Subdifferential enlargements and continuity properties of the $\mathcal{VU}$-decomposition in convex optimization

Shuai Liu, Claudia Sagastizábal and Mikhail Solodov

Abstract. We review the concept of $\mathcal{VU}$-decomposition of nonsmooth convex functions, which is closely related to the notion of partly smooth functions. As $\mathcal{VU}$-decomposition depends on the subdifferential at the given point, the associated objects lack suitable continuity properties (because the subdifferential lacks them), which poses an additional challenge to the already difficult task of constructing superlinearly convergent algorithms for nonsmooth optimization. We thus introduce certain ε-$\mathcal{VU}$-objects, based on an abstract enlargement of the subdifferential, which have better continuity properties. We note that the standard ε-sudifferential belongs to the introduced family of enlargements, but we argue that this is actually not the most appropriate choice from the algorithmic point of view. Specifically, strictly smaller enlargements are desirable, as well as enlargements tailored to specific structure of the function (when there is such structure). Various illustrative examples are given.

Mathematics Subject Classification (2000). Primary 49J52; Secondary 49J53, 65K10.

Keywords. Subdifferential, enlargements, VU-decomposition, sublinear functions, partial smooth.

1. Introduction

Designing superlinearly convergent algorithms for nonsmooth convex optimization has been an important challenge over the last thirty years or so [22]. The elusive fast convergence in nonsmooth settings remained out of reach until it was noticed

The first author was supported by CNPq grant 401119/2014-9. The second author was partially supported by CNPq Grant 303905/2015-8 and FAPERJ Grant 203.052/2016. The third author was supported in part by CNPq Grant 303724/2015-3 and by FAPERJ Grant 203.052/2016.

in [13] that for a nonsmooth convex function $f : \mathbb{R}^n \to \mathbb{R}$, second-order expansions do exist, when restricted to a certain special subspace. More precisely, near a given point $\bar{x}$, while the function has kinks on the subspace spanned by its subdifferential $\partial f(\bar{x})$, on the orthogonal complement of this subspace the graph of f appears "U"-shaped (in other words, smooth). Because this subspace concentrates the smoothness of f, it was called the $\mathcal{U}$-subspace in [14]. Likewise, its orthogonal complement reflects all the nonsmoothness of f about $\bar{x}$, and so it was called the $\mathcal{V}$-subspace, $\mathcal{V}\ (\bar{x}) := \mathrm{aff}(\partial f(\bar{x}) - g)$ where $g \in \partial f(\bar{x})$ and aff stands for the affine hull of a set.

From the algorithmic perspective, the importance of having a second-order expansion for f lies in the possibility of making a fast $\mathcal{U}$-Newton-move, thus enabling superlinear convergence. The $\mathcal{V}\,\mathcal{U}$-space decomposition gave rise to conceptual methods, that were made implementable for problems with good structural properties, such as max-eigenvalue minimization [27] and primal-dual gradient structured functions [23]. The ability of bundle methods to identify nonsmoothness structure for max-functions was demonstrated in [5]. An interesting connection with Sequential Quadratic Programming methods was revealed in [26]. The $\mathcal{VU}$-theory was applied in [21] to solve stochastic programming problems, and also in [11] and [12] to design fast algorithms for nonconvex maximum eigenvalue problems. More recently, for finite max-functions, [6] shows that it is possible to construct $\mathcal{VU}$-approximations in a derivative-free setting; see also [8].

The notion of partial smoothness, introduced in [15] for nonconvex nonsmooth functions, includes the $\mathcal{VU}$-decomposition concepts as a particular case. A partly smooth function appears as being smooth when moving along certain *manifold of activity*, whose tangent subspace is the $\mathcal{U}$-subspace. The question of finding the activity manifold is related to the problem of identifying active constraints in nonlinear programming [16], and has been explored in [7], [4], [19], [20].

Although partial smoothness and $\mathcal{VU}$-decomposition appear as very powerful theoretical tools, they have not yet been exploited up to their full potential in algorithmic developments. Recall that bundle methods (which are some of the most practical and efficient tools for the nonsmooth setting) are designed as black-box (first-order) methods, working with an *oracle* which provides functional values and only *one* subgradient for any given point. From a numerical perspective, properly identifying the $\mathcal{VU}$-subspaces (or the activity manifold) on the basis of such scarce/limited information is very hard. However, in our understanding, there is also another important reason for the difficulties, and it is inherent to the $\mathcal{VU}$-concepts (or partly smooth concepts) themselves. Specifically, at issue is the lack of suitable continuity of $\mathcal{VU}$-objects when they are seen as set-valued mappings of x. Roughly speaking, if the function happens to have a kink of nondifferentiability at one iterate x^k, the corresponding $\mathcal{V}^{\,k}$-subspace should have some positive dimension; whereas if at x^{k+1} the function is differentiable, the $\mathcal{V}^{\,k+1}$-subspace shrinks to zero. We emphasize that this may occur even for two iterates arbitrarily close to each other. Such oscillations are undesirable in any algorithm, as it becomes prone to erratic/unstable behaviour.

Motivated by the observations above, in this work we introduce ε-$\mathcal{VU}$-objects that have better continuity properties, when considered as multifunctions of x and ε. Note that a natural, and straightforward, proposal would be to consider the lineality space of the ε-subdifferential of f at x, $\mathcal{V}_{\varepsilon}(x) := \operatorname{aff}(\partial_\varepsilon f(\bar{x}) - g)$, where $g \in \partial_\varepsilon f(\bar{x})$. However, we shall argue that this may not necessarily be the most appropriate choice. To take advantage of possible structural properties of f, as well as to allow more options and flexibility, we shall introduce a certain abstract enlargement, denoted by $\delta_\varepsilon f(\bar{x})$, for which the ε-subdifferential is one but not the only possible choice. Among other things, we allow enlargements of $\partial f(\bar{x})$ which can be smaller than $\partial_\varepsilon f(\bar{x})$. We believe this can be desirable for a number of reasons, ranging from tighter approximation to easier computation when compared to the ε-subdifferential. Indeed, we actually prefer the enlargement $\delta_\varepsilon f(\bar{x})$ to be *strictly* contained in $\partial_\varepsilon f(\bar{x})$. This is because we often find $\partial_\varepsilon f(\bar{x})$ to be too big, causing $\mathcal{V}_\varepsilon(\bar{x})$ too "fat", and consequently $\mathcal{U}_\varepsilon(\bar{x})$ too "thin". We conclude this discussion by noting that there is another well-known candidate for enlargement of the subdifferential, namely the so-called ε-enlargement of the maximally monotone operator (the subdifferential is maximally monotone), see [1]. However, this enlargement is even larger than the ε-subdifferential.

This work is organized as follows. Section 2 gives an overview of important concepts regarding the $\mathcal{VU}$-theory and partial smoothness. Section 3 is devoted to the abstract enlargement of the subdifferential, and studies under which conditions the resulting ε-$\mathcal{VU}$-subspaces are outer and/or inner semicontinuous multifunctions of (x, ε). Section 4 illustrates how the abstract enlargement is more versatile than the ε-subdifferential, and how it can be used to fully exploit known structure of a function in a model example. This section finishes showing the impact that the chosen enlargement and the corresponding ε-$\mathcal{VU}$-subspaces have in an algorithmic framework. Sections 5 and 6 consider enlargements suitable for functions defined as the pointwise maximum of convex functions and sublinear functions, respectively. The work concludes with final remarks and comments on future research.

2. A fast overview of $\mathcal{VU}$-theory

We now recall the essential notions of the $\mathcal{VU}$-theory, which is at the heart of developing superlinearly convergent variants of bundle methods [23].

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

Given a convex function $f\colon \mathbb{R}^n \to \mathbb{R}$ and a point $\bar{x} \in \mathbb{R}^n$, let g be any subgradient in $\partial f(\bar{x})$. The $\mathcal{VU}$-decomposition [14] of $\mathbb{R}^n$ at $\bar{x}$ is defined by

$$\mathcal{V}(\bar{x}) = \operatorname{span}(\partial f(\bar{x}) - g) = \operatorname{aff}(\partial f(\bar{x})) - g, \quad \mathcal{U}(\bar{x}) = \mathcal{V}(\bar{x})^{\perp}. \qquad (2.1)$$

Given any $g^\circ \in \mathrm{ri}\,\partial f(\bar{x})$, the interior of $\partial f(\bar{x})$ relative to its affine hull, it is known that the subspaces in question can also be characterized as

$$\mathcal{U}(\bar{x}) = \{w \in \mathbb{R}^n \colon f'(\bar{x}; -w) = -f'(\bar{x}; w)\} = N_{\partial f(\bar{x})}(g^\circ),$$
$$\mathcal{V}(\bar{x}) = T_{\partial f(\bar{x})}(g^\circ),$$

where $f'(x; d)$ is the directional derivative of f at x in the direction d; $N_D(z)$ stands for the normal cone and $T_D(z)$ for the tangent cone, respectively, to the convex set D at $z \in D$.

As $\mathbb{R}^n = \mathcal{U}(\bar{x}) \oplus \mathcal{V}(\bar{x})$, each $x \in \mathbb{R}^n$ can be decomposed into $x = x_{\mathcal{U}(\bar{x})} \oplus x_{\mathcal{V}(\bar{x})}$. The $\mathcal{U}$-Lagrangian of f depends on the $\mathcal{V}(\bar{x})$-component $\bar{g}_{\mathcal{V}(\bar{x})}$ of a given subgradient $\bar{g} \in \partial f(\bar{x})$:

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

Each $\mathcal{U}$-Lagrangian is a proper closed convex function that is differentiable at $u = 0$ with $\mathcal{U}$-gradient given by $\nabla L_{\mathcal{U}}(0; \bar{g}_{\mathcal{V}(\bar{x})}) = \bar{g}_{\mathcal{U}(\bar{x})} = P_{\mathcal{U}(\bar{x})}(\partial f(\bar{x}))$, the projection of $\partial f(\bar{x})$ on $\mathcal{U}(\bar{x})$. When f has the primal-dual gradient structure at $\bar{x}$, $L_{\mathcal{U}}$ is twice continuously differentiable around 0 and f has a second-order expansion in u along the smooth trajectory corresponding to $\mathcal{U}$.

2.2. Fast tracks

A conceptual superlinearly convergent $\mathcal{VU}$-algorithm makes a minimizing step in the $\mathcal{V}$-subspace, followed by a $\mathcal{U}$-Newton step along the direction of the $\mathcal{U}$-gradient. To make this idea implementable, a fundamental result is [24, Thm 5.2]. It states that, at least locally, $\mathcal{V}$-steps can be replaced by proximal steps. Proximal steps, in turn, can be approximated (with any desired precision) by bundle methods.

Let $p(x)$ denote the proximal point of f at x, which also depends on a prox-parameter $\mu > 0$ (although not reflected in our notation). I.e., $p(x)$ is the unique minimizer of $f(y) + \mu/2|y - x|^2$. Then, by [25, Cor. 4.3], to compute the $\mathcal{U}$-gradient it suffices to find $s(x)$, the element of minimum norm in $\partial f(p(x))$. Under reasonable assumptions, the primal-dual pair $(p(x), s(x))$ is a *fast track* leading to $(\bar{x}, 0)$, a minimizer and the null subgradient.

Since bundle methods can be used to approximate proximal steps, see, e.g., [3], all the $\mathcal{VU}$-related quantities can be approximated by suitably combining the ingredients above. Specifically, let φ denote a convex piecewise linear model of f, available at the current iteration. Then the bundle method quadratic programming problem solution

$$\hat{d} := \arg\min_{d} \varphi(x + d) + \frac{1}{2}\mu|d|^2$$

yields an approximation $\hat{p} = x + \hat{d}$ for the proximal point $p(x)$. Once $\hat{p}$ is known, since φ is a simple max-function, the full subdifferential $\partial\varphi(\hat{p})$ is readily available. It is then possible to compute its null space, and its basis matrix U, to approximate

$\mathcal{U}(\hat{p})$. Also, it is possible to obtain an approximation for $s(x)$, by solving a second quadratic programming problem

$$\hat{s} := \arg\min_{s} \left\{ |s|^2 : s \in \partial\varphi(x+\hat{d}) \right\} .$$

The $\mathcal{U}$-step is given by $d_{\mathcal{U}} := UH^{-1}U^{\top}\hat{s}$, for a positive definite matrix H gathering second-order information of f about $\hat{p}$.

However, since we are dealing with approximations, the $\mathcal{V}$ and $\mathcal{U}$-steps yielding the next iterate, $x^{+} = \hat{p} - d_{\mathcal{U}}$, are done only when the approximation is deemed good enough; specifically, when

$$f(\hat{p}) - \varphi(\hat{p}) \leq \frac{m}{2\mu}|\hat{s}|^2 \tag{2.2}$$

for a given parameter $m \in (0,1)$.

The steps above are the main ingredients of the globally and locally superlinearly convergent $\mathcal{VU}$-algorithm developed in [23] (barring a simple line search that we omit here, to simplify the exposition).

2.3. Smooth manifolds and partial smoothness

A concept closely related to $\mathcal{VU}$-theory is the one of *partial smoothness*, relative to certain smooth manifold.

In [18], a set $\mathcal{M} \subset \mathbb{R}^n$ is said to be a $\mathcal{C}^k$-smooth manifold of codimension m around $\bar{x} \in \mathcal{M}$ if there is an open set $Q \subset \mathbb{R}^n$ such that

$$\mathcal{M} \cap Q = \{x \in Q \colon \phi_i(x) = 0,\ i = 1, \cdots, m\},$$

where ϕ_i are $\mathcal{C}^k$ functions (i.e., k times continuously differentiable) with the set $\{\nabla\phi_i(\bar{x}),\ i = 1, \ldots, m\}$ being linearly independent. The normal (orthogonal to tangent) subspace to $\mathcal{M}$ at $\bar{x}$ is given by

$$N_{\mathcal{M}}(\bar{x}) = \operatorname{span}\{\nabla\phi_i(\bar{x}),\ i = 1, \cdots, m\}.$$

A proper closed convex function f is said to be *partly smooth* at $\bar{x}$, relative to $\mathcal{M}$, a C^2-manifold around $\bar{x}$, if $\partial f(\bar{x}) \neq \emptyset$, and

1. Smoothness: f restricted to $\mathcal{M}$ is C^2 around $\bar{x}$;
2. Normals are parallel to subdifferential: $N_{\mathcal{M}}(\bar{x}) = \mathcal{V}(\bar{x})$;
3. Continuity: ∂f is continuous at $\bar{x}$ relative to $\mathcal{M}$.

The function f is said to be *partly smooth relative to the manifold* $\mathcal{M}$ if f is partly smooth at each point in $\mathcal{M}$, relative to $\mathcal{M}$.

In [15] it is shown that if a function is partly smooth relative to a manifold, then the manifold contains the primal component of the fast track (proposed in $\mathcal{VU}$-decomposition) is partly smooth at $\bar{x}$, relative to a manifold that can be defined from the functional structure. For example, given a nonempty finite index set I and any C^1-convex functions $f_i,\ i \in I$, the max-function

$$f(x) := \max_{i \in I} f_i(x)$$

is partly smooth at $\bar{x}$ relative to the manifold

$$\mathcal{M}_{\bar{x}} = \{x \colon I(x) = I(\bar{x})\} ,$$

where

$$I(x) := \{i \in I \colon f_i(x) = f(x)\} \tag{2.3}$$

is the activity set, provided the set of active gradients $\{\nabla f_i(\bar{x}) \colon i \in I(\bar{x})\}$ is linearly independent [15, Cor. 4.8].

Regarding the $\mathcal{VU}$-algorithm, the smooth manifold appears in connection with the sequence of $\mathcal{V}$-steps: finding an acceptable $\hat{p}$ can be seen as a succession of corrector steps, bringing the iterate x close to the smooth manifold. Pursuing further with this geometrical interpretation, once (2.2) holds, the $\mathcal{U}$-step represents a predictor step, taken along a direction tangent to the manifold. In this respect, approximating the $\mathcal{U}$-basis amounts to essentially approximating the smooth manifold.

3. Approximating the $\mathcal{V}$-subspace

We start with the following considerations. The superlinear convergence rate of the $\mathcal{VU}$-algorithm in [23] relies on the quality of the approximate $\mathcal{U}$-matrices, particularly with respect to their ability to asymptotically span the true $\mathcal{U}$ $(\bar{x})$-subspace.

Similarly to identification of active constraints in nonlinear programming, determining those bases can prove difficult. As mentioned, the $\mathcal{U}$-basis is related to the manifold of partial smoothness, which in turn depends on certain activity set. For the max-function example, in particular, this requires correctly identifying the activity index $I(\bar{x})$ (without knowing $\bar{x}$, of course). In fact, even for any given x, identifying *reliably* all the values in $\{f_1(x)\, , f_2(x)\, , f_3(x)\, , \ldots\, , f_m(x)\}$ that are *equal* is already tricky from the numerical point of view (what "equal" really means for two non-integer numbers in a computer?).

Another example of the aforementioned difficulty is given by the max-eigenvalue function, where one would need to find the exact multiplicity of the maximum eigenvalue of a matrix. Indeed, let $\mathcal{S}_n$ denote the Euclidean space of the n-by-n real symmetric matrices, endowed with the Frobenius norm and inner product. The eigenvalues of $X \in \mathcal{S}_n$ are denoted by $\lambda_1(X) \geq \lambda_2(X) \geq \cdots \geq \lambda_n(X)$ (listed in decreasing order by multiplicity) and $E_i(X)$ is the eigenspace associated with $\lambda_i(X)$. If $E_1(\bar{X})$ is m-dimensional, then $f(X) := \lambda_1(X)$ is partly smooth at $\bar{X}$, relative to the manifold (cf. [15, Ex. 3.6])

$$\mathcal{M}_{\bar{X}} = \{X \in \mathcal{S}_n \colon \lambda_1(X) \text{ has multiplicity } m\} \qquad (1 \leq m \leq n).$$

Exact identifications in this setting, within some realistic algorithmic process, are extremely hard.

As already mentioned above, in addition to possible difficulties with exact identifications, there is also the issue of (lack of) continuity of the resulting objects. We discuss this in more detail next.

3.1. Semicontinuity notions

In order to understand the phenomenon of instability, and to propose a solution, we need to analyze the continuity properties of the $\mathcal{VU}$-subspaces, when considered as a (multi)function of x. Recall that for a set-valued mapping $S\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$,

- S is inner semicontinuous (isc) at $\bar{x}$ if $\liminf_{x^k\to\bar{x}} S(x^k) \supset S(\bar{x})$, i.e.,

$$\begin{gathered}\text{given } \bar{s}\in S(\bar{x}) \text{ and any sequence } \{x^k\}\to\bar{x}\\ \text{there exists a selection } s: x\mapsto S(x) \text{ for all } x \text{ such that}\\ s(x^k)\to\bar{s} \text{ as } k\to\infty\,.\end{gathered}$$

- S is outer semicontinuous (osc) at $\bar{x}$ if $\bar{s} := \limsup_{x^k\to\bar{x}} S(x^k) \subset S(\bar{x})$, i.e.,

$$\begin{gathered}\text{given any sequence } \{x^k\}\to\bar{x}\\ \text{for each selection } \{s^k\in S(x^k)\}\to\bar{s} \text{ as } k\to\infty\\ \text{it holds that } \bar{s}\in S(\bar{x})\,.\end{gathered}$$

Proposition 3.1. *Consider a set-valued mapping $S\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and the multifunction resulting from applying the affine hull operation:*

$$A\colon x\mapsto \operatorname{aff} S(x) := \left\{\sum_{i=1}^{p}\lambda_i s_i : s_i\in S(x)\,, \sum_{i=1}^{p}\lambda_i = 1\,, \lambda_i\in\mathbb{R}\,, \textit{and } p\in\mathbb{N}\right\}.$$

If S is isc at $\bar{x}$, then so is A.

Proof. Any fixed $\bar{a}\in A(\bar{x})$ is of the form $\bar{a} = \sum_{i=1}^p \lambda_i\bar{s}_i$, with $\bar{s}_i\in S(\bar{x})\,, \sum_{i=1}^p\lambda_i = 1$, $\lambda_i\in\mathbb{R}$, and $p\in\mathbb{N}$. For each $i=1,\dots,p$, since $\bar{s}_i\in S(\bar{x})$, the isc of S at $\bar{x}$ ensures that for any sequence $x^k\to\bar{x}$ there exists a selection $s_i(x)$ such that $s_i^k = s_i(x^k)$ converges to $\bar{s}_i$. The isc of A follows, because the sequence $a^k := \sum_{i=1}^p \lambda_i s_i^k \in A(x^k)$ defines a selection converging to $\sum_{i=1}^p\lambda_i\bar{s}_i = \bar{a}$.

□

As explained below, for set-valued mappings, composing with the affine hull does not preserve outer semicontinuity.

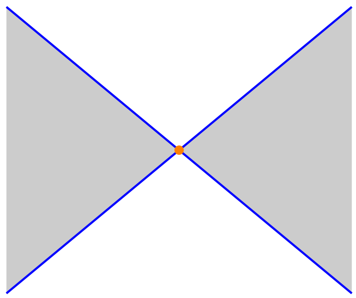
Remark 3.2 (On outer semicontinuity of affine hulls). A set-valued mapping $S\colon\mathbb{R}^n\rightrightarrows\mathbb{R}^m$ is osc everywhere if and only if its graph $\operatorname{gph} S := \{(x,u)\colon u\in S(x)\}$ is closed (cf. [29, Thm. 5.7]). However, given an osc mapping $S(x)$, its affine hull mapping $\operatorname{aff} S(x)$ does not necessarily have closed graph. For example, consider a mapping $S\colon\mathbb{R}\rightrightarrows\mathbb{R}$ with graph shown in Figure 1(a). Although it has a closed graph, the graph of $\operatorname{aff} S(x)$ is not closed.

Moreover, even an osc maximally monotone mapping does not necessarily have an osc affine hull mapping. Consider a function $f\colon[-1,1]\mapsto\mathbb{R}$

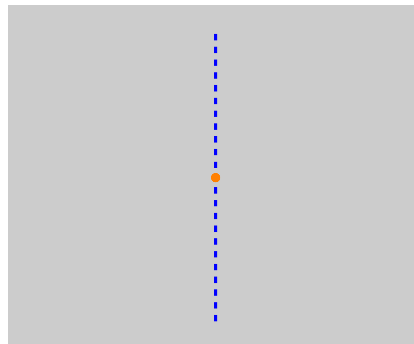
$$f(x) = \begin{cases}\frac{1}{k}x, & \text{if } x\in(\frac{1}{k+1},\frac{1}{k}] \text{ for all } k\in\mathbb{N},\\ 0, & \text{if } x = 0.\end{cases}$$

Define

$$S(x) := \partial f(x) = \begin{cases}\frac{1}{k}, & \text{if } x\in(\frac{1}{k+1},\frac{1}{k}),\\ \left[\frac{1}{k+1},\frac{1}{k}\right], & \text{if } x = \frac{1}{k+1}.\end{cases}$$



(A) The graph of $S(x)$



(B) The graph of aff $(S(x))$

FIGURE 1. $S(x)$ is osc everywhere but aff $(S(x))$ is not osc at one point.

Then $S(0) = \{0\}$, S is maximally monotone and

$$\operatorname{aff} S(x) = \begin{cases} \frac{1}{k}, & \text{if } x \in (\frac{1}{k+1}, \frac{1}{k}), \\ \mathbb{R}, & \text{if } x = \frac{1}{k+1}. \end{cases}$$

There exists a sequence $x^k := \frac{1}{k+1}$ such that as $k \to +\infty$, we have $x^k \to 0$ but $\operatorname{aff} S\left(x^k\right) \to \mathbb{R} \not\subset \operatorname{aff} S(0) = \{0\}$.

3.2. Subdifferential enlargements

The $\mathcal{V}$-subspace at x is defined by taking the affine hull of $(\partial f(x) - g)$, with any $g \in \partial f(x)$. By Prop. 3.1 and Rem. 3.2, the inner semicontinuity of the operator S is inherited by its affine hull, but not its outer semicontinuity. We face a paradoxical situation because, as a multifunction, the subdifferential is osc but not isc; cf. [10, p. VI.6.2]. Under these circumstances, the chances for the $\mathcal{V}$-subspace to enjoy some continuity property on x are slim. In Sec. 4 below we illustrate with a simple example that the lack of continuity of the concept can lead to drastic changes from one point to another: in that example, the $\mathcal{V}$-subspaces shrink from the whole space to the origin. The corresponding oscillations in the algorithmic process have the undesirable result of slowing down the convergence speed.

In order to stabilize the erratic behaviour of the $\mathcal{V}$-subspaces, we define *approximations*, denoted by $\mathcal{V}_\varepsilon$ below, that are given by taking the affine hull of a set-valued mapping with better continuity properties than the subdifferential.

Thanks to the "enlarging" parameter ε, the ε-subdifferential from Convex Analysis is both inner and outer semicontinuous as a set-valued mapping of $\varepsilon > 0$ and x [10, §XI.4.1]. One could then consider replacing $\partial f(x)$ by the ε-subdifferential in the definition of $\mathcal{V}\ (x)$ given in (2.1). However, for a number of reasons that will become clear in the subsequent sections, we shall introduce an *abstract* set-valued function (enlargement), denoted by $\delta_\varepsilon f(x)$. The abstract enlargement must satisfy a certain minimal set of conditions, which we identified as relevant for our purposes. These conditions are shown below to hold for the usual ε-subdifferential as well, i.e., the abstract concept includes this important enlargement as a particular case. But it also allows for other enlargements, in particular smaller than the ε-subdifferential, as well as those that explicitly use the structure of the function when it is available.

Given a convex function $f\colon \mathbb{R}^n \to \mathbb{R}$, and an abstract enlargement of $\partial f(x)$, the corresponding $\mathcal{V}_\varepsilon\mathcal{U}_\varepsilon$-decomposition is defined as follows:

$$\mathcal{V}_\varepsilon(x) := \operatorname{aff}\left(\delta_\varepsilon f(x) - s\right),\ s \in \delta_\varepsilon f(x), \qquad \mathcal{U}_\varepsilon(x) := \mathcal{V}_\varepsilon(x)^\perp. \tag{3.1}$$

We require $\delta_\varepsilon f(x)$ to be convex and closed, and to satisfy the following "sandwich" inclusions for all $\varepsilon \geq 0$:

$$\partial f(x) \subseteq \delta_\varepsilon f(x) \subseteq \partial_\varepsilon f(x)\,. \tag{3.2}$$

Proposition 3.3. *The following holds for the abstract enlargement $\delta_\varepsilon f(x)$ and the corresponding subspaces defined in* (3.1)*:*

(i) $\mathcal{U}_\varepsilon(x) = N_{\delta_\varepsilon f(x)}(s^\circ)$, *for all $s^\circ \in \operatorname{ri}\delta_\varepsilon f(x)$.*

(ii) *The $\mathcal{V}$-subspace is enlarged and the $\mathcal{U}$-subspace is shrunk: $\mathcal{V}(x) \subseteq \mathcal{V}_\varepsilon(x)$ and $\mathcal{U}(x) \supseteq \mathcal{U}_\varepsilon(x)$.*

(iii) *If the enlargement satisfies* (3.2)*, then it is an outer semicontinuous multifunction of x and ε and, in particular,*

$$\limsup_{\substack{x^k \to x\\ \varepsilon \to 0}} \delta_\varepsilon f(x^k) = \partial f(x)\,. \tag{3.3}$$

(iv) *If the enlargement is isc at $\bar{x}$ and $\bar{\varepsilon} > 0$, so is $\mathcal{V}\ _\varepsilon(x)$.*

(v) *If the enlargement satisfies* (3.2) *and is isc, then it is also continuous.*

(vi) *If the enlargement satisfies* (3.2)*, is isc, and*

$$\lim_{\substack{(x,\varepsilon)\to(\bar{x},\bar{\varepsilon})\\ \varepsilon\geq 0}} \mathcal{V}_\varepsilon(x) = \mathcal{V}_{\bar{\varepsilon}}(\bar{x}),$$

then

$$\lim_{\substack{(x,\varepsilon)\to(\bar{x},\bar{\varepsilon})\\ \varepsilon\geq 0}} \mathcal{U}_\varepsilon(x) = \mathcal{U}_{\bar{\varepsilon}}(\bar{x}).$$

Proof. Item (i) is obtained by using the same arguments as those in [14, Proposition 2.2]. Item (ii) is straightforward from (3.2), whose inclusions are preserved when taking the affine hull of the sets, which yield the corresponding $\mathcal{V}$ spaces.
Regarding item (iii), both ∂f and $(x, \varepsilon) \mapsto \partial_\varepsilon f(x)$ are outer semicontinuous multifunctions. In view of (3.2), $(x, \varepsilon) \mapsto \delta_\varepsilon f(x)$ is also osc and (3.3) holds.
Item (iv) is just Prop. 3.1, written with $S(x) = \delta_\varepsilon f(x)$, while item (v) follows from item (iii).
To show item (vi), let $\bar{s}$ be an arbitrary element in $\delta_\varepsilon f\,(\bar{x})$ and $s_i,\ i = 1, \cdots, m$ be all the elements in $\delta_\varepsilon f\,(\bar{x})$ such that $\{s_i - \bar{s}\}_{i=1}^m$ is a set that contains the maximal number of linearly independent vectors. Define the $n \times n$ matrix

$$A_x^\varepsilon := [s_1 - \bar{s}, s_2 - \bar{x}, \cdots, 0_{n\times(n-m)}].$$

By definition, $\mathcal{V}_\varepsilon(x) = \{\sum_{i=1}^m \alpha_i(s_i - \bar{s}) \colon \alpha_i \in \mathbb{R}, i = 1, \cdots, m\} = \operatorname{Range}(A_x^\varepsilon)$ and then $\mathcal{U}_\varepsilon(x) = \ker(A_x^\varepsilon)$. The convergence of $\mathcal{V}_\varepsilon(x)$ corresponds to the convergence of A_x^ε. Applying [29, Thm. 4.32], we have that the kernel of A_x^ε, i.e., $\mathcal{U}_\varepsilon(x)$, also converges. □

Item (iii) in the above proposition shows that outer semicontinuity of the relaxed $\mathcal{V}$-space can be obtained if the sandwich inclusion (3.2) holds. The enlargements proposed for several classes of functions in the next sections will all satisfy that important relation.

4. Impact of the chosen enlargement

We now illustrate the different choices that can be made for the abstract enlargement using the following model function:

$$h(x_1, x_2) = h_1(x_1) + h_2(x_2), \quad \text{where} \quad \begin{cases} h_1(x_1) = |x_1| \\ h_2(x_2) = \frac{1}{2}x_2^2\,. \end{cases} \tag{4.1}$$

Note that h is separable, h_1 is nonsmooth, and h_2 is smooth. This function is the simplest instance of the "half-and-half" functions created by Lewis and Overton to analyze BFGS behavior in a nonsmooth setting; [17, Sec. 5.5].

4.1. The ε-subdifferential enlargement

The (smoothness and) second-order of the function h is all concentrated in the x_2-component, where the function has constant curvature, equal to 1. The subdifferential is easy to compute:

$$\partial h(x) = \partial h_1(x_1) \times \{x_2\} \qquad \text{for } \partial h_1(x_1) = \begin{cases} \{-1\} & \text{if } x_1 < 0 \\ [-1, 1] & \text{if } x_1 = 0, \\ \{1\} & \text{if } x_1 > 0\,. \end{cases}$$

As a result, the $\mathcal{V}$ and $\mathcal{U}$ subspaces are

$$\begin{array}{llll} \text{if } x_1 \neq 0 \text{ then} & \mathcal{V}(x_1, x_2) = \{(0,0)\} & \text{and} & \mathcal{U}(x_1, x_2) = \mathbb{R}^2 \\ \text{while} & \mathcal{V}(0,0) = \mathbb{R} \times \{0\} & \text{and} & \mathcal{U}(0,0) = \{0\} \times \mathbb{R}\,. \end{array} \tag{4.2}$$

To compute the ε-subdifferential, first combine Remark 3.1.5 and Example 1.2.2 from [10, Ch. XI] to obtain that

$$\begin{aligned} \partial_\varepsilon h(x) &= \bigcup_{\varepsilon_1 \in [0,\varepsilon]} \partial_{\varepsilon_1} h_1(x_1) \times \partial_{\varepsilon - \varepsilon_1} h_2(x_2) \\ &= \bigcup_{\varepsilon_1 \in [0,\varepsilon]} \partial_{\varepsilon_1} h_1(x_1) \times \left\{ x_2 + \xi : \frac{1}{2}\xi^2 \leq \varepsilon - \varepsilon_1 \right\}, \end{aligned} \tag{4.3}$$

and use the well-known formula below

$$\partial_{\varepsilon_1} h_1(x_1) = \begin{cases} [-1, -1 - \varepsilon_1/x_1] & \text{if } x_1 < -\varepsilon_1/2, \\ [-1, 1] & \text{if } -\varepsilon_1/2 \le x_1 \le \varepsilon_1/2, \\ [1 - \varepsilon_1/x_1, 1] & \text{if } x_1 > \varepsilon_1/2. \end{cases} \tag{4.4}$$

The expression (4.3)-(4.4) puts in evidence the following drawback of the ε-subdifferential in our setting: it enlarges $\partial h(x)$ also at points where h is actually smooth. Such a "fattening" is detrimental/harmful from the $\mathcal{VU}$-decomposition perspective: if in our example we were to take $\delta_\varepsilon h(x) = \partial_\varepsilon h(x)$, we would end up with the following, extreme, decomposition in (3.1):

$$\mathcal{V}_\varepsilon(x) = \mathbb{R}^2 \quad \text{and} \quad \mathcal{U}_\varepsilon(x) = \{0\}\,,$$

for all x, which would make it impossible to make the desirable Newton-like $\mathcal{U}$-steps.

4.2. A separable enlargement

Since often we only need enlargements near kinks, we can use the knowledge that $h = h_1 + h_2$, with h_2 being smooth. This yields a more suitable enlargement, corresponding to the particular set resulting from taking $\varepsilon_1 = \varepsilon$ in the union in (4.3):

$$\delta_\varepsilon h(x_1, x_2) := \partial_\varepsilon h_1(x_1) \times \{\nabla h_2(x_2)\}\ . \tag{4.5}$$

In view of (4.2), condition (3.2) holds, and the enlargement is osc, by item (iii) in Prop. 3.3. Moreover, since (3.2) holds and $\delta_\varepsilon h$ has one component that is a continuous enlargement, and the other component is single-valued, the enlargement is continuous, by (v) in Prop. 3.3. Finally, item (vi) also holds, since for all $x \in \mathbb{R}^2$,

$$\mathcal{V}_\varepsilon(x) = \operatorname{aff}\Big(\delta_\varepsilon h(x) - g_0\Big) = \mathbb{R} \times \{0\} \quad \text{and, hence,} \quad \mathcal{U}_\varepsilon(x) = \{0\} \times \mathbb{R}\,. \tag{4.6}$$

When compared with the original (exact) $\mathcal{VU}$-decomposition in (4.2), we have gained not only stability in the decomposition (the subspaces are the same for all x), but also continuity in the $\mathcal{V}_\varepsilon$-subspace.

4.3. A not-so-large enlargement

A small enlargement makes the $\mathcal{U}_\varepsilon$-subspace larger, and this somehow favours a faster speed of convergence, via the $\mathcal{U}$-steps. Since considering the large set $\partial_\varepsilon h(x)$ also led to an unsuitable space decomposition, looking for options that are smaller may be of interest.

We analyze a choice of this type for our example, noting that in Sec 6 we shall consider a general definition for this set, suitable for any sublinear function, not only for $h_1(x_1) = |x_1|$.

As a geometrical motivation to our approach, consider the enlargements of ∂h_1 given in Figure 2. For fixed ε, the graph of the ε-subdifferential (4.4) is shown on the left.

The simpler (polyhedral) multifunction shown on the right is the graph of $\delta_\varepsilon h_1(x)$. We observe that the enlargement $\delta_\varepsilon h_1$ remains a singleton at points far

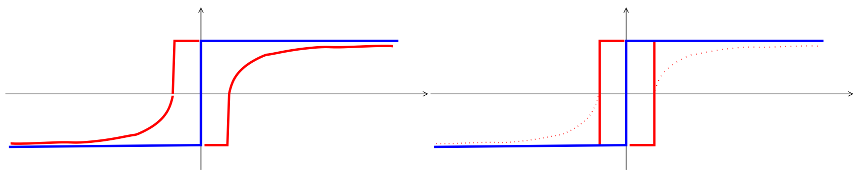


FIGURE 2. Two enlargements for $\partial|x|$

from the origin. This is consistent with the fact that the function h_1 is smooth in that region. By contrast, the ε-subdifferential expands ∂h_1 at all the points where h_1 is differentiable.

To see the impact of this difference in the resulting $\mathcal{V}$-approximations, consider the enlargement

$$\delta_\varepsilon h(x_1, x_2) := \delta_\varepsilon h_1(x_1) \times \{\nabla h_2(x_2)\},$$

where we define

$$\begin{aligned}\delta_\varepsilon h_1(x_1) &:= \overline{\mathrm{co}}\,\{s \in \mathrm{ext}\,\partial h_1(0)\colon\ \langle s, x_1\rangle \geq h_1(x_1) - \varepsilon\} \\ &= \begin{cases} \{-1\} & \text{if } x_1 < -\varepsilon/2, \\ [-1,1] & \text{if } -\varepsilon/2 \leq x_1 \leq \varepsilon/2, \\ \{1\} & \text{if } x_1 > \varepsilon/2. \end{cases}\end{aligned} \tag{4.7}$$

In the above, $\mathrm{ext}\, S$ stands for the extreme points of the convex set S, $\mathrm{co}\, D$ denotes the convex hull of the set D, and $\overline{S}$ is the closure of the set S.

From (4.7), (4.4) and (4.3) we see that $\delta_\epsilon h(x_1, x_2)$ satisfies condition (3.2). Hence, by Prop. 3.3, it is osc. The enlargement is isc everywhere except when $|x_1| = \frac{\varepsilon}{2}$.

The resulting space decomposition is

$$\mathcal{V}_\varepsilon(x) = \begin{cases} \mathbb{R} \times \{0\}, \\ \{(0,0)\}, \end{cases} \quad \text{and } \mathcal{U}_\varepsilon(x) = \begin{cases} \{0\} \times \mathbb{R}, & \text{if } -\varepsilon/2 \leq x_1 \leq \varepsilon/2, \\ \mathbb{R}^2, & \text{otherwise}. \end{cases} \tag{4.8}$$

We can verify that $\mathcal{V}_\epsilon(x)$ and $\mathcal{U}_\epsilon(x)$ are continuous everywhere except when $|x_1| = \frac{\varepsilon}{2}$. With the separable enlargement, $\mathcal{V}_\varepsilon$ is always $\mathbb{R} \times 0$. By contrast, with the smaller enlargement, $\mathcal{V}_\varepsilon$ is maintained as the null subspace at points not close to the origin. From the $\mathcal{V}\,\mathcal{U}$-optimization point of view, this enlargement appears to be the best one.

4.4. The enlargement as a stabilization device of the $\mathcal{V}\mathcal{U}$-scheme

For our simple example and the enlargement given by (4.5), we next show how the approximated subspaces and the level of information that is made available by the oracle impacts on the calculation of primal and dual iterates $(\hat{p}, \hat{s})$ approximating the fast track pair $(p(x), s(x))$, as well as on the determination of the $\mathcal{U}$-subspace spanning matrices.

4.4.1. Exact $\mathcal{V}\,\mathcal{U}$-approach. Suppose the oracle delivers the full subdifferential at h_1 at any given point. In this case the $\mathcal{V}$-step can be computed exactly, componentwise:

$$p_1(x_1) = \begin{cases} x_1 + \dfrac{1}{\mu} & \text{if } x_1 < -\dfrac{1}{\mu} \\ 0 & \text{if } -\dfrac{1}{\mu} \leq x_1 \leq \dfrac{1}{\mu} \\ x_1 - \dfrac{1}{\mu} & \text{if } x_1 > \dfrac{1}{\mu} \end{cases} \quad \text{and} \quad p_2(x_2) = \frac{\mu}{1+\mu} x_2 \,. \tag{4.9}$$

The shortest element in the subdifferential at $p(x)$ is also straightforward:

$$s_1(x_1) = \begin{cases} -1 & \text{if } x_1 < -\dfrac{1}{\mu} \\ 0 & \text{if } -\dfrac{1}{\mu} \leq x_1 \leq \dfrac{1}{\mu} \\ 1 & \text{if } x_1 > \dfrac{1}{\mu} \end{cases} \quad \text{and} \quad s_2(x_2) = p_2(x) \,.$$

The matrices spanning the respective $\mathcal{U}(p(x))$-subspaces are

$$U = [\,0 \quad 1\,]^\top \text{ and } U = I, \text{ the } 2{\times}2 \text{ identity matrix.}$$

Finally, take the "exact Hessian" is given by

$$H = \begin{cases} 1 & \text{when } \mathcal{V}(p(x)) = \mathbb{R} \times \{0\} \\ \begin{bmatrix} \sigma & 0 \\ 0 & 1 \end{bmatrix} & \text{when } \mathcal{V}(p(x)) = \{(0,0)\} \,. \end{cases}$$

The (small) parameter $\sigma > 0$ above is introduced to ensure positive-definiteness in the Newton system giving the $\mathcal{U}$-direction that corrects the $\mathcal{V}$-step; recall that $Hd = -U^\top \hat{s}$. Since the next iterate is $x_1^+ = \hat{p} + Ud$, it follows that

$$x_1^+ = \begin{cases} x_1 + \dfrac{1}{\mu} + \dfrac{1}{\sigma} & \text{if } x_1 < -\dfrac{1}{\mu} \\ 0 & \text{if } -\dfrac{1}{\mu} \leq x_1 \leq \dfrac{1}{\mu} \\ x_1 - \dfrac{1}{\mu} - \dfrac{1}{\sigma} & \text{if } x_1 > \dfrac{1}{\mu} \end{cases} \quad \text{and} \quad x_2^+ = 0 \,.$$

The problem under consideration is $\min h(x)$, which is solved by $\bar{x} = (0,0)$. The exact $\mathcal{V}\,\mathcal{U}$-scheme above finds $\bar{x}_2$ in one iteration for any initial point x_2. The first component, $\bar{x}_1$, would be found in one iteration if x_1 was taken sufficiently close to the solution ($|x_1| \leq 1/\mu$). Otherwise, assuming μ is kept fixed along iterations, convergence speed is driven by the choice of σ. This is an artificial parameter, which should be zero to reflect the second order structure of h, but was set to a positive (preferably small) value to make H positive definite. For instance, when $x_1 < -1/\mu$, if σ is such that $-2/\mu \leq x_1 + 1/\sigma \leq 0$, then $x_1^+ \in [-1/\mu, 1/\mu]$ and $x_1^{++} = 0$. However, if σ is too small and $x_1 + 1/\sigma > 0$, this makes $x_1^+ > 1/\mu$ and $x_1^{++} = x_1$. This undesirable oscillation comes from the lack of continuity of the $\mathcal{VU}$-subspaces with respect to the variable x. The ε-objects are meant to help in this sense, as shown next.

4.4.2. Exact $\mathcal{V}$-step with $\mathcal{U}_\varepsilon$-step. For less simple functions, we do not have access to the full subdifferential of h at all points. Hence, neither $p(x)$ nor $s(x)$, or any of the objects defined at each iteration are computable in an exact manner. We introduce gradually approximations in those objects, in a manner that reveals the utility of the proposed enlargements.

Suppose we do know the full enlargement $\delta_\varepsilon h(p(x))$ and also $p(x)$, but cannot compute explicitly $s(x)$.

Since $\hat{p} = p(x)$ amounts to using h as its model φ to estimate the $\mathcal{V}$-step, the relation in (2.2) holds for any choice of $\hat{s}$. In particular, if we project 0 onto the enlargement $\delta_\varepsilon h(p(x))$ to compute $\hat{s}$. Plugging the prox-expression (4.9) in (4.5), we see that

$$\hat{s}_1(x_1) = \begin{cases} -1 - \dfrac{\varepsilon\mu}{\mu x_1 + 1} & \text{if } x_1 < -\dfrac{1}{\mu} - \varepsilon \\ 0 & \text{if } -\dfrac{1}{\mu} \le x_1 \le \dfrac{1}{\mu} + \varepsilon \\ 1 - \dfrac{\varepsilon\mu}{\mu x_1 - 1} & \text{if } x_1 > \dfrac{1}{\mu} + \varepsilon \end{cases}$$

and $\hat{s}_2(x_2) = p_2(x_2) = \dfrac{\mu}{1+\mu} x_2$.

Because the $\mathcal{V}_\varepsilon\mathcal{U}_\varepsilon$-subspaces in (4.6) are "stable" with respect to x, we can take the same matrix $H = 1$ and $U = [\,0 \quad 1\,]^\top$ for all x_1 (without distinguishing the three cases, as above).

The $\mathcal{U}$-step uses the direction $d_1 = 0$ and $d_2 = p_2(x_2)$, which gives the update

$$x^+ = p(x) - \begin{pmatrix} 0 \\ p_2(x_2) \end{pmatrix} = \begin{pmatrix} p_1(x_1) \\ 0 \end{pmatrix}.$$

As before, $\bar{x}_2$ is found in one iteration. Regarding $\bar{x}_1$, the next iteration is

$$x_1^+ = \begin{cases} x_1 + \dfrac{1}{\mu} & \text{if } x_1 < -\dfrac{1}{\mu} \\ 0 & \text{if } -\dfrac{1}{\mu} \le x_1 \le \dfrac{1}{\mu} \\ x_1 - \dfrac{1}{\mu} & \text{if } x_1 > \dfrac{1}{\mu}\,. \end{cases}$$

If μ is kept fixed, it takes (integer value of) $|\mu x_1|$ iterations to find $\bar{x}_1$. This termination result is consistent with the $\mathcal{VU}$-superlinear convergence feature, since the result stating that proximal points are on the fast track requires that $\mu(x-\bar{x}) \to 0$ as $x \to \bar{x}$.

4.4.3. Approximating both steps. Suppose now the $\mathcal{V}$-step yielding $p_1(x_1)$ is not computable because we cannot solve exactly the (implicit) inclusion

$$\mu(x_1 - p_1(x_1)) \in \partial h_1(p_1(x_1))\,.$$

Assuming that we do have access to the full ε-subdifferential, we consider two options, both using the ε-subdifferential enlargement as a replacement of the subdifferential.

Implicit $\mathcal{V}_\varepsilon$ step. In this first option, we solve the implicit inclusion

$$\mu(x_1 - \hat{p}_1) \in \delta_\varepsilon h_1(\hat{p}_1)\,.$$

Through some algebraic calculations, we get

$$\hat{p}_1 \in \begin{cases} \left[\frac{1}{2}\left(x_1 + \frac{1}{\mu} - \sqrt{\left(x_1 + \frac{1}{\mu}\right)^2 + \frac{4\varepsilon}{\mu}}\right), x_1 + \frac{1}{\mu}\right], & \text{if } x_1 \le \frac{1}{\mu} - \frac{\varepsilon}{2} \\ \left[x_1 - \frac{1}{\mu}, x_1 + \frac{1}{\mu}\right], & \text{if } \frac{1}{\mu} - \frac{\varepsilon}{2} \le x_1 \le \frac{\varepsilon}{2} - \frac{1}{\mu} \\ \left[x_1 - \frac{1}{\mu}, \frac{1}{2}\left(x_1 - \frac{1}{\mu} + \sqrt{\left(x_1 - \frac{1}{\mu}\right)^2 + \frac{4\varepsilon}{\mu}}\right)\right], & \text{if } x_1 \ge \frac{\varepsilon}{2} - \frac{1}{\mu} \end{cases} \tag{4.10}$$

if $\frac{\varepsilon}{2} \ge \frac{1}{\mu}$; and otherwise,

$$\hat{p}_1 = \begin{cases} \left[\frac{1}{2}\left(x_1 + \frac{1}{\mu} - \sqrt{\left(x_1 + \frac{1}{\mu}\right)^2 + \frac{4\varepsilon}{\mu}}\right), x_1 + \frac{1}{\mu}\right], & \text{if } x_1 \le \frac{\varepsilon}{2} - \frac{1}{\mu} \\ [B_1, B_2]\,, & \text{if } \frac{\varepsilon}{2} - \frac{1}{\mu} \le x_1 \le \frac{1}{\mu} - \frac{\varepsilon}{2} \\ \left[x_1 - \frac{1}{\mu}, \frac{1}{2}\left(x_1 - \frac{1}{\mu} + \sqrt{\left(x_1 - \frac{1}{\mu}\right)^2 + \frac{4\varepsilon}{\mu}}\right)\right], & \text{if } x_1 \ge \frac{1}{\mu} - \frac{\varepsilon}{2} \end{cases} \tag{4.11}$$

where $B_1 = \frac{1}{2}\left(x_1 + \frac{1}{\mu} - \sqrt{\left(x_1 + \frac{1}{\mu}\right)^2 + \frac{4\varepsilon}{\mu}}\right)$ and
$B_2 = \frac{1}{2}\left(x_1 - \frac{1}{\mu} + \sqrt{\left(x_1 - \frac{1}{\mu}\right)^2 + \frac{4\varepsilon}{\mu}}\right)$.

By the expressions in (4.10) and (4.11), the solution $\hat{p}_1$ is not unique but an interval. For this reason, we need to append the inclusion with a *selection mechanism*, which takes $\hat{p}_1$ as the closest solution to x_1. The motivation behind this rule is to approximately find the fixed point of the proximal mapping. The unique solution for both (4.10) and (4.11) is:

$$\hat{p}_1 = \begin{cases} \frac{1}{2}\left(x_1 + \frac{1}{\mu} - \sqrt{\left(x_1 + \frac{1}{\mu}\right)^2 + \frac{4\varepsilon}{\mu}}\right), & \text{if } x_1 \le -\varepsilon \\ x_1, & \text{if } -\varepsilon \le x_1 \le \varepsilon \\ \frac{1}{2}\left(x_1 - \frac{1}{\mu} + \sqrt{\left(x_1 - \frac{1}{\mu}\right)^2 + \frac{4\varepsilon}{\mu}}\right), & \text{if } x_1 \ge \varepsilon \end{cases} \tag{4.12}$$

Explicit $\mathcal{V}_{\varepsilon}$ step. We can also solve the explicit inclusion

$$\mu(x_1 - \hat{p}_1) \in \delta_\varepsilon h_1(x_1),$$

to obtain

$$\hat{p}_1 = \begin{cases} \left[x_1 + \frac{1}{\mu} + \frac{\varepsilon}{x\mu}, x_1 + \frac{1}{\mu}\right], & \text{if } x_1 \leq -\frac{\varepsilon}{2} \\ \left[x_1 - \frac{1}{\mu}, x_1 + \frac{1}{\mu}\right], & \text{if } -\frac{\varepsilon}{2} \leq x_1 \leq \frac{\varepsilon}{2} \\ \left[x_1 - \frac{1}{\mu}, x_1 - \frac{1}{\mu} + \frac{\varepsilon}{x_1\mu}\right], & \text{if } x_1 \geq \frac{\varepsilon}{2} \end{cases} \tag{4.13}$$

and define the resulting update x^+. Once again, we see from (4.13) that the solution $\hat{p}_1$ is not unique but an interval. Applying the same selection mechanism,

$$\hat{p}_1 = \begin{cases} x_1 + \frac{1}{\mu} + \frac{\varepsilon}{x_1\mu}, & \text{if } x_1 \leq -\varepsilon \\ x_1, & \text{if } -\varepsilon \leq x_1 \leq \varepsilon \\ x_1 - \frac{1}{\mu} + \frac{\varepsilon}{x_1\mu}, & \text{if } x_1 \geq \varepsilon \end{cases} \tag{4.14}$$

Iteration update for all the cases. The approximation of the Newton direction is the same as in Section 4.4.2, i.e., $d = \begin{pmatrix} 0 \\ p_2(x_2) \end{pmatrix}$. Consequently, in all cases we have the same update $x_1^+ = p_1(x_1)$ and $x_2^+ = 0$. Consider the expressions of $\hat{p}_1$ in (4.12) and (4.14). Simple algebraic calculations can verify that $\hat{p}_1$ becomes closer to 0 compared with x_1 except in two cases where $\hat{p}_1 = x_1$ when $|x| \leq \varepsilon$. When this happens, we can decrease ε in our algorithm so that the sequence of updates x_1^+ eventually converges to 0.

For several classes of functions with special structure we now consider different subdifferential enlargements and the corresponding $\mathcal{V}_\varepsilon\mathcal{U}_\varepsilon$ objects.

5. Maximum of convex functions

When defining our enlargements, a very important principle we follow is that their computation must also be practically less difficult than computing the ε-subdifferential. Furthermore, we prefer the enlargement $\delta_\varepsilon f$ to be strictly contained in $\partial_\varepsilon f$ because often we find $\partial_\varepsilon f$ is too big, causing $\mathcal{V}_\varepsilon$ too "fat" and then $\mathcal{U}_\varepsilon$ too "thin".

5.1. ε-activity sets

We start with a result examining the persistence of the set of active indices (2.3), which characterizes the manifold of partial smoothness of a max-function.

We use the notation $\mathbb{I}_S$ for the indicator function of the set S, i.e., $\mathbb{I}_S(x) = 0$ if $x \in S$ and it is $+\infty$ otherwise.

Proposition 5.1. *Consider a max-function of the form* $f(x) := \max_{i\in I} f_i(x) + \mathbb{I}_C(x)$ *where* I *is a non-empty finite index set,* C *is a subset of* $\mathbb{R}^n$ *and* $f_i \colon \mathbb{R}^n \to \mathbb{R}$ *is continuous for all* $i \in I$.

For $\varepsilon \geq 0$, *let*

$$I_\varepsilon(x) := \{i \in I \colon f_i(x) \geq f(x) - \varepsilon\} \tag{5.1}$$

and denote $I(x) := I_0(x)$.

The following holds.

(i) *For any* $\bar{x} \in C$, *there exists* $\bar{\varepsilon} \geq 0$ *such that* $I_\varepsilon(x) \subset I(\bar{x})$ *for all* $x \in B(\bar{x}, \bar{\varepsilon})$ *and* $\varepsilon \in [0, \bar{\varepsilon}]$.

(ii) *If, in particular,* $\bar{x} \in \operatorname{int} C$ *and* $\varepsilon \neq 0$, *then the inclusion holds as an equality.*

Proof. Given $\bar{x} \in C$, to show item (i) we prove that there exists $r \geq 0$ such that $I_\varepsilon(x) \subset I(\bar{x})$ for all $x \in B(\bar{x}, r)$ and $\varepsilon \in [0, r]$. For any $i \in I \setminus I(\bar{x})$, the number $c_i := f(\bar{x}) - f_i(\bar{x})$ is positive as f is finite at $\bar{x}$. Define $g_i(x) := \max_{i\in I} f_i(x) - f_i(x) - \frac{c_i}{2}$. Then $g_i(\bar{x}) = \max_{i\in I} f_i(\bar{x}) - f_i(\bar{x}) - \frac{c_i}{2} > 0$. The continuity of g_i yields a neighborhood $B(\bar{x}, \gamma_i)$ such that $g_i(x) > 0$ for any $x \in B(\bar{x}, \gamma_i)$. Consequently, $f_i(x) < \max_{i\in I} f_i(x) - \frac{c_i}{2} \leq f(x) - \frac{c_i}{2}$ and $i \in I \setminus I_{\frac{c_i}{2}}(x)$. Take $r := \min_{i \in I\setminus I(\bar{x})} \left\{\frac{c_i}{2}, \gamma_i\right\}$, then $i \in I \setminus I_\varepsilon(x)$ for all $(x, \varepsilon) \in B(\bar{x}, r) \times [0, r]$. Therefore, $I_\varepsilon(x) \subset I(\bar{x})$ for all $x \in B(\bar{x}, r)$ and $\varepsilon \in [0, r]$, as stated.
Regarding item (ii), for any $\bar{x} \in \operatorname{int} C$, there exists $\gamma > 0$ such that $B(\bar{x}, \gamma) \subset C$. Suppose for contradiction that for any $t \in (0, +\infty)$ there exist $w \in B(\bar{x}, t)$ and $\varepsilon \in (0, t)$ such that $I_\varepsilon(w) \neq I(\bar{x})$. Let $\beta := \min\{\gamma, r\}$, $y \in B(\bar{x}, \beta)$ and $\varepsilon \in (0, \beta)$ such that $I_\varepsilon(y) \neq I(\bar{x})$. Then it can only be that $I_\varepsilon(y) \subsetneqq I(\bar{x})$. Let j be an index in $I(\bar{x}) \setminus I_\varepsilon(y)$. Then $f_j(y) < f(y) - \varepsilon = \max_{i\in I} f_i(y) - \varepsilon$. By continuity, there exists $B(y, \alpha)$ such that $f_j(x) < \max_{i\in I} f_i(y) - \varepsilon$ for all $x \in B(y, \alpha)$. It then follows that $f_j(x) < f(x) - \varepsilon$ for all $x \in B(y, \alpha) \cap B(\bar{x}, \beta)$. And hence $j \notin I_\varepsilon(x)$. Consider the point $w \in B(y, \alpha)$ which is on the line segment $\{\lambda\bar{x} + (1-\lambda)y \colon \lambda \in [0,1]\}$ and has the shortest distance to $\bar{x}$. By continuity there exists another neighborhood $B(w, p)$ such that $j \notin I_\varepsilon(x)$ for all $x \in B(w, p)$. This process can be done repetitively in a finite number of times until the existence of a ball containing $\bar{x}$ such that $j \notin I_\varepsilon(\bar{x})$. This is impossible because we have $j \in I(\bar{x}) \subset I_\varepsilon(\bar{x})$. □

The relaxed activity sets are used below to define suitable enlargements, first for polyhedral functions, and afterwards for general max-functions.

5.2. Polyhedral functions: enlargement and $\mathcal{V}_\varepsilon\mathcal{U}_\varepsilon$-subspaces

We start our development with the important class of polyhedral functions $f \colon \mathbb{R}^n \mapsto \overline{\mathbb{R}}$:

$$f(x) = \mathfrak{f}(x) + \mathbb{I}_{\mathfrak{D}}(x), \text{ for } \left\{ \begin{array}{lcl} \mathfrak{f}(x) & := & \max_{i\in I}\left\{\left\langle a^i, x\right\rangle + b^i\right\} \\ \mathfrak{D} & := & \left\{x \in \mathbb{R}^n \colon \left\langle c^j, x\right\rangle \leq d^j,\ j \in J\right\} \end{array} \right., \tag{5.2}$$

where the index sets are finite with $I \neq \emptyset$. Because the function $\mathfrak{f}(x)$ has full domain, $\partial f(x) = \partial \mathfrak{f}(x) + N_{\mathfrak{D}}(x)$ and, hence,

$$\partial f(x) = \left\{ \sum_{i \in I(x)} \alpha_i a^i + \sum_{j \in J(x)} \beta_j c^j \;\middle|\; \begin{array}{c} \displaystyle\sum_{i \in I(x)} \alpha_i = 1, \\ \alpha_i \geq 0 \;\; (i \in I(x)), \\ \beta_j \geq 0 \;\; (j \in J(x)) \end{array} \right\}. \tag{5.3}$$

It can be seen in [15, Ex. 3.4] that

$$\mathcal{V}(x) = \left\{ \sum_{i \in I(x)} \alpha_i a^i + \sum_{j \in J(x)} \beta_j c^j : \sum_{i \in I(x)} \alpha_i = 0, \; \beta_j \in \mathbb{R} \right\}, \tag{5.4}$$

The "active" index sets above are

$$I(x) = \left\{ i \in I : \left\langle a^i, x \right\rangle + b^i = f(x) \right\} \text{ and } J(x) = \left\{ j \in J : \left\langle c^j, x \right\rangle = d^j \right\}. \tag{5.5}$$

To derive the expression of the ε-subdifferential, we combine once again results from [9, Ch. XI, Sec.3.14]. First, by Theorem 3.1.1 therein, for any $\varepsilon_f \geq 0$

$$\partial_\varepsilon f(x) = \bigcup_{\varepsilon_f \in [0, \varepsilon]} \left\{ \partial_{\varepsilon_f} \mathfrak{f}(x) + \partial_{\varepsilon - \varepsilon_f} \mathbb{I}_{\mathfrak{D}}(x) \right\}.$$

Second, as shown in Example 3.5.3 and letting $e_i := \mathfrak{f}(x) - \left\langle a^i, x \right\rangle - b^i \geq 0$ for $i \in I$,

$$\partial_{\varepsilon_f} \mathfrak{f}(x) = \left\{ \sum_{i \in I} \alpha_i a^i : \alpha_i \geq 0, \sum_{i \in I} \alpha_i = 1, \sum_{i \in I} \alpha_i e_i \leq \varepsilon_f \right\}. \tag{5.6}$$

Third, by Example 3.1.4, $\partial_{\varepsilon - \varepsilon_f} \mathbb{I}_{\mathfrak{D}}(x) = N_{\mathfrak{D}, \varepsilon - \varepsilon_f}(x)$ the set of approximate normal elements. Finally, by Example 3.1.3, letting $E_j := d^j - \left\langle c^j, x \right\rangle \geq 0$ for $j \in J$,

$$\partial_{\varepsilon - \varepsilon_f} \mathbb{I}_{\mathfrak{D}}(x) = \left\{ \sum_{j \in J} \beta_j c^j : \beta_j \geq 0, \sum_{j \in J} \beta_j E_j \leq \varepsilon - \varepsilon_f \right\},$$

Combining all the expressions above,

$$\partial_\varepsilon f(x) = \bigcup_{\varepsilon_f \in [0, \varepsilon]} \left\{ \sum\nolimits_{i \in I} \alpha_i a^i + \sum\nolimits_{j \in J} \beta_j c^j \;\middle|\; \begin{array}{lll} \displaystyle\sum_{i \in I} \alpha_i e_i \leq \varepsilon_f & \alpha_i \geq 0 & \displaystyle\sum_{i \in I} \alpha_i = 1 \\ \displaystyle\sum_{j \in J} \beta_j E_j \leq \varepsilon - \varepsilon_f & \beta_j \geq 0 & \end{array} \right\}. \tag{5.7}$$

Like with the one dimensional example in Section 4, we could choose as enlargement the set in the union (5.7) that corresponds to $\varepsilon_f = \varepsilon$, i.e.,

$$\delta_\varepsilon f(x) = \partial_\varepsilon \mathfrak{f}(x) + \partial_0 \mathbb{I}_{\mathfrak{D}}(x) = \partial_\varepsilon \mathfrak{f}(x) + N_{\mathfrak{D}}(x).$$

This enlargement satisfies the sandwich inclusion (3.2), but it is not continuous everywhere. Another, similar, option is to take $\delta_\varepsilon f(x) = \partial_{\frac{\varepsilon}{2}} \mathfrak{f}(x) + N_{\mathfrak{D}, \frac{\varepsilon}{2}}(x)$, which is actually continuous everywhere except when $\varepsilon = 0$.

However, since we prefer smaller enlargements, we consider instead the ε-activity index set from (5.1), which in this setting has the form

$$I_\varepsilon(x) = \left\{i \in I \colon f^i(x) = \left\langle a^i, x\right\rangle + b^i \geq f(x) - \varepsilon\right\}, \tag{5.8}$$

and define the associated enlargement

$$\delta_\varepsilon f(x) = \left\{ \sum_{i \in I_\varepsilon(x)} \alpha_i a^i + \sum_{j \in J(x)} \beta_j c^j \;\middle|\; \begin{array}{ll} \sum_{i \in I_\varepsilon(x)} \alpha_i = 1 & \alpha_i \geq 0 \\ & \beta_j \geq 0 \end{array} \right\}, \tag{5.9}$$

which satisfies (3.2). The corresponding subspace (3.1) can be explicitly represented as follows (see [15]):

$$\mathcal{V}_\varepsilon(x) = \left\{ \sum_{i \in I_\varepsilon(x)} \alpha_i a^i + \sum_{j \in J(x)} \beta_j c^j \colon \sum_{i \in I_\varepsilon(x)} \alpha_i = 0,\ \beta_j \in \mathbb{R} \right\}. \tag{5.10}$$

Applying Prop. 5.1, we see that, for any $\bar{x} \in \mathfrak{D}$, the relaxed activity set satisfies $I_\varepsilon(x) \subset I(\bar{x})$ when x is sufficiently close to $\bar{x}$ and ε is small enough, where the inclusion holds as an equation if $\bar{x} \in \operatorname{int} \mathfrak{D}$.

In view of the expressions (5.9) and (5.3), the inclusion $\delta_\varepsilon f(x) \subset \partial f(\bar{x})$ always holds, and it becomes an identity if $\bar{x} \in \operatorname{int} \mathfrak{D}$. Hence, by (5.10), for all $\bar{x} \in \mathfrak{D}$ when (x, ε) are sufficiently close to $(\bar{x}, 0)$,

$$\mathcal{V}_\varepsilon(x) \subset \mathcal{V}(\bar{x}) \quad \text{and} \quad \mathcal{U}_\varepsilon(x) \supset \mathcal{U}(\bar{x}),$$

and

$$\limsup_{\substack{(x,\varepsilon) \to (\bar{x},0) \\ \varepsilon \geq 0}} \mathcal{V}_\varepsilon(x) = \mathcal{V}(\bar{x}) \quad \text{and} \quad \liminf_{\substack{(x,\varepsilon) \to (\bar{x},0) \\ \varepsilon \geq 0}} \mathcal{U}_\varepsilon(x) = \mathcal{U}(\bar{x}). \tag{5.11}$$

While, if $\bar{x} \in \operatorname{int} \mathfrak{D}$, when (x, ε) are sufficiently close to $(\bar{x}, 0)$,

$$\mathcal{V}_\varepsilon(x) = \mathcal{V}(\bar{x}) \quad \text{and} \quad \mathcal{U}_\varepsilon(x) = \mathcal{U}(\bar{x}), \tag{5.12}$$

and

$$\lim_{\substack{(x,\varepsilon) \to (\bar{x},0) \\ \varepsilon \geq 0}} \mathcal{V}_\varepsilon(x) = \mathcal{V}(\bar{x}) \quad \text{and} \quad \lim_{\substack{(x,\varepsilon) \to (\bar{x},0) \\ \varepsilon \geq 0}} \mathcal{U}_\varepsilon(x) = \mathcal{U}(\bar{x}). \tag{5.13}$$

5.3. Polyhedral functions: manifold and manifold relaxation

In [15, Ex. 3.4] it is shown that any polyhedral function is partly smooth at any $\bar{x}$, relative to

$$\mathcal{M}_{\bar{x}} := \{x \in \mathbb{R}^n \colon I(x) = I(\bar{x}) \text{ and } J(x) = J(\bar{x})\}. \tag{5.14}$$

We now give various equivalent characterizations for the manifold of partial smoothness. The last one is particularly suitable for defining a manifold relaxation based on the $\mathcal{U}_\varepsilon$-subspace.

The *normals parallel to subdifferential* property of partly smooth functions is equivalent to $T_{\mathcal{M}}(x) = \mathcal{U}(x)$. This condition obviously holds if $\mathcal{M}$ is just the affine subspace $x + \mathcal{U}(x)$. Particularly, this is the case for polyhedral functions locally.

Proposition 5.2. *For polyhedral functions, we have $\mathcal{M}_{\bar{x}} \subset \bar{x} + \mathcal{U}(\bar{x})$. Moreover, there exists $\varepsilon > 0$ such that $\mathcal{M}_{\bar{x}} \cap B(\bar{x}, \varepsilon) = \bar{x} + (\mathcal{U}(\bar{x}) \cap B(0, \varepsilon))$.*

Proof. For any $x \in \mathcal{M}_{\bar{x}}$ we show $x - \bar{x} \in \mathcal{V}(\bar{x})^{\perp}$. From the subdifferential characterization (5.3), we see that $\partial f(x) = \partial f(\bar{x})$ and therefore $\mathcal{V}(x) = \mathcal{V}(\bar{x})$. From the definition of $\mathcal{M}_{\bar{x}}$, (5.4), and (5.5) we have $\langle x - \bar{x}, v\rangle = 0$ for any $v \in \mathcal{V}(x)$, and the inclusion $\mathcal{M}_{\bar{x}} \subset \bar{x} + \mathcal{U}(\bar{x})$ follows.

To prove the next statement, we only need to show $I(\bar{x} + d) = I(\bar{x})$ and $J(\bar{x} + d) = J(\bar{x})$ for any $d \in \mathcal{U}(\bar{x}) \cap B(0, \varepsilon)$ and for some ε. From the continuity of the functions defining f, there exists a neighborhood $B(\bar{x}, \varepsilon)$ such that $I(\bar{x} + d) \subset I(\bar{x})$ and $J(\bar{x} + d) \subset J(\bar{x})$ for any $\bar{x} + d \in B(\bar{x}, \varepsilon)$. It suffices to show the other way around. As $d \in \mathcal{V}(\bar{x})^{\perp}$, in view of (5.4) (written with $x = \bar{x}$), we have $\left\langle \sum_{i\in I(\bar{x})} \alpha_i a^i + \sum_{j\in J(\bar{x})} \beta_j c^j, d \right\rangle = 0$ for all α_i such that $\sum_{i\in I(x)} \alpha_i = 0$ and for all $\beta_j \in \mathbb{R}$. It follows that $\left\langle a^i, d\right\rangle = \left\langle a^k, d\right\rangle$ for all $i, k \in I(\bar{x})$ and $\left\langle c^j, d\right\rangle = 0$ for all $j \in J(\bar{x})$. Consequently, $f(\bar{x}) + \left\langle a^i, d\right\rangle$ is not dependent on the choice of $i \in I(\bar{x})$. As $I(\bar{x}+d) \subset I(\bar{x})$, there must be an index $t \in I(\bar{x}) \cap I(\bar{x}+d)$. Therefore $\langle a^t, \bar{x}\rangle + b^t + \langle a^t, d\rangle = f(\bar{x} + d) = f(\bar{x}) + \langle a^t, d\rangle$. This shows that all $i \in I(\bar{x})$ are also elements of $I(\bar{x}+d)$. For any $j \in J(\bar{x})$, it follows that $\left\langle c^j, \bar{x} + d\right\rangle = \left\langle c^j, \bar{x}\right\rangle$ and thus $j \in J(\bar{x} + d)$. Consequently, $J(\bar{x} + d) \subset J(\bar{x})$ and the proof is finished. □

The manifold (5.14) is characterized by the active indices. In view of the relations between $I(x)$ and $I_\varepsilon(x)$ (cf. (5.8)), one may consider the following relaxation:

$$\mathcal{M}_x^\varepsilon := \{y \in \mathbb{R}^n \colon I_\varepsilon(y) = I_\varepsilon(x),\ J(y) = J(x)\}\,.$$

However, we do not know whether $\mathcal{M}_x$ is contained in $\mathcal{M}_x^\varepsilon$ or not. We therefore consider another option for relaxing the smooth manifold $\mathcal{M}_x$. Define

$$\mathcal{M}_x^\varepsilon := x + \mathcal{U}_\varepsilon(x),$$

where $\mathcal{U}_\varepsilon(x)$ is defined based on the enlargement $\delta_\varepsilon f(x)$ from (5.9). Then $T_{\mathcal{M}_x^\varepsilon}(x) = \mathcal{U}_\varepsilon(x)$ as $\mathcal{U}_\varepsilon(x)$ is a subspace contained in $\mathcal{U}(x)$. Furthermore, in view of (5.11) and Prop. 5.2, there exists $\varepsilon' > 0$ such that

$$\liminf_{\substack{(x,\varepsilon)\to(\bar{x},0)\\ \varepsilon\geq 0}} \mathcal{M}_x^\varepsilon \cap B(\bar{x}, \varepsilon') = \mathcal{M}_{\bar{x}} \cap B(\bar{x}, \varepsilon') \quad \text{for all } \bar{x} \in \mathfrak{D},$$

and

$$\lim_{\substack{(x,\varepsilon)\to(\bar{x},0)\\ \varepsilon\geq 0}} \mathcal{M}_x^\varepsilon \cap B(\bar{x}, \varepsilon') = \mathcal{M}_{\bar{x}} \cap B(\bar{x}, \varepsilon') \quad \text{for all } \bar{x} \in \operatorname{int} \mathfrak{D},$$

confirming the fact that $\mathcal{M}_x^\varepsilon$ enlarges the manifold $\mathcal{M}_x$.

Illustration on a simple example

Consider the function $f(x) = \max_{1\leq i\leq n} x_i$. It is known that

$$\partial f(x) = \{\alpha \in \Delta_n \colon\ \langle \alpha, x\rangle \geq f(x)\} = \{\alpha \in \Delta_n \colon \alpha_i = 0 \text{ if } x_i < f(x)\}\,,$$

where Δ_n is the unit simplex. Let $I(x) = \{i \in I \colon x_i = f(x)\}$ and

$$I_\varepsilon(x) = \{i \in I \colon x_i \geq f(x) - \varepsilon\}\,. \tag{5.15}$$

Then

$$\delta_\varepsilon f(x) = \text{co}\left\{e_i\colon x_i - \max_{i\in I} x_i + \varepsilon \geq 0\right\} = \text{co}\,\{e_i\colon i \in I_\varepsilon(x)\}\,,$$

where I_ε is defined in (5.15). For comparison,

$$\begin{aligned}\partial_\varepsilon f(x) &= \left\{s \in \Delta^n\colon\ \langle s, x\rangle \geq \max_{i\in I} x_i - \varepsilon\right\}\\ &= \left\{s \in \Delta^n\colon\ \sum_{i\in I} s_i\left(x_i - \max_{i\in I} x_i + \varepsilon\right) \geq 0\right\}.\end{aligned}$$

Consider the case of $\mathbb{R}^2$. Then

$$I_\varepsilon(x) = \begin{cases}\{1\}\,, & \text{if } x_1 > x_2 + \varepsilon\\ \{1,2\}\,, & \text{if } x_2 - \varepsilon \leq x_1 \leq x_2 + \varepsilon\\ \{2\}\,, & \text{if } x_1 < x_2 - \varepsilon\end{cases}$$

$$\delta_\varepsilon f(x) = \begin{cases}\begin{pmatrix}1\\0\end{pmatrix}, & \text{if } x_1 > x_2 + \varepsilon\\ \Delta^2, & \text{if } x_2 - \varepsilon \leq x_1 \leq x_2 + \varepsilon\\ \begin{pmatrix}0\\1\end{pmatrix} & \text{if } x_1 < x_2 - \varepsilon\end{cases} \tag{5.16}$$

We see that $\delta_\varepsilon f(x)$ is only a non-singleton if x_1 and x_2 are close enough. On the other hand, $\partial_\varepsilon f(x)$ as the intersection of a two dimensional simplex and some half space, is always non-singleton. Consider the two dimensional case, when $\bar{x} = (2,3) \in \mathbb{R}^2$ and $\varepsilon = 0.6$. From (5.16) we see $\delta_\varepsilon f(\bar{x})$ is just the point $(0,1)$. The set $\partial_\varepsilon f(\bar{x})$ is shown in Figure 3 as the solid blue line. This means that, at points where x_1 and x_2 are not "too close", the dimension of the enlarged subspace, $\mathcal{V}_\varepsilon(x)$, based on $\delta_\varepsilon f(x)$, is smaller than the subspace that would be obtained using $\partial_\varepsilon f(x)$.

In this case, $\mathcal{M}_{\bar{x}} = \{x : x_2 > x_1\}$, $\mathcal{V}_\varepsilon(x) = \{x : x_2 = -x_1\}$ if $x_2 - \varepsilon \leq x_1 \leq x_2 + \varepsilon$ and otherwise $\mathcal{V}_\epsilon(x) = \{0\}$, and $\mathcal{M}_x^\epsilon = x + \{x : x_2 = x_1\}$ if $x_2 - \varepsilon \leq x_1 \leq x_2 + \varepsilon$ and otherwise $\mathcal{M}_x^\epsilon = \mathbb{R}^2$.

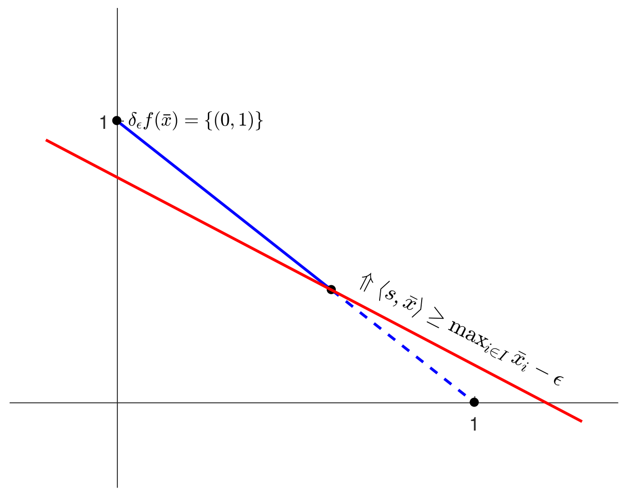
5.4. General max-functions: enlargement and $\mathcal{V}_\varepsilon\mathcal{U}_\varepsilon$-subspaces

Consider $f(x) := \max_{i\in I} f_i(x)$ where each $f_i(x)$ is convex and is C^1. The subdifferential is $\partial f(x) = \text{co}\,\{\nabla f_i(x)\colon i \in I(x)\}$, while the expression for the ε-subdifferential is similar to the one in (5.6), written with $e_i := f(x) - f^i(x) \geq 0$ for $i \in I$. To define a smaller enlargement, consider the ε-activity set in (5.1), so that

$$\delta_\varepsilon f(x) := \text{co}\,\{\nabla f_i(x)\colon i \in I_\varepsilon(x)\}$$

satisfies the sandwich inclusion (3.2). Moreover, since the max-function is now defined in the whole domain, Prop. 5.1(ii) applies with $C = \mathbb{R}^n$ and, hence, all the relations in (5.12) and (5.13) hold for this enlargement.

FIGURE 3. $\delta_\varepsilon f(\bar{x})$ and $\partial_\varepsilon f(\bar{x})$



Applying Prop. 3.3, we get the osc of $\delta_\varepsilon f(x)$. Regarding its inner semicontinuity, by Prop. 5.1, given any $\bar{x}$ when x and ε are sufficiently close to $\bar{x}$ and 0, there is always $I_\varepsilon(x) = I(\bar{x})$. Obviously the set-valued mapping $(x, \varepsilon) \mapsto \{\nabla f_i(x) \colon i \in I(\bar{x})\}$ is isc. It follows from [29, Thm. 5.9] that $\delta_\varepsilon f(x)$ is isc at $(\bar{x}, 0)$. Consequently, $\delta_\varepsilon f(x)$ is continuous at $(\bar{x}, 0)$.

The approximations of $\mathcal{VU}$ spaces can also be expressed as follows:

$$\mathcal{V}_\varepsilon(x) = \operatorname{span}\left\{\nabla f_i(x) - \nabla f_{i_\circ}(x) \colon i \in I_\varepsilon(x)\right\},$$

where $i_\circ$ is any fixed index in $I_\varepsilon(x)$, and

$$\mathcal{U}_\varepsilon(x) = \left\{y \in \mathbb{R}^n \colon \langle \nabla f_i(x), y\rangle = \langle \nabla f_{i_\circ}, y\rangle\,,\ i \in I_\varepsilon(x)\right\}.$$

Let (x, ϵ) be sufficiently close to $(\bar{x}, 0)$ and be fixed. Then $\mathcal{V}_\epsilon(x)$ and $\mathcal{U}_\epsilon(x)$ can be respectively considered as the range and kernel of a matrix A_x^ε whose columns are of the form $\nabla f_i(x) - \nabla f_{i_\circ}(x)$ for $i \in I_\varepsilon(x) = I(\bar{x})$ by Prop. 5.1, and the number of its columns is the constant $|I(\bar{x})|$. As $(x, \varepsilon) \to (\bar{x}, 0)$, due to the continuity of ∇f_i the sequence of matrices A_x^ε converges to another matrix $A_{\bar{x}}$ whose columns are of the form $\nabla f_i(\bar{x}) - \nabla f_{i_\circ}(\bar{x})$ for $i \in I(\bar{x})$. It follows from [29, Thm. 4.32] that the range and kernel of A_x^ε converges correspondingly. Consequently, $\mathcal{V}_\epsilon(x)$ and $\mathcal{U}_\epsilon(x)$ as set-valued mappings are both continuous at $(\bar{x}, 0)$ for all $\bar{x}$.

6. Sublinear functions

We next turn our attention to $\mathcal{H}$, the class of proper closed sublinear functions from $\mathbb{R}^n$ to $\overline{\mathbb{R}}$. Sublinear functions are positively homogeneous and convex; they constitute a large family including norms, quadratic seminorms, gauges of closed convex sets containing 0 in their interior, infimal convolutions of sublinear functions, and support functions of bounded sets. The class also comprises perspective functions, that are systematically studied in [2], in particular to show they provide constructive means to model general lower semicontinuous convex functions.

6.1. Enlargement and $\mathcal{V}_\varepsilon\mathcal{U}_\varepsilon$-subspaces

As shown in [9, Ch. V], sublinear functions enjoy remarkable properties. In particular, see [9, Ch. VI Remark 1.2.3], they are support functions (of the subdifferential at zero):

$$h(x) = \sup_{s\in\partial h(0)} \langle x, s\rangle \ . \tag{6.1}$$

Hence, the subdifferential is given by

$$\partial h(x) = \{s \in \partial h(0)\colon h(x) = \langle s, x\rangle\} \ , \tag{6.2}$$

see [29, Cor. 8.25]. Moreover, as shown in [9, Ch.VI Example 1.2.5], it holds that

$$\partial_\varepsilon h(x) = \{s \in \partial h(0)\colon\ \langle s, x\rangle \geq h(x) - \varepsilon\} \,, \tag{6.3}$$

from which we define the enlargement in (6.4), generalizing the one introduced in Sec. 4.3 for the model function (4.1).

Proposition 6.1. *Given $\varepsilon \geq 0$, consider the enlargement given by*

$$\delta_\varepsilon h(x) := \overline{\mathrm{co}}\,\{s \in \operatorname{ext}\partial h(0)\colon\ \langle s, x\rangle \geq h(x) - \varepsilon\} \,. \tag{6.4}$$

If $h \in \mathcal{H}$ is finite valued, the following relation holds (which is stronger than the "sandwich" relation (3.2)*):*

$$\partial h(x) \subset \delta_0 h(x) \subset \delta_\varepsilon h(x) \subset \partial_\varepsilon h(x) \,.$$

Proof. The rightmost inclusion is straightforward, because $\operatorname{ext}\partial h(0) \subset h(0)$ and the set in (6.3) is closed and convex: $\delta_\varepsilon h(x) \subset \partial_\varepsilon h(x)$. Together with the fact that $\delta_0 h(x) \subset \delta_\varepsilon h(x)$, by definition of the enlargement, we only need to show that $\partial h(x) \subset \delta_0 h(x)$. In view of (6.1), for the support function $h(x)$ to be finite everywhere, the set $\partial h(0)$ must be bounded. Plugging the identity $\partial h(0) = \overline{\mathrm{co}}\operatorname{ext}\partial h(0)$ from [28, Corollary 18.5.1] in (6.2) yields

$$\partial h(x) = \{s \in \overline{\mathrm{co}}\operatorname{ext}\partial h(0)\colon\ \langle s, x\rangle = h(x)\} \ .$$

Therefore, for any $s \in \partial h(x)$, there exist $t \in \mathbb{N}$, $s_i \in \operatorname{ext}\partial h(0)$, $\alpha_i \geq 0$, $i = 1, \cdots, t$, such that $\sum_{i=1}^t \alpha_i = 1$, $s = \sum_{i=1}^t \alpha_i s_i$ and $\langle s, x\rangle = \sum_{i=1}^t \alpha_i \langle s_i, x\rangle = h(x)$. Without loss of generality we can assume $\alpha_i > 0$ for $i = 1, \cdots, t$. Now to show $s \in \delta_0 h(x)$, by definition, we only need to show $\langle s_i, x\rangle = h(x)$ for all $i = 1, \cdots, t$. Suppose for contradiction that there exists an index j such that $\langle s_j, x\rangle \neq h(x)$. Then we must have $\langle s_j, x\rangle < h(x)$ because $h(x) = \sup_{s'\in\partial h(0)} \langle s', x\rangle$

and $s_j \in \operatorname{ext} \partial h(0) \subseteq \partial h(0)$. Consequently, $\langle s, x\rangle = \sum_{i=1}^{t} \alpha_i \langle s_i, x\rangle = \alpha_j \langle s_j, x\rangle + \sum_{i=1, i\neq j}^{t} \langle s_i, x\rangle < \alpha_j h(x) + \sum_{i=1, i\neq j}^{t} \alpha_i \langle s_i, x\rangle \leq \alpha_j h(x) + \sum_{i=1, i\neq j}^{t} \alpha_i h(x) = \alpha_j h(x) + (1-\alpha_j)\, h(x) = h(x)$, contradicting the fact that $\langle s, x\rangle = h(x)$. $\square$

6.2. A variety of sublinear functions

Since the sandwich inclusion in (3.2) holds, Prop. 3.3 ensures that the enlargement (6.4) is outer semicontinuous and satisfies (3.3). Inner continuity may hold locally in some particular cases, as shown below.

6.2.1. The absolute value function. For $h(x) = |x| = \sigma_{[-1,1]}(x)$, we have that $\partial h(0) = [-1, 1]$ and $\operatorname{ext} \partial h(0) = \{-1, 1\}$. It is not difficult to derive

$$\delta_\varepsilon h(x) = \begin{cases} \{-1\} & \text{if } x < -\varepsilon/2, \\ [-1, 1] & \text{if } -\varepsilon/2 \leq x \leq \varepsilon/2, \\ \{1\} & \text{if } x > \varepsilon/2, \end{cases}$$

which coincides with the right graph in Figure 2.

This enlargement is isc everywhere except when $|x| = \frac{\varepsilon}{2}$. For any pair $(\bar{x}, \bar{\varepsilon})$ such that $|\bar{x}| < \frac{\bar{\varepsilon}}{2}$, there exists $r > 0$ such that $|x| < \frac{\varepsilon}{2}$ for all $(x, \varepsilon) \in B(\bar{x}, r) \times [\max\{\bar{\varepsilon} - r, 0\}, \bar{\varepsilon} + r]$. Consequently, $\delta_\varepsilon h(x)$ remains a constant over $B(\bar{x}, r) \times [\max\{\bar{\varepsilon} - r, 0\}, \bar{\varepsilon} + r]$ and is continuous there. Same arguments apply if $\bar{x} > \frac{\bar{\varepsilon}}{2}$ or $\bar{x} < -\frac{\bar{\varepsilon}}{2}$. Now consider the pair such that $|\bar{x}| = \frac{\bar{\varepsilon}}{2}$. Then for all $r > 0$ there exists $(x, \varepsilon) \in B(\bar{x}, r) \times [\max\{\bar{\varepsilon} - r, 0\}, \bar{\varepsilon} + r]$ such that $|x| > \frac{\varepsilon}{2}$. According to [29, Theorem 4.10], $\delta_\varepsilon h(x)$ is isc at $(\bar{x}, \bar{\varepsilon})$ relative to $\{(x, \varepsilon)\colon \varepsilon \geq 0\}$ if and only if for every $\rho > 0$ and $\alpha > 0$ there is $B(\bar{x}, r) \times [\max\{\bar{\varepsilon} - r, 0\}, \bar{\varepsilon} + r]$ such that $[-1, 1] \cap \rho B \subset \delta_\varepsilon h(x) + \alpha B$ for all $(x, \varepsilon) \in B(\bar{x}, r) \times [\max\{\bar{\varepsilon} - r, 0\}, \bar{\varepsilon} + r]$, where B is the unit ball. Without loss of generality, take $(x', \varepsilon') \in B(\bar{x}, r) \times [\max\{\bar{\varepsilon} - r, 0\}, \bar{\varepsilon} + r]$ with $x' > \frac{\varepsilon'}{2}$ and thus $\delta_{\varepsilon'} h(x') = \{1\}$, and $\rho = \alpha = 1$. Then $[-1, 1] \subset \{1\} + [-1, 1] = [0, 2]$, which is impossible.

6.2.2. Finite valued sublinear polyhedral functions. The absolute-value function is a very simple illustration of a polyhedral function that is sublinear and finite everywhere. The enlargement defined in Sec 5 for general polyhedral functions can be particularized to the finite-valued sublinear case. For a polyhedral function to be sublinear, all the vectors b^i and d^j in (5.2) must be null, and for the function to be finite valued, the set J must be empty. As a result, given a nonempty finite index set I,

$$h(x) = \mathfrak{h}(x) \quad \text{for} \quad \mathfrak{h}(x) = \max_{i \in I} \left\{ \left\langle a^i, x \right\rangle \right\}.$$

Setting $b^i = 0$ and $J = \emptyset$ in (5.5) gives the expressions for the active and almost active index sets $I(x)$ and $I_\varepsilon(x)$. Likewise for the errors $e_i = h(x) = \left\langle a^i, x \right\rangle$ used in the ε-subdifferential expression (5.7). As shown in Sec 5.2, the enlargement (5.9) resulting from $I_\varepsilon(x)$ is osc.

For finite-valued polyhedral functions in $\mathcal{H}$, we have therefore two enlargements: $\delta_\varepsilon h(x)$ from (6.4), and the one from (5.9), given by

$$\delta_\varepsilon^{(5.9)}(x) = \mathrm{co}\left\{a^i,\ i \in I_\varepsilon(x)\right\} .$$

This enlargement is larger, because $\delta_\varepsilon h(x) = \mathrm{co}\left\{a^i,\ i \in I_\varepsilon(x) \cap I^*\right\}$, with $I^* = \left\{i \in I \colon a^i \in \mathrm{ext}\, \partial h(0)\right\}$. Together with the sandwich inclusion shown in Prop. 6.1, we have that

$$\partial h(x) \subset \delta_\varepsilon h(x) \subset \delta_\varepsilon^{(5.9)} h(x) .$$

Since, in addition, Prop. 5.1(ii) applies for all $\bar{x}$, there exists $\bar{\varepsilon} > 0$ such that

$$\text{for all } x \in B(\bar{x}, \bar{\varepsilon}) \text{ and all } \varepsilon \in [0, \bar{\varepsilon}] \quad \delta_\varepsilon^{(5.9)} h(x) = \partial h(x) .$$

As a result, locally both enlargements coincide and are continuous as multifunctions of (x, ε).

Polyhedral norms are finite-valued sublinear polyhedral functions:

$$h_1(x) := \|x\|_1 \quad \text{and} \quad h_\infty(x) := \|x\|_\infty .$$

The respective ε-subdifferentials are

$$\partial_\varepsilon h_1(x) = \{s \in \mathbb{R}^n \colon \|s\|_\infty \leq 1,\ \langle s, x\rangle \geq \|x\|_1 - \varepsilon\}$$

and

$$\partial_\varepsilon h_\infty(x) = \{s \in \mathbb{R}^n \colon \|s\|_1 \leq 1,\ \langle s, x\rangle \geq \|x\|_\infty - \varepsilon\} .$$

While the enlargements (6.4) are

$$\delta_\varepsilon h_1(x) = \mathrm{co}\ \{s \in \mathbb{R}^n \colon s_i = \pm 1,\ \langle s, x\rangle \geq \|x\|_1 - \varepsilon\}$$

and

$$\delta_\varepsilon h_\infty(x) = \mathrm{co}\ \{s \in \mathbb{R}^n \colon s_i = \pm 1,\ s_j = 0,\ j \neq i,\ 1 \leq i \leq n,\ \langle s, x\rangle \geq \|x\|_\infty - \varepsilon\} .$$

Figure 4 shows the enlargements for these functions, the δ_ε enlargement represented by the red line, and the ε-subdifferential by the gray area.

6.2.3. The maximum eigenvalue function. The maximum eigenvalue function

$$\lambda_1(X) = \max_{\{q \in \mathbb{R}^n, \|q\| = 1\}} q^\top X q$$

is also sublinear. In [27] a special enlargement was introduced to develop a second-order bundle method to minimize the composition with an affine function.

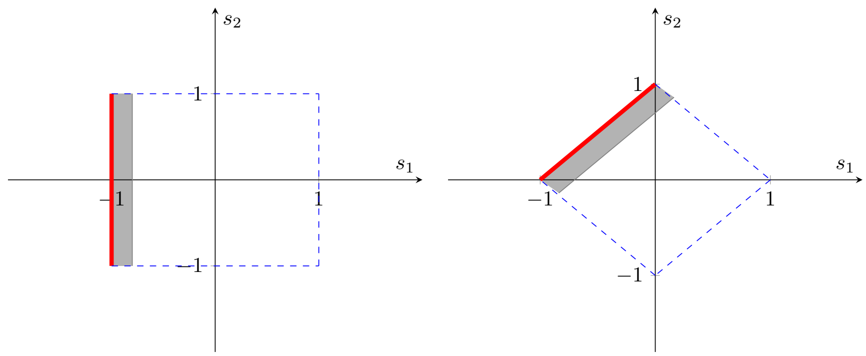
The ε-subdifferential is given by

$$\partial_\varepsilon \lambda_1(X) = \left\{\sum_{i=1}^t \alpha_i d_i d_i^\top : \|d_i\| = 1,\ \alpha_i \geq 0,\ \sum_{i=1}^t \alpha_i = 1,\ \sum_{i=1}^t \alpha_i d_i^\top X d_i \geq \lambda_1(X) - \varepsilon\right\} .$$

The expression for the subdifferential is obtained taking $\varepsilon = 0$.

In a manner similar to (5.1), the author of [27] considers an "activity" set including the ε-largest eigenvalues

$$I_\varepsilon(X) := \{i \in \{1, ..., n\} \colon \lambda_i(X) > \lambda_1(X) - \varepsilon\},$$



(A) $\delta_\varepsilon h_1(\bar{x})$ and $\partial_\varepsilon h_1(\bar{x})$ at $\bar{x} = (-2, 0)$

(B) $\delta_\varepsilon h_\infty(\bar{x})$ and $\partial_\varepsilon h_\infty(\bar{x})$ at $\bar{x} = (-2, 2)$

FIGURE 4. The ε-subdifferential and the δ_ε enlargement for h_1 and h_∞

whose eigenspace is the product of the "almost-active" eigenspaces:

$$E_\varepsilon(X) := \oplus_{i \in I_\varepsilon(X)} E_i(X) .$$

With these tools, the proposed enlargement has the expression

$$\delta_\varepsilon^{[27]} \lambda_1 (X) := \text{co} \left\{ dd^\top : \|d\| = 1,\ d \in E_\varepsilon (X) \right\} .$$

Although not apparent at first sight, by taking advantage of the structure of λ_1, this enlargement is much easier to compute than the ε-subdifferential. As shown in [27, Prop. 1], the enlargement satisfies (3.2) and is continuous, thanks to the nice analytical property in [27, Prop. 9], stating that, for some matrix X_ε,

$$\delta_\varepsilon^{[27]} \lambda_1 (X) = \partial \lambda_1 (X_\varepsilon) \quad \text{and, therefore,} \quad \mathcal{V}_\varepsilon^{[27]} (X) = \mathcal{V} (X_\varepsilon) \ .$$

The enlargement from (6.4) has the expression

$$\delta_\varepsilon \lambda_1 (X) = \text{co} \left\{ dd^\top : \|d\| = 1,\ d^\top X d \geq \lambda_1 (X) - \varepsilon \right\} ,$$

which gives a larger set than $\delta_\varepsilon^{[27]} \lambda_1 (X)$. To see this, take $d \in E_\varepsilon (X)$ such that $\|d\| = 1$. As any matrix $X \in \mathcal{S}_n$ is diagonalizable, there exist eigenvectors $d_j \in E_j (X)$, $j = 1, \cdots, \bar{r}_\varepsilon$ such that $d = \sum_{j=1}^{\bar{r}_\varepsilon} d_j$ and $\langle d_j, d_k \rangle = 0$ for $j \neq k$, where $\bar{r}_\varepsilon$ is the number of distinct eigenvalues among the first r_ε eigenvalues of X. Then $d^\top X d = \left(\sum_{j=1}^{\bar{r}_\varepsilon} d_j \right)^\top X \left(\sum_{j=1}^{\bar{r}_\varepsilon} d_j \right) = \sum_{j=1}^{\bar{r}_\varepsilon} d_j^\top X d_j$. As λ_j, $j = 1, \cdots \bar{r}_\varepsilon$ are distinct eigenvalues satisfying $\lambda_j \geq \lambda_1 - \varepsilon$, it follows that $d^\top X d \geq \sum_{j=1}^{\bar{r}_\varepsilon} \|d_j\|^2 (\lambda_1 - \varepsilon) = \lambda_1 - \varepsilon$. Therefore, $\delta_\varepsilon^{[27]} \lambda_1 (X) \subset \delta_\varepsilon \lambda_1 (X)$, as stated.

7. Concluding remarks

In Section 3, we considered a simple model function (4.1), and different versions of $\mathcal{V}_\varepsilon\mathcal{U}_\varepsilon$-subspaces, defined using different options for enlargements of the subdifferential. We argued that the usual ε-subdifferential is not the best choice for the task. In particular, comparing (4.6) and (4.8), we see that the latter is a better approximation from the point of view of $\mathcal{V}\mathcal{U}$-decomposition, because at points not near the origin, the function is smooth and it makes sense to let $\mathcal{U}_\varepsilon(x) = \mathbb{R}^2$ and $\mathcal{V}_\varepsilon(x) = \{(0,0)\}$, as in (4.8).

In fact, the situation is similar for the so-called *half-and-half* function, extending (4.1) to $\mathbb{R}^n$, with n even. Specifically, consider

$$f(x) := f_1(x) + f_2(x) \text{ with } \begin{cases} f_1(x) = \sqrt{x^\top A x} \\ f_2(x) = x^\top B x\,. \end{cases}$$

The matrix A has all the elements zero, except for ones on the diagonal at odd numbered locations ($A(i,i) = 1$ for i odd). The matrix B is diagonal with elements $B(i,i) = 1/i$ for all i. The minimizer of this partly smooth convex function is at $\bar{x} = 0$, where the $\mathcal{V}$ and $\mathcal{U}$ subspaces both have dimension $n/2$ (hence, the name "half-and-half").

Directly computing the ε-subdifferential of f is not easy. But we can exploit the structure of this function and define our enlargement to be

$$\delta_\varepsilon f(x) = \delta_\varepsilon f_1(x) + \nabla f_2(x)\,.$$

Since f_1 is sublinear, the enlargement could be $\partial_\varepsilon f_1(x)$ or the one defined in (6.4).

The possibility of choosing different enlargements, that was illustrated in this work on some structured classes of functions, can be further extended to a much larger group, by means of *composition.* Namely, suppose a function can be expressed as $f(x) = g\left(F(x)\right)$ where $g\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is a proper closed convex function and $F\colon \mathbb{R}^n \to \mathbb{R}^m$ a smooth mapping. If we have a good enlargement $\delta_\varepsilon g(\cdot)$, then we can simply define

$$\delta_\varepsilon f(\bar{x}) = \nabla F(\bar{x})^\top \delta_\varepsilon g\left(F(\bar{x})\right)\,.$$

The corresponding enlargement of the $\mathcal{V}$-space is

$$\mathcal{V}_\varepsilon^f(\bar{x}) = \nabla F(\bar{x})^\top \mathcal{V}_\varepsilon^g\left(F(\bar{x})\right)\,.$$

Exploiting those constructions in an implementable algorithmic framework is a subject of our current and future work.

Acknowledgement. The authors would like to thank Alfredo N. Iusem for raising the two counter examples given in Rem. 3.2, and Regina Burachik for her help in the proof of Prop. 3.1.

References

[1] R. S. Burachik and A. N. Iusem. *Set-Valued Mappings and Enlargements of Monotone Operators*. New York: Springer–Verlag, 2008.

[2] P. L. Combettes. "Perspective functions: Properties, constructions, and examples". *Set-Valued and Variational Analysis* (2016), pp. 1–18.

[3] R. Correa and C. Lemaréchal. "Convergence of some algorithms for convex minimization". *Mathematical Programming* 62.1-3 (1993), pp. 261–275.

[4] A. Daniilidis, D. Drusvyatskiy, and A. S. Lewis. "Orthogonal invariance and identifiability". *SIAM Journal on Matrix Analysis and Applications* 35.2 (2014), pp. 580–598.

[5] A. Daniilidis, C. Sagastizábal, and M. Solodov. "Identifying structure of nonsmooth convex functions by the bundle technique". *SIAM Journal on Optimization* 20.2 (2009), pp. 820–840.

[6] W. Hare. "Numerical Analysis of VU-Decomposition, U-Gradient, and U-Hessian Approximations". *SIAM Journal on Optimization* 24.4 (2014), pp. 1890–1913.

[7] W. L. Hare and A. S. Lewis. "Identifying active constraints via partial smoothness and prox-regularity". *Journal of Convex Analysis* 11.2 (2004), pp. 251–266.

[8] W. Hare, C. Sagastizábal, and M. Solodov. "A proximal bundle method for nonsmooth nonconvex functions with inexact information". *Computational Optimization and Applications* 63.1 (2016), pp. 1–28.

[9] J. B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I: Fundamentals*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 1996.

[10] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms II. Advanced Theory and Bundle Methods*. Vol. 306. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 1993.

[11] M. Huang, L.-P. Pang, and Z.-Q. Xia. "The space decomposition theory for a class of eigenvalue optimizations". *Computational Optimization and Applications* 58.2 (June 1, 2014), pp. 423–454.

[12] M. Huang et al. "A space decomposition scheme for maximum eigenvalue functions and its applications". *Mathematical Methods of Operations Research* 85.3 (June 1, 2017), pp. 453–490.

[13] C. Lemaréchal and C. Sagastizábal. "Practical Aspects of the Moreau–Yosida Regularization: Theoretical Preliminaries". *SIAM Journal on Optimization* 7.2 (1997), pp. 367–385.

[14] C. Lemaréchal, F. Oustry, and C. Sagastizábal. "The U-Lagrangian of a convex function". *Transactions of the American Mathematical Society* 352.2 (2000), pp. 711–729.

[15] A. S. Lewis. "Active Sets, Nonsmoothness, and Sensitivity". *SIAM Journal on Optimization* 13.3 (Jan. 2002), pp. 702–725.

[16] A. S. Lewis and S. J. Wright. "Identifying Activity". *SIAM Journal on Optimization* 21.2 (2011), pp. 597–614.

[17] A. S. Lewis and M. L. Overton. "Nonsmooth Optimization via BFGS". *Optimization Online 2172* (2008).

[18] A. S. Lewis and S. Zhang. "Partial smoothness, tilt stability, and generalized Hessians". *SIAM Journal on Optimization* 23.1 (2013), pp. 74–94.

[19] J. Liang, J. Fadili, and G. Peyré. "Activity Identification and Local Linear Convergence of Forward–Backward-type Methods". *SIAM Journal on Optimization* 27.1 (2017), pp. 408–437.

[20] J. Liang, J. Fadili, and G. Peyré. "Local Convergence Properties of Douglas–Rachford and Alternating Direction Method of Multipliers". *Journal of Optimization Theory and Applications* 172.3 (Mar. 1, 2017), pp. 874–913.

[21] Y. Lu et al. "Stochastic Methods Based on-Decomposition Methods for Stochastic Convex Minimax Problems". *Mathematical Problems in Engineering* 2014 (2014).

[22] R. Mifflin and C. Sagastizábal. "Optimization Stories". In: vol. Extra Volume ISMP 2012, ed. by M. Grötschel. DOCUMENTA MATHEMATICA, 2012. Chap. A Science Fiction Story in Nonsmooth Optimization Originating at IIASA, p. 460.

[23] R. Mifflin and C. Sagastizábal. "A VU-algorithm for convex minimization". *Mathematical programming* 104.2-3 (2005), pp. 583–608.

[24] R. Mifflin and C. Sagastizábal. "Proximal points are on the fast track". *Journal of Convex Analysis* 9.2 (2002), pp. 563–580.

[25] R. Mifflin and C. Sagastizábal. "UV-smoothness and proximal point results for some nonconvex functions". *Optimization Methods and Software* 19.5 (2004), pp. 463–478.

[26] S. A. Miller and J. Malick. "Newton methods for nonsmooth convex minimization: connections among-Lagrangian, Riemannian Newton and SQP methods". *Mathematical programming* 104.2-3 (2005), pp. 609–633.

[27] F. Oustry. "A second-order bundle method to minimize the maximum eigenvalue function". *Mathematical Programming, Series B* 89.1 (2000), pp. 1–33.

[28] R. T. Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997, pp. xviii+451.

[29] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis: Grundlehren Der Mathematischen Wissenschaften*. Vol. 317. Springer, 1998.

Shuai Liu
IMECC-UNICAMP
Rua Sérgio Buarque de Holanda, 651
Campinas - SP 13083-859
Brazil

e-mail: `shuailiu@ime.unicamp.br`

Claudia Sagastizábal
IMECC-UNICAMP
Rua Sérgio Buarque de Holanda, 651
Campinas - SP 13083-859
Brazil
e-mail: sagastiz@unicamp.br

Mikhail Solodov
Instituto de Matemática Pura e Aplicada
Estrada Dona Castorina, 110
Rio de Janeiro - RJ 22460-320
Brazil
e-mail: solodov@impa.br