Journal of Convex Analysis
Volume 12 (2005), No. 1, 81–93

Relating $\mathcal{U}$-Lagrangians to Second-order Epi-derivatives and Proximal-tracks

Robert Mifflin*
Department of Mathematics, Washington State University,
Pullman, WA 99164-3113, USA
mifflin@math.wsu.edu

Claudia Sagastizábal†
IMPA, Estrada Dona Castorina 110, Jardim Botânico,
Rio de Janeiro RJ 22460-320, Brazil
and INRIA, BP 105, 78153 Le Chesnay, France
sagastiz@impa.br

With this paper, having a somewhat wordy title, we thank Claude Lemaréchal, the man who only speaks in words, for his "corne d'abondance" thereof.

Received November 20, 2003

We make use of $\mathcal{VU}$-space decomposition theory to connect three minimization-oriented objects. These objects are $\mathcal{U}$-Lagrangians obtained from minimizing a function over $\mathcal{V}$-space, proximal points depending on minimization over $\mathbb{R}^n = \mathcal{U} \oplus \mathcal{V}$, and epi-derivatives determined by lower limits associated with epigraphs. We relate second-order epi-derivatives of a function to the Hessian of its associated $\mathcal{U}$-Lagrangian. We also show that the function's proximal points are on a trajectory determined by certain $\mathcal{V}$-space minimizers.

Keywords: $\mathcal{U}$-Lagrangian, proximal point, second-order epi-derivatives

1991 Mathematics Subject Classification: Primary: 90C31, 49J52; Secondary: 65K10, 49J53

1. Introduction and motivation

In his seminal work from 1980 [17, Section 5.5.3 *Information du second ordre*], C. Lémarechal addressed the question of defining generalized second-order objects for functions lacking second-order derivatives:

> *Existe-t-il une géneralisation adéquate de la notion de Hessien?... Cette question est la plus passionnante qui se pose actuellement, et une réponse satisfaisante marquerait probablement pour longtemps une étape décisive dans les recherches fondamentales en programmation mathématique.*[1]

Indeed, the need for defining second-order objects for lower semicontinuous (lsc) functions appears both for theoretical and algorithmic reasons. An important theoretical example is given by "second-order" optimality conditions, such as in [3], [13], [14], [41]. As for

*Research supported by the National Science Foundation under Grant No. DMS-0071459
†Research supported by CNPq (Brazil) under Grant No. 303540/03-6
[1] "Does an adequate generalization for the notion of a Hessian exist?... This is today's most interesting question, to which a satisfactory answer would probably start a new, long-lasting and decisive era for basic research in mathematical programming."

ISSN 0944-6532 / \$ 2.50 © Heldermann Verlag

algorithmic reasons, they are essentially related to the extension of Newton-like methods for minimization of functions that are not continuously differentiable, [20], [1], [37], [5], [8], [36], [24], [30], [2], [22], [38].

Second-order Nonsmooth Analysis is a vast and complex subject. Without going into details, we mention here that, depending on the choice of tangent cone and/or convergence notion, it is possible to define B-derivatives [39]; proto-derivatives [40], epi-derivatives; pseudo-derivatives [4], [7]; second-order sub-derivatives, [16]; graphical derivatives; sub-Hessians, [33]; as well as other second-order objects. Chapter 13 in [42] gives an exhaustive presentation and unification of these (many) concepts.

In this paper, we focus on another approach, that was suggested early on in [18] and [23]. Consider the graph of a convex function f, near a point $\bar{x} \in \mathbb{R}^n$. Two distinctively different situations, calling for different techniques, may appear:

- Either *graph* f is a smooth curve, which for $n = 1$ means it is U-shaped (here, a Newton method, employing successive quadratic models for f is suitable)
- Or *graph* f is "sharp", which for $n = 1$ gives a V-shaped graph (here, a cutting-plane method, using successive piecewise-linear models for f is preferable).

As a result, it seems reasonable to look for second-order derivatives only where f is not "sharp", i.e., only on the $\mathcal{U}$-subspace, perpendicular to the $\mathcal{V}$-subspace, a subspace that is parallel to the subdifferential of f at $\bar{x}$. This is the basis for the so-called $\mathcal{U}$-Lagrangian theory, introduced in [21] and formalized in [19] for convex functions. Later on, in [12], [28], and [29], $\mathcal{VU}$-space decomposition theory was extended to certain nonconvex functions. In particular, [12] develops the *quadratic sub-Lagrangian* (cf. (5) below) as an extension of the $\mathcal{U}$-Lagrangian to prox-regular functions ([35]) that are prox-bounded.

For our development we consider trajectories $\chi(u)$ parameterized by $u \in \mathcal{U}$ which are given by the quadratic sub-Lagrangian. These trajectories converge to $\bar{x}$ and are tangent to $\mathcal{U}$ there. When the sub-Lagrangian has a Hessian at $0 \in \mathcal{U}$ (i.e., when f has a "$\mathcal{U}$-Hessian"), f has second-order epi-derivatives which agree with the sub-Lagrangian Hessian on $\mathcal{U}$. Furthermore, those trajectories that are C^2 give a second order expansion for $f(\chi(u))$. We also show that, near a minimizer, the proximal mapping sends points onto a particular trajectory. This is an important result, because it is known that a bundle mechanism can approximate proximal points with any desired accuracy; see [6], [15].

Our paper is organized as follows: Section 2 reviews Variational Analysis, $\mathcal{VU}$-theory and $\mathcal{U}$- and sub- Lagrangian definitions and results. Section 3, with our main results, is divided into two parts. In Subsection 3.1 we give the tangency property of a trajectory, the second-order expansion for its corresponding sub-Lagrangian, and show that the second-order epi-derivative of f and the $\mathcal{U}$-Hessian are equivalent second-order objects. Finally, in Subsection 3.2 we give the connection between proximal points and a special trajectory that we call a *proximal-track*.

2. Basic definitions and previous results

Here we recall from previous work important concepts and relations that we use in our development. We start with some basic Variational Analysis definitions. Then we review some elements of $\mathcal{VU}$-space decomposition and $\mathcal{U}$-Lagrangian theory from [19], as well as results for quadratic sub-Lagrangians from [12].

Our notation essentially follows that of [42] and [32]. In particular, from [32], given a sequence of vectors $\{z_k\}$ converging to 0,

- $\zeta_k = o(|z_k|) \iff \forall\varepsilon > 0\ \exists k_\varepsilon > 0$ such that $|\zeta_k| \le \varepsilon|z_k|$ for all $k \ge k_\varepsilon$.
- $\zeta_k = O(|z_k|) \iff \exists C > 0$ such that $|\zeta_k| \le C|z_k|$ for all $k \ge 1$.

2.1. Some notions from Variational Analysis

For a set $C \subset \mathbb{R}^n$ and a point $x \in C$:

- A vector v is *normal* to C at x if there are sequences $x^\nu \to_C x$ and $v^\nu \to v$ such that $\langle v^\nu, z - x^\nu\rangle \le o(|z - x^\nu|)$ for all $z \in C$.
- A set C is said to be *Clarke regular* at $x \in C$ when C is locally closed at x and each normal vector v satisfies $\langle v, z - x\rangle \le o(|z - x|)$ for all $z \in C$.

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be an lsc function, so that its epigraph, denoted and defined by $\operatorname{epi} f := \{(x, \beta) \in \mathbb{R}^n \times \mathbb{R} : \beta \ge f(x)\}$, is a closed set in $\mathbb{R}^{n+1}$. Take $\bar{x} \in \mathbb{R}^n$ where f is finite-valued.

- We use the Mordukhovich subdifferential ([31]) denoted by $\partial f(\bar{x})$ in [42]; see p. 301 and Definition 8.3 therein.
- The function f is said to be *subdifferentially regular* at $\bar{x}$ if $\operatorname{epi} f$ is a Clarke regular set at $(\bar{x}, f(\bar{x}))$; see [42, Definition 7.25, p. 260]. For such a function, the set $\partial f(\bar{x})$ is convex.
- The function f is said to be *prox-regular* at $\bar{x}$ for a subgradient $\bar{g} \in \partial f(\bar{x})$ (with parameter ρ) if there exists $\rho > 0$ such that

  $$f(x') \ge f(x) + \langle g, x' - x\rangle - \frac{\rho}{2}|x' - x|^2$$

  whenever x and x' are near $\bar{x}$ with $f(x)$ near $f(\bar{x})$ and $g \in \partial f(x)$ near $\bar{g}$.
  When this property holds for all subgradients in $\partial f(\bar{x})$, the function is said to be prox-regular at $\bar{x}$; see [35], [42, Definition 13.27, p. 610]. Moreover, in this case it can be shown that f is subdifferentially regular at $\bar{x}$, [9].

Convex functions are both subdifferentially regular and prox-regular, and in this case ∂f is the subdifferential from Convex Analysis. Lower C^2 and strongly amenable functions are also prox-regular; see [42, p. 613 and 612].

The epigraphical convergence theory developed in [42, Chapter 7] includes the following useful characterization of epi-limits:

- Let $\{q^\nu\}$ be a sequence of functions on $\mathbb{R}^n$, and let w be any point in $\mathbb{R}^n$. The value $q(w)$ is the *epi-limit* of the sequence q^ν at w if and only if

  $$\begin{cases} \operatorname{liminf}_\nu q^\nu(w^\nu) \ge q(w) & \text{for every sequence } w^\nu \to w\,, \\ \operatorname{limsup}_\nu q^\nu(w^\nu) \le q(w) & \text{for some sequence } w^\nu \to w\,. \end{cases}$$

For a function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ and point $\bar{x}$ with $h(\bar{x})$ finite we consider the second-order difference quotient: $\frac{h(\bar{x}+\tau\cdot) - h(\bar{x}) - \tau\langle y,\cdot\rangle}{\frac{1}{2}\tau^2}$ for $\tau > 0$ and $y \in \mathbb{R}^n$.

- The *second subderivative* of h at $\bar{x}$ relative to y in the direction w is denoted and defined by

$$\mathrm{d}^2 h(\bar{x}|y)(w) := \liminf_{\tau \searrow 0, w' \to w} \frac{h(\bar{x} + \tau w') - h(\bar{x}) - \tau \langle y, w' \rangle}{\frac{1}{2}\tau^2}.$$

When the second-order difference quotient has an epi-limit at w as $\tau \searrow 0$ then $\mathrm{d}^2 h(\bar{x}|y)(w)$ is this limit and it is then called the *second epi-derivative* of h at $\bar{x}$ relative to y in the direction w.

2.2. $\mathcal{VU}$-space decomposition

For a function f at a point $\bar{x} \in \mathbb{R}^n$ where f is finite, let g be any subgradient in $\partial f(\bar{x})$. Then, letting lin$Y$ denote the linear hull of a given set Y, the orthogonal subspaces

$$\mathcal{V} := \mathrm{lin}(\partial f(\bar{x}) - g) \qquad \text{and} \qquad \mathcal{U} := \mathcal{V}^{\perp} \tag{1}$$

define the $\mathcal{VU}$-space decomposition at $\bar{x}$ of [19, §2]. We use the compact notation $\oplus$ for such decomposition, and write $\mathbb{R}^n = \mathcal{U} \oplus \mathcal{V}$, as well as

$$\mathbb{R}^n \ni x = x_{\mathcal{U}} \oplus x_{\mathcal{V}} \in \mathcal{U} \times \mathcal{V}.$$

From (1), the relative interior of $\partial f(\bar{x})$, denoted by ri$\partial f(\bar{x})$, is the interior of $\partial f(\bar{x})$ relative to its affine hull, a manifold that is parallel to $\mathcal{V}$ (cf. [19, Definition 2.1 and Proposition 2.2]). Accordingly,

$$\bar{g} \in \mathrm{ri}\partial f(\bar{x}) \quad \Longrightarrow \quad \bar{g} + \Big(B(0,\eta) \cap \mathcal{V}\Big) \subset \partial f(\bar{x}) \text{ for some } \eta > 0, \tag{2}$$

where $B(0,\eta)$ denotes a ball in $\mathbb{R}^n$ centered at 0, with radius η.

Throughout the following we assume that *dim* $\mathcal{U} \geq 1$ and *dim* $\mathcal{V} \geq 1$.

2.3. $\mathcal{U}$-Lagrangians for convex functions

Suppose f is a convex function on $\mathbb{R}^n$. Given a subgradient $\bar{g} \in \partial f(\bar{x})$ with $\mathcal{V}$-component $\bar{g}_{\mathcal{V}}$, the $\mathcal{U}$-Lagrangian of f, depending on $\bar{g}_{\mathcal{V}}$, is defined by

$$\mathcal{U} \ni u \mapsto L_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}}) := \inf_{v \in \mathcal{V}} \Big\{ f(\bar{x} + u \oplus v) - \langle \bar{g}_{\mathcal{V}}, v \rangle_{\mathcal{V}} \Big\}, \tag{3}$$

where $\langle \cdot, \cdot \rangle_{\mathcal{V}(\mathcal{U})}$ denotes a scalar product induced in the subspace $\mathcal{V}(\mathcal{U})$, and similarly for the norms. When the infimum in (3) is attained, the set of corresponding $\mathcal{V}$-space minimizers is defined by

$$W(u; \bar{g}_{\mathcal{V}}) := \Big\{ v \in \mathcal{V} : L_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}}) = f(\bar{x} + u \oplus v) - \langle \bar{g}_{\mathcal{V}}, v \rangle_{\mathcal{V}} \Big\}.$$

When $W(u; \bar{g}_{\mathcal{V}})$ is nonempty, the associated $\mathcal{U}$-Lagrangian is a convex function that is differentiable at $u = 0$ with

$$\nabla L_{\mathcal{U}}(0; \bar{g}_{\mathcal{V}}) = \bar{g}_{\mathcal{U}} = g_{\mathcal{U}} \quad \text{for all } g \in \partial f(\bar{x}). \tag{4}$$

Finally, when $\bar{g} \in \mathrm{ri}\partial f(\bar{x})$, $W(u; \bar{g}_{\mathcal{V}})$ is nonempty with $W(0; \bar{g}_{\mathcal{V}}) = \{0\}$ and each $w(u) \in W(u; \bar{g}_{\mathcal{V}})$ being $o(|u|_{\mathcal{U}})$, see [19, Corollary 3.5]. In Lemma 3.1 below we extend this result to prox-regular functions, via the quadratic sub-Lagrangian from [12], which is the subject of our next subsection.

2.4. Quadratic sub-Lagrangians for prox-regular functions

Suppose f is a function that is finite at $\bar{x}$. Given a subgradient $\bar{g} \in \partial f(\bar{x})$ with $\mathcal{V}$-component $\bar{g}_{\mathcal{V}}$, the quadratic sub-Lagrangian of f, depending on a positive parameter R, is defined by

$$\mathcal{U} \ni u \mapsto \Phi_R(u; \bar{g}_{\mathcal{V}}) := \inf_{v \in \mathcal{V}} \left\{ f(\bar{x} + u \oplus v) - \langle \bar{g}_{\mathcal{V}}, v \rangle_{\mathcal{V}} + \frac{R}{2}|v|_{\mathcal{V}}^2 \right\} . \tag{5}$$

In Lemma 2.1 below we give conditions for the corresponding set of $\mathcal{V}$-space minimizers

$$W_R(u; \bar{g}_{\mathcal{V}}) := \left\{ v \in \mathcal{V} : \Phi_R(u; \bar{g}_{\mathcal{V}}) = f(\bar{x} + u \oplus v) - \langle \bar{g}_{\mathcal{V}}, v \rangle_{\mathcal{V}} + \frac{R}{2}|v|_{\mathcal{V}}^2 \right\} \tag{6}$$

to be nonempty.

The envelope function Φ_R extends many properties of the $\mathcal{U}$-Lagrangian to certain nonconvex functions f. The following lemma states some of these properties, that are relevant for our development.

Lemma 2.1. *Suppose that f is subdifferentially regular $\bar{x} \in \mathbb{R}^n$ and prox-regular there for $\bar{g} \in \partial f(\bar{x})$ with parameter $\rho > 0$, and that*

$$\forall x \in \mathbb{R}^n \quad f(x) \geq f(\bar{x}) + \langle \bar{g}, x - \bar{x} \rangle - \frac{\rho}{2}|x - \bar{x}|^2. \tag{7}$$

Then for any $R \geq \rho$, the function $\Phi_R(\cdot; \bar{g}_{\mathcal{V}})$ is well defined with $\Phi_R(0; \bar{g}_{\mathcal{V}}) = f(\bar{x})$. Furthermore, the following hold for any $R > \rho$:

- (i) *$\Phi_R(\cdot; \bar{g}_{\mathcal{V}})$ is strictly continuous and strictly differentiable at 0 (see Definitions 9.1 and 9.17 in [42]), with $\nabla \Phi_R(0; \bar{g}_{\mathcal{V}}) = \bar{g}_{\mathcal{U}}$.*
- (ii) *$W_R(u; \bar{g}_{\mathcal{V}})$ is nonempty for all u near 0, $W_R(0; \bar{g}_{\mathcal{V}}) = \{0\}$, and $W_R(\cdot; \bar{g}_{\mathcal{V}})$ is outer semicontinuous at 0.*

Proof. Condition A in [12, p. 1120] gathers together our assumptions. The facts that $\Phi_R(\cdot; \bar{g}_{\mathcal{V}})$ is well defined and $W_R(0; \bar{g}_{\mathcal{V}}) = \{0\}$ come from Theorem 5 in [12]. Item (i) is part of Theorem 14 in [12]. Nonemptiness and outer semicontinuity of $W_R(\cdot; \bar{g}_{\mathcal{V}})$ follow, respectively, from Proposition 6 in [12] and the continuity in item (i) combined with Theorem 7 in [12]. □

Condition (7) above is a strong form of *prox-boundedness* for f, see [42, Definition 1.23, p. 20]. This property is required for the proximal point mapping of a prox-regular function to be single valued; see Lemma 3.4 below.

Finally, we mention that for the case of convex functions, quadratic sub-Lagrangians can be traced back to the function $\phi_{\mathcal{V}}$ in [21, 3.2], which corresponds to Φ_R with $R = 1$ in the notation of this paper. Later on, in [19, Section 5], the same function was shown to agree up to second order with $L_{\mathcal{U}}$. More precisely, Lemma 5.1 in [19] (whose proof holds using R instead of 1 to define $\phi_{\mathcal{V}}$) gives the following relation: If f is convex and $\bar{g} \in \mathrm{ri}\partial f(\bar{x})$ then

$$\forall \varepsilon > 0 \ \exists \delta > 0 : |u|_{\mathcal{U}} \leq \delta \implies |\Phi_R(u; \bar{g}_{\mathcal{V}}) - L_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}})| \leq \varepsilon |u|_{\mathcal{U}}^2; \tag{8}$$

In particular, from (4), this means that $\nabla\Phi_R(0;\bar g_{\mathcal{V}}) = \bar g_{\mathcal{U}}$, an equality which is consistent with item (i) in Lemma 2.1. Also, if $L_{\mathcal{U}}(\cdot;\bar g_{\mathcal{V}})$ has a Hessian at 0, then $\Phi_R(\cdot;\bar g_{\mathcal{V}})$ has the same one; see Remark 3.3 below.

3. Main results

The results presented so far show how $\mathcal{VU}$ decomposition theory provides a set of smoothness for f, via the envelope functions $L_{\mathcal{U}}$ and Φ_R. Because of relation (4) and its left hand equality analog for Φ_R in Lemma 2.1(i), the gradient $\bar g_{\mathcal{U}}$ is called the *$\mathcal{U}$-gradient of f at $\bar x$*. Similarly, whenever the Hessian $\nabla^2\Phi_R(0;\bar g_{\mathcal{V}})$ exists, we call it a *$\mathcal{U}$-Hessian for f at $\bar x$ relative to $\bar g$ and R.*

We now show how the quadratic sub-Lagrangian captures a function's second-order epi-differential behavior with respect to $\mathcal{U}$ via ordinary second derivatives.

3.1. $\mathcal{U}$-Hessians give 2nd-order epi-derivatives

We start by showing $\mathcal{U}$-tangency properties of trajectories of the form $\chi(u) := \bar x + u \oplus v(u)$, where $v(u)$ is a $\mathcal{V}$-space minimizer defining a sub-Lagrangian corresponding to a particular subgradient at $\bar x$. A similar result can be found in [9, Chapter 4].

Lemma 3.1. *Suppose that f is subdifferentially regular $\bar x \in \mathbb{R}^n$ and prox-regular there for $\bar g \in \partial f(\bar x)$ with parameter $\rho > 0$, and that (7) holds. Also, suppose that $\bar g \in \mathrm{ri}\,\partial f(\bar x)$ and for $R > \rho$ consider a $\mathcal{V}$-space minimizer function $v(u) \in W_R(u;\bar g_{\mathcal{V}})$ from (6). Then the following hold for all u small enough:*

(i) *$v(u) \to 0$ as $u \to 0$.*

(ii) *If, in addition, $\Phi_R(\cdot;\bar g_{\mathcal{V}})$ has a Hessian at 0, then $v(u) = O(|u|^2_{\mathcal{U}})$ and*

$$f(\bar x + u \oplus v(u)) = f(\bar x) + \langle \bar g, u \oplus v(u)\rangle + \frac{1}{2}\langle u, Hu\rangle_{\mathcal{U}} + o(|u|^2_{\mathcal{U}}), \tag{9}$$

where $H := \nabla^2\Phi_R(0;\bar g_{\mathcal{V}})$.

Proof. The first assertion is straightforward from item (ii) in Lemma 2.1, since it implies that $v(u) \to W_R(0;\bar g_{\mathcal{V}}) = 0$ as $u \to 0$.
To see (ii), use Lemma 2.1 to write the following second-order expansion for $\Phi_R(\cdot;\bar g_{\mathcal{V}})$:

$$\begin{aligned}\Phi_R(u;\bar g_{\mathcal{V}}) &= \Phi_R(0;\bar g_{\mathcal{V}}) + \langle \nabla\Phi_R(0;\bar g_{\mathcal{V}}), u\rangle_{\mathcal{U}} + \frac{1}{2}\langle u, Hu\rangle_{\mathcal{U}} + o(|u|^2_{\mathcal{U}})\\ &= f(\bar x) + \langle \bar g_{\mathcal{U}}, u\rangle_{\mathcal{U}} + \frac{1}{2}\langle u, Hu\rangle_{\mathcal{U}} + o(|u|^2_{\mathcal{U}}).\end{aligned}$$

Together with the fact that, by (6),

$$\Phi_R(u;\bar g_{\mathcal{V}}) = f(\bar x + u \oplus v(u)) - \langle \bar g_{\mathcal{V}}, v(u)\rangle_{\mathcal{V}} + \frac{R}{2}|v(u)|^2_{\mathcal{V}},$$

we obtain that

$$f(\bar x + u \oplus v(u)) = f(\bar x) + \langle \bar g, u \oplus v(u)\rangle - \frac{R}{2}|v(u)|^2_{\mathcal{V}} + \frac{1}{2}\langle u, Hu\rangle_{\mathcal{U}} + o(|u|^2_{\mathcal{U}}). \tag{10}$$

From (2) we have that for $\eta > 0$ sufficiently small

$$\gamma := \bar{g} + \left(0 \oplus \frac{\eta v(u)}{|v(u)|_{\mathcal{V}}}\right) \in \partial f(\bar{x}),$$

which, by prox-regularity (with η sufficiently small to have γ near enough to $\bar{g}$), implies that

$$f(x') \geq f(\bar{x}) + \langle \gamma, x' - \bar{x} \rangle - \frac{\rho}{2}|x' - \bar{x}|^2$$

for all x' close $\bar{x}$ such that $f(x')$ is close to $f(\bar{x})$. In particular, for $x' = \bar{x} + u \oplus v(u)$ (10) and item (i) imply that $f(x') \to f(\bar{x})$ as $u \to 0$, so

$$\begin{aligned} f(\bar{x} + u \oplus v(u)) &\geq f(\bar{x}) + \left\langle \bar{g} + \left(0 \oplus \frac{\eta v(u)}{|v(u)|_{\mathcal{V}}}\right), u \oplus v(u) \right\rangle - \frac{\rho}{2}|u \oplus v(u)|^2 \\ &= f(\bar{x}) + \langle \bar{g}, u \oplus v(u) \rangle + \eta |v(u)|_{\mathcal{V}} - \frac{\rho}{2}\left(|u|^2_{\mathcal{U}} + |v(u)|^2_{\mathcal{V}}\right). \end{aligned}$$

Together with (10), the last inequality gives, after rearrangement of terms,

$$\frac{1}{2}\langle u, Hu \rangle_{\mathcal{U}} + \frac{\rho}{2}|u|^2_{\mathcal{U}} + o(|u|^2_{\mathcal{U}}) \geq \eta |v(u)|_{\mathcal{V}} + \frac{R - \rho}{2}|v(u)|^2_{\mathcal{V}},$$

which implies that $v(u) = O(|u|^2_{\mathcal{U}})$. Then the desired result (9) follows from (10). □

In the epigraphical setting, the second-order epi-derivative provides a second-order approximation in the sense of closeness of the epigraphs of the second-order difference quotient function and $\mathrm{d}^2 f(\bar{x}|y)(\cdot)$; see [34]. In contrast, a quadratic sub-Lagrangian $\Phi_R(u; \bar{g}_{\mathcal{V}})$ can provide a second-order approximation with respect to u in the classical sense, of local uniform convergence.

We now establish a relation between these second-order objects.

Theorem 3.2. *Suppose that f is subdifferentially regular at $\bar{x} \in \mathbb{R}^n$ and prox-regular there for $\bar{g} \in \partial f(\bar{x})$ with parameter $\rho > 0$, and that* (7) *holds. Also, suppose that $\bar{g} \in$* $\mathrm{ri}\partial f(\bar{x})$ *and for $R > \rho$ the sub-Lagrangian $\Phi_R(\cdot; \bar{g}_{\mathcal{V}})$ has a Hessian at* 0*. Then the second-order epi-derivative of f at $\bar{x}$ relative to $\bar{g}$ for each $w \in \mathcal{U}$ is given by*

$$\mathrm{d}^2 f(\bar{x}|\bar{g})(w) = \left\langle w, \nabla^2 \Phi_R(0; \bar{g}_{\mathcal{V}}) w \right\rangle_{\mathcal{U}}.$$

Proof. For convenience, we let $H := \nabla^2 \Phi_R(0; \bar{g}_{\mathcal{V}})$ and $\chi(u) := \bar{x} + u \oplus v(u)$, where $v(u) \in W_R(u; \bar{g}_{\mathcal{V}})$. Then for all $v \in \mathcal{V}$

$$\Phi_R(u; \bar{g}_{\mathcal{V}}) = f(\chi(u)) - \langle \bar{g}_{\mathcal{V}}, v(u) \rangle_{\mathcal{V}} + \frac{R}{2}|v(u)|^2_{\mathcal{V}} \leq f(\bar{x} + u \oplus v) - \langle \bar{g}_{\mathcal{V}}, v \rangle_{\mathcal{V}} + \frac{R}{2}|v|^2_{\mathcal{V}}.$$

Subtracting $f(\bar{x}) + \langle \bar{g}_{\mathcal{U}}, u \rangle_{\mathcal{U}}$ from both sides of the inequality above gives

$$f(\chi(u)) - f(\bar{x}) - \langle \bar{g}, \chi(u) - \bar{x} \rangle + \frac{R}{2}|v(u)|^2_{\mathcal{V}} \leq f(\bar{x} + u \oplus v) - f(\bar{x}) - \langle \bar{g}, u \oplus v \rangle + \frac{R}{2}|v|^2_{\mathcal{V}}. \quad (11)$$

Then, since the left hand side above involves $\Phi_R(u; \bar{g}_{\mathcal{V}})$, which can be expanded up to second order as in (9) of Lemma 3.1(ii), we obtain that for all $u \in \mathcal{U}$ small enough and all $v \in \mathcal{V}$

$$\frac{1}{2}\langle u, Hu \rangle_{\mathcal{U}} + o(|u|^2_{\mathcal{U}}) \leq f(\bar{x} + u \oplus v) - f(\bar{x}) - \langle \bar{g}, u \oplus v \rangle + \frac{R}{2}|v|^2_{\mathcal{V}}. \quad (12)$$

Suppose $w \in \mathcal{U}$ so that $w = w_{\mathcal{U}} \oplus 0$. Take any sequence (w', τ) with $w' \to w$ and $\tau \searrow 0$, and let $u' := w'_{\mathcal{U}}$. Then

$$w' = w'_{\mathcal{U}} \oplus w'_{\mathcal{V}} = u' \oplus w'_{\mathcal{V}}$$

and

$$\bar{x} + \tau w' = \bar{x} + \tau\Big(u' \oplus w'_{\mathcal{V}}\Big) = \bar{x} + \Big(\tau u'\Big) \oplus \Big(\tau w'_{\mathcal{V}}\Big).$$

From (12) with $u = \tau u' \in \mathcal{U}$, τ small enough and $v = \tau w'_{\mathcal{V}} \in \mathcal{V}$ we obtain

$$\frac{1}{2}\left\langle \tau u', H\tau u' \right\rangle_{\mathcal{U}} + o(|\tau u'|^2_{\mathcal{U}}) \leq f(\bar{x} + \tau w') - f(\bar{x}) - \tau \left\langle \bar{g}, w' \right\rangle + \frac{R}{2}|\tau w'_{\mathcal{V}}|^2_{\mathcal{V}}. \tag{13}$$

Dividing both sides of this inequality by $\frac{1}{2}\tau^2$ yields

$$\left\langle u', Hu' \right\rangle_{\mathcal{U}} + \frac{o(\tau^2|u'|^2)}{\frac{1}{2}\tau^2} \leq \frac{f(\bar{x} + \tau w') - f(\bar{x}) - \tau \left\langle \bar{g}, w' \right\rangle}{\frac{1}{2}\tau^2} + R|w'_{\mathcal{V}}|^2_{\mathcal{V}}.$$

Note that since $w' \to w \in \mathcal{U}$, $w'_{\mathcal{V}} \to 0$, and the definition of u' implies that $u' \to w_{\mathcal{U}}$. Hence, passing to the limit as $w' \to w$ and $\tau \searrow 0$ and using the fact that $w = w_{\mathcal{U}} \oplus 0$ gives the following inequality involving the second subderivative:

$$\left\langle w, Hw \right\rangle_{\mathcal{U}} \leq \mathrm{d}^2 f(\bar{x}|\bar{g})(w).$$

To show that the left hand side is an epi-limit for $w \in \mathcal{U}$ we reexamine the above proof concentrating on the left hand sides of the inequalities. Given $w \in \mathcal{U}$ we define a sequence w_e^τ converging to w as follows: Let u^τ be any sequence such that $u^\tau \to w_{\mathcal{U}}$ as $\tau \searrow 0$ and let

$$w_e^\tau := u^\tau \oplus \frac{1}{\tau} v(\tau u^\tau) \text{ which implies } \bar{x} + \tau w_e^\tau = \chi(\tau u^\tau).$$

From Lemma 3.1(ii) we have

$$v(\tau u^\tau) = O(|\tau u^\tau|^2_{\mathcal{U}}), \tag{14}$$

so $\frac{1}{\tau} v(\tau u^\tau) \to 0$ with τ and, hence, $w_e^\tau \to w_{\mathcal{U}} \oplus 0 = w$.
Furthermore, since $\bar{x} + \tau w_e^\tau = \chi(\tau u^\tau)$, the left hand side in (11) with $u = \tau u^\tau$, divided by $\frac{1}{2}\tau^2$, can be written as

$$\left[\frac{f(\bar{x} + \tau w_e^\tau) - f(\bar{x}) - \tau \left\langle \bar{g}, w_e^\tau \right\rangle}{\frac{1}{2}\tau^2}\right] + \frac{\frac{R}{2}|v(\tau u^\tau)|^2_{\mathcal{V}}}{\frac{1}{2}\tau^2}.$$

By the above argument, as $\tau \searrow 0$ this two term expression converges to $\left\langle w, Hw \right\rangle_{\mathcal{U}}$. Its second term converges to zero by (14). Therefore, its first term converges to $\left\langle w, Hw \right\rangle_{\mathcal{U}}$ and the proof is complete. $\square$

Similar expressions are given in [27] for convex finite max functions and in [28] for more general "pdg-structured" functions. In addition, the first reference gives second order epi-derivatives for three specific examples while the second gives them for an example that is not prox-regular, cf. [11, Section 7].

Remark 3.3. Suppose in the statements of Lemma 3.1 and Theorem 3.2 we replace the first sentence by the assumption that f is convex on $\mathbb{R}^n$ and replace W_R, Φ_R, R, and ρ by W, $L_{\mathcal{U}}$, 0, and 0, respectively. Then we can conclude that if $L_{\mathcal{U}}(\cdot, \bar{g}_{\mathcal{V}})$ has a Hessian at 0 for some $\bar{g} \in \mathrm{ri}\,\partial f(\bar{x})$ then for all $w \in \mathcal{U}$

$$\mathrm{d}^2 f(\bar{x}|\bar{g})(w) = \left\langle w, \nabla^2 L_{\mathcal{U}}(0; \bar{g}_{\mathcal{V}}) w \right\rangle_{\mathcal{U}},$$

using in the proofs the Φ_R-like properties of $L_{\mathcal{U}}$ from Subsection 2.3.

3.2. Proximal points correspond to $\mathcal{V}$-space minimizers

The following Lemma, extracted from [42, Proposition 13.37, p. 617], gives basic properties of the proximal point mapping for a prox-regular function; see also [35]. It depends on condition (7), that implies prox-boundedness.

Lemma 3.4. *Suppose that f is prox-regular at $\bar{x} \in \mathbb{R}^n$ for $\bar{g} = 0 \in \partial f(\bar{x})$ with parameter $\rho > 0$ and that (7) holds. Then for each $\mu > 0$ sufficiently large there is a neighborhood of $\bar{x}$ on which the proximal point mapping*

$$p_\mu(x) := \operatorname{argmin}_w \left\{ f(w) + \frac{\mu}{2}|w - x|^2 \right\}$$

is well defined, single valued and Lipschitz continuous. In addition,

$$g_\mu(x) := \mu\,(x - p_\mu(x)) \in \partial f(p_\mu(x)) \quad \textit{and} \quad p_\mu(\bar{x}) = \bar{x}.$$

□

We now relate the proximal point mapping to a very particular trajectory $\bar{x} + u \oplus v_{\partial f}(u)$, where the $\mathcal{V}$-space minimizer function $v_{\partial f}(u)$ is the **same** for all relative interior subgradients at $\bar{x}$.

Theorem 3.5. *Suppose that f is prox-regular at $\bar{x} \in \mathbb{R}^n$ with parameter $\rho > 0$ and that (7) holds for all $\bar{g} \in \operatorname{ri}\partial f(\bar{x})$. In addition, suppose $0 \in \operatorname{ri}\partial f(\bar{x})$ and for some $R > \rho$ there is a function $v_{\partial f} : \mathcal{U} \to \mathcal{V}$ such that, for all u small enough, $v_{\partial f}(u) \in W_R(u; \bar{g}_\mathcal{V})$ for all $\bar{g} \in \operatorname{ri}\partial f(\bar{x})$. Then, for all $\mu > 0$ sufficiently large and x close enough to $\bar{x}$,*

$$p_\mu(x) = \bar{x} + \pi_\mathcal{U}(x) \oplus v_{\partial f}\big(\pi_\mathcal{U}(x)\big) \textit{ where } \pi_\mathcal{U}(x) := (p_\mu(x) - \bar{x})_\mathcal{U}.$$

Proof. For x close enough to $\bar{x}$, we write its proximal point using $\mathcal{VU}$ coordinates, as follows:

$$p_\mu(x) = \bar{x} + \pi_\mathcal{U}(x) \oplus \pi_\mathcal{V}(x) \text{ where } \pi_\mathcal{U}(x) := (p_\mu(x) - \bar{x})_\mathcal{U} \text{ and } \pi_\mathcal{V}(x) := (p_\mu(x) - \bar{x})_\mathcal{V}.$$

By Lemma 3.4, as $x \to \bar{x}$, $x - p_\mu(x) \to \bar{x} - p_\mu(\bar{x}) = 0$, and likewise, for the components $\pi_\mathcal{U}(x)$ and $\pi_\mathcal{V}(x)$. Since $\pi_\mathcal{U}(x) \to 0$, by Lemma 3.1(i), for any relative interior $\bar{g}$, any $v \in W_R(\pi_\mathcal{U}(x); \bar{g}_\mathcal{V})$ converges to 0. In particular, $v_{\partial f}\big(\pi_\mathcal{U}(x)\big) \to 0$. As a result, the function $\gamma_\mathcal{V} : \mathbb{R}^n \to \mathcal{V}$ defined by

$$\gamma_\mathcal{V}(x) := \mu(\bar{x} - x)_\mathcal{V} + \left(\frac{R - \mu}{2}\right)\Big(v_{\partial f}\big(\pi_\mathcal{U}(x)\big) + \pi_\mathcal{V}(x)\Big)$$

converges to 0 as $x \to \bar{x}$. From (2) written with $\bar{g} = 0 \in \operatorname{ri}\partial f(\bar{x})$, we obtain that

$$\gamma := 0 \oplus \gamma_\mathcal{V}(x) \in \partial f(\bar{x}) \quad \text{for } x \text{ close enough to } \bar{x}$$

(in fact, $\gamma \in \operatorname{ri}\partial f(\bar{x})$, by definition of relative interior). Thus, by Lemma 2.1, the function $\Phi_R(\cdot; \gamma_\mathcal{V}(x))$ corresponding to the subgradient γ is well defined. In particular, at $u = \pi_\mathcal{U}(x)$, using (6) and letting $\chi(\pi_\mathcal{U}) := \bar{x} + \pi_\mathcal{U}(x) \oplus v_{\partial f}\big(\pi_\mathcal{U}(x)\big)$,

$$\Phi_R\Big(\pi_\mathcal{U}(x); \gamma_\mathcal{V}(x)\Big) = f(\chi(\pi_\mathcal{U}(x))) - \big\langle \gamma_\mathcal{V}(x), v_{\partial f}\big(\pi_\mathcal{U}(x)\big)\big\rangle_\mathcal{V} + \frac{R}{2}|v_{\partial f}\big(\pi_\mathcal{U}(x)\big)|^2_\mathcal{V}.$$

Since $\pi_{\mathcal{V}}(x) \in \mathcal{V}$, definition (5) of the sub-Lagrangian implies that

$$\Phi_R\Big(\pi_{\mathcal{U}}(x); \gamma_{\mathcal{V}}(x)\Big) \leq f(\bar{x} + \pi_{\mathcal{U}}(x) \oplus \pi_{\mathcal{V}}(x)) - \langle \gamma_{\mathcal{V}}(x), \pi_{\mathcal{V}}(x)\rangle_{\mathcal{V}} + \frac{R}{2}|\pi_{\mathcal{V}}(x)|^2_{\mathcal{V}}.$$

As a result,

$$\begin{aligned} &f(\chi(\pi_{\mathcal{U}}(x))) - \big\langle \gamma_{\mathcal{V}}(x), v_{\partial f}\big(\pi_{\mathcal{U}}(x)\big)\big\rangle_{\mathcal{V}} + \frac{R}{2}|v_{\partial f}\big(\pi_{\mathcal{U}}(x)\big)|^2_{\mathcal{V}} \\ &\quad \leq f(p_\mu(x)) - \langle \gamma_{\mathcal{V}}(x), \pi_{\mathcal{V}}(x)\rangle_{\mathcal{V}} + \frac{R}{2}|\pi_{\mathcal{V}}(x)|^2_{\mathcal{V}}. \end{aligned} \tag{15}$$

By the definition of the proximal point mapping in Lemma 3.4,

$$f(p_\mu(x)) + \frac{\mu}{2}|p_\mu(x) - x|^2 \leq f(\chi(\pi_{\mathcal{U}}(x))) + \frac{\mu}{2}|\chi(\pi_{\mathcal{U}}(x)) - x|^2. \tag{16}$$

Combining the two inequalities above yields, after rearrangement of terms,

$$\begin{aligned} \frac{R}{2}|v_{\partial f}\big(\pi_{\mathcal{U}}(x)\big)|^2_{\mathcal{V}} \;\leq\; & \frac{\mu}{2}\Big(|\chi(\pi_{\mathcal{U}}(x)) - x|^2 - |p_\mu(x) - x|^2\Big) \\ & + \big\langle \gamma_{\mathcal{V}}(x), v_{\partial f}\big(\pi_{\mathcal{U}}(x)\big) - \pi_{\mathcal{V}}(x)\big\rangle_{\mathcal{V}} \\ & + \frac{R}{2}|\pi_{\mathcal{V}}(x)|^2_{\mathcal{V}}. \end{aligned} \tag{17}$$

We now show that the inequality above is in fact an equality. To abbreviate notation, we drop the argument "(x)" in $\pi_{\mathcal{U}}(x)$, $p_\mu(x)$, $v_{\partial f}\big(\pi_{\mathcal{U}}(x)\big)$, $\pi_{\mathcal{V}}(x)$, and $\gamma_{\mathcal{V}}(x)$, and write instead $\pi_{\mathcal{U}}$, p_μ, $v_{\partial f}(\pi_{\mathcal{U}})$, $\pi_{\mathcal{V}}$, and $\gamma_{\mathcal{V}}$. First we expand the difference of squares factor in (17) and use the fact that $\chi(\pi_{\mathcal{U}})$ and p_μ have the same $\mathcal{U}$-component:

$$\begin{aligned} |\chi(\pi_{\mathcal{U}}) - x|^2 - |p_\mu - x|^2 &= \langle \chi(\pi_{\mathcal{U}}) - p_\mu, \chi(\pi_{\mathcal{U}}) + p_\mu - 2x\rangle \\ &= \Big\langle v_{\partial f}(\pi_{\mathcal{U}}) - \pi_{\mathcal{V}}, \Big(\chi(\pi_{\mathcal{U}}) + p_\mu - 2x\Big)_{\mathcal{V}}\Big\rangle_{\mathcal{V}} \\ &= \langle v_{\partial f}(\pi_{\mathcal{U}}) - \pi_{\mathcal{V}}, v_{\partial f}(\pi_{\mathcal{U}}) + \pi_{\mathcal{V}} - 2(x - \bar{x})_{\mathcal{V}}\rangle_{\mathcal{V}} \\ &= |v_{\partial f}(\pi_{\mathcal{U}})|^2_{\mathcal{V}} - |\pi_{\mathcal{V}}|^2_{\mathcal{V}} + 2\,\langle v_{\partial f}(\pi_{\mathcal{U}}) - \pi_{\mathcal{V}}, (x - \bar{x})_{\mathcal{V}}\rangle_{\mathcal{V}}\,. \end{aligned}$$

Then

$$\frac{\mu}{2}\Big(|\chi(\pi_{\mathcal{U}}) - x|^2 - |p_\mu - x|^2\Big) = \frac{\mu}{2}\Big(|v_{\partial f}(\pi_{\mathcal{U}})|^2_{\mathcal{V}} - |\pi_{\mathcal{V}}|^2_{\mathcal{V}}\Big) - \langle v_{\partial f}(\pi_{\mathcal{U}}) - \pi_{\mathcal{V}}, \mu(\bar{x} - x)_{\mathcal{V}}\rangle_{\mathcal{V}}\,.$$

Now we use the definition of $\gamma_{\mathcal{V}}$ to write the second right hand side term in (17) as follows:

$$\begin{aligned} \langle \gamma_{\mathcal{V}}, v_{\partial f}(\pi_{\mathcal{U}}) - \pi_{\mathcal{V}}\rangle_{\mathcal{V}} &= \Big\langle \mu(\bar{x} - x)_{\mathcal{V}} + \frac{R-\mu}{2}\Big(v_{\partial f}(\pi_{\mathcal{U}}) + \pi_{\mathcal{V}}\Big), v_{\partial f}(\pi_{\mathcal{U}}) - \pi_{\mathcal{V}}\Big\rangle_{\mathcal{V}} \\ &= \langle \mu(\bar{x} - x)_{\mathcal{V}}, v_{\partial f}(\pi_{\mathcal{U}}) - \pi_{\mathcal{V}}\rangle_{\mathcal{V}} + \frac{R-\mu}{2}\Big(|v_{\partial f}(\pi_{\mathcal{U}})|^2_{\mathcal{V}} - |\pi_{\mathcal{V}}|^2_{\mathcal{V}}\Big) \end{aligned}$$

Using these expressions in the right hand side in (17), we obtain that (17) holds with equality. Since the (in)equality in (17) cannot be strict, we deduce that neither the inequality in (15) nor the one in (16) can be strictly satisfied. In particular, since $p_\mu(x)$ is unique, from (16) we obtain that $p_\mu(x) = \chi(\pi_{\mathcal{U}}(x))$, i.e., that $\pi_{\mathcal{V}}(x) = v_{\partial f}\big(\pi_{\mathcal{U}}(x)\big)$. $\square$

Trajectories obtained using the special function $\mathfrak{v}_{\partial f}(\cdot)$ are called *fast tracks* in [26] whenever $\mathfrak{v}_{\partial f}(\cdot)$ and $L_{\mathcal{U}}(\cdot;0)$ are C^2 functions. The proximal-track result in Theorem 3.5, obtained without requiring a second order assumption on Φ_R, is similar to the fast track results in [26, Theorem 5.2] for f convex and in [29, Theorem 9] for f having a strongly transversal pdg-structure. Another related result can be found in [10], where it is shown that, for a convex function f, fast tracks, partly smooth functions, and identifiable surfaces are equivalent concepts. It should be noted, however, that the proximal-track in Theorem 3.5 above may not be "fast" unless $\Phi_R(\cdot;0)$ and $\mathfrak{v}_{\partial f}(\cdot)$ have continuous Hessians. These are desirable $\mathcal{VU}$-smoothness conditions that are important for rapid convergence of minimization algorithms.

Remark 3.6. If in Theorem 3.5 we replace the first sentence by the assumption that f is convex on $\mathbb{R}^n$, delete ρ, and replace R, W_R, and Φ_R by 0, W, and $L_{\mathcal{U}}$, respectively (which are well defined objects for all subgradients $\bar{g} \in \mathrm{ri}\partial f(\bar{x})$), we can use its proof to obtain the same result for the proximal-track of a convex function.

Remark 3.7. In the nonconvex case $\partial f(\bar{x})$ can be unbounded and making assumptions for all $\bar{g} \in \mathrm{ri}\partial f(\bar{x})$ may be somewhat strong. We could weaken the assumptions of Theorem 3.5 as follows:

> Suppose that f is prox-regular at $\bar{x} \in \mathbb{R}^n$ with parameter $\rho > 0$ and that (7) holds for all $\bar{g} \in \mathrm{ri}\partial f(\bar{x})$ **near** $0 \in \partial f(\bar{x})$. In addition, suppose $0 \in \mathrm{ri}\partial f(\bar{x})$, and for some $R > \rho$, $\mathfrak{v}_{\partial f}(u) \in W_R(u;\bar{g}_{\mathcal{V}})$ for all $\bar{g} \in \mathrm{ri}\partial f(\bar{x})$ **near** 0.

Then use the fact that the subgradient γ in the proof is near enough to 0 for x close enough to $\bar{x}$ to obtain the proximal-track result.

Acknowledgements. We thank W. Hare for beneficial comments, including a suggestion for improvement of the statement of Theorem 3.5.

References

[1] J.-F. Bonnans, J. Ch. Gilbert, C. Lemaréchal, C. Sagastizábal: A family of variable metric proximal methods, Math. Program., Ser. A 68(1) (1995) 15–47.

[2] J. V. Burke, Maijian Qian: On the super-linear convergence of the variable metric proximal point algorithm using Broyden and BFGS matrix secant updating, Math. Program., Ser. A 88(1) (2000) 157–181.

[3] A. Ben-Tal, J. Zowe: Necessary and sufficient optimality conditions for a class of nonsmooth minimization problems, Math. Program. 24(1) (1982) 70–91.

[4] R. Cominetti, R. Correa: A generalized second-order derivative in nonsmooth optimization, SIAM J. Control Optimization 28(4) (1990) 789–809.

[5] X. Chen, M. Fukushima: Proximal quasi-Newton methods for nondifferentiable convex optimization, Math. Program., Ser. A 85(2) (1999) 313–334.

[6] R. Correa, C. Lemaréchal: Convergence of some algorithms for convex minimization, Math. Program., Ser. B 62(2) (1993) 261–275.

[7] R. Cominetti: On pseudo-differentiability, Trans. Amer. Math. Soc. 2324(2) (1991) 843–865.

[8] M. Fukushima, L. Qi: A globally and superlinearly convergent algorithm for nonsmooth convex minimization, SIAM J. Optimization 6(4) (1996) 1106–1120.

[9] W. L. Hare: The Quadratic Sub-Lagrangian of Prox-Regular Functions, M. Sc. Dissertation, Department of Mathematical and Statistical Sciences, University of Alberta, Canada (2000),available at http://www.cecm.sfu.ca/˜whare.

[10] W. L. Hare: Recent functions and sets of smooth substructure: Relationships and examples, preprint, available at http://www.cecm.sfu.ca/˜whare.

[11] W. L. Hare, A. S. Lewis: Identifying active contraints via partial smoothness and prox-regularity, J. Convex Analysis 11(2) (2003) 251–266.

[12] W. L. Hare, R. A. Poliquin: The quadratic sub-Lagrangian of a prox-regular function, Nonlinear Anal. 47(2) (2001) 1117–1128.

[13] J.-B. Hiriart-Urruty: Refinements of necessary optimality conditions in nondifferentiable programming I, Appl. Math. Optimization 5 (1979) 63–82.

[14] J.-B. Hiriart-Urruty: Refinements of necessary optimality conditions in nondifferentiable programming II, Math. Program. Study 19 (1982) 120–139.

[15] J.-B. Hiriart-Urruty, C. Lemaréchal: Convex Analysis and Minimization Algorithms, Grundlehren der Mathematischen Wissenschaften 305-306, Springer, Berlin (1993) (two volumes).

[16] A. D. Ioffe: Variational analysis of a composite function: A formula for the lower second order epi-derivative, J. Math. Anal. Appl. (1991) 379–405.

[17] C. Lemaréchal: Extensions Diverses des Méthodes de Gradient et Applications, Thèse d'Etat, Université de Paris IX (1980).

[18] C. Lemaréchal, R. Mifflin: Global and superlinear convergence of an algorithm for one-dimensional minimization of convex functions, Math. Program. 24 (1982) 241–256.

[19] C. Lemaréchal, F. Oustry, C. Sagastizábal: The $\mathcal{U}$-Lagrangian of a convex function, Trans. Amer. Math. Soc. 352(2) (2000) 711–729.

[20] C. Lemaréchal, C. Sagastizábal: An approach to variable metric bundle methods, in: Systems Modelling and Optimization, J. Henry, J.-P. Yvon (eds.), Lecture Notes in Control and Information Sciences 197, Springer (1994) 144–162.

[21] C. Lemaréchal, C. Sagastizábal: Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries, SIAM J. Optimization 7(2) (1997) 367–385.

[22] C. Lemaréchal, C. Sagastizábal: Variable metric bundle methods: From conceptual to implementable forms, Math. Program., Ser. B 76(3) (1997) 393–410.

[23] R. Mifflin: On superlinear convergence in univariate nonsmooth minimization, Math. Program., Ser. A 49(2) (1991) 273–279.

[24] R. Mifflin: A quasi-second-order proximal bundle algorithm, Math. Program., Ser. A 73(1) (1996) 51–72.

[25] R. Mifflin, C. Sagastizábal: On $\mathcal{VU}$-theory for functions with primal-dual gradient structure, SIAM J. Optimization 11(2) (2000) 547–571.

[26] R. Mifflin, C. Sagastizábal: Proximal points are on the fast track, J. Convex Analysis 9(2) (2002) 563–579.

[27] R. Mifflin, C. Sagastizábal: On the relation between $\mathcal{U}$-Hessians and second-order epiderivatives, European Journal of Operations Research 157 (2004) 28–38.

[28] R. Mifflin, C. Sagastizábal: Primal-dual gradient structured functions: Second-order results; links to epi-derivatives and partly smooth functions, SIAM J. Optimization 13(4) (2003) 1174–1194.

[29] R. Mifflin, C. Sagastizábal: $\mathcal{VU}$-Smoothness and proximal point results for nonconvex functions, Optimization Methods and Software 19 (2004) 463–478.

[30] R. Mifflin, D. F. Sun, L. Q. Qi: Quasi-Newton bundle-type methods for nondifferentiable convex optimization, SIAM J. Optimization 8(2) (1998) 583–603.

[31] B. Sh. Mordukhovich: Maximum principle in the problem of time optimal response with nonsmooth constraints, Prikl. Mat. Meh. 40(6) (1976) 1014–1023.

[32] J. M. Ortega, W. C. Rheinboldt: Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York (1970).

[33] J.-P. Penot: Subhessians, superhessians and conjugation, Nonlinear Analysis: Theory, Methods and Applications 23(6) (1994) 689–702.

[34] R. A. Poliquin, R. T. Rockafellar: Second-order nonsmooth analysis in nonlinear programming, in: Recent Advances in Nonsmooth Optimization, World Sci. Publishing, River Edge (1995) 322–349.

[35] R. A. Poliquin, R. T. Rockafellar: Prox-regular functions in variational analysis, Trans. Amer. Math. Soc. 348(5) (1996) 1805–1838.

[36] L. Qi, X. Chen: A preconditioning proximal Newton method for nondifferentiable convex optimization, Math. Program., Ser. B 76(3) (1997) 411–429.

[37] L. Qi: Superlinearly convergent approximate Newton methods for LC^1 optimization problems, Math. Program., Ser. A 64(3) (1994) 277–294.

[38] A. I. Rauf, M. Fukushima: Globally convergent BFGS method for nonsmooth convex optimization, J. Optim. Theory Appl. 104(3) (2000) 539–558.

[39] S. M. Robinson: Local structure of feasible sets in nonlinear programming, Part III: Stability and sensitivity, Math. Program. Study 30 (1987) 45–66.

[40] R. T. Rockafellar: Proto-differentiability of set-valued mappings and its applications in optimization, Ann. Inst. Henri Poincaré, Anal. Non Linéaire 6 (1989) 449–482.

[41] R. T. Rockafellar: Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives, Math. Oper. Res. 14 (1989) 462–484.

[42] R. T. Rockafellar, R. J.-B. Wets: Variational Analysis, Grundlehren der Mathematischen Wissenschaften 317, Springer (1998).