On the relation between $\mathcal{U}$-Hessians and second-order epi-derivatives

Robert Mifflin* and Claudia Sagastizábal†

June 18, 2002

Abstract

We consider second order objects for a convex function defined as the maximum of a finite number of C^2-functions. Variational analysis yields explicit formulæ for the second-order epi-derivatives of such max-functions. Another second-order object can be defined by means of a space decomposition that allows us to identify a subspace on which a Lagrangian related to a max-function is smooth. This decomposition yields an explicit expression for the so-called $\mathcal{U}$-Hessian of the function, defined as the Hessian of the related Lagrangian. We show that the second-order epi-derivative and the $\mathcal{U}$-Hessian are equivalent second-order objects. Thus, the $\mathcal{U}$-Lagrangian of a max-function captures the function's second-order epi-differential behavior with ordinary second derivatives.

Keywords Second-order epi-derivatives, $\mathcal{U}$-Hessians, epigraphical and pointwise convergence.

AMS Subject Classification Primary: 49J52, 49K35. Secondary: 65K10, 90C47.

1 Introduction

In optimization, second-order derivatives constitute a significant tool, both for understanding optimality (by means of optimality conditions) and for developing algorithms (by means of local quadratic approximations). For functions $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ which lack differentiability, concepts such as "gradients" and "Hessians" need to be generalized. As pointed out in [RW98, Ch. 4], *the precise meaning of such basic notions in analysis is dictated by the choice of a notion of limit for sequences of functions.* Along these lines, the classical notion of pointwise convergence may sometimes be inadequate for approximating functions in optimization problems or for defining

*Department of Pure and Applied Mathematics, Washington State University, Pullman, WA 99164-3113. Research supported by the National Science Foundation under Grant No. DMS-0071459, `mifflin@math.wsu.edu`

†IMPA, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro RJ 22460-320, Brazil. On leave from INRIA, BP 105, 78153 Le Chesnay, France. Research supported by FAPERJ under Grant No.E26/150.581/00 and by CAPES (Brazil), `sagastiz@impa.br`

their derivatives. A good (and complementary) option considers the limiting behavior, not of the functions themselves, but of their epigraphs. Being sets, epigraphs require the use of set convergence theory, specifically Painlevé-Kuratowski convergence, [RW98, Ch. 7].

Derivatives obtained by taking epigraphical limits of difference quotients are the so-called *epi-derivatives*. In this paper we focus on the second-order epi-derivatives of a function $f : \mathbb{R}^n \to \mathbb{R}$ defined as the pointwise maximum of a finite number of C^2-functions:

$$f(x) = \max\{f_i(x) : i = 0, \cdots, m\}. \tag{1}$$

In particular, if f is also convex, it is regular and strictly continuous. Under these circumstances, as mentioned in [RW98, p. 586], the second-order epi-derivative of f at $\bar{x}$ for $\bar{g} \in \partial f(\bar{x})$ cannot be finite everywhere unless $\partial f(\bar{x})$ is a singleton, i.e., unless f is differentiable at $\bar{x}$. In order to capture the key second-order behavior of nonsmooth functions, it is necessary to accept infinite values for second-order derivatives. For example, the second order epi-derivative depending on a subgradient $\bar{g}$ is finite only on the normal cone $N_{\partial f(\bar{x})}(\bar{g})$; see Lemma 2 below.

Recently, an alternative approach yielding finite (pointwise) second-order derivatives for a certain Lagrangian associated with a nonsmooth convex function f was introduced. It comes from considering a suitable subspace, tangentially containing f's "smooth" behavior. This function is the *$\mathcal{U}$-Lagrangian* of f, introduced in [LOS00] and further studied in [MS99], [MS00b], [MS00a], [Ous00]. The basic idea is to decompose $\mathbb{R}^n$ into two orthogonal subspaces $\mathcal{V}$ and $\mathcal{U}$ depending on a fixed $\bar{x}$ in such a way that near $\bar{x}$ f's nonsmoothness is concentrated essentially in $\mathcal{V}$:

$$\mathcal{V} := \operatorname{lin}(\partial f(\bar{x}) - g) \qquad \text{and} \qquad \mathcal{U} := \mathcal{V}^{\perp}. \tag{2}$$

Since $\mathbb{R}^n = \mathcal{U} \oplus \mathcal{V}$, every $x \in \mathbb{R}^n$ can be decomposed into components $x_{\mathcal{U}}$ and $x_{\mathcal{V}}$, in which case we write $x = x_{\mathcal{U}} \oplus x_{\mathcal{V}}$. The $\mathcal{U}$-Lagrangian, denoted by $L_{\mathcal{U}}(\cdot, \bar{g}_{\mathcal{V}})$ depends on a particular subgradient $\bar{g}$, and is defined in the following constructive manner:

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

The introduction of this associated function allows us to restrict the second-order study of f to that of $L_{\mathcal{U}}$ on $\mathcal{U}$. In particular, when $\bar{g} \in \operatorname{ri}\partial f(\bar{x})$, the relative interior of the subdifferential of f at $\bar{x}$, the subspace $\mathcal{U}$ is the normal cone $N_{\partial f(\bar{x})}(\bar{g})$, so the second order analysis of f is restricted to $N_{\partial f(\bar{x})}(\bar{g})$, a set of directions for which the second-order epi-derivatives are finite.

The primary feature of the $\mathcal{U}$-Lagrangian is that is has a gradient at $u = 0$. Therefore, under appropriate assumptions, $L_{\mathcal{U}}$ may have a *generalized Hessian*, defined as the pointwise limit of a difference quotient, see [HU83] and also [RW98, Def. 13.1 p. 580]. Our main result shows that for a convex function given by (1), the second-order epi-derivative of f on $\mathcal{U}$ space and the generalized Hessian of the $\mathcal{U}$-Lagrangian are equivalent second-order objects when the latter exists. In other words, the $\mathcal{U}$-Lagrangian can be considered as the function that captures the second-order epi-differential behavior of f via ordinary second derivatives.

This paper is organized as follows. Section 2 contains all of the relevant results characterizing the first and second order epi-derivatives for a max-function given by (1). For completeness, we also give expressions for the parabolic derivatives introduced in [BTZ82]. The material in this section follows the presentation and notation in [RW98], unifying the earlier results on this subject in [Roc88], [PR94]. and [PR95]. Section 3 recalls some elements from the $\mathcal{VU}$-theory and gives Theorem 1, the link between epi-derivatives and $\mathcal{U}$-Hessians. Finally, in Section 4, we consider three convex functions from [MS99]. Example 1 does not satisfy Theorem 1's linear independencce assumption and does not have a $\mathcal{U}$-Hessian, but does have second order epi-derivatives in all directions. Examples 2 and 3 show that the second epi-derivative and the $\mathcal{U}$-Hessian coincide even in some cases that do not satisfy the assumptions of Theorem 1.

2 Epi-derivatives for finite max-functions

Before giving expressions for the epi-derivatives of a max-function given by (1), we recall the formal definition of epigraphical convergence.

2.1 Characterization of epi-limits

The epigraphical convergence theory developed in [RW98, Ch. 7] includes the following useful characterization of epi-limits:

Proposition 7.2 in [RW98, p. 241]

Let $\{q^\nu\}$ be a sequence of functions on $\mathbb{R}^n$, and let w be any point in $\mathbb{R}^n$. Then the *lower epi-limit* of the sequence at w is given by the expression

$$\text{e–}\liminf_{\nu} q^\nu(w) := \min\{\alpha \in \overline{\mathbb{R}} : \exists w^\nu \to w \text{ with } \liminf_{\nu} q^\nu(w^\nu) = \alpha\}\,.$$

Likewise, for the *upper epi-limit*,

$$\text{e–}\limsup_{\nu} q^\nu(w) := \min\{\alpha \in \overline{\mathbb{R}} : \exists w^\nu \to w \text{ with } \limsup_{\nu} q^\nu(w^\nu) = \alpha\}\,.$$

When these two limits coincide their common value is denoted by $q(w)$ and is called the *epi-limit* of the sequence at w. Thus, q^ν epi-converges to q if and only if at each point w one has

$$\begin{cases} \liminf_\nu q^\nu(w^\nu) \geq q(w) & \text{for every sequence } w^\nu \to w\,, \\ \limsup_\nu q^\nu(w^\nu) \leq q(w) & \text{for some sequence } w^\nu \to w\,. \end{cases}$$

$\square$

Note that in the expression above the limit $q(w)$ can be infinity.

2.2 First order derivatives

We now give expressions for the epi-derivatives of a function as in (1). We follow the notation in [RW98, Ex. 8.31, p. 321], except for the index 0.

Let $I(x) := \{i \in \{0, \cdots, m\} : f_i(x) = f(x)\}$ be the set of active indices at x; given $\bar{x} \in dom\, f = \mathbb{R}^n$, we use the notation $\bar{I}$ for the set $I(\bar{x})$. Without loss of generality we may assume that $\bar{I} = \{0, \cdots, m_1\}$ where $0 \leq m_1 \leq m$.

Consider $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $\bar{x} \in dom f$ with $f(\bar{x})$ finite. The *first subderivative* function of f at $\bar{x}$ is defined by the lower epi-limit of the first-order difference quotient $\frac{f(\bar{x}+\tau\cdot)-f(\bar{x})}{\tau}$:

$$\mathrm{d}f(\bar{x}) := \operatorname*{e\text{–}liminf}_{\tau \searrow 0} \frac{f(\bar{x} + \tau\cdot) - f(\bar{x})}{\tau},$$

i.e., for each $w \in \mathbb{R}^n$,

$$\mathrm{d}f(\bar{x})(w) = \liminf_{\tau \searrow 0, w' \to w} \frac{f(\bar{x} + \tau w') - f(\bar{x})}{\tau}.$$

When the first-order difference quotient above epi-converges to $\mathrm{d}f(\bar{x})$ as $\tau \searrow 0$, f is said to be *epi-differentiable*, and $\mathrm{d}f(\bar{x})(w)$ is the *first epi-derivative* of f at $\bar{x}$ for w.

In particular, for the objective function in (1), see [RW98, Ex. 8.31, p. 321], the Clarke subdifferential of f at $\bar{x}$ is given by $\partial f(\bar{x}) = conv\{\nabla f_i(\bar{x}) : i \in \bar{I}\}$ and the first epi-derivative is given by

$$\mathrm{d}f(\bar{x})(w) = \max_{i \in \bar{I}} \{\langle \nabla f_i(\bar{x}), w\rangle\}, \tag{4}$$

the well-known directional derivative. In addition, since f is regular (as defined in [Cla83, Def. 2.3.4]), $\mathrm{d}f(\bar{x})$ is the support function of $\partial f(\bar{x})$, [RW98, Theorem 8.30, p. 320]. Thus, if we let

$$I(\bar{x}, w) := \{i \in \bar{I} : \mathrm{d}f(\bar{x})(w) = \langle \nabla f_i(\bar{x}), w\rangle\}$$

then

$$\langle \nabla f_j(\bar{x}), w\rangle < \mathrm{d}f(\bar{x})(w) = \langle \nabla f_i(\bar{x}), w\rangle \text{ for all } j \in \bar{I}\backslash I(\bar{x}, w) \text{ and all } i \in I(\bar{x}, w). \tag{5}$$

Lemma 1 *Let f be a max-function as defined in* (1), *$\bar{x} \in dom f = \mathbb{R}^n$, $\bar{g} \in \partial f(\bar{x})$, and $w \in \mathbb{R}^n$. For the sets*

$$N_{\partial f(\bar{x})}(\bar{g}) := \{z \in \mathbb{R}^n : \langle z, g\rangle \leq \langle z, \bar{g}\rangle \text{ for all } g \in \partial f(\bar{x})\} \text{ and } \partial f(\bar{x}|w) := conv\{\nabla f_i(\bar{x}) : i \in I(\bar{x}, w)\}$$

the following hold:

(i) if $w \in N_{\partial f(\bar{x})}(\bar{g})$ then for any nonnegative numbers α_i for $i \in \bar{I}$, which sum to 1 and satisfy $\bar{g} = \sum_{i \in \bar{I}} \alpha_i \nabla f_i(\bar{x})$ it follows that $\alpha_i = 0$ for all $i \in \bar{I}\backslash I(\bar{x}, w)$;

(ii) $w \in N_{\partial f(\bar{x})}(\bar{g}) \iff \bar{g} \in \partial f(\bar{x}|w)$;

(iii) $I(\bar{x}, w) = \bar{I}$ *for all* $w \in \mathcal{U}$, *the subspace defined in* (2);

(iv) $\mathcal{U} \subseteq N_{\partial f(\bar{x})}(\bar{g})$ *for all* $\bar{g} \in \partial f(\bar{x})$, *with equality holding if and only if* $\bar{g} \in \mathrm{ri}\partial f(\bar{x})$.

Proof. [(i)] Suppose for purposes of contradiction, that $\alpha_j > 0$ for some $j \in \bar{I} \backslash I(\bar{x}, w)$. Since $w \in N_{\partial f(\bar{x})}(\bar{g})$ and $\nabla f_i(\bar{x}) \in \partial f(\bar{x})$ for all $i \in \bar{I}$, the definition of $N_{\partial f(\bar{x})}(\bar{g})$ and (4) imply that $\mathrm{d}f(\bar{x})(w) \leq \langle w, \bar{g} \rangle$. Together with the expression for $\bar{g}$ this implies that

$$\mathrm{d}f(\bar{x})(w) \leq \textstyle\sum_{i \in \bar{I}} \alpha_i \langle w, \nabla f_i(\bar{x}) \rangle \tag{6}$$

with $\alpha_i \geq 0$ and $\sum_{i \in \bar{I}} \alpha_i = 1$. Now, since $\alpha_j > 0$ and $j \in \bar{I} \backslash I(\bar{x}, w)$, (5) implies that the right hand side of (6) is strictly less than $\mathrm{d}f(\bar{x})(w)$, which is a contradiction.
[(ii)] The forward implication follows from item (i). To show the reverse implication suppose that $\bar{g} \in \partial f(\bar{x}|w)$. Then, by the definitions of $I(\bar{x}, w)$ and $\partial f(\bar{x}|w)$, $\langle w, \bar{g} \rangle = \mathrm{d}f(\bar{x})(w)$ and, since $\mathrm{d}f(\bar{x})(\cdot)$ is the support function of $\partial f(\bar{x})$, $w \in N_{\partial f(\bar{x})}(\bar{g})$.
[(iii)] Suppose $i \in \bar{I}$ and $w \in \mathcal{U}$. Since $0 \in \bar{I}$, $\nabla f_0(\bar{x}) \in \partial f(\bar{x})$, so taking $g = \nabla f_0(\bar{x})$ in (2) gives $\nabla f_i(\bar{x}) - \nabla f_0(\bar{x}) \in \mathcal{V}$. Then, since $\mathcal{U} = \mathcal{V}^\top$, $\langle \nabla f_i(\bar{x}) - \nabla f_0(\bar{x}), w \rangle = 0$. Thus, $\langle \nabla f_i(\bar{x}), w \rangle$ is the same value (depending on w) for all $i \in \bar{I}$ and the desired result follows from the definition of $I(\bar{x}, w)$.
[(iv)] From [LOS00, Proposition 2.2], $\mathcal{U} \subseteq N_{\partial f(\bar{x})}(\bar{g})$ and there is equality if $\bar{g} \in \mathrm{ri}\partial f(\bar{x})$. If the normal cone $N_{\partial f(\bar{x})}(\bar{g}) = \mathcal{U}$ then $N_{\partial f(\bar{x})}(\bar{g})$ is a subspace and [LS96, Proposition 2.1] (see also [Roc70, Theorem 2.7]) implies that $\bar{g} \in \mathrm{ri}\partial f(\bar{x})$. □

2.3 Second order derivatives

Consider $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, $\bar{x} \in dom f$ with $f(\bar{x})$ finite, and any $g \in \mathbb{R}^n$. The *second subderivative* of f at $\bar{x}$ for $g \in \mathbb{R}^n$ is defined by the lower epi-limit of the second-order difference quotient $\frac{f(\bar{x}+\tau\cdot) - f(\bar{x}) - \tau\langle g, \cdot\rangle}{\frac{1}{2}\tau^2}$:

$$\mathrm{d}^2 f(\bar{x}|g) := \operatorname*{e\text{–}liminf}_{\tau \searrow 0} \frac{f(\bar{x} + \tau\cdot) - f(\bar{x}) - \tau \langle g, \cdot \rangle}{\frac{1}{2}\tau^2}.$$

In other words, for each $w \in \mathbb{R}^n$,

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

When the second-order difference quotient above epi-converges to $\mathrm{d}^2 f(\bar{x}|g)$ as $\tau \searrow 0$, f is said to be *twice epi-differentiable*, and $\mathrm{d}^2 f(\bar{x}|g)$ is the *second epi-derivative* of f at $\bar{x}$ for g. If, in addition, $\mathrm{d}^2 f(\bar{x})$ is a proper function (i.e., not everywhere equal to $+\infty$), then f is properly twice epi-differentiable; [RW98, Def. 13.6, p. 586].

It important to note that, in this epigraphical setting, the second-order epi-derivative provides a second-order approximation $f(\bar{x} + \tau w) \approx f(\bar{x}) + \tau \langle g, w \rangle + \frac{1}{2}\tau^2 \mathrm{d}^2 f(\bar{x}|g)(w)$, not in the usual sense of local uniform convergence, but of closeness of the epigraphs of the difference quotient function $\frac{f(\bar{x}+\tau\cdot)-f(\bar{x})-\tau\langle g,\cdot\rangle}{\frac{1}{2}\tau^2}$ and $\mathrm{d}^2 f(\bar{x}|g)(\cdot)$, see [PR95].

Lemma 2 *Let f be a max-function as defined in* (1). *For $\bar{x} \in dom f = \mathbb{R}^n$ and any $\bar{g} \in \partial f(\bar{x})$, f is properly twice epi-differentiable, $dom\, \mathrm{d}^2 f(\bar{x}|\bar{g}) = N_{\partial f(\bar{x})}(\bar{g})$ and the following hold:*

(i) for all $w \in N_{\partial f(\bar{x})}(\bar{g})$ the second epi-derivative has the expression

$$\mathrm{d}^2 f(\bar{x}|\bar{g})(w) = \max\left\{\left\langle w, \sum_{i\in I(\bar{x},w)} \alpha_i \nabla^2 f_i(\bar{x}) w\right\rangle : \alpha \in \Delta_{\bar{g}}(\bar{x},w)\right\},$$

$$\textit{with } \Delta_{\bar{g}}(\bar{x},w) := \left\{\alpha \in \mathbb{R}^{|I(\bar{x},w)|} : \alpha_i \geq 0 \textit{ for } i \in I(\bar{x},w),\ \sum_{i\in I(\bar{x},w)} \alpha_i = 1 \textit{ and } \bar{g} = \sum_{i\in I(\bar{x},w)} \alpha_i \nabla f_i(\bar{x})\right\},$$

which is a singleton if the set $\{(\nabla f_i(\bar{x})^\top, 1)\}_{i\in I(\bar{x},w)} \subset \mathbb{R}^{n+1}$ is linearly independent;

(ii) if, in addition, the set $\{\nabla f_i(\bar{x}) - \nabla f_0(\bar{x})\}_{0\neq i\in\bar{I}}$ is linearly independent then for all $w \in \mathcal{U}$ the set $\Delta_{\bar{g}}(\bar{x},w)$ is a singleton $\{\bar{\alpha}\}$,

$$\mathrm{d}^2 f(\bar{x}|\bar{g})(w) = \left\langle w, \sum_{i\in\bar{I}} \bar{\alpha}_i \nabla^2 f_i(\bar{x}) w\right\rangle$$

and $\bar{\alpha}_i > 0$ for all $i \in \bar{I}$ if $\bar{g} \in \mathrm{ri}\partial f(\bar{x})$.

Proof. The facts that $dom\,\mathrm{d}^2 f(\bar{x}|\bar{g}) = N_{\partial f(\bar{x})}(\bar{g})$ and item (i) hold follow directly from combining (i) and (ii) of our Lemma 1 with [RW98, Ex. 13.16 p. 600], written with $v = \bar{g}$ and noting that $K(\bar{x}, v)$ therein stands for $N_{\partial f(\bar{x})}(\bar{g})$.
To see (ii), first note that, by Lemma 1(iii) and (iv), the index set $I(\bar{x},w)$ is $\bar{I}$ for all $w \in \mathcal{U} \subseteq N_{\partial f(\bar{x})}(\bar{g})$. Then, the linear independence assumption implies that the multiplier vector set $\Delta_{\bar{g}}(\bar{x},w)$ is a singleton. Taking this singleton to be $\bar{\alpha}$ and using part (i) gives the expression for $\mathrm{d}^2 f(\bar{x}|\bar{g})(w)$. Finally, from [HUL93, Remark III.2.1.4], we have that

$$\mathrm{ri}\, conv\{\nabla f_i(\bar{x}) : i \in \bar{I}\} = \left\{\sum_{i\in\bar{I}} \alpha_i \nabla f_i(\bar{x}) : \alpha_i > 0 \text{ for } i \in \bar{I}, \sum_{i\in\bar{I}} \alpha_i = 1\right\}.$$

Thus, all $|\bar{I}|$ components of the multiplier vector in $\Delta_{\bar{g}}(\bar{x},w)$ are positive if $\bar{g} \in \mathrm{ri}\partial f(\bar{x})$. □

Explicit formulæ for the second order epi-derivatives of a finite max-function follow from [Roc88, Theorem 4.5] and were first given in [PR94, Theorem 2] in the context of protoderivative calculus for the mapping ∂f. Strict second order epi-differentiability under the assumptions in Lemma 2(ii) above was established in [PR95, Proposition 4.10].

2.4 Parabolic derivatives

We now turn our attention to parabolic derivatives, defined somewhat like traditional second order directional derivatives, but taking (epi) limits along parabolas instead of straight lines.

Consider $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $\bar{x} \in dom f$ with $f(\bar{x})$ finite. The *parabolic second subderivative* of f at $\bar{x}$ for $w \in \mathbb{R}^n$ is defined by the lower epi-limit of the parabolic second-order difference quotient $\frac{f(\bar{x}+\tau w+\frac{1}{2}\tau^2\cdot)-f(\bar{x})-\tau \mathrm{d}f(\bar{x})(w)}{\frac{1}{2}\tau^2}$:

$$\mathrm{d}^2 f(\bar{x})(w|\cdot) := \text{e–}\liminf_{\tau \searrow 0} \frac{f(\bar{x}+\tau w+\frac{1}{2}\tau^2\cdot)-f(\bar{x})-\tau \mathrm{d}f(\bar{x})(w)}{\frac{1}{2}\tau^2} .$$

In other words, with respect to $z \in \mathbb{R}^n$,

$$\mathrm{d}^2 f(\bar{x})(w|z) = \liminf_{\tau \searrow 0, z' \to z} \frac{f(\bar{x}+\tau w+\frac{1}{2}\tau^2 z')-f(\bar{x})-\tau \mathrm{d}f(\bar{x})(w)}{\frac{1}{2}\tau^2} .$$

When the parabolic second-order difference quotient above epi-converges to a proper function $\mathrm{d}^2 f(\bar{x})(w|\cdot)$ as $\tau \searrow 0$, f is said to be *parabolically epi-differentiable* at $\bar{x}$ for w, [RW98, Def. 13.59, p. 633].

It was first shown in [Roc88, Theorem 4.7] that second order parabolic derivatives and epi-derivatives are somewhat dual to each other. Below we give the result as stated in [RW98] for a special fully amenable function.

A function is said to be *fully amenable* at $\bar{x} \in dom\, f$ if there is an open neighborhood of $\bar{x}$ on which f can be represented in the form $f = g \circ F$, with g piecewise linear-quadratic and F a C^2-mapping satisfying a certain basic constraint qualification, see [RW98, Def. 10.23].

In particular, see [RW98, Ex. 10.24(e), p. 443], for a max-function as in (1) the constraint qualification is always satisfied. Moreover, f is fully amenable, i.e., $f = g \circ F$, for $F := (f_0, \cdots, f_m)$ and $g(u) := vecmax(u) := \max(u_0, \cdots, u_m)$, a piecewise linear function.

For finite-valued functions, [BTZ82] considered parabolic derivatives defined as the pointwise limit of the second-order parabolic difference quotient above, with $\mathrm{d}f(\bar{x})(w)$ therein replaced by the classical directional derivative. For (finite-valued) fully amenable functions this concept coincides with the parabolic second epi-derivative, see [Roc88].

Lemma 3 *Let f be a max-function as defined in* (1)*. For $\bar{x} \in dom f = \mathbb{R}^n$, f is parabolically epi-differentiable at $\bar{x}$ and the following hold for all $w, z \in \mathbb{R}^n$:*

(i) the function $z \mapsto \mathrm{d}^2 f(\bar{x})(w|z)$ is conjugate to the function

$$g \mapsto \begin{cases} -\mathrm{d}^2 f(\bar{x}|g)(w) & \textit{if } g \in \partial f(\bar{x}|w) \\ +\infty & \textit{otherwise;} \end{cases}$$

(ii) the parabolic epi-derivative has the expression

$$\mathrm{d}^2 f(\bar{x})(w|z) = \max\left\{\left\langle w, \nabla^2 f_i(\bar{x}) w\right\rangle + \left\langle \nabla f_i(\bar{x}), z\right\rangle : i \in I(\bar{x}, w)\right\} .$$

Proof. Item (i) is just the first part of [RW98, Theorem 13.67, pp. 636-637].
To see (ii), we write f in (1) in a fully amenable representation, as above. From [RW98, p. 318 and p. 600, respectively], $\partial g(u) = conv\{e_i \in \mathbb{R}^{m+1} : i \text{ such that } u_i = vecmax\, u\}$ and $\mathrm{d}^2 g(u)(w) = 0$ for all $u \in \mathbb{R}^{m+1}$ and $w \in \mathbb{R}^n$. Therefore, equation 13(40) in [RW98, Theorem 13.67] gives, for all $w\,, z \in \mathbb{R}^n$, the following expression for the parabolic epi-derivative :

$$\mathrm{d}^2 f(\bar{x})(w|z) = \max\left\{\sum_{i=0}^{m} \alpha_i \left(\left\langle w, \nabla^2 f_i(\bar{x}) w\right\rangle + \langle \nabla f_i(\bar{x}), z\rangle\right) : \alpha \in Y(\bar{x}|w)\right\}$$

where $Y(\bar{x}|w) := \arg\max\{\sum_{i=0}^{m} \alpha_i \langle \nabla f_i(\bar{x}), w\rangle : \alpha \in conv\{e_i \in \mathbb{R}^{m+1} : i \in \bar{I}\}$. For any $\alpha \in Y(\bar{x}|w)$ the definition of $Y(\bar{x}|w)$ implies that $\alpha_i = 0$ for all $i \in \{0, \ldots, m\}\backslash\bar{I}$ and the relations in (5) further imply that $\alpha_j = 0$ for all $j \in \bar{I}\backslash I(\bar{x}, w)$, so the maximization in the expression for $\mathrm{d}^2 f(\bar{x})(w|z)$ is over

$$\{\alpha \in \mathbb{R}^{m+1} : \sum_{i \in I(\bar{x},w)} \alpha_i = 1\,, \alpha_i \geq 0 \text{ for } i \in I(\bar{x}, w)\,, \alpha_i = 0 \text{ for } i \in \{0, \ldots, m\}\backslash I(\bar{x}, w)\}\,.$$

As a function of α, the maximization problem in this expression is a linear program with a simplicial feasible set, thus, it is enough to take the maximum over the set of unit vector extreme points, which gives the desired result. □

3 Link with $\mathcal{VU}$-theory

We start by reviewing $\mathcal{VU}$-space decomposition and $\mathcal{U}$-Lagrangians from [LOS00] and [MS00b]. Throughout this section we assume that $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function.

3.1 $\mathcal{VU}$-space decomposition

Consider the subspaces $\mathcal{V}$ and $\mathcal{U}$ defined in (2). Letting $\bar{V}$ be a basis matrix for $\mathcal{V}$, not necessarily orthonormal, and letting $\bar{U}$ be an orthonormal basis matrix for $\mathcal{U}$ every $x \in \mathbb{R}^n$ can be decomposed into components $x_{\mathcal{U}}$ and $x_{\mathcal{V}}$ as follows:

$$\begin{array}{rcllll} \mathbb{R}^n \ni x & = & \bar{U} & (\bar{U}^\top x) & + & \bar{V} \quad \left([\bar{V}^\top \bar{V}]^{-1}\bar{V}^\top x\right) \\ & = & \bar{U} & x_{\mathcal{U}} & + & \bar{V} \quad x_{\mathcal{V}} \\ & = & & x_{\mathcal{U}} & \oplus & \quad x_{\mathcal{V}} \qquad \in \mathbb{R}^{dim\,\mathcal{U}} \times \mathbb{R}^{dim\,\mathcal{V}}\,. \end{array}$$

The reason why we do not assume that $\bar{V}$ is orthonormal is because typical $\mathcal{V}$-basis matrix approximations made by bundle methods are not orthonormal.

3.2 $\mathcal{U}$-Lagrangians and $\mathcal{U}$-Hessians

Given a subgradient $\bar{g} \in \partial f(\bar{x})$ the $\mathcal{U}$-Lagrangian of f is defined by (3). Using the above matrix notation, $L_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}}) = \min_{v \in \mathbb{R}^{dim\,\mathcal{V}}} \{f(\bar{x} + \bar{U}u + \bar{V}v) - \bar{g}^\top \bar{V}v\}$ for $u \in \mathbb{R}^{dim\,\mathcal{U}}$. The associated set of $\mathcal{V}$-space minimizers is defined by

$$W(u; \bar{g}_{\mathcal{V}}) := \{\bar{V}v : L_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}}) = f(\bar{x} + \bar{U}u + \bar{V}v) - \bar{g}^\top \bar{V}v\}\,.$$

From [LOS00], $W(u; \bar{g}_{\mathcal{V}})$ is nonempty if $\bar{g} \in \mathrm{ri}\,\partial f(\bar{x})$, but this is not a necessary condition, see [MS99] and [MS00b]. Each $\mathcal{U}$-Lagrangian is a convex function that is differentiable at $u = 0$ with

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

Whenever $\nabla^2 L_{\mathcal{U}}(0; \bar{g}_{\mathcal{V}})$ exists (possibly as a generalized Hessian, as defined in [HU83]), those $x(u) := \bar{x} + u \oplus v$ with $\bar{V}v \in W(u)$ yield the following:

$$\begin{aligned} f(x(u)) &= f(\bar{x}) + \nabla L_{\mathcal{U}}(0; \bar{g}_{\mathcal{V}})^\top u + \bar{g}^\top \bar{V} v + \frac{1}{2} u^\top \nabla^2 L_{\mathcal{U}}(0; \bar{g}_{\mathcal{V}}) u + o(\|u\|^2) \\ &= f(\bar{x}) + \langle \bar{g}, u \oplus v\rangle + \frac{1}{2} u^\top \nabla^2 L_{\mathcal{U}}(0; \bar{g}_{\mathcal{V}}) u + o(\|u\|^2). \end{aligned} \tag{7}$$

Because of this expansion, we call $\nabla^2 L_{\mathcal{U}}(0; \bar{g}_{\mathcal{V}})$ a "$\mathcal{U}$-Hessian" of f at $\bar{x}$ relative to $\bar{g}$.

Altogether, we have our main result relating $\mathcal{U}$-Hessians to second-order epi-derivatives.

Theorem 1 *Suppose f is a convex function and a max-function as defined in* (1). *Let $\bar{x} \in$ dom $f = \mathbb{R}^n$ and let $\bar{g}$ be a subgradient such that $\bar{g} \in \mathrm{ri}\,\partial f(\bar{x})$. If the set $\{\nabla f_i(\bar{x}) - \nabla f_0(\bar{x})\}_{0 \neq i \in \bar{I}}$ is linearly independent then for all $w \in \mathcal{U}$*

$$\mathrm{d}^2 f(\bar{x}|\bar{g})(w_{\mathcal{U}} \oplus 0) = \left\langle w_{\mathcal{U}} \oplus 0, \left[\sum_{i \in \bar{I}} \bar{\alpha}_i \nabla^2 f_i(\bar{x})\right] (w_{\mathcal{U}} \oplus 0) \right\rangle = w_{\mathcal{U}}^\top \nabla^2 L_{\mathcal{U}}(0; \bar{g}_{\mathcal{V}}) w_{\mathcal{U}}\,,$$

with $\{\bar{\alpha}\} = \Delta_{\bar{g}}(\bar{x}, w_{\mathcal{U}} \oplus 0)$ and $w_{\mathcal{U}} = \bar{U}^\top w$.

Proof. The left hand side equality follows from Lemma 2(ii). To obtain the right hand side equality, we apply the results stated in [MS00b] for functions having `pdg` structure with $m_2 = 0$. Since $\bar{g} \in \mathrm{ri}\,\partial f(\bar{x})$, [MS00b, Theorem 6.1(i)] implies that the set $\bar{I} = \{0, \ldots, m_1\}$ is strongly $\mathcal{V}$-optimal, so the desired result follows from [MS00b, Theorems 6.3 and 5.2] with $\bar{K} := \bar{I}$. □

4 Examples

We consider the convex example functions in [MS99, Ex. 14, 15, 17], but with all vectors considered to be column vectors as in [MS00b] and with $x = (x_1, x_2)^\top \in \mathbb{R}^2$. They are all considered at the point $\bar{x} = (0, 0)^\top$ which is also a minimizing point for each function.

The first example has a singleton subdifferential at $\bar{x}$, second-order epi-derivatives in all directions, but no $\mathcal{U}$-Hessian. This example appears in [OY94, p. 103].

Example 1 (Example 14 in [MS99]) Consider the function

$$f(x_1, x_2) = \max\{f_0(x_1, x_2), f_1(x_1, x_2)\} := \max\left\{\frac{1}{2}x_1^2 + x_2^2,\ x_1^2 + \frac{1}{2}x_2^2\right\}.$$

Then $\partial f(\bar{x}) = \{\bar{g} = (0, 0)^\top\}$, $\mathcal{V} = \{(0, 0)^\top\}$, $\mathcal{U} = \mathbb{R}^2$, and $L_{\mathcal{U}} \equiv f$.
At $\bar{x} = (0, 0)^\top$ the active index set $\bar{I}$ is $\{0, 1\}$.

We now give the parabolic and second order epi-derivatives for this example.

Since $\nabla f_0(\bar{x}) = \nabla f_1(\bar{x}) = (0, 0)^\top$, the first epi-derivative is $\mathrm{d}f(\bar{x})(w) = 0$ for all $w \in \mathbb{R}^2$, so $I(\bar{x}, w) = \{0, 1\}$ and, since $\bar{g} = (0, 0)^\top$, the normal cone is $N_{\partial f(\bar{x})}(\bar{g}) = \mathbb{R}^2$, and the set $\Delta_{\bar{g}}(\bar{x}, w) = \{\alpha \in [0, 1]^2 : \alpha_1 = 1 - \alpha_0\}$. Therefore, taking into account the two structure Hessians, by Lemma 2(i), for all $w \in \mathbb{R}^2$

$$\mathrm{d}^2 f(\bar{x}|\bar{g})(w) = \max_{\alpha_1 \in [0,1]}\left\{\left\langle w, \begin{pmatrix} 2 - \alpha_1 & 0 \\ 0 & 1 + \alpha_1 \end{pmatrix} w\right\rangle\right\} = \begin{cases} \left\langle w, \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} w\right\rangle & \text{if } |w_1| \le |w_2| \\ \left\langle w, \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} w\right\rangle & \text{if } |w_1| \ge |w_2|\,. \end{cases}$$

Moreover, since both gradients are zero vectors, from Lemma 3(ii), $\mathrm{d}^2 f(\bar{x})(w|z) = \mathrm{d}^2 f(\bar{x}|\bar{g})(w)$ for all $w, z \in \mathbb{R}^2$. However, there is no $\mathcal{U}$-Hessian for this function, because $\mathcal{U} = \mathbb{R}^2$, $L_{\mathcal{U}} \equiv f$ and f has no Hessian. Moreover, since the $\mathcal{U}$-Hessian of this f at $(0, 0)^\top$ does not exist and $\mathcal{U} = \mathbb{R}^2$, Theorem 5.3$(ii)$ of [LOS00] implies that the Moreau-Yosida regularization of this function does not have a Hessian at $(0, 0)^\top$.

For the next two examples all of the underlying C^2-functions are active at $\bar{x} = (0, 0)^\top$ and

$$\mathcal{V} = \operatorname{lin}\{(0, 1)^\top\} \quad \text{and} \quad \mathcal{U} = \operatorname{lin}\{(1, 0)^\top\}\,. \tag{8}$$

Since $\mathcal{U} = \operatorname{lin}\{(1, 0)^\top\}$ we take $\bar{U} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ as a basis matrix for $\mathcal{U}$.

Example 2 (Example 15 in [MS99]) Let $f(x_1, x_2) := \max\{\frac{1}{2}(x_1^2 + x_2^2), x_2\}$. Then for (1) the relevant C^2-functions and gradients are

$$\begin{array}{lcl} f_0(x_1, x_2) := \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 & \text{and} & f_1(x_1, x_2) := x_2\,; \\ \nabla f_0(x_1, x_2) = (x_1, x_2)^\top & \text{and} & \nabla f_1(x_1, x_2) = (0, 1)^\top. \end{array}$$

Note that at $\bar{x} = (0, 0)^\top$ the active index set is $\bar{I} = \{0, 1\}$, so

$$\partial f(\bar{x}) = \operatorname{conv}\{(0, 0)^\top, (0, 1)^\top\} = \{(0, \gamma)^\top : \gamma \in [0, 1]\}$$

and $\mathcal{V}$ and $\mathcal{U}$ are given by (8).

From (4) for all $w \in \mathbb{R}^2$ the first epi-derivative is given by $\mathrm{d}f(\bar{x})(w) = \max\{0, w_2\}$, so

$$I(\bar{x}, w) = \begin{cases} \{0\} & \text{if } w_2 < 0 \\ \{0,1\} & \text{if } w_2 = 0 \\ \{1\} & \text{if } w_2 > 0\,. \end{cases}$$

For the subgradients $\bar{g} = (0, \gamma)^\top \in \partial f(\bar{x})$, $\gamma \in [0,1]$ the normal cone is given by

$$N_{\partial f(\bar{x})}(\bar{g}) = \begin{cases} \{(w_1, w_2) : w_2 \leq 0\} & \text{if } \gamma = 0 \\ \{(w_1, w_2) : w_2 = 0\} = \mathcal{U} & \text{if } 0 < \gamma < 1 \\ \{(w_1, w_2) : w_2 \geq 0\} & \text{if } \gamma = 1\,. \end{cases}$$

For $w \in N_{\partial f(\bar{x})}(\bar{g})$ the multiplier sets $\Delta_{\bar{g}}(\bar{x}, w)$ defined in Lemma 2(i) are singletons given by

$$\Delta_{\bar{g}}(\bar{x}, w) = \begin{cases} \{\alpha_0 = 1\} & \text{if } \gamma = 0 \text{ and } w_2 \leq 0 \\ \{(\alpha_0, \alpha_1) = (1-\gamma, \gamma)\} & \text{if } 0 < \gamma < 1 \text{ and } w_2 = 0 \\ \{\alpha_1 = 1\} & \text{if } \gamma = 1 \text{ and } w_2 \geq 0\,. \end{cases}$$

The structure Hessians are given by $\nabla^2 f_0(\bar{x}) = I$ and $\nabla^2 f_1(\bar{x}) = 0$. So we also have from Lemma 2 (writing only the finite values) that

$$\mathrm{d}^2 f(\bar{x}|\bar{g})(w) = \begin{cases} w_1^2 + w_2^2 & \text{if } \gamma = 0 \text{ and } w_2 \leq 0 \\ (1-\gamma)w_1^2 & \text{if } 0 < \gamma < 1 \text{ and } w_2 = 0 \\ 0 & \text{if } \gamma = 1 \text{ and } w_2 \geq 0 \end{cases}$$

and from Lemma 3 for all $w, z \in \mathbb{R}^2$ the parabolic derivatives are given by

$$\mathrm{d}^2 f(\bar{x})(w|z) = \begin{cases} w_1^2 + w_2^2 & \text{if } w_2 < 0 \\ \max(w_1^2, z_2) & \text{if } w_2 = 0 \\ z_2 & \text{if } w_2 > 0\,. \end{cases}$$

From [MS99] for $u \in \mathbb{R}^1 = \mathbb{R}^{dim\, \mathcal{U}}$ and $\bar{g}_{\mathcal{V}} = \gamma$,

$$L_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}}) = \begin{cases} (1-\gamma)(1 - \sqrt{1-u^2}) & \text{if } 0 < \gamma \leq 1 \\ \frac{1}{2}u^2 & \text{if } \gamma = 0 \end{cases}$$

and

$$\nabla^2 L_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}}) = \begin{cases} (1-\gamma)(1-u^2)^{-3/2} & \text{if } 0 < \gamma \leq 1 \\ 1 & \text{if } \gamma = 0\,, \end{cases}$$

so for $u = 0$ the $\mathcal{U}$-Hessian is given by $\nabla^2 L_{\mathcal{U}}(0; \bar{g}_{\mathcal{V}}) = 1 - \gamma$ for all $\gamma \in [0,1]$.
If $w = (w_1, w_2)^\top \in \mathcal{U}$ then, using the matrix notation, $w_{\mathcal{U}} = \bar{U}^\top w = w_1$ and $w_2 = 0$, so

$$w_{\mathcal{U}}^\top \nabla^2 L_{\mathcal{U}}(0; \bar{g}_{\mathcal{V}}) w_{\mathcal{U}} = (1-\gamma) w_1^2\,,$$

which agrees with the expression above for $\mathrm{d}^2 f(\bar{x}|\bar{g})(w)$ when $w_2 = 0$ for all subgradients $\bar{g}$, including the two extremes ones, corresponding to $\gamma = 0$ and $\gamma = 1$, which are not in $\mathrm{ri}\,\partial f(\bar{x})$ and, hence, do not satisfy the corresponding assumption of Theorem 1.

Example 3 (Example 17 in [MS99]) Let $f(x_1,x_2) := \max\{\frac{1}{2}x_1^2, |x_2|\}$. Then for (1) the relevant C^2-functions and gradients are given by

$$\begin{array}{llll} f_0(x_1,x_2) = \frac{1}{2}x_1^2\,, & f_1(x_1,x_2) = x_2 & \text{and} & f_2(x_1,x_2) = -x_2\,; \\ \nabla f_0(x_1,x_2) = (x_1,0)^\top\,, & \nabla f_1(x_1,x_2) = (0,1)^\top & \text{and} & \nabla f_2(x_1,x_2) = (0,-1)^\top\,. \end{array}$$

Note that at $\bar{x} = (0,0)^\top$, the active set is $\bar{I} = \{0,1,2\}$,

$$\partial f(\bar{x}) = \operatorname{conv}\{(0,0)^\top, (0,1)^\top, (0,-1)^\top\} = \{(0,\gamma)^\top : \gamma \in [-1,1]\}\,,$$

and $\mathcal{V}$ and $\mathcal{U}$ are given by (8).
From (4) for all $w \in \mathbb{R}^2$ the first epi-derivative is given by $\mathrm{d}f(\bar{x})(w) = \max\{0, w_2, -w_2\}$, so

$$I(\bar{x},w) = \begin{cases} \{2\} & \text{if } w_2 < 0 \\ \{0,1,2\} & \text{if } w_2 = 0 \\ \{1\} & \text{if } w_2 > 0\,. \end{cases}$$

For the subgradients $\bar{g} = (0,\gamma)^\top \in \partial f(\bar{x})$, $\gamma \in [-1,1]$ the normal cone is given by

$$N_{\partial f(\bar{x})}(\bar{g}) = \begin{cases} \{(w_1,w_2) : w_2 \le 0\} & \text{if } \gamma = -1 \\ \{(w_1,w_2) : w_2 = 0\} = \mathcal{U} & \text{if } -1 < \gamma < 1 \\ \{(w_1,w_2) : w_2 \ge 0\} & \text{if } \gamma = 1\,. \end{cases}$$

For $w \in N_{\partial f(\bar{x})}(\bar{g})$ the multiplier sets $\Delta_{\bar{g}}(\bar{x},w)$ from Lemma 2(i) are given by

$$\Delta_{\bar{g}}(\bar{x},w) = \begin{cases} \{\alpha_2 = 1\} & \text{if } \gamma = -1 \text{ and } w_2 \le 0 \\ \left\{(1-2\alpha_1+\gamma, \alpha_1, \alpha_1-\gamma) : \alpha_1 \in \left[\max\{0,\gamma\}, \dfrac{\gamma+1}{2}\right]\right\} & \text{if } -1<\gamma<1 \text{ and } w_2 = 0 \\ \{\alpha_1 = 1\} & \text{if } \gamma = 1 \text{ and } w_2 \ge 0\,. \end{cases}$$

Note that for $-1 < \gamma < 1$ and $w_2 = 0$, i.e., $\bar{g} \in \operatorname{ri}\partial f(\bar{x})$ and $w \in \mathcal{U}$, this multiplier set is not a singleton, so the linear independence condition of Lemma 2(ii) does not hold. This is easy to verify, since here $\nabla f_i(\bar{x}) - \nabla f_0(\bar{x}) = (0, (-1)^{i-1})^\top$ for $i = 1,2$.
The structure Hessians are given by $\nabla^2 f_0(\bar{x}) = I$ and $\nabla^2 f_1(\bar{x}) = \nabla^2 f_2(\bar{x}) = 0$. So we have from Lemma 2 (writing only the finite values) that

$$\mathrm{d}^2 f(\bar{x}|\bar{g})(w) = \begin{cases} 0 & \text{if } \gamma = -1 \text{ and } w_2 \le 0 \\ (1-|\gamma|)w_1^2 & \text{if } -1<\gamma<1 \text{ and } w_2 = 0 \\ 0 & \text{if } \gamma = 1 \text{ and } w_2 \ge 0 \end{cases}$$

and from Lemma 3 for all $w, z \in \mathbb{R}^2$ the parabolic derivatives are given by

$$\mathrm{d}^2 f(\bar{x})(w|z) = \begin{cases} -z_2 & \text{if } w_2 < 0 \\ \max\{w_1^2, |z_2|\} & \text{if } w_2 = 0 \\ z_2 & \text{if } w_2 > 0\,. \end{cases}$$

From [MS99] for all $u \in \mathbb{R}^1 = \mathbb{R}^{dim\,\mathcal{U}}$ and $\bar{g}_{\mathcal{V}} = \gamma \in [-1, 1]$,

$$L_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}}) = \frac{1}{2}(1 - |\gamma|)u^2, \quad \nabla L_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}}) = (1 - |\gamma|)u, \text{ and } \quad \nabla^2 L_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}}) = 1 - |\gamma|.$$

If $w = (w_1, w_2)^\top \in \mathcal{U}$ then, using the matrix notation, $w_{\mathcal{U}} = \bar{U}^\top w = w_1$ and $w_2 = 0$, so

$$w_{\mathcal{U}}^\top \nabla^2 L_{\mathcal{U}}(0; \bar{g}_{\mathcal{V}}) w_{\mathcal{U}} = (1 - |\gamma|)w_1^2,$$

which agrees with the expression above for $\mathrm{d}^2 f(\bar{x}|\bar{g})(w)$ when $w_2 = 0$ for all subgradients $\bar{g}$, including the two extremes ones, corresponding to $\gamma = -1$ and $\gamma = 1$, which are not in $\mathrm{ri}\,\partial f(\bar{x})$. In addition, note that the $\mathcal{U}$-Hessian coincides with the second order epi-derivative even though the linear independence assumption of Theorem 1 does not hold.

References

[BTZ82] A. Ben-Tal and J. Zowe, *Necessary and sufficient optimality conditions for a class of nonsmooth minimization problems*, Math. Program. **24** (1982), no. 1, 70–91.

[Cla83] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, 1983, Reprinted by Siam, 1990.

[HU83] J.-B. Hiriart-Urruty, *The approximate first-order and second-order directional derivatives for a convex function*, Mathematical Theories of Optimization (J.-P. Cecconi and T. Zolezzi, eds.), Lecture Notes in Mathematics, no. 979, Springer-Verlag, 1983, pp. 154–166.

[HUL93] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms*, Grund. der math. Wiss, no. 305-306, Springer-Verlag, 1993, (two volumes).

[LOS00] C. Lemaréchal, F. Oustry, and C. Sagastizábal, *The $\mathcal{U}$-Lagrangian of a convex function*, Trans. Amer. Math. Soc. **352** (2000), no. 2, 711–729.

[LS96] C. Lemaréchal and C. Sagastizábal, *More than first-order developments of convex functions: primal-dual relations*, Journal of Convex Analysis **3** (1996), no. 2, 1–14.

[MS99] R. Mifflin and C. Sagastizábal, *$\mathcal{VU}$-decomposition derivatives for convex max-functions*, Ill-posed Variational Problems and Regularization Techniques (R. Tichatschke and M. Théra, eds.), Lecture Notes in Economics and Mathematical Systems, no. 477, Springer-Verlag Berlin Heidelberg, 1999, pp. 167–186.

[MS00a] R. Mifflin and C. Sagastizábal, *Functions with primal-dual gradient structure and $\mathcal{U}$-Hessians*, Nonlinear Optimization and Related Topics (G. Di Pillo and F. Giannessi, eds.), Applied Optimization, no. 36, Kluwer Academic Publishers B.V., 2000, pp. 219–233.

[MS00b] R. Mifflin and C. Sagastizábal, *On $\mathcal{VU}$-theory for functions with primal-dual gradient structure*, SIAM Journal on Optimization **11** (2000), no. 2, 547–571.

[Ous00] F. Oustry, *A second-order bundle method to minimize the maximum eigenvalue function*, Math. Program. **89** (2000), no. 1, Ser. A, 1–33.

[OY94] M.L. Overton and X.J. Ye, *Towards second-order methods for structured nonsmooth optimization*, Advances in Optimization and Numerical Analysis (S. Gomez and J.P. Hennart, eds.), Kluwer, 1994, pp. 97–109.

[PR94] R.A. Poliquin and R.T. Rockafellar, *Proto-derivative formulas for basic subgradient mappings in mathematical programming*, Set-Valued Anal. **2** (1994), no. 1-2, 275–290.

[PR95] R.A. Poliquin and R.T. Rockafellar, *Second-order nonsmooth analysis in nonlinear programming*, Recent advances in nonsmooth optimization, World Sci. Publishing, River Edge, NJ, 1995, pp. 322–349.

[Roc70] R.T. Rockafellar, *Convex analysis*, Princeton Mathematics Ser., no. 28, Princeton University Press, 1970.

[Roc88] R.T. Rockafellar, *First and second-order epi-differentiability in nonlinear programming*, Trans. Amer. Math. Soc. **307** (1988), 75–108.

[RW98] R.T. Rockafellar and R.J.-B. Wets, *Variational analysis*, Grund. der math. Wiss, no. 317, Springer-Verlag, 1998.