TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 352, Number 2, Pages 711–729
S 0002-9947(99)02243-6
Article electronically published on September 21, 1999

THE $\mathcal{U}$-LAGRANGIAN OF A CONVEX FUNCTION

CLAUDE LEMARÉCHAL, FRANÇOIS OUSTRY, AND CLAUDIA SAGASTIZÁBAL

Abstract. At a given point $\overline{p}$, a convex function f is differentiable in a certain subspace $\mathcal{U}$ (the subspace along which $\partial f(\overline{p})$ has 0-breadth). This property opens the way to defining a suitably restricted second derivative of f at $\overline{p}$. We do this via an intermediate function, convex on $\mathcal{U}$. We call this function the $\mathcal{U}$-Lagrangian; it coincides with the ordinary Lagrangian in composite cases: exact penalty, semidefinite programming. Also, we use this new theory to design a conceptual pattern for superlinearly convergent minimization algorithms. Finally, we establish a connection with the Moreau-Yosida regularization.

1. Introduction

This paper deals with higher-order expansions of a nonsmooth function, a problem addressed in [4], [5], [7], [9], [13], [25], and [31] among others.

The initial motivation for our present work lies in the following facts. When trying to generalize the classical second-order Taylor expansion of a function f at a nondifferentiability point $\overline{p}$, the major difficulty is by far the nonlinearity of the first-order approximation. Said otherwise, the gradient vector $\nabla f(\overline{p})$ is now a set $\partial f(\overline{p})$ and we have to consider difference quotients between sets, say

$$\frac{\partial f(\overline{p}+h) - \partial f(\overline{p})}{\|h\|}\,. \tag{1.1}$$

Giving a sensible meaning to the minus-sign in this expression is a difficult problem, to say the least; it has received only abstract answers so far; see [1], [3], [10], [12], [16], [18], [23], [24], [30]. However, here are two crucial observations (already mentioned in [22]):

– There is a subspace $\mathcal{U}$ (the "ridge") in which the first-order approximation $f'(\overline{p};\cdot)$ (the directional derivative) is linear.
– Defining a second-order expansion of f is unnecessary along directions not in $\mathcal{U}$. Consider for example the case where $f = \max_i f_i$ with smooth f_i's; then a minimization algorithm of the SQP-type will converge superlinearly, even if the second-order behaviour of f is identified in the ridge only ([26], [6]).

Here, starting from results presented in [14] and [15], we take advantage of these observations. After some preliminary theory in §2, we define our key-objects in §3: the $\mathcal{U}$-Lagrangian and its derivatives. In §4 we give some specific examples (further studied in [17], [20]): how the $\mathcal{U}$-Lagrangian specializes in an NLP and an SDP

Received by the editors July 18, 1996 and, in revised form, August 1, 1997.
1991 *Mathematics Subject Classification.* Primary 49J52, 58C20; Secondary 49Q12, 65K10.
Key words and phrases. Nonsmooth analysis, generalized derivative, second-order derivative, composite optimization.

framework, and how it could help designing superlinearly convergent algorithms for general convex functions. Finally, we show in §5 a connection between our objects thus defined and the Moreau-Yosida regularization. Indeed, the present paper clarifies and formalizes the theory sketched in §3.2 of [15]; for a related subject see also [29], [25].

Our notation follows closely that of [28] and [11]. The space $\mathbb{R}^n$ is equipped with a scalar product $\langle \cdot, \cdot \rangle$, and $\| \cdot \|$ is the associated norm; in a subspace $\mathcal{S}$, we will write $\langle \cdot, \cdot \rangle_{\mathcal{S}}$ and $\| \cdot \|_{\mathcal{S}}$ for the induced scalar product and norm. The open ball of $\mathbb{R}^n$ centered at x with radius r is $B(x, r)$; and once again, we use the notation $B_{\mathcal{S}}(x, r)$ in a subspace $\mathcal{S}$. We denote by $x_{\mathcal{S}}$ the projection of a vector $x \in \mathbb{R}^n$ onto the subspace $\mathcal{S}$. Throughout this paper, we consider the following situation:

$$f \text{ is a finite-valued convex function, } \overline{p} \text{ and } \overline{g} \in \partial f(\overline{p}) \text{ are fixed.} \tag{1.2}$$

We will also often assume that $\overline{g}$ lies in the relative interior of $\partial f(\overline{p})$.

2. The $\mathcal{VU}$ decomposition

We start by defining a decomposition of the space $\mathbb{R}^n = \mathcal{U} \oplus \mathcal{V}$, associated with a given $\overline{p} \in \mathbb{R}^n$. We give three equivalent definitions for the subspaces $\mathcal{U}$ and $\mathcal{V}$; each has its own merit to help the intuition.

Definition 2.1. (*i*) Define $\mathcal{U}_1$ as the subspace where $f'(\overline{p}; \cdot)$ is linear and take $\mathcal{V}_1 := \mathcal{U}_1^{\perp}$. Because $f'(\overline{p}; \cdot)$ is sublinear, we have

$$\mathcal{U}_1 := \{d \in \mathbb{R}^n : f'(\overline{p}; d) = -f'(\overline{p}; -d)\} \, ;$$

if necessary, see for instance Proposition V.1.1.6 in [11]. In other words, $\mathcal{U}_1$ is the subspace where $f(\overline{p} + \cdot)$ appears to be "differentiable" at 0. Note that this definition of $\mathcal{U}_1$ does not rely on a particular scalar product.

(*ii*) Define $\mathcal{V}_2$ as the subspace parallel to the affine hull of $\partial f(\overline{p})$ and take $\mathcal{U}_2 := \mathcal{V}_2^{\perp}$. In other words, $\mathcal{V}_2 := \operatorname{lin}(\partial f(\overline{p}) - \overline{g})$ for an arbitrary $\overline{g} \in \partial f(\overline{p})$, and $d \in \mathcal{U}_2$ means $\langle \overline{g} + v, d \rangle = \langle \overline{g}, d \rangle$ for all $v \in \mathcal{V}_2$.

(*iii*) Define $\mathcal{U}_3$ and $\mathcal{V}_3$ respectively as the normal and tangent cones to $\partial f(\overline{p})$ at an arbitrary g° in the relative interior of $\partial f(\overline{p})$. It is known (see, for example, Proposition 2.2 in [14]) that the property $g^\circ \in \operatorname{ri} \partial f(\overline{p})$ is equivalent to these cones being subspaces. □

To visualize these definitions, the reader may look at Figure 1 in §3.2 (where $\overline{g} = g^\circ \in \operatorname{ri} \partial f(\overline{p})$). We recall the definition of the relative interior: $g^\circ \in \operatorname{ri} \partial f(\overline{p})$ means

$$g^\circ + (B(0, \eta) \cap \mathcal{V}_2) \subset \partial f(\overline{p}) \quad \text{for some } \eta > 0. \tag{2.1}$$

We start with a preliminary result, showing in particular that Definition 2.1 does define the same pair $\mathcal{VU}$ three times.

Proposition 2.2. *In Definition 2.1,*

(i) *the subspace $\mathcal{U}_3$ is actually given by*

$$\{d \in \mathbb{R}^n : \langle g - g^\circ, d \rangle = 0 \text{ for all } g \in \partial f(\overline{p})\} = \mathrm{N}_{\partial f(\overline{p})}(g^\circ) \tag{2.2}$$

and is independent of the particular $g^\circ \in \operatorname{ri} \partial f(\overline{p})$;

(ii) $\mathcal{U}_1 = \mathcal{U}_2 = \mathcal{U}_3 =: \mathcal{U}$;

(iii) $\mathcal{U} \subset \mathrm{N}_{\partial f(\overline{p})}(\overline{g})$ *for all* $\overline{g} \in \partial f(\overline{p})$.

Proof. (i) To prove (2.2), take $g^\circ \in \operatorname{ri}\partial f(\overline{p})$ and set $N := \mathrm{N}_{\partial f(\overline{p})}(g^\circ)$. By definition of a normal cone, N contains the left-hand side in (2.2); we only need to establish the converse inclusion. Let $d \in N$ and $g \in \partial f(\overline{p})$; it suffices to prove $\langle g-g^\circ, d\rangle \geq 0$. Indeed, (assuming $g-g^\circ \neq 0$), $v := -\frac{g-g^\circ}{\|g-g^\circ\|} \in \mathcal{V}_2$, hence (2.1) and $d \in N$ imply that

$$0 \geq \langle g^\circ + \eta v - g^\circ, d\rangle = -\frac{\eta}{\|g-g^\circ\|}\langle g-g^\circ, d\rangle \quad \text{for some } \eta > 0$$

and we are done.

To see the independence on the particular g°, replace g° in (2.2) by some other $\gamma^\circ \in \operatorname{ri}\partial f(\overline{p})$:

$$\mathrm{N}_{\partial f(\overline{p})}(\gamma^\circ) = \left\{d \in \mathbb{R}^n : \langle g, d\rangle = \langle \gamma^\circ, d\rangle = \langle g^\circ, d\rangle, \text{ for all } g \in \partial f(\overline{p})\right\} = \mathcal{U}_3.$$

(ii) Write

$$\mathcal{U}_1 = \left\{d \in \mathbb{R}^n : \max_{g\in\partial f(\overline{p})}\langle g, d\rangle = \min_{g\in\partial f(\overline{p})}\langle g, d\rangle\right\} \tag{2.3}$$

to see from (i) that $\mathcal{U}_1 = \mathcal{U}_3$. Then we only need to prove $\mathcal{U}_1 \subset \mathcal{U}_2 \subset \mathcal{U}_3$.

Let $d \in \mathcal{U}_1$. For an arbitrary $v = \sum_j \lambda_j(g_j - \overline{g}) \in \mathcal{V}_2$ with $g_j \in \partial f(\overline{p})$, we have from (2.3)

$$\langle v, d\rangle = \sum_j \lambda_j(\langle g_j, d\rangle - \langle \overline{g}, d\rangle) = 0\,;$$

hence $d \in \mathcal{V}_2^\perp = \mathcal{U}_2$.

Let $d \in \mathcal{U}_2$. We have $\langle g, d\rangle = \langle \overline{g}, d\rangle$ for all $g \in \partial f(\overline{p})$. It follows that $\langle g, d\rangle = \langle g^\circ, d\rangle$ and this, together with (i), implies $d \in \mathcal{U}_3$.

(iii) Let $d \in \mathcal{U} = \mathcal{U}_3$. Given $\overline{g} \in \partial f(\overline{p})$, we have $\langle g^\circ, d\rangle = \langle g, d\rangle = \langle \overline{g}, d\rangle$ for all $g \in \partial f(\overline{p})$; hence $d \in \mathrm{N}_{\partial f(\overline{p})}(\overline{g})$. □

Using projections, every $x \in \mathbb{R}^n$ can be decomposed as $x = (x_{\mathcal{U}}, x_{\mathcal{V}})^T$. Throughout this paper we use the notation $x_{\mathcal{U}} \oplus x_{\mathcal{V}}$ for the vector with components $x_{\mathcal{U}}$ and $x_{\mathcal{V}}$. In other words, $\oplus$ stands for the linear mapping from $\mathcal{U} \times \mathcal{V}$ onto $\mathbb{R}^n$ defined by

$$\mathcal{U} \times \mathcal{V} \ni (u, v) \mapsto u \oplus v := \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^n. \tag{2.4}$$

With this convention, $\mathcal{U}$ and $\mathcal{V}$ are themselves considered as vector spaces. We equip them with the scalar product induced by $\mathbb{R}^n$, so that

$$\langle g, x\rangle = \langle g_{\mathcal{U}} \oplus g_{\mathcal{V}}, x_{\mathcal{U}} \oplus x_{\mathcal{V}}\rangle = \langle g_{\mathcal{U}}, x_{\mathcal{U}}\rangle_{\mathcal{U}} + \langle g_{\mathcal{V}}, x_{\mathcal{V}}\rangle_{\mathcal{V}},$$

with similar expressions for norms.

Remark 2.3. The projection $x \mapsto x_{\mathcal{U}}$, as well as the operation $(u, v) \mapsto \overline{p} + u \oplus v$, will appear recurrently in all our development. Consider the three convex functions h_1, h_2 and h defined by

$$\begin{array}{rcll} \mathcal{U} \ni u & \mapsto & h_1(u) := f(\overline{p} + u \oplus v), & \text{with } v \in \mathcal{V} \text{ arbitrary;} \\ \mathcal{V} \ni v & \mapsto & h_2(v) := f(\overline{p} + u \oplus v), & \text{with } u \in \mathcal{U} \text{ arbitrary;} \\ \mathcal{U} \times \mathcal{V} \ni (u, v) & \mapsto & h(u, v) := f(\overline{p} + u \oplus v). & \end{array}$$

Their subdifferentials have the expressions

$$\begin{array}{rcl} \partial h_1(u) & = & \{g_{\mathcal{U}} : g \in \partial f(\overline{p} + u \oplus v)\}, \\ \partial h_2(v) & = & \{g_{\mathcal{V}} : g \in \partial f(\overline{p} + u \oplus v)\}, \\ \partial h(x_{\mathcal{U}}, x_{\mathcal{V}}) & = & \{g_{\mathcal{U}} \oplus g_{\mathcal{V}} : g \in \partial f(\overline{p} + x)\}. \end{array}$$

Proving these formulae is a good exercise to become familiar with the operation $\oplus$ of (2.4) and with our $\mathcal{VU}$ notation. Just consider the adjoint of $\oplus$ and of the projections onto the various subspaces involved. □

In the $\mathcal{VU}$ language, (2.1) gives the following elementary result.

Proposition 2.4. *Suppose in* (1.2) *that* $\overline{g} \in \operatorname{ri} \partial f(\overline{p})$. *Then there exists* $\eta > 0$ *small enough such that*

$$\overline{g} + 0 \oplus \frac{\eta v}{\|v\|_{\mathcal{V}}} \in \partial f(\overline{p})$$

for any $0 \neq v \in \mathcal{V}$. *In particular,*

$$f(\overline{p} + u \oplus v) \geq f(\overline{p}) + \langle \overline{g}_{\mathcal{U}}, u \rangle_{\mathcal{U}} + \langle \overline{g}_{\mathcal{V}}, v \rangle_{\mathcal{V}} + \eta \|v\|_{\mathcal{V}}, \tag{2.5}$$

for any $(u, v) \in \mathcal{U} \times \mathcal{V}$.

Proof. Just translate (2.1): with v as stated, $u \oplus v \overline{g}_{\mathcal{U}}\big(\overline{g}_{\mathcal{V}} + \frac{\eta v}{\|v\|_{\mathcal{V}}}\big) \in \partial f(\overline{p})$ and the rest follows easily. □

3. The $\mathcal{U}$-Lagrangian

In this section we formalize the theory outlined in §3.2 of [15]. Along with the $\mathcal{VU}$ decomposition, we introduced there the "tangential" regularization $\phi_{\mathcal{V}}$. Here, we find it convenient to consider $\phi_{\mathcal{V}}$ as a function defined on $\mathcal{U}$ only; in addition, we drop the quadratic term appearing in (13) of [15]. As will be seen in §4, these modifications result in some sort of Lagrangian, which we denote by $L_{\mathcal{U}}$ instead of $\phi_{\mathcal{V}}$.

3.1. **Definition and basic properties.** Following the above introduction, we define the function $L_{\mathcal{U}}$ as follows:

$$\mathcal{U} \ni u \mapsto L_{\mathcal{U}}(u) := \inf_{v \in \mathcal{V}} \{ f(\overline{p} + u \oplus v) - \langle \overline{g}_{\mathcal{V}}, v \rangle_{\mathcal{V}} \}. \tag{3.1}$$

Associated with (3.1) we have the set of minimizers

$$W(u) := \operatorname*{Argmin}_{v \in \mathcal{V}} \{ f(\overline{p} + u \oplus v) - \langle \overline{g}_{\mathcal{V}}, v \rangle_{\mathcal{V}} \}. \tag{3.2}$$

It will be seen below that an important question is whether $W(u)$ is nonempty.

Remark 3.1. The function $L_{\mathcal{U}}$ of (3.1) will be called the *$\mathcal{U}$-Lagrangian.* Note that it depends on the particular $\overline{g}$, a notation $L_{\mathcal{U}}(u, \overline{g})$ is also possible. In fact, since $\overline{g}$ lies in the dual of $\mathbb{R}^n$, it connotes a dual variable; this will become even more visible in §4.1 (just observe here that $\overline{g} \mapsto -L_{\mathcal{U}}$ is a conjugate function).

At this point, the idea behind (3.1) can be roughly explained. As is commonly known, smoothness of a convex function is related to strong convexity of its conjugate. In our context, a useful property is the "radial" strong convexity of f^* at $\overline{g}$, say,

$$f^*(\overline{g} + s) \geq f^*(\overline{g}) + \langle s, \overline{p} \rangle + \frac{1}{2} c \|s\|^2 + o(\|s\|^2)$$

for some $c > 0$. However, the above inequality is hopeless for an s of the form $s = 0 \oplus v$ (see §4 in [14]; see also [2] for related developments). To obtain radial strong convexity on $\mathcal{V}$, we introduce the function

$$f^*(\overline{g} + s) + \frac{1}{2} c \|s_{\mathcal{V}}\|^2_{\mathcal{V}}. \tag{3.3}$$

Its conjugate (restricted to $\mathcal{U}$) is precisely $L_{\mathcal{U}}$ when $c = +\infty$ (a value which yields the "strongest" possible convexity); Theorem 3.3 will confirm the smoothness of $L_{\mathcal{U}}$.

The value $c = 1$ in (3.3) may be deemed more natural – and indeed, it will be useful in §5; in fact, Lemma 5.1 will show that the choice of c has minor importance for second order. □

Theorem 3.2. *Assume* (1.2).

(i) *The function* $L_{\mathcal{U}}$ *defined in* (3.1) *is convex and finite everywhere.*
(ii) *A minimum point* $w \in W(u)$ *in* (3.2) *is characterized by the existence of some* $g \in \partial f(\overline{p} + u \oplus w)$ *such that* $g_{\mathcal{V}} = \overline{g}_{\mathcal{V}}$.
(iii) *In particular,* $0 \in W(0)$ *and* $L_{\mathcal{U}}(0) = f(\overline{p})$.
(iv) *If* $\overline{g} \in \operatorname{ri} \partial f(\overline{p})$, *then* $W(u)$ *is nonempty for each* $u \in \mathcal{U}$ *and* $W(0) = \{0\}$.

Proof. (*i*) The infimand in (3.1) is $h(u, v) - \langle \overline{g}_{\mathcal{V}}, v\rangle_{\mathcal{V}}$, where the function h was defined in Remark 2.3. It is clearly finite-valued and convex on $\mathcal{U} \times \mathcal{V}$, and the subgradient inequality at $(u, v) = (0, 0)$ gives

$$h(u, v) - \langle \overline{g}_{\mathcal{V}}, v\rangle_{\mathcal{V}} \geq f(\overline{p}) + \langle \overline{g}_{\mathcal{U}}, u\rangle_{\mathcal{U}} \quad \text{for any } v \in \mathcal{V}.$$

It follows that $L_{\mathcal{U}}$ is nowhere $-\infty$ and, being a partial infimum of a jointly convex function, it is convex as well, see for example §IV.2.4 in [11].

(*ii*) The optimality condition for $w \in W(u)$ is $0 \in \partial h_2(w) - \overline{g}_{\mathcal{V}}$, with h_2 as in Remark 2.3. Knowing the expression of ∂h_2, we obtain $0 = g_{\mathcal{V}} - \overline{g}_{\mathcal{V}}$, for some $g \in \partial f(\overline{p} + u \oplus w)$.

(*iii*) In particular, for $u = 0$, we can take $w = 0$ and $g = \overline{g} \in \partial f(\overline{p} + 0 \oplus 0)$ in (*ii*). This proves that $v = 0$ satisfies the optimality condition for (3.1); then $L_{\mathcal{U}}(0) = f(\overline{p})$.

(*iv*) Apply (2.5): there exists $\eta > 0$ such that, for any $v \neq 0$,

$$h(u, v) - \langle \overline{g}_{\mathcal{V}}, v\rangle_{\mathcal{V}} \geq f(\overline{p}) + \langle \overline{g}_{\mathcal{U}}, u\rangle_{\mathcal{U}} + \eta \|v\|_{\mathcal{V}}.$$

Thus, the infimand in (3.1) is inf-compact on $\mathcal{V}$ and the set $W(u)$ is nonempty. At $u = 0$, we have

$$h(0, v) - \langle \overline{g}_{\mathcal{V}}, v\rangle_{\mathcal{V}} \geq f(\overline{p}) + \eta \|v\|_{\mathcal{V}},$$

which shows that $v = 0$ is the unique minimizer. □

3.2. **First-order behaviour.** The primary interest of the $\mathcal{U}$-Lagrangian is that it has a gradient at 0. Besides, its subdifferential is obtained from the optimality condition in Theorem 3.2(*ii*).

Theorem 3.3. *Assume* (1.2).

(i) *Let* u *be such that* $W(u) \neq \emptyset$. *Then the subdifferential of* $L_{\mathcal{U}}$ *at this* u *has the expression*

$$\partial L_{\mathcal{U}}(u) = \{g_{\mathcal{U}} : g_{\mathcal{U}} \oplus \overline{g}_{\mathcal{V}} \in \partial f(\overline{p} + u \oplus w)\}, \tag{3.4}$$

where w *is an arbitrary point in* $W(u)$.
(ii) *In particular,* $L_{\mathcal{U}}$ *is differentiable at* 0, *with* $\nabla L_{\mathcal{U}}(0) = \overline{g}_{\mathcal{U}}$.

Proof. (*i*) Using again the notation of Remark 2.3, write the infimand in (3.1) as $h(u, v) - \langle 0 \oplus \overline{g}_{\mathcal{V}}, u \oplus v\rangle$. For the subdifferential of the marginal function $L_{\mathcal{U}}$,

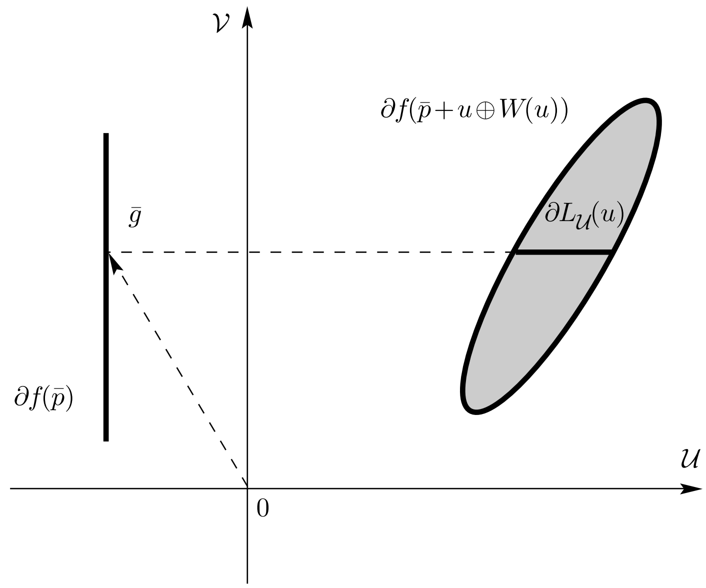

FIGURE 1. Subdifferential of $L_{\mathcal{U}}$

Corollary VI.4.5.3 in [11] gives the calculus rule

$$\begin{aligned} s \in \partial_u L_{\mathcal{U}}(u) \quad &\iff \quad s \oplus 0 \in \partial_{u,v}(h - \langle 0 \oplus \overline{g}_{\mathcal{V}}, \cdot\rangle)(u, w) \\ &\iff \quad s \oplus 0 \in \partial_{u,v} h(u, w) - 0\overline{g}_{\mathcal{V}} \\ &\iff \quad s \oplus \overline{g}_{\mathcal{V}} \in \partial_{u,v} h(u, w), \end{aligned}$$

where $w \in W(u)$ is arbitrary. From the expression of $\partial_{u,v} h = \partial h$ in Remark 2.3, this is (3.4).

(*ii*) Because of Theorem 3.2(*iii*), (3.4) holds at $u = 0$ and becomes $\partial L_{\mathcal{U}}(0) = \{g_{\mathcal{U}} : g_{\mathcal{U}} \oplus \overline{g}_{\mathcal{V}} \in \partial f(\overline{p})\}$. This latter set clearly contains $\overline{g}_{\mathcal{U}}$. Actually, it does not contain any other point, due to Definition 2.1(*ii*): $\partial f(\overline{p}) \subset \overline{g} + \mathcal{V}$, i.e., all subgradients at $\overline{p}$ have the same $\mathcal{U}$-component, namely $\overline{g}_{\mathcal{U}}$. ☐

This result is illustrated in Figure 1. We stress the fact that the set in the right-hand-side of (3.4) does not depend on the particular $w \in W(u)$. In other words, (3.4) expresses the following: to obtain the subgradients of $L_{\mathcal{U}}$ at u, take those subgradients g of f at $\overline{p} + u \oplus W(u)$ that have the same $\mathcal{V}$-component as $\overline{g}$ (namely $\overline{g}_{\mathcal{V}}$); then take their $\mathcal{U}$-component. Remembering that $\mathcal{U}$ is in effect a subset of $\mathbb{R}^n$, we can also write more informally

$$\partial L_{\mathcal{U}}(u) = [\partial f(\overline{p} + u \oplus W(u)) \cap (\overline{g} + \mathcal{U})]_{\mathcal{U}}.$$

This operation somewhat simplifies when $\overline{g}_{\mathcal{V}} = 0$:

$$\text{(3.5)} \qquad \text{if } \overline{g}_{\mathcal{V}} = 0, \text{ then } \partial L_{\mathcal{U}}(u) = \partial f(\overline{p} + u \oplus W(u)) \cap \mathcal{U}.$$

See the end of §3.2 below for additional comments on the "trajectories" $\overline{p}+u\oplus W(u)$. Another observation is that, for all $u \in \mathcal{U}$,

$$f'(\overline{p}; u \oplus 0) = \langle \overline{g}, u \oplus 0\rangle = \langle \overline{g}_{\mathcal{U}}, u\rangle_{\mathcal{U}} = \langle \nabla L_{\mathcal{U}}(0), u\rangle_{\mathcal{U}}.$$

In other words, $L_{\mathcal{U}}$ agrees, up to first order, with the restriction of f to $\overline{p}+\mathcal{U}$. Continuing with our $\mathcal{U}$-terminology, we will say that $\overline{g}_{\mathcal{U}}$ is the *$\mathcal{U}$-gradient* of f at $\overline{p}$, and note that $\overline{g}_{\mathcal{U}}$ is actually independent of the particular $\overline{g} \in \partial f(\overline{p})$ (recall Proposition 2.2(i)).

Remark 3.4. We add that, because f is locally Lipschitzian, this $\mathcal{U}$-differentiability property holds also tangentially to $\mathcal{U}$:

$$f(\overline{p}+h) = f(\overline{p}) + \langle \overline{g}, h\rangle + o(\|h\|) \quad \text{whenever} \quad \|h_{\mathcal{V}}\|_{\mathcal{V}} = o(\|h_{\mathcal{U}}\|_{\mathcal{U}}). \tag{3.6}$$

This remark will be instrumental when coming to higher order; then we will have to *select* h appropriately, to allow a specification of the remainder term in (3.6); see Theorem 3.9. □

As already mentioned, the existence of $\nabla L_{\mathcal{U}}(0)$ is of paramount importance, since it suppresses the difficulty pointed out in the introduction of this paper; now the difference quotient in (1.1) takes the form

$$\frac{\partial L_{\mathcal{U}}(u) - \overline{g}_{\mathcal{U}}}{\|u\|_{\mathcal{U}}},$$

which does make sense. Here is a useful first consequence: $W(u) = o(\|u\|_{\mathcal{U}})$.

Corollary 3.5. *Assume* (1.2). *If* $\overline{g} \in \operatorname{ri} \partial f(\overline{p})$, *then*

$$\forall \varepsilon > 0 \; \exists \delta > 0 : \; \|u\|_{\mathcal{U}} \le \delta \Rightarrow \|w\|_{\mathcal{V}} \le \varepsilon \|u\|_{\mathcal{U}} \textit{ for any } w \in W(u).$$

Proof. Use Theorem 3.3(ii) to write the first-order expansion of $L_{\mathcal{U}}$:

$$L_{\mathcal{U}}(u) = L_{\mathcal{U}}(0) + \langle \nabla L_{\mathcal{U}}(0), u\rangle_{\mathcal{U}} + o(\|u\|_{\mathcal{U}}) = f(\overline{p}) + \langle \overline{g}_{\mathcal{U}}, u\rangle_{\mathcal{U}} + o(\|u\|_{\mathcal{U}}).$$

For any $w \in W(u)$ we have $L_{\mathcal{U}}(u) = f(\overline{p}+u\oplus w) - \langle \overline{g}_{\mathcal{V}}, w\rangle_{\mathcal{V}}$; therefore, (2.5) written for $v = w$, gives $L_{\mathcal{U}}(u) \ge f(\overline{p}) + \langle \overline{g}_{\mathcal{U}}, u\rangle_{\mathcal{U}} + \eta\|w\|_{\mathcal{V}}$. Altogether, we obtain

$$o(\|u\|_{\mathcal{U}}) = L_{\mathcal{U}}(u) - f(\overline{p}) - \langle \overline{g}_{\mathcal{U}}, u\rangle_{\mathcal{U}} \ge \eta\|w\|_{\mathcal{V}}. \quad \square$$

Let us sum up our results so far.

- Given $\overline{g} \in \partial f(\overline{p})$, we define via (3.1) a convex function $L_{\mathcal{U}}$ (Theorem 3.2(i)), which is differentiable at 0 and coincides up to first order with the restriction of f to $\overline{p}+\mathcal{U}$ (Theorem 3.3(ii)).
- When $W(\cdot) \ne \emptyset$, this $\mathcal{U}$-Lagrangian is indeed the restriction of f to a "thick surface" $\{\overline{p} + \cdot \oplus W(\cdot)\}$, parametrized by $u \in \mathcal{U}$.
- We also define, via Theorem 3.2(ii), a "thick selection" of ∂f on this thick surface, made up of those subgradients that have the same $\mathcal{V}$-component as $\overline{g}$.
- As a function of the parameter u, this thick selection behaves like a subdifferential, namely $\partial L_{\mathcal{U}}$ (Theorem 3.3(i)).
- When $\overline{g} \in \operatorname{ri} \partial f(\overline{p})$, our thick surface has $\mathcal{U}$ as "tangent space" at $\overline{p}$ (Corollary 3.5; we use quotation marks because W is multivalued).

Remark 3.6. We note in passing two extreme cases in which our theory becomes trivial:
– when f is differentiable at $\overline{p}$, then $\mathcal{U} = \mathbb{R}^n$, $\mathcal{V} = \{0\}$ and $L_{\mathcal{U}} \equiv f$;
– when $\partial f(\overline{p})$ has full dimension, then $\mathcal{U} = \{0\}$ and there is no $\mathcal{U}$-Lagrangian. □

3.3. **Higher-order behaviour.** Proceeding further in our differential analysis of $L_\mathcal{U}$, we now study the behaviour of $\partial L_\mathcal{U}$ near 0. A very basic property of this set is its radial Lipschitz continuity. We say that f has a radially Lipschitz subdifferential at $\overline{p}$ when there is a $D > 0$ and a $\delta > 0$ such that

$$\partial f(\overline{p} + d) \subset \partial f(\overline{p}) + B(0, D\|d\|), \quad \text{for all } d \in B(0, \delta). \tag{3.7}$$

This is equivalent to an upper quadratic growth condition on the function itself (recall Corollary 3.5 in [14]): there is a $C > 0$ and an $\varepsilon > 0$ such that

$$f(\overline{p} + d) \leq f(\overline{p}) + f'(\overline{p}; d) + \frac{1}{2}C\|d\|^2, \quad \text{for all } d \in B(0, \varepsilon). \tag{3.8}$$

This property is transmitted from f to $L_\mathcal{U}$:

Proposition 3.7. *Assume* (1.2). *Assume also that* $W(u)$ *is nonempty for* u *small enough, and that* (3.7) $\equiv$ (3.8) *is satisfied. Then*

(i) $\partial L_\mathcal{U}(u) \subset \overline{g}_\mathcal{U} + B_\mathcal{U}(0, 2C\|u\|_\mathcal{U})$, *for some* $\delta > 0$ *and all* $u \in B_\mathcal{U}(0, \delta)$*;*
(ii) $L_\mathcal{U}(u) \leq L_\mathcal{U}(0) + \langle \overline{g}_\mathcal{U}, u\rangle_\mathcal{U} + \frac{1}{2}R\|u\|^2_\mathcal{U}$, *for some* $\rho > 0$, $R > 0$ *and all* $u \in B_\mathcal{U}(0, \rho)$.

Proof. Remember that $\nabla L_\mathcal{U}(0) = \overline{g}_\mathcal{U}$. Because the subdifferential is an outer-semicontinuous mapping, we can choose $\delta > 0$ such that for all $u \in B_\mathcal{U}(0, \delta)$ and $g_\mathcal{U} \in \partial L_\mathcal{U}(u)$, $\|g_\mathcal{U} - \overline{g}_\mathcal{U}\|_\mathcal{U} \leq \frac{\varepsilon C}{2}$ (see § VI.6.2 of [11] for example). On the other hand, assume δ so small that $W(u)$ contains some w; from Theorem 3.2(*ii*), $g_\mathcal{U} \oplus \overline{g}_\mathcal{V} \in \partial f(\overline{p} + u \oplus w)$.

Now $\mathcal{U} \subset \mathrm{N}_{\partial f(\overline{p})}(\overline{g})$ (Proposition 2.2(*iii*)). Using the notation $s := (g_\mathcal{U} - \overline{g}_\mathcal{U}) \oplus 0$, so that $g_\mathcal{U} \oplus \overline{g}_\mathcal{V} = \overline{g} + s \in \partial f(\overline{p} + u \oplus w)$, we are in the conditions of Corollary 3.3 in [14] written with $\varphi = f$, $z_0 = \overline{p}$, $g_0 = \overline{g}$, $x = \overline{p} + u \oplus w$. Inequality (14) therein becomes

$$\|g_\mathcal{U} - \overline{g}_\mathcal{U}\|^2_\mathcal{U} = \|s\|^2 \leq 2C\langle s, u \oplus w\rangle = 2C\langle g_\mathcal{U} - \overline{g}_\mathcal{U}, u\rangle_\mathcal{U} \leq 2C\|g_\mathcal{U} - \overline{g}_\mathcal{U}\|_\mathcal{U}\|u\|_\mathcal{U},$$

which is (i). As for (ii), it is equivalent to (i) (Corollary 3.5 in [14]). □

Back to the f-context, Proposition 3.7 says: for small $u \in \mathcal{U}$ and all $w \in W(u)$, there holds

$$\{g_\mathcal{U} : g_\mathcal{U} \oplus \overline{g}_\mathcal{V} \in \partial f(\overline{p} + u \oplus w)\} \subset \overline{g}_\mathcal{U} + B_\mathcal{U}(0, 2C\|u\|_\mathcal{U})$$

as well as

$$f(\overline{p} + u \oplus w) \leq f(\overline{p}) + \langle \overline{g}, u \oplus w\rangle + \frac{1}{2}R\|u\|^2_\mathcal{U}.$$

Now, we have a function $L_\mathcal{U}$, which is differentiable at 0, and whose second-order difference quotients inherit the qualitative properties of those of f. The stage is therefore set to consider the case where $L_\mathcal{U}$ has a generalized Hessian at 0, in the sense of [9] (see also [15], §3). Generally speaking, we say that a convex function φ has at z_0 a generalized Hessian $\mathrm{H}\varphi(z_0)$ when

(i) the gradient $\nabla\varphi(z_0)$ exists;
(ii) there exists a symmetric positive semidefinite operator $\mathrm{H}\varphi(z_0)$ such that

$$\varphi(z_0 + d) = \varphi(z_0) + \langle \nabla\varphi(z_0), d\rangle + \frac{1}{2}\langle \mathrm{H}\varphi(z_0)d, d\rangle + o(\|d\|^2);$$

(iii) or equivalently,

$$\partial\varphi(z_0 + d) \subset \nabla\varphi(z_0) + \mathrm{H}\varphi(z_0)d + B(0, o(\|d\|)). \tag{3.9}$$

Definition 3.8. Assume (1.2). We say that f has at $\overline{p}$ a *$\mathcal{U}$-Hessian* $\mathrm{H}_{\mathcal{U}}f(\overline{p})$ (associated with $\overline{g}$) if $L_{\mathcal{U}}$ has a generalized Hessian at 0; then we set

$$\mathrm{H}_{\mathcal{U}}f(\overline{p}) := \mathrm{H}L_{\mathcal{U}}(0). \quad \square$$

When it exists, the $\mathcal{U}$-Hessian $\mathrm{H}_{\mathcal{U}}f(\overline{p})$ is therefore a symmetric positive semi-definite operator from $\mathcal{U}$ to $\mathcal{U}$. Its existence means the possibility of expanding f along the thick surface $\overline{p} + \cdot \oplus W(\cdot)$ introduced at the end of §3.2.

Theorem 3.9. *Take $\overline{g} \in \operatorname{ri}\partial f(\overline{p})$ and let the $\mathcal{U}$-Hessian $\mathrm{H}_{\mathcal{U}}f(\overline{p})$ exist. For $u \in \mathcal{U}$ and $h \in u \oplus W(u)$, there holds*

$$f(\overline{p}+h) = f(\overline{p}) + \langle \overline{g}, h\rangle + \frac{1}{2}\langle \mathrm{H}_{\mathcal{U}}f(\overline{p})u, u\rangle_{\mathcal{U}} + o(\|h\|^2). \tag{3.10}$$

Proof. We know from Theorem 3.2(iv) that $W(u) \neq \emptyset$. Then apply the definition of $L_{\mathcal{U}}$ and expand $L_{\mathcal{U}}$ to obtain for all u and $w \in W(u)$:

$$\begin{aligned} L_{\mathcal{U}}(u) &= f(\overline{p} + u \oplus w) - \langle \overline{g}_{\mathcal{V}}, w\rangle_{\mathcal{V}} \\ &= L_{\mathcal{U}}(0) + \langle \nabla L_{\mathcal{U}}(0), u\rangle_{\mathcal{U}} + \tfrac{1}{2}\langle \mathrm{H}_{\mathcal{U}}f(\overline{p})u, u\rangle_{\mathcal{U}} + o(\|u\|^2_{\mathcal{U}}) \\ &= f(\overline{p}) + \langle \overline{g}_{\mathcal{U}}, u\rangle_{\mathcal{U}} + \tfrac{1}{2}\langle \mathrm{H}_{\mathcal{U}}f(\overline{p})u, u\rangle_{\mathcal{U}} + o(\|u\|^2_{\mathcal{U}}). \end{aligned}$$

In view of Corollary 3.5, $o(\|u\|^2_{\mathcal{U}}) = o(\|h\|^2)$; (3.10) follows, adding $\langle \overline{g}_{\mathcal{V}}, w\rangle_{\mathcal{V}}$ to both sides. $\square$

To the second-order expansion (3.10), there corresponds a first-order expansion of *selected* subgradients along the thick surface $\overline{p} + \cdot \oplus W(\cdot)$: with the notation and assumptions of Theorem 3.9,

$$\{g_{\mathcal{U}} : g_{\mathcal{U}} \oplus \overline{g}_{\mathcal{V}} \in \partial f(\overline{p}+h)\} \subset \overline{g}_{\mathcal{U}} + \mathrm{H}_{\mathcal{U}}f(\overline{p})u + B_{\mathcal{U}}(0, o(\|h\|)).$$

With reference to Remark 3.4, the expansion (3.10) makes (3.6) more explicit, for increments $h = h_{\mathcal{U}} \oplus h_{\mathcal{V}}$ such that $h_{\mathcal{V}} \in W(h_{\mathcal{U}})$. The aim of the next section is to disclose some intrinsic interest of these particular h's.

4. Examples of application

This section shows how the $\mathcal{U}$-concepts developed in §3 generalize well-known objects. We will first consider special situations: max-functions (§4.1) and semi-definite programming (§4.2). Then in §4.3 we outline a conceptual minimization algorithm.

4.1. **Exact penalty.** Consider an ordinary nonlinear programming problem

$$\begin{cases} \min \psi(p), \\ f_i(p) \le 0, \quad i = 1, \ldots, m, \end{cases} \tag{4.1}$$

with convex C^2 data ψ and f_i. Take an optimal $\overline{p}$ and suppose that the KKT conditions hold: with $L(p,\lambda) := \psi(p) + \sum_i \lambda_i f_i(p)$, defined for $(p,\lambda) \in \mathbb{R}^n \times \mathbb{R}^m$, there exist Lagrange multipliers $\overline{\lambda}_i$ such that

$$\begin{cases} [\nabla_p L(\overline{p}, \overline{\lambda}) =] \;\; \nabla\psi(\overline{p}) + \displaystyle\sum_{i=1}^{m} \overline{\lambda}_i \nabla f_i(\overline{p}) = 0, \\ \overline{\lambda}_i \ge 0 \quad \overline{\lambda}_i f_i(\overline{p}) = 0, \qquad \text{for } i = 1, \ldots, m. \end{cases} \tag{4.2}$$

We will use the notation $\gamma := \nabla\psi$, $g_i := \nabla f_i$, $\overline{\gamma} := \nabla\psi(\overline{p})$, $\overline{g}_i := \nabla f_i(\overline{p})$.

Consider now an exact penalty function associated with (4.1): with $f_0(p) \equiv 0$ (and $g_0(p) := \nabla f_0(p) \equiv 0$), set

$$f(p) := \psi(p) + \pi \max\{f_0(p), \dots, f_m(p)\}, \tag{4.3}$$

where $\pi > 0$ is a penalty parameter. Call

$$J(p) := \big\{j \in \{0, \dots, m\} : \psi(p) + \pi f_j(p) = f(p)\big\}$$

the set of indices realizing the max at p. Standard subdifferential calculus gives

$$\partial f(p) = \gamma(p) + \pi \mathrm{conv}\{g_j(p) : j \in J(p)\}.$$

In NLP language, instead of maximal functions, one speaks of active constraints. We therefore set

$$\overline{I} := \big\{i \in \{1, \dots, m\} : f_i(\overline{p}) = 0\big\}$$

(naturally, we assume $\overline{I} \neq \emptyset$; otherwise, the problem lacks interest). It is easy to see that $J(\overline{p}) = \overline{I} \cup \{0\}$; correspondingly, we associate with $J(\overline{p})$ the "multipliers"

$$\overline{\mu}_i := \overline{\lambda}_i \text{ for } i \in \overline{I} \quad \text{and} \quad \overline{\mu}_0 := \pi - \sum_{i \in \overline{I}} \overline{\lambda}_i. \tag{4.4}$$

For π large enough, it is well known that $\overline{p}$ solving (4.1) also minimizes f of (4.3). We proceed to apply the theory of §3 to the present situation: f is the exact penalty function of (4.3), $\overline{p}$ is optimal and $\overline{g} = 0$. We will show that the $\mathcal{U}$-Lagrangian $L_{\mathcal{U}}$ coincides up to second order with the restriction to $\mathcal{U}$ of the ordinary Lagrangian $L(\overline{p} + \cdot, \overline{\lambda})$. All along this subsection, we make the following assumptions:

- the active gradients $\{\overline{g}_i\}_{i \in \overline{I}}$ are linearly independent (hence $\overline{\lambda}$ is unique in the KKT conditions (4.2)),
- $\overline{\lambda}_i > 0$ for $i \in \overline{I}$ (strict complementarity),
- and $\pi > \sum_{i \in \overline{I}} \overline{\lambda}_i$, i.e., $\overline{\mu}_0 > 0$ in (4.4).

The following development should be considered as a mere illustration of the $\mathcal{U}$-theory. This is why we content ourselves with the above simplifying assumptions, which are relaxed in the more complete work of [17].

We start with a basic result, stating in particular that $\mathcal{U}$ is the space tangent to the surface defined by the active constraints (well-defined thanks to our simplifying assumptions).

Proposition 4.1. *With the above notation and assumptions, we have the following relations for $p = \overline{p}$:*

(i) $\partial f(\overline{p}) = \overline{\gamma} + \big\{\sum_{i \in \overline{I}} \mu_i \overline{g}_i : \mu_i \geq 0, \sum_{i \in \overline{I}} \mu_i \leq \pi\big\}$;
(ii) the subspaces $\mathcal{U}$ and $\mathcal{V}$ of Definition 2.1 are

$$\mathcal{V} = \mathrm{lin}\{\overline{g}_i\}_{i \in \overline{I}}, \quad \mathcal{U} = \{d \in \mathbb{R}^n : \langle \overline{g}_i, d\rangle = 0, i \in \overline{I}\};$$

(iii) $\overline{g} := 0 \in \mathrm{ri}\, \partial f(\overline{p})$.

Proof. (*i*) We have

$$\begin{aligned} \partial f(\overline{p}) &= \overline{\gamma} + \pi \mathrm{conv}\{\overline{g}_i : i \in \overline{I} \cup \{0\}\} \\ &= \overline{\gamma} + \big\{\pi\alpha_0 0 + \sum_{i \in \overline{I}} \pi \alpha_i \overline{g}_i : \alpha_i \geq 0, \alpha_0 + \sum_{i \in \overline{I}} \alpha_i = 1\big\}. \end{aligned}$$

The formula is then straightforward, setting $\mu_i := \pi\alpha_i$ and eliminating the unnecessary vector 0.

(ii) Apply Definition 2.1(ii): $\mathcal{V} = \text{lin}\{\partial f(\overline{p}) - \overline{\gamma}\}$ because $\overline{\gamma} \in \partial f(\overline{p})$. Together with (i), the results clearly follow.

(iii) Consider the set $\mathcal{B} := \{\sum_{\overline{I}} \mu_i \overline{g}_i : \mu_i \geq -\overline{\mu}_i, \sum_{\overline{I}} \mu_i \leq \overline{\mu}_0\}$, where $\overline{\mu}$ was defined in (4.4). Because of (ii), $\mathcal{B} \subset \mathcal{V}$. Because of strict complementarity and $\overline{\mu}_0 > 0$, $\mathcal{B}$ is a relative neighborhood of $0 = \overline{g} \in \mathcal{V}$. Finally, because of (4.2) and (4.4),

$$\begin{aligned} \mathcal{B} &= \overline{\gamma} + \mathcal{B} + \sum_{\overline{I}} \overline{\lambda}_i \overline{g}_i \\ &= \overline{\gamma} + \{\sum_{\overline{I}} (\mu_i + \overline{\mu}_i)\overline{g}_i : \mu_i + \overline{\mu}_i \geq 0, \sum_{\overline{I}} (\mu_i + \overline{\mu}_i) \leq \pi\}. \end{aligned}$$

In view of (i), $\mathcal{B} \subset \partial f(\overline{p})$ and we are done. ☐

Lemma 4.2. *With the notation and assumptions of this subsection, let p be close to $\overline{p}$. Then $J(p) \subset J(\overline{p}) = \overline{I} \cup \{0\}$ and the system in $\{\mu_j\}_{J(p)}$*

$$(4.5) \qquad \begin{cases} \langle \overline{g}_i, \gamma(p)\rangle + \displaystyle\sum_{j\in J(p)} \mu_j \langle \overline{g}_i, g_j(p)\rangle = 0 \quad \text{for all } i \in \overline{I}, \\ \displaystyle\sum_{j\in J(p)} \mu_j = \pi \end{cases}$$

has a solution, which is unique, if and only if $J(p) = J(\overline{p}) = \overline{I} \cup \{0\}$. The solution $\mu(p)$ satisfies $\mu_j(p) > 0$ for all $j \in J(p) = J(\overline{p})$. Moreover, $\mu(\overline{p}) = \overline{\mu}$ of (4.4) *and $p \mapsto \mu(p)$ is differentiable at $p = \overline{p}$.*

Proof. Let $j \notin J(\overline{p})$. By continuity, $f_j(p) < f_i(p)$ for all $i \in J(\overline{p})$, hence $J(p) \subset J(\overline{p})$.

Now consider (4.5). First, observe that, because of (4.2), $\overline{\mu}$ of (4.4) is a solution at $p = \overline{p}$.

(a) Assume first that $J(p) = J(\overline{p}) = \overline{I} \cup \{0\}$. Since $g_0(p) \equiv 0$, the variable μ_0 is again directly given by $\mu_0(p) = \pi - \sum_{\overline{I}} \mu_j(p)$. As for the μ_j's, $j \in \overline{I}$, they are given by an $\overline{I} \times \overline{I}$ linear system, whose matrix is $(\langle \overline{g}_i, g_j(p)\rangle)_{ij}$. Because the $\overline{g}_i$'s are linearly independent, this matrix is positive definite. The solution $\mu(p)$ is unique; it is also close to $\overline{\mu}$, is therefore positive and sums up to less than π: $\mu_0(p) > 0$. In particular, $\mu(\overline{p}) = \overline{\mu}$ is the unique solution at $p = \overline{p}$. The differentiability property then comes from the Implicit Function Theorem.

(b) On the other hand, assume the set $I_0 := J(\overline{p}) \backslash J(p)$ is nonempty and suppose (4.5) has a solution $\{\mu_j^*\}_{j\in J(p)}$. Set $\mu_j^* := 0$ for $j \in I_0$; then μ^* also solves (4.5) with $J(p)$ replaced by $J(\overline{p})$. This contradicts part *(a)* of the proof. ☐

The next result reveals a nice interpretation of $W(\cdot)$ in (3.2): it makes a local description of the surface defined by the active constraints.

Theorem 4.3. *Use the notation and assumptions of this subsection. For $u \in \mathcal{U}$ small enough, $W(u)$ defined in* (3.2) *is a singleton $w(u)$, which is the unique solution of the system with unknown $v \in \mathcal{V}$*

$$(4.6) \qquad f_i(\overline{p} + u \oplus v) = 0, \quad \text{for all } i \in \overline{I}.$$

Proof. According to Theorem 3.2(ii) and (3.5), an arbitrary $p \in \overline{p} + u \oplus W(u)$ is characterized by $\partial f(p) \cap \mathcal{U} \neq \emptyset$; there are convex multipliers $\{\alpha_j\}_{j\in J(p)}$ such that $\gamma(p) + \pi \sum_{J(p)} \alpha_j g_j(p) \in \mathcal{U}$. Setting $\mu_j := \pi\alpha_j$, this means that the system (4.5) has a nonnegative solution. Now, in view of Proposition 4.1(*iii*) and Corollary 3.5, $p-\overline{p}$ is small; we can apply Lemma 4.2, $J(p)=\overline{I}\cup\{0\}$, and this is just (4.6).

Uniqueness of such a p is then easy to prove. Substituting f_i for h_2 in Remark 2.3, the gradients of the functions $v\mapsto f_i(\overline{p}+u\oplus v)$ are $g_i(\overline{p}+u\oplus v)_{\mathcal{V}}$, which are linearly independent for $(u,v)=(0,0)$. By the Implicit Function Theorem, (4.6) has a unique solution $w(u)$ for small u. ☐

Now we are in a position to give specific expressions for the derivatives of the $\mathcal{U}$-Lagrangian.

Theorem 4.4. *Use the notation and assumptions of this subsection.*

(i) *The $\mathcal{U}$-Lagrangian is differentiable in a neighborhood of* 0. *With $\mu(\cdot)$ and $w(\cdot)$ defined in Lemma 4.2 and Theorem 4.3 respectively, and with*

$$p(u):=\overline{p}+u\oplus w(u),$$

we have for $u\in\mathcal{U}$ small enough

$$\nabla L_{\mathcal{U}}(u)\oplus 0=\gamma(p(u))+\sum_{j\in\overline{I}}\mu_j(p(u))g_j(p(u)). \tag{4.7}$$

(ii) *The Hessian $\nabla^2 L_{\mathcal{U}}(0)$ exists. Using the matrix-like decomposition*

$$\nabla^2_{pp}L(\overline{p},\overline{\lambda})=\begin{pmatrix} H_{\mathcal{U}\mathcal{U}} & H_{\mathcal{U}\mathcal{V}}\\ H_{\mathcal{V}\mathcal{U}} & H_{\mathcal{V}\mathcal{V}}\end{pmatrix}$$

for the Hessian of the Lagrangian, we have $\nabla^2 L_{\mathcal{U}}(0)=H_{\mathcal{U}\mathcal{U}}$.

Proof. (*i*) Put together Lemma 4.2 and Theorem 4.3. Observe, in particular, that the right-hand side of (4.7) lies in $\mathcal{U}$. Then invoke (3.5).

(*ii*) In view of Lemma 4.1(*iii*) and Corollary 3.5, $w(u)=o(\|u\|_{\mathcal{U}})$, hence $p(\cdot)$ has a Jacobian at 0; in fact, $\mathrm{J}p(0)u=u\oplus 0$ for all $u\in\mathcal{U}$. Then, using Lemma 4.2, (4.7) clearly shows that $\nabla L_{\mathcal{U}}$ is differentiable at 0. Compute from (4.7) the differential $\nabla^2 L_{\mathcal{U}}(0)u$ for $u\in\mathcal{U}$:

$$\begin{aligned}(\nabla^2 L_{\mathcal{U}}(0)u)\oplus 0 &= \nabla^2\psi(\overline{p})\mathrm{J}p(0)u+\sum_{\overline{I}}\overline{\lambda}_j\nabla^2 f_j(\overline{p})\mathrm{J}p(0)u\\ &\qquad +\sum_{\overline{I}}\langle\nabla\mu_j(\overline{p}),\mathrm{J}p(0)u\rangle\overline{g}_j\\ &= \nabla^2_{pp}L(\overline{p},\overline{\lambda})(u\oplus 0)+\sum_{\overline{I}}\langle\nabla\mu_j(\overline{p}),\mathrm{J}p(0)u\rangle\overline{g}_j.\end{aligned}$$

Thus, $\nabla^2 L_{\mathcal{U}}(0)u$ is the $\mathcal{U}$-part of the right-hand side. The second term is a sum of vectors in $\mathcal{V}$, which does not count; we do obtain (*ii*). ☐

In Remark 3.1 we have said that $\overline{g}$ in §3 plays the role of a dual variable. This is suggested by the relation $0=\overline{g}_0+\sum_{\overline{I}}\overline{\lambda}_i\overline{g}_i\in\partial f(\overline{p})$ which, in the present NLP context, establishes a correspondence between $\overline{g}=0$ and the multipliers $\overline{\lambda}_i$ or $\overline{\mu}_i$. Taking some nonzero $\overline{g}'\in\operatorname{ri}\partial f(\overline{p})$ does not change the situation much; this just amounts to applying the theory to $f-\langle\overline{g}',\cdot\rangle$, which is still minimal at $\overline{p}$ – but of course the multipliers are changed, say, to $\overline{\lambda}_i'$ or $\overline{\mu}_i'$. Denoting by $g(p(u))$ the right-hand side in (4.7), the correspondence $\overline{g}\leftrightarrow\overline{\lambda}\leftrightarrow\overline{\mu}$ can even be extended to $g(p(u))\leftrightarrow\overline{\lambda}(u)\leftrightarrow\mu(u)$.

4.2. **Eigenvalue optimization.** Consider the problem of minimizing with respect to $x \in \mathbb{R}^m$ the largest eigenvalue λ_1 of a real symmetric $n \times n$ matrix A, depending affinely on x. Most of the relevant information for the function $\lambda_1 \circ A$ can be obtained by analyzing the maximum eigenvalue function $\lambda_1(A)$, which is convex (and finite-valued). We briefly describe here how the $\mathcal{U}$-theory applies to this context. For a detailed study, we refer to [20] where an interesting connection is established with the geometrical approach of [21].

For the sake of consistency, we keep the notation $\overline{p} := A(\overline{x})$ for the reference matrix where the analysis is performed. If $\overline{r}$ denotes the multiplicity of $\lambda_1(\overline{p})$, then

$$\mathcal{W}_{\overline{r}} := \{p : \ p \text{ is a symmetric matrix and } \lambda_1(p) \text{ has multiplicity } \overline{r}\}$$

is the smooth manifold Ω of [21].

First, the subspaces $\mathcal{U}$ and $\mathcal{V}$ in Definition 2.1 are just the tangent and normal spaces to $\mathcal{W}_{\overline{r}}$ at $\overline{p}$ (Corollary 4.8 in [20]). Similarly to Theorem 4.3, Theorem 4.11 in [20] shows that the set $W(u)$ of (3.2) is a singleton $w(u)$, characterized by

$$\overline{p} + u \oplus w(u) \in \mathcal{W}_{\overline{r}}.$$

As for second order, the $\mathcal{U}$-Lagrangian (3.1) is twice continuously differentiable in a neighbourhood of $0 \in \mathcal{U}$. Finally, use again the matrix-like decomposition

$$\begin{pmatrix} H_{\mathcal{U}\mathcal{U}} & H_{\mathcal{U}\mathcal{V}} \\ H_{\mathcal{V}\mathcal{U}} & H_{\mathcal{V}\mathcal{V}} \end{pmatrix}$$

for the Hessian of the Lagrangian introduced in Theorem 5 of [21]. Then Theorem 4.12 in [20] shows that $\nabla^2 L_{\mathcal{U}}(0) = H_{\mathcal{U}\mathcal{U}}$ is the reduced Hessian matrix (5.31) in [21].

4.3. **A conceptual superlinear scheme.** The previous subsections have shown that our $\mathcal{U}$-objects become classical when f has some special form. It is also demonstrated in [17] and [20] how these $\mathcal{U}$-objects can provide interpretations of known minimization algorithms. Here we go back to a general f and we design a superlinearly convergent conceptual algorithm for minimizing f. Again, we obtain a general formalization of known techniques from classical optimization.

Given p close to a minimum point $\overline{p}$, the problem is to compute some p_+, superlinearly closer to $\overline{p}$. We propose a conceptual scheme, in which we compute first the $\mathcal{V}$-component of the increment $p_+ - p$, and then its $\mathcal{U}$-component. This idea of decomposing the move from p to p_+ in a "vertical" and a "horizontal" step can be traced back to [8].

Algorithm 4.5. *$\mathcal{V}$-Step.* Compute a solution $\delta v \in \mathcal{V}$ of

$$\min\{f(p + 0 \oplus \delta v) : \delta v \in \mathcal{V}\} \tag{4.8}$$

and set $p' := p + 0 \oplus \delta v$.
$\mathcal{U}$-Step. Make a Newton step in $p' + \mathcal{U}$: compute the solution $\delta u \in \mathcal{U}$ of

$$g'_{\mathcal{U}} + \mathrm{H}_{\mathcal{U}} f(\overline{p})\delta u = 0, \tag{4.9}$$

where $g' \in \partial f(p')$ is such that $g'_{\mathcal{V}} = 0$, so that $g'_{\mathcal{U}} \in \partial L_{\mathcal{U}}((p' - \overline{p})_{\mathcal{U}})$.
Update. Set $p_+ := p' + \delta u \oplus 0 = p + \delta u \oplus \delta v$.

Remark 4.6. This algorithm needs the subspace $\mathcal{U}$ associated with $\overline{p}$, as well as the $\mathcal{U}$-Hessian $\mathrm{H}_{\mathcal{U}} f(\overline{p})$, which must exist and be positive definite. The knowledge of $\mathcal{U}$ may be considered as a bold requirement; constructing appropriate approximations of it is for sure a key to obtain implementable forms. As for existence and positive

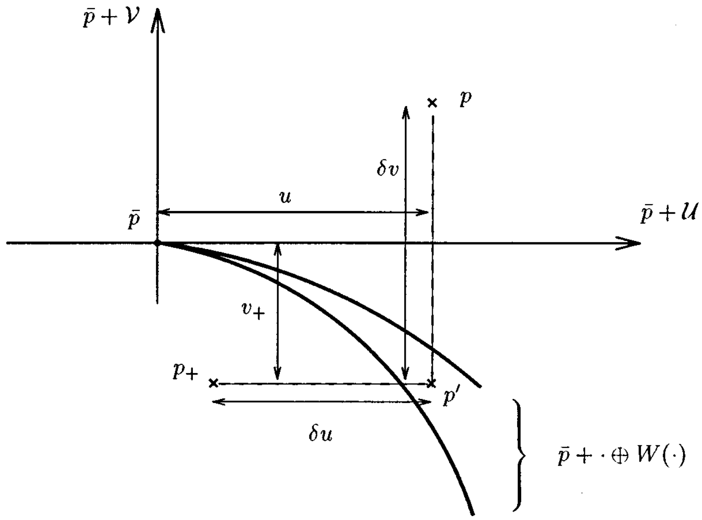

FIGURE 2. Conceptual algorithm

definiteness of $\mathrm{H}_{\mathcal{U}}f(\overline{p})$, it is a natural assumption. Quasi-Newton approximations of it might be suitable, as well as other approaches in the lines of [27]. ☐

The next result supports our scheme.

Theorem 4.7. *Using the notation of §3, assume that* $\overline{g} := 0 \in \operatorname{ri}\partial f(\overline{p})$, *and that* f *has at* $\overline{p}$ *a positive definite* $\mathcal{U}$*-Hessian. Then the point* p_+ *constructed by Algorithm 4.5 satisfies* $\|p_+ - \overline{p}\| = o(\|p - \overline{p}\|)$.

Proof. We denote by $u := (p - \overline{p})_{\mathcal{U}}$ the $\mathcal{U}$-component of $p - \overline{p}$ (see Figure 2). For $\delta v \in \mathcal{V}$, make the change of variables $v := (p-\overline{p})_{\mathcal{V}} + \delta v$, so that (4.8) can be written $\min_{v \in \mathcal{V}} f(\overline{p} + u \oplus v)$. Denoting by v_+ a solution, we have

$$v_+ = (p - \overline{p})_{\mathcal{V}} + \delta v = (p_+ - \overline{p})_{\mathcal{V}} \in W(u)$$

and Corollary 3.5 implies that

$$\|(p_+ - \overline{p})_{\mathcal{V}}\|_{\mathcal{V}} = o(\|u\|_{\mathcal{U}}) = o(\|p - \overline{p}\|). \tag{4.10}$$

From the definition (3.9) of $\mathrm{H}_{\mathcal{U}}f(\overline{p})$ and observing that $\nabla L_{\mathcal{U}}(0) = 0$, we have

$$\partial L_{\mathcal{U}}(u) \ni g'_{\mathcal{U}} = 0 + \mathrm{H}_{\mathcal{U}}f(\overline{p})u + o(\|u\|_{\mathcal{U}}).$$

Subtracting from (4.9), $\mathrm{H}_{\mathcal{U}}f(\overline{p})(u + \delta u) = o(\|u\|_{\mathcal{U}})$ and, since $\mathrm{H}_{\mathcal{U}}f(\overline{p})$ is invertible, $\|u + \delta u\|_{\mathcal{U}} = o(\|u\|_{\mathcal{U}})$. Then, writing

$$(p_+ - \overline{p})_{\mathcal{U}} = (p_+ - p')_{\mathcal{U}} + (p' - p)_{\mathcal{U}} + (p - \overline{p})_{\mathcal{U}} = u + \delta u,$$

we do have $\|(p_+ - \overline{p})_{\mathcal{U}}\|_{\mathcal{U}} = o(\|u\|_{\mathcal{U}}) = o(\|p - \overline{p}\|)$. With (4.10), the conclusion follows. ☐

5. $\mathcal{U}$-HESSIAN AND MOREAU-YOSIDA REGULARIZATIONS

The whole business of §3 was to develop a theory ending up with the definition of a $\mathcal{U}$-Hessian (Definition 3.8). Our aim now is to assess this concept: we give a necessary and sufficient condition for the existence of $\mathrm{H}_{\mathcal{U}}f$, in terms of Moreau-Yosida regularization ([32], [19]).

We denote by F the Moreau-Yosida regularization of f, asssociated with the Euclidean metric,

$$F(x) := \min_{y\in\mathbb{R}^n} \left\{ f(y) + \tfrac{1}{2}\|x-y\|^2 \right\}. \tag{5.1}$$

The unique minimizer in (5.1), called the *proximal* point of x, is denoted by

$$p(x) := \operatorname*{argmin}_{y\in\mathbb{R}^n} \left\{ f(y) + \tfrac{1}{2}\|x-y\|^2 \right\}. \tag{5.2}$$

It is well known that F has a (globally) Lipschitzian gradient, satisfying

$$\nabla F(x) = x - p(x) \in \partial f(p(x)). \tag{5.3}$$

Given $\overline{p}$ and $\overline{g}$ satisfying (1.2), we are interested in the behaviour of F near

$$\overline{x} := \overline{p} + \overline{g} \tag{5.4}$$

(recall, for example, Theorem 2.8 of [15]: $\overline{g} = \nabla F(\overline{x})$ and $\overline{x}$ is such that $p(\overline{x}) = \overline{p}$). More precisely, restricting our attention to $\overline{x}+\mathcal{U}$, we will give an equivalence result and a formula linking the so restricted Hessian of F, with the $\mathcal{U}$-Hessian of f at $\overline{p}$. To prove our results, we introduce an intermediate function, similar to $\phi_\mathcal{V}$ in §3.2 of [15], but adapted to our $\mathcal{U}$-context:

$$\mathcal{U} \ni u \mapsto \phi_\mathcal{V}(u) := \min_{v\in\mathcal{V}} \{ f(\overline{p} + u \oplus v) - \langle \overline{g}_\mathcal{V}, v\rangle_\mathcal{V} + \frac{1}{2}\|v\|^2_\mathcal{V} \}. \tag{5.5}$$

We start by showing that this function agrees up to second order with $L_\mathcal{U}$.

Lemma 5.1. *With the notation above, assume that the conclusion of Corollary 3.5 holds for at least one $w \in W(u)$ – for example, let $\overline{g}$ be in* $\operatorname{ri}\partial f(\overline{p})$. *Then*

$$\forall \varepsilon > 0\ \exists \delta > 0 :\ \|u\|_\mathcal{U} \le \delta \Rightarrow |\phi_\mathcal{V}(u) - L_\mathcal{U}(u)| \le \varepsilon \|u\|^2_\mathcal{U}.$$

In particular,

$$\nabla \phi_\mathcal{V}(0) = \overline{g}_\mathcal{U} \quad \textit{and} \quad \exists \mathrm{H} L_\mathcal{U}(0) \iff \exists \mathrm{H}\phi_\mathcal{V}(0) = \mathrm{H}L_\mathcal{U}(0). \tag{5.6}$$

Proof. Clearly $\phi_\mathcal{V}(u) \ge L_\mathcal{U}(u)$. To obtain an opposite inequality, write the minimand in (5.5) for $v = w \in W(u)$:

$$\begin{aligned}\phi_\mathcal{V}(u) &\le f(\overline{p} + u \oplus w) - \langle \overline{g}_\mathcal{V}, w\rangle_\mathcal{V} + \frac{1}{2}\|w\|^2_\mathcal{V} \\ &= L_\mathcal{U}(u) + \frac{1}{2}\|w\|^2_\mathcal{V}.\end{aligned}$$

Taking, in particular, w such that $\|w\|_\mathcal{V} = o(\|u\|_\mathcal{U})$ (or applying Corollary 3.5), the results follow. ☐

The reason for introducing $\phi_\mathcal{V}$ is that its Moreau-Yosida regularization $\Phi_\mathcal{V}$ is obtained from the restriction $F_\mathcal{U}$ of F to $\overline{x}+\mathcal{U}$ by a mere translation.

Proposition 5.2. *Assume* (1.2). *The two functions*

$$\mathcal{U} \ni d_\mathcal{U} \mapsto \begin{cases} \Phi_\mathcal{V}(d_\mathcal{U}) := \min_{u\in\mathcal{U}} \{\phi_\mathcal{V}(u) + \frac{1}{2}\|d_\mathcal{U} - u\|^2_\mathcal{U}\}, \\ F_\mathcal{U}(d_\mathcal{U}) := F(\overline{x} + d_\mathcal{U} \oplus 0), \end{cases}$$

satisfy

$$F_\mathcal{U}(d_\mathcal{U}) = \Phi_\mathcal{V}(\overline{g}_\mathcal{U} + d_\mathcal{U}) + \frac{1}{2}\|\overline{g}_\mathcal{V}\|^2_\mathcal{V} \quad \textit{for all } d_\mathcal{U} \in \mathcal{U}.$$

Proof. Take $d_{\mathcal{U}} \in \mathcal{U}$. Recalling (5.4), compute $F_{\mathcal{U}}(d_{\mathcal{U}}) = F(\overline{p} + (\overline{g}_{\mathcal{U}} + d_{\mathcal{U}}) \oplus \overline{g}_{\mathcal{V}})$ in the following tricky way:

$$\begin{aligned} F_{\mathcal{U}}(d_{\mathcal{U}}) &= \min_{(u,v)\in\mathcal{U}\times\mathcal{V}} \{f(\overline{p} + u \oplus v) + \frac{1}{2}\|(\overline{g}_{\mathcal{U}} + d_{\mathcal{U}} - u) \oplus (\overline{g}_{\mathcal{V}} - v)\|^2\} \\ &= \min_{u\in\mathcal{U}} \Big\{ \min_{v\in\mathcal{V}} \{ f(\overline{p} + u \oplus v) + \tfrac{1}{2}\|\overline{g}_{\mathcal{V}} - v\|_{\mathcal{V}}^2 \} + \tfrac{1}{2}\|\overline{g}_{\mathcal{U}} + d_{\mathcal{U}} - u\|_{\mathcal{U}}^2 \Big\} \\ &= \min_{u\in\mathcal{U}} \{ \phi_{\mathcal{V}}(u) + \frac{1}{2}\|\overline{g}_{\mathcal{V}}\|_{\mathcal{V}}^2 + \frac{1}{2}\|\overline{g}_{\mathcal{U}} + d_{\mathcal{U}} - u\|_{\mathcal{U}}^2 \} \\ &= \Phi_{\mathcal{V}}(\overline{g}_{\mathcal{U}} + d_{\mathcal{U}}) + \tfrac{1}{2}\|\overline{g}_{\mathcal{V}}\|_{\mathcal{V}}^2 . \quad \square \end{aligned}$$

Since $L_{\mathcal{U}}$ is so close to $\phi_{\mathcal{V}}$ (Lemma 5.1), its Moreau-Yosida regularization is close to $\Phi_{\mathcal{V}}$, i.e., to $F_{\mathcal{U}}$, up to a translation. This explains the next result, which is the core of this section.

Theorem 5.3. *Make the assumptions of Lemma 5.1.*

(i) *If* $\mathrm{H}_{\mathcal{U}} f(\overline{p})$ *exists, then* $\nabla^2 F_{\mathcal{U}}(0)$ *exists and is given by*

$$\nabla^2 F_{\mathcal{U}}(0) = \mathcal{I}_{\mathcal{U}} - (\mathcal{I}_{\mathcal{U}} + \mathrm{H}_{\mathcal{U}} f(\overline{p}))^{-1}\,; \tag{5.7}$$

here $\mathcal{I}_{\mathcal{U}}$ *denotes the identity in* $\mathcal{U}$*.*

(ii) *Conversely, assume that* $\nabla^2 F_{\mathcal{U}}(0)$ *exists. If* (3.7) $\equiv$ (3.8) *holds, then* $\mathrm{H}_{\mathcal{U}} f(\overline{p})$ *exists and is given by*

$$\mathrm{H}_{\mathcal{U}} f(\overline{p}) = (\mathcal{I}_{\mathcal{U}} - \nabla^2 F_{\mathcal{U}}(0))^{-1} - \mathcal{I}_{\mathcal{U}}. \tag{5.8}$$

If, in addition, $\mathrm{H}_{\mathcal{U}} f(\overline{p})$ *is positive definite – for example, if* f *is strongly convex–, we also have*

$$\mathrm{H}_{\mathcal{U}} f(\overline{p}) = (\nabla^2 F_{\mathcal{U}}(0)^{-1} - \mathcal{I}_{\mathcal{U}})^{-1}.$$

Proof. (*i*) When $\mathrm{H}_{\mathcal{U}} f(\overline{p})$ exists, use (5.6) to see that

$$\mathrm{H}_{\mathcal{U}} f(\overline{p}) = \mathrm{H}L_{\mathcal{U}}(0) = \mathrm{H}\phi_{\mathcal{V}}(0). \tag{5.9}$$

Then we can apply Theorem 3.1 of [15] to $\phi_{\mathcal{V}}$. We see from (5.6) that the proximal point giving $\Phi_{\mathcal{V}}(\overline{g}_{\mathcal{U}})$ is $0 \in \mathcal{U}$, so we have

$$\nabla^2 \Phi_{\mathcal{V}}(\overline{g}_{\mathcal{U}}) = \mathcal{I}_{\mathcal{U}} - (\mathcal{I}_{\mathcal{U}} + \mathrm{H}\phi_{\mathcal{V}}(0))^{-1}.$$

In view of Proposition 5.2 and (5.9), this is just (5.7).

(*ii*) Combine Proposition 3.7(*i*) with Lemma 5.1 to see that (3.7) $\equiv$ (3.8) also holds for $\phi_{\mathcal{V}}$ at $0 \in \mathcal{U}$; furthermore, $\nabla\phi_{\mathcal{V}}(0)$ exists. Then we can apply Theorem 3.14 of [15] to $\phi_{\mathcal{V}}$: when $\nabla^2\Phi_{\mathcal{V}}(\overline{g}_{\mathcal{U}}) = \nabla^2 F_{\mathcal{U}}(0)$ exists, then $\mathrm{H}\phi_{\mathcal{V}}(0) = \mathrm{H}_{\mathcal{U}} f(\overline{p})$ exists. We can write (5.7) and invert it to obtain (5.8).

Finally, suppose that f is strongly convex: for some $c > 0$ and all $(u, w) \in \mathcal{U} \times \mathcal{V}$,

$$\begin{aligned} f(\overline{p} + u \oplus w) &\geq f(\overline{p}) + \langle \overline{g}, u \oplus w\rangle + \frac{c}{2}\|u \oplus w\|^2 \\ &\geq f(\overline{p}) + \langle \overline{g}_{\mathcal{U}}, u\rangle_{\mathcal{U}} + \langle \overline{g}_{\mathcal{V}}, w\rangle_{\mathcal{V}} + \frac{c}{2}\|u\|_{\mathcal{U}}^2. \end{aligned}$$

Take $w \in W(u)$ and subtract $\langle \overline{g}_{\mathcal{V}}, w\rangle_{\mathcal{V}}$ from both sides

$$L_{\mathcal{U}}(u) \geq L_{\mathcal{U}}(0) + \langle \nabla L_{\mathcal{U}}(0), u\rangle_{\mathcal{U}} + \frac{c}{2}\|u\|_{\mathcal{U}}^2,$$

hence $\mathrm{H}_{\mathcal{U}} f(\overline{p}) = \mathrm{H}L_{\mathcal{U}}(0)$ is certainly positive definite. Computing its inverse from (5.8) and applying (20) from [15], we obtain the last relation. $\square$

A consequence of this result is that, when $\nabla^2 F(\overline{x})$ exists, then $\mathrm{H}_{\mathcal{U}} f(\overline{p})$ exists; $\nabla^2 F_{\mathcal{U}}(0)$ is just the $\mathcal{U}\mathcal{U}$-block of $\nabla^2 F(\overline{x})$. Furthermore, $x \mapsto p(x)$ has at $\overline{x}$ a Jacobian of the form

$$\mathrm{J}p(\overline{x}) = \mathcal{I} - \nabla^2 F(\overline{x}) = \begin{pmatrix} P & 0 \\ 0 & 0 \end{pmatrix}$$

(recall Corollary 2.6 in [15]). If f satisfies (3.8) at $\overline{p}$, then

$$P = (\mathcal{I} - \nabla^2 F(\overline{x}))_{\mathcal{U}\mathcal{U}} = \mathcal{I}_{\mathcal{U}} - \nabla^2 F_{\mathcal{U}}(0) = (\mathrm{H}_{\mathcal{U}} f(\overline{p}) + \mathcal{I}_{\mathcal{U}})^{-1}$$

is positive definite.

6. Conclusion

The distinctive difficulty of nonsmooth optimization is that the graph of f near a minimum point $\overline{p}$ behaves like an elongated, gully-shaped valley. Such a valley is relatively easy to describe in the composite case (max-functions, maximal eigenvalues): it consists of those points where the non-differentiability of f stays qualitatively the same as at $\overline{p}$; see the considerations developed in [22]. In the general case, however, even an appropriate definition of this valley is already not clear. We believe that the main contribution of this paper lies precisely here: we have generalized the concept of the gully-shaped valley to arbitrary (finite-valued) convex functions. To this aim, we have adopted the following process:

– First, we have used the tangent space to the active constraints, familiar in the NLP world; this was $\mathcal{U}$ of Definition 2.1.
– Then we have defined the gully-shaped valley, together with its parametrization by $u \in \mathcal{U}$, namely the mapping $W(\cdot)$ of (3.2).
– At the same time, we have singled out in (3.5) a selection of subgradients of f, together with a potential function $L_{\mathcal{U}}$. A nice feature is that our definitions are *constructive* via (3.1).
– This has allowed us to reduce the second-order study of f, restricted to the valley, to that of $L_{\mathcal{U}}$ (in $\mathcal{U}$).
– We have shown how our generalizations reduce to known objects in composite optimization, and how they can be used for the design of superlinearly convergent algorithms.
– Finally, we have related our new objects with the Moreau-Yosida regularization of f.

Acknowledgment

We are deeply indebted to R. Mifflin, for his careful reading and numerous helpful suggestions. The $\mathcal{U}$-terminology is due to him.

References

1. J.P. Aubin, *Contingent derivatives of set-valued maps and existence of solutions to nonlinear inclusions and differential equations*, Mathematical Analysis and Applications (L. Nachbin, ed.), Academic Press, 1981, pp. 159–229. MR **83m:**58014
2. D. Azé, *On the remainder of the first order development of convex functions*, Ann. Sci. Math. Quebec **23** (1999), 1–13.
3. H.T. Banks and M.Q. Jacobs, *A differential calculus for multifunctions*, Journal of Mathematical Analysis and Applications **29** (1970), 246–272. MR **42:**846

4. A. Ben-Tal and J. Zowe, *Necessary and sufficient optimality conditions for a class of nonsmooth minimization problems*, Mathematical Programming **24** (1982), no. 1, 70–91. MR **83m:**90075
5. R.W. Chaney, *On second derivatives for nonsmooth functions*, Nonlinear Analysis: Theory, Methods and Applications **9** (1985), no. 11, 1189–1209. MR **87c:**49018
6. T.F. Coleman and A.R. Conn, *Nonlinear programming via an exact penalty function: asymptotic analysis*, Mathematical Programming **24** (1982), 123–136. MR **84e:**90087a
7. R. Cominetti and R. Correa, *A generalized second-order derivative in nonsmooth optimization*, SIAM Journal on Control and Optimization **28** (1990), no. 4, 789–809. MR **91h:**49017
8. A.R. Conn, *Constrained optimization using a nondifferentiable penalty function*, SIAM Journal on Numerical Analysis **10** (1973), no. 4, 760–784. MR **49:**12094
9. 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. 144–177. MR **84i:**49029
10. ———, *A new set-valued second order derivative for convex functions*, Fermat Days 85: Mathematics for Optimization (J.B. Hiriart-Urruty, ed.), Mathematics Studies, no. 129, North-Holland, 1986, pp. 157–182. MR **88d:**90092
11. J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms*, Grundlehren der mathematischen Wissenschaften, no. 305-306, Springer-Verlag, 1993. MR **95m:**90001; MR **95m:**90002
12. A.D. Ioffe, *Nonsmooth analysis and the theory of fans*, Convex Analysis and Optimization (J.P. Aubin and R.B. Vinter, eds.), Pitman, 1982, pp. 93–118. MR **83h:**58012
13. M. Kawasaki, *An envelope-like effect of infinitely many inequality constraints on second-order conditions for minimization problems*, Mathematical Programming **41** (1988), no. 1, 73–96. MR **89d:**90191
14. C. Lemaréchal and C. Sagastizábal, *More than first-order developments of convex functions: primal-dual relations*, Journal of Convex Analysis **3** (1996), 255–268. MR **98k:**49048
15. ———, *Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries*, SIAM Journal on Optimization **7** (1997), no. 2, 367–385. MR **98e:**49085
16. C. Lemaréchal and J. Zowe, *The eclipsing concept to approximate a multi-valued mapping*, Optimization **22** (1991), no. 1, 3–37. MR **92a:**49024
17. R. Mifflin and C. Sagastizábal, *$\mathcal{VU}$-decomposition derivatives for convex max-functions*, Ill-Posed Problems and Variational Inequalities (R. Tichatschke and M. Théra, eds.) Lecture Notes, Springer-Verlag (to appear).
18. B.S. Mordukhovich, *Generalized differential calculus for nonsmooth and set-valued mappings*, Journal of Mathematical Analysis and Applications **183** (1993), no. 1, 250–288. MR **95i:**49029
19. J.J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bulletin de la Société Mathématique de France **93** (1965), 273–299. MR **34:**1829
20. F. Oustry, *The $\mathcal{U}$-Lagrangian of the maximum eigenvalue function*, SIAM Journal of Optimization **9** (1999), 526–549.
21. M. L. Overton and R.S. Womersley, *Second derivatives for optimizing eigenvalues of symmetric matrices*, SIAM J. Matrix Anal. Appl. **16** (1995), 697–718. MR **96c:**65062
22. 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. MR **95e:**90099
23. J.-P. Penot, *Differentiability of relations and differential stability of perturbed optimization problems*, SIAM Journal on Control and Optimization **22** (1984), no. 4, 529–551, **26** (1988), 997–998. MR **85i:**49041; MR **89d:**49027.
24. R.A. Poliquin, *Proto-differentiation of subgradient set-valued mappings*, Canadian Journal of Mathematics **42** (1990), no. 3, 520–532. MR **91g:**49007
25. R.A. Poliquin and R.T. Rockafellar, *Generalized hessian properties of regularized nonsmooth functions*, SIAM Journal on Optimization **6** (1996), no. 4, 1121–1137. MR **97j:**49025
26. M.J.D. Powell, *The convergence of variable metric methods for nonlinearly constrained optimization calculations*, Nonlinear Programming 3 (O.L. Mangasarian, R.R. Meyer, and S.M. Robinson, eds.), 1978, pp. 27–63. MR **80c:**90138
27. L.Q. Qi and D.F. Sun, *A nonsmooth version of Newton's method*, Mathematical Programming **58** (1993), no. 3, 353–367. MR **94b:**90077

28. R.T. Rockafellar, *Convex analysis*, Princeton Mathematics Ser., no. 28, Princeton University Press, 1970. MR **43:**445
29. ———, *Maximal monotone relations and the second derivatives of nonsmooth functions*, Annales de l'Institut Henri Poincaré, Analyse non linéaire **2** (1985), no. 3, 167–186. MR **87c:**49021
30. ———, *Proto-differentiability of set-valued mappings and its applications in optimization*, Annales de l'Institut Henri Poincaré, Analyse Non Linéaire **6** (1989), 449–482. MR **90k:**90140
31. ———, *Generalized second derivatives of convex function and saddle functions*, Transactions of the American Mathematical Society **322** (1990), no. 1, 51–78. MR **91b:**90190
32. K. Yosida, *Functional analysis*, Springer Verlag, 1965. MR **31:**5054

INRIA, 655 avenue de l'Europe, 38330 Montbonnot, France
E-mail address: `Claude.Lemarechal@inria.fr`

INRIA, 655 avenue de l'Europe, 38330 Montbonnot, France
E-mail address: `Francois.Oustry@inria.fr`

INRIA, BP 105, 78153 Le Chesnay, France
E-mail address: `Claudia.Sagastizabal@inria.fr`