The ε-strategy in variational analysis: illustration with the closed convexification of a function

Jean-Baptiste Hiriart-Urruty, Marco A. López and Michel Volle

Abstract

In this work, we concentrate our interest and efforts on general variational (or optimization) problems which do not have solutions necessarily, but which do have approximate solutions (or solutions within $\varepsilon > 0$). We shall see how to recover all the (exact) minimizers of the relaxed version of the original problem (by closed-convexification of the objective function) in terms of the ε-minimizers of the original problem. Applications to two approximation problems in a Hilbert space setting will be shown.

Introduction

This paper is mainly devoted to deriving, exploiting and illustrating the possible relationship between the minimizers of a general variational (or optimization) problem (actually the ε-minimizers) and those of a relaxed version of it (by relaxation we mean here closed-convexification of the objective function).

In section 1 we put in perspective what we call the $\varepsilon-strategy\ in\ mathematics$, i.e. how to benefit from $\varepsilon-$perturbations of a given problem. We recall in that respect some classical results or approaches like Ekeland's variational principle (§1.1), viscosity solutions for Hamilton-Jacobi equations (§1.2), and the crucial role played by the notion of $\varepsilon-$subdifferential in Convex analysis and numerical optimization (§1.3).

Section 2 is devoted to the links between a general function J and its *closed convex hull* $\overline{\text{co}}\, J$. Taking the closed convex hull is a widely spread and known operation in Variational analysis; we however list here some new or less known features of that fascinating operation (§2.1). Section 2.2 is a central one in the present paper: it gives ways to derive $\text{argmin}(\overline{\text{co}}\, J)$ from the $\varepsilon-$argmin J's. Two different and not completely comparable approaches are presented: one in the finite-dimensional context (§2.2.1) due to Benoist and Hiriart-Urruty [4] and another one, more recent, by López and Volle [23], also valid in an infinite-dimensional setting (§2.2.2). Since one cannot explain the obtained results without drawing pictures, even with functions of a real variable, we show in section 2.2.3 various illustrative examples.

Section 3 is devoted to some applications (of results in §2.2) to some Hilbertian approximation-minimization problems. In §3.1 we revisit two approximation problems, birthplaces of long-standing conjectures (surveyed in [18] and [19]), and show how a result like that of Berens [5] can be retrieved from results in section 2.2. In §3.2, we put the two considered approximation problems in a general framework and end by, again, a result allowing to recover solutions of the relaxed version of a presented problem $(\mathcal{P})$ by filtering the $\varepsilon-$solutions of this problem $(\mathcal{P})$ (Proposition 10).

1. Benefiting from ε-perturbations in variational analysis

In mathematics, it happens that one cannot tackle a problem directly, for several reasons: because the original problem has no solution, because the definition of solution itself is ambiguous, because the posed problem may have several solutions while the underlying physical or mechanical problem clearly indicates only one, etc. Therefore, an often used strategy is to circumvent the intrinsic difficulty of the problem by *perturbing it by $\varepsilon > 0$*. The perturbed problem often turns out to be easier to solve or just more meaningful than the original one. Then, since ε is small, as always in mathematics, it remains to see what happens when $\varepsilon \to 0$ or how to filter the $\varepsilon-$mathematical objects associated with the $\varepsilon-$perturbed problem. In doing so, one expects to get something of interest for the original problem, which typically corresponds to $\varepsilon = 0$.

In this paper, we concentrate our interest on variational (or optimization) problems which do not have solutions necessarily, but which do have approximate solutions (or solutions within $\varepsilon > 0$). The question we address is: what to do with such $\varepsilon-$solutions? We shall see in section 2 how to recover all the minimizers of the relaxed version of an abstract variational problem in terms of $\varepsilon-$minimizers of the original variational problem (especially when this original problem has no solution).

Before going further, we illustrate with some examples what we call the *$\varepsilon-$strategy in mathematics.* This very specific choice of examples reflects our personal interests, but no doubt that the readers have their own examples.

1.1. Ekeland's variational principle, ε-minimizers

This celebrated result is typically an answer to the question: what to do when the given variational problem has no solution, or when we cannot assert it has solutions? Ekeland's variational principle starts with $\varepsilon-$minimizers of the variational problem (for $\varepsilon > 0$) and gets the existence of (exact) minimizers of some perturbed version of the initial problem. This result is nowadays widely used in Nonlinear analysis, we even teach it at the Master level, and it has given rise to many descendants. The 30-years old reference [12] still remains a good reference for the statement of Ekeland's variational principle and some of its first applications.

A penalization process or a regularization process applied to a variational problem also are well-known ways of circumventing difficulties for tackling the original problem. Both methods consist in replacing a given minimization problem "Minimize $f(x)$, $x \in X$" by a sequence of simpler ones, says "Minimize $f_n(x)$, $x \in X$". In appropriate situations (cf. [1], [3]), any limit of a sequence of ε_n-minimizers of f_n, with $\varepsilon_n \to \varepsilon$, is an ε-minimizer of f. This way of doing is widely used in numerical optimization.

1.2. Viscosity solutions for Hamilton-Jacobi equations

Initial work on the so-called Hamilton-Jacobi equations was concerned with first-order partial differential equations like

$$(1) \qquad H(x, u(x), \nabla u(x)) = 0.$$

Since this equation may not have solutions (in the classical sense) or may have infinitely many, a way to solve it was to "perturb it by $\varepsilon > 0$", which in the present particular case amounted to introduce an $\varepsilon-$term in the equation:

$$(1_\varepsilon) \qquad -\varepsilon \triangle u(x) + H(x, u(x), \nabla u(x)) = 0.$$

The strategy therefore was:

– Solve (1_ε), under suitable assumptions, so that a "regular solution" pops out.

– Pass to the limit on $\varepsilon > 0$ and get hopefully at the limit a "solution" of (1).

This way of doing, i.e. adding a "viscosity term" $-\varepsilon\triangle u(x)$ to the equation (1), was the historical basis for the terminology "viscosity solution" which we now use for equations like (1) (see [9]).

1.3. The approximate subdifferential of a convex function

Let X be a Banach space and $f : X \to \mathbb{R} \cup \{+\infty\}$ an arbitrary function. At a point x_0, where f is finite, the subdifferential of f at x_0 is the collection of slopes of continuous affine minorants of f coinciding with f at x_0 :

$$(2) \qquad s \in \partial f(x_0) \iff f(x) \geq f(x_0) + \langle s, x - x_0 \rangle \text{ for all } x \in X.$$

We shall make use of this concept mainly for lower-semicontinuous convex f.

To ensure that $\partial f(x_0)$ is nonempty is not an easy matter. Moreover, calculus rules like the one giving $\partial(f + g)(x_0)$ as the sum $\partial f(x_0) + \partial g(x_0)$ requires additional assumptions of f and g.

In order to be more comfortable with these drawbacks, an "enlargement by $\varepsilon > 0$" of $\partial f(x_0)$ was proposed; it is called the $\varepsilon-$subdifferential of f at x_0, and it is defined as follows:

$$(2_\varepsilon) \qquad s \in \partial_\varepsilon f(x_0) \iff f(x) \geq f(x_0) + \langle s, x - x_0 \rangle - \varepsilon \text{ for all } x \in X.$$

First good news: If $f^{**}(x_0) = f(x_0)$ (where f^{**} stands for the biconjugate of f, see later on), as it is when f is lower-semicontinuous convex, $\partial_\varepsilon f(x_0)$ is nonempty whenever $\varepsilon > 0$ (even if $\partial f(x_0)$ is empty) and, of course, we recover $\partial f(x_0)$ from $\partial_\varepsilon f(x_0)$, $\varepsilon > 0$, by filtering all of them: $\partial f(x_0) = \cap_{\varepsilon>0}\partial_\varepsilon f(x_0)$.

More interesting is that one can derive calculus rules on subdifferentials, without any assumption on the involved functions, by using $\varepsilon-$subdifferentials instead of (exact) subdifferentials. A typical result in that direction is the following calculus rule on the subdifferential of the sum of convex functions.

Theorem 1 (Hiriart-Urruty and Phelps, [22]) *Let f and g be lower-semicontinuous convex functions from a Banach space X into $\mathbb{R} \cup \{+\infty\}$. At a point x_0 where both f and g are finite, we have*:

$$\partial(f + g)(x_0) = \bigcap_{\varepsilon>0} \overline{\partial_\varepsilon f(x_0) + \partial_\varepsilon g(x_0)}^{weak^*}.$$

If moreover X is reflexive,

$$(3) \qquad \partial(f + g)(x_0) = \bigcap_{\varepsilon>0} \overline{\partial_\varepsilon f(x_0) + \partial_\varepsilon g(x_0)}.$$

This is typical of what one can expect by adopting the ε-strategy: since $\partial(f+g)(x_0)$ cannot be obtained from $\partial f(x_0) + \partial g(x_0)$, perturb by $\varepsilon > 0$, take the corresponding ε-subdifferential $\partial_\varepsilon f(x_0) + \partial_\varepsilon g(x_0)$, and filter all of them with the intersection operation. For more on the calculus rules using ε-subdifferentials, see [21]. In [15] the assumption of lower semi-continuity of f and g is replaced by a slightly weaker "closedness" qualification.

The behaviour as a set-valued mapping of $x \in X$ also differs drastically when one moves from $\partial f(x)$ to $\partial_\varepsilon f(x)$. For a continuous convex function on the open convex subset Ω of X, the best semicontinuity result on the subdifferential set-valued mapping from X into X^* is that $x \longmapsto \partial f(x)$ is norm-to-weak* upper (or outer) semicontinuous on Ω. For $\varepsilon > 0$, the bevaviour of $x \longmapsto \partial_\varepsilon f(x)$ is much nicer.

Theorem 2 ([16]) *Let $f : X \to \mathbb{R}$ be a convex Lipschitz function and let $\varepsilon > 0$. Then there exists $K > 0$ such that*

$$\text{(4)} \qquad \text{haus}\,(\partial_\varepsilon f(x), \partial_\varepsilon f(x')) \le K \,\|x - x'\| \ \text{ for all } x, x' \text{ in } X.$$

Here, haus(C, D) *stands for the* HAUSDORFF *distance between C and D.*

For a non Lipschitz function f, the above result is adapted by considering $\partial_\varepsilon f(x) \cap r\mathbb{B}^*$ ($r > 0$ large enough, $\mathbb{B}^*$ closed unit ball in X^*) instead of $\partial_\varepsilon f(x)$ itself. In short, while the difficulty for obtaining the continuity of the set-valued mapping $x \longmapsto \partial f(x)$ is "$+\infty$" (no hope, as a general rule), it is null (i.e., no problem at all!) for the set-valued mapping $x \longmapsto \partial_\varepsilon f(x)$ whenever $\varepsilon > 0$.

The "enlargement by viscosity" of $\partial f(x)$ to get $\partial_\varepsilon f(x)$ ended with a concept close to the initial one $\partial f(x)$, but behaving in a much more regular way. This has had consequences in the theoretical study of convex optimization problems, as also –indeed much more– when dealing with numerical minimization procedures [20, Vol. II].

2. Relaxation by closed convexification

"Relaxing a problem" has various meanings in mathematics, depending on the areas where it is defined, depending also on what one relaxes (a functional, the underlying space, etc.). In a variational context, when dealing with the minimization of $J : X \to \mathbb{R} \cup \{+\infty\}$, the most general way of looking at relaxation is to consider the lower-semicontinuous hull of J. It allows us not to be limited by the linear structure of X and, therefore, to consider situations where the set of possible choices is only a metric space (or a mere topological space) as it happens to be in shape optimization. In those cases

in which we are dealing with normed vector spaces (context of Variational calculus), usually one considers the closed convex hull of J. In the present approach, we are not going to consider the more general framework, but indeed the way in which we relax is passing from J to its closed convex hull. Henceforth, the context is a follows:

(5)   *X is a Banach space; $J : X \rightarrow \mathbb{R} \cup \{+\infty\}$ is not identically equal to $+\infty$, and it is minorized by some continuous affine minorant.*

The topological dual space is endowed with a topology so that a duality pairing (X, X^*) is settled for the LEGENDRE-FENCHEL transformation (see below). In particular, the biconjugate J^{**} of J is defined on X (and not on X^{**}).

2.1. The closed convex hull operation

For a given $J : X \rightarrow \mathbb{R} \cup \{+\infty\}$, getting its closed convex hull $\overline{\mathrm{co}}\, J : X \rightarrow \mathbb{R} \cup \{+\infty\}$ is a complicated, but at the same time fascinating, operation. There are at least two ways of constructing $\overline{\mathrm{co}}\, J$:

- the "internal construction": consider all the convex combinations of elements of the epigraph epi J of J, so that co(epi J) is built, and then close it; the set $\overline{\mathrm{co}}$(epi J) is the epigraph of a function, namely of $\overline{\mathrm{co}}\, J$.

- the "external construction": consider all the continuous affine functions a_J minorizing J; then $\overline{\mathrm{co}}\, J = \sup a_J$.

The fact that we get exactly the same function, via the internal construction or the external one, is one of the key results in Convex analysis. In terms of the LEGENDRE-FENCHEL transformation (or conjugacy) $J \rightsquigarrow J^*$, since J is assumed to satisfy (5), we have that $J^{**} := (J^*)^*$ equals $\overline{\mathrm{co}}\, J$. We therefore use the notations $\overline{\mathrm{co}}\, J$ and J^{**} indistinctly.

The closed-convex-hull operation is a *global* one, in the sense that it requires to know –a priori– the behaviour of J on the whole X. In particular, the behaviour of $J(x)$ "at infinity", i.e. when $\|x\| \rightarrow \infty$, is of utmost importance; that is the main source of difficulties in delineating $\overline{\mathrm{co}}\, J$.

Let us review here the main known properties of $\overline{\mathrm{co}}\, J$, those inherited from J and those not.

- The *infimal values*. We have:

(6) $\displaystyle \inf_{x \in X} J(x) = \inf_{x \in X} (\overline{\mathrm{co}}\, J)(x)$ (an equality in $\mathbb{R} \cup \{-\infty\}$).

This is simply due to the fact that $\inf_{x \in X} J(x) = -J^*(0)$ and $J^* = (\overline{\mathrm{co}}\, J)^*$.

- The *set of minimizers.* If we denote by $\operatorname{argmin} J$ the set of $x \in X$ minimizing J on X (possibly, the empty set), we easily see that

$$(7) \qquad \overline{\operatorname{co}}(\operatorname{argmin} J) \subset \operatorname{argmin}(\overline{\operatorname{co}} J),$$

a very weak result indeed, especially if $\operatorname{argmin} J$ is empty! We shall come back to this point in the next section; see also the examples from one-dimensional variational calculus below.

Playing with the equality $J^* = (\overline{\operatorname{co}} J)^*$ and the "seesaw rule": $x \in \partial J^*(s)$ if and only if $s \in \partial(\overline{\operatorname{co}} J)(x)$, we get:

$$(8) \qquad (\overline{x} \text{ minimizes } J \text{ on } X) \Leftrightarrow (J(\overline{x}) = (\overline{\operatorname{co}} J)(\overline{x}) \text{ and } \overline{x} \in \partial J^*(0)),$$

still not very informative. There is however a situation where the coincidence relation serves in filtering global minimizers of J from just critical (or stationary) points. Suppose, for the sake of simplicity, that X is a Hilbert space. We denote by $\nabla J(x)$ the gradient of J at x, whenever it is GÂTEAUX-differentiable at x.

Theorem 3 *Let $J : X \to \mathbb{R} \cup \{+\infty\}$ be* GÂTEAUX-*differentiable at $\overline{x}$, where X is a Hilbert space. Then*:

$$(9) \qquad \begin{pmatrix} \overline{x} \text{ is a global minimizer} \\ \text{of } J \text{ on } X \end{pmatrix} \Leftrightarrow \begin{pmatrix} \nabla J(\overline{x}) = 0 \text{ and} \\ J(\overline{x}) = (\overline{\operatorname{co}} J)(\overline{x}) \end{pmatrix}.$$

This had been observed by DEDIEU in the context of Calculus of variations [11]; a short pedagogical proof in a finite dimensional setting was published in [17]. Indeed, the proof is easy from (8) and the following observation: if J is GÂTEAUX-differentiable at $\overline{x}$, then either $\partial J(\overline{x})$ is empty or $\partial J(\overline{x}) = \{\nabla J(\overline{x})\}$. At a critical point $\overline{x}$, we are precisely in the latter case, $\partial J(\overline{x}) = \{0\}$, so that $\overline{x} \in \partial J^*(0)$.

It is interesting to note how a local condition ($\nabla J(\overline{x}) = 0$), combined with a global one ($J(\overline{x}) = (\overline{\operatorname{co}} J)(\overline{x})$) serve to distinguish global minimizers of J from just critical points. It is also worth noting that the theorem above belongs to the realm of differentiable optimization. Indeed, if the property "$\overline{x}$ is a local minimizer of J" replaces "$\nabla J(\overline{x}) = 0$", in absence of differentiability of J at $\overline{x}$, then the equivalence (9) breaks down. That means that any generalization of "$\nabla J(\overline{x}) = 0$" of the form "$0 \in \partial^g J(\overline{x})$" for some generalized subdifferential $\partial^g J$ of J will not do the job..., an upsetting fact!

We continue with the review of the properties of the closed convex hull of a function.

- The *continuity property.* Even if J is the restriction of a $\mathcal{C}^\infty$ function on a compact convex subset C of $\mathbb{R}^n$ (and $+\infty$ out of C), the (convex) function $\overline{\text{co}}J$ may not be continuous at some boundary points of C.

- The *differentiability property.* If $J : \mathbb{R} \to \mathbb{R}$ is differentiable on $\mathbb{R}$, then so is $\overline{\text{co}}\,J$ (even if $(\overline{\text{co}}\,J)(x) < J(x)$ for all $x \in \mathbb{R}$). There however are $\mathcal{C}^\infty$ functions $J : \mathbb{R}^2 \to \mathbb{R}$ for which $\overline{\text{co}}\,J$ is no more differentiable on $\mathbb{R}^2$. The general possible relationship between the subdifferential of $\overline{\text{co}}\,J$ and that of J is studied in [4]. Let us just quote a result from there.

Theorem 4 *Suppose that $J:\mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is lower-semicontinuous and* Gâteaux-*differentiable on $\mathbb{R}^n$ (for instance, if J is* Fréchet-*differentiable on $\mathbb{R}^n$). Assume moreover that J is asymptotically epi-pointed on $\mathbb{R}^n$ in the sense that J is minorized by $n+1$ affine functions $\langle s_i, \cdot \rangle - r_i$ with affinely independent slopes s_i (an equivalent condition is that the domain of J^* has a nonempty interior). Then $\overline{\text{co}}\,J$ is a $\mathcal{C}^1$ function on $\mathbb{R}^n$.*

- *Behaviour at infinity.* Indeed $\overline{\text{co}}\,J \leq J$ on X. However, $\overline{\text{co}}\,J$ ends by "behaving like J at infinity". The following result and proof are taken from [6].

Theorem 5 *We have*

$$(10) \qquad \liminf_{\|x\| \to \infty} \frac{J(x) - (\overline{\text{co}}\,J)(x)}{\|x\|} = 0.$$

Proof. Since $J(x) \geq (\overline{\text{co}}\,J)(x)$ for all $x \in X$, the above $\liminf$ is $\ell \geq 0$ (possibly $+\infty$). Suppose $\ell > 0$. Therefore, there exist $c > 0$ and $A > 0$ such that

$$(11) \qquad \inf_{\|y\| \geq \|x\|} \frac{J(y) - (\overline{\text{co}}\,J)(y)}{\|y\|} \geq c \text{ whenever } \|x\| > A.$$

Thus, $J(x) - (\overline{\text{co}}\,J)(x) \geq c\,\|x\|$ when $\|x\| > A$, while $J(x) - (\overline{\text{co}}\,J)(x) \geq 0$ otherwise. In short,

$$J(x) \geq (\overline{\text{co}}\,J)(x) + c(\|x\| - A) \text{ for all } x \in X.$$

This comparison result between two functions, the one on the right-hand side being convex, yields

$$(\overline{\text{co}}\,J)(x) \geq (\overline{\text{co}}\,J)(x) + c(\|x\| - A) \text{ for all } x \in X.$$

This does not hold true for $\|x\| > A$. Hence, the hypothesis at the beginning of the proof, $\ell > 0$, is wrong. ∎

In spite of (10), note that the gap between $(\overline{\text{co}}\, J)(x)$ and $J(x)$ when $\|x\| \to \infty$ may be larger and larger; think of $J : \; x \in \mathbb{R} \mapsto J(x) := \sqrt{|x|}$ for that.

- *Numerical computation.* The numerical computation of $(\overline{\text{co}}\, J)(x)$, via that of J^* (on $\mathbb{R}^n$, or on some grids of $\mathbb{R}^n$) is not broached here. For that, we refer to [7] and to the recent survey-paper [24].

We finish this section by showing a classical example in Calculus of variations where there is no minimizer for J while there is just one for the relaxed function $\overline{\text{co}}\, J$ (on the same underlying space X). In Calculus of variations, under suitable assumptions (cf. [13], [10]), the relaxed form of a criterion like $u \in X \mapsto J(u) := \int_{[a,b]} \ell(t, u(t), u'(t))\, dt$ turns out to be $\int_{[a,b]} (\overline{\text{co}}\, \ell)(t, u(t), u'(t))\, dt$, where the inner closed-convexification of ℓ is with respect to the velocity, i.e. the closed-convexification of $\ell(t, u, \cdot)$. Let $X = H^1(0,1)$ be the standard HILBERT-SOBOLEV space equipped with the inner product

$$(u \mid v) := \int_{[0,1]} [u(t)v(t) + u'(t)v'(t)]\, dt,$$

and the associated Hilbertian norm $\|\cdot\| = \sqrt{(\cdot \mid \cdot)}$. We consider

$$u \in X \mapsto J(u) := \int_{[0,1]} \left\{ \left| (u'(t))^2 - 1 \right| + u(t)^2 \right\} dt.$$

The function J is continuous and $1-$coercive on X (i.e. $J(u)/\|u\| \to +\infty$ when $\|u\| \to +\infty$). Its relaxed version $\overline{\text{co}}\, J$ is defined as follows:

$$u \in X \mapsto (\overline{\text{co}}\, J)(u) := \int_{[0,1]} \left\{ [(u'(t))^2 - 1]^+ + u(t)^2 \right\} dt.$$

By considering "serrated functions[1]" u, one easily sees that $\inf_{u \in X} J(u) = 0$, while there is no $\overline{u} \in X$ such that $J(\overline{u}) = 0$. As for the relaxed function $\overline{\text{co}}\, J$, it has only one minimizer, namely $\overline{u} \equiv 0$.

This is a typical situation in Calculus of variations where J has no minimizer, while $\overline{\text{co}}\, J$ does have one. There is no way of recovering $\text{argmin}(\overline{\text{co}}\, J)$ from $\text{argmin}\, J$ (the empty set!). But there are $\varepsilon-$minimizers of J on X, i.e. $u_\varepsilon \in X$ for which $J(u_\varepsilon) \le \inf_X J + \varepsilon$ (here, $\varepsilon > 0$); so there is some hope to get something by applying the $\varepsilon-$strategy: consider the set of $\varepsilon-$minimizers of J, and filter in some way to get the set of minimizers of $\overline{\text{co}}\, J$. This is what is done in the next section.

---

[1]The so-called "serrated functions" are continuous piecewise linear functions, with slopes ± 1, smaller and smaller as the number of switching points for slopes is higher.

2.2. Deriving argmin($\overline{\mathrm{co}}\,J$) from the ε-argmin J's

In this section, without loss of generality, we consider a situation a little more precise than the one in (5):

(12)   *X is a Banach space; $J : X \to \mathbb{R} \cup \{+\infty\}$ is not identically equal to $+\infty$, and bounded from below on X.*

Thus, $-J^*(0) = \inf_{x \in X} J(x) > -\infty$, and there are minimizing sequences (x_n) for J, i.e. (x_n) for which $J(x_n) \to \inf_X J$ when $n \to \infty$. The following result then holds true in the usual context of Calculus of variations (cf. [13, 10]): every cluster point of a minimizing sequence for J is a minimizer of $\overline{\mathrm{co}}\,J$; every minimizer of $\overline{\mathrm{co}}\,J$ is the limit of a minimizing sequence for J.

For deriving argmin($\overline{\mathrm{co}}\,J$) from argmin J or from the $\varepsilon - \mathrm{argmin}\,J$'s, $\varepsilon > 0$,

$$\varepsilon - \mathrm{argmin}\,J = \{x \in X \mid J(x) \le \inf_X J + \varepsilon\},$$

there are two competing approaches, not completely comparable, one more appropriate to the finite dimensional context that uses exact minimizers, the other one valid in the infinite dimensional setting but using approximate minimizers.

2.2.1. An approach in the finite-dimensional context

Here, $X = \mathbb{R}^n$ and we assume (at least) that $J : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is not identically equal to $+\infty$, lower-semicontinuous and bounded from below on $\mathbb{R}^n$. The behaviour at infinity of J is taken into account by the asymptotic function J'_∞ of J:

(13) $$d \in \mathbb{R}^n \mapsto J'_\infty(d) := \liminf_{\substack{t \to 0^+ \\ v \to d}} t J(v/t).$$

In other words, $J'_\infty(d)$ is "the slope at infinity of J in the d direction". If J is moreover convex, we recover the more classical expression

$$J'_\infty(d) = \lim_{t \to +\infty} \frac{J(x_0 + td) - J(x_0)}{t} = \sup_{t > 0} \frac{J(x_0 + td) - J(x_0)}{t},$$

where x_0 is any point at which J is finite.

The property of being asymptotically epi-pointed was originally defined for J in terms of the cone $K = \mathrm{epi}\,J'_\infty$. More palpable properties can be derived, dealing directly with the function J. Here they are:

J is said to be *asymptotically epi-pointed* on $\mathbb{R}^n$ if one of the following equivalent conditions holds true:

1) J is minorized by $n + 1$ affine functions $\langle s_i, \cdot \rangle - r_i$ with affinely independent s_i;

2) there exists $s_0 \in \mathbb{R}^n, \sigma_0 > 0$ and $r_0 \in \mathbb{R}^n$ such that

$$J(x) \geq \langle s_0, x \rangle + \sigma_0 \|x\| - r_0, \text{ for all } x \in \mathbb{R}^n;$$

3) there exists $s_0 \in \mathbb{R}^n$ such that

$$\liminf_{\|x\| \to \infty} \frac{J(x) - \langle s_0, x \rangle}{\|x\|} > 0;$$

4) the domain of J^*, dom J^*, has a nonempty interior.

Then, knowing that $\operatorname{argmin}(\overline{\operatorname{co}} J) = \{\overline{x} \in \mathbb{R}^n \mid 0 \in \partial(\overline{\operatorname{co}} J)(\overline{x})\}$ and the rules relating $\partial(\overline{\operatorname{co}} J)$ and ∂J (cf. [4]), we get:

Theorem 6 *Assume that J is asymptotically epi-pointed on $\mathbb{R}^n$. Then*:

(14) $$\operatorname{argmin}(\overline{\operatorname{co}} J) = \operatorname{co}(\operatorname{argmin} J) + \operatorname{co}(\operatorname{argmin} J'_\infty)$$
$$= \operatorname{co}(\operatorname{argmin} J + \operatorname{argmin} J'_\infty).$$

Beware! (14) can just be: $\emptyset = \emptyset$.

Besides the convex hull operation, necessary since $\operatorname{argmin}(\overline{\operatorname{co}} J)$ is convex and $\operatorname{argmin} J$ and $\operatorname{argmin} J'_\infty$ are not, we observe in (14) two contributions to the construction of $\operatorname{argmin}(\overline{\operatorname{co}} J)$: one directly from J itself, another one from the behaviour at infinity of J (via J'_∞).

A particular case of application of Theorem 6: if $\overline{\operatorname{co}} J$ is $0-$coercive on $\mathbb{R}^n$ (a stronger requirement than the $0-$coercitivity of J), then we have:

(15) $$\operatorname{argmin}(\overline{\operatorname{co}} J) = \operatorname{co}(\operatorname{argmin} J).$$

We do not know of any generalization of Theorem 6 to an infinite dimensional setting.

2.2.2. Another approach in an infinite dimensional context

We come back here to the general situation under assumption (12). In general, $(\overline{\operatorname{co}} J)'_\infty \leq \overline{\operatorname{co}} J'_\infty$. The behaviour of $\overline{\operatorname{co}} J$ at infinity is described by $\{d \in X \mid (\overline{\operatorname{co}} J)'_\infty(d) \leq 0\}$ (called *recession cone* of $\overline{\operatorname{co}} J$) or, in its dual form, by the normal cone to dom J^* at 0,

(16) $$N(\operatorname{dom} J^*, 0) = (\mathbb{R}_+ \operatorname{dom} J^*)^\circ$$
$$= \{x \in X \mid \langle s, x \rangle \leq 0 \text{ for all } s \in \operatorname{dom} J^*\}.$$

(Remember that we have assumed $0 \in \operatorname{dom} J^*$.)

With the help of $\varepsilon-$minimizers of J on X and $N(\operatorname{dom} J^*, 0)$, it is now possible to recover $\operatorname{argmin}(\overline{\operatorname{co}} J)$. Here is a first series of results in that direction.

Theorem 7 (López and Volle, [23])   (a) *As a general rule, we have:*

$$(17) \qquad \operatorname{argmin}(\overline{\operatorname{co}} J) = \bigcap_{\substack{\varepsilon > 0 \\ s \in \operatorname{dom} J^*}} \overline{\operatorname{co}} \left(\varepsilon - \operatorname{argmin} J + s^\circ \right),$$

where s° stands for the closed half-space of X normal to s, i.e.

$$s^\circ = \{ x \in X \mid \langle s, x \rangle \leq 0 \}.$$

(b) *If either the cone generated by $\operatorname{dom} J^*$, i.e. $\mathbb{R}_+ \operatorname{dom} J^*$, is $w^* -$closed, or the topological relative interior of $\mathbb{R}_+ \operatorname{dom} J^*$ is nonempty, then:*

$$(18) \qquad \operatorname{argmin}(\overline{\operatorname{co}} J) = \bigcap_{\varepsilon > 0} \overline{\operatorname{co}} \{ \varepsilon - \operatorname{argmin} J + N(\operatorname{dom} J^*, 0) \}.$$

In particular, if the cone generated by $\operatorname{dom} J^*$ is the whole of X^*, we have:

$$(19) \qquad \operatorname{argmin}(\overline{\operatorname{co}} J) = \bigcap_{\varepsilon > 0} \overline{\operatorname{co}} \left(\varepsilon - \operatorname{argmin} J \right).$$

Like in (14), we note here two contributions to the construction of $\operatorname{argmin}(\overline{\operatorname{co}} J)$: one coming directly from the $\varepsilon-$minimizers of J, another one taking into account the behaviour of $\overline{\operatorname{co}} J$ (via the recession cone $N(\operatorname{dom} J^*, 0)$). Contrary to (14), the behaviour of J itself does not show up directly, but that of $\overline{\operatorname{co}} J$ instead. In short, what enters into the picture in the formulae, in a somewhat hidden form, is the "collection of slopes s of continuous affine minorants of $\overline{\operatorname{co}} J$ (or, equivalently, of J)".

Recall that when X is reflexive, the cone generated by $\operatorname{dom} J^*$ is the whole X^* whenever $\overline{\operatorname{co}} J$ is $0-$coercive on X.

A result, a little bit sharper than (18), can be obtained when X is reflexive; it involves $\overline{\operatorname{co}} \left(\varepsilon - \operatorname{argmin} J \right) + N(\operatorname{dom} J^*, 0)$ rather than the *a priori* larger $\overline{\operatorname{co}} \left[(\varepsilon - \operatorname{argmin} J) + N(\operatorname{dom} J^*, 0) \right]$.

Theorem 8 *Assume that X is a reflexive Banach space.*

(a) *If the interior of $\mathbb{R}_+ \operatorname{dom} J^*$ (for the strong topology) is nonempty, then:*

$$(20) \qquad \operatorname{argmin}(\overline{\operatorname{co}} J) = \bigcap_{\varepsilon > 0} [\overline{\operatorname{co}} \left(\varepsilon - \operatorname{argmin} J \right) + N(\operatorname{dom} J^*, 0)].$$

(b) *If there exists $s \in X^*$, $\alpha > 0$ and $r \in \mathbb{R}$ such that*

$$(21) \qquad J(x) \geq \alpha \|x\| + \langle s, x \rangle - r \text{ for all } x \in X,$$

then (20) holds true.

Proof. (a) In order to abreviate the notations, we set $M_\varepsilon J := \varepsilon - \operatorname{argmin} J$. For a set A in X (resp. in X^*), we denote by i_A the indicator function of A and by $\sigma_A := i_A^*$ the support function, defined on X^* (resp. on X). The proof is decomposed in three steps.

Step 1. $\operatorname{dom} J^*$ (hence $\mathbb{R}_+ \operatorname{dom} J^*$) is contained in the domain of $\sigma_{M_\varepsilon J}$ (also called the barrier cone of $M_\varepsilon J$ in Convex analysis). Indeed, let $p \in \operatorname{dom} J^*$; we have

$$J^*(p) = \sup_{x \in \operatorname{dom} J} [\langle p, x \rangle - J(x)] < +\infty.$$

Now, for $x \in M_\varepsilon J$, $J(x) \leq \inf_X J + \varepsilon = -J^*(0) + \varepsilon$, so that

$$\langle p, x \rangle \leq J^*(p) - J^*(0) + \varepsilon.$$

Hence we obtain that $\sigma_{M_\varepsilon J}(p) < +\infty$.

Step 2. For $\varepsilon > 0$, the set $S_\varepsilon := \bigcap_{s \in \operatorname{dom} J^*} \overline{\operatorname{co}}(M_\varepsilon J + s^\circ)$ is the subdifferential at 0 of the convex function $\sigma_{M_\varepsilon J} + i_{\mathbb{R}_+ \operatorname{dom} J^*}$. Indeed, by definition,

$$x \in S_\varepsilon \Leftrightarrow x \in \overline{\operatorname{co}}(M_\varepsilon J + s^\circ) \text{ for all } s \in \operatorname{dom} J^*,$$

and this is equivalent to

(22) $$\langle p, x \rangle \leq \sigma_{M_\varepsilon J}(p) + \sigma_{s^\circ}(p),$$

for all $p \in X^*$ and all $s \in \operatorname{dom} J^*$ (or $s \in \mathbb{R}_+ \operatorname{dom} J^*$).

Now, $\sigma_{s^\circ}(p) = 0$ if $p = \alpha s$ for some $\alpha \geq 0$, and $\sigma_{s^\circ}(p) = +\infty$ in other cases. Thus, (22) can be summarized in the following inequality:

(23) $$\langle p, x \rangle \leq \sigma_{M_\varepsilon J}(p) + i_{\mathbb{R}_+ \operatorname{dom} J^*}(p) \text{ for all } p \in X^*.$$

Since $\left(\sigma_{M_\varepsilon J} + i_{\mathbb{R}_+ \operatorname{dom} J^*}\right)(0) = 0$, we recognize in (23) the definition of

$$x \in \partial\left(\sigma_{M_\varepsilon J} + i_{\mathbb{R}_+ \operatorname{dom} J^*}\right)(0).$$

Step 3. It results from Step 1 and from the assumption made in the theorem that the convex function $i_{\mathbb{R}_+ \operatorname{dom} J^*}$ is finite and (strongly) continuous at a point where the other convex function $\sigma_{M_\varepsilon J}$ is finite. By a classical calculus rule on the subdifferential sum of two convex functions ([13], [2], and [25], for example), we have

$$S_\varepsilon = \partial\left(\sigma_{M_\varepsilon J} + i_{\mathbb{R}_+ \operatorname{dom} J^*}\right)(0) = \partial\sigma_{M_\varepsilon J}(0) + \partial\left(i_{\mathbb{R}_+ \operatorname{dom} J^*}\right)(0).$$

But $x_1 \in \partial\sigma_{M_\varepsilon J}(0)$ exactly means that $x_1 \in \overline{\operatorname{co}}(M_\varepsilon J)$, while

$$\partial\left(i_{\mathbb{R}_+ \operatorname{dom} J^*}\right)(0) = N(\operatorname{dom} J^*, 0).$$

As a final result,

$$S_\varepsilon = \overline{\mathrm{co}}\left(M_\varepsilon J\right) + N\left(\mathrm{dom}\, J^*, 0\right).$$

It then remains to apply (a) of Theorem 7, which expresses that

$$\mathrm{argmin}\left(\overline{\mathrm{co}}\, J\right) = \bigcap_{\varepsilon > 0} S_\varepsilon.$$

(b) Following the assumption (21) on J, one easily checks that $J^*\left(p\right)$ is finite whenever $\|p - s\|_* \leq \alpha$ (ball of X^* for the dual norm). Therefore, part (b) of Theorem 7 applies. ∎

Comments

- If s is taken to be 0 in the assumption (21) of Theorem 8, the function $\overline{\mathrm{co}}\, J$ is indeed $0-$coercive on X, so that $N\left(\mathrm{dom}\, J^*, 0\right) = \{0\}$ and

$$\mathrm{argmin}\left(\overline{\mathrm{co}}\, J\right) = \bigcap_{\varepsilon > 0} \overline{\mathrm{co}}\left(\varepsilon - \mathrm{argmin}\, J\right),$$

  a result we already noticed earlier.

- Beware that the formulae in Theorem 7 and Theorem 8 could just say, like in Theorem 6, that $\emptyset = \emptyset$!

- Theorem 6 and Theorem 8 (b) both apply to asymptotically epi-pointed functions on $\mathbb{R}^n$, but the resulting formulae on $\mathrm{argmin}\left(\overline{\mathrm{co}}\, J\right)$ are different; one could not get one from the other.

2.2.3. Various illustrative examples

Here we illustrate either Theorem 6 or Theorems 7–8 with various one-dimensional examples and a two-dimensional one.
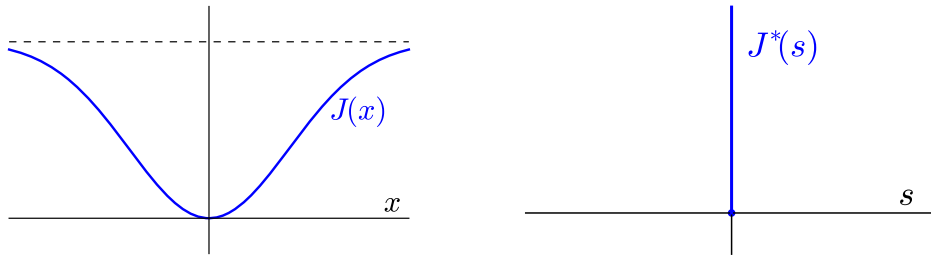
Example 1

argmin $J = \{0\}$ $N(\mathrm{dom}\, J^*, 0) = \mathbb{R}_+$ argmin $J'_\infty = \mathbb{R}_+$
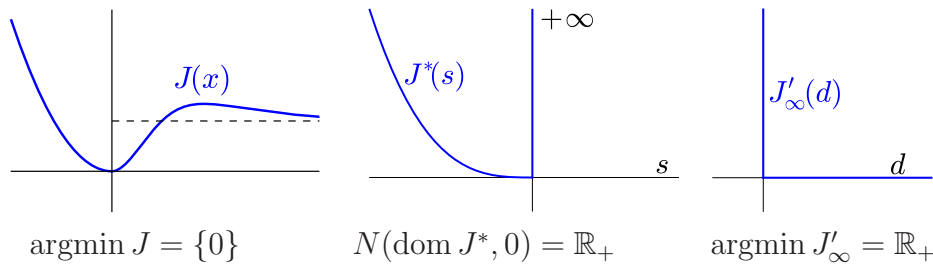
In this example, J is asymptotically epi-pointed. Both Theorem 6 and Theorems 7-8 apply:

$$\operatorname{argmin}\left(\overline{\operatorname{co}} J\right) = \mathbb{R}_+;$$

$$\operatorname{co}\left(\operatorname{argmin} J + \operatorname{argmin} J'_\infty\right) = \mathbb{R}_+;$$

$$\bigcap_{\varepsilon>0} \overline{\operatorname{co}}\left[(\varepsilon - \operatorname{argmin} J) + N(\operatorname{dom} J^*, 0)\right]$$

$$= \bigcap_{\varepsilon>0}\left[\overline{\operatorname{co}}(\varepsilon - \operatorname{argmin} J) + N(\operatorname{dom} J^*, 0)\right] = \mathbb{R}_+.$$

Note that $\bigcap_{\varepsilon>0} \overline{\operatorname{co}}\left(\varepsilon - \operatorname{argmin} J\right)$ is just $\{0\}$.
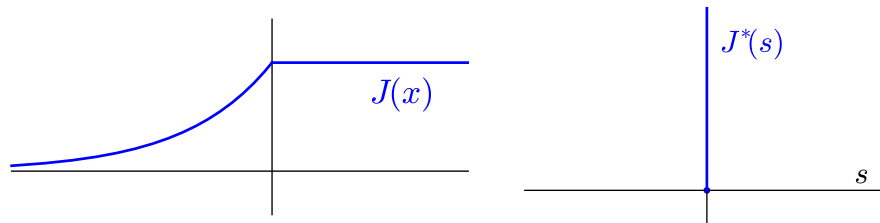
Example 2

J is not asymptotically epi-pointed. Since $\operatorname{dom} J^* = \{0\}$, $N(\operatorname{dom} J^*, 0) = \mathbb{R}$. We have $\overline{\operatorname{co}} J \equiv 0$ and:

$$\operatorname{argmin}(\overline{\operatorname{co}} J) = \mathbb{R};$$

$$\varepsilon - \operatorname{argmin} J + N(\operatorname{dom} J^*, 0) = \mathbb{R} \text{ for all } \varepsilon > 0.$$

Here again,

$$\bigcap_{\varepsilon>0} \overline{\operatorname{co}}\left(\varepsilon - \operatorname{argmin} J\right) = \{0\}.$$
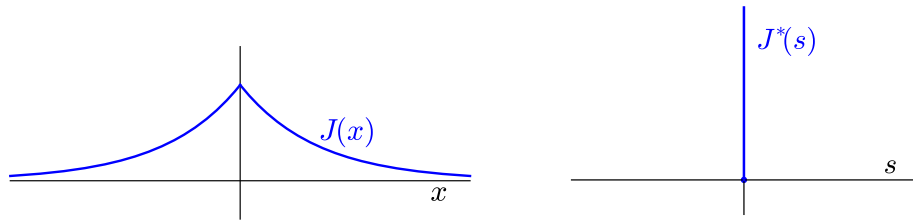
Example 3

For this case,

$$\varepsilon - \operatorname{argmin} J + N(\operatorname{dom} J^*, 0) = \mathbb{R} \ \textit{for all } \varepsilon > 0,$$

while:

$$\bigcap_{\varepsilon > 0} \overline{\operatorname{co}} \, (\varepsilon - \operatorname{argmin} J) = \emptyset;$$
$$\operatorname{argmin} J = \emptyset;$$
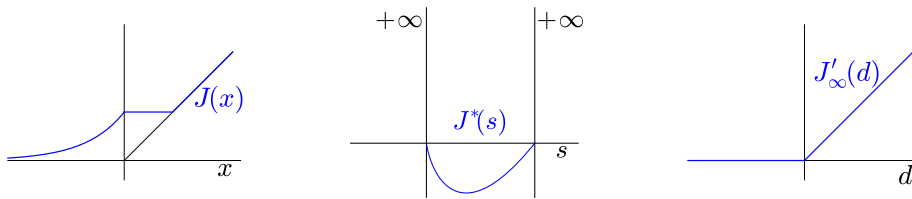$$\operatorname{argmin}(\overline{\operatorname{co}} \, J) = \mathbb{R}.$$

Example 4

We have $\overline{\operatorname{co}} \, J \equiv 0$ *and:*

$$\overline{\operatorname{co}} \, (\varepsilon - \operatorname{argmin} J) = \mathbb{R} \ \textit{for all } \varepsilon > 0;$$
$$\operatorname{argmin} J = \emptyset;$$
$$\operatorname{argmin}(\overline{\operatorname{co}} \, J) = \mathbb{R}.$$

In this particular instance, the contribution to the construction of $\operatorname{argmin}(\overline{\operatorname{co}} \, J)$ *of the behaviour of* J *"at infinity" is null.*
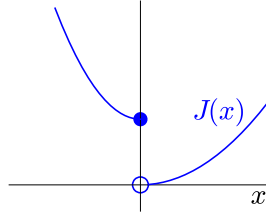
Example 5

Here, J *is asymptotically epi-pointed, but* $\operatorname{argmin} J = \emptyset$.
*Also,* $N(\operatorname{dom} J^*, 0) = \mathbb{R}_-$, *so that*

$$\bigcap_{\varepsilon > 0} \overline{\operatorname{co}} \, [(\varepsilon - \operatorname{argmin} J) + N(\operatorname{dom} J^*, 0)] =$$

$$= \bigcap_{\varepsilon > 0} [\overline{\operatorname{co}}(\varepsilon - \operatorname{argmin} J) + N(\operatorname{dom} J^*, 0)] = \emptyset = \operatorname{argmin}(\overline{\operatorname{co}} \, J).$$

Example 6 *All the examples shown above are with continuous J; but The-orem 6 and Theorems 7–8 can apply to discontinuous J: to lower-semi-continuous J for Theorem 6 (the lower-semicontinuity was a requirement from the beginning in §2.2.1); to arbitrary discontinuous J for Theorems 7–8. Here:*

J is not lower-semicontinuous;
$\operatorname{argmin} J = \emptyset$;
$0 \in \operatorname{int}(\operatorname{dom} J^*),$ *so that* $N(\operatorname{dom} J^*, 0) = \{0\}$;
$\bigcap_{\varepsilon > 0} \overline{\operatorname{co}}\,(\varepsilon - \operatorname{argmin} J) = \{0\} = \operatorname{argmin}(\overline{\operatorname{co}} J)$ *indeed.*

Example 7 (from [4]).
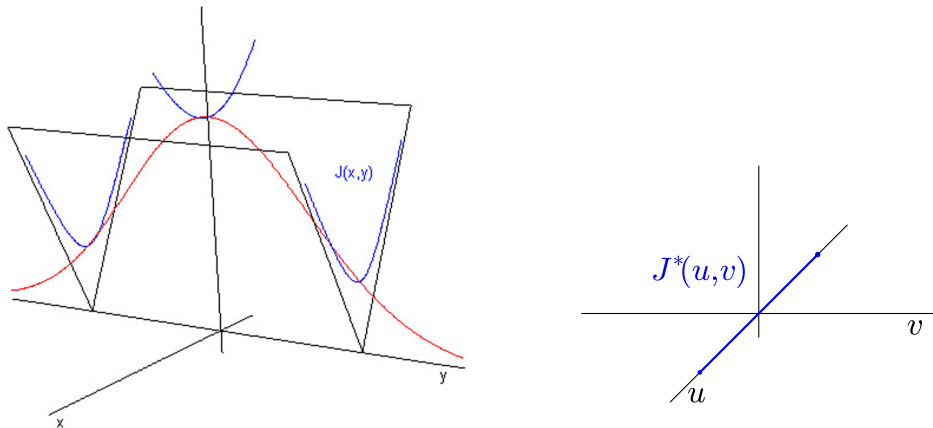
J is defined on $\mathbb{R}^2$ *by*

$$J(x,y) = \sqrt{x^2 + \exp\left(-y^2\right)}.$$

It is not asymptotically epi-pointed on $\mathbb{R}^2$. *We easily see that* J^* *is the indicator function of the segment* $[-1, 1] \times \{0\}$, *whence* $N(\operatorname{dom} J^*, 0)$ *is the line* $\{0\} \times \mathbb{R}$. *The set of* ε-*minimizers of J is a little bit complicated to delineate, but* $\overline{\operatorname{co}} J$ *is fairly easy:* $\overline{\operatorname{co}} J(x,y) = |x|$ *for* $(x,y) \in \mathbb{R}^2$.

Indeed, $\operatorname{argmin} J = \emptyset$ *and*

$$\{0\} \times \mathbb{R} = \operatorname{argmin}(\overline{\operatorname{co}} J) = \bigcap_{\varepsilon > 0} \overline{\operatorname{co}}\left[\varepsilon - \operatorname{argmin} J + \{0\} \times \mathbb{R}\right].$$

3. Applications to Hilbertian approximation-minimization problems

3.1. Applications to two Hilbertian approximation problems

3.1.1.

Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space, equipped with the associated norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$, and let S be a nonempty closed subset of H (nothing more is required on S). Many problems in Approximation, Optimization, Optimal control, etc. can be formulated in an abstract form like this: given $a \in H$, find (if any!) $\overline{x} \in S$ closest to a (the proximity is understood in the sense of the Hilbertian norm $\| \cdot \|$). For that, let us introduce some notations and definitions:

- $a \in H \mapsto d_S(a) := \inf\{\|x - a\| \mid x \in S\}$, the distance function to S;

- $a \in H \mapsto P_S(a) := \{x \in S \mid d_S(a) = \|x - a\|\}$, the projection set-valued mapping on S.

Since $P_S(a)$ might be empty, we also consider, for $\varepsilon > 0$,

$$(24) \qquad P_S^\varepsilon(a) := \{x \in S \mid \|x - a\| \leq d_S(a) + \varepsilon\}.$$

This collection of sets will play a key-role below, as expected if we keep in mind the general thread of the present paper.

One of our favorite functions in such a context is the ASPLUND function φ_S defined on H as follows:

$$(25) \qquad x \in H \mapsto \varphi_S(x) := \frac{1}{2}[\|x\|^2 - d_S^2(x)].$$

This function, finite and continuous on H, turns out to be *convex* whatever S be (convex or not, connected or not, discrete or not). If S is convex, a classical result from Convex analysis in Hilbert spaces is that

$$(26) \qquad \partial \varphi_S(x) = \{\nabla \varphi_S(x)\} = \{p_S(x)\}, \quad \text{for all } x \in S,$$

where $p_S(x)$ stands here for the unique element in $P_S(x)$. The relation between the assumption "$P_S(x)$ is single-valued for all $x \in H$" and the one "S is convex" is partially surveyed in [18].

Extending (26) to the nonconvex case (i.e., when S is not necessarily convex) is hopeless in general, especially if H is not finite-dimensional. One simply has:

$$(27) \qquad \overline{\text{co}}\, P_S(x) \subset \partial \varphi_S(x) \quad \text{for all } x \in H,$$

a weak result indeed if $P_S(x)$ is empty!

There however is a result in Approximation theory, due to BERENS (see [5, 14]) which says the following:

$$(28) \qquad \partial\varphi_S(x) = \bigcap_{\varepsilon>0} \overline{\mathrm{co}}\, P_S^\varepsilon(x) \quad \text{for all } x \in H.$$

It is interesting to note, from the pedagogical viewpoint (paradox in infinite dimensional Hilbert spaces), the following differences:

$$\overline{\mathrm{co}}\, P_S(x) = \overline{\mathrm{co}} \bigcap_{\varepsilon>0} P_S^\varepsilon(x) \text{ may be empty, while}$$
$$\bigcap_{\varepsilon>0} \overline{\mathrm{co}}\, P_S^\varepsilon(x) \text{ is (always) nonempty and weakly compact.}$$

We shall interpret and recover (28) in the context of our results. For that, given a fixed $a \in H$, define

$$a \in H \mapsto f_{S,a}(x) := \begin{cases} \frac{1}{2}\|x - a\|^2 \text{ if } x \in S, \\ +\infty \text{ if not.} \end{cases}$$

($f_{S,a} = \frac{1}{2}\| \cdot -a\|^2 + i_S$ in short). This function $f_{S,a}$ is not convex (except if S is convex), and we shall consider its relaxed form $\overline{\mathrm{co}}\, f_{S,a} = (f_{S,a})^{**}$. We get some information about this relaxed function $\overline{\mathrm{co}}\, f_{S,a}$ since we are able to calculate the LEGENDRE-FENCHEL conjugate of $f_{S,a}$. Indeed, for $p \in H$,

$$\begin{aligned} f_{S,a}^*(p) &= \sup_{x\in S} \left[\langle p, x \rangle - \frac{1}{2}\|x - a\|^2 \right] \\ &= \sup_{x\in S} \left[\langle p + a, x \rangle - \frac{1}{2}\|x\|^2 \right] - \frac{1}{2}\|a\|^2 \\ &\quad \dots \text{ (some easy calculations)} \\ &= \varphi_S(p + a) - \frac{1}{2}\|a\|^2 \quad \text{(the } \varphi_S \text{ function pops out!)} \end{aligned}$$

As a first consequence,

$$(29) \qquad \partial f_{S,a}^*(0) = \partial\varphi_S(a).$$

Moreover,

- $\overline{x} \in \operatorname{argmin}\overline{\mathrm{co}}\, f_{S,a} \;\Leftrightarrow\; \overline{x} \in \partial f_{S,a}^*(0)$
  $\qquad\qquad\qquad\qquad\quad \Leftrightarrow\; \overline{x} \in \partial\varphi_S(a) \quad \text{(according to (29));}$

- $\overline{x} \in \operatorname{argmin}\overline{\mathrm{co}}\, f_{S,a} \;\Leftrightarrow\; \overline{x} \in \bigcap_{\varepsilon>0} \overline{\mathrm{co}}\, (\varepsilon - \operatorname{argmin} f_{S,a})$
  $\qquad\qquad\qquad\qquad\quad \text{(according to Theorems 7-8).}$

As a final result,

$$\partial\varphi_S(a) = \bigcap_{\varepsilon>0} \overline{\text{co}} \left\{ x \in S \mid \frac{1}{2}\|x-a\|^2 \leq \frac{1}{2}d_S^2(a) + \varepsilon \right\}$$
$$= \bigcap_{\alpha>0} \overline{\text{co}} \left\{ x \in S \mid \|x-a\| \leq d_S(a) + \alpha \right\},$$

which is exactly Berens' result (28).

3.1.2.

The situation considered here is parallel to the one in the previous subsection. Let S be a bounded closed subset of H. We define:

- $a \in H \mapsto \Delta_S(a) := \sup\{\|x-a\| \mid x \in S\}$, the farthest distance function to S;

- $a \in H \mapsto Q_S(a) := \{x \in S \mid \Delta_S(a) = \|x-a\|\}$, the farthest set-valued mapping on S.

We also consider, for $\varepsilon > 0$,

$$(30) \qquad\qquad Q_S^\varepsilon(a) := \{x \in S \mid \|x-a\| \geq \Delta_S(a) - \varepsilon\}.$$

Our favorite function in the present context is what we call the second Asplund function ψ_S, defined on H as follows:

$$x \in H \mapsto \psi_S(x) := \frac{1}{2} \left[\Delta_S^2(x) - \|x\|^2 \right].$$

This function, finite and continuous on H, is *convex* whatever S be (the function Δ_S^2 is strongly convex on H, with 1 as a modulus of strong convexity). The relation between the assumption "$Q_S(x)$ is single-valued for all $x \in H$" and the one "S is a singleton" is surveyed in [19].

As a general rule, we have the inclusion

$$(31) \qquad\qquad -\overline{\text{co}}\, Q_S(x) \subset \partial\psi_S(x) \text{ for all } x \in H;$$

the equality holds true when S is compact (and the closing operation is unnecessary if H is finite-dimensional). The next result sharpens a similar one proved in [19, Proposition 3.5].

Proposition 9 *For all $a \in H$,*

$$(32) \qquad \partial \psi_S(a) = -\bigcap_{\varepsilon > 0} \overline{\mathrm{co}}\, Q_S^\varepsilon(a),$$

and

$$(33) \qquad \partial\left(\frac{1}{2}\Delta_S^2\right)(a) = a - \bigcap_{\varepsilon > 0} \overline{\mathrm{co}}\, Q_S^\varepsilon(a).$$

Proof. Given $a \in H$, define

$$x \in H \mapsto g_{S,a}(x) := \begin{cases} -\frac{1}{2}\|x + a\|^2 \text{ if } x \in -S, \\ +\infty \text{ if not.} \end{cases}$$

($g_{S,a} = -\frac{1}{2}\| \cdot + a\|^2 + i_{-S}$ in short.)

The function $g_{S,a}$ is not convex (except in the very particular case where S is a singleton); we therefore consider its relaxed version $\overline{\mathrm{co}}\, g_{S,a} = (g_{S,a})^{**}$. More is known about this relaxed form via the LEGENDRE-FENCHEL conjugate of $g_{S,a}$. Actually, for $p \in H$,

$$\begin{aligned} g_{S,a}^*(p) &= \sup_{x \in -S}\left[\langle p, x\rangle + \frac{1}{2}\|x + a\|^2\right] \\ &= \sup_{x \in -S}\left[\langle p + a, x\rangle + \frac{1}{2}\|x\|^2\right] + \frac{1}{2}\|a\|^2 \\ &\quad \dots \text{ (some easy calculations)} \\ &= \psi_S(p + a) + \frac{1}{2}\|a\|^2. \end{aligned}$$

This conjugacy result is one reason why the function ψ_S was introduced. The convexity of ψ_S is easily derived since $\psi_S = g_{S,0}^*$.

As a consequence of the result above, we have:

$$(34) \qquad \partial g_{S,a}^*(0) = \partial \psi_S(a).$$

Now,

- $\overline{x} \in \operatorname{argmin} \overline{\mathrm{co}}\, g_{S,a} \iff \overline{x} \in \partial g_{S,a}^*(0)$
  $\iff \overline{x} \in \partial \psi_S(a)$ (according to (34));

- $\overline{x} \in \operatorname{argmin} \overline{\mathrm{co}}\, g_{S,a} \iff \overline{x} \in \bigcap_{\varepsilon > 0} \overline{\mathrm{co}}\, (\varepsilon - \operatorname{argmin} g_{S,a})$
  (according to Theorems 7-8).

To get the formula (32), it remains to observe that:

$$\varepsilon - \operatorname{argmin} g_{S,a} = -\left\{ x \in S \mid \frac{1}{2}\|x - a\|^2 \geq \frac{1}{2}\Delta_S^2(a) - \varepsilon \right\};$$

hence

$$\bigcap_{\varepsilon > 0} \overline{\operatorname{co}} \left\{ x \in S \mid \frac{1}{2}\|x - a\|^2 \geq \frac{1}{2}\Delta_S^2(a) - \varepsilon \right\} =$$
$$= \bigcap_{\alpha > 0} \overline{\operatorname{co}} \left\{ x \in S \mid \|x - a\| \geq \Delta_S(a) - \alpha \right\}.$$

Finally, a standard subdifferential calculus rule on the sum of two convex functions gives:

$$\partial\left(\frac{1}{2}\Delta_S^2\right)(a) = \partial\psi_S(a) + a,$$

which is exactly (33). ∎

3.2. Extension to a general framework

The two examples in the previous subsection 3.1 can be cast in the same general framework. Still in a Hilbert space setting $(H, \langle \cdot, \cdot \rangle)$, consider

$$\begin{cases} \text{a function } f : H \to \mathbb{R} \cup \{+\infty\}, \\ \text{a nonempty subset } S \text{ of } H, \\ \text{an element } a \in H. \end{cases}$$

With these data is associated the following variational problem:

$$(\mathcal{P}) \qquad \textit{Minimize } f(x) - \langle a, x \rangle \textit{ over } S.$$

We assume that f is finite at some point of S and that the "tilted" version $f - \langle a, \cdot \rangle$ of f is bounded from below on S. With the notations used in section 2, $X = H$ and $J = f + i_S - \langle a, \cdot \rangle$. We denote by $M_\varepsilon(f, S, a)$ the set of ε-minimizers in $(\mathcal{P})$, just $M(f, S, a)$ for the set of exact minimizers.

As one can imagine, many variational problems can be moulded in the $(\mathcal{P})$-format. For example, we recover the case treated in Subsection 3.1.1 by considering

$$f(x) := \frac{1}{2}\|x\|^2 + \frac{1}{2}\|a\|^2$$

(so that $f + i_S - \langle a, \cdot \rangle$ is what we called $f_{S,a}$ there), while the case treated in Subsection 3.1.2 corresponds to

$$f(x) := -\frac{1}{2}\|x\|^2 - \frac{1}{2}\|a\|^2$$

and S changed into $-S$ (so that $f + i_{-S} - \langle a, \cdot \rangle$ is what we called $g_{S,a}$ there). Even some variational principles are formulated in the $(\mathcal{P})$ format; for example, the STEGALL minimization principle [8, p. 43] says the following: assuming that f is lower-semicontinuous and bounded from below on S, and that S is bounded and closed, there exists a dense set of points a in H having the property that $M(f, S, a)$ is single-valued.

Since $\overline{\mathrm{co}} \, (f + i_S - \langle a, \cdot \rangle)$ is just $\overline{\mathrm{co}} \, (f + i_S) - \langle a, \cdot \rangle$ (one of the rare cases where the $\overline{\mathrm{co}}$ operation behaves nicely with the addition), the relaxed version of $(\mathcal{P})$ is as follows:

$$(\mathcal{P}^{relax}) \qquad Minimize \; \overline{\mathrm{co}} \, (f + i_S) \, (x) - \langle a, x \rangle \; over \; H.$$

We correspondingly denote by $M^{\text{relax}}(f, S, a)$ the solution set in $(\mathcal{P}^{\text{relax}})$. In everything linking both problems $(\mathcal{P})$ and $(\mathcal{P}^{\text{relax}})$, a key-role will be played by the function

$$(35) \qquad\qquad\qquad \theta_S := (f + i_S)^*,$$

a generalization of ASPLUND's functions in subsections 3.1.1 and 3.1.2. To begin with, it is easy to observe that

$$(36) \qquad\qquad \overline{\mathrm{co}} \, M(f, S, a) \subset M^{\text{relax}}(f, S, a) = \partial \theta_S(a).$$

Equality holds true in (36) if some topological properties are imposed on f and S.

Proposition 10 *Assume that S is compact and that f is finite and lower-semicontinuous on S. Then, for all $a \in H$, we have:*

$$(37) \qquad\qquad \emptyset \neq \partial \theta_S(a) = M^{relax}(f, S, a) = \overline{\mathrm{co}} \, M(f, S, a).$$

Proof. Just observe that θ_S can be put in the sup form as

$$\theta_S = \sup_{x \in S} \left[\langle x, \cdot \rangle - f(x) \right],$$

and apply the (most) general formula on the subdifferential of the supremum of a collection of convex functions [25, p. 97]. $\blacksquare$

With the inclusion (36) we also deduce the following:

Proposition 11 *Assume θ_S is GÂTEAUX-differentiable on H. Then, for any $a \in H$, $M(f, S, a)$ is either empty or reduces to a singleton. If, moreover, S is compact and f is finite and lower-semicontinuous on S, then $M(f, S, a)$ is a singleton for all $a \in H$.*

We now formulate a general result allowing to recover $M^{\mathrm{relax}}(f, S, a)$ by "filtering" the sets of ε-minimizers $M_\varepsilon(f, S, a)$, $\varepsilon > 0$. This is just an application of Theorems 7-8.

Proposition 12 *Assume that θ_S is finite-valued on H (or just $\mathbb{R}_+ \operatorname{dom} \theta_S = H$). Then, for all $a \in H$,*

$$(38) \qquad M^{\mathrm{relax}}(f, S, a) = \bigcap_{\varepsilon > 0} \overline{\mathrm{co}}\, M_\varepsilon(f, S, a).$$

References

[1] Attouch, H., Buttazzo, G. and Michaille, G.: *Variational analysis in Sobolev and BV spaces. Applications to PDEs and optimization.* MPS/SIAM Series on Optimization **6**. Society for Industrial and Applied Mathematics (SIAM); Mathematical Programming Society (MPS), Philadelphia, PA, 2006.

[2] Azé, D.: *Éléments d'analyse convexe et variationnelle.* Editions Ellipses, Paris, 1997.

[3] Beer, G. and Lucchetti, R.: Convex optimization and the epi-distance topology. *Trans. Amer. Math. Soc.* **327** (1991), 795–813.

[4] Benoist, J. and Hiriart-Urruty, J.-B.: What is the subdifferential of the closed convex hull of a function? *SIAM J. Math. Anal.* **27** (1996), no. 6, 1661–1679.

[5] Berens, H.: Best approximations in Hilbert space. In *Approximation theory, III (Proc. Conf., Univ. Texas, Austin, Tex., 1980)*, 1–20. Academic Press, New York-London, 1980.

[6] Brighi, B.: Sur l'enveloppe convexe d'une fonction de la variable réelle. *Revue de Mathématiques Spéciales, Vuibert*, 104ème année, (1994), no. 8, 547–550.

[7] Brighi, B. and Chipot, M.: Approximated convex envelope of a function. *SIAM J. Numer. Anal.* **31** (1994), no. 1, 128-148.

[8] Clarke, F. H., Ledyaev, Yu. S., Stern, R. J. and Wolenski, P. R.: *Nonsmooth analysis and control theory.* Graduate Texts in Mathematics **178**. Springer-Verlag, New York, 1998.

[9] Crandall, M. G. and Lions, P. L.: Viscosity solutions of Hamilton-Jacobi equations. *Trans. Amer. Math. Soc.* **277** (1983), no. 1, 1–42.

[10] DACOROGNA, B.: *Direct methods in the calculus of variations*. Applied Mathematical Sciences **78**. Springer, New York, 2008.

[11] DEDIEU, J.-P.: *Une condition nécessaire et suffisante d'optimalité en optimisation non convexe et en calcul des variations*. Séminaire d'Analyse numérique, Université Paul Sabatier (1979-1980).

[12] EKELAND, I.: Nonconvex minimization problems. *Bull. Amer. Math. Society (N.S.)* **1** (1979), 443–474.

[13] EKELAND, I. AND TEMAM, R.: *Convex analysis and variational problems*. Classics in Applied Mathematics **28**. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999.

[14] GODINI, G.: On a problem of H. Berens. *Math. Ann.* **263** (1983), no. 3, 279–281.

[15] HANTOUTE, A., LÓPEZ, M. A. AND ZĂLINESCU, C.: Subdifferential calculus rules in convex analysis: a unifying approach via pointwise supremum functions. *SIAM J. Optim.* **19** (2008), 863–882.

[16] HIRIART-URRUTY, J.-B.: Lipschitz r-continuity of the approximate subdifferential of a convex function. *Math. Scan.* **47** (1980), 123–134.

[17] HIRIART-URRUTY, J.-B.: When is a point x satisfying $\nabla f(x) = 0$ a global minimum of f? *Amer. Math. Monthly* **93** (1986), no. 7, 556–558.

[18] HIRIART-URRUTY, J.-B.: Ensembles de Tchebychev vs. ensembles convexes: l'état de la situation vu via l'analyse convexe non lisse. *Ann. Sci. Math. Québec* **22** (1998), 47–62.

[19] HIRIART-URRUTY, J.-B.: La conjecture des points les plus éloignés revisitée. *Ann. Sci. Math. Québec* **29** (2005), 197–214.

[20] HIRIART-URRUTY, J.-B. AND LEMARECHAL, C.: *Convex analysis and minimization algorithms*. Grudlehren der mathematischen Wissenschaften **305** and **306**. Springer-Verlag, Berlin Heidelberg, 1993. Second printing in 1996.

[21] HIRIART-URRUTY, J.-B., MOUSSAOUI, M., SEEGER, A. AND VOLLE, M.: Subdifferential calculus without qualification conditions, using approximate subdifferentials: a survey. *Nonlinear Anal.* **24** (1995), no. 12, 1727–1754.

[22] HIRIART-URRUTY, J.-B. AND PHELPS, R. R.: Subdifferential calculus using ε-subdifferentials. *J. Funct. Anal.* **118** (1993), no. 1, 154–166.

[23] LÓPEZ, M. A. AND VOLLE, M.: A formula for the set of optimal solutions of a relaxed minimization problem. Applications to subdifferential calculus. *J. Convex Analysis* **17** (2010), no. 3-4, 1057–1075.

[24] LUCET, Y.: What shape is your conjugate? A survey of computational convex analysis and its applications. *SIAM J. Optim.* **20** (2009), no. 1, 216–250.

[25] Zălinescu, C.: *Convex analysis in general vector spaces.* World Scientific Publishing, River Edge, NJ, 2002.

Jean-Baptiste Hiriart-Urruty
Institut de Mathématiques
Université Paul Sabatier
118, route of Narbonne
31062 Toulouse Cedex 9, France
jbhu@cict.fr

Marco A. López
Dep. of Statistics and Operations Research
Faculty of Sciences, Alicante University
Ctra. San Vicente del Raspeig s/n
03071 Alicante, Spain
marco.antonio@ua.es

Michel Volle
Université d'Avignon
Laboratoire de Mathématiques
33, rue Louis Pasteur
84000 Avignon, France
michel.volle@univ-avignon.fr