GENERALIZED HESSIAN PROPERTIES OF REGULARIZED NONSMOOTH FUNCTIONS

R. A. Poliquin[1] **and R. T. Rockafellar**[2]

Abstract. The question of second-order expansions is taken up for a class of functions of importance in optimization, namely Moreau envelope regularizations of nonsmooth functions f. It is shown that when f is prox-regular, which includes convex functions and the extended-real-valued functions representing problems of nonlinear programming, the many second-order properties that can be formulated around the existence and stability of expansions of the envelopes of f or of their gradient mappings are linked by surprisingly extensive lists of equivalences with each other and with generalized differentiation properties of f itself. This clarifies the circumstances conducive to developing computational methods based on envelope functions, such as second-order approximations in nonsmooth optimization and variants of the proximal point algorithm. The results establish that generalized second-order expansions of Moreau envelopes, at least, can be counted on in most situations of interest in finite-dimensional optimization.

Keywords. Prox-regularity, amenable functions, primal-lower-nice functions, Hessians, first- and second-order expansions, strict proto-derivatives, proximal mappings, Moreau envelopes, regularization, subgradient mappings, nonsmooth analysis, variational analysis, proto-derivatives, second-order epi-derivatives, Attouch's theorem.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 49A52, 58C06, 58C20; Secondary 90C30

May, 1995

---

[1] Work supported by the Natural Sciences and Engineering Research Council of Canada under grant OGP41983. Dept. of Math. Sciences, Univ. of Alberta, Edmonton, Alberta, Canada T6G 2G1; rene@fenchel.math.ualberta.ca.

[2] Work supported by the National Science Foundation under grant DMS–9200303. Dept. of Math., Univ. of Washington, Seattle, WA 98195 USA; rtr@math.washington.edu.

1. Introduction

Any problem of optimization in $I\!R^n$ can be posed abstractly as one of minimizing a function $f : I\!R^n \to \overline{I\!R} := I\!R \cup \pm\infty$ over all of $I\!R^n$; constraints are represented by infinite penalization. It is natural then that f be lower semicontinuous (l.s.c.). In this setting, useful for many purposes, consider a point $\bar{x}$ where f achieves its minimum, $f(\bar{x})$ being finite. For any $\lambda > 0$, $\bar{x}$ also minimizes the Moreau envelope function

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

which provides a *regularization* of f. While f may have ∞ values and exhibit discontinuities, e_λ is finite and locally Lipschitz continuous, and it approximates f in the sense that e_λ increases pointwise to f as $\lambda \searrow 0$. Among other regularizing effects, e_λ has one-sided directional derivatives at all points, even Taylor expansions of degree 2 almost everywhere, and at $\bar{x}$ it is differentiable with

$$e_\lambda(\bar{x}) = f(\bar{x}), \qquad \nabla e_\lambda(\bar{x}) = 0. \tag{1.2}$$

When f is convex, e_λ is convex too and actually of class C^{1+}: differentiable with locally Lipschitz continuous gradient mapping ∇e_λ.

The prospect is raised that if the properties of e_λ around $\bar{x}$ were adequately understood in their relation to f, it might be possible to develop methods for minimizing e_λ which in effect would open new approaches to minimizing f despite nonsmoothness. Here we analyze e_λ in ways suggested by this notion. Although we do not explore the algorithmic potential directly, we try to identify the circumstances in which e_λ has favorable second-order properties around $\bar{x}$ for the support of numerical work. These include the existence of a Hessian matrix at $\bar{x}$ and the presence of continuity of some sort at $\bar{x}$ in certain generalized second-derivatives of e_λ.

In this effort we are following the lead of Lemaréchal and Sagastizábal [13], who have recently drawn attention to the numerical motivations, noting the connections between Moreau envelopes e_λ and basic techniques like the proximal point algorithm. They replaced the canonical squared Euclidean norm in (1.1) by a quadratic form corresponding to any symmetric, positive definite matrix, as obviously would be useful for computational purposes. In theoretical discussions the canonical notation can be retained for simplicity, since the extension can also be viewed as amounting just to a linear change of variables.

Other work building on [13] has been carried out by Qi in [30]. For motivations coming from recent efforts at establishing superlinearly convergent algorithms for solving nonsmooth optimization problems, we refer to [1], [5], [8], [9], [18], [28], and [29].

Lemaréchal and Sagastizábal concentrated in [13] on the case of finite, convex functions f. In this paper, however, we drop both finiteness and convexity as prerequisites on f so as to be able to cover a greater range of problems of optimization. We aim in particular at including the extended-real-valued functions f that correspond to standard models in nonlinear programming, for instance the minimization of

$$f(x) = \begin{cases} f_0(x) & \text{for } x \in C, \\ \infty & \text{for } x \notin C, \end{cases} \tag{1.3}$$

when C is specified by a system of constraints, say

$$C = \big\{ x \,\big|\, \big(f_1(x), \ldots, f_m(x)\big) \in D \big\} \tag{1.4}$$

with D a closed, convex set in $\mathbb{R}^m$ and the functions f_i of class $\mathcal{C}^2$, and with f_0 likewise such a function or perhaps nonsmooth of the form

$$f_0(x) = \max\big\{ f_{01}(x), \ldots, f_{0s}(x) \big\} \tag{1.5}$$

for functions f_{0k} of class $\mathcal{C}^2$.

To achieve such coverage we must be careful in selecting the general class of functions f on which to focus our efforts. Obviously the class has to be very broad, if all l.s.c. convex functions are to be encompassed along with functions of the kind just described. Yet if the class were taken to be too broad, it would not have enough structure to secure properties of the kind we are looking for in the envelopes e_λ.

Fortunately a natural choice can be made: the class of *prox-regular* functions f. Such functions, introduced and investigated in [25], include all "strongly amenable" functions, hence all the examples mentioned so far and much more. The theory of their generalized derivatives is available in a form that makes translation of properties between f and e_λ relatively straightforward. The key facts about prox-regular functions developed in [25] will be reviewed in Section 2. Prox-regularity of f at $\bar{x}$ ensures in particular that e_λ is of class $\mathcal{C}^{1+}$ on a neighborhood of $\bar{x}$.

Functions that are $\mathcal{C}^{1+}$ have been the focus of much research recently. The reader interested in the study of generalized second-order directional derivatives and Hessians of these functions will want to consult the work of Cominetti and Correa [6], Hiriart-Urruty [10], Jeyakumar and Yang [12], Páles and Zeidan [17], and Yang and Jeyakumar [39]. Note that here the function e_λ is not only $\mathcal{C}^{1+}$ but also lower-$\mathcal{C}^2$.

In the exploration of second-order aspects of e_λ, the following notions will be useful.

Definition 1.1. *Consider a function g on $I\!R^n$ and a point $\hat{x}$ where g is differentiable.*

(a) *g has a second-order expansion at $\hat{x}$ if there is a finite, continuous function h such that the second-order difference quotient functions*

$$\Delta^2_{\hat{x},t} g(\xi) := \big[g(\hat{x}+t\xi) - g(\hat{x}) - t\langle \nabla g(\hat{x}), \xi\rangle\big] \big/ \tfrac{1}{2}t^2$$

converge to h uniformly on bounded sets as $t \searrow 0$. The expansion is strict if g is differentiable not only at $\hat{x}$ but on a neighborhood of $\hat{x}$, and the functions

$$\Delta^2_{x,t} g(\xi) := \big[g(x+t\xi) - g(x) - t\langle \nabla g(x), \xi\rangle\big] \big/ \tfrac{1}{2}t^2$$

converge to h uniformly on bounded sets as $t \searrow 0$ and $x \to \hat{x}$.

(b) *g has a Hessian matrix H at $\hat{x}$, this being a symmetric $n \times n$ matrix, if g has a second-order expansion with $h(\xi) = \langle \xi, H\xi\rangle$. The Hessian is strict if the expansion is strict.*

(c) *g is twice differentiable at $\hat{x}$ if its first partial derivatives exist on a neighborhood of $\hat{x}$ and are themselves differentiable at $\hat{x}$, i.e., the second partial derivatives of g exist at $\hat{x}$. Then $\nabla^2 g(\hat{x})$ denotes the matrix formed by these second partial derivatives.*

A second-order expansion in the sense of Definition 1.1 automatically requires the function h also to be positively homogeneous of degree 2: $h(\lambda\xi) = \lambda^2 h(\xi)$ for $\lambda > 0$, and in particular, $h(0) = 0$. In traditional notation it means that

$$g(\hat{x}+t\xi) \;=\; g(\hat{x}) + t\langle \nabla g(\hat{x}), \xi\rangle + \tfrac{1}{2}t^2 h(\xi) + o\big(t^2|\xi|^2\big)$$

for such a function h that is finite and continuous. The existence of a Hessian corresponds to h actually being quadratic. Note that, in general, this property might be present without the Hessian matrix H being the matrix $\nabla^2 g(\hat{x})$ of classical twice differentiability, since the latter requires more for its definition than the mere assumption that the first partial derivatives of g exist at $\hat{x}$.

Our fundamental theorem, in Section 3, says that when f is prox-regular at a point $\bar{x} \in \operatorname{argmin} f$ ($= \operatorname{argmin} e_\lambda$), the existence of a second-order expansion of e_λ at $\bar{x}$ (when λ is sufficiently small) is equivalent to the "twice epi-differentiability" of f at $\bar{x}$. This is especially illuminating because it shows one cannot expect the regularized function e_λ to have significant second-order properties without some such second-order property already being present in f, moreover a kind of property satisfied by many, but definitely not all, convex or lower-$\mathcal{C}^2$ functions. This is in sharp contrast with the first-order theory, where no first-order assumptions on f are needed to conclude that e_λ is differentiable (in fact

C^{1+}) around a point $\bar{x} \in \operatorname{argmin} f$ when f is prox-regular. Twice epi-differentiability is enjoyed by all fully amenable functions in particular, so second-order expansions of e_λ nonetheless are the rule for most applications.

Proceeding from this theorem, we are able to demonstrate another special feature of the Moreau envelopes e_λ when f is prox-regular. A Hessian H exists for e_λ at $\bar{x} \in \operatorname{argmin} f$ if and only if the gradient mapping ∇e_λ is differentiable at $\bar{x}$, in which case H is the matrix $\nabla(\nabla e_\lambda)(\bar{x}) = \nabla^2 e_\lambda(\bar{x})$ of second-partial derivatives—which therefore must be symmetric. This is somewhat surprising, since the ordinary criterion for the symmetry of the matrix of second derivatives of a function is that the function be of class $\mathcal{C}^2$, whereas here the first derivatives are only known to be Lipschitz continuous, hence perhaps only differentiable almost everywhere.

We show further that the existence of a Hessian for e_λ does not require the finiteness of f, but merely that the corresponding second-order epi-derivative function for f be a "generalized quadratic function" (i.e., a quadratic function on a certain subspace, but ∞ elsewhere). This covers situations where, for example, f has the nonlinear programming form (1.3)–(1.5) and $\bar{x}$ is a point where f is not differentiable, much less twice differentiable. But in particular, if f does have a Hessian at a point $\bar{x} \in \operatorname{argmin} f$, then e_λ inherits this property.

In Section 4 we establish that e_λ cannot have a strict second-order expansion at $\bar{x} \in \operatorname{argmin} f$ unless it actually has a strict Hessian there. We characterize this situation in many ways, tying it to various properties of ∇e_λ and f. Our efforts disclose that this is the one and only case, short of e_λ actually being $\mathcal{C}^2$ around $\bar{x}$, where second-order properties of e_λ and f exhibit a degree of local stability.

2. Prox-Regularity and Derivatives

Continuing under the assumption that f is a proper, l.s.c. function on $\mathbb{R}^n$, we denote by $\partial f(x)$ the set of limiting proximal subgradients of f at any point $x \in \operatorname{dom} f$. For a recent exposition of the theory of subgradients, see Loewen [16].

Definition 2.1. *The function f is prox-regular at $\hat{x}$ relative to $\hat{v}$ if $\hat{x} \in \operatorname{dom} f$, $\hat{v} \in \partial f(\hat{x})$, and there exist $\varepsilon > 0$ and $r > 0$ such that*

$$f(x') > f(x) + \langle v, x' - x \rangle - \frac{r}{2}\left|x' - x\right|^2 \tag{2.1}$$

whenever $\left|x' - \hat{x}\right| < \varepsilon$ and $\left|x - \hat{x}\right| < \varepsilon$, with $x' \neq x$ and $\left|f(x) - f(\hat{x})\right| < \varepsilon$, while $\left|v - \hat{v}\right| < \varepsilon$ with $v \in \partial f(x)$.

It is obvious that if f is convex, it is prox-regular at $\hat{x}$ for every subgradient $\hat{v} \in \partial f(\hat{x})$. The same is true for lower-$\mathcal{C}^2$ functions, primal-lower-nice (p.l.n.) functions, and *strongly amenable* functions, cf. [25]. Strong amenability of f at $\hat{x}$ refers to the existence of a representation $f = g \circ F$ in a neighborhood of $\hat{x}$ for a mapping $F : I\!R^n \to I\!R^m$ of class C^2 and a proper, l.s.c., convex function $g : I\!R^m \to \overline{I\!R}$ satisfying at $\hat{x}$ with respect to the convex set $D = \operatorname{dom} g$ the basic constraint qualification that

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

Amenability itself merely requires F to be of class $\mathcal{C}^1$, whereas *full* amenability is the subcase of strong amenability where g is also piecewise linear-quadratic.

Full amenability already covers most applications that arise in the framework of nonlinear programming and its extensions. A function f as in (1.3)–(1.5), for instance, is fully amenable at any point $\hat{x} \in C$ at which the constraint qualification (2.2) is satisfied with $F = (f_1, \ldots, f_m)$, cf. [33] and [35]. If D consists of the vectors $u = (u_1, \ldots, u_m)$ with $u_i \leq 0$ for $i = 1, \ldots, s$ but $u_i = 0$ for $i = s+1, \ldots, m$, so that C is specified by constraints $f_i(x) \leq 0$ and $f_i(x) = 0$, this constraint qualification reduces to the standard one of Mangasarian-Fromovitz. For more on amenability, see [22]–[24]; for composite functions more broadly, see [3], [11]–[12], [14]–[15], [19]–[27].

Note that when f is *convex*, the parameter $\varepsilon > 0$ in the definition of prox-regularity can be taken *arbitrarily large*, while the parameter $r > 0$ can be taken *arbitrarily small*. This lends a stronger interpretation to some of the results below.

For any $\varepsilon > 0$, we speak of the *f-attentive ε-localization* of ∂f at $(\hat{x}, \hat{v})$ in referring to the mapping $T : I\!R^n \rightrightarrows I\!R^n$ defined by

$$T(x) = \begin{cases} \{v \in \partial f(x) \,\big|\, |v - \hat{v}| < \varepsilon\} & \text{when } |x - \hat{x}| < \varepsilon \text{ and } |f(x) - f(\hat{x})| < \varepsilon, \\ \emptyset & \text{otherwise.} \end{cases} \tag{2.3}$$

The graph of T is the intersection of $\operatorname{gph} \partial f$ with the product of an f-attentive neighborhood of $\hat{x}$ and an ordinary neighborhood of $\hat{v}$. (The f-attentive topology on $I\!R^n$ is the weakest topology in which f is continuous.)

In cases where f is subdifferentially continuous at $\hat{x}$ for $\hat{v}$, in the sense that $f(x) \to f(\hat{x})$ automatically when $(x, v) \to (\hat{x}, \hat{v})$ in the graph of ∂f, the $|f(x) - f(\hat{x})| < \varepsilon$ enforcing f-attentiveness in (2.3) could be dropped without affecting any or the results below. All amenable functions, in particular, l.s.c. convex functions and lower-$\mathcal{C}^2$ functions, are subdifferentially continuous.

Theorem 2.2 [25](Thm. 3.2)**.** *The function f is prox-regular at a point $\hat{x}$ relative to $\hat{v}$ if and only if the vector $\hat{v}$ is a proximal subgradient of f at $\hat{x}$ and there exist $\varepsilon > 0$ and $r > 0$*

such that, for the f-attentive ε-localization T of ∂f, the mapping $T+rI$ is monotone, i.e., one has $\langle v_1-v_0, x_1-x_0\rangle \geq -r\big|x_1-x_0\big|^2$ whenever $\big|x_i-\hat{x}\big|<\varepsilon$, $\big|f(x_i)-f(\hat{x})\big|<\varepsilon$, and $\big|v_i-\hat{v}\big|<\varepsilon$ with $v_i\in\partial f(x_i)$, $i=0,1$.

Concepts of generalized second-order differentiability beyond those already given in Definition 1.1 will be crucial in our analysis. Recall f is *twice epi-differentiable* at $\hat{x}$ for a vector $\hat{v}$ if $\hat{v}\in\partial f(\hat{x})$ and the second-order difference quotient functions $\Delta^2_{\hat{x},\hat{v},t}f : I\!\!R^n \to \overline{I\!\!R}$ defined by

$$\Delta^2_{\hat{x},\hat{v},t}f(\xi)=\big[f(\hat{x}+t\xi)-f(\hat{x})-t\langle\hat{v},\xi\rangle\big]\big/\tfrac{1}{2}t^2 \ \text{ for } \ t>0 \tag{2.4}$$

epi-converge to a proper function as $t\searrow 0$. The limit is then the *second epi-derivative* function $f''_{\hat{x},\hat{v}} : I\!\!R^n \to \overline{I\!\!R}$; see [23], [33] and [35]. (Epi-convergence of functions refers to set convergence of their epigraphs.) We say that f is *strictly* twice epi-differentiable at $\hat{x}$ for $\hat{v}$ if more generally the functions

$$\Delta^2_{x,v,t}f(\xi)=\big[f(x+t\xi)-f(x)-t\langle v,\xi\rangle\big]\big/\tfrac{1}{2}t^2 \ \text{ for } \ t>0 \tag{2.5}$$

epi-converge as $t\searrow 0$, $x\to\hat{x}$ with $f(x)\to f(\hat{x})$, and $v\to\hat{v}$ with $v\in\partial f(x)$. When $\partial f(\hat{x})$ is a singleton consisting of $\hat{v}$ alone (as when f is strictly differentiable at $\hat{x}$), the notation $f''_{\hat{x},\hat{v}}$ can be simplified to $f''_{\hat{x}}$ and we can just speak of f being twice epi-differentiable, or strictly twice epi-differentiable, at $\hat{x}$, the particular vector $\hat{v}$ being implicit.

Next, a set-valued mapping $T : I\!\!R^n \rightrightarrows I\!\!R^n$ is *proto-differentiable* at $\hat{x}$ for $\hat{v}$ if $\hat{v}\in T(\hat{x})$ and the set-valued mappings

$$\Delta_{\hat{x},\hat{v},t}T : \xi \mapsto \big[T(\hat{x}+t\xi)-\hat{v}\big]\big/t \ \text{ for } \ t>0 \tag{2.6}$$

graph-converge as $t\searrow 0$ (i.e., their graphs converge as subsets of $I\!\!R^n\times I\!\!R^n$). If so, the limit mapping is denoted by $T'_{\hat{x},\hat{v}}$ and called the *proto-derivative* of T at $\hat{x}$ for $\hat{v}$; see [24], [34], [36]. This mapping assigns to each $\xi\in I\!\!R^n$ a subset $T'_{\hat{x},\hat{v}}(\xi)$ of $I\!\!R^n$, which could be empty for some choices of ξ. We say that T is *strictly* proto-differentiable at $\hat{x}$ for $\hat{v}$ if more generally the mappings

$$\Delta_{x,v,t}T : \xi \mapsto \big[T(x+t\xi)-v\big]\big/t \ \text{ for } \ t>0 \tag{2.7}$$

graph-converge as $t\searrow 0$, $x\to\hat{x}$ and $v\to\hat{v}$ with $v\in T(x)$. Again, in the case where $T(\hat{x})$ is a singleton consisting of $\hat{v}$ only (as for instance in the case where T is actually single-valued everywhere), the notation $T'_{\hat{x},\hat{v}}(\xi)$ can be simplified to $T'_{\hat{x}}(\xi)$, and we can just speak of T being proto-differentiable at $\hat{x}$, or strictly proto-differentiable at $\hat{x}$.

Theorem 2.3 [25](Thm. 6.1). *Suppose that f is prox-regular at $\hat{x}$ for $\hat{v}$ with respect to ε and r, and let T be the f-attentive ε-localization of ∂f at $(\hat{x}, \hat{v})$. Then f is twice epi-differentiable at $\hat{x}$ for $\hat{v}$ if and only if T is proto-differentiable at $\hat{x}$ for $\hat{v}$, in which event one has*

$$T'_{\hat{x},\hat{v}}(\xi) = \partial\big[\tfrac{1}{2} f''_{\hat{x},\hat{v}}\big](\xi) \text{ for all } \xi.$$

Furthermore, f is strictly twice epi-differentiable at $\hat{x}$ for $\hat{v}$ if and only if T is strictly proto-differentiable at $\hat{x}$ for $\hat{v}$.

Henceforth we concentrate on the case of minimizing points $\bar{x} \in \operatorname{argmin} f$. Such points have $\bar{v} = 0$ as subgradient $\bar{v} \in \partial f(\bar{x})$. As a companion to the envelope function e_λ we utilize the *proximal mapping* $P_\lambda : I\!R^n \rightrightarrows I\!R^n$ defined by

$$P_\lambda(x) \;:=\; \operatorname*{argmin}_{x'} \left\{ f(x') + \frac{1}{2\lambda}\big|x' - x\big|^2 \right\}. \tag{2.8}$$

Theorem 2.4 [25](Thms. 4.4, 4.6, 5.2). *Suppose that f is prox-regular at $\bar{x} \in \operatorname{argmin} f$ for $\bar{v} = 0$ with respect to ε and r, and let T be the f-attentive ε-localization of ∂f at $(\bar{x}, 0)$. Then for any $\lambda \in (0, 1/r)$ there is a convex neighborhood X_λ of $\bar{x}$ such that*

(a) *the mapping P_λ is single-valued and Lipschitz continuous on X_λ with*

$$P_\lambda = (I + \lambda T)^{-1}, \qquad \big|P_\lambda(x') - P_\lambda(x)\big| \le \frac{\lambda}{1 - \lambda r}\big|x' - x\big|, \qquad P_\lambda(\bar{x}) = \bar{x},$$

(b) *the function e_λ is $\mathcal{C}^{1+}$ and lower-$\mathcal{C}^2$ on X_λ with*

$$e_\lambda + \frac{r}{2(1 - \lambda r)}|\cdot|^2 \text{ convex }, \qquad \nabla e_\lambda = \lambda^{-1}\big[I - P_\lambda\big] = \big[\lambda I + T^{-1}\big]^{-1}.$$

3. Second-Order Expansions

Before looking at second-order expansions of the Moreau envelope functions e_λ, we inspect the case of f itself.

Theorem 3.1. *Suppose that f is prox-regular at $\hat{x}$ for $\hat{v}$. Then the following are equivalent and imply that f is lower-$\mathcal{C}^2$ on a neighborhood of $\hat{x}$, with*

$$f(\hat{x} + t\xi) \;=\; f(\hat{x}) + t\langle \hat{v}, \xi\rangle + \tfrac{1}{2}t^2 f''_{\hat{x},\hat{v}}(\xi) + o\big(t^2|\xi|^2\big). \tag{3.1}$$

(a) *f is differentiable at $\hat{x}$, $\nabla f(\hat{x}) = \hat{v}$, and f has a second-order expansion at $\hat{x}$.*

(b) *f is twice epi-differentiable at $\hat{x}$ for $\hat{v}$ with $f''_{\hat{x},\hat{v}}$ finite everywhere.*

Proof. The equivalence of (b) with the expansion (3.1) is asserted by [25](Thm. 6.7) with the function $f''_{\hat{x},\hat{v}}$ being in this case lower-$\mathcal{C}^2$ (hence continuous), and the same for f on some neighborhood of $\hat{x}$. This in turn implies (a) through Definition 1.1 with $h = f''_{\hat{x},\hat{v}}$. Conversely, if we had (a) the functions $\Delta^2_{\hat{x},\hat{v},t} f$ would in particular epi-converge to h, so $h = f''_{\hat{x},\hat{v}}$. This would give (b). $\square$

It might be conjectured that when f is prox-regular at $\bar{x} \in \operatorname{argmin} f$ for $\bar{v} = 0$ and also twice epi-differentiable for these elements, the second-order epi-derivative function $f''_{\bar{x},\bar{v}}$ ought to be *convex* (in addition to being nonnegative, l.s.c. and positively homogeneous of degree 2). But this is false, even if $f''_{\bar{x},\bar{v}}$ is finite everywhere. An example in $\mathbb{R}^2$ is

$$f(x_1, x_2) = |x_1 x_2| = \max\{-x_1 x_2,\, x_1 x_2\,\}$$

at the origin. This function is lower-$\mathcal{C}^2$, hence certainly prox-regular everywhere, but its second-order epi-derivative function at the origin is f itself, which is not convex.

Theorem 3.2. *Suppose that f is lower-$\mathcal{C}^2$ on a neighborhood of $\hat{x}$ and also differentiable on such a neighborhood (or equivalently, that f can be represented around $\hat{x}$ as the difference between a differentiable convex function and a $\mathcal{C}^2$ function). Then the following are equivalent:*

(a) *f has a Hessian matrix H at $\hat{x}$;*

(b) *∇f is differentiable at $\hat{x}$ with Jacobian matrix H;*

(c) *f is twice differentiable at $\hat{x}$, and $H = \nabla^2 f(\hat{x})$.*

Proof. The parenthetical alternative to the hypothesis on f is known from [37]. From this it is clear that differentiable, lower-$\mathcal{C}^2$ functions f are $\mathcal{C}^1$ with $\partial f(x) = \big\{\nabla f(x)\big\}$. Then Theorem 2.3 is applicable with T simply a localization of ∇f. Denote by H_a and H_b the matrices that appear in (a) and (b) respectively. Conditions (a) and (b) both imply that f has a second-order expansion at $\hat{x}$ with $h(\xi) = \langle \xi, H_a \xi\rangle = \langle \xi, H_b \xi\rangle$. Theorem 3.1 tells us that f is twice epi-differentiable at $\hat{x}$ for $\hat{v} = \nabla f(\hat{x})$ with $f''_{\hat{x},\hat{v}}(\xi) = h(\xi)$. This in turn means by Theorem 2.3 that

$$H_b(\xi) = \nabla^2 f(\hat{x})(\xi) = \big(\partial f\big)'_{\hat{x},\hat{v}}(\xi) = \partial\big[\tfrac{1}{2} f''_{\hat{x},\hat{v}}\big](\xi) = H_a(\xi) = \tfrac{1}{2}\big[H_b + {H_b}^{T}\big](\xi) \text{ for all } \xi$$

(recall that H_a is symmetric). This yields the equivalence of (a) and (b), while (c) just restates (b). $\square$

Theorem 3.2 is not new—it has recently been proved even in a Banach-space version by Borwein and Noll [2](Theorem 3.1) for convex functions f, from which it readily extends to lower-$\mathcal{C}^2$ functions. Our proof is different, however, in its reliance on fundamental relationships between second-order epi-derivatives of functions and proto-derivatives of their subgradient mappings.

Corollary 3.3. *For a function f of the kind in Theorem 3.2, and in particular for any differentiable convex function f, twice differentiability at $\hat{x}$ automatically implies the symmetry of the second-derivative matrix $\nabla^2 f(\hat{x})$.*

This fact is interesting because second-derivative matrices do not have to be symmetric for differentiable functions in general. The standard criterion for ensuring symmetry is continuous twice differentiability of f, but that is not assumed in the situation at hand. Also, apart from the symmetry issue, the existence of a Hessian H for a differentiable function f does not automatically mean, unless f is lower-C^2 as in Theorem 3.2, the existence of the second-derivative matrix $\nabla^2 f(\hat{x})$. This is confirmed by the example of

$$f(x) = \begin{cases} |x|^2 \sin(1/|x|^2) & \text{for } x \neq 0, \\ 0 & \text{for } x = 0, \end{cases}$$

which is everywhere differentiable and has Hessian $H = 0$ at $\hat{x} = 0$, but has ∇f not even bounded on a neighborhood of 0, much less differentiable there; no Jacobian exists at all for ∇f at 0. These facts highlight the special nature of lower-$\mathcal{C}^2$ functions, and more specifically finite convex functions, in the absence of such pathology.

Our basic result about second-order expansions of envelope functions will involve first-order expansions of their gradient mappings. For this we need another definition.

Definition 3.4. *A single-valued mapping G from an open neighborhood of $\hat{x} \in I\!R^n$ into $I\!R^m$ has a first-order expansion at a point $\hat{x} \in O$ if there is a continuous mapping D such that the difference quotient mappings*

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

converge to D uniformly on bounded sets as $t \searrow 0$. The expansion is strict if actually the mappings

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

converge to D uniformly on bounded sets as $t \searrow 0$ and $x \to \hat{x}$.

When a first-order expansion exists, the mapping D giving the approximating term must not only be continuous but positively homogeneous: $D(\lambda\xi) = \lambda D(\xi)$ for $\lambda > 0$, and in particular $D(0)=0$. The expansion can be expressed by

$$G(\hat{x}+t\xi) \;=\; G(\hat{x})+tD(\xi)+o\big(|t\xi|\big).$$

Differentiability of G at $\hat{x}$ in the classical sense is the case where D happens to be a linear mapping. If D is linear and the expansion is strict, one has the *strict differentiability* of G at $\hat{x}$.

In general, the existence of a first-order expansion means the *directional differentiability* of G at $\hat{x}$. It is equivalent (in our finite-dimensional setting) to the existence, for every vector $\hat{\xi}\in\mathbb{R}^n$, of the *directional derivative* limit

$$\lim_{\substack{\xi\to\hat{\xi}\\ t\searrow 0}} \frac{G(\hat{x}+t\xi)-G(\hat{x})}{t}.$$

Likewise, a strict first-order expansion means *strict* directional differentiability at $\hat{x}$ and corresponds to the existence for every $\hat{\xi}$ of the more complicated limit where $\hat{x}$ is replaced by x, and $x\to\hat{x}$ along with $\xi\to\hat{\xi}$ and $t\searrow 0$. Either way, the mapping D in Definition 3.4 gives for each $\hat{\xi}$ the directional derivative $D(\hat{\xi})$.

The existence of a first-order expansion can also be identified with a special case of proto-differentiability (a concept applicable to single-valued mappings in particular). Indeed, G has a first-order expansion at $\hat{x}$ in terms of D if and only if G is proto-differentiable at $\hat{x}$ and the proto-derivative mapping $G'_{\hat{x}}$ has no empty values. Then $D=G'_{\hat{x}}$, so we get

$$G(\hat{x}+t\xi) \;=\; G(\hat{x})+tG'_{\hat{x}}(\xi)+o\big(|t\xi|\big).$$

Theorem 3.5. *Suppose that f is prox-regular at $\bar{x}\in\operatorname{argmin} f$ for $\bar{v}=0$ with respect to ε and r, and let $\lambda\in(0,1/r)$. Then the following properties are equivalent:*

(a) *f is twice epi-differentiable at $\bar{x}$ for $\bar{v}=0$;*

(b) *e_λ is twice epi-differentiable at $\bar{x}$ for $\bar{v}=0$;*

(c) *e_λ has a second-order expansion at $\bar{x}$;*

(d) *∇e_λ has a first-order expansion at $\bar{x}$;*

(e) *∇e_λ is proto-differentiable at $\bar{x}$;*

(f) *P_λ has a first-order expansion at $\bar{x}$;*

(g) *P_λ is proto-differentiable at $\bar{x}$.*

In that event the expansions

$$\begin{aligned}
e_\lambda(\bar{x}+t\xi) \;&=\; e_\lambda(\bar{x})+t^2 d_\lambda(\xi)+o\big(|t\xi|^2\big),\\
\nabla e_\lambda(\bar{x}+t\xi) \;&=\; t\nabla d_\lambda(\xi)+o\big(|t\xi|\big),\\
P_\lambda(\bar{x}+t\xi) \;&=\; \bar{x}+t\big[I-\lambda\nabla d_\lambda(\xi)\big]+o\big(|t\xi|\big),
\end{aligned}$$

hold for a function d_λ that is both $\mathcal{C}^{1+}$ and lower-$\mathcal{C}^2$, the gradient mapping ∇d_λ being Lipschitz continuous globally. In fact d_λ is the Moreau λ-envelope for $\frac{1}{2}f''_{\bar{x},\bar{v}}$; one has

$$d_\lambda(\xi) \;=\; \min_{\xi'}\left\{\frac{1}{2}f''_{\bar{x},\bar{v}}(\xi') + \frac{1}{2\lambda}\big|\xi' - \xi\big|^2\right\} \;\text{ for all } \xi. \tag{3.2}$$

Moreover

$$\partial\big[\tfrac{1}{2}d_\lambda\big] \;=\; \big[\lambda I + D^{-1}\big]^{-1} = \lambda^{-1}\big[I - (I+\lambda D)^{-1}\big] \;\text{ for }\; D := \partial\big[\tfrac{1}{2}f''_{\bar{x},\bar{v}}\big]. \tag{3.3}$$

Proof. The equivalence of (a) and (b) is asserted along with (3.2) and (3.3) by [25](Cor. 6.6). On the other hand, the equivalence of (b) and (c) follows from applying Theorem 3.1 to e_λ in place of f, which is possible because of the properties of e_λ in Theorem 2.4(b). In the same context, (b) is equivalent to (e) through Theorem 2.3.

As noted prior to the statement of the theorem, the existence of a first-order expansion is a special case of proto-differentiability; (d) implies (e). Conversely, the local Lipschitz continuity of ∇e_λ guarantees that the difference quotient mappings

$$\xi \mapsto \big[\nabla e_\lambda(\bar{x} + t\xi) - \nabla e_\lambda(\bar{x})\big]/t$$

enjoy the same Lipschitz constant on some neighborhood of $\xi = 0$ for sufficiently small $t > 0$, so that graphical convergence of these mappings, as demanded by proto-differentiability, is impossible unless they converge pointwise. But uniformly Lipschitz continuous mappings that converge pointwise must converge uniformly on bounded sets, moreover to a limit mapping that likewise is single-valued Lipschitz continuous with the same constant on a neighborhood of 0. In this case the limit mapping is positively homogeneous, so the latter properties hold not just on a neighborhood of 0 but globally on $\mathbb{R}^n$. Thus, (e) implies (d) and the single-valuedness and global Lipschitz continuity of the proto-derivative $(e_\lambda)'_{\bar{x}}$. The relationships with P_λ come immediately now out of Theorem 2.4. □

Corollary 3.6. *The properties in Theorem 3.5 hold when f is fully amenable at $\bar{x}$.*

Proof. Fully amenable functions are twice epi-differentiable everywhere, cf. [33]. □

Moving on now to the question of the existence and characterization of Hessian expansions, we require two special concepts.

Definition 3.7. *A function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a generalized (purely) quadratic function if it is expressible in the form*

$$h(x) = \begin{cases} \frac{1}{2}\langle x, Qx\rangle & \textit{if } x \in N, \\ \infty & \textit{if } x \notin N, \end{cases} \tag{3.4}$$

where N is a linear subspace of $\mathbb{R}^n$ and Q is a symmetric $n \times n$ matrix. A possibly set-valued mapping $D : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a generalized linear mapping if its graph is a linear subspace of $\mathbb{R}^n \times \mathbb{R}^m$.

Generalized quadratic functions can be interpreted as "quadratic functions with $+\infty$ as a possible eigenvalue" (the subspace $N^\perp$ complementary to N being the eigenspace for that eigenvalue). They are precisely the functions h on $\mathbb{R}^n$ whose subgradient mappings ∂h are generalized linear mappings, the graph of ∂h necessarily being in fact an n-dimensional subspace of $\mathbb{R}^n \times \mathbb{R}^n$; see [32](Section 4).

Theorem 3.8. *Suppose that f is prox-regular at $\bar{x} \in \operatorname{argmin} f$ for $\bar{v} = 0$ with respect to ε and r, and let $\lambda \in (0, 1/r)$. Then the following properties are equivalent:*

(a) *f is twice epi-differentiable at $\bar{x}$ for $\bar{v}$, and $f''_{\bar{x},\bar{v}}$ is generalized quadratic;*

(b) *e_λ has a Hessian matrix H_λ at $\bar{x}$;*

(c) *∇e_λ is differentiable at $\bar{x}$ with Jacobian matrix H_λ;*

(d) *e_λ is twice differentiable at $\bar{x}$ for $\bar{v}$, and $H_\lambda = \nabla^2 e_\lambda(\bar{x})$;*

(e) *P_λ is differentiable at $\bar{x}$ with Jacobian matrix $I - \lambda H_\lambda$;*

(f) *the f-attentive ε-localization T of ∂f at $(\bar{x}, \bar{v})$ is proto-differentiable there, and the proto-derivative mapping $T'_{\bar{x},\bar{v}}$ is generalized linear.*

Proof. We have (b), (c) and (e) equivalent as the special case of Theorem 3.5(a)(b)(f) in which d_λ is quadratic. Next, (d) is just another way of saying (c), which is also the same as asserting that the mapping ∇e_λ is proto-differentiable at $\bar{x}$ for $\bar{v} = 0$ and the graph of the proto-derivative mapping is a linear subspace of $\mathbb{R}^n$ (cf. the remarks about proto-differentiability in the proof of Theorem 3.5). Through the last formula in Theorem 2.4, which sets up a linear transformation in $\mathbb{R}^n \times \mathbb{R}^n$ under which the graph of ∇e_λ corresponds locally to that of T, proto-differentiability of ∇e_λ corresponds to that of T. Thus, (c) is equivalent to (f). But by Theorem 2.3, (f) is equivalent to (a). $\square$

Theorem 3.9. *Suppose that f is prox-regular at $\bar{x} \in \operatorname{argmin} f$ for $\bar{v} = 0$ with respect to ε and r. Then there exist $\lambda_0 > 0$ and a neighborhood U of $(\bar{x}, \bar{v})$ such that for all $0 < \lambda < \lambda_0$ and $(x, v) \in \operatorname{gph} T \cap U$ (with T the f-attentive ε-localization of ∂f at $(\bar{x}, \bar{v})$) the following properties are equivalent:*

(a) *f is twice epi-differentiable at x for v, and $f''_{x,v}$ is generalized quadratic;*

(b) *e_λ has a Hessian matrix H_λ at $x + \lambda v$;*

(c) *∇e_λ is differentiable at $x + \lambda v$ with Jacobian matrix H_λ;*

(d) *e_λ is twice differentiable at $x+\lambda v$ for v, and $H_\lambda = \nabla^2 e_\lambda(x+\lambda v)$;*

(e) *P_λ is differentiable at $x+\lambda v$ with Jacobian matrix $I - \lambda H_\lambda$;*

(f) *the f-attentive ε-localization T of ∂f at (x,v) is proto-differentiable there, and the proto-derivative mapping $T'_{x,v}$ is generalized linear.*

Proof. We can assume that $\bar{x}=0$, $f(\bar{x})=0$. Fix x and $v \in \partial f(x)$ with $|x|<\varepsilon/4$, $|v|<\varepsilon$ and $|f(x)|<\varepsilon$. We claim that by choosing $R>r$ sufficiently large, we can make the function $\bar{f}(x') := f(x') - \langle v, x'-x\rangle + (R/2)|x'-x|^2$ have $x \in \operatorname{argmin} \bar{f}$. To see this, notice that because f is prox-regular at 0 for 0 we have for any R and any x' with $|x'|<\varepsilon$ that

$$\begin{aligned}\bar{f}(x') &\geq f(x) + \langle v, x'-x\rangle - (r/2)|x'-x|^2 - \langle v, x'-x\rangle + (R/2)|x'-x|^2\\ &= f(x) + \big((R-r)/2\big)|x'-x|^2 \geq f(x) = \bar{f}(x).\end{aligned}$$

On the other hand we have for any R and any x' with $|x'| \geq \varepsilon$ (and hence in particular with $|x'-x| \geq (3/4)\varepsilon$) that

$$\begin{aligned}\bar{f}(x') &\geq f(0) - \langle v, x'-x\rangle + (R/2)|x'-x|^2 \;=\; f(x) - f(x) - \langle v, x'-x\rangle + (R/2)|x'-x|^2\\ &\geq f(x) + (R/2)|x'-x|^2 - \varepsilon|x'-x| - \varepsilon.\end{aligned}$$

This last quantity is greater that $f(x)$ provided $|x'-x| \geq (1/R)\Big[\varepsilon + \sqrt{\varepsilon^2 + 2R\varepsilon}\;\Big]$. By choosing R large enough we can have this last quantity less than $(3/4)\varepsilon$. Then $\bar{f}(x') \geq \bar{f}(x)$ for all x', so $x \in \operatorname{argmin} \bar{f}$.

With R having been chosen large enough that $x \in \operatorname{argmin} \bar{f}$, and with the observation that $\bar{f}$ is prox-regular at x for 0, we now turn our attention to the Moreau λ-envelope of $\bar{f}$, which we denote by $\bar{f}_\lambda$. This is given by

$$\begin{aligned}\bar{f}_\lambda(w) &= \inf_{x'}\left\{g(x') + \frac{1}{2\lambda}|x'-w|^2\right\}\\ &= \inf_{x'}\left\{f(x') - \langle v, x'-x\rangle + \frac{R}{2}|x'-x|^2 + \frac{1}{2\lambda}|x'-w|^2\right\}\\ &= \inf_{x'}\left\{f(x') + \frac{1}{2\mu}|u-x'|^2\right\} - \frac{1}{2\mu}|u|^2 + \frac{1}{2\lambda}|w|^2 + \langle v, x\rangle + \frac{R}{2}|x|^2\\ &= e_\mu(u) - \frac{1}{2\mu}|u|^2 + \frac{1}{2\lambda}|w|^2 + \langle v, x\rangle + \frac{R}{2}|x|^2,\end{aligned}$$

where $\mu = \lambda/(\lambda R + 1)$ and $u = \mu[v + Rx + (1/\lambda)w]$; to obtain the penultimate step in the preceding calculation simply note that

$$\frac{1}{2\mu}|u-x'|^2 = \frac{1}{2\mu}|u|^2 - \langle v + Rx + (1/\lambda)w, x'\rangle + \big(\frac{R}{2} + \frac{1}{2\lambda}\big)|x'|^2.$$

From the formula for $\bar{f}_\lambda$ we see that the second-order properties of $\bar{f}_\lambda$ at x for 0 translate to second-order properties of e_μ at $\mu[v + Rx + (1/\lambda)x] = \mu v + x$ for v (since by Theorem 2.4, $\nabla e_\mu(x + \mu v) = v$). Besides, $\bar{f}$ is twice epi-differentiable at x for 0 if and only if f is twice epi-differentiable at x for v.

Now we apply the equivalent properties in Theorem 3.8 not just at $(\bar{x}, \bar{v})$ but locally.

$\square$

Corollary 3.10. *Suppose that f is prox-regular at a point $\hat{x}$ for $\hat{v}$ with constants ε and r. Then there is a neighborhood U of $(\hat{x}, \hat{v})$ such that for all $(x, v) \in \operatorname{gph} T \cap U$ (with T the f-attentive ε-localization of ∂f at $(\hat{x}, \hat{v})$) the following properties are equivalent:*

(a) *f is twice epi-differentiable at x for v and $f''_{x,v}$ is generalized quadratic;*

(b) *T is proto-differentiable at x for v, and the proto-derivative mapping $T'_{x,v}$ is generalized linear.*

Proof. Take $(x, v) \in \operatorname{gph} T$ with $|x - \hat{x}| < (\varepsilon/2)$. We have $f(x') \geq f(x) + \langle v, x' - x\rangle - (r/2)|x' - x|^2$ whenever $|x' - \hat{x}| < \varepsilon$. Therefore, for some neighborhood U of x the function $g(x') := f(x') - \langle v, x' - x\rangle + (r/2)|x' - x|^2 + \delta_U(x')$ has $x \in \operatorname{argmin} g$ and $0 \in \partial g(x)$. In addition, the function g is prox-regular at x for 0. Now apply Theorem 3.8 to g at x for 0; the equivalent properties for g easily translate into equivalent properties for f. $\square$

Corollary 3.11. *Suppose that f is prox-regular at $\bar{x} \in \operatorname{argmin} f$ for $\bar{v} = 0$ with respect to ε and r.*

(a) *If for some $\lambda \in (0, 1/r)$ the function e_λ has a Hessian matrix at $\bar{x}$, this must be true for every $\lambda \in (0, 1/r)$.*

(b) *If for some $\lambda \in (0, 1/r)$ the mapping ∇e_λ is differentiable at $\bar{x}$, this must be true for every $\lambda \in (0, 1/r)$.*

(c) *If for some $\lambda \in (0, 1/r)$ the mapping P_λ is differentiable at $\bar{x}$, this must be true for every $\lambda \in (0, 1/r)$.*

Corollary 3.12. *Suppose f is lower-$\mathcal{C}^2$ on a neighborhood of $\bar{x} \in \operatorname{argmin} f$ and there exists $s > 0$ such that $f(x) \leq f(\bar{x}) + s|x - \bar{x}|^2$ for all x in some neighborhood of $\bar{x}$. Then for $\lambda > 0$ sufficiently small, e_λ has a Hessian at $\bar{x}$ if and only if f itself has a Hessian at $\bar{x}$.*

Proof. The bound given on the growth of f around $\bar{x}$ has the consequence that if the second-order epi-derivative function $f''_{\bar{x},\bar{v}}$ exists it must be finite. $\square$

4. Stability of Approximation

The existence of a second-order expansion of e_λ at $\bar{x}$ provides a type of second-order approximation which may be useful for a number of purposes. To get full benefit from such an approximation in a numerical context, however, it must display some degree of stability. In a standard framework of analysis relying on differentiability, the property to focus on is that of e_λ being of class $\mathcal{C}^2$ on a neighborhood of $\bar{x}$; then Hessian matrices exist at points x near $\bar{x}$ and depend continuously on x. We must ascertain what property of f is necessary and sufficient for this, but also try to explore less demanding notions of stability which might nonetheless provide a numerical handle in the environment where e_λ is not $\mathcal{C}^2$, just C^{1+} and lower-$\mathcal{C}^2$ around $\bar{x}$ (through Theorem 2.4).

Strict second-order expansions of e_λ (Definition 1.1) and strict first-order expansions of ∇e_λ (Definition 3.4) offer promising territory. Strict differentiability, as the case of a strict first-order expansion where the approximating mapping is linear, is known to have the interpretation as the single-point localization of the $\mathcal{C}^1$ property, and something similar might be anticipated for the existence of a strict Hessian in relation to the $\mathcal{C}^2$ property. In exploring this, we appeal to the strict versions of proto-differentiability and twice epi-differentiability defined before Theorem 2.3.

Theorem 4.1. *Suppose that f is prox-regular at $\bar{x} \in \operatorname{argmin} f$ for $\bar{v} = 0$ with respect to ε and r, and let $\lambda \in (0, 1/r)$. Then the following properties are equivalent and imply that $f''_{\bar{x},\bar{v}}$ is generalized quadratic and $T'_{\bar{x},\bar{v}}$ is generalized linear:*

(a) *f is strictly twice epi-differentiable at $\bar{x}$ for $\bar{v}$;*

(b) *e_λ has a strict Hessian at $\bar{x}$;*

(c) *∇e_λ is strictly differentiable at $\bar{x}$;*

(d) *e_λ is twice differentiable at $\bar{x}$, and $\nabla^2 e_\lambda(x) \to \nabla^2 e_\lambda(\bar{x})$ as $x \to \bar{x}$ in the set of points where e_λ is twice differentiable;*

(e) *e_λ is strictly twice epi-differentiable at $\bar{x}$ for $\bar{v}$;*

(f) *∇e_λ is strictly proto-differentiable at $\bar{x}$ for $\bar{v}$;*

(g) *P_λ is strictly differentiable at $\bar{x}$;*

(h) *P_λ is strictly proto-differentiable at $\bar{x}$;*

(i) *the f-attentive ε-localization T of ∂f at $(\bar{x}, \bar{v})$ is strictly proto-differentiable there.*

Proof. Theorem 2.3 as applied to e_λ in the context of the properties of this function in Theorem 2.4(b) yields the equivalence of (b) with (c) and more generally that of (e) with (f). In principle, (f) is a weaker property than (c), but we shall argue next that, in this context, (f) implies (c) through (d).

The linear mappings $\xi \mapsto \nabla^2 e_\lambda(x)\xi$ at points x where e_λ is twice differentiable are limits of difference quotient mappings for ∇e_λ at such points, which in particular can be construed as graphical limits. The definition of strict proto-differentiability of ∇e_λ at $\bar{x}$ ensures (through diagonalization) that such linear mappings must graph-converge to the proto-derivative mapping $(\nabla e_\lambda)'_{\bar{x}}$ as $x \to \bar{x}$. Graphical convergence of linear mappings coincides with pointwise convergence of such mappings and requires the limit to be linear; thus it corresponds to convergence of the coefficient matrices. Therefore (f) implies (d).

When (d) holds, however, powerful information is provided through the fact in Theorem 2.4(b) that the mapping ∇e_λ is Lipschitz continuous around $\bar{x}$, hence differentiable almost everywhere in some neighborhood of $\bar{x}$ (Rademacher's Theorem). This means that e_λ is twice differentiable at almost every x in some neighborhood of $\bar{x}$. The norms $\left|\nabla^2 e_\lambda(x)\right|$ are uniformly bounded in such a neighborhood. Indeed, any upper bound to these norms serves as a local Lipschitz constant for ∇e_λ, and vice versa. In this framework we see that for the locally Lipschitz continuous mapping $D_\lambda(x) = \nabla e_\lambda(x) - \nabla e_\lambda(\bar{x})$ condition (c) implies the existence for any $\varepsilon > 0$ of $\delta > 0$ such that ε serves as a Lipschitz constant for D_λ on the δ-ball around $\bar{x}$. But this is equivalent to D_λ being strictly differentiable at $\bar{x}$ with the zero matrix as its Jacobian. Hence (d) implies (c).

So far we have established the equivalence of conditions (b)–(f). Clearly (g) and (h) are equivalent to (b) and (f) through the formula in Theorem 2.4 connecting P_λ with ∇e_λ. On the other hand, (a) and (i) are equivalent by Theorem 2.3. It remains only to argue that (i) corresponds to (f). This is true because strict proto-differentiability is a geometric property of the graph of a mapping and is preserved under any linear transformation of the graph. The formula connecting ∇e_λ with T in Theorem 2.4 sets up an invertible linear transformation between the graphs of these mappings.

Finally note that the equivalence of (g) and (h) shows that for the mapping P_λ strict proto-differentiability automatically entails generalized linearity (actually true linearity) of the proto-derivative mapping. Since the graphs of T and P_λ are "isomorphic" under a certain linear transformation in graph space, the same must hold for T as well. Once we know that $T'_{\bar{x},\bar{v}}$ is generalized linear, we have that $f''_{\bar{x},\bar{v}}$ is generalized quadratic. □

Theorem 4.2. *Under the hypothesis of Theorem 4.1, and with T the f-attentive ε-localization of ∂f at $(\bar{x}, \bar{v})$, the following conditions can be added to the list of equivalences:*

(a) *$f''_{x,v}$ epi-converges (to something) as $(x, v) \to (\bar{x}, \bar{v})$ with $f(x) \to f(\bar{x})$ in the set of pairs (x, v) with $v \in \partial f(x)$ for which f is twice epi-differentiable;*

(b) $(e_\lambda)''_x$ *epi-converges as* $x \to \bar{x}$ *in the set where* e_λ *is twice epi-differentiable;*

(c) $(\nabla e_\lambda)'_x$ *graph-converges as* $x \to \bar{x}$ *in the set where* ∇e_λ *is proto-differentiable;*

(d) $(\nabla P_\lambda)'_x$ *graph-converges as* $x \to \bar{x}$ *in the set where* P_λ *is proto-differentiable;*

(e) $T'_{x,v}$ *graph-converges as* $x \to \bar{x}$ *and* $v \to \bar{v}$ *in the set of pairs* (x, v) *with* $v \in T(x)$ *for which* T *is proto-differentiable;*

(f) $f''_{x,v}$ *epi-converges as* $(x, v) \to (\bar{x}, \bar{v})$ *with* $f(x) \to f(\bar{x})$ *in the set of pairs* (x, v) *with* $v \in \partial f(x)$ *for which* f *is twice epi-differentiable and* $f''_{x,v}$ *is generalized quadratic;*

(g) $T'_{x,v}$ *graph-converges as* $x \to \bar{x}$ *and* $v \to \bar{v}$ *in the set of pairs* (x, v) *with* $v \in T(x)$ *for which* T *is proto-differentiable and* $T'_{x,v}$ *is generalized linear.*

Proof. The proof of Theorem 4.1 already shows that (c) can be added to the list of equivalences. We have (c) equivalent to (d) and (e) because of the formulas relating these mappings in Theorem 2.4: not only proto-differentiability but proto-derivatives correspond when the graphs of mappings can be identified under an invertible linear transformation. Under our assumption that f is prox-regular at $\bar{x}$ for $\bar{v}$ with respect to ε and r, we know that $T + rI$ is monotone (Theorem 2.2), so that $T'_{x,v} + rI$ is monotone when the proto-derivative exists. But by Theorem 2.3, $T'_{x,v}$ is the subgradient mapping for $\frac{1}{2}\left(f''_{x,v} + r|\cdot|^2\right)$ so its monotonicity implies the convexity of $\frac{1}{2}\left(f''_{x,v} + r|\cdot|^2\right)$ [19]. Graphical convergence of $T'_{x,v} + rI$ as $(x, v) \in (\bar{x}, \bar{v})$ with $v \in T(x)$ is equivalent then by Attouch's theorem to epi-convergence of $\frac{1}{2}\left(f''_{x,v} + r|\cdot|^2\right)$ as $(x, v) \in (\bar{x}, \bar{v})$ with $v \in T(x)$; the latter is the same as having $(x, v) \in (\bar{x}, \bar{v})$ with $v \in \partial f(x)$ and $f(x) \to f(\bar{x})$ by [25](Prop. 2.3). Hence (e) is equivalent to (a).

Similarly by Attouch's theorem, we have the equivalence of (c) with (b) because this likewise can be reduced to the convex case because of the convexity property asserted in Theorem 2.4(b). Finally, (f) and (g) come into the picture as viable substitutes for (a) and (e), respectively, because the graph of T is a Lipschitz manifold of dimension n around $(\bar{x}, \bar{v})$ [25](Thm. 4.7) and consequently has a linear tangent space at almost every pair (x, v) in some neighborhood of $(\bar{x}, \bar{v})$, cf. [32]. At such (x, v), this subspace is the graph of $T'_{x,v}$, which therefore is generalized linear, and the function $\frac{1}{2}f''_{x,v}$ having $T'_{x,v}$ as its subgradient mapping is accordingly a generalized quadratic function. $\square$

Again we note, as in Corollary 3.11, that if one of the λ-dependent properties in Theorems 4.1 and 4.2 holds for some $\lambda \in (0, 1/r)$, they all hold for *every* $\lambda \in (0, 1/r)$.

Corollary 4.3. *Suppose that* f *is prox-regular at a point* $\hat{x}$ *for* $\hat{v}$ *with constants* ε *and* r. *Then there is a neighborhood* U *of* $(\hat{x}, \hat{v})$ *such that for all* $(x, v) \in \operatorname{gph} T \cap U$ *(where* T *is* *the f-attentive ε-localization of ∂f at $(\hat{x}, \hat{v})$) the following properties are equivalent and imply that $f''_{x,v}$ is generalized quadratic and $T'_{x,v}$ is generalized linear:*

(a) *f is strictly twice epi-differentiable at x for v*

(b) *T is strictly proto-differentiable at x for v;*

(c) *$f''_{x',v'}$ epi-converges (to something) as $(x', v') \to (x, v)$ with $f(x') \to f(x)$ in the set of pairs (x', v') with $v' \in \partial f(x')$ for which f is twice epi-differentiable;*

(d) *$T'_{x',v'}$ graph-converges (to something) as $x' \to x$ and $v' \to v$ in the set of pairs (x', v') with $v' \in T(x')$ for which T is proto-differentiable;*

(e) *$f''_{x',v'}$ epi-converges (to something) as $(x', v') \to (x, v)$ with $f(x') \to f(x)$ in the set of pairs (x', v') with $v' \in \partial f(x')$ for which f is twice epi-differentiable and $f''_{x',v'}$ is generalized quadratic;*

(f) *$T'_{x',v'}$ graph-converges (to something) as $x' \to x$ and $v' \to v$ in the set of pairs (x', v') with $v' \in T(x')$ for which T is proto-differentiable and $T'_{x',v'}$ is generalized linear.*

Proof. The proof is similar to that of Corollary 3.10 except that at the end we use Theorems 4.1 and 4.2 instead. $\square$

The existence of a strict Hessian for e_λ at $\bar{x}$ is a very natural property to look for in any numerical method for minimizing e_λ that relies on second-order approximations. It is remarkable that this property is equivalent to so many others in Theorems 4.1 and 4.2, especially those of f and the localization T of ∂f. In taking any step further toward stability or continuity properties of generalized derivatives of f it appears one has to pass to the case described in the next theorem, our last.

We have already made use of the fact proved in [25](Thm. 4.7) that when f is prox-regular at $\bar{x}$ for $\bar{v}$, an f-attentive localization of the graph of ∂f around $(\bar{x}, \bar{v})$ is a Lipschitz manifold. Such a manifold, as defined in [32], can be identified locally under some invertible linear transformation from $\mathbb{R}^n \times \mathbb{R}^n$ onto itself with the graph of a Lipschitz continuous mapping. Similarly we can speak of a *smooth manifold* in the case where identification is made with the graph of a $\mathcal{C}^1$ mapping.

Theorem 4.4. *Suppose that f is prox-regular at $\bar{x} \in \operatorname{argmin} f$ for $\bar{v} = 0$ with respect to ε and r. Let T be the f-attentive ε-localization of ∂f at $(\bar{x}, \bar{v})$. Then there exists $\lambda_0 > 0$ such that for all $0 < \lambda < \lambda_0$ the following properties are equivalent;*

(a) *e_λ is $\mathcal{C}^2$ on a neighborhood of $\bar{x}$;*

(b) *e_λ is strictly twice epi-differentiable on a neighborhood of $\bar{x}$;*

(c) P_λ *is* $\mathcal{C}^1$ *on a neighborhood of* $\bar{x}$*;*

(d) P_λ *is strictly proto-differentiable on a neighborhood of* $\bar{x}$*;*

(e) f *is twice epi-differentiable at* x *for* v *with respect to all* (x, v) *sufficiently near to* $(\bar{x}, \bar{v})$ *in the graph of* T*, moreover* $f''_{x,v}$ *depends epi-continuously on* (x, v)*, i.e.,* $f''_{x',v'}$ *epi-converges to* $f''_{x,v}$ *as* $(x', v') \to (x, v)$ *with* $f(x') \to f(x)$ *and* $v' \in \partial f(x')$*;*

(f) f *is strictly twice epi-differentiable at* x *for* v *with respect to all* (x, v) *sufficiently near to* $(\bar{x}, \bar{v})$ *in the graph of* T*;*

(g) *the graph of* T *is a smooth manifold relative to some neighborhood of* $(\bar{x}, \bar{v})$*;*

(h) T *is strictly proto-differentiable at* x *for* v *with respect to all* (x, v) *sufficiently near to* $(\bar{x}, \bar{v})$ *in the graph of* T*.*

Proof. We can assume that $\bar{x} = 0$, $f(\bar{x}) = 0$. Fix x and $v \in \partial f(x)$ with $|x| < \varepsilon/4$, $|v| < \varepsilon$ and $|f(x)| < \varepsilon$. As in the proof of Theorem 3.9, we take $R > r$ sufficiently large so that $x \in \operatorname{argmin} \bar{f}$ where $\bar{f}(x') := f(x') - \langle v, x' - x\rangle + (R/2)|x' - x|^2$. Note that $\bar{f}$ is prox-regular at x for 0. Again from the proof of Theorem 3.9 we have

$$\bar{f}_\lambda(w) = e_\mu(u) - \frac{1}{2\mu}|u|^2 + \frac{1}{2\lambda}|w|^2 + \langle v, x\rangle + \frac{R}{2}|x|^2,$$

where $\mu = \lambda/(\lambda R + 1)$ and $u = \mu[v + Rx + (1/\lambda)w]$.

From the formula for $\bar{f}_\lambda$ we see that $\bar{f}_\lambda$ is strictly twice epi-differentiable at x for 0 if and only if e_μ is strictly twice epi-differentiable at $\mu[v + Rx + (1/\lambda)x] = \mu v + x$ for v (since by Theorem 2.4, $\nabla e_\mu(x + \mu v) = v$). Besides, $\bar{f}$ is strictly twice epi-differentiable at x for 0 if and only if f is strictly twice epi-differentiable at x for v.

Now we apply the equivalent properties in Theorems 4.1 and 4.2 not just at $(\bar{x}, \bar{v})$ but locally, keeping in mind that a mapping is $\mathcal{C}^1$ on an open set if and only if it is strictly differentiable there, and the fact that there is a one-to-one correspondence between the points x' close to 0 and the points $x + \mu v$ with $v \in T(x)$ (by Theorem 2.4). The local identification of the graph of T with a Lipschitz manifold is achieved through parameterization by P_λ, so the smoothness of P_λ under the present circumstances corresponds to having the graph of T be in fact a smooth manifold. $\square$

Corollary 4.5. *If* f *is* $\mathcal{C}^2$ *on a neighborhood of a point* $\bar{x} \in \operatorname{argmin} f$*, then for all* $\lambda > 0$ *sufficiently small, the function* e_λ *is* $\mathcal{C}^2$ *on a neighborhood of* $\bar{x}$*.*

Corollary 4.6. *Suppose* f *is lower-*$\mathcal{C}^2$ *on a neighborhood of a point* $\bar{x} \in \operatorname{argmin} f$ *and there exists* $s > 0$ *such that* $f(x) \le f(\bar{x}) + s|x - \bar{x}|^2$ *for all* x *in some neighborhood of* $\bar{x}$*. Then for* $\lambda > 0$ *sufficiently small,* e_λ *is* $\mathcal{C}^2$ *around* $\bar{x}$ *if and only if* f *itself is* $\mathcal{C}^2$ *around* $\bar{x}$*.*

Proof. By adding $s_0|x-\bar{x}|^2$ to f for sufficiently high s_0, we can reduce to the case where f is actually convex on a neighborhood of $\bar{x}$. As already observed in the proof of Corollary 3.12, the bound on the growth of f around $\bar{x}$ has ensures that if the second-order epi-derivative function $f''_{\bar{x},\bar{v}}$ exists it must be finite. But when f is convex, its derivatives $f''_{x,v}$ are convex as well. Such a function, being positively homogeneous of degree 2, is finite everywhere if it is finite on a neighborhood of 0. When convex functions epi-converge to a convex function that is finite on a neighborhood of 0, they must eventually themselves be finite on a neighborhood of 0, cf. [38]. In the setting of Theorem 4.4(c), therefore, we must have finite quadratic, convex functions that depend epi-continuously on (x, v). In light of Theorem 3.1, this is the same as f being $\mathcal{C}^2$ around $\bar{x}$. ☐

Corollary 4.7. *A function f is $\mathcal{C}^2$ on a neighborhood of a point $\hat{x}$ if and only if f is prox-regular at $\hat{x}$ and there exist $\hat{v} \in \partial f(\hat{x})$ and $\varepsilon > 0$ such that, whenever*

$$v \in \partial f(x) \text{ with } |v-\hat{v}| < \varepsilon, \quad |x-\hat{x}| < \varepsilon, \quad f(x) + f(\hat{x}) + \varepsilon,$$

f is strictly twice epi-differentiable at x for v, the function $f''_{\hat{x},\hat{v}}$ itself being finite. Then the relation $v \in \partial f(x)$ reduces locally to $v = \nabla f(x)$, while $f''_{x,v}(\xi) = \langle \xi, \nabla^2 f(x)\xi \rangle$.

Proof. The necessity is obvious. For the sufficiency we can use the fact that $\hat{v}$ is in particular a proximal subgradient of f at $\hat{x}$ to add a quadratic function to f so as to reduce to the case where $\hat{x} \in \operatorname{argmin} f$ and $\hat{v} = 0$. Then the equivalences in Theorem 4.4 come into force. The finiteness of $f''_{\hat{x},\hat{v}}$ puts us at the same time in the picture of Theorems 3.1 and 3.2; the generalized quadratic functions $f''_{x,v}$ for (x, v) near enough to $(\hat{x}, \hat{v})$ have to be ordinary quadratic functions. ☐

References

1. J. F. Bonnans, J.-Ch. Gilbert, C. Lemaréchal and C. A. Sagastizábal, "A family of variable metric proximal methods," Math. Prog. **68** (1995), 15–47.
2. J. M. Borwein and D. Noll, "Second-order differentiability of convex functions in Banach spaces," Trans. Amer. Math. Soc. **342** (1994), 43–81.
3. J. V. Burke and R. A. Poliquin, "Optimality conditions for non-finite convex composite functions," Mathematical Programming **57** (1992), 103–120.
4. F. H. Clarke, "Generalized gradients and applications," Trans. Amer. Math. Soc. **205** (1975), 247–262.
5. X. Chen, L. Qi and R. Womersley, "Newton's method for quadratic stochastic programs with recourse," J. Computational Appl. Math., forthcoming.

6. R. Cominetti and R. Correa, "A generalized second-order derivative in nonsmooth optimization," SIAM J. Control and Optimization **28** (1990), 789–809.
7. S. Dolecki, G. Salinetti and R. J.-B. Wets, "Convergence of functions: equi-semicontinuity," Trans. Amer. Math. Soc. **276** (1983), 409–429.
8. F. Facchinei, "Minimization of SC^1 functions and the Maratos effect," Oper. Res. Letters, forthcoming.
9. M. Fukushima and L. Qi, "A globally and superlinearly convergent algorithm for nonsmooth convex minimization," AMR 95/1 Applied Math. Report, Univ. of New South Wales, 1995.
10. J.-B. Hiriart-Urruty, "Characterization of the plenary hull of the generalized Jacobian matrix," Math. Prog. Study **17** (1982), 1–12.
11. A. D. Ioffe, "Variational analysis of a composite function: a formula for the lower second order epi-derivative," J. Math. Anal. Appl. **160** (1991), 379–405.
12. V. Jeyakumar and X. Q. Yang, "Convex composite minimization with $C^{1,1}$ functions," J. Optim. Theory Appl., forthcoming.
13. C. Lemaréchal and C. Sagastizábal, "Practical aspects of the Moreau-Yosida regularization I: Theoretical properties," preprint 1994.
14. A. Levy, R. A. Poliquin and L. Thibault, "Partial extension of Attouch's theorem with applications to proto-derivatives of subgradient mappings," Trans. Amer. Math. Soc. **347** (1995), 1269–1294.
15. A. Levy and R. T. Rockafellar, "Variational conditions and the proto-differentiation of partial subgradient mappings," Nonlinear Anal. Th. Meth. Appl., forthcoming.
16. P. D. Loewen, *Optimal Control via Nonsmooth Analysis*, CRM-AMS Lecture Notes in Mathematics 2, Amer. Math. Soc., 1993.
17. Z. Páles and V. Zeidan, "Generalized Hessian for $\mathcal{C}^{1+}$ functions in infinite dimensional normed spaces," preprint 1994.
18. J. S. Pang and L. Qi, "A globally convergent Newton method for convex SC^1 minimization problems," J. Optim. Theory Appl., forthcoming.
19. R. A. Poliquin, "Subgradient monotonicity and convex functions," Nonlinear Analysis, Theory, Methods & Applications, **14** (4) (1990), 305-317.
20. R. A. Poliquin, "Integration of subdifferentials of nonconvex functions," Nonlinear Anal. Th. Meth. Appl. **17** (1991), 385–398.
21. R. A. Poliquin, "An extension of Attouch's Theorem and its application to second-order epi-differentiation of convexly composite functions," Trans. Amer. Math. Soc.

332 (1992), 861–874.

22. R. A. Poliquin and R. T. Rockafellar, "Amenable functions in optimization," Nonsmooth Optimization Methods and Applications F. Giannessi (ed.), Gordon & Breach, Philadelphia, 1992, 338–353.
23. R. A. Poliquin and R. T. Rockafellar, "A calculus of epi-derivatives applicable to optimization," Canadian J. Math. **45** (4) (1993), 879–896.
24. R. A. Poliquin and R. T. Rockafellar, "Proto-derivative formulas for basic subgradient mappings in mathematical programming," Set-Valued Analysis **2** (1994), 275–290.
25. R. A. Poliquin and R. T. Rockafellar, "Prox-regular functions in variational analysis," Trans. Amer. Math. Soc., forthcoming.
26. R. A. Poliquin and R. T. Rockafellar, "Second-order nonsmooth analysis in nonlinear programming," in *Recent Advances in Optimization* (D. Du, L. Qi and R. Womersley, eds.), World Scientific Publishers, 1995.
27. R. A. Poliquin, J. Vanderwerff and V. Zizler, "Convex composite representation of lower semicontinuous functions and renormings," C. R. Acad. Sci. Paris **317** Série I (1993), 545–549.
28. L. Qi, "Superlinearly convergent approximate Newton methods for LC^1 optimization problems," Math. Prog. **64** (1994), 277–294.
29. L. Qi and R. Womersley, "An SQP algorithm for extended linear-quadratic problems in stochastic programming," Annals of O. R., forthcoming.
30. L. Qi, "Second-order analysis of the Moreau-Yosida approximation of a convex function," AMR 94/20, Applied Mathematics Report, Univ. of New South Wales, 1994.
31. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
32. R. T. Rockafellar, "Maximal monotone relations and the second derivatives of nonsmooth functions," Ann. Inst. H. Poincaré: Analyse Non Linéaire **2** (1985), 167–184.
33. R. T. Rockafellar, "First- and second-order epi-differentiability in nonlinear programming," Trans. Amer. Math. Soc. **307** (1988), 75–107.
34. R. T. Rockafellar, "Proto-differentiability of set-valued mappings and its applications in optimization," Analyse Non Linéaire, H. Attouch et al. (eds.), Gauthier-Villars, Paris (1989), 449–482.
35. R. T. Rockafellar, "Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives," Math. of Oper. Research **14** (1989), 462–484.
36. R. T. Rockafellar, "Generalized second derivatives of convex functions and saddle functions," Trans. Amer. Math. Soc. **320** (1990), 810–822.

37. R. T. Rockafellar, "Favorable classes of Lipschitz continuous functions in subgradient optimization," Progress in Nondifferentiable Optimization, E. Nurminski (ed.), IIASA Collaborative Proceedings Series, IIASA, Laxenburg, Austria, 1982, 125–144.

38. G. Salinetti and R. J-B Wets, "On the relation between two types of convergence for convex functions," J. Math. Analysis and Appl. **60** (1977), 211–226.

39. X. Q. Yang and V. Jeyakumar, "Generalized second-order directional derivatives and optimization with $C^{1,1}$ functions," Optimization **26** (1992), 165–185.