# EPIGRAPHICAL ANALYSIS

H. ATTOUCH
*Université de Perpignan, 66025 Perpignan-Cedex, France*

R. J.-B. WETS*
*University of California-Davis, CA 95616*

**Abstract.** We provide a succinct overview of the basic tools of epigraphical analysis, and of the accompanying calculus. We show that it provides a very rich and unified tool to study a large class of problems that includes variational problems, generalized equations, differential inclusions and limit problems.

* This research was supported in part by the National Science Foundation.

## 1. Introduction

The study of variational problems has been much aided by casting them in the following (simple) framework:

$$\text{find } x \text{ that minimizes } f(x) \text{ on } X,$$

where $X$ is the underlying space and f takes its values in $(-\infty,\infty]$; the value $\infty$ being used to model the constraints that may be imposed on the choice of $x$. Probably starting with the work of Fenchel (on convex functions), it has become more and more obvious that the properties of such problems (stability, dualizability,...) and the properties of their solutions, depend intimately on the properties of the *epigraph* of $f$. This is in contrast to the "classical" approach that would be mostly concerned with the properties of the *graph* of $f$. The analysis via epigraphs requires a number of (nonclassical) tools, some of which have been brought to the fore during the last couple of decades and others that are waiting further development. It is the purpose of this article to introduce the basic tools of epigraphical analysis. A complete treatment of the subject is far beyond the scope of this contribution, it would require a monograph of some length. Indeed one would need to broach such topics as: epigraphical calculus (algebraic operations), epi-convergence (and associated topological questions), epi-derivatives and epi-integration, as well as discuss existing and potential applications in a variety of areas of pure and applied analysis. We shall limit ourselves to a brief survey concentrating on elementary calculus rules and definitions of limits. We also conclude this introduction by giving a series of examples that suggest the wide applicability of (and the need for) epigraphical analysis.

We consider problems of the type (for which Rockafellar and Wets [**39**] use the term *variational system*):

$$\text{minimize } f(x,\theta) \text{ for } x \in X,$$

where $(x,\theta) \mapsto f(x,\theta) : X \times \Theta \to \overline{\mathbb{R}}$. The variable $x$ representing the decision variables (state, control, etc.) with $X$ of finite or infinite (distributed systems, stochastic problems, for example) dimensions. The space of parameters $\Theta$ could arise from perturbations (parametric optimization), approximations (numerical solutions procedures, etc.), or could correspond to physical characteristics that are intrinsic to the problem at hand. To each value of $\theta$ correspond: the value of the infimum,

$$(\inf f)(\theta) := \inf[f(x,\theta),\ x \in X],$$

and the set of (optimal) solutions of the problem, and the multifunction (set-valued map) of solutions of the problem,

$$\operatorname{argmin} f(\cdot,\theta) := \{x \in X \mid f(x,\theta) \le \inf f(\theta)\}.$$

We are interested in the continuity and differentiability properties of the *infimal function*, $\theta \mapsto \inf f(\theta)$, also called the marginal function or the value function, and of the argmin-multifunction (set-valued map) $\theta \mapsto \operatorname{argmin} f(\theta)$, and, when differentiable or subdifferentiable, in computing their derivatives.

As indicated earlier, the epigraphical approach consists in deriving the properties of those maps from those of the multifunction

$$\theta \mapsto \operatorname{epi} f(\cdot,\theta) := \{(x,\alpha) \in X \times \mathbb{R} \mid \alpha \geq f(x,\theta)\},$$

i.e., an epigraph-valued multifunction, called the *epigraphical multifunction.*

Epigraphical analysis can be regarded as a specialization of results about multifunctions to the map $\theta \mapsto \operatorname{epi} f(\cdot,\theta)$. And in many ways this is an appropriate viewpoint, not withstanding the fact that the study of epigraphical multifunctions actually provides much of the motivation for the study of multifunctions. However, because of the special nature of epigraphs, of the questions raised in a functional setting, and in particular because of the framework forced upon us by the applications, it is useful and instructive to have a theory specifically developed for, and directly applicable to, epigraphical multifunctions. In this we are helped by the fact that the subspace of epigraphical multifunctions is closed with respect to the addition of their defining functions, epi-addition (Minkowski-addition of the epigraphs), differentiation, taking (epi-)limits, integration, etc., i.e., all the basic analytical operations.

To conclude this introduction, let us consider a few typical situations. We give a couple examples of each one of the following cases: $\theta$ is an approximation parameter (**A.**), $\theta$ is a perturbation parameter(**B.**), and $\theta$ represents some physical characteristics of the problem(**C.**).

**Example A.1 Stochastic optimization.** The problem is to minimize $f(x) = \int_{\Xi} h(x,\xi)P(d\xi)$ where $P$ is a probability measure on $\Xi \subset \mathbb{R}^N$, and $h$ is an extended real-valued function defined on $X \times \Xi$. For example,

$$h(x,\xi) = \inf[\, g(x,y,\xi) \mid y \in D(x,\xi)],$$

where $D(x,\xi)$ represents the constraints as they depend on $x$ and on the random elements of the problem $\xi$. This is a difficult problem to solve numerically because of the lack of regularity of $f$ (in general, $f$ takes on the value $+\infty$, and is not smooth on its effective domain), and the lack of techniques and calculus for dealing effectively with multi-dimensional integration. One is lead to consider approximations of the type:

$$\text{minimize} \quad \int_{\Xi} h^{\theta}(x,\xi)P_{\theta}(d\xi)$$

with the $h^\theta$ converging to $h$ and the $P_\theta$ (usually discrete) measures approximating $P$. The parameter space $\Theta = \mathbb{N}$. We want to know: *(i)* if the solutions of the approximating problems converge to the solution of the given problem, and *(ii)* at what rate (error bounds).

**Example A.2 Finite elements approximations.** Here $X$ is a Sobolev space of type $H^1(\Omega)$ or $H^1_o(\Omega)$, and

$$f(u) = \tfrac{1}{2}a(u,u) - \langle h, u\rangle$$

where $a$ is a symmetric, continuous, coercive bilinear form on $X \times X$. Let $u^* \in \operatorname{argmin} f$. The numerical procedures for calculating $u^*$ rely on the following approximations. The space $X$ is replaced by a finite dimensional subspace $X_\theta$ obtained by taking linear combinations of a finite number of "elements". The approximating problems take the form:

$$\text{minimize } [\tfrac{1}{2}a(u,u) - \langle h, u\rangle \mid u \in X_\theta].$$

We are interested in the convergence of the approximating solutions $u^\theta$ to $u^*$, and in estimating $\|u^\theta - u^*\|$. (Note that, this latter estimate will depend on the regularity of $u^*$.)

**Example B.1 Mathematical programming.** Our point of departure is the optimization problem:

$$\begin{aligned} \text{minimize}\quad & f_0(x) \\ \text{subject to}\quad & f_i(x) \le 0, \qquad i = 1,\dots,s; \\ & f_i(x) = 0, \qquad i = s+1,\dots,m; \\ & x \in \mathbb{R}^n. \end{aligned}$$

A very useful characterization of optimality is provided by the differential inclusion:

$$0 \in \partial f_0(x) + \sum y_i \partial f_i(x),$$

where the $(y_i,\ i = 1,\dots,m)$ are (Lagrange) multipliers associated with the constraints. The existence of such multiplers, and thus the possibility of making use of the preceding relation, is intimately connected with the properties of the infimal function of the (variational) system obtained by parametrizing the constraints, namely:

$$f(x,\theta) = \begin{cases} f_0(x) & \text{if } f_i(x) \le \theta_i, \quad i = 1,\dots s \\ & \text{and } f_i(x) = \theta_i, \quad i = s+1,\dots,m, \\ \infty & \text{otherwise.} \end{cases}$$

In particular, those multipliers can often be identified with the subgradients of the infimal function at $\theta = 0$. Constraint qualifications can be viewed as providing sufficient (sometimes necessary) conditions to guarantee continuity, or subdifferentiability of $\theta \mapsto \inf f(\theta)$ at 0, or in a neighborhood of 0.

**Example B.2 Duality in convex optimization.** The problem is as in Example B.1, assuming now that the functions $f_i$ are convex for $i = 0, \ldots, s$, and affine for $i = s+1, \ldots, m$, and $X$ is any normed linear space. As in Example B.1, the problem is embedded in a family of parametrized problems, with the same type of perturbations or by perturbing other quantities, but in such way that the function $(x, \theta) \mapsto f(x, \theta)$ is convex on $X \times \Theta$ with $\Theta$ another normed linear space. The dual problem is the conjugate (Legendre-Fenchel transform) of the infimal function, i.e.,

$$\text{minimize } [g(v) = \sup\{\langle v, \theta\rangle - (\inf f)(\theta) \mid \theta \in \Theta\} \text{ for } v \in V].$$

The bilinear form $\langle\cdot,\cdot\rangle$ brings the spaces $V$ and $\Theta$ in duality. Let us consider the following simple calculus of variations problem: for $\Omega$ a subset of $\mathbb{R}^n$,

$$\text{minimize } \frac{1}{2}\int_\Omega a(s)|\text{grad} u(s)|^2 ds - \int_\Omega hu\, ds \text{ for } u \in H_0^1(\Omega).$$

One, of a number of possible dual problems is obtained by perturbing the integrand by an $L^2$ function $p$:

$$\text{minimize } \frac{1}{2}\int_\Omega a(s)|\text{grad} u(s) + p|^2 ds - \int_\Omega hu\, ds \text{ for } u \in H_0^1(\Omega).$$

The dual problem is then:

$$\text{minimize } \frac{1}{2}\int_\Omega |p^*(s)|^2 ds \text{ with } p^* \in L^2(\Omega)^n, \text{div}\, p^* = h.$$

This approach to duality can be combined with such results as the continuity (with respect to the epi-topology) of the Legendre-Fenchel transform to derive stability and convergence properties for dual variables, cf. Back [14], [15], Attouch, Azé and Wets [4].

**Example B.3 Dynamic programming.** We consider the following system: for $t \in [0, T]$,

$$dx = h(t, x, u)dt, \; x(0) = x_0 \text{ with } u(t) \in U(t), \tag{B.3}$$

where $u$ is the control function and $x$ is the trajectory of the system. We are interested in choosing $u$ so as to

$$\text{minimize } g(x(T)) \text{ for } x(\cdot) \text{ satisfying (B.3)}.$$

One can reformulate (B.3) as a differential inclusion:

$$x'(t) \in F(t, x(t)), \text{ for } F(x, t) = \{f(t, x, u) \mid u \in U(t)\}.$$

The dynamic programming method associates to the preceding dynamical system the value function

$$v(t,x) = \inf\{g(x(T)) \mid x'(s) \in F(s,x(s)) \text{ on } [t,T], x(t) = x\}.$$

If $v$ can be calculated it is relatively easy to derive the optimal control $u^*$ that yields the optimal trajectory. When $v$ is sufficiently smooth, one way of computing $v$ is to solve the Hamilton-Jacobi equation

$$\frac{\partial v}{\partial t} = H(t,x,-\frac{\partial v}{\partial x}) \text{ with } H(t,x,p) = \sup\{\langle p,y\rangle \mid y \in F(t,x)\}.$$

If $v$ is not sufficiently smooth, we need to rely on nonsmooth analysis to give an interpretation to the preceding equation as has been done by introducing viscosity solutions , cf. Crandall, Evans and Lions [**22**], or contingent derivatives, cf. Frankowska [**27**].

**Example C.1 Homogenization.** We are dealing with composite materials, or materials with many small holes, fibered or stratified materials. In such materials, the *physical parameters* (such as conductivity, elasticity coefficients, etc.) are discontinuous, and oscillate rapidly between the different values that characterize the various components. A variational formulation of such problems takes the following form:

$$\text{minimize } \{\int_\Omega j\left(\frac{x}{\theta}\text{grad}u(x)\right)dx - \int_\Omega hu\,dx \mid u \in H_0^1(\Omega)\},$$

where $j$ is $Y$-periodic, $Y = \prod_{i=1}^n (a_i, b_i]$, and $\theta$ is a parameter near 0. The "homogenized" problem

$$\text{minimize } \left\{\int_\Omega j_{\text{hom}}(\text{grad}u(x))dx - \int_\Omega hu\,dx \mid u \in H_0^1(\Omega)\right\},$$

is obtained by letting the parameter tend to zero and taking a variational limit (an epi-limit, cf. Section 3) of the energy functionals, in particular,

$$j_{\text{hom}}(z) = \inf\left[\int_Y j(y, \text{grad}w(y) + z)dy \text{ with } w \;\; Y\text{-periodic}\right].$$

The calculation of $j_{\text{hom}}$ again requires solving a variational system (that depends on a parameter $z$).

**Example C.2 Optimum design.** Consider the following situation: two media $\Omega_1$ and $\Omega_2$ are separated by an isolating screen $\Sigma$ (a smooth manifold of co-dimension 1) with "holes" $D = \cup D_i$ that allows for the transfer of heat. With heat source $h$, the state equations of the system are:

$$-\Delta y = h \text{ on } \Omega_1 \cup \Omega_2 \cup D,$$

$$\frac{\partial y}{\partial n} = 0 \text{ on } \Sigma \setminus D,$$

with the boundary conditions determined by the temperature $g$, i.e., $y = g$ on the boundary of $\Omega = \Omega_1 \cup \Omega_2 \cup \Sigma$. The problem is to "design" $D$ so that the resulting solution $y_D$ of the preceding system is as close as possible to a desired state $z(\cdot)$. The performance criterion is then:

$$J(D) = \int_\Omega |y_D(x) - z(x)|^2 dx.$$

A more sophisticated version of this problem would allow for a control mechanism to regulate the exchanges between $\Omega_1$ and $\Omega_2$, cf. Attouch and Picard [5]. In general, such problems do not have an optimal solution. This comes from the an "homogenization" phenomenon: for given $J$ and a class of admissible designs, the sets $\{D_\nu\ \nu \in \mathbb{N}\}$ of a minimizing sequences could tend to be more and more fragmented (clouds of tiny holes). Epi-convergence is used to study the limiting behavior and to characterize the limiting "problems".

## 2. Epigraphical Operations.

Let $X$ be a vector space. To any extended real-valued function $f : X \to \overline{\mathbb{R}} = [-\infty, \infty]$, we can associate

$$\operatorname{epi} f = \{(x, \alpha) \mid \alpha \geq f(x)\},$$

the *epigraph* of $f$, and

$$\operatorname{epi}_s f = \{(x, \alpha) \mid \alpha > f(x)\},$$

the *strict epigraph* of $f$; these sets are empty if $f \equiv \infty$. The classical notions of (Minkowski) sum of sets

$$C + D = \{x + y \mid x \in C,\ y \in D\},$$

and scalar multiplication of sets, for $\lambda \in \mathbb{R}$,

$$\lambda C = \{\lambda x \mid x \in C\},$$

when applied to epigraphs generate the epigraphs of new functions that are called the epi-sum and the ($\lambda$-)epi-multiple. In particular, if we have two functions $f$ and $g$, defined on $X$ and with values in the extended reals, the (Minkowski) addition

$$\operatorname{epi}_s f + \operatorname{epi}_s g =: \operatorname{epi}_s (f \underset{e}{+} g) \tag{2.1}$$

define the strict epigraph of a function that we denote by $f \underset{e}{+} g$ and call the *epi-sum* of $f$ and $g$. The subscript "e" referring to the fact that the operation takes place on epigraphs. A functional definition is given by the next relation:

$$\begin{aligned} (f \underset{e}{+} g)(x) &= \inf\{f(u) + g(v) \mid u + v = x\} \\ &= \inf\{f(u) + g(x - u) \mid u \in X\}; \end{aligned} \tag{2.2}$$

we use here the (epigraphical) convention that $\infty - \infty = \infty$. If, instead of strict epigraphs, we want to relate the epi-sum of $f$ and $g$ to the sum of the epigraphs, we have the following identity:

$$\mathrm{epi}(f \underset{e}{+} g) = \mathrm{vcl}(\mathrm{epi}\, f + \mathrm{epi}\, g), \tag{2.3}$$

where vcl means vertical closure: for a set $C \subset X \times \mathbb{R}$,

$$\mathrm{vcl}\, C := \{(x, \alpha) \mid \alpha \geq \inf[\mu \mid (x, \mu) \in C]\}.$$

Similarly, we define the (scalar) *epi-multiplication* $\underset{e}{*}$ as follows: for $x$ in $X$, $\lambda > 0$

$$(\lambda \underset{e}{*} f)(x) := \lambda f(\lambda^{-1} x) \tag{2.4}$$

since then

$$\mathrm{epi}(\lambda \underset{e}{*} f) = \lambda(\mathrm{epi}\, f), \tag{2.5}$$

i.e., the epigraph of $\lambda \underset{e}{*} f$ is obtained as the $\lambda$-multiple of the epigraph of $f$. For $\lambda = 0$, the function $0^{+} \underset{e}{*} f$ is the *recession* function of $f$:

$$0^{+} \underset{e}{*} f(x) := (\mathrm{rc}\, f)(x) := \liminf_{x' \to x, \lambda \downarrow 0} \lambda f(\lambda^{-1} x'). \tag{2.6}$$

The function $0^{+} \underset{e}{*} f$ is to be viewed as a limit of the collection $\{\lambda \underset{e}{*} f, \lambda > 0\}$ when $\lambda$ goes to 0; more precisely a lower epi-limit, cf. Section 4.

**Remark 2.1.** The terminology and the notations are different from that used in the past. In the literature one finds the epi-sum $f \underset{e}{+} g$, denoted by $f \,\square\, g$ (or $f \nabla g$) and called the *inf-convolution* of $f$ and $g$. The reference to "convolution" is formal, whereas the epigraphical terminology refers to the geometric interpretation of these operations. The need for change became imperative as these notions came to play the central role in epigraphical analysis. This viewpoint is vividly illustrated by the next statement.

**Theorem 2.2.** *For all $\lambda, \mu > 0$, $f, g$ and $h$ extended real valued functions*

$$f \underset{e}{+} (g \underset{e}{+} h) = (f \underset{e}{+} g) \underset{e}{+} h,$$
$$f \underset{e}{+} g = g \underset{e}{+} f,$$
$$f \underset{e}{+} \delta_{\{0\}} = f,$$

*where $\delta_{\{0\}}$ is the indicator function of $\{0\}$. Thus, the space of extended real-valued functions with the epi-addition is an algebraic semi-group. Also,*

$$\lambda \underset{e}{*} (f \underset{e}{+} g) = \lambda \underset{e}{*} f \underset{e}{+} \lambda \underset{e}{*} g,$$
$$\lambda \underset{e}{*} (\mu \underset{e}{*} f) = \lambda\mu \underset{e}{*} f,$$
$$1 \underset{e}{*} f = f,$$

*and if f is convex,*

$$(\lambda + \mu) \underset{e}{*} f = \lambda \underset{e}{*} f \underset{e}{+} \mu \underset{e}{*} f. \tag{2.7}$$

*Finally, if f and g are convex, so is* $\lambda \underset{e}{*} (f \underset{e}{+} g)$.

**Proof.** The argument is straightforward. For illustration purpose, we derive the following identity. When $\lambda > 0$, it follows from the definitions that

$$\begin{aligned}
\mathrm{epi}_s [\lambda \underset{e}{*} (f \underset{e}{+} g)] &= \lambda \mathrm{epi}_s (f \underset{e}{+} g) \\
&= \lambda(\mathrm{epi}_s f + \mathrm{epi}_s g) \\
&= \lambda \mathrm{epi}_s f + \lambda \mathrm{epi}_s g \\
&= \mathrm{epi}_s (\lambda \underset{e}{*} f) + \mathrm{epi}_s (\lambda \underset{e}{*} g) \\
&= \mathrm{epi}_s (\lambda \underset{e}{*} f \underset{e}{+} \lambda \underset{e}{*} g).
\end{aligned}$$

Moreover, convexity is preserved by these operations since the sum of two convex sets is convex and so is the scalar multiple of a convex set. □

The next two theorems are concerned with the basic variational properties of these operations. For $f : X \to \overline{\mathbb{R}}$, let

$$\operatorname{argmin} f := \{x \mid f(x) = \inf f\}. \tag{2.8}$$

**Theorem 2.3.** *Given $\lambda > 0$, f and g two extended real valued functions defined on X. Then*

*(i)* $\operatorname{argmin} f + \operatorname{argmin} g \subset \operatorname{argmin}(f \underset{e}{+} g)$ (2.9)
*i.e., if x minimizes f on X and y minimizes g, then x + y minimizes* $f \underset{e}{+} g$*;*

*(ii)* $\lambda(\operatorname{argmin} f) = \operatorname{argmin}(\lambda \underset{e}{*} f)$, (2.10)
*i.e., if x minimizes f on X, then* $(\lambda x)$ *minimizes* $(\lambda \underset{e}{*} f)$.

**Proof.** If $x \in \operatorname{argmin} f$, and $y \in \operatorname{argmin} g$, then for all $u, v$ in $X$

$$f(x) + g(y) \leq f(u) + f(v).$$

Hence, for all $z$ in $X$

$$\inf\{f(u) + g(v) \mid u + v = x + y\} \leq f(x) + g(y) \leq \inf\{f(u) + g(v) | u + v = z\}$$

and this means that

$$(f \underset{e}{+} g)(x + y) \leq (f \underset{e}{+} g)(z)$$

from which (2.9) follows. To prove (ii), if $x \in \operatorname{argmin} f$, for all $\lambda > 0$ and $v$ in $X$, we have

$$\lambda f(\lambda^{-1}(\lambda x)) \leq \lambda f(\lambda^{-1} v).$$

And thus

$$(\lambda \underset{e}{*} f)(\lambda x) \leq (\lambda \underset{e}{*} f)(v) \quad \text{for all } v \in X,$$

or equivalently (2.10). □

**Theorem 2.4.** *Suppose $\lambda > 0$, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, and $f$, $g$ proper extended real valued functions defined on $X$ with $\inf f > -\infty$, $\inf g > -\infty$. Then*

*(i)* $\varepsilon_1\text{-}\operatorname{argmin} f + \varepsilon_2\text{-}\operatorname{argmin} g \subset (\varepsilon_1 + \varepsilon_2)\text{-}\operatorname{argmin}(f \underset{e}{+} g)$ (2.11)

*where*

$$\varepsilon_1\text{-}\operatorname{argmin} f := \{x \mid f(x) \leq \inf f + \varepsilon_1\}$$
$$\varepsilon_2\text{-}\operatorname{argmin} g := \{x \mid g(x) \leq \inf g + \varepsilon_2\}$$

*(ii)* $\lambda(\varepsilon_1\text{-}\operatorname{argmin} f) = (\lambda\varepsilon_1)\text{-}\operatorname{argmin}(\lambda \underset{e}{*} f)$ (2.12)

*(iii) for all* $\varepsilon > 0, \operatorname{argmin}(f \underset{e}{+} g) \subset \varepsilon\text{-}\operatorname{argmin} f + \varepsilon\text{-}\operatorname{argmin} g$

**Proof.** For (i) and (ii), the proof of Theorem 2.3 also works here. If $x \in \operatorname{argmin}(f \underset{e}{+} g)$, there exist, by definition, $u_\varepsilon, v_\varepsilon$ such that $u_\varepsilon + v_\varepsilon = x$, and

$$\begin{aligned} f(u_\varepsilon) + g(u_\varepsilon) &\leq (f \underset{e}{+} g)(x) + \varepsilon, \\ &\leq \inf_{u \in X} f(u) + \inf_{v \in X} g(v) + \varepsilon, \end{aligned}$$

where the last inequality follows from the fact that the infimum of $f \underset{e}{+} g$ is the sum of the infima. And from this the assertion readily follows. □

Before we continue with our development, we describe two classes of problems where these epigraphical operations play an important and natural role.

**The Average Problem.** Let $f_1, f_2, \ldots, f_n : X \to \overline{\mathbb{R}}$ correspond to the realizations of random optimization problems. Typically, these come from uncertainty in the data due to a random environment; for example: conductivity coefficients in heterogeneous material, spatial porosity parameters for soils, data measurement errors. For $i = 1, \ldots, n$, let $x_i \in \operatorname{argmin} f_i$. The average of the solutions

$$\frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

is the solution of a new minimization problem

$$\text{minimize } \left[\frac{1}{n} \underset{e}{*} (f_1 \underset{e}{+} \cdots \underset{e}{+} f_n)\right] \tag{2.13}$$

called the *average problem*. We would be interested in knowing:

(i) the form of the limit average problem as $n \to \infty$;

(ii) if $\frac{1}{n}(x_1 + \cdots + x_n)$ provides a reasonable approximation to the optimal solution of the limit problem. □

**Asymptotic Expansion of Solutions.** Let us consider the following family of (parametric) optimization problems

$$\min_{u \in X} f_\theta(u) \quad \text{where } \theta \in \Theta,$$

$f_\theta : X \to (-\infty, \infty]$, $f_\theta \not\equiv \infty$, and $\Theta$ is a subset of a linear space that contains 0. Let us assume that for all $\theta$ in $\Theta$, the problems have all the desirable properties, in particular that there exists

$$u_\theta \in \operatorname{argmin} f_\theta .$$

We are interested in the asymptotic behavior of $u_\theta$ as $\theta$ goes to 0.

One could just be interested in the topological aspects of the problem. Namely, does the sequence $\{u_\theta, \theta \to 0\}$ converge, for which topologies, and can one characterize this limit as the solution of a (new) optimization problem? These questions are answered in the framework provided by the theory of epi-convergence that we review quickly in the next section.

For numerical reasons, as well as for calculating error bounds and estimates, one often needs more information about how (at which rate, e.g.) the $u_\theta$ approach their limit. This can be done either by introducing epi-metrics, in particular the epi-distance, to measure the rate at which the functions $f_\theta$ converge to $f_0$. This leads to quantitative stability results, cf. Section 4. Or, in certain particular cases one may hope for an asymptotic expansion of the solution $u_\theta$ which respect to $\theta$

$$u_\theta = u_0 + \theta u_1 + \theta^2 u_2 + \cdots .$$

The question is to determine $u_0, u_1, u_2, \ldots$ and to characterize them as solutions of new variational problems. Let us proceed here at a formal level and assume that such a development exists. From Theorem 2.3, assuming that

$$u_i \in \operatorname{argmin} f_i \qquad i = 0, 1, 2, \ldots,$$

we would have

$$u_\theta \in \operatorname{argmin}\{f_0 \underset{e}{+} \theta \underset{e}{*} f_1 \underset{e}{+} \theta^2 \underset{e}{*} f_2 + \cdots\}.$$

This brings us to consider an expansion for $f_\theta$ of the following type:

$$f_\theta = f_0 \underset{e}{+} \theta \underset{e}{*} f_1 \underset{e}{+} \theta^2 \underset{e}{*} f_2 + \dots, \tag{2.14}$$

or equivalently

$$\operatorname{epi} f_\theta = \operatorname{epi} f_0 + \theta(\operatorname{epi} f_1) + \theta^2(\operatorname{epi} f_2) + \cdots . \tag{2.15}$$

We refer to these as *asymptotic epigraphical expansions.* □

For the same reasons as those for which the equation: given $f$, $f_0$, find $f_1$ such that

$$f = f_0 \underset{e}{+} f_1,$$

has in general no solution, one cannot, in general, find epigraphical expansions. The space of extended real-valued functions with the epi-addition is an additive semigroup, not a group, cf. Theorem 2.2. An interesting question is to characterize the class of funtcions for which the preceding equation has a solution. It follows from the recent work of Hiriart-Urruty and Mazure [**29**], and M. Volle [**40**], that a solution $f_1$, if it exists, is given by the formula:

$$f_1(x) = \sup_u \{f(u) - f_0(x-u)\}.$$

A promising approach to the existence question, as well as the actual calculation, of asymptotic expansions for the solution $u_\epsilon$, relies on the second order epigraphical differential calculus, cf. Section 5.

We shall return to this after we have introduced the appropriate topological concepts that will allow us to understand in which sense we must interpret limits, approximations, etc.

Many of the classical operations on functions find their natural interpretation in the epigraphical setting. In particular, let $f_1$ and $f_2$ be two extended real-valued functions defined on $X$. Then

$$\operatorname{epi}(f_1 \vee f_2) = \operatorname{epi} f_1 \cap \operatorname{epi} f_2, \tag{2.16}$$

$$\operatorname{epi}(f_1 \wedge f_2) = \operatorname{epi} f_1 \cup \operatorname{epi} f_2, \tag{2.17}$$

where for all $x$,

$$(f_1 \vee f_2)(x) := \sup\{f_1(x), f_2(x)\},$$

$$(f_1 \wedge f_2)(x) := \inf\{f_1(x), f_2(x)\}.$$

Also, let $g : X \times Y \to \overline{\mathbb{R}}$ with $Y$ another vector space and define the *infimal function*

$$h(y) := \inf_{x \in X} g(x,y). \tag{2.18}$$

This function, also called the marginal value function or the perturbation function, plays an important role in optimization (mathematical programming, calculus of variation, dynamic programming, optimal control, etc.). The operation, identified by (2.18), is called *epi-projection*, in view of the following identity

$$\operatorname{epi}_s h = \operatorname{prj}_{Y\times\mathbb{R}}(\operatorname{epi}_s g), \tag{2.19}$$

where $\operatorname{prj}_{Y\times\mathbb{R}}$ is the projection operator defined by

$$(x,y,\alpha)\mapsto \operatorname{prj}_{Y\times\mathbb{R}}(x,y,\alpha) := (y,\alpha).$$

To establish (2.19), simply note that

$$(y,\alpha)\in \operatorname{epi}_s h \text{ iff } \exists x\in X \text{ such that } (x,y,\alpha)\in \operatorname{epi}_s g,$$

or equivalently

$$\alpha > h(y) \text{ iff } \inf_{x\in X} g(x,y)\le \alpha.$$

The epi-addition and the epi-multiplication have been introduced historically as "dual" operations corresponding respectively to the classical sum of (convex) functions and scalar multiplication of functions. Suppose that $X$ is paired with $X^*$ through the bilinear form $\langle\cdot,\cdot\rangle$, and $f$ is an extended real-valued function defined on $X$. The *Legendre-Fenchel transform* $f\mapsto f^*$, where $f^*: X^*\to\overline{\mathbb{R}}$ is defined by

$$f^*(x^*) := \sup_{x\in X}\{\langle x^*,x\rangle - f(x)\}.$$

This function, called the *conjugate* of $f$, is clearly a convex function, since in view of (2.16),

$$\operatorname{epi} f^* = \bigcap_{x\in X} \operatorname{epi}(\langle\cdot,x\rangle - f(x))$$

and for all $x$, $\operatorname{epi}(\langle\cdot,x\rangle - f(x))$ is a (convex) half-space.

**Proposition 2.5.** *Let $f,g: X\to\overline{\mathbb{R}}$ be any two functions, and $\lambda$ a positive scalar. Then*

$$(f \underset{e}{+} g)^* = f^* + g^*, \tag{2.20}$$

*and*

$$(\lambda f)^* = \lambda \underset{e}{*} f^*. \tag{2.21}$$

**Proof.** From the definitions, we readily have

$$\begin{aligned}(f \underset{e}{+} g)^*(x^*) &= \sup_{x\in X}\{\langle x^*, x\rangle - \inf_{u\in X}[f(u) + g(x-u)]\}\\ &= \sup_{x\in X,\, u\in X}\{\langle x^*, u\rangle - f(u) + \langle x^*, x-u\rangle - g(x-u)\}\\ &= \sup_{v\in X,\, u\in X}\{\langle x^*, u\rangle - f(u) + \langle x^*, v\rangle - g(v)\}\\ &= f^*(x^*) + g^*(x^*).\end{aligned}$$

Also,

$$\begin{aligned}(\lambda f)^*(x^*) &= \sup_{x\in X}\{\lambda\langle \lambda^{-1}x^*, x\rangle - \lambda f(x)\}\\ &= \lambda f^*(\lambda^{-1}x^*) = (\lambda \underset{e}{*} f^*)(x^*).\end{aligned}$$

when $\lambda > 0$. □

From this proposition, and the fact that $f^{**} \leq f$, we also have that

$$(f+g)^* \leq (f^* \underset{e}{+} g^*)^{**} \tag{2.22}$$

and

$$(\lambda \underset{e}{*} f)^* \leq (\lambda f)^{**}. \tag{2.23}$$

To obtain equalities one needs $f^{**} = f$, $g^{**} = g$. This requires $f$ and $g$ to be proper, lower semiontinuous and convex, or identically $+\infty$ or $-\infty$. If, moreover $(f^* \underset{e}{+} g^*)^{**} = f^* \underset{e}{+} g^*$, then (2.22) becomes

$$(f+g)^* = f^* \underset{e}{+} g^*. \tag{2.24}$$

## 3. Epigraphical regularization

The epi-addition has a smoothing effect. By this, one means that a function obtained as an epi-sum is at least as "smooth" as any of its elements, up to $C^{1,1}$ regularity; this is a general phenomenon when dealing with unilateral variational problem. This provides a natural bridge between epigraphical analysis and classical analysis. As we shall see, by choosing appropriately the regularizing term in an epi-sum, one can endow the resulting function with the appropriate differentiability properties; this smoothing effect could in some way, justify calling the epi-addition an "inf-convolution".

To begin with, let us consider the epi-sum of an arbitrary function and a "kernel" of the type $(\lambda p)^{-1}\|\cdot\|^p$.

**Definition 3.1.** *Let $(X, \|\cdot\|)$ be a normed linear space. Given $p \in [1, \infty)$ and $\lambda > 0$, the epi-regularization of index $\lambda$ of a function $f : X \to \overline{\mathbb{R}}$ is defined as*

$$f_\lambda := f \underset{e}{+} \frac{1}{\lambda p}\|\cdot\|^p. \tag{3.1}$$

*Explicitly*

$$f_\lambda(x) = \inf_{x \in x}\{f(u) + \frac{1}{\lambda p}\|x - u\|^p\}. \tag{3.2}$$

This terminology is justified by the following result.

**Proposition 3.2.** *Let $(X, \|\cdot\|)$ be a normed linear space $\lambda > 0, p \in [1, \infty)$, and $f : X \to \overline{\mathbb{R}}$. Suppose that for some $\alpha_0 \in \mathbb{R}$, $\alpha_1 \in \mathbb{R}_+$, $f$ majorizes $\alpha_0 - \alpha_1\|\cdot\|^p$, i.e.,*

$$\text{for all } x \in X,\ f(x) \geq \alpha_0 - \alpha_1\|x\|^p.$$

*Then, provided $0 < \lambda < 2^{1-p}(\alpha_1 p)^{-1}$,*

*$f_\lambda$ is lipschitz continuous if $p = 1$*

*$f_\lambda$ is locally lipschitz continuous if $p > 1$,*

*and the lipschitz constants depend only on the value of the function at one point, and on $\alpha_0, \alpha_1$ and $\lambda$; moreover this dependence is continuous.*

**Proof.** A number of related results have been obtained recently, see Attouch [**3**], Fougères and Truffert [**26**], Attouch, Azé and Wets [4]. This particular theorem can be found in Attouch and Wets [9, Lemma 3.2]. ▢

**Remark 3.3.** Geometrically, the preceding result takes the following form:

$$\text{epi}_s f_\lambda = \text{epi}_s f + \text{epi}_s (\lambda p)^{-1}\|\cdot\|^p.$$

Now observe that $\text{epi}_s (\lambda p)^{-1}\|\cdot\|^p$ is a "smooth" set if $p > 1$. The statement would thus follow from the general fact: "if $D$ is a smooth set, then $C + D$ inherits the regularity of $D$". We prove next a result of this type that has various implications.

**Proposition 3.4.** *Let $H$ be a hilbert space and $\varphi : H \to \mathbb{R}_+$ a potential, i.e., $\varphi$ is continuously differentiable, strictly convex and coercive, $\varphi(0) = 0$ and for any $u$ and any sequence $\{u_\nu, \nu \in \mathbb{N}\}$ one has*

$$[u_\nu \underset{w}{\longrightarrow} u \text{ and } \varphi(u_\nu) \to \varphi(u)] \Longrightarrow u_\nu \underset{s}{\longrightarrow} u.$$

*Then, for any function $f : H \to (-\infty, \infty]$, convex, proper and lower semicontinuous, its $\varphi$-epi-regularization*

$$f_\varphi = f \underset{e}{+} \varphi.$$

*is continuously differentiable.*

*Moreover, for any nonempty closed convex set $C \subset H$ one has:*

$$C + B_\varphi = \{x \in H \mid (\varphi \underset{e}{+} \delta_C)(x) \leq 1\}, \tag{3.3}$$

*where*

$$B_\varphi := \{x \in H \mid \varphi(x) \leq 1\},$$

*and $\delta_C$ is the indicator function of the set $C$. This means that $C + B_\varphi$ is a smooth set in the following sense: it is equal to the level set $\mathrm{lev}_\alpha\, g$ with $\alpha > \inf g$, of a convex continuously differentiable function $g$.*

**Proof.** From convex calculus (recall that $\varphi$ is coercive), we know that the infimum in the expression (2.2) that defines $(f \underset{e}{+} \varphi)(x)$ is achieved at a unique point that we denote by $J_\varphi(x)$.

$$\begin{aligned}(f \underset{e}{+} \varphi)(x) &= \min_{u \in H}[f(u) + \varphi(x-u)] \\ &= f(J_\varphi(x)) + \varphi(x - J_\varphi(x)),\end{aligned}$$

and

$$\partial(f \underset{e}{+} \varphi(x) = \partial f(J_\varphi(x)) \cap \nabla\varphi(x - J_\varphi(x)),$$

see, e.g., Laurent [**31**], Aubin and Ekeland [**12**].

Thus, $\partial(f \underset{e}{+} \varphi)(x)$ is reduced to a single element, and hence $f \underset{e}{+} \varphi$ is gâteaux differentiable. Fréchet differentiability of $f \underset{e}{+} \varphi$ will follow if we prove that the gâteaux derivative is continuous. In view of the preceding relations, this boils down to showing that

$$x \mapsto J_\varphi(x) \text{ is continuous},$$

using here the fact that $\nabla\varphi$ is continuous.

Let $\{x_\nu,\, \nu \in \mathbb{N}\}$ be a sequence converging strongly to $x$. The coercivity of $\varphi$ guarantees the boundedness of the corresponding sequence $\{J_\varphi(x_\nu),\ \nu \in \mathbb{N}\}$. For all $u$ in $H$, we have

$$f(J_\varphi(x_\nu)) + \varphi(x_\nu - J_\varphi(x_\nu)) \leq f(u) + \varphi(x_\nu - u),$$

and after extracting a weakly convergent subsequence such that $J_\varphi(x_{\nu'}) \underset{w}{\longrightarrow} z$, we pass to the limit to obtain

$$f(z) + \varphi(x - z) \leq f(u) + \varphi(x - u), \qquad \forall u \in H.$$

Note that we only used the convexity and lower semicontinuity of $f$. The preceding inequality implies that

$$z = J_\varphi(x),$$

and that the whole sequence $\{J_\varphi(x_\nu),\ \nu \in \mathbb{N}\}$ converge to $J_\varphi(x)$. As a consequence of this, we obtain the convergence of the infimal values:

$$f(J_\varphi(x_\nu)) + \varphi(x_\nu - J_\varphi(x_\nu)) \to f(J_\varphi(x)) + \varphi(x - J_\varphi(x)).$$

Since

$$f(J_\varphi(x)) \le \liminf_\nu f(J_\varphi(x_\nu))$$
$$\varphi(x_\nu - J_\varphi(x_\nu)) \le \varphi(x - J_\varphi(x)),$$

we have that

$$\limsup_\nu(\varphi(x_\nu) - J_\varphi(x_\nu)) \le \limsup_\nu[f(J_\varphi(x_\nu)) + \varphi(x_\nu - J_\varphi(x_\nu))] - \liminf_\nu f(J_\varphi(x_\nu)),$$

which with the preceding inequalities yields

$$\limsup_\nu \varphi(x_\nu - J_\varphi(x_\nu)) \le \varphi(x - J_\varphi(x)).$$

Hence

$$x_\nu - J_\varphi(x_\nu) \text{ converges weakly to } x - J_\varphi(x),$$
$$\varphi(x_\nu - J_\varphi(x_\nu)) \text{ converges to } \varphi(x - J_\varphi(x)).$$

Now, use the fact that $\varphi$ is a potential to conclude that $J_\varphi(x_\nu) \xrightarrow[s]{} J_\varphi(x)$. And this completes the proof of the fréchet differentiability of $f \underset{e}{+} \varphi$.

We now turn to (3.3). Note that

$$\begin{aligned} C + B_\varphi &= \{z \in H \mid \exists x \in C,\ \varphi(z - x) \le 1)\}, \\ &= \{z \in H \mid \min_{u \in H}[\delta_C(u) + \varphi(z - u)] \le 1\}, \\ &= \{z \in H \mid (\delta_C \underset{e}{+} \varphi)(z) \le 1\}. \end{aligned}$$

The assertion now follows from the first part of the proof with $f = \delta_C$. □

Let us consider a special case of this result. For any $\lambda > 0$, define $\varphi_\lambda(x) = (2\lambda)^{-1}\|x\|^2$. Clearly, $\varphi_\lambda$ is continuously differentiable on a hilbert space. For any proper lower semicontinuous convex functions $f : H \to (-\infty, \infty]$ and $\lambda > 0$, we set:

$$f_\lambda := f \underset{e}{+} (2\lambda)^{-1}\|\cdot\|^2,$$
$$f_\lambda(x) = \min_{u \in H}[f(u) + \frac{1}{2\lambda}\|x - u\|^2].$$

In view of the previous proposition, the function $f_\lambda$ is continuously differentiable. This particular result was first obtained by Brézis [**18**]. These functions $\{f_\lambda,\ \lambda > 0\}$ play a fundamental role in optimization theory and are called *Moreau-Yosida approximates*. The reference to Yosida approximation (which usually refers to approximating operators) is justified by the following property:

$$\nabla f_\lambda(x) = (\partial f)_\lambda(x) := \frac{1}{\lambda}(x - (I + \partial f)^{-1}x),$$

where $(\partial f)_\lambda$ is the Yosida approximate of the maximal monotone operator $\partial f$.

## 4. Epigraphical Limits

We only review here a few of the main features of this theory which has received a lot of attention during the last decade. For more about epi-convergence, the reader could refer to the book on "Variational Convergence for Functions and Operators", Attouch [**3**].

Let $(X, \|\cdot\|)$ be a normed linear space, $\{f^\nu : X \to \overline{\mathbb{R}},\ \nu \in \mathbb{N}\}$ a sequence of extended real-valued functions defined on $X$. We say that the $f^\nu$ *epi-converge* to $f$ at $x$, if

(i) for any sequence $\{x^\nu,\ \nu \in \mathbb{N}\}$ converging to $x$, $\liminf_\nu f^\nu(x^\nu) \geq f(x)$, and

(ii) there exists $\{\hat{x}^\nu,\ \nu \in \mathbb{N}\}$ converging to $x$ such that $\limsup_\nu f^\nu(\hat{x}^\nu) \leq f(x)$.

Note that these conditions imply that $f$ is lower semicontinuous. We then say that $f$ is the *epi-limit* of the $f^\nu$, and write $f = \operatorname{epi-lim}_{\nu\to\infty} f^\nu$. We refer to this type of convergence as epi-convergence, since it is equivalent to the set-convergence of the epigraphs. Interest in epi-convergence stems from the fact that from a variational viewpoint it is the weakest type of convergence that possesses the following properties:

**Proposition 4.1.** (Rockafellar and Wets [**39**]) *Suppose $\{f; f^\nu : X \to \overline{\mathbb{R}},\ \nu = 1,\ldots\}$ is a collection of functions such that $f = \operatorname{epi-lim}_{\nu\to\infty} f^\nu$. Then*

$$\limsup_{\nu\to\infty}(\inf f^\nu) \leq \inf f, \tag{4.1}$$

*and, if*

$$x^k \in \operatorname{argmin} f^{\nu_k} \text{ for some subsequence } \{\nu_k,\ k = 1,\ldots\}$$

*and $x = \lim_{k\to\infty} x^k$, it follows that*

$$x \in \operatorname{argmin} f,$$

*and*

$$\lim_{k\to\infty}(\inf f^{\nu_k}) = \inf f;$$

*so in particular if there exists a compact set $D \subset X$ such that for some subsequence $\{\nu_k,\ k = 1, \ldots\}$,*

$$\operatorname{argmin} f^{\nu_k} \cap D \neq \emptyset,$$

*then the minimum of $f$ is attained at some point in $D$.*

*Moreover, if* $\operatorname{argmin} f \neq \emptyset$, *then* $\lim_{\nu\to\infty}(\inf f^\nu) = \inf f$ *if and only if* $x \in \operatorname{argmin} f$ *implies the existence of sequences* $\{\varepsilon_\nu \geq 0,\ \nu = 1, \ldots\}$ *and* $\{x^\nu \in X,\ \nu = 1, \ldots\}$ *with*

$$\lim_{\nu\to\infty} \varepsilon_\nu = 0, \text{ and } \lim_{\nu\to\infty} x^\nu = x,$$

*such that for all* $\nu = 1, \ldots$

$$x^\nu \in \varepsilon_\nu\text{-}\operatorname{argmin} f^\nu := \{x \mid f^\nu(x) \leq \varepsilon_\nu + \inf f^\nu\}.$$

We will not provide an overview of epi-convergence (sometimes called Γ-convergence, DeGiorgi [**24**]). We shall limit ourselves to a brief introduction to the recently developed quantitative theory and stating one result that illustrates the relationship between taking epigraphical limits and operations. We begin with this latter.

**Proposition 4.2.** *Suppose $(X, \|\cdot\|)$ is a normed linear space, and $\{f; f^\nu,\ \nu \in \mathbb{N}\}$, $\{g; g^\nu,\ \nu \in \mathbb{N}\}$ two collections of extended real-valued functions defined on $X$, such that*

$$f = \underset{\nu\to\infty}{\textit{epi-lim}}\, f^\nu, \text{ and } g = \underset{\nu\to\infty}{\textit{epi-lim}}\, g^\nu.$$

*Suppose also that the $\{f; f^\nu,\ \nu \in \mathbb{N}\}$, are equi-inf-compact, i.e., for all $\alpha$ the level sets*

$$\operatorname{lev}_\alpha f^\nu := \{x \mid f^\nu(x) \leq \alpha\},$$

*are contained in a compact set, for $\nu$ sufficiently large; and suppose that $\{\inf g^\nu,\ \nu \in \mathbb{N}\}$ are (equi-)bounded below, i.e., there exists $\beta \in \mathbb{R}$ such that for $\nu$ sufficiently large $\inf g^\nu \geq \beta$. Then*

$$f \underset{e}{+} g = \underset{\nu\to\infty}{\textit{epi-lim}}(f^\nu \underset{e}{+} g^\nu).$$

**Proof.** It is immediate that $\text{epi-}\limsup_{\nu\to\infty}(f^\nu \underset{e}{+} g^\nu) \leq f \underset{e}{+} g$. The converse inequality $f \underset{e}{+} g \leq \text{epi-}\liminf_{\nu\to\infty}(f^\nu \underset{e}{+} g^\nu)$ is derived from the argument that follows. Whenever $\gamma > \liminf_{\nu\to\infty}(f^\nu \underset{e}{+} g^\nu)(z)$, there exits a sequence $\{z^\nu, \nu \in \mathbb{N}\}$, passing to a subsequence if necessary, such that

$$\liminf_{\nu\to\infty}(f^\nu \underset{e}{+} g^\nu)(z^\nu) < \gamma.$$

By definition of epi-sum, it follows that there exist $\{x^\nu,\ \nu \in \mathbb{N}\}$, $\{y^\nu,\ \nu \in \mathbb{N}\}$, such that $x^\nu + y^\nu = z^\nu$, and

$$f^\nu(x^\nu) + g^\nu(y^\nu) < \gamma.$$

Since the $g^\nu$ are equi-bounded below, for $\nu$ sufficiently large, we have that $f^\nu(x^\nu) \leq \gamma - \beta$. This means that from some $\nu$ on, all $x^\nu$ belong to a compact set contained in $\mathrm{lev}_{\gamma-\beta} f^\nu$. Again passing to a subsequence, if necessary, let $x$ denote the limit of the sequence $\{x^\nu,\ \nu \in \mathbb{N}\}$. The corresponding sequence $\{y^\nu,\ \nu \in \mathbb{N}\}$ then also converges to $z - x := y$. From these observations, it follows that

$$(f \underset{e}{+} g)(z) \leq f(x) + g(y) \leq \liminf_{\nu\to\infty}[f^\nu(x^\nu) + g^\nu(y^\nu)] \leq \gamma.$$

The first inequality follows from the definition of epi-sum, and the second one from the epi-convergence of the $f^\nu$ and $g^\nu$ to $f$ and $g$. □

We now turn to the definition of a distance between functions that would be compatible with the topology of epi-convergence. Let $d$ the distance function generated by the norm on $X$, and let $d(x, C)$ denote the *distance from $x$ to $C$*; if $C = \emptyset$, set $d(x, C) = \infty$. For any $\rho \geq 0$, $\rho B$ denotes the ball of radius $\rho$ and for any set $C$,

$$C_\rho := C \cap \rho B.$$

For $C, D \subset X$, the *"excess" function* of $C$ on $D$ is defined as,

$$e(C, D) := \sup_{x \in C} d(x, D),$$

with the convention that $e = 0$ if $C = \emptyset$. For any $\rho \geq 0$, the *$\rho$-(Hausdorff-)distance* between $C$ and $D$ is given by

$$\mathrm{haus}_\rho(C, D) = \sup\{e(C_\rho, D),\ e(D_\rho, C)\}.$$

**Definition 4.3.** (Attouch and Wets [9]) *For $\rho \geq 0$, the $\rho$-(Hausdorff-)epi-distance between two extended real valued functions $f, g$ defined on $X$, is*

$$\mathrm{haus}_\rho(f, g) := \mathrm{haus}_\rho(\mathrm{epi}\, f, \mathrm{epi}\, g),$$

*where the unit ball of $X \times \mathbb{R}$ is the set $B := B_{X\times\mathbb{R}} = \{(x, \alpha) : \|x\| \leq 1,\ |\alpha| \leq 1\}$.*

Convergence with respect to the epi-distances is somewhat stronger than epi-convergence, at least in the infinite dimensional case, when $X$ is a reflexive Banach space and epi-limits are defined in terms of Mosco-convergence, i.e. epi-convergence with respect to both the strong and the weak topology on $X$. Let $\{f^\nu : X \to \overline{\mathbb{R}}, \nu = 1, \ldots\}$ be a sequence of functions. We say that $f$ is the *Mosco-epi-limit* of this sequence, if for all $x$ in $X$:

for any sequence $\{x^\nu,\ \nu = 1, \ldots\}$ converging weakly to $x$, $\liminf_\nu f^\nu(x^\nu) \geq f(x)$,

and

there exists $\{\hat{x}^\nu,\ \nu = 1, \ldots\}$ converging strongly to $x$ such that $\limsup_\nu f^\nu(\hat{x}^\nu) \leq f(x)$.

**Proposition 4.4.** *Suppose $X$ is a reflexive Banach space, $\{f; f^\nu,\ \nu = 1, \ldots\}$ a collection of proper, extended real valued, lower semicontinuous, convex functions defined on $X$. Then,*

$$\lim_{\nu \to \infty} \text{haus}_\rho(f, f^\nu) = 0,$$

*for all $\rho$ sufficiently large implies*

$$f = \text{Mosco-}\mathit{epi\text{-}lim}_{\nu \to \infty} f^\nu.$$

*Moreover, if $X(=\mathbb{R}^n)$ is finite dimensional, the epi-distance topology is the epi-topology, i.e., the topology of epi-convergence (without any convexity assumptions).*

The proof can be found in Attouch and Wets [9], where one can also find an example of a sequence of functions defined on a hilbert space that epi-converges but has no limit with respect to the epi-distance. This, however, does not paint the full picture. It is also shown in [9, Theorem 4.7] that most, if not all, of the known applications of epi-convergence in infinite dimensions are such that Mosco-epi-convergence and convergence with respect to the epi-distance coincide.

These pseudo-distances are not the only ones that could be used to quantify epi-convergence. In Attouch and Wets [7], we introduce a notion of distance based on epi-graphical regularization. For fixed $p \in [1, \infty)$ and $\lambda > 0$, we work with the following epigraphical regularization: $f_\lambda := f \underset{e}{+} (\lambda p)^{-1} \|\cdot\|^p$. This leads to the following notion of distance between two functions $f$ and $g$:

$$d_{\lambda,\rho}(f, g) = \sup_{\|x\| \leq \rho} |f_\lambda(x) - g_\lambda(x)|.$$

Assuming that $f$ and $g$ are proper, this quantity is well defined since both $f_\lambda$ and $g_\lambda$ are then bounded on bounded sets. The next theorem puts forward the relationship between these pseudo-distances and the epi-distance.

**Theorem 4.5.** (Attouch and Wets [**9**]) *Let $f$ and $g$ be two extended real valued functions defined on a normed linear space $(X, \|\cdot\|)$, such that for some $\alpha_0 \geq 0$ and $\alpha_1 \in \mathbb{R}$,*

$$f \geq -\alpha_0\|\cdot\|^p - \alpha_1, \text{ and } g \geq -\alpha_0\|\cdot\|^p - \alpha_1,$$

*for $1 \leq p < \infty$. Then for $0 < \lambda < (\alpha_0 p)^{-1}2^{1-p}$, and $\rho \geq 0$*

$$d_{\lambda,\rho}(f,g) \leq \beta(\lambda,\rho)\mathrm{haus}_{\gamma(\lambda,\rho)}(f,g),$$

*with the constants $\gamma$ and $\beta$ depending only on the norm of a point at which $f$ and $g$ are finite and the corresponding values of $f$ and $g$, on $\alpha_0, \alpha_1$ and of course, on $\rho$ and $\lambda$. Similarly, for all $0 < \lambda < (\alpha_0 p)^{-1}2^{1-p}$, and*

$$\rho \geq \max[d(0, \mathrm{epi}\, f), d(0, \mathrm{epi}\, g)],$$

*we have*

$$\mathrm{haus}_\rho(f,g) \leq d_{\lambda,9\rho}(f,g) + 2(\lambda p)^{1/p}\left[\frac{\alpha_0 2^{p-1}(9\rho)^p + 9\rho + \alpha_1}{1 - \alpha_0 p\lambda 2^{p-1}}\right]^{1/p}.$$

We are thus dealing with the same uniformities. The epi-distance is relatively easy to calculate or to estimate, one can use the Kenmochi conditions [**9**, Section 2] or the results of Azé and Penot [**13**] who derive bounds for the $\rho$-Hausdorff distance between sets obtained as the results of various operations (union, intersection, addition, etc.). On the other hand, the distances $d_{\lambda,\rho}$ are better suited for theoretical investigations; for example, one can demonstrate that the Legendre-Fenchel transform is an isometry for those distances [**7**].

We have used these distances to derive hölderian and lipschitzian properties for the optimal and $\varepsilon$-solutions of optimization problems, cf. Attouch and Wets [**8**], [**10**]. For example, in the normalized case ($x_f = 0$ and $f(x_f) = 0$), when $f$ is quadratically "conditioned" at 0, i.e., $f(x) \geq \|x\|^2 = \varphi(x)$ for $\|x\| \leq 1$, and $g$ is some approximation or perturbation of $f$, with $x_g$ a corresponding minimizer, Theorem 4.1 of [**8**] asserts that

$$\|x_g - x_f\| \leq [4\mathrm{haus}_\rho(f,g)]^{1/2}$$

provided that the epi-distance (of parameter $\rho$) $\mathrm{haus}_\rho(f,g)$ is sufficiently small. We also showed that this hölderian stability result is optimal.

These, and related considerations, seem to suggest that the epi-distance topology (or equivalently, the topology generated by the pseudo-distance $d_{\lambda,\rho}$) plays a *central role* in the theory of epi-convergence. It allows for the development of a quantitative theory of epi-convergence (used to derive error bounds for the solution of approximating optimization problems), the Legendre-Fenchel transform is bicontinuous with respect to this topology (in fact, isometric properties can be derived), and, convergence with respect to the epi-distance is sufficient general so that all significant applications are covered. We know a few of its properties, for example, the space of extended real-valued functions equipped with the uniform structure generated by the epi-distances $\{\text{haus}_\rho,\ \rho > 0\}$ is complete when $X$ is finite dimensional or if $X$ is a reflexive Banach space and the functions are proper, lower semicontinuous and convex. Certainly a much more thorough study of this topology is warranted.

## 5. Epigraphical Differential Calculus.

The concept of epi-derivative has emerged from the need to push the subdifferential calculus beyond that for first order derivatives; consult, for example, the work of Aubin [**11**] on the differentiability of multifunctions (as it applies to subgradient multifunctions), and of Rockafellar [**37**], [**38**] on the second order generalized derivatives.

**Definition 5.1.** *Suppose $f$ is finite at $x$ with $f$ an extended real-valued function defined on a normed linear space $X$. If*

$$D_\uparrow f(x;\cdot) = \operatorname*{epi\text{-}lim}_{h \to 0} \frac{1}{h}[f(x+h\cdot) - f(x)]$$

*is a well-defined function (possibly with values $\pm\infty$), then $f$ is said to be epi-differentiable at $x$. The function $y \mapsto D_\uparrow f(x;y)$ is the directional epi-derivative of $f$ at $x$.*

The difference with the classical definition of directional derivatives is the use of epi-limits instead of pointwise limits. We are lead to this in the most natural way if we approach the problem geometrically.

For $\{C_\nu \subset X,\ \nu \in N\}$ be a filtered family $(N, \mathcal{H})$, let

$$\begin{aligned}\operatorname*{Lim\,inf}_{\mathcal{H}} C_\nu &= \bigcap_{H \in \mathcal{H}^\#} \operatorname{cl}\Big(\bigcup_{\nu \in H} C_\nu\Big)\\ &= \{x \in X \mid \lim_{\mathcal{H}} d(x, C_\nu) = 0\}\end{aligned}$$

where $\mathcal{H}^\# = \{H' \subset N \mid H' \subset H \neq \emptyset,\ \forall H \in \mathcal{H}\}$ in the *grill* of $\mathcal{H}$, and

$$\begin{aligned}\operatorname*{Lim\,sup}_{\mathcal{H}} C_\nu &= \bigcap_{H \in \mathcal{H}} \operatorname{cl}\Big(\bigcup_{\nu \in H} C_\nu\Big)\\ &= \{x \in X \mid \liminf_{\mathcal{H}} d(x, C_\nu) = 0\}.\end{aligned}$$

We have the following notions of tangency: for $x^0 \in \operatorname{cl} K$, $K \subset X$,

$$T_K(x^0) := \operatorname*{Lim\,sup}_{h \downarrow 0+} h^{-1}[K - x^0], \quad \text{(the Boulingand tangent cone)},$$
$$A_K(x^0) := \operatorname*{Lim\,inf}_{h \downarrow 0+} h^{-1}[K - x^0], \quad \text{(adjacent cone)},$$
$$C_K(x^0) := \operatorname*{Lim\,inf}_{(h \downarrow 0+,\, x' \xrightarrow[K]{} x^0)} h^{-1}[K - x^0], \quad \text{(the Clarke tangent cone)},$$

where $x' \xrightarrow[K]{} x^0$ means that the convergence takes place with $x'$ in $K$. Obviously all these cones are closed, and

$$T_K(x^0) \subset A_K(x^0) \subset C_K(x^0).$$

A particularly nice feature of $C_K$ is that it is convex. We say that $K$ is *proto-differentiable* at $x^0$ if $T_K = A_K$, i.e., the Painlevé-Kuratowski limit

$$\operatorname*{Lim}_{h \downarrow 0+} h^{-1}[K - x^0]$$

exists. All of this leads to natural definitions of differentiability for multifunctions. Let $X$ and $Y$ be two normed linear spaces and $\Gamma : X \rightrightarrows Y$ a multifunction (set-valued map). By

$$\operatorname{gph} \Gamma := \{(x, y) \mid y \in \Gamma(x),\ x \in X\},$$

we denote the graph of $\Gamma$. For $y^0 \in \Gamma(x^0)$, we define the differentials of $\Gamma$ at $(x^0, y^0)$ as follows

$$D^t\Gamma(x^0, y^0) := T_{\operatorname{gph}\Gamma}(x^0, y^0),$$
$$D^a\Gamma(x^0, y^0) := A_{\operatorname{gph}\Gamma}(x^0, y^0),$$
$$D^c\Gamma(x^0, y^0) := C_{\operatorname{gph}\Gamma}(x^0, y^0);$$

it is proto-differentiable at $(x^0, y^0)$ if $D^t\Gamma(x^0, y^0) = D^a\Gamma(x^0, y^0)$.

The epi-derivative is related to the subdifferentials of multifunctions in the following way: given $f : X \to [-\infty, \infty]$, we associate to it the multifunction

$$x \mapsto \Gamma(x) : X \rightrightarrows \mathbb{R}$$

where

$$\Gamma(x) = \begin{cases} [f(x), \infty) & \text{if } f(x) < \infty \\ \emptyset & \text{if } f(x) = \infty. \end{cases}$$

Then

$$\operatorname{gph} \Gamma = \operatorname{epi} f,$$

and

$$h^{-1}[\mathrm{epi}\, f - (x^0, f(x^0)] = \mathrm{epi}\left[\frac{f(x^0 + h\cdot) - f(x^0)}{h}\right].$$

We have

$$\begin{aligned} D^{\mathrm{t}}\Gamma(x^0, f(x^0)) &= T_{\mathrm{epi}\, f}(x^0, f(x^0)) \\ &= \mathrm{epi}\text{-}\liminf_{h \downarrow 0+} h^{-1}[f(x^0 + h\cdot) - f(x^0)], \end{aligned}$$

where the epi-liminf is the function whose epigraph is the Limsup of the epigraphs, see [**3**] for example,

$$\begin{aligned} D^{a}\Gamma(x^0, f(x^0)) &= A_{\mathrm{epi}\, f}(x^0, f(x^0)) \\ &= \mathrm{epi}\text{-}\limsup_{h \downarrow 0+} h^{-1}[f(x^0 + h\cdot) - f(x^0)] \end{aligned}$$

where the epi-lim sup is the function whose epigraph is the Liminf of the epigraphs, and

$$\begin{aligned} D^{c}\Gamma(x^0, f(x^0)) &= C_{\mathrm{epi}\, f}(x^0, f(x^0)) \\ &= \operatorname*{epi\text{-}limsup}_{h \downarrow 0+, (x', \alpha') \underset{\mathrm{epi}\, f}{\longrightarrow} (x^0, f(x^0))} h^{-1}[f(x' + h\cdot) - \alpha']. \end{aligned}$$

Moreover $\Gamma$ is proto-differentiable at $(x^0, f(x^0))$ if and only if $f$ is epi-differentiable at $x^0$.

In the case of proper, lower semicontinuous convex functions, the notion of a second order epi-derivative also follows in a natural way from the multivalued subdifferential calculus. Let $f : H \to (-\infty, \infty]$ be a proper, lower semicontinuous convex function defined on the hilbert space $H$. In computing the differentials of the subgradient map $x \mapsto \partial f(x)$ at a point $(x^0, v^0) \in \mathrm{gph}\, \partial f$, we utilize the following facts:

(i) $h^{-1}[\mathrm{gph}\, \partial f - (x^0, v^0)] = \partial h^{-2}[f(x^0 + h\cdot) - f(x^0) - h\langle \cdot, v^0 \rangle]$

(i) convergence of subdifferentiable multifunctions is equivalent to the Mosco-epi-convergence of the associated proper, lower semicontinuous, convex functions, see Attouch [**3**]. Recall that Mosco-epi-convergence requires epi-convergence with respect to both the strong and the weak topology on $H$.

From this follows a notion for second order epi-derivative (Rockafellar [**38**], Ndoutoume [**34**])

$$Df(x^0, v^0; \cdot) = \operatorname*{Mosco\text{-}epi\text{-}lim}_{h \downarrow 0+} \frac{1}{h^2}[f(x^0 + h\cdot) - f(x^0) - h\langle \cdot, v^0 \rangle].$$

To conclude, let us give the definition of epi-integral. Aumann, see [**20**] for example, introduced the following definition for the integral of a multifunction. Let $\Gamma : (\Omega, \mathcal{A}, \mu) \rightrightarrows Y$ be a measurable multifunction. Then

$$\int \Gamma(\omega)\mu(d\omega) := \mathrm{cl}\left\{ \int u(\omega)\mu(d\omega) \,\middle|\, u \text{ a } \mathcal{L}^1\text{-selection of } \Gamma \right\}.$$

Taking closure is at variance with Auman's definition; we do it here because we deal with closed-valued multifunctions whose integrals should also be closed. A function $f: X \times \Omega \to (-\infty, \infty]$ is a *normal integrand* if the epigraphical multifunction $\omega \mapsto \operatorname{epi} f(\cdot,\omega)$ is a closed-valued measurable multifunction and the *epi-integral* is defined as

$$\operatorname{epi} ⨙ f(\cdot,\omega)\mu(d\omega) := \int \operatorname{epi} f(\cdot,\omega)\mu(d\omega).$$

(The operation $⨙ \cdot \mu(d\omega)$ has also been called the continuous inf-convolution.) One can prove the following formulas, when $f$ is $\mu$-atom convex [**2**]:

$$\begin{aligned} ⨙ f(x,\omega)\mu(d\omega) &= \inf\left\{ \int f(u(\omega),\omega)\mu(d\omega) \,\Big|\, \int u(\omega)\mu(d\omega) \to x \right\} \\ &= \operatorname{cl}\operatorname{co}\left\{ \inf\left[ \int f(u(\omega),\omega))\mu(d\omega) \,\Big|\, \int u(\omega)\mu(d\omega) = x \right] \right\} \end{aligned}$$

where co takes the convex closure, and cl the lower semicontinuous closure. The first one of these two expressions follows from the second one, which in turn can be derived from the identity

$$\left(⨙ f(\cdot,\omega)\mu(d\omega)\right)^{*}(v) = \int_{\Omega} f^{*}(v,\omega)\mu(d\omega).$$

**References.**


[1] Zvi Artstein & Sergiu Hart, "Law of large numbers for random sets and allocation processes," *Mathematics of Operations Research* **6** (1981), 482–492.

[2] Zvi Artstein & Roger J-B Wets, "Approximating the integral of a multifunction," *J. Multivariate Analysis* **24** (1988), 285–308.

[3] Hedy Attouch, *Variational Convergence for Functions and Operators*, Pitman, London. 1984., Applicable Mathematics Series .

[4] Hedy Attouch, Dominique Aze & Roger J-B Wets, "Convergence of convex-concave saddle functions: continuity properties of the Legendre-Fenchel transform with applications to convex programming and mechanics," *Annales de l'Institut H. Poincaré: Analyse Nonlinéaire* (1988 (to appear)), Techn. Report A.V.A.M.A.C.-Perpignan, # 85-08, 1985.

[5] Hedy Attouch & Colette Picard, "On the control of transmission problems across perforated walls," in *Proceedings of the First International Workshop: Sensors and Actuators in Distributed Parameter Systems*, A. El Jai & M. Amouroux, eds., IFAC-Université de Perpignan, Perpignan, 1987, 31–58.

[6] Hedy Attouch & Hassan Riahi, "The epi-continuation method for minimization problems. Relation with the degree theory of F. Browder for maximal monotone operators," AVAMAC #87-02, Perpignan, 1987.

[7] Hedy Attouch & Roger J-B Wets, "Isometries for the Legendre-Fenchel transform," *Transactions of the American Mathematical Society* **296** (1986), 33–60.

[8] Hedy Attouch & Roger J-B Wets, "Quantitative stability of variational systems: II. A framework for nonlinear conditioning," IIASA Working Paper 88-9, Laxenburg, Austria., February 1988.

[9] Hedy Attouch & Roger J-B Wets, "Quantitative stability of variational systems: I. The epigraphical distance," IIASA Working Paper 88-8, Laxenburg, February 1988.

[10] Hedy Attouch & Roger J-B Wets, "Lipschitzian stability of the epsilon-approximate solutions in convex optimization," IIASA Working Paper WP-87-25, Laxenburg, Austria, March 1987.

[11] Jean-Pierre Aubin, "Graphical convergence of set-valued maps," WP-87, IIASA Laxenburg, September 1987.

[12] Jean-Pierre Aubin & Ivar Ekeland, *Applied Nonlinear Analysis*, J. Wiley Intersciences, New York, 1984.

[13] Dominique Aze & Jean-Paul Penot, "Operations on convergent families of sets and functions," Techn. Report AVAMAC #87-05, Perpignan, 1987.

[14] Kerry Back, "Convergence of Lagrange multipliers and dual variables for convex optimization problems," *Mathematics of Operations Research* **13** (1988), 74–79.

[15] Kerry Back, "Continuity of the Fenchel transform of convex functions," *Proceedings of the American Mathematical Society* **97** (August 1986), 661–667.

[16] Gerald Beer, "On Mosco convergence of convex sets," *Bulletin of the Australian Mathematical Society* (1987)

[17] Gerald Beer & Petar Kenderov, "On the argmin multifuction for lower semicontinuous functions," *Proceedings of the American Mathematical Society* **102** (1988), 107–113.

[18] Haim Brezis, *Opérateurs Maximaux Monotones et Semi-groupes de Contractions dans les Espaces de Hilbert*, North Holland, Amsterdam, 1973.

[19] Giuseppe Buttazzo, "Su una definizione dei Γ-limiti," *Bollettino Unione Matematica Italiana* **14-B** (1977), 722–744.

[20] Charles Castaing & Michel Valadier, *Convex Analysis and Measurable Multifunctions*, Springer Verlag,Lecture Notes in Mathematics 580, Berlin, 1977.

[21] Frank H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York, 1983.

[22] Michael Crandall, L. Craig Evans & Pierre-Louis Lions, "Some properties of the viscosity solutions of Hamilton-Jacobi equations," *Tansactions of the American Mathematical Society* **282** (1984), 487–502.

[23] Gianni Dal Maso, Ennio De Giorgi & Luciano Modica, "Weak convergence of measures on spaces of lower semicontinuous functions," ISAS Report 8/86/M, Trieste, 1986.

[24] Ennio De Giorgi, "Convergence problems of functionals and operators," in *Recent Methods in Nonlinear Analysis*, E. De Giorgi, E. Magenes & U. Mosco, eds., Pitagora Editrice, Bologna, 1979, 131–188.

[25] Szymon Dolecki, "Convergence of minima in convergence spaces," *Optimization* **17** (1986), 553–572.

[26] Alain Fougeres & Annik Truffert, "Régularisation s.c.i. et épiconvergence: approximations inf-convolutives associées à un référentiel," *Annali di Matematica Puri ed Applicati* (1986), preprint AVAMAC# 84-13.

[27] Halina Frankowska, "Optimal trajectories associated to a solution of contingent Hamilton-Jacobi equation," Working Paper WP-87-069, I.I.A.S.A., Laxenburg, June 1987.

[28] Christian Hess, "Quelques théorèmes limites pour des ensembles aléatoires bornés ou non," *Séminaire d'Analyse Convexe, Université du Languedoc* **14** (1984), 12:1–12:50.

[29] Jean-Baptiste Hiriart-Urruty & Marie-Laurence Mazure, "Formulations variationnelles de l'addition parallèle et de la soustraction parallèle d'opérateurs semi-definis positifs," *Comptes Rendus de l'Académie des Sciences de Paris* **302** (1986), 527–530.

[30] Jean-Luc Joly, "Une famille de topologies sur l'ensemble des fonstions convexes pour lesquelles la polarité est bicontinue," *J. Mathématiques Pures et Appliquées* **52** (1973), 421–441.

[31] Pierre-Jean Laurent, *Approximation et Optimisation*, Hermann, Paris, 1972.

[32] Jean-Jacques Moreau, "Rafle par un convexe variable, 1ère partie.," *Séminaire d'Analyse Convexe, Université de Montpellier* **1** (1971), 15:1–15:xx.

[33] Jean-Jacques Moreau, "Rafle par un convexe variable, 2ème partie.," *Séminaire d'Analyse Convexe, Université de Montpellier* **2** (1972), 3:1–3:xx.

[34] James-Louis Ndoutoume, "Calcul différentiel généralisé du second ordre pour des fonctions convexes," AVAMAC, Université de Perpignan, Vol.2, no. 2., 1986.

[35] Jean-Paul Penot, "Preservation of persistence and stability under intersection and operations," Manuscript, Université de Pau, 1986, (to appear).

[36] Stephen M. Robinson, "Local epi-continuity and local optimization," *Mathematical Programming* **37** (1987), 208–222.

[37] R.T. Rockafellar, "First and second-order epi-differentiability in nonlinear programming," Manuscript, University of Washington, Seattle, 1987.

[38] R.T. Rockafellar, "Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives," Manuscript, University of Washington, Seattle, 1987.

[39] R.T. Rockafellar & Roger J-B Wets, "Variational systems, an introduction," in *Multifunctions and Integrands: Stochastic Analysis, Approximation and Optimization*, G. Salinetti, ed., Springer Verlag,Lecture Notes in Mathematics 1091, Berlin, 1984, 1–54.

[40] Michel Volle, "Equations inf-convolutives," Manuscript, Université de Limoges, February 1988.