Journal of Convex Analysis
Volume 8 (2001), No. 2, 423–446

The Conjugates, Compositions and Marginals of Convex Functions

S. P. Fitzpatrick
Department of Mathematics and Statistics,
University of Western Australia, Nedlands 6907, Australia
fitzpatr@maths.uwa.edu.au

S. Simons
Department of Mathematics,
University of California, Santa Barbara, CA 96106-3080, USA
simons@math.ucsb.edu

Received November 6, 2000
Revised manuscript received May 18, 2001

Continuing the work of Hiriart-Urruty and Phelps, we discuss (in both locally convex spaces and Banach spaces) the formulas for the conjugates and subdifferentials of the precomposition of a convex function by a continuous linear mapping and the marginal function of a convex function by a continuous linear mapping. We exhibit a certain (incomplete) duality between the operations of precomposition and marginalization. Our results lead easily to Thibault's proof of the maximal monotonicity of the subdifferential of a proper, convex lower semicontinuous function on a Banach space. We show that some of the Hiriart-Urruty—Phelps results on ε–subdifferentials have analogs in terms of the "ε–enlargement" of the subdifferential. We obtain new results on the conjugates and subdifferentials of sums of convex functions without constraint qualifications and also of episums of convex functions. We discuss constrained minimization on *non–closed* convex subsets of a Banach space.

1. Introduction

This paper is inspired by the paper [5] by Hiriart-Urruty and Phelps, in which a set of calculus rules is presented for the subdifferentials of convex functions on locally convex spaces. Before discussing in detail the results that we are going to present, we will say a little about the order of presentation, since it is rather unusual for this kind of analysis. We consider the marginal function of a convex function through a continuous linear map, and the precomposition of a convex function by a continuous linear map, *before* considering the sums and episums of convex functions. Typically, authors consider sums and episums first and *deduce* results on the precompositions and marginal functions. The advantage of our order of presentation is that it leads to simpler proofs of results on the sums or episums of more than two functions. There are also more technical advantages, that are discussed in the preambles to Theorems 5.1 and 7.1.

Much of our analysis is for locally convex spaces (we assume that the vector spaces are over the real numbers, as is usual in convex analysis, and that all locally convex spaces are Hausdorff). We mention here the paper [4] by Hiriart-Urruty, Moussaoui, Seeger and Volle, which contains a number of other results on subdifferentials and approximate subdifferentials in the locally convex situation. The deeper results in this paper are in the Banach space context, where we use the Brøndsted–Rockafellar theorem for appropriately chosen renormings to derive characterizations of various subdifferentials. This is also a

ISSN 0944-6532 / $ 2.50 © Heldermann Verlag

good place to mention the paper [18] by Thibault, in which the Brøndsted–Rockafellar theorem is used to derive from the results of [5] limiting calculus rules for the exact subdifferential.

In Section 2 of the paper, we discuss the formulas for the conjugates of the two common ways of operating on a convex function by a continuous linear mapping A from E to F: for a convex function h on F we have the precomposition $h \circ A$ defined on E, while for a convex function f on E we have the marginal function f/A of f through A, defined on F by

$$f/A(y) := \inf\{f(x) : x \in A^{-1}y\}.$$

The first of these results, Lemma 2.4, is well known. We have not seen the second of them, Theorem 2.7, in the literature — the formula here is more indirect, and proceeds through the epigraphs of the functions concerned. We have organized our presentation in order to exhibit a certain duality between the operations of precomposition and marginalization. As Example 2.6 and Conjecture 6.3 show, this duality is not complete.

In Section 3, we discuss the subdifferential of $h \circ A$ under various conditions. Theorem 3.3 gives a characterization of this subdifferential in terms of a net from the subdifferential of h, in the case when F is a Banach space and h is lower semicontinuous on F. We also sketch in Remark 3.4 a short proof by Thibault that shows how (16), the formally weakest condition in Theorem 3.3, implies that the subdifferential of a proper, convex lower semicontinuous function on a Banach space is maximal monotone.

Section 4 is also related to monotonicity. We define the *ε–enlargement* of a multifunction and show in Theorem 4.2 that if F is a Banach space and h is proper, convex and lower semicontinuous then the characterization of the subdifferential of $h \circ A$ due to Hiriart-Urruty and Phelps in terms of the ε–subdifferential of h (see Theorem 3.1) has an analog in terms of the ε–enlargement of the subdifferential of h. The proof of this result is not easy, relying as it does on the approximation result proved in Theorem 3.3, and also on the maximal monotonicity of the subdifferential of a proper, convex lower semicontinuous function on a Banach space that we mentioned above.

In Section 5, the first of two bootstrapping sections, we obtain results on the conjugates and subdifferentials of sums of convex functions without constraint qualifications. Some of these extend results from [5], as well as more recent results of Thibault, [18], and Revalski and Théra, [10].

Section 6 is "dual" to Section 3, and contains results on the subdifferentials of marginal functions. The main result here is Theorem 6.2, in which we characterize the subdifferential of f/A in terms of a net from the subdifferential of f, in the case when E is a Banach space and f is lower semicontinuous on E.

Section 7 is the second bootstrapping section. The main result here is Theorem 7.1, in which we obtain results on the conjugates and subdifferentials of episums of convex functions.

In the final Section 8, we use Corollary 5.2 to obtain necessary and sufficient conditions for a proper, convex, lower semicontinuous function on a Banach space to attain a minimum at a certain point of a *non–closed* convex subset.

The authors would like to express their gratitude to M. Théra and L. Thibault for sending them preprints of some of their recent work, including [10] and [18]. Most of the work in this paper was done while the second–named author was visiting the University of Western Australia in Perth, and he would like to thank the University of Western Australia for its hospitality.

2. Conjugate functions in locally convex spaces

Let E be a nontrivial locally convex space and $\mathcal{PC}(E)$ stand for the set of all convex functions $f\colon\ E \mapsto \mathbb{R} \cup \{\infty\}$ such that $\operatorname{dom} f \neq \emptyset$, where the *effective domain* of f, $\operatorname{dom} f$, is defined by

$$\operatorname{dom} f := \{x \in E\colon\ f(x) \in \mathbb{R}\}.$$

(The "$\mathcal{P}$" stands for "proper", which is the adjective frequently used to denote the fact that the effective domain of a function is nonempty.) We write E^* for the dual space of E. If $f \in \mathcal{PC}(E)$, the *Fenchel conjugate*, f^*, of f is the function from E^* into $\mathbb{R} \cup \{\infty\}$ defined by

$$f^*(x^*) := \sup_E (x^* - f).$$

We write $\mathcal{PCPC}(E)$ for the set of all those $f \in \mathcal{PC}(E)$ such that $\operatorname{dom} f^* \neq \emptyset$. (The extra "$\mathcal{PC}$" stands for "proper conjugate".) If $f \in \mathcal{PC}(E)$, we write "$\operatorname{epi} f$" and "$\operatorname{stepi} f$" for the *epigraph* and *strict epigraph* of f, defined by

$$\operatorname{epi} f := \{(x,\lambda) \in E \times \mathbb{R}\colon\ f(x) \le \lambda\} \quad \text{and} \quad \operatorname{stepi} f := \{(x,\lambda) \in E \times \mathbb{R}\colon\ f(x) < \lambda\}.$$

The main result in this section is Theorem 2.7, in which we give an (indirect) formula for the conjugate of the precomposition of a convex function by a linear map.

Lemma 2.1. *Let F be a vector space, $S\colon\ F \mapsto \mathbb{R}$ be sublinear, $\varphi \in \mathcal{PC}(F)$ and $x_0 \in F$. Then there exists a linear functional L on F such that $L \le S$ on F and*

$$\inf_{x \in F} [L(x - x_0) + \varphi(x)] = \inf_{x \in F} [S(x - x_0) + \varphi(x)].$$

Proof. This follows from the *sandwich theorem* (see König, [6, Theorem 1.7, p. 112]) and a translation argument, or from the more general result of [15, Theorem 5.3]. □

Theorem 2.2 is a basic equivalence.

Theorem 2.2. *Let F be a locally convex space, $\varphi \in \mathcal{PCPC}(F)$, and $(x_0, \alpha) \in F \times \mathbb{R}$. Then (1) $\Longleftrightarrow$ (2).*

$$\left.\begin{array}{r} \textit{For all } u^* \in F^* \textit{ and } \varepsilon > 0, \textit{there exists } x \in F \textit{ such that} \\ \langle x_0 - x, u^* \rangle + \varphi(x) < \alpha + \varepsilon. \end{array}\right\} \tag{1}$$

$$(x_0, \alpha) \in \overline{\operatorname{stepi} \varphi}. \tag{2}$$

(The overbar in (2) stands for the closure in the product topology of $F \times \mathbb{R}$.)

Proof. ($\Longrightarrow$) Suppose that (2) is false. Then there exist a continuous seminorm p on F and $\varepsilon > 0$ such that

$$p(x - x_0) < 1 \text{ and } |\lambda - \alpha| < \varepsilon \quad \Longrightarrow \quad \varphi(x) \ge \lambda,$$

from which it follows that

$$p(x - x_0) < 1 \quad \Longrightarrow \quad \varphi(x) \geq \alpha + \varepsilon. \tag{3}$$

Since $\varphi \in \mathcal{PCPC}(F)$, there exists $w^* \in F^*$ such that $\varphi^*(w^*) \in \mathbb{R}$. Let

$$\beta := (\alpha + \varepsilon + \varphi^*(w^*) - \langle x_0, w^* \rangle) \vee 0 \quad \text{and} \quad q := \beta p + |w^*|.$$

q is a continuous seminorm on F. We shall prove that

$$x \in F \quad \Longrightarrow \quad q(x - x_0) + \varphi(x) \geq \alpha + \varepsilon. \tag{4}$$

Since $q \geq 0$ on F, (4) is immediate from (3) if $p(x - x_0) < 1$, so we can and will suppose that $p(x - x_0) \geq 1$. Then

$$\begin{aligned} q(x - x_0) + \varphi(x) &\geq \beta p(x - x_0) - \langle x - x_0, w^* \rangle + \varphi(x) \\ &\geq \beta - \langle x - x_0, w^* \rangle + \varphi(x) \\ &\geq \alpha + \varepsilon + \varphi^*(w^*) - \langle x_0, w^* \rangle - \langle x - x_0, w^* \rangle + \varphi(x) \\ &= \alpha + \varepsilon + \varphi^*(w^*) - \langle x, w^* \rangle + \varphi(x), \end{aligned}$$

which gives (4). It follows from (4) and Lemma 2.1 that there exists $u^* \in F^*$ such that

$$x \in F \quad \Longrightarrow \quad \langle x_0 - x, u^* \rangle + \varphi(x) \geq \alpha + \varepsilon.$$

Thus (1) fails, which completes the proof of ($\Longrightarrow$).

($\Longleftarrow$) Let $u^* \in F^*$ and $\varepsilon > 0$. Since $\{x \in F\colon\ |\langle x - x_0, u^* \rangle| < \varepsilon/2\}$ is a neighborhood of x_0 in F, it follows from (2) that there exists $(x, \lambda) \in \operatorname{stepi} \varphi$ such that $|\langle x - x_0, u^* \rangle| < \varepsilon/2$ and $|\lambda - \alpha| < \varepsilon/2$. Thus $\langle x_0 - x, u^* \rangle + \varphi(x) - \alpha < \langle x_0 - x, u^* \rangle + \lambda - \alpha < \varepsilon/2 + \varepsilon/2 = \varepsilon$, and (1) now follows. □

We will actually use the following special case of Theorem 2.2. Significant choices for $\mathcal{C}$ are the *weak* topology*, $w(E^*, E)$ and the *Mackey topology*, $m(E^*, E)$. If E is a reflexive Banach space then we can take $\mathcal{C}$ to be the norm topology of E^*.

Corollary 2.3. *Let E be a locally convex space with dual E^*. Let $\mathcal{C}$ be any locally convex topology on E^* giving E as dual. Let $\varphi \in \mathcal{PCPC}(E^*, \mathcal{C})$ and $(x_0^*, \alpha) \in E^* \times \mathbb{R}$. Then (5) $\Longleftrightarrow$ (6).*

$$\left.\begin{array}{r} \textit{For all } u \in E \textit{ and } \varepsilon > 0, \textit{there exists } x^* \in E^* \textit{ such that} \\ \langle u, x_0^* - x^* \rangle + \varphi(x^*) < \alpha + \varepsilon. \end{array}\right\} \tag{5}$$

$$(x_0^*, \alpha) \in \overline{\operatorname{stepi} \varphi}. \tag{6}$$

(The overbar in (6) stands for the closure in the product topology of $(E^*, \mathcal{C}) \times \mathbb{R}$.)

If $A\colon\ E \mapsto F$ and $f\colon E \mapsto [-\infty, \infty]$ then, for all $y \in F$, we write

$$f/A(y) := \inf f(A^{-1}y),$$

with the convention that $f/A(y) := \infty$ if $A^{-1}y = \emptyset$. The function f/A is the "marginal function of f through A" – sometimes known as the "image of f under A" or "Shur complement of f relative to A". We now give the formula for the conjugate of a marginal function under the appropriate topological and convexity assumptions. Lemma 2.4 below is a slightly more detailed form of Aubin and Ekeland, [1, Proposition 4.4.10(a), p. 206]. We have chosen to write f/A instead of Af or f_A, which are notations used by some authors, since it fits in quite naturally with the definition, and also with (8).

Lemma 2.4. *Let E and F be locally convex spaces, $A\colon\ E \mapsto F$ be continuous and linear, $f \in \mathcal{PC}(E)$ and $f^* \circ A^* \in PC(F^*)$. Then $f/A \in \mathcal{PCPC}(F)$ and $(f/A)^* = f^* \circ A^*$.*

Proof. We first observe that if $y \in F$, $x \in A^{-1}y$ and $y^* \in \operatorname{dom} f^* \circ A^*$ then

$$f(x) \geq \langle x, A^*y^*\rangle - f^*(A^*y^*) = \langle y, y^*\rangle - (f^* \circ A^*)(y^*). \tag{7}$$

Taking the infimum over x, $f/A(y) \geq \langle y, y^*\rangle - (f^* \circ A^*)(y^*) > -\infty$. Thus f/A maps F into $\mathbb{R} \cup \{\infty\}$. It can be verified by direct computation that f/A is convex on F. It is also clear that

$$x \in E \quad \Longrightarrow \quad f(x) \geq (f/A)(Ax), \tag{8}$$

from which it follows that $f/A \in \mathcal{PC}(F)$. Finally, if $y^* \in F^*$ then

$$\begin{aligned} (f/A)^*(y^*) &= \sup_{y \in F}[\langle y, y^*\rangle - f/A(y)] = \sup_{y \in F}\sup_{x \in A^{-1}y}[\langle y, y^*\rangle - f(x)] \\ &= \sup_{y \in F}\sup_{x \in A^{-1}y}[\langle x, A^*y^*\rangle - f(x)] = \sup_{x \in E}[\langle x, A^*y^*\rangle - f(x)] = f^*(A^*y^*). \end{aligned}$$

So $(f/A)^* = f^* \circ A^* \in \mathcal{PC}(F^*)$, and consequently $f/A \in \mathcal{PCPC}(F)$. □

It is tempting to hope, by "duality", that under certain conditions we could strengthen Lemma 2.5 below and prove that $h^*/A^* = (h \circ A)^*$; however, Example 2.6 shows that, even in the simplest of situations, this hope is unjustified.

Lemma 2.5. *Let E and F be locally convex spaces, $A\colon\ E \mapsto F$ be continuous and linear, $h \in \mathcal{PCPC}(F)$ and $h \circ A \in \mathcal{PC}(E)$. Then*

$$h^*/A^* \in \mathcal{PCPC}(E^*, w(E^*, E)) \quad \text{and} \quad h^*/A^* \geq (h \circ A)^* \text{ on } E^*.$$

Proof. We give two proofs of this result. The first uses Lemma 2.4, however it contains a number of technical computations using biconjugates. The second, slightly longer, proof is independent of Lemma 2.4, and avoids these computations.

For the first proof, we observe by direct computation that

$$(h^*)^* \circ (A^*)^* \leq h \circ A \quad \text{on } (E^*)^* = E. \tag{9}$$

Since this implies that $(h^*)^* \circ (A^*)^* \in \mathcal{PC}((E^*)^*)$, we can apply Lemma 2.4 with E replaced by $(F^*, w(F^*, F))$, F replaced by $(E^*, w(E^*, E))$, A by A^* and f by $h^* \in \mathcal{PC}(F^*)$. Thus $h^*/A^* \in \mathcal{PCPC}(E^*)$ (as required) and $(h^*/A^*)^* = (h^*)^* \circ (A^*)^*$ on $(E^*)^* = E$. We see from (9) that $(h^*/A^*)^* \leq h \circ A$ on E, thus $h^*/A^* \geq ((h^*/A^*)^*)^* \geq (h \circ A)^*$ on E^*. This completes the first proof of the Lemma.

For the second proof, we start off by observing that if $x^* \in E^*$, $y^* \in (A^*)^{-1}x^*$ and $x \in \operatorname{dom} h \circ A$ then

$$h^*(y^*) \geq \langle Ax, y^*\rangle - h(Ax) = \langle x, x^*\rangle - (h \circ A)(x). \tag{10}$$

Taking the infimum over y^*, $h^*/A^*(x^*) \geq \langle x, x^*\rangle - (h\circ A)(x) > -\infty$. Thus h^*/A^* maps E^* into $\mathbb{R}\cup\{\infty\}$. It can be verified by direct computation that h^*/A^* is convex on E^*, and it is also clear that $y^* \in F^* \implies h^*(y^*) \geq (h^*/A^*)(A^*y^*)$, from which it follows that $h^*/A^* \in \mathcal{PC}(E^*)$. Returning to (10) and taking the supremum over $x \in \operatorname{dom} h\circ A$ before taking the infimum over y^*, we obtain $h^*/A^*(x^*) \geq (h\circ A)^*(x^*)$. Thus $h^*/A^* \geq (h\circ A)^*$ on E^*, as required. Finally, taking $x \in \operatorname{dom} h\circ A$ and writing $\widehat{x}$ for the canonical image of x in $(E^*)^*$,

$$\begin{aligned}(h^*/A^*)^*(\widehat{x}) &= \sup_{x^*\in E^*}[\langle x^*, \widehat{x}\rangle - (h^*/A^*)(x^*)] \\ &\leq \sup_{x^*\in E^*}[\langle x, x^*\rangle - (h\circ A)^*(x^*)] \leq (h\circ A)(x) < \infty,\end{aligned}$$

and consequently $h^*/A^* \in \mathcal{PCPC}(E^*)$. □

We write $\mathcal{PCLSC}(F)$ for the set of all those elements h of $\mathcal{PC}(F)$ that are lower semicontinuous.

Example 2.6. Let $E = \mathbb{R}$, $F = \mathbb{R}^2$ and $A\colon\ E \mapsto F$ be defined by $A\lambda := (0, \lambda)$. Write C for the epigraph of the exponential function, and h for the support functional of C. Then $h \in \mathcal{PCLSC}(F)$, and $h^* \in \mathcal{PCLSC}(F^*)$ is the indicator function of C. However h^*/A^* is the indicator function of $(0, \infty)$ and is, therefore, not lower semicontinuous. Consequently, since $(h\circ A)^*$ is lower semicontinuous, we cannot have $h^*/A^* = (h\circ A)^*$.

We now strengthen the assumptions of Lemma 2.5 by assuming that $h \in \mathcal{PCLSC}(F)$. Theorem 2.7 shows that, even though it is not generally true that $h^*/A^* = (h\circ A)^*$, they are "close" (if we measure them by their epigraphs). Before starting on Theorem 2.7, we will make some general comments about $\mathcal{PCLSC}(F)$. The *Fenchel–Moreau formula* asserts that if $h \in \mathcal{PCLSC}(F)$ then

$$y \in F \quad\implies\quad h(y) = \sup_{y^*\in F^*}[\langle y, y^*\rangle - h^*(y^*)].$$

It follows from this that $\mathcal{PCLSC}(F) \subset \mathcal{PCPC}(F)$.

Theorem 2.7. *Let E and F be locally convex spaces, $A\colon\ E \mapsto F$ be continuous and linear, $h \in \mathcal{PCLSC}(F)$ and $h\circ A \in \mathcal{PC}(E)$. Let $\mathcal{C}$ be any locally convex topology on E^* giving E as dual. Then*

$$\operatorname{epi}(h\circ A)^* = \overline{\operatorname{epi} h^*/A^*} = \overline{\operatorname{stepi} h^*/A^*}.$$

(The overbar above stands for the closure in the product topology of $(E^*, \mathcal{C}) \times \mathbb{R}$.)

Proof. It is obvious from Lemma 2.5 that $\operatorname{epi}(h\circ A)^* \supset \operatorname{epi} h^*/A^*$. Since $\operatorname{epi}(h\circ A)^*$ is closed in $E^* \times \mathbb{R}$, it follows that $\operatorname{epi}(h\circ A)^* \supset \overline{\operatorname{epi} h^*/A^*}$. We also have $\operatorname{epi} h^*/A^* \supset \operatorname{stepi} h^*/A^*$, so it only remains to prove that

$$\operatorname{epi}(h\circ A)^* \subset \overline{\operatorname{stepi} h^*/A^*}. \tag{11}$$

So suppose that $(x_0^*, \alpha) \in \operatorname{epi}(h\circ A)^*$, that is to say, $(h\circ A)^*(x_0^*) \leq \alpha$. Let $u \in E$ and $\varepsilon > 0$. Then

$$\langle u, x_0^*\rangle - h(Au) = \langle u, x_0^*\rangle - (h\circ A)(u) < \alpha + \varepsilon.$$

Since $h \in \mathcal{PCLSC}(F)$, the Fenchel–Moreau formula gives that

$$\text{there exists } y^* \in F^* \text{ such that } \langle u, x_0^* \rangle - [\langle Au, y^* \rangle - h^*(y^*)] < \alpha + \varepsilon,$$

that is to say

$$\text{there exists } y^* \in F^* \text{ such that } \langle u, x_0^* - A^* y^* \rangle + h^*(y^*) < \alpha + \varepsilon.$$

Combining with (8),

$$\text{there exists } y^* \in F^* \text{ such that } \langle u, x_0^* - A^* y^* \rangle + (h^*/A^*)(A^* y^*) < \alpha + \varepsilon.$$

We note from Lemma 2.5 that $h^*/A^* \in \mathcal{PCPC}(E^*, w(E^*, E))$, which implies that $h^*/A^* \in \mathcal{PCPC}(E^*, \mathcal{C})$. We now derive from Corollary 2.3, with $\varphi := h^*/A^*$ and $x^* := A^* y^*$, that

$$(x_0^*, \alpha) \in \overline{\operatorname{stepi} h^*/A^*}.$$

This gives (11), and completes the proof of Theorem 2.7. □

Remark 2.8. It follows from Theorem 2.7 that if E is a locally convex space, for all $k = 1, \ldots, K$, $h_k \in \mathcal{PCLSC}(E)$ and $\bigcap_{k=1}^{K} \operatorname{dom} h_k \neq \emptyset$ then

$$\operatorname{epi}(h_1 + \cdots + h_K)^* = \overline{\operatorname{epi} h_1^* + \cdots + \operatorname{epi} h_K^*}.$$

We do not give details of this here, since we will be proving a more general result in Theorem 5.1(a).

Remark 2.9. A couple of issues are raised by Theorem 2.2 and Corollary 2.3. The first is the method of proof of Theorem 2.2. The Eidelheit separation theorem could also be used instead of Lemma 2.1. However, one then has to contend with the problem of the "vertical hyperplane". The other is the requirement that $\varphi \in \mathcal{PCPC}(E^*)$ in Corollary 2.3. If we define

$$\varphi^{**}(x^*) := \sup_{x \in E} \big[\langle x, x^* \rangle - \varphi^*(x) \big] \qquad (x^* \in E^*)$$

then it follows from Corollary 2.3 that if $\alpha \in \mathbb{R}$ and $\varepsilon > 0$ then

$$x_0^* \in E^* \text{ and } \varphi^{**}(x_0^*) \leq \alpha \quad \Longrightarrow \quad x_0^* \in \overline{\{x^* \in E^* \colon \ \varphi(x^*) < \alpha + \varepsilon\}}. \tag{12}$$

We leave details of this to the reader. However, (12) fails if we assume merely that $\varphi \in \mathcal{PC}(E^*)$. Suppose that there is a discontinuous linear functional g on E^*. Let f be the indicator function of a closed half–space C of E^*, and $\varphi := f + g$. Now $f^* \colon E \mapsto \mathbb{R} \cup \{\infty\}$, and g^* is identically $+\infty$ on E. Further, f and g satisfy the Moreau–Rockafellar constraint qualification, so $\varphi^* = f^* \underset{e}{+} g^*$, and consequently $\varphi^* = +\infty$ on E. It follows that $\varphi^{**} = -\infty$ on E^*, from which

$$\{x^* \in E^* \colon \ \varphi^{**}(x^*) \leq 0\} = E^*.$$

On the other hand,

$$\overline{\{x^* \in E^* \colon \ \varphi(x^*) \leq 1\}} \subset \overline{\{x^* \in E^* \colon \ f(x^*) = 0\}} = C.$$

We mention this because the discussion preceding (2.1) of [5] seems to imply that (12) holds when we only assume that $\varphi \in \mathcal{PC}(E^*)$. In reality, this does not cause any problem in [5] since $\varphi \in \mathcal{PCPC}(E^*)$ when (2.1) is used there.

3. Subdifferentials of precompositions

If $f \in \mathcal{PC}(E)$ and $x \in E$, the *subdifferential* of f at x is the subset of E^* defined by

$$\partial f(x) := \{z^* \in E^*: \quad y \in E \Longrightarrow f(x) + \langle y - x, z^* \rangle \le f(y)\}.$$

If, further, $\varepsilon > 0$, the *ε–subdifferential* of f at x is the larger subset of E^* defined by

$$\partial_\varepsilon f(x) := \{z^* \in E^*: \quad y \in E \Longrightarrow f(x) + \langle y - x, z^* \rangle \le f(y) + \varepsilon\}.$$

It is then easy to see that

$$x^* \in \partial f(x) \iff f(x) + f^*(x^*) \le \langle x, x^* \rangle$$

and

$$x^* \in \partial_\varepsilon f(x) \iff f(x) + f^*(x^*) \le \langle x, x^* \rangle + \varepsilon.$$

We now show how the results of the previous section on conjugate functions lead to results on subdifferentials and ε-subdifferentials. Theorem 3.1 was proved by Hiriart–Urruty and Phelps in [5, Theorem 3.1, p. 161–163], using results on the sums of functions (see also Hiriart-Urruty, Moussaoui, Seeger and Volle, [4, Corollary 7.1, p. 1742]) . We shall reverse the procedure and deduce (in Theorem 5.1) more general results on sums from Theorem 3.1. This leads to much easier proofs in the case where more than two functions are being added. Our approach using precompositions also allows us to give a fairly simple proof of Theorem 3.3, in which E is not required to be a Banach space.

Theorem 3.1. *Let E and F be locally convex spaces, $A\colon\ E \mapsto F$ be continuous and linear, $h \in \mathcal{PCLSC}(F)$, $h \circ A \in \mathcal{PC}(E)$ and $x \in E$. Let $\mathcal{C}$ be any locally convex topology on E^* giving E as dual. Then*

$$\partial(h \circ A)(x) = \bigcap_{\varepsilon > 0} \overline{A^*\partial_\varepsilon h(Ax)}.$$

(The overbar above stands for the closure in $(E^*, \mathcal{C})$. Of course, since $A^*\partial_\varepsilon h(Ax)$ is convex, this is identical with the closure in $(E^*, w(E^*, E))$.)

Proof. The inclusion "$\supset$" is immediate since, for all $\varepsilon > 0$, $\partial_\varepsilon(h \circ A)(x)$ is closed in $\mathcal{C}$, and so $\overline{A^*\partial_\varepsilon h(Ax)} \subset \partial_\varepsilon(h \circ A)(x)$. We now prove "$\subset$". Suppose that $x^* \in \partial(h \circ A)(x)$. Let U be a $\mathcal{C}$–neighborhood of x^* in E^* and $\varepsilon > 0$. Then

$$(x^*, \langle x, x^* \rangle - h(Ax)) = (x^*, \langle x, x^* \rangle - (h \circ A)(x)) = (x^*, (h \circ A)^*(x^*)) \in \operatorname{epi}(h \circ A)^*.$$

Thus from Theorem 2.7, for all $\varepsilon > 0$, there exists $(u^*, \lambda) \in \operatorname{epi} h^*/A^*$ such that

$$u^* \in U, \quad |\langle x, u^* - x^* \rangle| < \varepsilon/2 \quad \text{and} \quad |\lambda - [\langle x, x^* \rangle - h(Ax)]| < \varepsilon/2,$$

from which

$$u^* \in U \quad \text{and} \quad h^*/A^*(u^*) < \langle x, u^* \rangle - h(Ax) + \varepsilon.$$

It follows from the definition of h^*/A^* that there exists $y^* \in F^*$ such that $A^*y^* = u^* \in U$ and $h^*(y^*) < \langle x, u^* \rangle - h(Ax) + \varepsilon = \langle Ax, y^* \rangle - h(Ax) + \varepsilon$, from which $y^* \in \partial_\varepsilon h(Ax)$. Thus $x^* \in \overline{A^*\partial_\varepsilon h(Ax)}$. Since this holds for all $\varepsilon > 0$, this completes the proof of the Theorem. $\square$

We now turn our attention to the situation when F is a Banach space. We will need the Brøndsted–Rockafellar theorem, which we state as Theorem 3.2 below (see [2, p. 608] or Phelps, [9, Theorem 3.17, p. 48]):

Theorem 3.2. *Let F be a Banach space, $h \in \mathcal{PCLSC}(F)$, $\alpha, \beta > 0$, $u \in F$ and $u^* \in \partial_{\alpha\beta} h(u)$. Then there exists $(z, z^*) \in G(\partial h)$ such that $\|z - u\| \le \alpha$ and $\|z^* - u^*\| \le \beta$.*

If $x \in E$ and $\emptyset \ne Y \subset E$, $\operatorname{dist}(x, Y)$ stands for the distance from x to Y, that is to say, $\operatorname{dist}(x, Y) := \inf_{y \in Y} \|x - y\|$. Further, $G(\partial h)$ stands for the graph of the multifunction ∂h, that is to say, $G(\partial h) := \{(z, z^*) \in F \times F^* \colon\ z^* \in \partial h(z)\}$.

Theorem 3.3. *Let E be a locally convex space, F be a Banach space, $A\colon E \mapsto F$ be continuous and linear, $h \in \mathcal{PCLSC}(F)$, $h \circ A \in \mathcal{PC}(E)$ and $(x, x^*) \in E \times E^*$. Let $\mathcal{C}$ be any locally convex topology on E^* giving E as dual. Then the conditions (13)–(16) are equivalent:*

$$x^* \in \partial(h \circ A)(x). \tag{13}$$

$$\left.\begin{array}{c}\text{There exists a net } (z_\gamma, z_\gamma^*) \text{ of elements of } G(\partial h) \text{ such that } \|z_\gamma - Ax\| \to 0, \\ A^* z_\gamma^* \to x^* \text{ in } \mathcal{C},\ \langle z_\gamma - Ax, z_\gamma^*\rangle \to 0 \text{ and } \operatorname{dist}(z_\gamma, A(E))\|z_\gamma^*\| \to 0.\end{array}\right\} \tag{14}$$

$$\left.\begin{array}{c}\text{There exists a net } (z_\gamma, z_\gamma^*) \text{ of elements of } G(\partial h) \text{ such that } \|z_\gamma - Ax\| \to 0, \\ A^* z_\gamma^* \to x^* \text{ in } \mathcal{C},\ \langle z_\gamma, z_\gamma^*\rangle \to \langle x, x^*\rangle \text{ and } \operatorname{dist}(z_\gamma, A(E))\|z_\gamma^*\| \to 0.\end{array}\right\} \tag{15}$$

$$\left.\begin{array}{c}\text{There exists a net } (z_\gamma, z_\gamma^*) \text{ of elements of } G(\partial h) \text{ such that } \|z_\gamma - Ax\| \to 0, \\ A^* z_\gamma^* \to x^* \text{ in } \mathcal{C} \text{ and } \langle z_\gamma, z_\gamma^*\rangle \to \langle x, x^*\rangle.\end{array}\right\} \tag{16}$$

Proof. ((14)$\Longrightarrow$(15)) This follows from the equality

$$\langle z_\gamma, z_\gamma^*\rangle = \langle z_\gamma - Ax, z_\gamma^*\rangle + \langle x, A^* z_\gamma^*\rangle.$$

((15)$\Longrightarrow$(16)) This is immediate.

((16)$\Longrightarrow$(13)) Let (z_γ, z_γ^*) be as in the statement of (16). Let v be an arbitrary element of E. Then $Av \in F$ hence, for all γ, since $(z_\gamma, z_\gamma^*) \in G(\partial h)$,

$$h(z_\gamma) + \langle v, A^* z_\gamma^*\rangle - \langle z_\gamma, z_\gamma^*\rangle = h(z_\gamma) + \langle Av - z_\gamma, z_\gamma^*\rangle \le h(Av).$$

Passing to the limit and using the lower semicontinuity of h,

$$h(Ax) + \langle v, x^*\rangle - \langle x, x^*\rangle \le h(Av),$$

that is to say,

$$(h \circ A)(x) + \langle v - x, x^*\rangle \le (h \circ A)(v),$$

Thus $(x, x^*) \in G(\partial(h \circ A))$. This completes the proof of (13).

((13)$\Longrightarrow$(14)) Let $\varepsilon > 0$, and q be any $\mathcal{C}$–continuous seminorm on E^*. We will first find $(z, z^*) \in G(\partial h)$ such that

$$\|z - Ax\| < \varepsilon, \quad |\langle z - Ax, z^*\rangle| < \varepsilon, \tag{17}$$

$$q(A^*z^* - x^*) < 1, \tag{18}$$

and

$$\operatorname{dist}(z, A(E))\|z^*\| < \varepsilon. \tag{19}$$

Let $V := \{v \in E\colon\ v \le q \text{ on } E^*\}$. If y^* is a fixed element of F^*, then

$$\sup_{v\in V} |\langle Av, y^*\rangle| = \sup_{v\in V} |\langle v, A^*y^*\rangle| \le q(A^*y^*) < \infty$$

hence, from the uniform boundedness theorem, $A(V)$ is bounded in F. Let

$$M := \sup_{v\in V} \|Av\| < \infty.$$

Choose $\alpha, \beta > 0$ so that $2\alpha(\beta + 1) < \varepsilon$ and $(2M + 2q(x^*) + 1)\beta < 1$. From Theorem 3.1, there exists $u^* \in \partial_{\alpha\beta}h(Ax)$ such that $q(A^*u^* - x^*) < 1/2$. Define an equivalent norm $|||\ |||$ on F (which depends on ε and q) by

$$|||y||| := \|y\| + |\langle y, u^*\rangle| + (1 + \|u^*\|)\operatorname{dist}(y, A(E)) \qquad (y \in F),$$

and write $|||\ |||$ for the corresponding dual norm on F^* also. Since $|||y||| \ge |\langle y, u^*\rangle|$, it follows that $|||u^*||| \le 1$. From the Brøndsted–Rockafellar theorem, there exists $(z, z^*) \in G(\partial h)$ such that $|||z - Ax||| \le \alpha$ and $|||z^* - u^*||| \le \beta$, from which $|||z^*||| \le |||z^* - u^*||| + |||u^*||| \le \beta + 1$. We will now show that (z, z^*) has the required properties. (17) follows since $\|z - Ax\| \le |||z - Ax||| \le \alpha < \varepsilon$ and

$$\begin{aligned} |\langle z - Ax, z^*\rangle| &\le |\langle z - Ax, z^* - u^*\rangle| + |\langle z - Ax, u^*\rangle| \\ &\le |||z - Ax|||\,|||z^* - u^*||| + |||z - Ax||| \\ &\le \alpha\beta + \alpha < \varepsilon. \end{aligned}$$

The proof of (18) is a little more complicated. From the one–dimensional Hahn–Banach theorem, there exists $v \in V$ such that $\langle v, A^*z^* - A^*u^*\rangle = q(A^*z^* - A^*u^*)$. Then, using the fact that $q(A^*u^* - x^*) < 1/2$,

$$\begin{aligned} |||Av||| &= \|Av\| + |\langle Av, u^*\rangle| + 0 = \|Av\| + |\langle v, A^*u^*\rangle| \\ &\le \|Av\| + q(A^*u^*) < \|Av\| + q(x^*) + 1/2 \le M + q(x^*) + 1/2. \end{aligned}$$

Consequently,

$$\begin{aligned} q(A^*z^* - A^*u^*) &= |\langle v, A^*z^* - A^*u^*\rangle| = |\langle Av, z^* - u^*\rangle| \\ &\le |||Av|||\beta < (M + q(x^*) + 1/2)\beta < 1/2. \end{aligned}$$

(18) now follows since $q(A^*z^* - x^*) \le q(A^*z^* - A^*u^*) + q(A^*u^* - x^*) < 1/2 + 1/2 = 1$. We now establish (19). For all $y \in F$, $|||y||| \le 2(1 + \|u^*\|)\|y\|$, hence

$$\|z^*\| \le 2(1 + \|u^*\|)|||z^*||| \le 2(1 + \|u^*\|)(\beta + 1).$$

Thus, since $\operatorname{dist}(z, A(E)) = \operatorname{dist}(z - Ax, A(E))$,

$$\begin{aligned} \operatorname{dist}(z, A(E))\|z^*\| &\le \operatorname{dist}(z - Ax, A(E))2(1 + \|u^*\|)(\beta + 1) \\ &= 2(1 + \|u^*\|)\operatorname{dist}(z - Ax, A(E))(\beta + 1), \\ &\le 2|||z - Ax|||(\beta + 1) \le 2\alpha(\beta + 1) < \varepsilon \end{aligned}$$

(using the definition of $|||z - Ax|||$), which gives (19). Using (17)–(19), it is easy to find a net (z_γ, z_γ^*) of elements of $G(\partial h)$ satisfying the conditions of (14). □

Remark 3.4. We now give an argument due to Thibault (see [18, Proposition 4]) which shows how condition (16) implies that if F is a Banach space, $h \in \mathcal{PCLSC}(F)$ and

$$(z, z^*) \in G(\partial h) \quad \Longrightarrow \quad \langle z, z^* \rangle \geq 0 \tag{20}$$

then

$$y \in \operatorname{dom} h \quad \Longrightarrow \quad h(y) \geq h(0).$$

It follows easily by translation arguments that if $f \in \mathcal{PCLSC}(F)$ then ∂f is maximal monotone — see [11] for Rockafellar's original proof of this, and [12], [13] and [14] for more general results. (Maximal monotonicity is defined in the introduction to Section 4 below). So let $y \in \operatorname{dom} h$. Define $A\colon\ \mathbb{R} \mapsto F$ by $A\lambda := \lambda y$. By semicontinuity, the function $\lambda \mapsto (h \circ A)(\lambda) + \lambda^2/2$ attains its minimum over $\mathbb{R}$ at some $x \in \mathbb{R}$, from which it follows easily that $-x \in \partial(h \circ A)(x)$. From (16), there exists a net (z_γ, z_γ^*) of elements of $G(\partial h)$ such that

$$\|z_\gamma - Ax\| \to 0, \quad A^* z_\gamma^* \to -x \text{ in } \mathbb{R} \tag{21}$$

and

$$\langle z_\gamma, z_\gamma^* \rangle \to \langle x, -x \rangle = -x^2. \tag{22}$$

(20) and (22) give $x = 0$, hence (21) gives $\|z_\gamma\| \to 0$ and $A^* z_\gamma^* \to 0$. Now, for all γ,

$$h(y) \geq h(z_\gamma) + \langle y - z_\gamma, z_\gamma^* \rangle = h(z_\gamma) + \langle 1, A^* z_\gamma^* \rangle - \langle z_\gamma, z_\gamma^* \rangle.$$

Passing to the limit and using the semicontinuity of h gives $h(y) \geq h(0)$, as required.

4. ε–enlargements of multifunctions and subdifferentials

Let E be a locally convex space and $S\colon\ E \mapsto 2^{E^*}$ be a multifunction (i.e, a set–valued map). We say that S is *monotone* if

$$(x, x^*) \text{ and } (y, y^*) \in G(S) \quad \Longrightarrow \quad \langle x - y, x^* - y^* \rangle \geq 0,$$

where $G(S)$ is the *graph* of S, that is to say, $G(S) := \{(x, x^*)\colon\ x \in E,\ x^* \in Sx\}$. We say that S is *maximal monotone* if S is monotone, and S has no proper monotone extension. Now let S be monotone and $\varepsilon > 0$. The *ε–enlargement* $S^\varepsilon\colon\ E \mapsto 2^{E^*}$ is defined by the following rule: if $(v, v^*) \in E \times E^*$ then $v^* \in S^\varepsilon v$ when

$$(x, x^*) \in G(S) \quad \Longrightarrow \quad \langle x - v, x^* - v^* \rangle \geq -\varepsilon.$$

It is clear from the definitions of monotonicity and maximal monotonicity that $Sv \subset S^\varepsilon v$, and that S is maximal monotone if, and only if, for all $v \in E$, $\bigcap_{\varepsilon>0} S^\varepsilon v \subset Sv$. Several authors have studied ε-enlargements: we refer the reader to Revalski and Théra [10], Burachik and Svaiter [3] and Svaiter [16].

So if $f \in \mathcal{PCLSC}(E)$, we now have two extensions of ∂f, namely $\partial_\varepsilon f$ and $(\partial f)^\varepsilon$. If $(v, v^*) \in G(\partial_\varepsilon f)$ and $(x, x^*) \in G(\partial f)$ then $v, x \in \operatorname{dom} f$ and

$$\langle x - v, x^* - v^* \rangle = [f(x) - f(v) + \langle v - x, v^* \rangle] + [f(v) - f(x) + \langle x - v, x^* \rangle] \geq -\varepsilon + 0.$$

Consequently, for all $v \in E$, $\partial_\varepsilon f(v) \subset (\partial f)^\varepsilon(v)$. It was also shown in Martines–Legaz and Théra, [7] that this inclusion can easily be strict: we can take $E := \mathbb{R}$, and $f(x) := x^2$. Then, for all $(v, v^*) \in \mathbb{R}^2$,

$$\inf_{(x,x^*) \in G(\partial f)} \langle x - v, x^* - v^* \rangle = -(2v - v^*)^2/8$$

but

$$\sup_{x\in\mathbb{R}}[f(v) - f(x) - \langle v - x, v^*\rangle] = (2v - v^*)^2/4.$$

Of course the difference between $\partial_\varepsilon f$ and $(\partial f)^\varepsilon$ is that the former is defined with reference to the values of f, while the latter is a much more general concept, defined with reference only to the values of the multifunction ∂f. We will show in Theorem 4.2 that the analog of Theorem 3.1 with $\partial_\varepsilon f$ replaced by $(\partial f)^\varepsilon$ is in fact true, and we will give another example, in Remark 4.3, where $\partial_\varepsilon f$ can almost be replaced by $(\partial f)^\varepsilon$, namely the Brøndsted–Rockafellar theorem.

Lemma 4.1. *Let E be a locally convex space, F be a Banach space, $A\colon\ E \mapsto F$ be continuous and linear, $h \in \mathcal{PCLSC}(F)$ and $h \circ A \in \mathcal{PC}(E)$. Let $\varepsilon > 0$ and $v \in E$. Then*

$$A^*(\partial h)^\varepsilon(Av) \subset \partial(h \circ A)^\varepsilon(v).$$

Proof. Let $v^* \in A^*(\partial h)^\varepsilon(Av)$. Then there exists $w^* \in (\partial h)^\varepsilon(Av)$ such that $v^* = A^*w^*$. Let (x, x^*) be an arbitrary element of $G(\partial(h\circ A))$. From (16), there exists a net (z_γ, z_γ^*) of elements of $G(\partial h)$ such that $\|z_\gamma - Ax\| \to 0$, $A^*z_\gamma^* \to x^*$ in $w(E^*, E)$ and $\langle z_\gamma, z_\gamma^*\rangle \to \langle x, x^*\rangle$. Then, for all γ,

$$\langle z_\gamma, z_\gamma^*\rangle - \langle v, A^*z_\gamma^*\rangle - \langle z_\gamma, w^*\rangle + \langle Av, w^*\rangle = \langle z_\gamma - Av, z_\gamma^* - w^*\rangle \geq -\varepsilon.$$

Passing to the limit,

$$\langle x, x^*\rangle - \langle v, x^*\rangle - \langle Ax, w^*\rangle + \langle Av, w^*\rangle \geq -\varepsilon,$$

hence

$$\langle x - v, x^* - A^*w^*\rangle \geq -\varepsilon,$$

that is to say,

$$\langle x - v, x^* - v^*\rangle \geq -\varepsilon.$$

Since this holds for all $(x, x^*) \in G(\partial(h \circ A))$, $v^* \in \partial(h \circ A)^\varepsilon(v)$, as required. □

Theorem 4.2. *Let E be a locally convex space, F be a Banach space, $A\colon\ E \mapsto F$ be continuous and linear, $h \in \mathcal{PCLSC}(F)$, $h \circ A \in \mathcal{PC}(E)$ and $x \in E$. Let $\mathcal{C}$ be any locally convex topology on E^* giving E as dual. Then*

$$\partial(h \circ A)(x) = \bigcap_{\varepsilon>0} \overline{A^*(\partial h)^\varepsilon(Ax)}.$$

(The overbar above stands for the closure in $(E^*, \mathcal{C})$.)

Proof. The inclusion "$\subset$" is immediate from Theorem 3.1 since $\partial_\varepsilon h(Ax) \subset (\partial h)^\varepsilon(Ax)$. On the other hand, for all $\varepsilon > 0$, we have from Lemma 4.1 that $A^*(\partial h)^\varepsilon(Ax) \subset \partial(h \circ A)^\varepsilon(x)$. Since $\partial(h \circ A)^\varepsilon(x)$ is $\mathcal{C}$–closed, it follows that

$$\overline{A^*(\partial h)^\varepsilon(Ax)} \subset \partial(h \circ A)^\varepsilon(x).$$

□

Consequently,

$$\bigcap_{\varepsilon>0} \overline{A^*(\partial h)^\varepsilon(Ax)} \subset \bigcap_{\varepsilon>0} \partial(h \circ A)^\varepsilon(x).$$

Since $\partial(h \circ A)$ is maximal monotone (see Remark 3.4),

$$\bigcap_{\varepsilon>0} \partial(h \circ A)^\varepsilon(x) = \partial(h \circ A)(x),$$

which completes the proof of the theorem. □

Remark 4.3. If E is a Banach space, the following analog of the Brøndsted–Rockafellar theorem holds for many (but not all) maximal monotone operators S: *Let* $\alpha,\ \beta > 0$, $0 < \varepsilon < \alpha\beta$ *and* $(v, v^*) \in G(S^\varepsilon) \setminus G(S)$. *Then there exists* $(x, x^*) \in G(S)$ *such that* $0 < \|x - v\| < \alpha$ *and* $0 < \|x^* - v^*\| < \beta$ (in fact, we can also make $\|x-v\|/\|x^*-v^*\|$ as near as we please to α/β and $\langle x - v, x^* - v^*\rangle/(\|x-v\|\|x^*-v^*\|)$ as near as we please to -1). (See [14, Theorem 8.6, p. 277-278].) The precise description of the class of maximal monotone multifunctions S for which this is valid is too complicated to enter into here, but it is certainly true if $S = \partial f$ for some $f \in \mathcal{PCLSC}(E)$ (see [14, Theorem 12.6, p. 287]). Thus $\partial_\varepsilon f$ can almost be replaced by $(\partial f)^\varepsilon$ in the Brøndsted–Rockafellar theorem. Another class of maximal monotone multifunction for which a similar "inequality–splitting" property is true is the class of (possibly discontinuous) maximal monotone linear operators of dense type (see [14, Theorem 11.2, p. 282]). It is *not* true for every continuous maximal monotone linear operator (see [14, Example 11.5, p. 283-284]).

5. Bootstrapping to sums

We show in this section how the results of the previous sections can be bootstrapped to incorporate the sums of functions. Since these results depend on the same substitution, we will combine them together into one composite theorem, Theorem 5.1. Theorem 5.1 being somewhat overburdened with symbols, we present in Corollary 5.2 a simplified version in which all the spaces are identical, and all the maps A_k are taken to be the identity map. Here are some comments on the individual results.

We have not seen Theorem 5.1(a) or Corollary 5.2(a) in the literature.

Theorem 5.1(b) or its consequence Corollary 5.2(b) imply the result proved by Hiriart–Urruty and Phelps in [5, Theorem 2.1, p. 160–161], namely that *if* E *is a locally convex space,* $f, g \in \mathcal{PCLSC}(E)$, $\operatorname{dom} f \cap \operatorname{dom} g \neq \emptyset$ *and* $x \in E$, *then*

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

Two different proofs of this were given in Hiriart-Urruty, Moussaoui, Seeger and Volle, [4, Theorem 3.1, p. 1732–1734].

Theorem 5.1(c) generalizes certain aspects of [18, Theorem 1]. (See Remark 5.3 for further comments on this.) The strongest of the conditions that is implied by (23) is (24), in which the final convergence relation has been "decoupled". This is the only one of the equivalent conditions that is specific to sums, and does not follow directly from the results on precompositions.

A new definition is necessary to understand the background for Theorem 5.1(d) and Corollary 5.2(d). If $S\colon\ E \mapsto 2^{E^*}$ and $T\colon\ E \mapsto 2^{E^*}$ are monotone then the *extended sum* $S \underset{ext}{+} T$ of S and T is defined by Revalski and Théra in [10] by

$$(S \underset{ext}{+} T)(x) := \bigcap_{\varepsilon>0} \overline{S^\varepsilon x + T^\varepsilon x}.$$

The motivation for this definition is that if S and T are maximal monotone then so is $S \underset{ext}{+} T$ under fairly weak assumptions. Theorem 5.1(d) or its consequence Corollary 5.2(d) imply the result proved by Revalski and Théra in [10], namely that *if E is a Banach space, $f, g \in \mathcal{PCLSC}(E)$ and* $\operatorname{dom} f \cap \operatorname{dom} g \neq \emptyset$, *then* $\partial(f+g) = \partial f \underset{ext}{+} \partial g$.

As observed by Hiriart–Urruty and Phelps in [5], summing two lower semicontinuous convex functions and precomposing a lower semicontinuous convex function by a continuous affine (or linear) mapping are equivalent operations from the convex analysis viewpoint. Theorem 5.1 shows that proving results on compositions *first* provides an easy way of establishing results for sums of $K(>2)$ functions. It is also much easier to handle analytically the renorming of the single space F that we performed in Theorem 3.3 than the renorming of the product space $F_1 \times \cdots \times F_K$ that we would have to perform in a direct proof of Theorem 5.1(c).

Theorem 5.1. *Let $E, F_1, \ldots, F_K$ be locally convex spaces, for all $k = 1, \ldots, K$, $A_k\colon\ E \mapsto F_k$ be continuous and linear, $h_k \in \mathcal{PCLSC}(F_k)$ and $\bigcap_{k=1}^K \operatorname{dom} h_k \circ A_k \neq \emptyset$. Let $\mathcal{C}$ be any locally convex topology on E^* giving E as dual. (The overbar in (a) stands for the closure in $(E^*, \mathcal{C}) \times \mathrm{I\!R}$. After that, it stands for the closure in $(E^*, \mathcal{C})$.) Then:*

(a)
$$\begin{aligned} \operatorname{epi}(h_1 \circ A_1 + \cdots + h_K \circ A_K)^* &= \overline{\operatorname{epi} h_1^*/A_1^* + \cdots + \operatorname{epi} h_K^*/A_K^*} \\ &= \overline{\operatorname{stepi} h_1^*/A_1^* + \cdots + \operatorname{stepi} h_K^*/A_K^*}. \end{aligned}$$

(b) For all $x \in E$,

$$\partial(h_1 \circ A_1 + \cdots + h_K \circ A_K)(x) = \bigcap_{\varepsilon>0} \overline{A_1^* \partial_\varepsilon h_1(A_1 x) + \cdots + A_K^* \partial_\varepsilon h_K(A_K x)}.$$

(c) Suppose, in addition, that $F_1, \ldots, F_K$ are Banach spaces and $(x, x^*) \in E \times E^*$. Then the conditions (23)–(25) are equivalent:

$$x^* \in \partial(h_1 \circ A_1 + \cdots + h_K \circ A_K)(x). \tag{23}$$

$$\left.\begin{array}{c} \textit{For all } k,\ \textit{there exists a net } (z_{k,\gamma}, z_{k,\gamma}^*) \textit{ of elements of } G(\partial h_k) \textit{ such that} \\ \|z_{k,\gamma} - A_k x\| \to 0,\ A_1^* z_{1,\gamma}^* + \cdots + A_K^* z_{K,\gamma}^* \to x^* \textit{ in } \mathcal{C} \\ \textit{and, for all } k,\ \langle z_{k,\gamma} - A_k x, z_{k,\gamma}^* \rangle \to 0. \end{array}\right\} \tag{24}$$

$$\left.\begin{array}{c} \textit{For all } k,\ \textit{there exists a net } (z_{k,\gamma}, z_{k,\gamma}^*) \textit{ of elements of } G(\partial h_k) \textit{ such that} \\ \|z_{k,\gamma} - A_k x\| \to 0,\ A_1^* z_{1,\gamma}^* + \cdots + A_K^* z_{K,\gamma}^* \to x^* \textit{ in } \mathcal{C} \\ \textit{and } \langle z_{1,\gamma}, z_{1,\gamma}^* \rangle + \cdots + \langle z_{K,\gamma}, z_{K,\gamma}^* \rangle \to \langle x, x^* \rangle. \end{array}\right\} \tag{25}$$

(d) Suppose still that $F_1, \ldots, F_K$ are Banach spaces, and $x \in E$. Then

$$\partial(h_1 \circ A_1 + \cdots + h_K \circ A_K)(x) = \bigcap_{\varepsilon>0} \overline{A_1^* (\partial h_1)^\varepsilon (A_1 x) + \cdots + A_K^* (\partial h_K)^\varepsilon (A_K x)}.$$

Proof. Let $F := F_1 \times \cdots \times F_K$, for all $x \in E$, $A(x) := (A_1 x, \ldots, A_K x) \in F$ and,

$$\text{for all } (y_1, \ldots, y_K) \in F, \quad h(y_1, \ldots, y_K) := h_1(y_1) + \cdots + h_K(y_K) \in \mathbb{R} \cup \{\infty\}.$$

Then,

$$(y_1^*, \ldots, y_K^*) \in F_1^* \times \cdots \times F_K^* \quad \Longrightarrow \quad A^*(y_1^*, \ldots, y_K^*) = A_1^* y_1^* + \cdots + A_K^* y_K^*.$$

(a) We have

$$\begin{aligned}
(x^*, \lambda) \;&\in \operatorname{stepi} h^*/A^* \\
\iff\quad & \text{there exists } y^* \in F^* \text{ such that } A^*(y^*) = x^* \text{ and } h^*(y^*) < \lambda \\
\iff\quad & \text{there exists } (y_1^*, \ldots, y_K^*) \in F_1^* \times \cdots \times F_K^* \text{ such that} \\
& \qquad A_1^* y_1^* + \cdots + A_K y_K^* = x^* \text{ and } h_1^*(y_1^*) + \cdots + h_K^*(y_K^*) < \lambda \\
\iff\quad & \text{there exist } (y_1^*, \ldots, y_K^*) \in F_1^* \times \cdots \times F_K^*,\ (x_1^*, \ldots, x_K^*) \in (E^*)^K \text{ and} \\
& \qquad (\lambda_1, \ldots, \lambda_K) \in \mathbb{R}^K \text{ such that, for all } k,\ A_k^* y_k^* = x_k^* \text{ and } h_k^*(y_k^*) < \lambda_k, \\
& \qquad\qquad x_1^* + \cdots + x_K^* = x^* \text{ and } \lambda_1 + \cdots + \lambda_K = \lambda \\
\iff\quad & (x^*, \lambda) \in \operatorname{stepi} h_1^*/A_1^* + \cdots + \operatorname{stepi} h_K^*/A_K^*.
\end{aligned}$$

Assertion (a) now follows easily from Theorem 2.7.

(b) This follows from Theorem 3.1 since, as the reader can easily verify,

$$\partial_\varepsilon h(Ax) \subset \partial_\varepsilon h_1(A_1 x) \times \cdots \times \partial_\varepsilon h_K(A_K x) \subset \partial_{K\varepsilon} h(Ax).$$

(c) It is immediate from Theorem 3.3 that (23) and (25) are equivalent to each other and to:

$$\left.\begin{aligned}
&\text{For all } k,\ \text{there exists a net } (z_{k,\gamma}, z_{k,\gamma}^*) \text{ of elements of } G(\partial h_k) \text{ such that} \\
&\qquad \|z_{k,\gamma} - A_k x\| \to 0,\ A_1^* z_{1,\gamma}^* + \cdots + A_K^* z_{K,\gamma}^* \to x^* \text{ in } \mathcal{C} \\
&\qquad \text{and } \langle z_{1,\gamma} - A_1 x, z_{1,\gamma}^* \rangle + \cdots + \langle z_{K,\gamma} - A_K x, z_{K,\gamma}^* \rangle \to 0.
\end{aligned}\right\} \tag{26}$$

Since (24)$\Longrightarrow$(26), it only remains to prove that (26)$\Longrightarrow$(24). Now if (26) is true then, from (23), $x \in \bigcap_{k=1}^K \operatorname{dom} h_k \circ A_k$. For all k, since $(z_{k,\gamma}, z_{k,\gamma}^*) \in G(\partial h_k)$,

$$\langle z_{k,\gamma} - A_k x, z_{k,\gamma}^* \rangle \geq h_k(z_{k,\gamma}) - h_k(A_k x).$$

Using the lower semicontinuity of h_k and the fact that $\|z_{k,\gamma} - A_k x\| \to 0$, we derive that $\liminf_\gamma \langle z_{k,\gamma} - A_k x, z_{k,\gamma}^* \rangle \geq 0$. If we now combine this with the assumption in (26) that $\langle z_{1,\gamma} - A_1 x, z_{1,\gamma}^* \rangle + \cdots + \langle z_{K,\gamma} - A_K x, z_{K,\gamma}^* \rangle \to 0$, we obtain easily that, for all k, $\langle z_{k,\gamma} - A_k x, z_{k,\gamma}^* \rangle \to 0$. Thus we have established (24).

(d) This follows from Theorem 4.2 and the fact that

$$(\partial h)^\varepsilon(Ax) \subset (\partial h_1)^\varepsilon(A_1 x) \times \cdots \times (\partial h_K)^\varepsilon(A_K x) \subset (\partial h)^{K\varepsilon}(Ax).$$

The second inclusion is immediate. We outline the proof of the first, which is a bit subtler. If $(v_1^*, \ldots, v_K^*) \in (\partial h)^\varepsilon(Ax)$ then, for all $(z_1, z_1^*) \in G(\partial h_1), \ldots, (z_K, z_K^*) \in G(\partial h_K)$,

$$\langle z_1 - A_1 x, z_1^* - v_1^* \rangle + \cdots + \langle z_K - A_K x, z_K^* - v_K^* \rangle \geq -\varepsilon.$$

Consequently,

$$\inf_{(z_1,z_1^*)\in G(\partial h_1)}\langle z_1 - A_1x, z_1^* - v_1^*\rangle + \cdots + \inf_{(z_K,z_K^*)\in G(\partial h_K)}\langle z_K - A_Kx, z_K^* - v_K^*\rangle \geq -\varepsilon. \tag{27}$$

If $(A_kx, v_k^*) \in G(\partial h_k)$ then we can take $(z_k, z_k^*) = (A_kx, v_k^*)$, while if $(A_kx, v_k^*) \notin G(\partial h_k)$ then, by the maximal monotonicity of ∂h_k, there exists $(z_k, z_k^*) \in G(\partial h_k)$ such that $\langle z_k - A_kx, z_k^* - v_k^*\rangle < 0$ — so, in either case, $\inf_{(z_k,z_k^*)\in G(\partial h_k)}\langle z_k - A_kx, z_k^* - v_k^*\rangle \leq 0$. Combining this with (27), we obtain that,

$$\text{for all } k, \quad \inf_{(z_k,z_k^*)\in G(\partial h_k)}\langle z_k - A_kx, z_k^* - v_k^*\rangle \geq -\varepsilon,$$

that is to say $v_k^* \in (\partial h_k)^\varepsilon(A_kx)$. Thus $(v_1^*, \ldots, v_K^*) \in (\partial h_1)^\varepsilon(A_1x) \times \cdots \times (\partial h_K)^\varepsilon(A_Kx)$, as required. (The argument used above is an application of the "negative infimum" property of maximal monotone multifunctions. See [13, Lemma 8.1(c), p. 30].) □

Corollary 5.2. *Let E be a locally convex space, for all $k = 1, \ldots, K$, $h_k \in \mathcal{PCLSC}(E)$ and $\bigcap_{k=1}^K \operatorname{dom} h_k \neq \emptyset$. Let $\mathcal{C}$ be any locally convex topology on E^* giving E as dual. (The overbar in (a) stands for the closure in $(E^*, \mathcal{C}) \times \mathbb{R}$. After that, it stands for the closure in $(E^*, \mathcal{C})$.) Then:*

(a)

$$\begin{aligned}\operatorname{epi}(h_1 + \cdots + h_K)^* &= \overline{\operatorname{epi} h_1^* + \cdots + \operatorname{epi} h_K^*}\\ &= \overline{\operatorname{stepi} h_1^* + \cdots + \operatorname{stepi} h_K^*}.\end{aligned}$$

(b) For all $x \in E$,

$$\partial(h_1 + \cdots + h_K)(x) = \bigcap_{\varepsilon>0}\overline{\partial_\varepsilon h_1(x) + \cdots + \partial_\varepsilon h_K(x)}.$$

(c) Suppose, in addition, that E is a Banach space and $(x, x^*) \in E \times E^*$. Then the conditions (28)–(30) are equivalent:

$$x^* \in \partial(h_1 + \cdots + h_K)(x). \tag{28}$$

$$\left.\begin{array}{c}\textit{For all } k, \textit{ there exists a net } (z_{k,\gamma}, z_{k,\gamma}^*) \textit{ of elements of } G(\partial h_k) \textit{ such that}\\ \|z_{k,\gamma} - x\| \to 0,\ z_{1,\gamma}^* + \cdots + z_{K,\gamma}^* \to x^* \textit{ in } \mathcal{C}\\ \textit{and, for all } k,\ \langle z_{k,\gamma} - x, z_{k,\gamma}^*\rangle \to 0.\end{array}\right\} \tag{29}$$

$$\left.\begin{array}{c}\textit{For all } k, \textit{ there exists a net } (z_{k,\gamma}, z_{k,\gamma}^*) \textit{ of elements of } G(\partial h_k) \textit{ such that}\\ \|z_{k,\gamma} - x\| \to 0,\ z_{1,\gamma}^* + \cdots + z_{K,\gamma}^* \to x^* \textit{ in } \mathcal{C}\\ \textit{and } \langle z_{1,\gamma}, z_{1,\gamma}^*\rangle + \cdots + \langle z_{K,\gamma}, z_{K,\gamma}^*\rangle \to \langle x, x^*\rangle.\end{array}\right\} \tag{30}$$

(d) Suppose still that E is a Banach space, and $x \in E$. Then

$$\partial(h_1 + \cdots + h_K)(x) = \bigcap_{\varepsilon>0}\overline{(\partial h_1)^\varepsilon(x) + \cdots + (\partial h_K)^\varepsilon(x)}.$$

Remark 5.3. The reader might have noticed that we have not translated the (optional) condition "$\mathrm{dist}(z_\gamma, A(E))\|z^*_\gamma\| \to 0$" of Theorem 3.3 into the situation of Theorem 5.1(c). The reason for this is that it turns out to be extremely cumbersome in general. Consider now the special case where $E = F_1$ and that A_1 is the identity map. If $z = (z_1, \ldots, z_n) \in F$ then, for all $x \in E$ and $k = 2, \ldots, K$,

$$\begin{aligned} \|z_k - A_k z_1\| &\le \|z_k - A_k x\| + \|A_k\|\|z_1 - x\| \le (1 + \|A_k\|)(\|z_k - A_k x\| \vee \|z_1 - x\|) \\ &\le (1 + \|A_k\|)\|z - Ax\|_\infty. \end{aligned}$$

Consequently, if the distance in F is measured in $\|\ \|_\infty$,

$$\|z_k - A_k z_1\| \le (1 + \|A_k\|)\mathrm{dist}(z, A(E)).$$

Thus we can then add the following optional condition into (24) or (25):

$$\text{for all } k = 2, \ldots, K, \quad \|z_{k,\gamma} - A_k z_{1,\gamma}\|\|(z^*_{1,\gamma}, \ldots, z^*_{K,\gamma})\|_1 \to 0.$$

This corresponds to the condition that appears in [18, Theorem 1(i)]. Turning now to the special situation of Corollary 5.2(c), it follows that we can always add the following optional condition into (29) or (30):

$$\text{for all } j, k \in \{1, \ldots, K\}, \quad \|z_{k,\gamma} - z_{j,\gamma}\|\|(z^*_{1,\gamma}, \ldots, z^*_{K,\gamma})\|_1 \to 0.$$

Remark 5.4. As we observed in the comments preceding Corollary 2.3, if E is a reflexive Banach space then we can take $\mathcal{C}$ to be the norm topology of E^* in Corollary 2.3, Theorem 2.7, Theorem 3.1, Theorem 3.3, Theorem 4.2, Theorem 5.1 and Corollary 5.2. Since this topology is metrizable, it follows easily that the nets in Theorem 3.3, Theorem 5.1(c) and Corollary 5.2(c) can be taken to be sequences.

6. Subdifferentials of marginal functions

We now return to the discussion of marginal functions that we considered briefly in Lemmas 2.4 and 2.5. The following result was proved by Hiriart–Urruty and Phelps in [5, Theorem 4.1, p. 164–165]. It can be thought of as an analog of Theorem 3.1 for marginal functions. It is worth pointing out that the argument of Example 2.6 shows that the function f/A that appears in Theorems 6.1 and 6.2 is not necessarily lower semicontinuous.

Theorem 6.1. *Let E and F be locally convex spaces, $A\colon\ E \mapsto F$ be continuous and linear, $f \in \mathcal{PC}(E)$, $f^* \circ A^* \in \mathcal{PC}(F)$ and $y \in F$. Then*

$$\partial(f/A)(y) = \bigcap_{\varepsilon > 0} (A^*)^{-1}\partial_\varepsilon f(A^{-1}y).$$

Proof. It follows from Lemma 2.4 and the definition of f/A that

$$\begin{aligned} y^* \in \partial(f/A)(y) &\iff f/A(y) + (f/A)^*(y^*) \le \langle y, y^* \rangle \\ &\iff \text{for all } \varepsilon > 0,\ f/A(y) + f^*(A^* y^*) < \langle y, y^* \rangle + \varepsilon \\ &\iff \text{for all } \varepsilon > 0,\ \text{there exists } x \in A^{-1}y \text{ such that} \\ &\qquad\qquad f(x) + f^*(A^* y^*) < \langle y, y^* \rangle + \varepsilon \\ &\iff \text{for all } \varepsilon > 0,\ \text{there exists } x \in A^{-1}y \text{ such that} \\ &\qquad\qquad f(x) + f^*(A^* y^*) \le \langle x, A^* y^* \rangle + \varepsilon \\ &\iff \text{for all } \varepsilon > 0,\ \text{there exists } x \in A^{-1}y \text{ such that } A^* y^* \in \partial_\varepsilon f(x). \end{aligned}$$

This is the required result. □

We now turn our attention to the situation when E is a Banach space. Our next result is an analog of Theorem 3.3 for marginal functions.

Theorem 6.2. *Let E be a Banach space, F be a locally convex space, $A\colon\ E \mapsto F$ be continuous and linear, $f \in \mathcal{PCLSC}(E)$, $f^* \circ A^* \in \mathcal{PC}(F^*)$ and $(y, y^*) \in F \times F^*$. Then the conditions (31)–(34) are equivalent:*

$$y^* \in \partial(f/A)(y). \tag{31}$$

$$\left.\begin{array}{c}\text{There exist nets } (z_\gamma, z_\gamma^*) \text{ of elements of } G(\partial f) \text{ and } x_\gamma \text{ of elements of}\\ \operatorname{dom} f \cap A^{-1}y \text{ such that } \|z_\gamma^* - A^*y^*\| \to 0,\ \langle z_\gamma, z_\gamma^* - A^*y^*\rangle \to 0,\\ f(x_\gamma) - f(z_\gamma) \to 0 \text{ and } Az_\gamma \to y \text{ in } F.\end{array}\right\} \tag{32}$$

$$\left.\begin{array}{c}\text{There exist nets } (z_\gamma, z_\gamma^*) \text{ of elements of } G(\partial f) \text{ and } x_\gamma \text{ of elements of}\\ \operatorname{dom} f \cap A^{-1}y \text{ such that } \|z_\gamma^* - A^*y^*\| \to 0,\ \langle z_\gamma, z_\gamma^*\rangle \to \langle y, y^*\rangle,\\ f(x_\gamma) - f(z_\gamma) \to 0 \text{ and } Az_\gamma \to y \text{ in } F.\end{array}\right\} \tag{33}$$

$$\left.\begin{array}{c}\text{There exists a net } (z_\gamma, z_\gamma^*) \text{ of elements of } G(\partial f)\\ \text{such that } z_\gamma^* \to A^*y^* \text{ in } w(E^*, E),\\ \langle z_\gamma, z_\gamma^*\rangle \to \langle y, y^*\rangle \text{ and } f/A(y) \le \liminf_\gamma f(z_\gamma).\end{array}\right\} \tag{34}$$

Proof. ((32)$\Longrightarrow$(33)) This follows from the equality

$$\langle z_\gamma, z_\gamma^*\rangle = \langle z_\gamma, z_\gamma^* - A^*y^*\rangle + \langle Az_\gamma, y^*\rangle.$$

((33)$\Longrightarrow$(34)) This follows from (8) since, for all γ, $f/A(y) = f/A(Ax_\gamma) \le f(x_\gamma)$, and so

$$f/A(y) \le \liminf_\gamma f(x_\gamma) = \liminf_\gamma \big[(f(x_\gamma) - f(z_\gamma)) + f(z_\gamma)\big] = \liminf_\gamma f(z_\gamma).$$

((34)$\Longrightarrow$(31)) Let (z_γ, z_γ^*) be as in the statement of (34). It follows from Lemma 2.4 and the $w(E^*, E)$–lower semicontinuity of f^* that

$$\begin{aligned} f/A(y) + (f/A)^*(y^*) &\le \liminf_\gamma f(z_\gamma) + f^*(A^*y^*) \le \liminf_\gamma f(z_\gamma) + \liminf_\gamma f^*(z_\gamma^*)\\ &\le \liminf_\gamma [f(z_\gamma) + f^*(z_\gamma^*)] \le \liminf_\gamma \langle z_\gamma, z_\gamma^*\rangle = \langle y, y^*\rangle. \end{aligned}$$

Thus $y^* \in \partial(f/A)(y)$, and (31) is satisfied.

((31)$\Longrightarrow$(32)) Let $\varepsilon > 0$ and q be a continuous seminorm on F. We will first find $(z, z^*) \in G(\partial f)$ and $x \in \operatorname{dom} f \cap A^{-1}y$ such that

$$\|z^* - A^*y^*\| < \varepsilon, \quad |\langle z, z^* - A^*y^*\rangle| < \varepsilon, \tag{35}$$

$$q(Az - y) < 1, \tag{36}$$

and

$$|f(x) - f(z)| < \varepsilon. \tag{37}$$

Since $q \circ A$ is a continuous seminorm on E, $\sup_{w\in E,\,\|w\|\le 1} q\circ A(w) < \infty$. Write M for this number, and choose $\alpha, \beta > 0$ so that

$$\alpha\beta + \beta < \varepsilon, \quad \alpha(M \vee q(y)) < 1 \quad \text{and} \quad \alpha\beta + \alpha(\|A^*y^*\| \vee |\langle y, y^*\rangle|) < \varepsilon.$$

From Theorem 6.1, there exists $x \in A^{-1}y$ such that $A^*y^* \in \partial_{\alpha\beta} f(x)$. Let B be the unit ball of E, and define an equivalent norm $|||\ |||$ on E (which depends on ε and q) by taking the set $\operatorname{co}\{B \cup \{x\} \cup \{-x\}\}$ as unit ball, and write $|||\ |||$ for the corresponding dual norm on E^* also. Then, for all $w^* \in E^*$, $|||w^*||| = \|w^*\| \vee |\langle x, w^*\rangle|$. Since $Ax = y$, it follows that,

$$\text{for all } v^* \in F^*, \quad |||A^*v^*||| = \|A^*v^*\| \vee |\langle y, v^*\rangle|. \tag{38}$$

From the Brøndsted–Rockafellar theorem, Theorem 3.2, there exists $(z, z^*) \in G(\partial f)$ such that

$$|||z - x||| \le \alpha \quad \text{and} \quad |||z^* - A^*y^*||| \le \beta.$$

(35) follows since $\|z^* - A^*y^*\| \le |||z^* - A^*y^*||| \le \beta < \varepsilon$ and

$$\begin{aligned} |\langle z, z^* - A^*y^*\rangle| &\le |\langle z - x, z^* - A^*y^*\rangle| + |\langle x, z^* - A^*y^*\rangle| \\ &\le |||z - x|||\,|||z^* - A^*y^*||| + |||z^* - A^*y^*||| \\ &\le \alpha\beta + \beta < \varepsilon. \end{aligned}$$

The proof of (36) is a little more complicated. From the one–dimensional Hahn–Banach theorem, there exists $v^* \in F^*$ such that $v^* \le q$ on F and $\langle Az - y, v^*\rangle = q(Az - y)$. Then

$$\|A^*v^*\| = \sup_{w\in E,\,\|w\|\le 1} \langle w, A^*v^*\rangle = \sup_{w\in E,\,\|w\|\le 1} \langle Aw, v^*\rangle \le \sup_{w\in E,\,\|w\|\le 1} q(Aw) = M.$$

Thus, from (38),

$$|||A^*v^*||| \le M \vee |\langle y, v^*\rangle| \le M \vee q(y).$$

Consequently, since $y = Ax$,

$$q(Az - y) = \langle Az - y, v^*\rangle = \langle z - x, A^*v^*\rangle \le |||z - x|||\,|||A^*v^*||| \le \alpha(M \vee q(y)) < 1,$$

which gives (36). Finally, we prove (37). Since $A^*y^* \in \partial_{\alpha\beta} f(x)$ and $z^* \in \partial f(z)$,

$$-\langle z - x, z^* - A^*y^*\rangle - \langle z - x, A^*y^*\rangle = \langle x - z, z^*\rangle \le f(x) - f(z) \le \langle x - z, A^*y^*\rangle + \alpha\beta,$$

from which

$$-|||z - x|||\,|||z^* - A^*y^*||| - |||z - x|||\,|||A^*y^*||| \le f(x) - f(z) \le \alpha\beta + |||z - x|||\,|||A^*y^*|||,$$

and so, using (38),

$$|f(x) - f(z)| \le \alpha\beta + \alpha|||A^*y^*||| = \alpha\beta + \alpha(\|A^*y^*\| \vee |\langle y, y^*\rangle|) < \varepsilon,$$

which gives (37). Using (35)–(37), we can easily find nets (z_γ, z_γ^*) of elements of $G(\partial f)$ and x_γ of $\operatorname{dom} f \cap A^{-1}y$ satisfying the conditions of (32). □

Conjecture 6.3. By analogy with Theorem 4.2, it is reasonable to conjecture that *if E is a Banach space, F is a locally convex space, $A\colon\ E \mapsto F$ is continuous and linear, $f \in \mathcal{PCLSC}(E)$, $f^* \circ A^* \in \mathcal{PC}(F)$ and $y \in F$ then*

$$\partial(f/A)(y) = \bigcap_{\varepsilon>0} (A^*)^{-1}(\partial f)^{\varepsilon}(A^{-1}y).$$

7. Bootstrapping to episums

Let E be a locally convex space and, for all $k = 1, \ldots, K$, $f_k \in \mathcal{PCLSC}(E)$. If $y \in E$, write

$$(f_1 \underset{e}{+} \cdots \underset{e}{+} f_K)(y) := \inf\{f_1(x_1) + \cdots + f_K(x_K)\colon\ x_1 + \cdots + x_K = y\}.$$

The function $f_1 \underset{e}{+} \cdots \underset{e}{+} f_K$ is the *episum* or *inf–convolution* of $f_1, \ldots, f_K$. We show in this section how the results of the previous sections can be bootstrapped to incorporate the episums of convex functions, and more general situations where the episum is combined with marginalization.

Since these results depend on the same substitution, we will combine them together into one composite theorem, Theorem 7.1. Theorem 7.1 being somewhat overburdened with symbols, we present in Corollary 7.2 a simplified version in which all the spaces are identical, and all the maps A_k are taken to be the identity map. Here are some comments on the individual results.

Corollary 7.2(a) is certainly well known.

Theorem 7.1(b) or its consequence Corollary 7.2(b) imply the result proved by Hiriart–Urruty and Phelps in [5, Theorem 1.1, p. 156–157], namely that *if E is a locally convex space, $f, g \in \mathcal{PC}(E)$, f and g are bounded below by the same continuous affine function and $x \in E$, then*

$$\partial(f \underset{e}{+} g)(x) = \bigcap_{\varepsilon>0} \partial_\varepsilon f(x) \cap \partial_\varepsilon g(x).$$

We have not seen Theorem 7.1(c) or Corollary 7.2(c) in the literature. The strongest of the conditions that is implied by (39) is (40), in which one of the convergence relations has been "decoupled". This is the only one of the equivalent conditions that is specific to episums, and does not follow from the results on marginal functions.

As Theorem 7.1 shows, proving results on marginal functions *first* provides an easy way of establishing results for episums of $K(> 2)$ functions. It is also much easier to handle analytically the renorming of the single space E that we performed in Theorem 6.2 than the renorming of the product space $E_1 \times \cdots \times E_K$ that we would have to perform in a direct proof of Theorem 7.1(c).

Theorem 7.1. *Let $E_1, \ldots, E_K$ and F be locally convex spaces, for all $k = 1, \ldots, K$, $A_k\colon\ E_k \mapsto F$ be continuous and linear, $f_k \in \mathcal{PC}(E_k)$ and*

$$f_1^* \circ A_1^* + \cdots + f_K^* \circ A_K^* \in \mathcal{PC}(F^*).$$

If $y \in F$, write

$$g(y) := \inf\{f_1(x_1) + \cdots + f_K(x_K)\colon\ A_1x_1 + \cdots + A_Kx_K = y\}.$$

Then:

(a) $g \in \mathcal{PCPC}(F)$ and $g^* = f_1^* \circ A_1^* + \cdots + f_K^* \circ A_K^*$.

(b) For all $y \in F$,

$$\partial g(y) = \bigcap_{\varepsilon>0} \bigcup_{A_1x_1+\cdots+A_Kx_K=y} (A_1^*)^{-1}\partial_\varepsilon f_1(x_1) \cap \cdots \cap (A_K^*)^{-1}\partial_\varepsilon f_K(x_K).$$

(c) Suppose, in addition, that $E_1, \ldots, E_K$ are Banach spaces, for all $k = 1, \ldots, K$ we have $f_k \in \mathcal{PCLSC}(E_k)$, and $(y, y^*) \in F \times F^*$. Then the conditions (39)—(42) are equivalent:

$$y^* \in \partial g(y). \tag{39}$$

$$\left.\begin{array}{c} \text{For all } k, \text{ there exist nets } (z_{k,\gamma}, z_{k,\gamma}^*) \text{ of elements of } G(\partial f_k) \text{ and} \\ x_{k,\gamma} \text{ of elements of } \operatorname{dom} f_k \text{ such that } \|z_{k,\gamma}^* - A_k^* y^*\| \to 0, \\ \langle z_{k,\gamma}, z_{k,\gamma}^* - A_1^* y^* \rangle \to 0, \\ f_1(x_{1,\gamma}) - f_1(z_{1,\gamma}) + \cdots + f_K(x_{K,\gamma}) - f_K(z_{K,\gamma}) \to 0, \\ A_1 x_{1,\gamma} + \cdots + A_K x_{K,\gamma} = y \text{ and} \\ A_1 z_{1,\gamma} + \cdots + A_K z_{K,\gamma} \to y \text{ in } F. \end{array}\right\} \tag{40}$$

$$\left.\begin{array}{c} \text{For all } k, \text{ there exist nets } (z_{k,\gamma}, z_{k,\gamma}^*) \text{ of elements of } G(\partial f_k) \text{ and} \\ x_{k,\gamma} \text{ of elements of } \operatorname{dom} f_k \text{ such that } \|z_{k,\gamma}^* - A_k^* y^*\| \to 0, \\ \langle z_{1,\gamma}, z_{1,\gamma}^* \rangle + \cdots + \langle z_{K,\gamma}, z_{K,\gamma}^* \rangle \to \langle y, y^* \rangle, \\ f_1(x_{1,\gamma}) - f_1(z_{1,\gamma}) + \cdots + f_K(x_{K,\gamma}) - f_K(z_{K,\gamma}) \to 0, \\ A_1 x_{1,\gamma} + \cdots + A_K x_{K,\gamma} = y \text{ and} \\ A_1 z_{1,\gamma} + \cdots + A_K z_{K,\gamma} \to y \text{ in } F. \end{array}\right\} \tag{41}$$

$$\left.\begin{array}{c} \text{For all } k, \text{ there exists a net } (z_{k,\gamma}, z_{k,\gamma}^*) \text{ of elements of } G(\partial f_k) \\ \text{such that } z_{k,\gamma}^* \to A_k^* y^* \text{ in } w(E_k^*, E_k), \\ \langle z_{1,\gamma}, z_{1,\gamma}^* \rangle + \cdots + \langle z_{K,\gamma}, z_{K,\gamma}^* \rangle \to \langle y, y^* \rangle \text{ and} \\ g(y) \le \liminf_\gamma (f_1(z_{1,\gamma}) + \cdots + f_K(z_{K,\gamma})). \end{array}\right\} \tag{42}$$

Proof. Let $E := E_1 \times \cdots \times E_K$, for all $(x_1, \ldots, x_K) \in E$,

$$A(x_1, \ldots, x_K) := A_1 x_1 + \cdots + A_K x_K \in F$$

and

$$f(x_1, \ldots, x_K) := f_1(x_1) + \cdots + f_K(x_K) \in \mathbb{R} \cup \{\infty\}.$$

Then

$$y^* \in F^* \quad \Longrightarrow \quad A^* y^* = (A_1^* y^*, \ldots, A_K^* y^*) \in E_1^* \times \cdots \times E_K^*$$

and

$$(x_1^*, \ldots, x_K^*) \in E_1^* \times \cdots \times E_K^* \quad \Longrightarrow \quad f^*(x_1^*, \ldots, x_K^*) = f_1^*(x_1^*) + \cdots + f_K^*(x_K^*).$$

(a) This is immediate from Lemma 2.4, since $g = f/A$.

(b) This follows from Theorem 6.1 since, for all $(x_1, \ldots, x_K) \in E$,

$$\partial_\varepsilon f(x_1, \ldots, x_K) \subset \partial_\varepsilon f_1(x_1) \times \cdots \times \partial_\varepsilon f_K(x_K) \subset \partial_{K\varepsilon} f(x_1, \ldots, x_K).$$

(c) It is immediate from Theorem 6.2 that (39), (41) and (42) are equivalent to:

$$\left.\begin{array}{c} \text{For all } k, \text{ there exist nets } (z_{k,\gamma}, z_{k,\gamma}^*) \text{ of elements of } G(\partial f_k) \text{ and} \\ x_{k,\gamma} \text{ of elements of } \operatorname{dom} f_k \text{ such that } \|z_{k,\gamma}^* - A_k^* y^*\| \to 0, \\ \langle z_{1,\gamma}, z_{1,\gamma}^* - A_1^* y^* \rangle + \cdots + \langle z_{K,\gamma}, z_{K,\gamma}^* - A_K^* y^* \rangle \to 0, \\ f_1(x_{1,\gamma}) - f_1(z_{1,\gamma}) + \cdots + f_K(x_{K,\gamma}) - f_K(z_{K,\gamma}) \to 0, \\ A_1 x_{1,\gamma} + \cdots + A_K x_{K,\gamma} = y \text{ and} \\ A_1 z_{1,\gamma} + \cdots + A_K z_{K,\gamma} \to y \text{ in } F. \end{array}\right\} \tag{43}$$

Since (40)$\Longrightarrow$(43), it only remains to prove that (43)$\Longrightarrow$(40). Now if (43) is true then, from (39), $y^* \in \operatorname{dom} g^* = \bigcap_{k=1}^K \operatorname{dom} f_k^* \circ A_k^*$. For all k and γ, since $(z_{k,\gamma}, z_{k,\gamma}^*) \in G(\partial f_k)$,

$$\langle z_{k,\gamma}, z_{k,\gamma}^* - A_k^* y^* \rangle \geq f_k^*(z_{k,\gamma}^*) - f_k^*(A_k^* y^*).$$

Using the lower semicontinuity of f_k^* and the fact that $\|z_{k,\gamma}^* - A_k^* y^*\| \to 0$, we derive that $\liminf_\gamma \langle z_{k,\gamma}, z_{k,\gamma}^* - A_k^* y^* \rangle \geq 0$. If we now combine this with the assumption in (43) that $\langle z_{1,\gamma}, z_{1,\gamma}^* - A_1^* y^* \rangle + \cdots + \langle z_{K,\gamma}, z_{K,\gamma}^* - A_K^* y^* \rangle \to 0$, we obtain easily that, for all k, $\langle z_{k,\gamma}, z_{k,\gamma}^* - A_k^* y^* \rangle \to 0$. Thus we have established (40). □

Corollary 7.2. *Let E be a locally convex space, for all $k = 1, \ldots, K$, $f_k \in \mathcal{PC}(E)$ and*

$$f_1^* + \cdots + f_K^* \in \mathcal{PC}(E^*).$$

Then:

(a) $f_1 \underset{e}{+} \cdots \underset{e}{+} f_K \in \mathcal{PCPC}(E)$ and $(f_1 \underset{e}{+} \cdots \underset{e}{+} f_K)^* = f_1^* + \cdots + f_K^*$.

(b) For all $y \in E$,

$$\partial(f_1 \underset{e}{+} \cdots \underset{e}{+} f_K)(y) = \bigcap_{\varepsilon>0} \bigcup_{x_1+\cdots+x_K=y} \partial_\varepsilon f_1(x_1) \cap \cdots \cap \partial_\varepsilon f_K(x_K).$$

(c) Suppose, in addition, that E is a Banach space, for all $k = 1, \ldots, K$, $f_k \in \mathcal{PCLSC}(E)$ and $(y, y^*) \in E \times E^*$. Then the conditions (44)—(47) are equivalent:

$$y^* \in \partial(f_1 \underset{e}{+} \cdots \underset{e}{+} f_K)(y). \tag{44}$$

$$\left.\begin{array}{c} \text{For all } k, \text{ there exist nets } (z_{k,\gamma}, z_{k,\gamma}^*) \text{ of elements of } G(\partial f_k) \text{ and} \\ x_{k,\gamma} \text{ of elements of } \operatorname{dom} f_k \text{ such that } \|z_{k,\gamma}^* - y^*\| \to 0, \\ \langle z_{k,\gamma}, z_{k,\gamma}^* - y^* \rangle \to 0, \\ f_1(x_{1,\gamma}) - f_1(z_{1,\gamma}) + \cdots + f_K(x_{K,\gamma}) - f_K(z_{K,\gamma}) \to 0, \\ x_{1,\gamma} + \cdots + x_{K,\gamma} = y \text{ and} \\ z_{1,\gamma} + \cdots + z_{K,\gamma} \to y \text{ in } F. \end{array}\right\} \tag{45}$$

$$\left.\begin{array}{c} \text{For all } k, \text{ there exist nets } (z_{k,\gamma}, z_{k,\gamma}^*) \text{ of elements of } G(\partial f_k) \text{ and} \\ x_{k,\gamma} \text{ of elements of } \operatorname{dom} f_k \text{ such that } \|z_{k,\gamma}^* - y^*\| \to 0, \\ \langle z_{1,\gamma}, z_{1,\gamma}^* \rangle + \cdots + \langle z_{K,\gamma}, z_{K,\gamma}^* \rangle \to \langle y, y^* \rangle, \\ f_1(x_{1,\gamma}) - f_1(z_{1,\gamma}) + \cdots + f_K(x_{K,\gamma}) - f_K(z_{K,\gamma}) \to 0, \\ x_{1,\gamma} + \cdots + x_{K,\gamma} = y \text{ and} \\ z_{1,\gamma} + \cdots + z_{K,\gamma} \to y \text{ in } E. \end{array}\right\} \tag{46}$$

$$\left.\begin{array}{c} \text{For all } k, \text{ there exist a net } (z_{k,\gamma}, z_{k,\gamma}^*) \text{ of elements of } G(\partial f_k) \\ \text{such that } z_{k,\gamma}^* \to y^* \text{ in } w(E^*, E), \\ \langle z_{1,\gamma}, z_{1,\gamma}^* \rangle + \cdots + \langle z_{K,\gamma}, z_{K,\gamma}^* \rangle \to \langle y, y^* \rangle \text{ and} \\ (f_1 \underset{e}{+} \cdots \underset{e}{+} f_K)(y) \leq \liminf_\gamma (f_1(z_{1,\gamma}) + \cdots + f_K(z_{K,\gamma})). \end{array}\right\} \tag{47}$$

8. Minimization over non–closed convex sets

In the following result, we show how Corollary 5.2 can be used to discuss constrained optimization on a non–closed convex set.

Theorem 8.1. *Let E be a Banach space, X be a (not necessarily closed) convex subset of E, $f \in \mathcal{PCLSC}(E)$ and $x \in \operatorname{dom} f \cap X$. Then the conditions (48)–(51) are equivalent:*

$$f(x) = \min_X f. \tag{48}$$

$$\left.\begin{array}{c}\text{There exists a net } (y_\gamma, y_\gamma^*) \text{ of elements of } G(\partial f) \text{ such that } \|y_\gamma - x\| \to 0,\\ \operatorname{dist}(y_\gamma, X)\|y_\gamma^*\| \to 0,\ \langle y_\gamma - x, y_\gamma^*\rangle \to 0\\ \text{and, for all } v \in X,\ \liminf_\gamma \langle v - x, y_\gamma^*\rangle \ge 0.\end{array}\right\} \tag{49}$$

$$\left.\begin{array}{c}\text{There exists a net } (y_\gamma, y_\gamma^*) \text{ of elements of } G(\partial f) \text{ such that } \|y_\gamma - x\| \to 0,\\ \operatorname{dist}(y_\gamma, X)\|y_\gamma^*\| \to 0 \text{ and, for all } v \in X,\ \liminf_\gamma \langle v - y_\gamma, y_\gamma^*\rangle \ge 0.\end{array}\right\} \tag{50}$$

$$\left.\begin{array}{c}\text{For all } \varepsilon > 0 \text{ and } v \in X, \text{ there exists } (y, y^*) \in G(\partial f)\\ \text{such that } \|y - x\| < \varepsilon \text{ and } \langle v - y, y^*\rangle > -\varepsilon.\end{array}\right\} \tag{51}$$

Proof. It is immediate that (49)$\Longrightarrow$(50)$\Longrightarrow$(51).

(51)$\Longrightarrow$(48) Let v be an arbitrary element of X and $\varepsilon > 0$. Since f is lower semicontinuous, there exists $\delta > 0$ such that

$$\|y - x\| < \delta \quad \Longrightarrow \quad f(y) > f(x) - \varepsilon/2.$$

From (51), there exists $(y, y^*) \in G(\partial f)$ such that $\|y - x\| < \delta$ and $\langle v - y, y^*\rangle > -\varepsilon/2$. Consequently,

$$f(v) \ge f(y) + \langle v - y, y^*\rangle \ge f(x) - \varepsilon/2 - \varepsilon/2 = f(x) - \varepsilon.$$

Since ε is arbitrary, it follows that $f(v) \ge f(x)$, which gives (48).

(48)$\Longrightarrow$(49) Let $f(x) = \min_X f$. We shall prove that if $\varepsilon > 0$ and $v_1, \ldots, v_n \in X$ then there exists $(y, y^*) \in G(\partial f)$ such that

$$\left.\begin{array}{c}\|y - x\| < \varepsilon,\ |\langle y - x, y^*\rangle| < \varepsilon,\ \operatorname{dist}(y, X)\|y^*\| < \varepsilon \text{ and,}\\ \text{for all } i = 1, \ldots, n,\ \langle v_i - x, y^*\rangle > -\varepsilon.\end{array}\right\} \tag{52}$$

Once this has been done, it is easy to construct the net required in (49). So let $\varepsilon > 0$ and $v_1, \ldots, v_n \in X$. Write $Y := \operatorname{co}\{x, v_1, \ldots, v_n\} \subset X$. Then $f(x) = \min_Y f$, thus $0 \in \partial(f+g)(x)$, where g is the indicator function of Y. However, g is lower semicontinuous so, from (29), enhanced as in Remark 5.3, there exist $(y, y^*) \in G(\partial f)$ and $(z, z^*) \in G(\partial g)$ such that

$$\left.\begin{array}{c}\|y - x\| < \varepsilon,\ |\langle y - x, y^*\rangle| < \varepsilon,\ \|y - z\|\|(y^*, z^*)\|_1 < \varepsilon,\ |\langle z - x, z^*\rangle| < \varepsilon/2\\ \text{and, for all } i = 1, \ldots, n,\ |\langle v_i - x, y^* + z^*\rangle| < \varepsilon/2.\end{array}\right\} \tag{53}$$

Now $\partial g(z)$ is the normal cone to Y at z and, for all $i = 1, \ldots, n$, $v_i \in Y$ and consequently $\langle v_i, z^*\rangle \le \langle z, z^*\rangle$, from which,

$$\langle v_i - x, y^*\rangle \ge -\langle v_i - x, z^*\rangle - \varepsilon/2 = \langle x - v_i, z^*\rangle - \varepsilon/2 \ge \langle x - z, z^*\rangle - \varepsilon/2 > -\varepsilon.$$

(52) follows by combining this with (53) and observing that $\operatorname{dist}(y, X) \le \|y - z\|$ and $\|y^*\| \le \|(y^*, z^*)\|_1$. $\square$

Acknowledgements. We are indebted to the anonymous referee for a number of insightful comments about the original version of this paper.

References

[1] J.–P. Aubin, I. Ekeland: Applied Nonlinear Analysis, Wiley, New York et al. 1984.

[2] A. Brøndsted, R. T. Rockafellar: On the subdifferentiability of convex functions, Proc. Amer. Math. Soc. 16 (1965) 605–611.

[3] R. S. Burachik, B. F. Svaiter: ϵ-enlargements of maximal monotone operators in Banach spaces, Set-Valued Analysis 7 (1999) 117–132.

[4] J.-B. Hiriart-Urruty, M. Moussaoui, A. Seeger, M. Volle: Subdifferential calculus without qualification conditions, using approximate subdifferentials: a survey, Nonlinear Anal. 24 (1995) 1727–1754.

[5] J.-B. Hiriart-Urruty, R. R. Phelps: Subdifferential calculus using ε-subdifferentials, J. Functional Analysis 118 (1993) 154–166.

[6] H. König: On some basic theorems in convex analysis, in: Modern Applied Mathematics, Optimization and Operations Research, B. Korte (ed.), Lect. Summer Sch., Bonn 1979, North-Holland (1982) 107–244.

[7] J.-E. Martinez-Legaz, M. Théra: ε–subdifferentials in terms of subdifferentials, Set-Valued Analysis 4 (1996) 327–332.

[8] M. Moussaoui, A. Seeger: Sensitivity analysis of optimal value functions of convex parametric programs with possibly empty solution sets, SIAM J. Optim. 4 (1994) 659–675.

[9] R. R. Phelps: Convex Functions, Monotone Operators and Differentiability, Lecture Notes in Mathematics 1364, Springer-Verlag (Second Edition), 1993.

[10] J. P. Revalski, M. Théra: Enlargements and sums of monotone operators, Nonlinear Analysis, to appear.

[11] R. T. Rockafellar: On the maximal monotonicity of subdifferential mappings, Pac. J. Math. 33 (1970) 209-216.

[12] S. Simons: Subdifferentials of convex functions, in: Recent Developments in Optimization Theory and Nonlinear Analysis, Y. Censor, S. Reich (eds.), American Mathematical Society, Providence, Rhode Island, Contemporary Mathematics 204 (1997) 217–246.

[13] S. Simons: Minimax and Monotonicity, Lecture Notes in Mathematics 1693, Springer-Verlag, 1998.

[14] S. Simons: Maximal monotone multifunctions of Brøndsted–Rockafellar type, Set-Valued Analysis 7 (1999) 255–294.

[15] S. Simons: Constraint qualifications for convex functions, submitted for publication.

[16] B. F. Svaiter: A family of enlargements of maximal monotone operators, Set-Valued Analysis 8 (2000) 311–328.

[17] L. Thibault: A generalized sequential formula for subdifferentials of sums of convex functions defined on Banach space, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag 429 (1995) 340–345.

[18] L. Thibault: Limiting convex subdifferential calculus with applications to integration and maximal monotonicity of subdifferential, Canadian Mathematical Society Conference Proceedings 27 (2000) 279–289.