &|\nabla f(x+\tau w) - v - \tau Aw| \leq \tau\varepsilon/\rho \\ &\qquad\qquad \text{when } \tau \in (0,\delta),\ |w| \leq \rho,\ x+\tau w \in \operatorname{dom}\nabla f, \end{aligned} \tag{2.12}$$

where $v = \nabla f(x)$. We can take δ small enough that $x + \delta\rho\mathbb{B} \subset \operatorname{int}\operatorname{dom} f$. Then f is Lipschitz continuous on $x + \delta\rho\mathbb{B}$, almost all points of which must belong to $\operatorname{dom}\nabla f$. It follows through Fubini's theorem that, for all most every w on the boundary of $\rho\mathbb{B}$ (with respect to surface measure) the line segment from x to $x + \delta w$ must have almost all of its points in $\operatorname{dom}\nabla f$ (with respect to linear measure). Hence for almost every $w \in \rho\mathbb{B}$, say $w \in D$, f is differentiable at $x + tw$ for almost all $t \in [0,\delta]$. For such $w \in D$ the function $\psi_w : t \mapsto f(x+tw) - f(x) - t\langle v, w\rangle - \frac{1}{2}t^2\langle Aw, w\rangle$ is Lipschitz continuous on $[0,\delta]$ with

$\psi_w'(t) = \langle \nabla f(x+tw) - v - tAw, w\rangle$ a.e. in t. That implies for $w \in D$ and $\tau \in (0,\delta)$ that $\psi_w(\tau) = \psi_w(0) + \int_0^\tau \psi_w'(t)dt$, where $\psi_w(0) = 0$, and therefore that $|\psi_w(\tau)| \leq \int_0^\tau |\psi_w'(t)|dt$. The estimate in (2.12) turns this into

$$\begin{aligned} &\left| f(x+\tau w) - f(x) - \tau\langle v, w\rangle - \tfrac{1}{2}\tau^2\langle Aw, w\rangle \right| \\ &\qquad\qquad \leq \int_0^\tau \left|\langle \nabla f(x+tw) - v - tAw, w\rangle\right| dt \leq \int_0^\tau \tau\varepsilon d\tau = \tfrac{1}{2}\tau^2\varepsilon, \end{aligned}$$

which on dividing by $\frac{1}{2}\tau^2$ becomes $\left|\Delta_\tau^2 f(x)(w) - \langle Aw, w\rangle\right| \leq \varepsilon$. We have demonstrated that, when $\tau \in (0,\delta)$, this inequality holds for all w in a dense subset D of $\rho I\!B$. It then holds for all $w \in \rho I\!B$ by continuity. Thus, (2.11) has been verified, and the functions $\Delta_\tau^2 f(x)$ converge as claimed. This convergence means that the expansion (2.2) holds.

The version of the expansion (2.2) that we've arrived at uses the matrix A from (2.3), yet it only depends on the symmetric part A_s of A. Our earlier argument told us, though, that if (2.2) holds for A_s, then (2.3) holds for A_s. We thus have (2.2) for both A and A_s. Invoking Theorem 2.3, we find that $[A - A_s]w = o(|w|)$, but that requires $A - A_s = 0$. In other words, A has to have been symmetric and positive semidefinite. □

Corollary 2.9. *The matrix $\nabla^2 f(x)$, whenever it exists through f being twice differentiable in the extended sense, is symmetric and positive semidefinite and furnishes the second-order expansion*

$$f(x+w) = f(x) + \langle \nabla f(x), w\rangle + \tfrac{1}{2}\langle \nabla^2 f(x)w, w\rangle + o(|w|^2). \tag{2.13}$$

3. Generalized expansions based on semi-differentiation

Through Theorem 2.8 and Corollary 2.9 we have a complete and satisfying picture of Taylor-like expansions of finite convex functions on open sets. In first-order convex analysis, however, we know how to get a generalized expansion generalized first-order expansion, not just Taylor-like, in terms of one-sided directional derivatives even at points where f isn't differentiable. At every $x \in \operatorname{int}\operatorname{dom} f$ the limit

$$\lim_{\substack{w' \to w \\ \tau \searrow 0}} \frac{f(x+\tau w') - f(x)}{\tau} = \lim_{\substack{w' \to w \\ \tau \searrow 0}} \Delta_\tau f(x)(w') \tag{3.1}$$

exists finitely for every vector w. In denoting it by $df(x)(w)$ we get

$$f(x+w) = f(x) + df(x)(w) + o(|w|). \tag{3.2}$$

Indeed, the limit in (3.1) exists finitely for every w if and only if the difference quotient functions $\Delta_\tau f(x)$ converge uniformly on bounded subsets of $I\!R^n$ to a function that's finite and continuous; cf. [1; 7.21].

The expansion property in (3.2) is termed the *semi-differentiability* of f at x, with $df(x)$ the corresponding *semi-derivative* function. It reduces to differentiability exactly when $df(x)$ is a linear function, or in other words when $x \in \operatorname{dom} \nabla f$, in which case $df(x)(w) = \langle \nabla f(x), w \rangle$. For points $x \in \operatorname{int} \operatorname{dom} f$ that don't belong to $\operatorname{dom} \nabla f$, although $df(x)$ isn't linear, it's at least *sublinear*, i.e., convex and positively homogeneous. Specifically, $df(x)$ is the support function of the nonempty, compact, convex set $\partial f(x)$:

$$df(x)(w) = \sup\big\{ \langle v, w \rangle \,\big|\, v \in \partial f(x) \big\}.$$

Are there analogs of the generalized first-order expansion (3.2) at the second-order level in convex analysis? Let's extend the definition of the second-order difference quotient $\Delta^2_\tau f(x)(w)$ in (2.7) by substituting $df(x)(w)$ for $\langle v, w \rangle$ there, so as to allow x to be any point in $\operatorname{int} \operatorname{dom} f$, not necessarily in $\operatorname{dom} \nabla f$. The question is whether at such a point x the limit

$$\lim_{\substack{w' \to w \\ \tau \searrow 0}} \frac{f(x + \tau w') - f(x) - df(x)(w')}{\frac{1}{2}\tau^2} \;=\; \lim_{\substack{w' \to w \\ \tau \searrow 0}} \Delta^2_\tau f(x)(w') \tag{3.3}$$

exists finitely for every vector w, or equivalently (in our context of $I\!R^n$) whether, as $\tau \searrow 0$, the functions $\Delta^2_\tau f(x)$ converge uniformly on bounded sets to some finite, continuous function. If that's true, then in denoting the limit by $d^2 f(x)(w)$ we get

$$f(x + w) \;=\; f(x) + df(x)(w) + \tfrac{1}{2} d^2 f(x)(w) + o(|w|^2). \tag{3.4}$$

We speak then of *second-order semi-differentiability* of f at x, with $d^2 f(x)$ being the *second semi-derivative function*. That's certainly present at points x where f is twice differentiable in the extended sense analyzed above, with $d^2 f(x)(w) = \langle \nabla^2 f(x) w, w \rangle$. We would like to understand how far it might hold elsewhere in $\operatorname{int} \operatorname{dom} f$ as well.

A comment about our notation should be made before proceeding, so as not to create a discrepancy with the notation used in [1] without assumptions of convexity. For any function f, $df(x)(w)$ refers there to the value given by the "lim inf" in (3.1). The associated "lim sup" comes out then as $-d[-f](x)(w)$, so the property of semi-differentiability corresponds to having $-d[-f](x)(w) = df(x)(w)$ for all w. This equation necessitates the finiteness of the expressions and their continuity with respect to w, but in general they wouldn't be sublinear in w, just positively homogeneous. Similar conventions govern (3.3). The notation $d^2 f(x)(w)$ refers in [1] to the "lim inf" in (3.3). One says that f is twice semi-differentiable at x if f is (once) semi-differentiable at x and the "lim inf" is a "lim", i.e., the equation $-d^2[-f](x)(w) = d^2 f(x)(w)$ holds for all w. Here we're taking advantage of the fact that, when f is convex and $x \in \operatorname{int} \operatorname{dom} f$, first-order semi-differentiability prevails and can be taken for granted when contemplating second-order semi-differentiability.

A shortcoming of second-order semi-differentiability in the study of a convex function f is that the difference quotients in (3.3) can fail to be convex with respect to w, unless $df(x)(w)$ is actually linear in w. The notion can thereby escape the realm of convex analysis. On the other hand, unless $df(x)(w)$ is linear in w, the ability of f to be twice semi-differentiable at x is much more limited than might be hoped. For instance, a function that's the pointwise maximum of a finite collection of quadratic functions can fail to be twice semi-differentiable at points where the quadratics join together; see [1; 13.10].

In the next section we'll look at a more robust concept than second-order semi-differentiability which gets around these difficulties and even allows treatment of f at boundary points of $\operatorname{dom} f$. For now, we'll concentrate on what can be said about second-order semi-differentiability at points x with linear $df(x)$, where the difference quotients are of the original form in (2.7).

Proposition 3.1. *At any point $x \in \operatorname{dom} \nabla f$ where f is twice semi-differentiable, the function $d^2 f(x)$ is (finite) convex, nonnegative and positively homogeneous of degree two.*

Proof. The functions $\Delta^2_\tau f(x)$ for $\tau > 0$ are convex and nonnegative by Proposition 2.7, and these properties are preserved when they converge pointwise to another function, namely $d^2 f(x)$. It's easy to see from the formula in (2.7) that, for $\lambda > 0$, one has $\Delta^2_\tau f(x)(\lambda w) = \lambda^2 \Delta^2_{\lambda\tau} f(x)(w)$. In the limit, this gives $d^2 f(x)(\lambda w) = \lambda^2 d^2 f(x)(w)$, which is positive homogeneity of degree two. □

Corollary 3.2. *At any point $x \in \operatorname{dom} \nabla f$ where f is twice semi-differentiable, there is a closed convex set C with $0 \in \operatorname{int} C$ such that*

$$d^2 f(x) = \gamma_C^2 \quad \textit{for the gauge } \gamma_C. \tag{3.5}$$

Proof. A convex function is nonnegative and positively homogeneous of degree two if and only if its square root is nonnegative and positively homogeneous of degree one. The closed convex functions of the latter sort are the gauges γ_C of the closed convex sets C containing 0; cf. [1; 3.50]. Finiteness of γ_C corresponds to 0 being an interior point of C. □

When f is twice differentiable at x in the extended sense, not merely twice semi-differentiable there, the set C in Corollary 3.2 comes out as the (possibly degenerate) ellipsoid $\{w \mid \langle \nabla^2 f(x) w, w \rangle \le 1\}$. In the broader setting it corresponds to a kind of second-order subdifferential which has been studied by Hiriart-Urruty [6], [7]. In contrast to other second-derivative concepts, this one has no known extension to nonconvex functions.

Our aim now will be to develop a counterpart to Theorem 2.8, and for that we need an extension of semi-differentiability to ∂f. We'll say that ∂f is *semi-differentiable* at a point $x \in \operatorname{dom} \nabla f$ if the limit

$$\lim_{\substack{w' \to w \\ \tau \searrow 0}} \frac{\partial f(x + \tau w') - v}{\tau} = \lim_{\substack{w' \to w \\ \tau \searrow 0}} \Delta_\tau[\partial f](x)(w'), \quad \text{where } v = \nabla f(x), \tag{3.6}$$

exists nonemptily for every w. We're dealing here with a limit in *set convergence*, inasmuch as $\partial f(x + \tau w')$ and $\Delta_\tau[\partial f](x)(w')$ generally denote subsets of $I\!R^n$ (ones which happen to be closed and convex). Further explanation may therefore be helpful.

Recall that a sequence of closed sets $C^\nu \subset I\!R^n$ converges to a closed set $C \subset I\!R^n$ as $\nu \to \infty$ if and only if $d(z, C^\nu) \to d(z, C)$ for every $z \in I\!R^n$, where $d(z, C)$ denotes the distance of z from C. Such convergence can be characterized in many different ways, as laid out in Chapter 4 of [1]. In broader terms, the *outer* limit set $\limsup_\nu C^\nu$ consists of the points z such that $\liminf_\nu d(z, C^\nu) = 0$, whereas the *inner* limit set $\liminf_\nu C^\nu$ consists of the points z such that $\limsup_\nu d(z, C^\nu) = 0$. To say that $C^\nu \to C$ is to say that $\limsup_\nu C^\nu = C = \liminf_\nu C^\nu$.

In general, the set defined by "lim sup" in (3.6) in place of "lim" is denoted in [1] by $D[\partial f](x)(w)$, and the mapping $D[\partial f](x) : I\!R^n \rightrightarrows I\!R^n$ that is so defined is called the *graphical derivative* of ∂f at x. In speaking of semi-differentiability of f at x, we are requiring this mapping to be nonempty-valued and to agree with the mapping obtained from "lim inf."

Actually, semi-differentiability of ∂f is a bit too strong a property for our purpose of coordinating with second-order semi-differentiability of f. What we'll need is the notion of ∂f being *almost semi-differentiable* at a point $x \in \operatorname{dom} \nabla f$ in the sense that the limit in (3.6) exists nonemptily for *almost* every w. The extent to which this can be interpreted as producing an expansion of ∂f at x is as follows.

Proposition 3.3. *For ∂f to be almost semi-differentiable at a point $x \in \operatorname{dom} \nabla f$, it is necessary and sufficient that the graphical derivative mapping $D[\partial f](x)$ be monotone with $D[\partial f](x)(0)$ containing no more than 0. Then $D[\partial f](x)$ is maximal monotone, locally bounded everywhere, and single-valued almost everywhere, and*

$$\partial f(x + w) \subset v + D[\partial f](x)(w) + o(|w|)I\!B, \quad \textit{where } v = \nabla f(x). \tag{3.7}$$

Proof. From the general definition of $D[\partial f](x)(w)$ as the *outer* limit of the expressions in (3.6), $D[\partial f](x)$ can be identified with the (set-valued) mapping whose graph consists of all pairs (w, z) such that for some sequence $\tau^\nu \searrow 0$ there exist (w^ν, z^ν) with $z^\nu \in \Delta_{\tau^\nu}[\partial f](x)(w^\nu)$ and $(w^\nu, z^\nu) \to (w, z)$. In other words, the graph of $D[\partial f](x)$ is the outer limit of the graphs of the mappings $\Delta_\tau[\partial f](x)$ as $\tau \searrow 0$. Hence by the theory of set convergence it's the union of all limits of sequences of sets $C^\nu = \operatorname{gph}\Delta_{\tau^\nu}[\partial f](x)$ that converge as $\tau^\nu \searrow 0$; cf. [1; 4.19]. The mappings $\Delta_\tau[\partial f](x)$ are maximal monotone by Proposition 2.7, and it's known that if a sequence of graphs of maximal monotone mappings converges, the limit has to be the graph of another maximal monotone mapping; see [1; 12.32]. Thus,

$$\operatorname{gph} D[\partial f](x) = \bigcup \{ \operatorname{gph} T \,|\, T \in \mathcal{T} \}$$

with $\mathcal{T}$ the collection of all maximal monotone mappings having graph obtainable as the limit of $\operatorname{gph}\Delta_{\tau^\nu}[\partial f](x)$ for some sequence $\tau^\nu \searrow 0$.

When ∂f is almost semi-differentiable at x, the mappings $T \in \mathcal{T}$ must have the same nonempty value at almost every w. In particular they must have $\operatorname{dom} T = I\!R^n$ (since the domain of a maximal monotone mapping is almost convex, so that its interior is the interior of its closure; cf. [1; 12.41]). But a maximal monotone mapping is single-valued almost everywhere on the interior of its domain and is completely determined by its restriction to the points where it is single-valued; cf. [1; 12.66, 12.67]. Hence there can only be one $T \in \mathcal{T}$, namely $D[\partial f](x)$, which therefore is maximal monotone. Any maximal monotone mapping which, like $D[\partial f](x)$, has all of $I\!R^n$ as its domain, is locally bounded everywhere (cf. [1; 12.28]) as well as single-valued almost everywhere, as already noted.

Because $D[\partial f](x)$ is a positively homogeneous mapping by its definition, $D[\partial f](x)(0)$ is always a cone. Local boundedness makes this cone reduce to $\{0\}$. The inclusion in (3.7) follows simply on the basis of $D[\partial f](x)$ being locally bounded; cf. [1; 12.40].

On the other hand, if $D[\partial f](x)$ is monotone, the maximality of the monotone mappings $T \in \mathcal{T}$ again implies that $D[\partial f](x)$ must be the sole element of $\mathcal{T}$. The graphs of the maximal monotone mappings $\Delta_\tau[\partial f](x)$ therefore converge to the graph of $D[\partial f](x)$ as $\tau \searrow 0$. If in addition $D[\partial f](x)(0) = \{0\}$, the origin must be an interior point of $\operatorname{dom} D[\partial f](x)$, because maximal monotone mappings can't have nonempty bounded values except on the interiors of their domains [1; 12.38]. Since $D[\partial f](x)$ is positively homogeneous, this implies $\operatorname{dom} D[\partial f](x) = I\!R^n$. The convergence of $\operatorname{gph}\Delta_\tau[\partial f](x)$ to $\operatorname{gph} D[\partial f](x)$ as $\tau \searrow 0$ implies then that, at all points w where $D[\partial f](x)(w)$ is a singleton, one has $\Delta_\tau[\partial f](x)(w') \to D[\partial f](x)(w)$ as $w' \to w$ and $\tau \searrow 0$; this invokes a general fact about convergence of maximal monotone mappings in [1; 12.40]. We thus have ∂f almost semi-differentiable at x. $\square$

Proposition 3.4. *If, for some $x \in \operatorname{dom}\nabla f$, the mapping $D[\partial f](x)$ is single-valued everywhere, then in particular ∂f is semi-differentiable at x.*

Proof. By the same line of reasoning as for Proposition 3.3, there can only one $T \in \mathcal{T}$, so $D[\partial f](x)$ is that T and is monotone. The single-valuedness of $D[\partial f](x)$ ensures that

$D[\partial f](x)(0) = \{0\}$. Then ∂f is almost semi-differentiable at x by Proposition 3.3, and in fact, as seen toward the end of the proof of that result, one has by [1; 12.40] that $\Delta_\tau[\partial f](x)(w') \to D[\partial f](x)(w)$ as $w' \to w$ and $\tau \searrow 0$, as long as w is such that $D[\partial f](x)(w)$ is a singleton. Here we're assuming that's true for all w, so we get the convergence for all w and thus have semi-differentiability. $\square$

Theorem 3.5. *For $x \in \operatorname{dom} \nabla f$, one has f twice semidifferentiable at x if and only if ∂f is almost semi-differentiable at x. Then*

$$\partial[\tfrac{1}{2}d^2 f(x)] \;=\; D[\partial f](x). \tag{3.8}$$

Proof. The argument parallels the proof of Theorem 2.8 but depends on Proposition 3.3. We begin by assuming that f is twice semi-differentiable at x and letting $\varphi = \frac{1}{2}d^2 f(x)$ and $\varphi_\tau = \frac{1}{2}\Delta_\tau^2 f(x)$. By Proposition 3.1, φ is a finite convex function, positively homogeneous of degree two, so that $\partial\varphi$ is a maximal monotone mapping that's positively homogeneous of degree one, nonempty-valued and locally bounded everywhere. In particular, $\partial\varphi(0) = \{0\}$. By Proposition 2.7, on the other hand, each φ_τ is a closed, proper, convex function with $\partial\varphi_\tau = \Delta_\tau[\partial f](x)$. Our assumption means that, as $\tau \searrow 0$, φ_τ converges to φ uniformly on all bounded sets. That implies through [2; Theorem 24.5] the existence, for every bounded set $W \subset I\!R^n$ and $\varepsilon > 0$, of $\delta > 0$ such that $\partial\varphi_\tau(w) \subset \varphi(w)$ for all $w \in W$ when $\tau \in (0, \tau)$, or in other words,

$$\Delta_\tau[\partial f](x)(w) \;\subset\; \partial\varphi(w) + \varepsilon I\!B \;\text{ when }\; \tau \in (0,\delta),\;\; w \in W.$$

The graph of $D[\partial f](x)$ must therefore be contained within the graph of $\partial\varphi$, so $D[\partial f](x)$ must be monotone. Then by Proposition 3.3, ∂f is almost semi-differentiable at x and $D[\partial f](x)$ is maximal monotone. Because $\operatorname{gph} D[\partial f](x) \subset \operatorname{gph} \partial\varphi$ and $\partial\varphi$ is maximal monotone, we have to have $D[\partial f](x) = \partial\varphi$, i.e., (3.8).

To argue in the other direction, we assume now that ∂f is almost semi-differentiable at x, in which case the characterization in Proposition 3.3 is available. Consider any $\varepsilon > 0$ and $\rho > 0$. The expansion (3.7) can be written as $\partial f(x + \tau w) \subset v + D[\partial f](x)(\tau w) + o(\tau|w|)I\!B$ where $\tau > 0$ and $D[\partial f](x)(\tau w) = \tau D[\partial f](x)(w)$, and it accordingly yields the existence of $\delta > 0$ such that

$$\Delta_\tau[\partial f](x)(w) \;\subset\; D[\partial f](x)(w) + (\varepsilon/\rho)I\!B \;\;\text{ when }\; \tau \in (0,\delta),\;\; |w| \le \rho. \tag{3.9}$$

Let E be the set of w where $D[\partial f](x)$ is single-valued, and let F be the single-valued mapping obtained by restricting $D[\partial f](x)$ to E. From (3.9) we have

$$\begin{aligned} &|\nabla f(x + \tau w) - v - \tau F(w)| \le \tau\varepsilon/\rho \\ &\qquad\qquad \text{when }\; \tau \in (0,\delta),\;\; |w| \le \rho,\;\; w \in E,\;\; x + \tau w \in \operatorname{dom} \nabla f. \end{aligned} \tag{3.10}$$

We can suppose δ to be small enough that $x + \delta\rho I\!\!B \subset \operatorname{int}\operatorname{dom} f$. Then, as explained in the proof of Theorem 2.8, there's a subset D of full measure in $\rho I\!\!B$ such that, when $w \in D$, f is differentiable at $x + tw$ for almost all $t \in [0, \delta]$. For such w the function $\psi_w : t \mapsto f(x+tw) - f(x) - t\langle v, w\rangle - \frac{1}{2}t^2\langle F(w), w\rangle$ is Lipschitz continuous on $[0, \delta]$ with $\psi_w'(t) = \langle \nabla f(x+tw) - v - tF(w), w\rangle$ a.e. in t. Because $\psi_w(0) = 0$, we have $|\psi_w(\tau)| \le \int_0^\tau |\psi_w'(t)| dt$ which, as long as w belongs also to E, gives us through (3.10) the estimate

$$\begin{aligned} &\big| f(x+\tau w) - f(x) - \tau\langle v, w\rangle - \tfrac{1}{2}\tau^2\langle F(w), w\rangle \big| \\ &\qquad \le \int_0^\tau \big|\langle \nabla f(x+tw) - v - tF(w), w\rangle\big| dt \le \int_0^\tau \tau\varepsilon d\tau = \tfrac{1}{2}\tau^2\varepsilon. \end{aligned}$$

On dividing this by $\frac{1}{2}\tau^2$, we get $\big|\Delta_\tau^2 f(x)(w) - \langle F(w), w\rangle\big| \le \varepsilon$. We have thus demonstrated the existence, for any $\varepsilon > 0$ and $\rho > 0$, of $\delta > 0$ such that

$$\big|\Delta_\tau^2 f(x)(w) - \langle F(w), w\rangle\big| \le \varepsilon \text{ when } \tau \in (0, \delta),\ \ w \in D \cap E,$$

with $D \cap E$ being a set of full measure in $\rho I\!\!B$, hence dense in $\rho I\!\!B$.

It follows that, as $\tau \searrow 0$, the functions $\Delta_\tau^2 f(x)$ converge pointwise on $D \cap E$ to a finite function, namely $w \mapsto \langle F(w), w\rangle$. When convex functions converge pointwise to a finite value at all points of a dense subset of an open convex set O, they converge finitely everywhere on O and do so uniformly on compact subsets of O; cf. [2; Theorem 10.8]. Hence the functions $\Delta_\tau^2 f(x)$ converge uniformly on compact subsets of $\operatorname{int} \rho I\!\!B$ to a certain finite function, which is convex and consequently continuous. We have shown this for arbitrary $\rho > 0$ and can conclude therefore that f is twice semi-differentiable at x. □

The proof of Theorem 3.5 indicates that, on the set of vectors w where $D[\partial f](x)(w)$ reduces to a single value $F(w)$, one has $d^2 f(x)(w) = \langle F(w), w\rangle$. This is analogous to the case of f being twice differentiable at x in the extended sense, where $d^2 f(x)(w) = \langle \nabla^2 f(x)w, w\rangle$. The general rule is the following.

Corollary 3.6. *At any $x \in \operatorname{dom} \nabla f$ where f is twice semi-differentiable, one has*

$$d^2 f(x)(w) = \langle z, w\rangle \quad \textit{for every } z \in D[\nabla f](x)(w). \tag{3.12}$$

Proof. The formula for $d^2 f(x)$ in Corollary 3.2 gives us $\partial[\frac{1}{2}d^2 f(x)](w) = \gamma_C(w)\partial\gamma_C(w)$. Thus by (3.8), each $z \in D[\partial f](x)(w)$ has the form $z = \gamma_C(w)u$ for some $u \in \partial\gamma_C(w)$. Here $\langle u, w\rangle = \gamma_C(w)$ by the positive homogeneity of γ_C, so $\langle z, w\rangle = \gamma_C(w)^2 = d^2 f(x)(w)$. □

4. Second derivatives based on variational convergence

While Theorem 3.5 provides an equivalence and other insights, it offers no criterion for ascertaining whether the semi-differentiability in question is present. All we know is that f is twice semi-differentiable at almost every point $x \in \operatorname{dom} \nabla f$, inasmuch as that property includes the cases where f twice differentiable in the extended sense, which are covered by Corollary 2.5. Another limitation of the ideas discussed so far is that they apply to a convex function f only at interior points of $\operatorname{dom} f$, and even then, only with vigor at points of $\operatorname{dom} \nabla f$. Some of the biggest successes of first-order convex analysis have come through the treatment of directional derivatives and subgradients at boundary points of $\operatorname{dom} f$ and of course in allowing fully for the set-valuedness of ∂f.

The source of these limitations is fundamentally in the reliance on uniform convergence of difference quotients. To progress further, we have to replace such convergence by something else, and the natural candidate is variational convergence, i.e., epi-convergence of extended-real-valued functions along with graphical convergence of set-valued mappings. Both are based on set convergence, but in application to epigraphs and graphs.

A sequence of functions $\varphi^\nu : I\!R^n \rightrightarrows \overline{I\!R}$ is said to *epi-converge* to another such function φ if their epigraphs $\operatorname{epi} \varphi^\nu$ converge to $\operatorname{epi} \varphi$ as subsets of $I\!R^n \times I\!R$. The topic is developed at length in Chapter 7 of [1]. A sequence of set-valued mappings $T^\nu : I\!R^n \rightrightarrows I\!R^n$ is said to converge *graphically* to another such mapping T if their graphs $\operatorname{gph} T^\nu$ converge to $\operatorname{gph} T$ as subsets of $I\!R^n \times I\!R^n$. We've already been using this notion, although not by name. It had a major role in the proof of Proposition 3.3, in particular. Details about graphical convergence can be found in Chapter 5 of [1].

In convex analysis, there's a key fact that relates these two kinds of convergence. It was proved by Attouch [8] in 1977 and says that a sequence of closed, proper, convex functions φ^ν epi-converges to such a function φ if and only if their subgradient mappings $\partial\varphi^\nu$ converge graphically to $\partial\varphi$ and (for the sake of fixing the constant of integration) some pair $(w, z) \in \operatorname{gph} \partial\varphi$ can be approached by pairs $(w^\nu, z^\nu) \in \operatorname{gph} \partial\varphi^\nu$ in such a way that $\varphi^\nu(w^\nu) \to \varphi(w)$.

The possibility that Attouch's theorem could be employed in connecting second derivatives of f with first derivatives of ∂f was uncovered by Rockafellar [9] in 1985 in a special case and eventually brought to bloom in [10]. Our goal here is to explain the main results briefly and then to assess what they say about semi-differentiability.

Whereas previously we worked with difference quotients in (2.7) and (2.8) in which the vector v was understood to be $\nabla f(x)$, we must now work with general subgradients $v \in \partial f(x)$. This has to be reflected in our notation. Accordingly we define the second-order difference quotient function

$$\Delta_\tau^2 f(x \,|\, v) : w \mapsto \frac{f(x+\tau w) - f(x) - \tau\langle v, w\rangle}{\frac{1}{2}\tau^2}, \quad \text{where } v \in \partial f(x), \tag{4.1}$$

and the first-order difference quotient mapping

$$\Delta_\tau[\partial f](x \,|\, v) : w \mapsto \frac{\partial f(x+\tau w) - v}{\tau}, \quad \text{where } v \in \partial f(x). \tag{4.2}$$

At a point $x \in \operatorname{dom} \partial f$ and for any $v \in \partial f(x)$, we denote by $d^2 f(x \,|\, v)$ the function with epigraph obtained as the "lim sup" (outer limit) of the epigraphs of the functions $\Delta_\tau^2 f(x \,|\, v)$ as $\tau \searrow 0$ and say that f is *twice epi-differentiable at x for v* if actually this "lim sup" is a "lim" and is proper, i.e., if $\Delta_\tau^2 f(x \,|\, v)$ in fact epi-converges as $\tau \searrow 0$. Likewise, we denote by $D[\partial f](x \,|\, v)$ the mapping with graph obtained as the "lim sup" of the graphs of the mappings $\Delta_\tau[\partial f](x \,|\, v)$ as $\tau \searrow 0$ and say that ∂f is *proto-differentiable at x for v* if this "lim sup" is a "lim," i.e., if $\Delta_\tau[\partial f](x \,|\, v)$ converges graphically to $D[\partial f](x \,|\, v)$ as $\tau \searrow 0$.

These differentiation concepts based on variational convergence were introduced in [11] and [12], respectively.

Theorem 4.1 [10]. *Let $x \in \operatorname{dom} \partial f$ and $v \in \partial f(x)$. Then f is twice epi-differentiable at x for v if and only if ∂f is proto-differentiable at x for v, in which case*

$$\partial[\tfrac{1}{2} d^2 f(x \,|\, v)] = D[\partial f](x \,|\, v), \tag{4.3}$$

with $d^2 f(x \,|\, v)$ being a closed, proper, convex function that is nonnegative and positively homogeneous of degree two, and $D[\partial f](x \,|\, v)$ being a maximal monotone mapping that is positively homogeneous (of degree one).

Proof. Just as in Proposition 2.7, it's elementary that for each $\tau > 0$ the function $\Delta_\tau^2 f(x \,|\, v) : \mathbb{R}^n \to \overline{\mathbb{R}}$ is closed, proper and convex and nonnegative, while the mapping $\Delta_\tau[\partial f](x \,|\, v) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone, and

$$\partial[\tfrac{1}{2} \Delta_\tau^2 f(x \,|\, v)] = \Delta_\tau[\partial f](x \,|\, v). \tag{4.4}$$

In addition, we have that $\Delta_\tau^2 f(x \,|\, v)(0) = 0$ and $\Delta_\tau[\partial f](x \,|\, v)(0) = 0$. To get the result, all we have to do is apply Attouch's theorem in the statement of it supplied above, with $(w^\nu, z^\nu) = (w, z) = (0, 0)$. (Because the functions $\Delta_\tau^2 f(x \,|\, v)$ are nonnegative and vanish at the origin, there's no risk of them epi-converging to a function that's improper.) $\square$

An important feature of epi-convergence of convex functions is that it's preserved when passing to conjugate functions: if φ^ν epi-converges to φ, then $\varphi^{\nu *}$ epi-converges to φ^*. This fact, proved by Wijsman [13] in 1964, is ideal for answering questions about what happens to second derivatives when passing to conjugate functions. The next result comes by that route in recalling that $v \in \partial f(x)$ if and only if $x \in \partial f^*(v)$.

Theorem 4.2 [10]. *One has f twice epi-differentiable at x for v if and only if f^* is twice epi-differentiable at v for x, in which case*

$$[\tfrac{1}{2}d^2f(x\,|\,v)]^* = \tfrac{1}{2}d^2f^*(v\,|\,x). \tag{4.5}$$

Proof. All that's needed is to apply Theorem 4.2 to the functions $\varphi_\tau = \frac{1}{2}\Delta^2_\tau f(x\,|\,v)$ as $\tau \searrow 0$ while observing that $\varphi^*_\tau = \frac{1}{2}\Delta^2_\tau f^*(v\,|\,x)$. $\square$

What is the relationship between second-order epi-differentiation and second-order semi-differentiation? This has a helpful answer.

Theorem 4.3. *Let $x \in \operatorname{dom} \nabla f$ and $v = \nabla f(x)$. Then f is twice semi-differentiable at x if and only if f is twice epi-differentiable at x for v and $d^2f(x\,|\,v)(w)$ is finite for all w. The second-order semi-derivative function $d^2f(x)$ coincides then with the second-order epi-derivative function $d^2f(x\,|\,v)$.*

Proof. The crucial fact is that a sequence of convex functions φ^ν epi-converges to a *finite* convex function φ if and only it converges to φ uniformly on all bounded sets; cf. [1; 7.17]. By applying this to the convergence of second-order difference quotient functions, we get the relationship claimed because second-order semi-differentiability corresponds to such uniform convergence to a finite function that's continuous, and finite convex functions are automatically continuous. $\square$

Corollary 4.4. *Let $x \in \operatorname{dom} \nabla f$ and $v = \nabla f(x)$. Then f is twice semi-differentiable at x if and only if f^* is twice epi-differentiable at v for x and $d^2f^*(v\,|\,x)$ is positive-definite in the sense that*

$$d^2f^*(v\,|\,x)(z) > 0 \ \textit{for all}\ z \neq 0. \tag{4.6}$$

Proof. Combining the duality in Theorem 4.2 with the assertions of Theorem 4.3, we are able to translate the second-order epi-differentiability of f to that of f^* and the finiteness of $d^2f(x\,|\,v)$ to the coercivity of $d^2f^*(v\,|\,x)$. Because $d^2f^*(v\,|\,x)$ is positively homogeneous of degree two by Theorem 4.1, it's coercive if and only if (4.6) holds. $\square$

Corollary 4.5. *Let $x \in \operatorname{dom} \nabla f$ and $v = \nabla f(x)$. A necessary and sufficient condition for ∂f to be almost semi-differentiable at x is that ∂f be proto-differentiable at x for v with $D[\partial f](x\,|\,v)(0)$ containing only 0. Then $D[\partial f](x) = D[\partial f](x\,|\,v)$.*

Proof. This combines Theorem 4.3 with Proposition 3.3 and Theorem 3.5. $\square$

Further material on the relationship between Theorem 4.1 and "expansions" of f or ∂f is available in [14].

These characterizations focus ever greater attention on the question of how to know whether a function f or f^* is twice epi-differentiable somewhere, and if so, what formulas might be used to express the epi-derivatives. According to Theorem 4.3 and its corollaries, information about that provides information also on semi-differentiability.

In fact a large class of functions, central to finite-dimensional optimization, has been shown in [11] to be twice epi-differentiable, namely the "fully amenable" functions. The results and formulas recently been explained also in [1], and there is no need for further duplication of them here. Instead, we conclude with some observations about their applications.

The tie between second-order epi-derivatives of f and proto-derivatives of ∂f in Theorem 4.1 is particularly valuable for the study of perturbations of solutions to problems of optimization. Typically those solutions are characterized in terms of subgradients, and in looking at the way solutions depend on parameter vectors, one therefore ends up looking at set-valued mappings derived from subgradient mappings. Proto-differentiability of a solution mapping can then be deduced from proto-differentiability of a subgradient mapping. The fact in Theorem 4.1 that the proto-derivatives $D[\partial f](x \,|\, v)(w)$ are the subgradients $\partial h(w)$ of $h = \frac{1}{2}d^2 f(x \,|\, v)$ can in that case have the remarkable consequence that the proto-derivatives of the solution mapping can be calculated by solving an auxiliary optimization problem in which the original objective function has been replaced by one of its second-order epi-derivative functions. Such developments can be found in [15] and [16] as well as, to some extent, in [1].

Other recent developments on solution perturbations, in [17] and [18], utilize yet another concept of generalized second derivative, namely the "coderivative Hessian" mapping introduced by Mordukhovich [19], [20]. A basic inclusion between the proto-derivative mappings and coderivative mappings associated with ∂f has been established in [21].

The extent to which such developments, generally for nonconvex functions, might lead to special results for convex functions, hasn't really been explored. That's the status also of results on generalized second derivatives of parabolic type, taken along quadratic curves rather than linear rays. Parabolic second derivatives of fully amenable functions were demonstrated in [11] to obey a certain duality with second epi-derivatives. They have been featured in some recent work on optimality; cf. [22].

References.

1. R. T. Rockafellar and R. J-B Wets, *Variational Analysis*, Springer, Berlin, 1998.
2. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.
3. A. D. Alexandrov, "Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it," Leningrad Sate University Annals [Uchenye Zapiski], Mathematical Series, 6 (1939), 3–35 (in Russian).
4. H. Busemann and W. Feller, "Krümmungseigenschaften convexer Flächen," Acta Mathematica, 66 (1936), 1–47.
5. F. Mignot, "Contrôle dans les inéquations variationelles elliptiques," Journal of Functional Analysis, 22 (1976), 130–185.
6. J.-B. Hiriart-Urruty, "A new set-valued second-order derivative for convex functions," 157–182 in *Mathematics for Optimization* (J.-B. Hiriart-Urruty, ed.), Mathematics Study Series 129, North-Holland, Amsterdam, 1986.
7. J.-B. Hiriart-Urruty and A. Seeger, "Calculus rules on a new set-valued second-order derivatives for convex functions," Nonlinear Analysis: Th. Meth. & Appl. 13 (1989), 721–738.
8. H. Attouch, "Convergence de fonctions convexes, de sous-différentiels et semi-groupes," Comptes Rendus de l'Académie des Sciences de Paris 284 (1977), 539–542.
9. R. T. Rockafellar, "Maximal monotone relations and the second derivatives of nonsmooth functions," Annales de l'Institut H. Poincaré: Analyse Non Linéaire 2 (1985), 167–184.
10. R. T. Rockafellar, "Generalized second derivatives of convex functions and saddle functions," Transactions Amer. Math. Soc. 322 (1990), 51–77.
11. R. T. Rockafellar, "First and second-order epi-differentiability in nonlinear programming," Transactions Amer. Math. Soc. 307 (1988), 75–108.
12. R. T. Rockafellar, "Proto-differentiability of set-valued mappings and its applications in optimization," 449–482 in *Analyse Non Linéaire* (H. Attouch, J.-P. Aubin, F. Clarke and I. Ekeland, eds.), Gauthier-Villars, Paris, 1989.
13. R. J. Wijsman, "Convergence of sequences of convex sets, cones and functions," Bulletin Amer. Math. Soc. 123 (1964), 32–45.
14. R. Poliquin and R. T. Rockafellar, "Generalized Hessian properties of regularized nonsmooth functions," SIAM J. Optim. 6 (1996), 1121–1137.

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

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

17. R. Poliquin and R. T. Rockafellar, "Tilt stability of a local minimum," SIAM J. Optim. 8 (1998), 287–299.

18. A. B. Levy, R. A. Poliquin and R. T. Rockafellar, "Stability of locally optimal solutions," SIAM J. Optim. (forthcoming).

19. B. S. Mordukhovich, "Sensitivity analysis in nonsmooth optimization," 32–46 in *Theoretical aspects of industrial design* (D. A. Field and V. Komkov, eds.), SIAM Volumes in Applied Mathematics 58, 1992.

20. B. S. Mordukhovich, "Lipschitzian stability of constraint systems and generalized equations," Nonlinear Analysis Th. Meth. Appl. 22 (1994), 173–206.

21. R. T. Rockafellar and D. Zagrodny, "A derivative-coderivative inclusion in second-order nonsmooth analysis," Set-Valued Analysis 5 (1997), 1–17.

22. J.-F. Bonnans, R. Cominetti and A. Shapiro, "Second order necessary and sufficient optimality conditions under abstract constraints," SIAM J. Optimization, to appear.