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

CLAUDE LEMARÉCHAL, FRANÇOIS OUSTRY, AND CLAUDIA SAGASTIZÁBAL

Abstract. At a given point $p$, a convex function $f$ is differentiable in a certain subspace $\mathcal{U}$ (the subspace along which $\partial f(p)$ has 0-breadth). This property opens the way to defining a suitably restricted second derivative of $f$ at $p$. 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 $p$, the major difficulty is by far the nonlinearity of the first-order approximation. Said otherwise, the gradient vector $\nabla f(p)$ is now a set $\partial f(p)$ and we have to consider difference quotients between sets, say

$$\frac{\partial f(p + h) - \partial f(p)}{\|h\|}.$$  

(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 $f'(p; \cdot)$ (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
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 \( S \), we will write \( \langle \cdot, \cdot \rangle_S \) and \( \| \cdot \|_S \) 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 \( B_S(x,r) \) in a subspace \( S \). We denote by \( x_S \) the projection of a vector \( x \in \mathbb{R}^n \) onto the subspace \( S \). Throughout this paper, we consider the following situation:

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

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

2. The \( \mathcal{U} \mathcal{V} \) 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 \( \mathcal{V}_1 := \mathcal{U}_1^1 \). 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 \( \mathcal{U}_2 := \mathcal{V}_2^2 \). In other words, \( \mathcal{V}_2 := \text{lin}(\partial f(\overline{p}) - \overline{g}) \) for an arbitrary \( \overline{g} \in \partial f(\overline{p}) \), and \( d \in \mathcal{U}_2 \) means \( (\overline{g} + v, d) = (\overline{g}, d) \) 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 \( g^\circ \) in the relative interior of \( \partial f(\overline{p}) \). It is known (see, for example, Proposition 2.2 in [14]) that the property \( g^\circ \in \text{ri } \partial f(\overline{p}) \) is equivalent to these cones being subspaces.

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

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

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

**Proposition 2.2.** In Definition 2.1,

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

\[
(2.2) \quad \{ d \in \mathbb{R}^n : (g - g^\circ, d) = 0 \text{ for all } g \in \partial f(\overline{p}) \} = N_{\partial f(\overline{p})}(g^\circ)
\]

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

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

(iii) \( \mathcal{U} \subset N_{\partial f(\overline{p})}(\overline{g}) \) for all \( \overline{g} \in \partial f(\overline{p}) \).
Proof. (i) To prove (2.2), take \( g^\circ \in \partial f(\overline{p}) \) and set \( N := N_{\partial f(\overline{p})}(g^\circ) \). 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 \( g - g^\circ \neq 0 \)), \( v := -\frac{g - g^\circ}{\|g - g^\circ\|} \in V_2 \), hence (2.1) and \( d \in N \) imply that

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

and we are done.

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

\[
N_{\partial f(\overline{p})}(\gamma^\circ) = \{ d \in \mathbb{R}^n : \langle g, d \rangle = \langle \gamma^\circ, d \rangle = \langle g^\circ, d \rangle, \text{ for all } g \in \partial f(\overline{p}) \} = U_1.
\]

(ii) Write

\[
U_1 = \{ d \in \mathbb{R}^n : \max_{g \in \partial f(\overline{p})} \langle g, d \rangle = \min_{g \in \partial f(\overline{p})} \langle g, d \rangle \},
\]

to see from (i) that \( U_1 = U_3 \). Then we only need to prove \( U_1 \subset U_2 \subset U_3 \).

Let \( d \in U_1 \). For an arbitrary \( v = \sum_j \lambda_j (g_j - \overline{g}) \in V_2 \) with \( g_j \in \partial f(\overline{p}) \), we have from (2.3)

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

hence \( d \in \mathcal{J}^2 \Rightarrow U_2 \).

Let \( d \in U_2 \). 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 \( d \in U_3 \).

(iii) Let \( d \in U = U_3 \). 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 \( d \in N_{\partial f(\overline{p})}(\overline{g}) \).

\( \square \)

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

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

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

\[
\langle g, x \rangle = \langle g_U + g_V, x_U + x_V \rangle = \langle g_U, x_U \rangle + \langle g_V, x_V \rangle,
\]

with similar expressions for norms.

Remark 2.3. The projection \( x \mapsto x_U \), 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

\[
U \ni u \mapsto h_1(u) := f(\overline{p} + u \oplus v), \quad \text{with } v \in V \text{ arbitrary};
\]

\[
V \ni v \mapsto h_2(v) := f(\overline{p} + u \oplus v), \quad \text{with } u \in U \text{ arbitrary};
\]

\[
U \times V \ni (u, v) \mapsto h(u, v) := f(\overline{p} + u \oplus v).
\]

Their subdifferentials have the expressions

\[
\partial h_1(u) = \{ g_U : g \in \partial f(\overline{p} + u \oplus v) \},
\]

\[
\partial h_2(v) = \{ g_V : g \in \partial f(\overline{p} + u \oplus v) \},
\]

\[
\partial h(x_U, x_V) = \{ g_U \oplus g_V : g \in \partial f(\overline{p} + x) \}.
\]
Proving these formulae is a good exercise to become familiar with the operation $\oplus$ of (2.4) and with our $VU$ notation. Just consider the adjoint of $\oplus$ and of the projections onto the various subspaces involved.

In the $VU$ language, (2.1) gives the following elementary result.

**Proposition 2.4.** Suppose in (1.2) that $\bar{g} \in \operatorname{ri} \partial f(p)$. Then there exists $\eta > 0$ small enough such that
\[
\bar{g} + 0 \oplus \frac{\eta v}{\|v\|_V} \in \partial f(p)
\]
for any $0 \neq v \in V$. In particular,
\[
f(p + u \oplus v) \geq f(p) + \langle \bar{g}_U, u \rangle_U + \langle \bar{g}_V, v \rangle_V + \eta \|v\|_V,
\]
for any $(u, v) \in U \times V$.

**Proof.** Just translate (2.1): with $v$ as stated, $u \oplus v \bar{g} U + \frac{\eta v}{\|v\|_V} \in \partial f(p)$ and the rest follows easily. \qed

3. **The $U$-Lagrangian**

In this section we formalize the theory outlined in §3.2 of [15]. Along with the $VU$ decomposition, we introduced there the “tangential” regularization $\varphi_U$. Here, we find it convenient to consider $\varphi_U$ as a function defined on $U$ 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 $L_U$ instead of $\varphi_U$.

3.1. **Definition and basic properties.** Following the above introduction, we define the function $L_U$ as follows:

\[
(3.1) \quad U \ni u \mapsto L_U(u) := \inf_{v \in V} \{ f(p + u \oplus v) - \langle \bar{g}_V, v \rangle_V \}.
\]

Associated with (3.1) we have the set of minimizers

\[
(3.2) \quad W(u) := \operatorname{Argmin}_{v \in V} \{ f(p + u \oplus v) - \langle \bar{g}_V, v \rangle_V \}.
\]

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

**Remark 3.1.** The function $L_U$ of (3.1) will be called the $U$-Lagrangian. Note that it depends on the particular $\bar{g}$, a notation $L_U(u, \bar{g})$ is also possible. In fact, since $\bar{g}$ 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 $\bar{g} \mapsto -L_U$ 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 $\bar{g}$, say,

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

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 $V$, we introduce the function

\[
f^*(\bar{g} + s) + \frac{1}{2} c \|s_V\|^2_V.
\]
Its conjugate (restricted to $U$) is precisely $L_U$ when $c = +\infty$ (a value which yields the “strongest” possible convexity); Theorem 3.3 will confirm the smoothness of $L_U$.

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.

\textbf{Theorem 3.2.} Assume (1.2).

(i) The function $L_U$ 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 $g \in \partial f(p + u \oplus w)$ such that $g_v = \overline{g}_v$.

(iii) In particular, $0 \in W(0)$ and $L_U(0) = f(\overline{p})$.

(iv) If $\overline{p} \in \text{ri} \partial f(p)$, then $W(u)$ is nonempty for each $u \in U$ and $W(0) = \{0\}$.

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

$$h(u, v) - \langle \overline{g}_v, v \rangle_V \geq f(\overline{p}) + \langle \overline{g}_U, u \rangle_U$$

for any $v \in V$.

It follows that $L_U$ 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 $0 \in \partial h_2(w) - \overline{g}_v$, with $h_2$ as in Remark 2.3. Knowing the expression of $\partial h_2$, we obtain $0 = g_v - \overline{g}_v$, for some $g \in \partial f(p + u \oplus w)$.

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

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

$$h(u, v) - \langle \overline{g}_v, v \rangle_V \geq f(\overline{p}) + \langle \overline{g}_U, u \rangle_U + \eta \|v\|_V.$$

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

$$h(0, v) - \langle \overline{g}_v, v \rangle_V \geq f(\overline{p}) + \eta \|v\|_V,$$

which shows that $v = 0$ is the unique minimizer. \hfill \square

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

\textbf{Theorem 3.3.} Assume (1.2).

(i) Let $u$ be such that $W(u) \neq \emptyset$. Then the subdifferential of $L_U$ at this $u$ has the expression

$$\partial L_U(u) = \{g_U : g_U \oplus \overline{g}_V \in \partial f(p + u \oplus w)\},$$

where $w$ is an arbitrary point in $W(u)$.

(ii) In particular, $L_U$ is differentiable at 0, with $\nabla L_U(0) = \overline{g}_U$.

\textbf{Proof.} (i) Using again the notation of Remark 2.3, write the infimand in (3.1) as $h(u, v) - \langle 0 \oplus \overline{g}_V, u \oplus v \rangle$. For the subdifferential of the marginal function $L_U$,
Corollary VI.4.5.3 in [11] gives the calculus rule
\[ s \in \partial_u L_{\mathcal{U}}(u) \iff s \oplus 0 \in \partial_u (h - \langle 0 \oplus g_U, \cdot \rangle)(u, w) \]
\[ \iff s \oplus 0 \in \partial_u h(u, w) - 0g_V \]
\[ \iff s \oplus g_V \in \partial_{u, w} h(u, w), \]
where \( w \in W(u) \) is arbitrary. From the expression of \( \partial_{u, w} h = \partial h \) in Remark 2.3, this is (3.4).

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

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 \( p + u \oplus W(u) \) that have the same \( \mathcal{V} \)-component as \( g \) (namely \( g_V \)); 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
\[ \partial L_{\mathcal{U}}(u) = [\partial f(p + u \oplus W(u)) \cap (g + \mathcal{V})]_{\mathcal{U}}. \]
This operation somewhat simplifies when \( g_V = 0 \):
\[ (3.5) \quad \text{if } g_V = 0, \text{ then } \partial L_{\mathcal{U}}(u) = \partial f(p + u \oplus W(u)) \cap \mathcal{U}. \]
See the end of §3.2 below for additional comments on the “trajectories” \( p + u \oplus W(u) \).

Another observation is that, for all \( u \in \mathcal{U} \),
\[ f'(p; u \oplus 0) = \langle \mathcal{F}, u \oplus 0 \rangle = \langle \mathcal{F}_U, u \rangle_\mathcal{U} = \langle \nabla L_{\mathcal{U}}(0), u \rangle_\mathcal{U}. \]
In other words, \( L_\mathcal{U} \) agrees, up to first order, with the restriction of \( f \) to \( \mathcal{P} + \mathcal{U} \). Continuing with our \( \mathcal{U} \)-terminology, we will say that \( \overline{\mathcal{g}}_\mathcal{U} \) is the \( \mathcal{U} \)-gradient of \( f \) at \( \overline{\mathcal{P}} \), and note that \( \overline{\mathcal{g}}_\mathcal{U} \) is actually independent of the particular \( \overline{\mathcal{g}} \in \partial f(\overline{\mathcal{P}}) \) (recall Proposition 2.2(ii)).

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

\[
(3.6) \quad f(\overline{\mathcal{P}} + h) = f(\overline{\mathcal{P}}) + \langle \overline{\mathcal{g}}, h \rangle + o(\|h\|) \quad \text{whenever} \quad \|h\|_V = o(\|h\|_\mathcal{U}).
\]

This remark will be instrumental when considering 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

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

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

**Corollary 3.5.** Assume (1.2). If \( \overline{\mathcal{g}} \in \partial f(\overline{\mathcal{P}}) \), then

\[
\forall \varepsilon > 0 \exists \delta > 0 : \|u\|_\mathcal{U} \leq \delta \Rightarrow \|u\|_V \leq \varepsilon \|u\|_\mathcal{U} \quad \text{for any} \ w \in W(u).
\]

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

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

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

\[
o(\|u\|_\mathcal{U}) = L_\mathcal{U}(u) - f(\overline{\mathcal{P}}) - \langle \overline{\mathcal{g}}_\mathcal{U}, u \rangle_\mathcal{U} \geq \eta \|w\|_V.
\]

Let us sum up our results so far.

- Given \( \overline{\mathcal{g}} \in \partial f(\overline{\mathcal{P}}) \), 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 \( \overline{\mathcal{P}} + \mathcal{U} \) (Theorem 3.3(ii)).
- When \( W(\cdot) \neq 0 \), this \( \mathcal{U} \)-Lagrangian is indeed the restriction of \( f \) to a “thick surface” \( \{ \overline{\mathcal{P}} + \cdot \oplus W(\cdot) \} \), parametrized by \( u \in \mathcal{U} \).
- We also define, via Theorem 3.2(ii), a “thick selection” of \( \partial f \) on this thick surface, made up of those subgradients that have the same \( \mathcal{V} \)-component as \( \overline{\mathcal{g}} \).
- As a function of the parameter \( u \), this thick selection behaves like a subdifferential, namely \( \partial L_\mathcal{U} \) (Theorem 3.3(i)).
- When \( \overline{\mathcal{g}} \in \partial f(\overline{\mathcal{P}}) \), our thick surface has \( \mathcal{U} \) as “tangent space” at \( \overline{\mathcal{P}} \) (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 \( \overline{\mathcal{P}} \), then \( \mathcal{U} = \mathbb{R}^n \), \( \mathcal{V} = \{0\} \) and \( L_\mathcal{U} \equiv f \);
- when \( \partial f(\overline{\mathcal{P}}) \) has full dimension, then \( \mathcal{U} = \{0\} \) and there is no \( \mathcal{U} \)-Lagrangian.  \( \square \)
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 $\bar{p}$ when there is a $D > 0$ and a $\delta > 0$ such that
\[
\partial f(\bar{p} + d) \subset \partial f(\bar{p}) + B(0, D\|d\|), \quad \text{for all } d \in B(0, \delta).
\]
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
\[
f(\bar{p} + d) \leq f(\bar{p}) + f'(\bar{p}; d) + \frac{1}{2}C\|d\|^2, \quad \text{for all } d \in B(0, \varepsilon).
\]
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) is satisfied. Then
\begin{enumerate}[(i)]  
  \item $\partial L_{\mathcal{U}}(u) \subset \overline{g_{\mathcal{U}}} + B_{\mathcal{U}}(0, 2C\|u\|_{\mathcal{U}})$, for some $\delta > 0$ and all $u \in B_{\mathcal{U}}(0, \delta)$;
  \item $L_{\mathcal{U}}(u) \leq L_{\mathcal{U}}(0) + \langle \overline{g_{\mathcal{U}}}, u \rangle_{\mathcal{U}} + \frac{1}{2}R\|u\|^2_{\mathcal{U}}$, for some $\rho > 0$, $R > 0$ and all $u \in B_{\mathcal{U}}(0, \rho)$.
\end{enumerate}

**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 $g_{\mathcal{U}} \in \partial L_{\mathcal{U}}(u)$, $\|g_{\mathcal{U}} - \overline{g_{\mathcal{U}}}\|_{\mathcal{U}} \leq \frac{\delta C}{\rho}$ (see §VI.6.2 of [11] for example). On the other hand, assume $\delta$ so small that $W(u)$ contains some $w$; from Theorem 3.2(ii),

Now $\mathcal{U} \subset N_{\partial f(\bar{p})}(\bar{g})$ (Proposition 2.2(iii)). Using the notation $s := (g_{\mathcal{U}} - \overline{g_{\mathcal{U}}})_{\mathcal{U}} + 0$, so that $g_{\mathcal{U}} + \overline{g_{\mathcal{U}}} = \bar{g} + s \in \partial f(\bar{p} + u \oplus w)$, we are in the conditions of Corollary 3.3 in [14] written with $\varphi = f$, $z_0 = \bar{p}$, $g_0 = \bar{g}$, $x = \bar{p} + u \oplus w$. Inequality (14) therein becomes
\[
\|g_{\mathcal{U}} - \overline{g_{\mathcal{U}}}\|_{\mathcal{U}}^2 = \|s\|^2 \leq 2C(s, u \oplus w) = 2C(g_{\mathcal{U}} - \overline{g_{\mathcal{U}}}, u)_{\mathcal{U}} \leq 2C\|g_{\mathcal{U}} - \overline{g_{\mathcal{U}}}\|_{\mathcal{U}}\|u\|_{\mathcal{U}},
\]
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
\[
\{g_{\mathcal{U}} : g_{\mathcal{U}} + \overline{g_{\mathcal{U}}} \in \partial f(\bar{p} + u \oplus w)\} \subset \overline{g_{\mathcal{U}}} + B_{\mathcal{U}}(0, 2C\|u\|_{\mathcal{U}})
\]
as well as
\[
f(\bar{p} + u \oplus w) \leq f(\bar{p}) + \langle \bar{g}, u \oplus w \rangle + \frac{1}{2}R\|u\|^2_{\mathcal{U}}.
\]
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 $\varphi$ has at $z_0$ a generalized Hessian $H\varphi(z_0)$ when
\begin{enumerate}[(i)]  
  \item the gradient $\nabla \varphi(z_0)$ exists;
  \item there exists a symmetric positive semidefinite operator $H\varphi(z_0)$ such that
    \[
    \varphi(z_0 + d) = \varphi(z_0) + \langle \nabla \varphi(z_0), d \rangle + \frac{1}{2}(H\varphi(z_0)d, d) + o(\|d\|^2);
    \]
  \item or equivalently,
    \[
    \partial \varphi(z_0 + d) \subset \nabla \varphi(z_0) + H\varphi(z_0)d + B(0, o(\|d\|)).
    \]\end{enumerate}
Definition 3.8. Assume (1.2). We say that \( f \) has at \( \bar{p} \) a \( \mathcal{U} \)-Hessian \( H_uf(\bar{p}) \) (associated with \( \bar{g} \)) if \( L_uf \) has a generalized Hessian at 0; then we set

\[
H_uf(\bar{p}) := HL_uf(0).
\]

When it exists, the \( \mathcal{U} \)-Hessian \( H_uf(\bar{p}) \) is therefore a symmetric positive semidefinite operator from \( \mathcal{U} \) to \( \mathcal{U} \). Its existence means the possibility of expanding \( f \) along the thick surface \( \bar{p} + \odot W(\cdot) \) introduced at the end of §3.2.

Theorem 3.9. Take \( \bar{g} \in \text{ri } \partial f(\bar{p}) \) and let the \( \mathcal{U} \)-Hessian \( H_uf(\bar{p}) \) exist. For \( u \in \mathcal{U} \) and \( h \in u \odot W(u) \), there holds

\[
f(\bar{p} + h) = f(\bar{p}) + \langle \bar{g}, h \rangle + \frac{1}{2} \langle H_uf(\bar{p})u, u \rangle_\mathcal{U} + o(\|h\|^2).
\]

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

\[
L_uf(u) = f(\bar{p} + u \odot w) - \langle \bar{g}_o, w \rangle_{\mathcal{V}} = L_uf(0) + \langle \nabla L_uf(0), u \rangle_\mathcal{U} + \frac{1}{2} \langle H_uf(\bar{p})u, u \rangle_\mathcal{U} + o(\|u\|^2) = f(\bar{p}) + \langle \bar{g}_o, u \rangle_\mathcal{U} + \frac{1}{2} \langle H_uf(\bar{p})u, u \rangle_\mathcal{U} + o(\|u\|^2).
\]

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

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

\[
\{g_uf : g_uf \odot \bar{g}_o \in \partial f(\bar{p} + h)\} \subset \bar{g}_uf + H_uf(\bar{p})u + B_uf(0, o(\|h\|)).
\]

With reference to Remark 3.4, the expansion (3.10) makes (3.6) more explicit, for increments \( h = h_uf \odot h_\mathcal{V} \) such that \( h_\mathcal{V} \in W(h_uf) \). 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 semidefinite programming (§4.2). Then in §4.3 we outline a conceptual minimization algorithm.

4.1. Exact penalty. Consider an ordinary nonlinear programming problem

\[
\begin{align*}
\min & \psi(p), \\
\text{s.t.} & \quad f_i(p) \leq 0, \quad i = 1, \ldots, m,
\end{align*}
\]

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

\[
\begin{align*}
[\nabla_p L(\bar{p}, \bar{\lambda}) =] \quad \nabla \psi(\bar{p}) + \sum \lambda_i \nabla f_i(\bar{p}) = 0, \\
\bar{\lambda}_i \geq 0 \quad \bar{\lambda}_i f_i(\bar{p}) = 0, \quad \text{for } i = 1, \ldots, m.
\end{align*}
\]

We will use the notation \( \gamma := \nabla \psi \), \( g_i := \nabla f_i \), \( \gamma := \nabla \psi(\bar{p}) \), \( \gamma_i := \nabla f_i(\bar{p}) \).
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
\[
(4.3) \quad f(p) := \psi(p) + \pi \max\{f_0(p), \ldots, f_m(p)\},
\]
where \( \pi > 0 \) is a penalty parameter. Call
\[
J(p) := \{ j \in \{0, \ldots, m\} : \psi(p) + \pi f_j(p) = f(p) \}
\]
the set of indices realizing the max at \( p \). Standard subdifferential calculus gives
\[
\partial f(p) = \gamma(p) + \pi \text{conv}\{g_j(p) : j \in J(p)\}.
\]
In NLP language, instead of maximal functions, one speaks of active constraints. We therefore set
\[
\pi > 0 \text{ is a penalty parameter. Call } J(p) := \{ j \in \{0, \ldots, m\} : \psi(p) + \pi f_j(p) = f(p) \}
\]
the set of indices realizing the max at \( p \). Standard subdifferential calculus gives
\[
\partial f(p) = \gamma(p) + \pi \text{conv}\{g_j(p) : j \in J(p)\}.
\]
In NLP language, instead of maximal functions, one speaks of active constraints. We therefore set
\[
\pi > 0 \text{ is a penalty parameter. Call } J(p) := \{ j \in \{0, \ldots, m\} : \psi(p) + \pi f_j(p) = f(p) \}
\]
the set of indices realizing the max at \( p \). Standard subdifferential calculus gives
\[
\partial f(p) = \gamma(p) + \pi \text{conv}\{g_j(p) : j \in J(p)\}.
\]
In NLP language, instead of maximal functions, one speaks of active constraints. We therefore set
\[
\pi > 0 \text{ is a penalty parameter. Call } J(p) := \{ j \in \{0, \ldots, m\} : \psi(p) + \pi f_j(p) = f(p) \}
\]
the set of indices realizing the max at \( p \). Standard subdifferential calculus gives
\[
\partial f(p) = \gamma(p) + \pi \text{conv}\{g_j(p) : j \in J(p)\}.
\]
In NLP language, instead of maximal functions, one speaks of active constraints. We therefore set
\[
\pi > 0 \text{ is a penalty parameter. Call } J(p) := \{ j \in \{0, \ldots, m\} : \psi(p) + \pi f_j(p) = f(p) \}
\]
the set of indices realizing the max at \( p \). Standard subdifferential calculus gives
\[
\partial f(p) = \gamma(p) + \pi \text{conv}\{g_j(p) : j \in J(p)\}.
\]
In NLP language, instead of maximal functions, one speaks of active constraints. We therefore set
\[
\pi > 0 \text{ is a penalty parameter. Call } J(p) := \{ j \in \{0, \ldots, m\} : \psi(p) + \pi f_j(p) = f(p) \}
\]
the set of indices realizing the max at \( p \). Standard subdifferential calculus gives
\[
\partial f(p) = \gamma(p) + \pi \text{conv}\{g_j(p) : j \in J(p)\}.
\]
In NLP language, instead of maximal functions, one speaks of active constraints. We therefore set
\[
\pi > 0 \text{ is a penalty parameter. Call } J(p) := \{ j \in \{0, \ldots, m\} : \psi(p) + \pi f_j(p) = f(p) \}
\]
the set of indices realizing the max at \( p \). Standard subdifferential calculus gives
\[
\partial f(p) = \gamma(p) + \pi \text{conv}\{g_j(p) : j \in J(p)\}.
\]
In NLP language, instead of maximal functions, one speaks of active constraints. We therefore set
\[
\pi > 0 \text{ is a penalty parameter. Call } J(p) := \{ j \in \{0, \ldots, m\} : \psi(p) + \pi f_j(p) = f(p) \}
\]
the set of indices realizing the max at \( p \). Standard subdifferential calculus gives
\[
\partial f(p) = \gamma(p) + \pi \text{conv}\{g_j(p) : j \in J(p)\}.
\]
In NLP language, instead of maximal functions, one speaks of active constraints. We therefore set
\[
\pi > 0 \text{ is a penalty parameter. Call } J(p) := \{ j \in \{0, \ldots, m\} : \psi(p) + \pi f_j(p) = f(p) \}
\]
the set of indices realizing the max at \( p \). Standard subdifferential calculus gives
\[
\partial f(p) = \gamma(p) + \pi \text{conv}\{g_j(p) : j \in J(p)\}.
\]
In NLP language, instead of maximal functions, one speaks of active constraints. We therefore set
\[
\pi > 0 \text{ is a penalty parameter. Call } J(p) := \{ j \in \{0, \ldots, m\} : \psi(p) + \pi f_j(p) = f(p) \}
\]
the set of indices realizing the max at \( p \). Standard subdifferential calculus gives
\[
\partial f(p) = \gamma(p) + \pi \text{conv}\{g_j(p) : j \in J(p)\}.
\]
In NLP language, instead of maximal functions, one speaks of active constraints. We therefore set
\[
\pi > 0 \text{ is a penalty parameter. Call } J(p) := \{ j \in \{0, \ldots, m\} : \psi(p) + \pi f_j(p) = f(p) \}
\]
the set of indices realizing the max at \( p \). Standard subdifferential calculus gives
\[
\partial f(p) = \gamma(p) + \pi \text{conv}\{g_j(p) : j \in J(p)\}.
\]
(ii) Apply Definition 2.1(ii): $\mathcal{V} = \text{lin}\{\partial f(p) - \mathbf{v}\}$ because $\mathbf{v} \in \partial f(p)$. Together with (i), the results clearly follow.

(iii) Consider the set $\mathcal{B} := \{\sum_{i} \mu_{i} \mathbf{g}_{i} : \mu_{i} \geq -p_{i}, \sum_{i} \mu_{i} \leq p_{0}\}$, where $p$ was defined in (4.4). Because of (ii), $\mathcal{B} \subset \mathcal{V}$. Because of strict complementarity and $p_{0} > 0$, $\mathcal{B}$ is a relative neighborhood of $0 = \mathbf{v} \in \mathcal{V}$. Finally, because of (4.2) and (4.4),

$$\mathcal{B} = \mathbf{v} + \mathcal{B} + \sum_{i} \lambda_{i} \mathbf{g}_{i} = \mathbf{v} + \{\sum_{i} (\mu_{i} + p_{i}) \mathbf{g}_{i} : \mu_{i} + p_{i} \geq 0, \sum_{i} (\mu_{i} + p_{i}) \leq \pi\}.$$

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

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

$$\begin{cases}
\langle \mathbf{g}_{i}, \gamma(p) \rangle + \sum_{j \in J(p)} \mu_{j} \langle \mathbf{g}_{i}, \mathbf{g}_{j}(p) \rangle = 0 & \text{for all } i \in \overline{T}, \\
\sum_{j \in J(p)} \mu_{j} = \pi
\end{cases}
$$

(4.5)

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

**Proof.** Let $j \notin J(\mathbf{p})$. By continuity, $f_{j}(p) < f_{i}(p)$ for all $i \in J(\mathbf{p})$, hence $J(p) \subset J(\mathbf{p})$.

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

(a) Assume first that $J(p) = J(\mathbf{p}) = \overline{T} \cup \{0\}$. Since $g_{0}(p) \equiv 0$, the variable $\mu_{0}$ is again directly given by $\mu_{0}(p) = \pi - \sum_{j} \mu_{j}(p)$. As for the $\mu_{j}$’s, $j \in \overline{T}$, they are given by an $\overline{T} \times \overline{T}$ linear system, whose matrix is $((\mathbf{g}_{i}, g_{j}(p)))_{ij}$. Because the $\mathbf{g}_{i}$’s are linearly independent, this matrix is positive definite. The solution $\mu(p)$ is unique; it is also close to $\mathbf{p}$, is therefore positive and sums up to less than $\pi$: $\mu_{0}(p) > 0$. In particular, $\mu(\mathbf{p}) = \mathbf{p}$ is the unique solution at $p = \mathbf{p}$. The differentiability property then follows from the Implicit Function Theorem.

(b) On the other hand, assume the set $I_{0} := J(\mathbf{p}) \setminus J(p)$ 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 $\mu^{*}$ also solves (4.5) with $J(p)$ replaced by $J(\mathbf{p})$. 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 $u \in U$ small enough, $W(u)$ defined in (3.2) is a singleton $w(u)$, which is the unique solution of the system with unknown $v \in \mathcal{V}$

$$f_{i}(p + u + v) = 0, \quad \text{for all } i \in \overline{T}.
$$

(4.6)

**Proof.** According to Theorem 3.2(ii) and (3.5), an arbitrary $p \in \mathbf{p} + u + W(u)$ is characterized by $\partial f(p) \cap U \neq \emptyset$; there are convex multipliers $\{\alpha_{j}\}_{j \in J(p)}$ such that $\gamma(p) + \pi \sum_{j} \alpha_{j} g_{j}(p) \in U$. 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, $J(p) = T \cup \{0\}$, 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 $g_i(\overline{p} + u \oplus v)_V$, which are linearly independent for $(u, v) = (0, 0)$. By the Implicit Function Theorem, (4.6) has a unique solution $w(u)$ for small $u$. \hfill \Box

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

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

- (i) The $U$-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 $u \in U$ small enough

\begin{equation}
\nabla L_U(u) \oplus 0 = \gamma(p(u)) + \sum_{j \in I} \mu_j(p(u))g_j(p(u)).
\end{equation}

- (ii) The Hessian $\nabla^2 L_U(0)$ exists. Using the matrix-like decomposition

\[\nabla^2_{pp} L(\overline{p}, \overline{\lambda}) = \begin{pmatrix} H_{UU} & H_{UV} \\ H_{VU} & H_{VV} \end{pmatrix}\]

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

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

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

\[
(\nabla^2 L_U(0)u) \oplus 0 = \nabla^2 \psi(\overline{p})Jp(0)u + \sum_{j \in I} \overline{\lambda}_j \nabla^2 f_j(\overline{p})Jp(0)u \\
+ \sum_{j \in I} \langle \nabla \mu_j(\overline{p}), Jp(0)u \rangle \overline{g}_j \\
= \nabla^2_{pp} L(\overline{p}, \overline{\lambda})(u \oplus 0) + \sum_{j \in I} \langle \nabla \mu_j(\overline{p}), Jp(0)u \rangle \overline{g}_j.
\]

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

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 $0 = \overline{g} + \sum_{i \in I} \overline{\lambda}_i \overline{g}_i \in \partial f(\overline{p})$ which, in the present NLP context, establishes a correspondence between $\overline{g} = 0$ and the multipliers $\overline{\lambda}_i$ or $\overline{g}_i$. Taking some nonzero $\overline{g}' \in ri \partial f(\overline{p})$ 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 $\overline{g}_i'$. Denoting by $g(p(u))$ the right-hand side in (4.7), the correspondence $\overline{g} \leftrightarrow \overline{\lambda} \leftrightarrow \overline{p}$ can even be extended to $g(p(u)) \leftrightarrow \overline{\lambda}(u) \leftrightarrow \mu(u)$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
4.2. Eigenvalue optimization. Consider the problem of minimizing with respect to \( x \in \mathbb{R}^m \) the largest eigenvalue \( \lambda_1 \) 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 \( p := A(x) \) for the reference matrix where the analysis is performed. If \( r \) denotes the multiplicity of \( \lambda_1(p) \), then \( W_r := \{ p : p \text{ is a symmetric matrix and } \lambda_1(p) \text{ has multiplicity } r \} \) is the smooth manifold \( \Omega \) of [21].

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

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

\[
\begin{pmatrix}
H_{u\mathcal{U}} & H_{u\mathcal{V}} \\
H_{v\mathcal{U}} & H_{v\mathcal{V}}
\end{pmatrix}
\]

for the Hessian of the Lagrangian introduced in Theorem 5 of [21]. Then Theorem 4.12 in [20] shows that \( \nabla^2 L_u(0) = H_{u\mathcal{U}} \) 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 \( \bar{p} \), the problem is to compute some \( p_+ \), superlinearly closer to \( p \). 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

\[
\min \{ f(p + 0 \oplus \delta v) : \delta v \in \mathcal{V} \}
\]

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

\[
g'_{u\mathcal{U}} + H_{u\mathcal{U}} f(p) \delta u = 0,
\]

where \( g' \in \partial f(p') \) is such that \( g'_{v\mathcal{U}} = 0 \), so that \( g'_{u\mathcal{U}} \in \partial L_u((p' - \bar{p}) \mathcal{U}) \).

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 \( \bar{p} \), as well as the \( \mathcal{U} \)-Hessian \( H_{u\mathcal{U}} f(\bar{p}) \), 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
definiteness of $H_{U}f(\overline{p})$, 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 $g := 0 \in \text{ri} \partial f(\overline{p})$, and that $f$ has at $\overline{p}$ a positive definite $U$-Hessian. Then the point $p_{+}$ constructed by Algorithm 4.5 satisfies

$$
\|p_{+} - p\| = o(\|p - \overline{p}\|).
$$

**Proof.** We denote by $u := (p - \overline{p})_{U}$ the $U$-component of $p - \overline{p}$ (see Figure 2). For $\delta v \in V$, make the change of variables $v := (p - \overline{p})_{V} + \delta v$, so that (4.8) can be written

$$
\min_{v \in V} f(p + u \oplus v).
$$

Denoting by $v_{+}$ a solution, we have

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

and Corollary 3.5 implies that

$$
\|v_{+}\|_{V} = o(\|u\|_{U}) = o(\|p - \overline{p}\|).
$$

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

$$
\partial L_{U}(u) \ni g_{U} = 0 + H_{U}f(\overline{p})u + o(\|u\|_{U}).
$$

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

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

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

5. $U$-HESIAN AND MOREAU-YOSIDA REGULARIZATIONS

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

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

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

$$
(5.2) \quad p(x) := \arg\min_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{1}{2} \|x - y\|^2 \right\}.
$$

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

$$
(5.3) \quad \nabla F(x) = x - p(x) \in \partial f(p(x)).
$$

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

$$
(5.4) \quad \pi := \p + \g
$$

(recall, for example, Theorem 2.8 of [15]: $\g = \nabla F(\pi)$ and $\pi$ is such that $p(\pi) = \p$).

More precisely, restricting our attention to $\pi + \mathcal{U}$, we will give an equivalence result and a formula linking the so restricted Hessian of $F$, with the $\mathcal{U}$-Hessian of $f$ at $\p$.

To prove our results, we introduce an intermediate function, similar to $\phi_V$ in §3.2 of [15], but adapted to our $\mathcal{U}$-context:

$$
(5.5) \quad \mathcal{U} \ni u \mapsto \phi_V(u) := \min_{v \in \mathcal{V}} \{ f(\p + u \oplus v) - (\g_V, v)_\mathcal{V} + \frac{1}{2} \|v\|^2_{\mathcal{V}} \}.
$$

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 $\p$ be in $\mathcal{U} \partial f(\p)$. Then

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

In particular,

$$
(5.6) \quad \nabla \phi_V(0) = \g_{\mathcal{U}} \quad \text{and} \quad \exists L_{\mathcal{U}}(0) \iff \exists \nabla L_{\mathcal{U}}(0).
$$

**Proof.** Clearly $\phi_V(u) \geq L_{\mathcal{U}}(u)$. To obtain an opposite inequality, write the minmand in (5.5) for $v = w \in W(u)$:

$$
\phi_V(u) \leq f(\p + u \oplus w) - (\g_V, w)_\mathcal{V} + \frac{1}{2} \|w\|^2_{\mathcal{V}} = L_{\mathcal{U}}(u) + \frac{1}{2} \|w\|^2_{\mathcal{V}}.
$$

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

The reason for introducing $\phi_V$ is that its Moreau-Yosida regularization $\Phi_V$ is obtained from the restriction $F_{\mathcal{U}}$ of $F$ to $\pi + \mathcal{U}$ by a mere translation.

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

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

satisfy

$$
F_{\mathcal{U}}(d_{\mathcal{U}}) = \Phi_V(\g_{\mathcal{U}} + d_{\mathcal{U}}) + \frac{1}{2} \|\g_V\|^2_{\mathcal{V}} \quad \text{for all } d_{\mathcal{U}} \in \mathcal{U}.
$$
Proof. Take \( d_U \in \mathcal{U} \). Recalling (5.4), compute \( F_{d_U}(d_U) = F(p + (\mathcal{g}_U + d_U) \oplus \mathcal{g}_V) \) in the following tricky way:

\[
F_{d_U}(d_U) = \min_{(u,v) \in (\mathcal{U} \times \mathcal{V})} \left\{ f(p + u + v) + \frac{1}{2} \|(\mathcal{g}_U + d_U - u) \oplus (\mathcal{g}_V - v)\|^2 \right\}
\]

hence \( H \)

Take \( w \in W(u) \) and subtract \( (\mathcal{g}_V, w)_V \) from both sides

\[
L_U(u) \geq L_U(0) + (\nabla L_U(0), u)_U + \frac{c}{2} \|u\|^2_{\mathcal{U}}
\]

hence \( H_{d_U}(p) = H_{d_U}(0) \) 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 $\nabla^2 F(\bar{x})$ exists, then $H_{UL}f(\bar{p})$ exists; $\nabla^2 F_U(0)$ is just the $UL$-block of $\nabla^2 F(\bar{x})$. Furthermore, $x \mapsto p(x)$ has at $\bar{x}$ a Jacobian of the form

$$Jp(\bar{x}) = I - \nabla^2 F(\bar{x}) = \begin{pmatrix} P & 0 \\ 0 & 0 \end{pmatrix}$$

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

$$P = (I - \nabla^2 F(\bar{x}))_{UL} = I_U - \nabla^2 F_U(0) = (H_{UL}f(\bar{p}) + I_U)^{-1}$$

is positive definite.

6. Conclusion

The distinctive difficulty of nonsmooth optimization is that the graph of $f$ near a minimum point $\bar{p}$ 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 $\bar{p}$; 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 $U$ of Definition 2.1.
- Then we have defined the gully-shaped valley, together with its parametrization by $u \in U$, 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 $L_U$. 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 $L_U$ (in $U$).
- 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 $U$-terminology is due to him.

References


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