Optimization,
Vol. 56, Nos. 1–2, February–April 2007, 149–169

The Hahn–Banach–Lagrange theorem†

S. SIMONS*

Department of Mathematics, University of California,
Santa Barbara, CA 93106-3080, USA

(*Received 6 November 2005; in final form 15 March 2006*)

This article is about a new version of the Hahn–Banach theorem, which we will call the "Hahn–Banach–Lagrange theorem", since it deals very effectively with certain problems of Lagrange type, as well as giving numerous results in functional analysis, convex analysis, and monotone operator theory. We will discuss several of these results in this article.

Keywords: Sublinear functional; Hahn–Banach theorems; Convex function; Conjugate function; Fenchel duality; Lagrange multipliers for convex problems; Minimax theorem

Mathematics Subject Classifications 2000: 46A22; 46N10; 90C46; 49J35

1. Introduction

The idea behind this article is to provide a unified and relatively nontechnical framework for treating the main existence theorems for continuous linear functionals in functional analysis, convex analysis, Lagrange multiplier theory, and minimax theory. While many of the results in this article are already known, our approach is new, and gives a large number of results with considerable economy of effort. More specifically, we prove a version of the Hahn–Banach theorem, the "Hahn–Banach–Lagrange theorem", that is sufficiently strong that all the aforementioned results follow fairly directly from it. Many of the results in this article, which are already known, are usually proved using the Eidelheit separation theorem in a product space. The proofs of these results given here avoid the problem of the elimination of the "vertical hyperplane". Furthermore, our approach leads naturally to the sharp numerical bounds for various problems discussed in sections 6, 8, and 10. (The sharp numerical bound discussed in

*Email: simons@math.ucsb.edu
†Dedicated to Diethard Pallaschke, on the occasion of his 65th birthday.

Optimization
ISSN 0233-1934 print/ISSN 1029-4945 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/02331930600819969

section 8 has applications to monotone multifunction theory that are beyond the scope of this study.)

The Hahn–Banach–Lagrange theorem appears in Theorem 2.9, which is bootstrapped from the special case contained in Lemma 2.4, which is, in turn, bootstrapped from the classical Hahn–Banach theorem for sublinear functionals contained in Lemma 2.3. Theorem 2.9 uses the concept of "*S*-convexity" introduced in Definition 2.7. The full force of this concept will be used only in Corollary 3.4, a result on convex functions with applications to a minimax theorem, and the analysis of Lagrange multipliers in Theorem 10.1(a). For all the other applications of Theorem 2.9 in this article, the reader can substitute "affine" for "*S*-convex". This change shortens the proof of Lemma 2.2 by a few lines.

In section 3, we sketch how Theorem 2.9 can be used to give the main existence theorems for linear functionals in functional analysis, and also how it gives the result, referred to above, that leads to a minimax theorem.

Sections 4–8 contain various applications of the Hahn–Banach–Lagrange theorem to convex analysis. In section 4, we show how Theorem 2.9 leads to a necessary and sufficient condition for the Fenchel duality inequality to hold in locally convex spaces. As a consequence, we deduce a sufficient condition in which neither function satisfies a semicontinuity condition and also a classical result due to Rockafellar. In section 5, we discuss an application of Theorem 2.9 to a "localized" version of the Fenchel–Moreau formula. In section 6, we give applications of the results of section 4 to the normed case, and obtain a *sharp lower bound* of an intuitive geometric character for the norm of the functional produced by the Fenchel duality theorem. In section 7, we bootstrap the results of section 4, and obtain the corresponding results on the conjugate of the sum of two convex functions. We also give an example of a very "unstable" conjugate duality. In section 8, we investigate some properties of the function $\frac{1}{2}\|\cdot\|^2$, and prove a sharp version of the Fenchel duality theorem that has been very useful in the investigation of monotone multifunctions.

In section 9, we give a short proof of a fundamental result on dual problems and Lagrangians due to Rockafellar.

In section 10, we show how the Hahn–Banach–Lagrange theorem can be used to obtain considerable insight on the existence of Lagrange multipliers for constrained convex minimization problems. In Theorem 10.1, we use Theorem 2.9 to obtain a *necessary and sufficient* condition for the existence of Lagrange multipliers, with a *sharp lower bound* on the norm of the multiplier. In Theorem 10.3, we deduce the classical sufficient "Slater condition", with the added bonus that we obtain a *sharp lower bound* on the norm of the multiplier.

Some (but not all) of the results in this article have appeared already in [14] and [15]. In both of these articles, the analysis was directed to applications to monotone multifunctions. In this article, we have concentrated more on the aspect of these results that are of interest in optimization theory.

Prof. Constantin Zălinescu [17] has written an excellent and comprehensive monograph on convex analysis, and the author is grateful to Prof. Zălinescu for many constructive comments, pointing out how many of the results on convex analysis presented here can be deduced from results in [17].

All vector spaces in this article are *real*.

2. The Hahn–Banach–Lagrange theorem

The Hahn–Banach–Lagrange theorem, Theorem 2.9 is the main topic of this article, and is proved using the technique of the "auxiliary sublinear functional". Most of the work for it is actually done in the technical Lemma 2.2.

Definition 2.1 Let E be a nontrivial vector space. We say that $S : E \mapsto \mathbb{R}$ is *sublinear* if

$$x, y \in E \Rightarrow S(x + y) \leq S(x) + S(y)$$

and

$$x \in E \text{ and } \lambda > 0 \Rightarrow S(\lambda x) = \lambda S(x).$$

Then, we note that it follows automatically that $S(0) = 0$.

LEMMA 2.2 *Let E be a nontrivial vector space and $S : E \mapsto \mathbb{R}$ be sublinear. Let D be a nonempty convex subset of a vector space, $a : D \mapsto E$ be affine and $\beta := \inf_D S \circ a \in \mathbb{R}$. For all $x \in E$, let*

$$T(x) := \inf_{d \in D,\, \lambda > 0} [S(x + \lambda a(d)) - \lambda\beta]. \tag{2.1}$$

Then $T : E \mapsto \mathbb{R}$, T is sublinear, $T \leq S$ on E and, for all $d \in D$, $-T(-a(d)) \geq \beta$.

Proof If $x \in E$, $d \in D$, and $\lambda > 0$ then

$$S(x + \lambda a(d)) - \lambda\beta \geq -S(-x) + \lambda S(a(d)) - \lambda\beta \geq -S(-x) > -\infty.$$

Taking the infimum over $d \in D$ and $\lambda > 0$, $T(x) \geq -S(-x) > -\infty$. Thus, $T : E \mapsto \mathbb{R}$. It is now easy to check that T is positively homogeneous, and so to prove that T is sublinear it remains to show that T is subadditive. To this end, let $x_1, x_2 \in E$. Let $d_1, d_2 \in D$ and $\lambda_1,\ \lambda_2 > 0$ be arbitrary. Write $x := x_1 + x_2$, $\lambda := \lambda_1 + \lambda_2$, $\mu_i := \lambda_i/\lambda$, and $d := \mu_1 d_1 + \mu_2 d_2$. Then, using the fact that $\mu_1 a(d_1) + \mu_2 a(d_2) = a(d)$,

$$\begin{aligned}
&[S(x_1 + \lambda_1 a(d_1)) - \lambda_1\beta] + [S(x_2 + \lambda_2 a(d_2)) - \lambda_2\beta] \\
&\quad \geq S(x + \lambda_1 a(d_1) + \lambda_2 a(d_2)) - \lambda\beta \\
&\quad = \lambda S(x/\lambda + \mu_1 a(d_1) + \mu_2 a(d_2)) - \lambda\beta, \\
&\quad = \lambda S(x/\lambda + a(d)) - \lambda\beta \\
&\quad = S(x + \lambda a(d)) - \lambda\beta \\
&\quad \geq T(x) = T(x_1 + x_2).
\end{aligned}$$

Taking the infimum over d_1, d_2, λ_1, and λ_2 gives $T(x_1) + T(x_2) \geq T(x_1 + x_2)$. Thus, T is subadditive, and consequently sublinear. Fix $d \in D$. Let x be an arbitrary element of E. Then, for all $\lambda > 0$, $T(x) \leq S(x) + \lambda[S(a(d)) - \beta]$. Letting $\lambda \to 0$, $T(x) \leq S(x)$. Thus, $T \leq S$ on E. Finally, let d be an arbitrary element of D. Then, taking $\lambda = 1$ in (2.1),

$$T(-a(d)) \leq S(-a(d) + a(d)) - \beta = -\beta,$$

Hence, $-T(-a(d)) \geq \beta$, which completes the proof of Lemma 2.2. ■

We now recall in Lemma 2.3 the classical Hahn–Banach theorem for sublinear functionals.

LEMMA 2.3 *Let E be a nontrivial vector space and $S : E \mapsto \mathbb{R}$ be sublinear. Then there exists a linear functional L on E such that $L \le S$ on E.*

Proof See [5, Theorem 3.4, p. 21] for a proof using cones, [11, Theorem 3.2, p. 56–57] for a proof using an extension by subspaces argument, and [6] and [12] for a proof using an ordering on sublinear functionals. ■

Remark 3.5 contains some comments on the appropriateness of the various methods of proof for Lemma 2.3.

Lemma 2.4 is an intermediate result that we establish on our way to Theorem 2.9. Splitting the proof of Theorem 2.9, as we have done here, enables us to avoid a considerable amount of computation compared to the original method used in [14].

LEMMA 2.4 *Let E be a nontrivial vector space and $S : E \mapsto \mathbb{R}$ be sublinear. Let D be a nonempty convex subset of a vector space and $a : D \mapsto E$ be affine. Then there exists a linear functional L on E such that $L \le S$ on E and*

$$\inf_D L \circ a = \inf_D S \circ a.$$

Proof Let $\beta := \inf_D S \circ a$. If $\beta = -\infty$, the result is immediate from Lemma 2.3 (take any linear functional L on E such that $L \le S$ on E). So we can suppose that $\beta \in \mathbb{R}$. Define the auxiliary sublinear functional, T, as in Lemma 2.2. From Lemma 2.3, there exists a linear functional L on E such that $L \le T$ on E. Since $T \le S$ on E, $L \le S$ on E, as required. Let $d \in D$. Then

$$L(a(d)) = -L(-a(d)) \ge -T(-a(d)) \ge \beta.$$

Taking the infimum over $d \in D$,

$$\inf_D L \circ a \ge \beta = \inf_C S \circ a.$$

On the other hand, since $L \le S$ on E, $\inf_D L \circ a \le \inf_D S \circ a$. ■

Remark 2.5 It is worth pointing out that the definition of the auxiliary sublinear functional used to prove Lemma 2.4 is "forced" in the sense that if L is linear, $L \le S$ on E and $\beta = \inf_D S \circ a = \inf_D L \circ a \in \mathbb{R}$ then, as the reader can easily verify, $L \le T$ on E.

Definition 2.6 Let C be a nonempty convex subset of a vector space and $\mathcal{PC}(C)$ stand for the set of all convex functions $k : C \mapsto \,]-\infty, \infty]$ such that $\operatorname{dom} k \ne \emptyset$, where $\operatorname{dom} k$, the *effective domain* of k, is defined by

$$\operatorname{dom} k := \{x \in C : k(x) \in \mathbb{R}\}.$$

(The "$\mathcal{P}$" stands for "proper", which is the adjective frequently used to denote the fact that a function is finite at least one point).

Definition 2.7 Let E be a nontrivial vector space and $S : E \mapsto \mathbb{R}$ be sublinear. Define the ordering "$\le_S$" on E by declaring that $y \le_S z$ if $S(y - z) \le 0$. Let C be a nonempty convex subset of a vector space and $j : C \mapsto E$. We say that j is *S-convex* if

$$x_1, x_2 \in C,\ \mu_1, \mu_2 > 0 \text{ and } \mu_1 + \mu_2 = 1 \Rightarrow j(\mu_1 x_1 + \mu_2 x_2) \le_S \mu_1 j(x_1) + \mu_2 j(x_2).$$

An affine function is clearly S-convex. As observed in the introduction, apart from the applications in Corollary 3.4 and Theorem 10.1(a), all the S-convex functions in this article will, in fact, be affine.

Remark 2.8 "S-convex" can mean different things under different circumstances. Consider the special case when $E=\mathbb{R}$. If $S(y):=|y|$, $S(y):=y$, $S(y):=-y$, or $S(y):=0$, respectively, then "S-convex" means "affine", "convex", "concave" or "arbitrary", respectively.

THEOREM 2.9 *Let E be a nontrivial vector space and $S: E\mapsto\mathbb{R}$ be sublinear. Let C be a nonempty convex subset of a vector space $k\in\mathcal{PC}(C)$ and $j: C\mapsto E$ be S-convex. Then there exists a linear functional L on E such that $L\le S$ on E and*

$$\inf_C[L\circ j+k]=\inf_C[S\circ j+k].$$

Proof Let $\tilde{E}:=E\times\mathbb{R}$, and define $\tilde{S}:\tilde{E}\mapsto\mathbb{R}$ by

$$\tilde{S}(y,\lambda):=S(y)+\lambda\quad\big((y,\lambda)\in\tilde{E}\big).$$

Then, as the reader can easily verify, $\tilde{S}$ is sublinear. Let

$$D:=\big\{(x,y,\lambda)\in C\times E\times\mathbb{R}\colon S(j(x)-y)\le 0,\ k(x)\le\lambda\big\},$$

and $a: D\mapsto\tilde{E}$ be defined by

$$a(x,y,\lambda):=(y,\lambda)\quad((x,y,\lambda)\in D).$$

Then D is a convex set and a is an affine function. Lemma 2.4 with E replaced by $\tilde{E}$, S by $\tilde{S}$, and C by D now gives a linear functional $\tilde{L}$ on $\tilde{E}$ such that

$$\tilde{L}\le\tilde{S}\text{ on }\tilde{E}\quad\text{and}\quad\inf_D\tilde{L}\circ a=\inf_D\tilde{S}\circ a.$$

Since $\tilde{L}\le\tilde{S}$ on $\tilde{E}$, there exists a linear functional L on E such that

$$L\le S\text{ on }E\quad\text{and}\quad(y,\lambda)\in\tilde{E}\Rightarrow\tilde{L}(y,\lambda)=L(y)+\lambda.$$

The result follows since, by direct computation,

$$\inf_D\tilde{L}\circ a=\inf_C[L\circ j+k]\quad\text{and}\quad\inf_D\tilde{S}\circ a=\inf_C[S\circ j+k].\qquad\blacksquare$$

3. Applications to functional analysis and minimax theorems

Corollary 3.1 is the *sandwich theorem* [6, Theorem 1.7, p. 112]. It follows immediately from Theorem 2.9 with $C:=E$ and $j(x):=x$.

COROLLARY 3.1 *Let E be a nontrivial vector space, $S: E\mapsto\mathbb{R}$ be sublinear, $k\in\mathcal{PC}(E)$ and $-k\le S$ on E. Then there exists a linear functional L on E such that $-k\le L\le S$ on E.*

Corollary 3.2 is the *extension form of the Hahn–Banach theorem*. It follows immediately from Theorem 2.9 with $C:=F$, $j(x):=x$, and $k:=-M$. It can also be deduced from Corollary 3.1 [6, Corollary 1.8, p. 112].

Corollary 3.2 *Let E be a nontrivial vector space, F be a linear subspace of E, $S: E \mapsto \mathbb{R}$ be sublinear, $M: F \mapsto \mathbb{R}$ be linear and $M \leq S$ on F. Then there exists a linear functional L on E such that $L \leq S$ on E and $L|_F = M$.*

Corollary 3.3 is the *Mazur–Orlicz theorem*. It follows immediately from Lemma 2.4 with $a(x) := x$. It can also be deduced from Theorem 2.9 or Corollary 3.1 [6, Theorem 1.9, p. 112].

Corollary 3.3 *Let E be a nontrivial vector space, $S: E \mapsto \mathbb{R}$ be sublinear and D be a nonempty convex subset of E. Then there exists a linear functional L on E such that $L \leq S$ on E and $\inf_D L = \inf_D S$.*

Corollary 3.4 was essentially proved by Fan–Glicksberg–Hoffman ([4, Theorem 1, p. 618], and leads to a short proof of the minimax theorem proved by Fan in [3] see ([13, Theorem 3.1, p. 17] for details of this). Corollary 3.4 follows easily from Theorem 2.9 with $E := \mathbb{R}^m$, $S(\mu_1, \ldots, \mu_m) := \mu_1 \vee \cdots \vee \mu_m$, $j(c) := (f_1(c), \ldots, f_m(c))$ and $k(c) := 0$.

Corollary 3.4 *Let C be a nonempty convex subset of a vector space and $f_1, \ldots, f_m$ be convex real functions on C. Then there exist $\lambda_1, \ldots, \lambda_m \geq 0$ such that $\lambda_1 + \cdots + \lambda_m = 1$ and*

$$\inf_C [f_1 \vee \cdots \vee f_m] = \inf_C [\lambda_1 f_1 + \cdots + \lambda_m f_m].$$

Remark 3.5 Since we have given Corollary 3.2 as (ultimately) a consequence of Lemma 2.3, in order to dispel any suspicion of circularity, it would seem better to avoid the ''extension by subspaces'' proof of Lemma 2.3 and using instead the ''cone'' argument of Kelly–Namioka [5] or the ''minimal sublinear functional'' argument outlined in the following text, which is most in tune with the other analysis in this article. It is easy to see from Zorn's lemma that, if $U: E \mapsto \mathbb{R}$ is sublinear, then there exists a sublinear functional S on E such that $S \leq U$ on E and S is minimal with respect to the pointwise ordering of $\mathbb{R}^E$. Let d be an arbitrary element of E. If we define $D := \{d\}$ and $a(d) := d$, then Lemma 2.2 yields a sublinear functional T on E such that $T \leq S$ on E and $-T(-d) \geq S(d)$. The minimality of S now gives $T = S$, thus we have proved that $-S(-d) \geq S(d)$. Since S is sublinear, it follows easily from this that S is linear. This gives us a proof of Lemma 2.3 that does not depend on an ''extension by subspaces'' argument. More details of this approach can be found in [6,12].

4. The Fenchel duality theorem in the locally convex case

In this section, we show how Theorem 2.9 leads to a necessary and sufficient condition for the Fenchel duality inequality (4.1) to hold. As a consequence, we deduce a sufficient condition in which neither function satisfies a semicontinuity condition, and also a classical result due to Rockafellar.

Let E be a nontrivial real Hausdorff locally convex space with dual E^*. $\mathcal{S}(E)$ stands for the family of all continuous seminorms on E. if $k \in \mathcal{PC}(E)$, the *Fenchel conjugate*, k^*, of k is the function from E^* into $]-\infty, \infty]$ defined by

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

Theorem 4.1 or its proof will be used not only in Corollary 4.2 and Theorem 4.3 but also in Corollary 6.1. It is worth noting that if we take the duality formula of Zălinescu [17, Corollary 2.8.5, p. 125] as known, then Theorem 4.1 can be deduced using the convex function $(x,y)\mapsto f(x)+g(y)$ on $E\times E$ and the linear map $(x,y)\mapsto x-y$ mapping $E\times E$ into E.

THEOREM 4.1 *Let E be a nontrivial Hausdorff locally convex space with dual E^*, and $f,g\in\mathcal{PC}(E)$. Then*

$$\textit{there exists } z^*\in E^* \textit{ such that } f^*(-z^*)+g^*(z^*)\le 0 \tag{4.1}$$

if, and only if,

$$\textit{there exists } S\in\mathcal{S}(E) \textit{ such that } \inf_{x,y\in E}[f(x)+g(y)+S(x-y)]\ge 0. \tag{4.2}$$

Proof Suppose first that statement (4.2) is satisfied. Then, for all $x,y\in E$,

$$\langle x,\,-z^*\rangle - f(x)+\langle y,z^*\rangle - g(y)\le f^*(-z^*)+g^*(z^*)\le 0,$$

consequently,

$$f(x)+g(y)+\langle x-y,z^*\rangle\ge 0, \tag{4.3}$$

and statement (4.2) follows with $S:=|z^*|$. Suppose, conversely, that statement (4.2) is satisfied. Then we apply Theorem 2.9 with $C:=E\times E$, $j(x,y):=x-y$ and $k(x,y):=f(x)+g(x)$ and obtain a linear functional L on E such that $L\le S$ and $\inf_{x,y\in E}[f(x)+g(y)+L(x-y)]\ge 0$, or equivalently, $\sup_{x,y\in E}[(-L)(x)-f(x)+L(y)-\ g(y)]\le 0$. (4.1) now follows (with $z^*=L$) by taking the supremum. ■

Corollary 4.2 or its proof will be used explicitly in our discussion of Fenchel–Moreau points in Theorem 5.2, and also in Corollary 6.4.

COROLLARY 4.2 *Let E be a nontrivial Hausdorff locally convex space with dual E^*, $x\in E$, $\lambda\in\mathbb{R}$ and $g\in\mathcal{PC}(E)$. Then there exists $z^*\in E^*$ such that $\langle x,z^*\rangle - g^*(z^*)\ge\lambda$ if, and only if, there exists $S\in\mathcal{S}(E)$ such that* $\inf_{y\in E}[g(y)+S(x-y)]\ge\lambda$.

Proof Define $f\in\mathcal{PC}(E)$ by

$$f(z):=\begin{cases}-\lambda & \text{if } z=x;\\ \infty, & \text{otherwise}\end{cases}$$

Then, for all $z^*\in E^*$, $f^*(-z^*)=-\langle x,z^*\rangle+\lambda$, and the result follows from Theorem 4.1. ■

Theorem 4.3 is a version of the Fenchel duality theorem for locally convex spaces in which neither of the functions satisfies any semicontinuity conditions. It will be bootstrapped in Theorem 7.1.

THEOREM 4.3 *Let E be a nontrivial Hausdorff locally convex space with dual E^*, f, $g\in\mathcal{PC}(E)$, $f+g\ge 0$ on E, and g be (finitely) bounded above in some neighborhood of a point of* $\operatorname{dom} f$. *Then*

$$\textit{there exists } z^*\in E^* \textit{ such that } f^*(-z^*)+g^*(z^*)\le 0. \tag{4.1}$$

Proof Choose $z\in\operatorname{dom} f$, $N\in\mathbb{R}$ and $T\in\mathcal{S}(E)$ such that $T(w)\le 1\Rightarrow g(z+w)\le N$, and define $M:=f(z)+N\ge f(z)+g(z)=(f+g)(z)\ge 0$, and $S:=MT\in\mathcal{S}(E)$.

Let x and y be arbitrary elements of E, λ be any real number such that $\lambda > T(x-y) \geq 0$, and write $w := (x-y)/\lambda$. Since $T(w) < 1$, our hypotheses imply that

$$M = f(z) + N \geq f(z) + g(z+w). \tag{4.4}$$

Now $x = y + \lambda w$, and so

$$\frac{x + \lambda z}{1+\lambda} = \frac{y + \lambda w + \lambda z}{1+\lambda} = \frac{y + \lambda(z+w)}{1+\lambda}.$$

Since $f + g \geq 0$ on E,

$$0 \leq f\left(\frac{x+\lambda z}{1+\lambda}\right) + g\left(\frac{y+\lambda(z+w)}{1+\lambda}\right) \leq \frac{f(x) + \lambda f(z) + g(y) + \lambda g(z+w)}{1+\lambda}.$$

Consequently, from (4.4),

$$f(x) + g(y) + \lambda M \geq f(x) + g(y) + \lambda[f(z) + \lambda g(z+w)] \geq 0.$$

Letting $\lambda \to T(x-y)$ in this, we have $f(x) + g(y) + MT(x-y) \geq 0$, that is to say, $f(x) + g(y) + S(x-y) \geq 0$. This gives (4.2), and the result now follows from Theorem 4.1. ■

Corollary 4.4 is an immediate consequence of Theorem 4.3 (see [9, Theorem 1, pp. 82–83] or [17, Theorem 2.8.3(iii), p. 123]). In fact, if we use [17, Lemma 2.2.8, p. 64], we can derive Theorem 4.3 from Corollary 4.4.

Corollary 4.4 *Let E be a nontrivial Hausdorff locally convex space with dual E^*, f, $g \in \mathcal{PC}(E)$, $f + g \geq 0$ on E and g be finite and continuous at a point of* dom f. *Then*

$$\textit{there exists } z^* \in E^* \textit{ such that } f^*(-z^*) + g^*(z^*) \leq 0. \tag{4.1}$$

5. Fenchel–Moreau points

Let E be a nontrivial real Hausdorff locally convex space with dual E^* and $g \in \mathcal{PC}(E)$. It follows easily from the definition of g^* in the previous section that, for all $x \in E$,

$$g(x) \geq \sup_{E^*}(\langle x, \cdot \rangle - g^*). \tag{5.1}$$

One of the fundamental results in convex analysis is the *Fenchel–Moreau formula* that, if $g \in \mathcal{PC}(E)$ is lower semicontinuous, then we *always have equality in* (5.1) [8, sections 5–6, pp. 26–39].

Now suppose that g is not necessarily lower semicontinuous. Let us say that $x \in E$ is a *Fenchel–Moreau point of* g if equality holds in (5.1). It is very tempting to speculate that every point of lower semicontinuity of g is a Fenchel–Moreau point of g. Remark 5.1 shows that this is false. However, we establish in Theorem 5.3 that every point of lower semicontinuity of g is a Fenchel–Moreau point provided that g is bounded below in a neighborhood of at least one point in its effective domain. Our proof goes by way of Theorem 5.2, which contains a simple characterization of the Fenchel–Moreau points of g.

Remark 5.1 Let E be an infinite–dimensional normed space. Fix $x^* \in E^* \backslash \{0\}$ and a discontinuous linear functional L on E. Define

$$g(x) := \begin{cases} \infty & \text{if } \langle x, x^* \rangle < 1; \\ L(x) & \text{if } \langle x, x^* \rangle \geq 1. \end{cases}$$

Clearly, $g \in \mathcal{PC}(E)$ and g is lower semicontinuous at 0. Let y^* be an arbitrary element of E^*. Since x^* and $y^* - L$ are linearly independent, there exist u, $v \in E$ such that

$$\langle u, x^* \rangle = 1, \langle v, x^* \rangle = 0, (y^* - L)(u) = 0, \text{ and } (y^* - L)(v) = 1.$$

Let $\lambda \in \mathbb{R}$, and set $x := u + \lambda v$. Then $\langle x, x^* \rangle = \langle u, x^* \rangle = 1$, and so $k(x) = L(x)$. Thus,

$$g^*(y^*) \geq \langle x, y^* \rangle - g(x) = (y^* - L)(x) = \lambda(y^* - L)(v) = \lambda.$$

Since this holds for all $\lambda \in \mathbb{R}$, $g^*(y^*) = \infty$. Thus, we have

$$g(0) = \infty > -\infty = \sup_{E^*}(0 - g^*),$$

and so 0 is not a Fenchel–Moreau point of g. (This example can also be justified using the Moreau–Rockafellar theorem on the conjugate of the sum of two convex functions that will appear in Corollary 7.2.)

We now come to the promised characterization of Fenchel–Moreau points.

Theorem 5.2 *Let E be a nontrivial Hausdorff locally convex space with dual E^*, $x \in E$ and $g \in \mathcal{PC}(E)$. Then x is a Fenchel–Moreau point of g if, and only if,*

$$\textit{for all } \lambda < g(x), \textit{ there exists } S \in \mathcal{S}(E) \textit{ such that } \inf_{y \in E}[g(y) + S(y - x)] \geq \lambda.$$

Proof This is immediate from Corollary 4.2. ■

Theorem 5.3 contains a positive result on Fenchel–Moreau points.

Theorem 5.3 *If $g \in \mathcal{PC}(E)$ is bounded below in a neighborhood of $z \in \operatorname{dom} g$, and lower semicontinuous at $x \in E$ then x is a Fenchel–Moreau point of g.*

Proof We will use Theorem 5.2. Let $\lambda < g(x)$. Choose $\nu \in \mathbb{R}$ and $T \in \mathcal{S}(E)$ such that

$$T(y - z) \leq 1 \Rightarrow g(y) > \nu \tag{5.2}$$

and

$$T(y - x) \leq 1 \Rightarrow g(y) > \lambda. \tag{5.3}$$

Write $\rho := g(z) - \nu > 0$. We first prove that

$$\inf_{y \in E}[g(y) + \rho T(y - x)] \geq \nu - \rho T(x - z). \tag{5.4}$$

To this end, let y be an arbitrary element of E. If $T(y - z) \leq 1$ then (5.2) implies that

$$g(y) + \rho T(y - z) \geq g(y) > \nu \geq \nu - \rho T(x - z).$$

If, on the other hand, $T(y - z) > 1$, let $\gamma := 1/T(y - z) \in]0, 1[$ and put $u := \gamma y + (1 - \gamma)z$. Then $T(u - z) = \gamma T(y - z) = 1$ and so, from (5.2) with y replaced by u,

$$\gamma g(y) + (1 - \gamma)g(z) \geq g(\gamma y + (1 - \gamma)z) = g(u) > \nu,$$

thus the definition of ρ implies that $\gamma(g(y)) - g(z)) + \rho \geq 0$. Substituting in the formula for γ and clearing of fractions yields $g(y) + \rho T(y - z) \geq g(z)$. Consequently, using (5.2) and the fact that $T(z - z) \leq 1$,

$$g(y) + \rho T(y - x) \geq g(y) + \rho T(y - z) - \rho T(x - z) \geq g(z) - \rho T(x - z) > \nu - \rho T(x - z).$$

This completes the proof of (5.4). Now choose μ so that $\mu \geq \rho$ and $\mu - \rho \geq \lambda - \nu + \rho T(x - z)$. We will prove that

$$\inf_{y \in E} [g(y) + \mu T(y - x)] \geq \lambda. \tag{5.5}$$

To this end, let y be an arbitrary element of E. If $T(y - x) \leq 1$ then (5.3) implies that $g(y) + \mu T(y - x) \geq g(y) > \lambda$. If, on the other hand, $T(y - x) > 1$ then, from (5.4) and the choice of μ,

$$\begin{aligned} g(y) + \mu T(y - x) &= g(y) + \rho T(y - x) + (\mu - \rho) T(y - x) \\ &\geq \nu - \rho T(x - z) + (\mu - \rho) \geq \lambda. \end{aligned}$$

This completes the proof of (5.5), and the result now follows from Theorem 5.2 with $S := \mu T \in \mathcal{S}(E)$. ■

Remark 5.4 Another way of stating the result of Theorem 5.3 is that if there is a point of lower semicontinuity of g that is not a Fenchel–Moreau point then g is unbounded below in every neighborhood of every point of $\operatorname{dom} g$. Of course, Theorem 5.3 provides a proof of the original Fenchel–Moreau formula when g is lower semicontinuous. Another way of proving Theorem 5.3 is to apply [17, Proposition 2.2.5, p. 62] to the lower semicontinuous envelope of g, which is also defined in [17, p. 62].

6. The Fenchel duality theorem in the normed case

In this section, we revisit the analysis of section 4 in the normed case, and obtain a sharp lower bound for the norm of the functional produced by the Fenchel duality theorem.

COROLLARY 6.1 *Let E be a nontrivial normed space with dual E^*, and $f, g \in \mathcal{PC}(E)$. Then*:

(a) *There exists $z^* \in E^*$ such that $f^*(-z^*) + g^*(z^*) \leq 0$ if, and only if,*

$$\textit{there exists } M \geq 0 \textit{ such that } \inf_{x, y \in E} [f(x) + g(y) + M\|x - y\|] \geq 0.$$

(b) *If $z^* \in E^*$ and $f^*(-z^*) + g^*(z^*) \leq 0$ then*

$$\sup_{x, y \in E, x \neq y} \frac{-f(x) - g(y)}{\|x - y\|} \leq \|z^*\| < \infty.$$

(c) *If $f + g > 0$ on E and*

$$\sup_{x, y \in E, x \neq y} \frac{-f(x) - g(y)}{\|x - y\|} < \infty$$

then

$$\min\{\|z^*\|: z^* \in E^*,\ f^*(-z^*) + g^*(z^*) \leq 0\} = \sup_{x, y \in E, x \neq y} \frac{-f(x) - g(y)}{\|x - y\|} \vee 0. \tag{6.1}$$

Proof

(a) This is immediate from Theorem 4.1.
(b) Inequality (4.3) implies that, for all $x, y \in E$, $-f(x) - g(y) \leq \langle x - y, z^* \rangle \leq \|x - y\| \|z^*\|$, which gives the required result.
(c) The inequality "$\geq$" in (6.1) follows from (b) and the fact that $\|z^*\| \geq 0$. Now write M for the right-hand side of (6.1). Then $M \geq 0$ and

$$x, y \in E \text{ and } x \neq y \Rightarrow \frac{-f(x) - g(y)}{\|x - y\|} \leq M \Rightarrow f(x) + g(y) + M\|x - y\| \geq 0.$$

If $x = y$, then $f(x) + g(y) + M\|x - y\| = (f + g)(x) \geq 0$. Thus, (4.2) is satisfied with $S := M\|\cdot\| \in \mathcal{S}(E)$. It now follows from the proof of Theorem 4.1 that there exists $z^* \in E^*$ satisfying (4.1) such that $z^* \leq M\|\cdot\|$ on E. Since this implies that $\|z^*\| \leq M$, we have established the inequality "$\leq$" in (6.1). ■

Remark 6.2 In this remark, we discuss a geometric interpretation of Corollary 6.1(c). We write $\mathcal{CA}(E)$ for the set of all continuous affine real functions on E. If $a \in \mathcal{CA}(E)$, then a can be written uniquely in the form $a = z^* + \alpha$, where $z^* \in E^*$ and $\alpha \in \mathbb{R}$. Since z^* is the derivative of a in any reasonable sense, we shall write $z^* = Da$. Turning now to Corollary 6.1(c), write $h = -f$ so that h is proper and concave and $h \leq g$ on E. Let $a \in \mathcal{CA}(E)$. Then it is easily seen by direct computation that, if $h \leq a \leq g$ on E, then $f^*(-Da) + g^*(Da) \leq 0$, and conversely, if $f^*(-z^*) + g^*(z^*) \leq 0$, then there exists $a \in \mathcal{CA}(E)$ such that $Da = z^*$ and $h \leq a \leq g$ on E. Furthermore,

$$\sup_{x, y \in E, x \neq y} \frac{-f(x) - g(y)}{\|x - y\|} = \sup_{x, y \in E, x \neq y} \frac{h(x) - g(y)}{\|x - y\|}$$

Now suppose, in addition, that $\sup_E h > \inf_E g$ to avoid the "$\vee$" in (6.1). Then the conclusion of Corollory 6.1(c) is

$$\min\{\|Da\| \colon a \in \mathcal{CA}(E), h \leq a \leq g \text{ on } E\} = \sup_{x, y \in E, x \neq y} \frac{h(x) - g(y)}{\|x - y\|}.$$

The quotient on the right-hand side of the equality above is, of course, the slope of the line segment going from the point $(y, g(y))$ on the graph of g to the point $(x, h(x))$ on the graph of h. The example $E = \mathbb{R}$ and $f = g = |\cdot|$ shows that the "$\vee$ 0" term in (6.1) is required.

Remark 6.3 It is shown in [13, section 14, pp. 49–51] how Corollary 6.1 leads to a proof of the version of the Fenchel duality theorem due to Attouch–Brezis [1, Theorem 1.1, pp. 126–130]. We emphasize that Theorem 4.1 and Corollary 6.1 give a *necessary and sufficient* condition for the existence of the linear functional, and not merely *sufficient* conditions.

Our next result will be used in Lemma 8.1.

Corollary 6.4 *Let E be a nontrivial normed space with dual E^*, $x \in E$, $\lambda \in \mathbb{R}$, $g \in \mathcal{PC}(E)$, $g(x) \geq \lambda$ and g be bounded above in some neighborhood of x (in particular, this is true if g is continuous at x). Then*

$$\min\{\|z^*\| \colon z^* \in E^*, g^*(z^*) - \langle x, z^* \rangle + \lambda \leq 0\} = \sup_{y \in E, y \neq x} \frac{\lambda - g(y)}{\|x - y\|} \vee 0 < \infty.$$

Proof Define $f \in \mathcal{PC}(E)$ as in Corollary 4.2. Then the result follows from Theorem 4.3 and Corollary 6.1 (b and c). ■

7. The conjugate of a sum

Let E be a nontrivial Hausdorff locally convex space with dual E^*, and f, $g \in \mathcal{PC}(E)$. If $y^*, z^* \in E^*$, and $x \in E$, then

$$\langle x, y^* + z^*\rangle - (f+g)(x) \le \langle x, y^*\rangle - f(x) + \langle x, z^*\rangle - g(x) \le f^*(y^*) + g^*(z^*).$$

Thus, taking the supremum over $x \in E$, we have $(f+g)^*(y^* + z^*) \le f^*(y^*) + g^*(z^*)$. Consequently,

$$x^* \in E^* \Rightarrow (f+g)^*(x^*) \le \inf\{f^*(y^*) + g^*(z^*)\colon y^*, z^* \in E^*, y^* + z^* = x^*\}.$$

In this section, we will be concerned with the question when

$$x^* \in E^* \Rightarrow (f+g)^*(x^*) = \min\{f^*(y^*) + g^*(z^*)\colon y^*, z^* \in E^*, y^* + z^* = x^*\}. \qquad (7.1)$$

We now bootstrap Theorem 4.3.

THEOREM 7.1 *Let E be a nontrivial Hausdorff locally convex space with dual E^*, f, $g \in \mathcal{PC}(E)$, and g be (finitely) bounded above in some neighborhood of a point of* $\operatorname{dom} f$. *Then implication* (7.1) *is true.*

Proof We can and will suppose that $(f+g)^*(x^*) \in \mathbb{R}$. Let $x^* \in E^*$. The Fenchel–Young inequality implies that $(f - x^* + (f+g)^*(x^*)) + g = (f+g) + (f+g)^*(x^*) - x^* \ge 0$ on E. Since $\operatorname{dom}(f - x^* + (f+g)^*(x^*)) = \operatorname{dom} f$, Theorem 4.3 implies that there exists $z^* \in E^*$ such that $(f - x^* + (f+g)^*(x^*))^*(-z^*) + g^*(z^*) \le 0$. However,

$$(f - x^* + (f+g)^*(x^*))^*(-z^*) = f^*(x^* - z^*) - (f+g)^*(x^*),$$

and so $f^*(x^* - z^*) - (f+g)^*(x^*) + g^*(z^*) \le 0$. Writing $y^* := x^* - z^*$, we have $y^* + z^* = x^*$ and $f^*(y^*) + g^*(z^*) \le (f+g)^*(x^*)$, which gives (7.1). ■

Corollary 7.2 is an immediate consequence of Theorem 7.1 (see [9, Theorem 3(a), p. 85] or [17, Theorem 2.8.3(iii), p. 123]):

COROLLARY 7.2 *Let E be a nontrivial Hausdorff locally convex space with dual E^*, f, $g \in \mathcal{PC}(E)$ and g be finite and continuous at a point of* $\operatorname{dom} f$. *Then* (7.1) *is true.*

The preceding two results indicate the (well-known) fact that results on the conjugate of a sum are very close to the Fenchel duality theorem. The purpose of the next example is to draw a distinction between these two results. Examples 7.3 and 7.4 were worked out in collaboration with Regina Burachik. We say that f and g *satisfy Fenchel duality* if there exists z^* such that $f^*(-z^*) + g^*(z^*) = (f+g)^*(0)$.

Example 7.3 We give an example of proper, convex lower semicontinuous functions f and g on $\mathbb{R}^2$ that satisfy Fenchel duality, but, for most $r \in (\mathbb{R}^2)^* = \mathbb{R}^2$, it is not true that there exist $p, q \in \mathbb{R}^2$ such that $p + q = r$ and $f^*(p) + g^*(q) = (f+g)^*(r)$.

Let $C = \{x \in \mathbb{R}^2\colon \|x\| \le 1\}$ and $x_0 = (1, 0) \in \mathbb{R}^2$. Write $A := C + x_0$, $B := C - x_0$, $f := \mathbb{I}_A$; and $f := \mathbb{I}_B$, where $\mathbb{I}_X$ is the indicator function of X. Since $f^*(0) = g^*(0) = (f+g)^*(0) = 0$, f and g satisfy Fenchel duality.

Now, for all $p, q \in \mathbb{R}^2$, $f^*(p) = \mathbb{I}_C^*(p) + \langle x_0, p\rangle = \|p\| + p_1$ and $g^*(q) = \mathbb{I}_C^*(q) - \langle x_0, q\rangle = \|q\| - q_1$. Consequently,

$$f^*(p) \geq 0 \quad \text{and} \quad (f^*(p) = 0 \Rightarrow p_1 \leq 0 \text{ and } p_2 = 0), \tag{7.2}$$

and

$$g^*(q) \geq 0 \quad \text{and} \quad (g^*(q) = 0 \Rightarrow q_1 \geq 0 \text{ and } q_2 = 0). \tag{7.3}$$

If $p, q \in \mathbb{R}^2$ are such that $p+q=r$ and $f^*(p) + g^*(q) = (f+g)^*(r)$, then, since $(f+g)^*(r) = 0$, (7.2) and (7.3) imply that $f^*(p)=0$ and $g^*(q)=0$, consequently $p_2=0$ and $q_2=0$, from which $r_2=0$. Thus, if $r_2 \neq 0$, then there do not exist $p, q \in \mathbb{R}^2$ such that $p+q=r$ and $f^*(p) + g^*(q) = (f+g)^*(r)$.

We can look at this another way: if f and $g-r$ satisfy Fenchel duality, then there exist $p \in \mathbb{R}^2$ such that $(f+g-r)^*(0) = f^*(p) + (g-r)^*(-p)$, that is to say, $f^*(p) + g^*(r-p) = 0$, and the preceding analysis shows that $r_2=0$. This argument can easily be reversed: if $r_2=0$ then there exists $p, q \in \mathbb{R}^2$ such that $p+q=r$ and $f^*(p) + g^*(q) = (f+g)^*(r)$, and f and $g-r$ satisfy Fenchel duality. At any rate, f and g fail "stable Fenchel–Rockafellar duality" in the sense of [2, Theorem 3.2(i)].

Example 7.4 Reference [2, Theorem 3.2(ii)] tells us that $\operatorname{epi} f^* + \operatorname{epi} g^*$ is not closed in $\mathbb{R}^2 \times \mathbb{R}$ in Example 7.3. We now give an explicit description of this set. If $p_1 \leq 0 \leq q_1$, then

$$f^*(p) = \frac{\|p\|^2 - p_1^2}{\|p\| - p_1} \leq \frac{p_2^2}{2|p_1|} \text{ and } g^*(q) = \frac{\|q\|^2 - q_1^2}{\|q\| + q_1} \leq \frac{q_2^2}{2|q_1|}. \tag{7.4}$$

Let r be an arbitrary element of $\mathbb{R}^2$ and $n > |r_1|$. Then, from (7.4),

$$f^*\left(\frac{r}{2} - ne_1\right) \leq \frac{r_2^2}{4(2n - r_1)} \text{ and } g^*\left(\frac{r}{2} + ne_1\right) \leq \frac{r_2^2}{4(2n + r_1)}.$$

Thus,

$$\left(\frac{r}{2} - ne_1, \frac{r_2^2}{4(2n - r_1)}\right) \in \operatorname{epi} f^* \text{ and } \left(\frac{r}{2} + ne_1, \frac{r_2^2}{4(2n + r_1)}\right) \in \operatorname{epi} g^*,$$

and so

$$\left(r, \frac{r_2^2}{4(2n - r_1)} + \frac{r_2^2}{4(2n + r_1)}\right) \in \operatorname{epi} f^* + \operatorname{epi} g^*.$$

Since epi f^* + epi g^* recedes vertically, it follows by letting $n \to \infty$ that

$$\{(r_1, r_2, \lambda): r_2 = 0, \lambda \geq 0\} \cup \{(r_1, r_2, \lambda): r_2 \neq 0, \lambda > 0\} \subset \operatorname{epi} f^* + \operatorname{epi} g^*.$$

It is also clear from (7.2) and (7.3) that $\operatorname{epi} f^* + \operatorname{epi} g^* \subset \mathbb{R}^2 \times \mathbb{R}^+$. Suppose now that $(r, 0) \in \operatorname{epi} f^* + \operatorname{epi} g^*$. Then there exist $(p, \lambda) \in \operatorname{epi} f^*$ and $(q, \mu) \in \operatorname{epi} g^*$ such that $(p+q, \lambda+\mu) = (r,0)$. Then $0 = \lambda + \mu \geq f^*(p) + g^*(q)$ so, from (7.2) and (7.3), $f^*(p)=0$ and $g^*(q)=0$. Arguing as in Example 7.3, $r_2=0$.

Combining all this together, we have

$$\operatorname{epi} f^* + \operatorname{epi} g^* = \{(r_1, r_2, \lambda): r_2 = 0, \lambda \geq 0\} \cup \{(r_1, r_2, \lambda): r_2 \neq 0, \lambda > 0\}$$

(which is obviously not closed).

We now investigate an even more unstable case of Fenchel duality. However, the analysis is a little more technical. We shall say that proper, convex, lower semicontinuous functions f and g on a Banach space E are *totally Fenchel unstable* if f and g satisfy Fenchel duality, but, for all $x^* \in E^*\backslash\{0\}$, f and $g - x^*$ do not satisfy Fenchel duality.

Example 7.5 We recall that if C is a convex subset of a Banach space E and $x \in C$ then x is a *support point* of C if there exists $x^* \in E^*\backslash\{0\}$ such that $\langle x, x^*\rangle = \sup\langle C, x^*\rangle$. We will give an example of a nonempty closed convex subset C of a Banach space E (actually ℓ_2) such that $C = -C$, and there exists an extreme point x_0 of C that is not a support point of C. Again, write $A := C + x_0$, $B := C - x_0$, $f := \mathbb{I}_A$, and $g := \mathbb{I}_B$. As in Example 7.3, f and g satisfy Fenchel duality.

Now, for all $y^*, z^* \in E^*$, $f^*(y^*) = \sup\langle C, y^*\rangle + \langle x_0, y^*\rangle = \sup\langle -C, -y^*\rangle - \langle x_0, -y^*\rangle = \sup\langle C, -y^*\rangle - \langle x_0, -y^*\rangle \geq 0$ and $g^*(y^*) = \sup\langle C, z^*\rangle - \langle x_0, z^*\rangle \geq 0$.

Let $y^*, z^* \in E^*$ be such that $y^* + z^* = x^*$ and $f^*(y^*) + g^*(z^*) = (f+g)^*(x^*)$. The fact that x_0 is an extreme point of C implies that $f + g = \mathbb{I}_{\{0\}}$, and so $(f+g)^*(x^*) = 0$. Thus $f^*(y^*) + g^*(z^*) = 0$, from which $f^*(y^*) = 0$ and $g^*(z^*) = 0$. Consequently, $\langle x_0, -y^*\rangle = \sup\langle C, -y^*\rangle$ and $\langle x_0, z^*\rangle = \sup\langle C, z^*\rangle$. Since x_0 is not a support point of C, $-y^* = 0$ and $z^* = 0$, thus $x^* = y^* + z^* = 0$. So we have established that f and g are totally Fenchel unstable.

Here is the promised example, which was suggested by Jon Borwein. Let $E = \ell_2$, $1 < p < 2$, and $C := \{x \in \ell_2\colon \|x\|_p \leq 1\}$. Since the function $\|\cdot\|_p$ is lower semicontinuous on ℓ_2, C is closed, and obviously C is convex and $C = -C$. Then x is an extreme point of C if, and only if, $x_p = 1$.

Let $x \in C$ and $\|x\|_p = 1$. We shall prove that x *is a support point of* C *if, and only if,* $x \in \ell_{2(p-1)}$. Suppose first that x is a support point of C. Then there exists $y \in \ell_2 = (\ell_2)^*$ such that $y \neq 0$ and

$$\langle x, y\rangle = \sup\langle C, y\rangle = \|y\|_q = \|x\|_p \|y\|_q.$$

Thus, we have equality in Hölder's inequality, and so there exists $\lambda > 0$ such that, for all $n \geq 1$, $|y_n|^q = (\lambda|x_n|)^p$. Since $y \in \ell_2$, $\sum_{n\geq 1} (\lambda|x_n|)^{2p/q} < \infty$, that is to say, $x \in \ell_{2(p-1)}$, as required. Suppose, conversely, that $x \in \ell_{2(p-1)}$. For all $n \geq 1$, let $y_n = \operatorname{sgn} x_n |x_n|^{p-1}$. Then, $y \in \ell_2 = (\ell_2)^*$. Further,

$$\langle x, y\rangle = \sum_{n\geq 1} x_n y_n = \sum_{n\geq 1} x_n \operatorname{sgn} x_n |x_n|^{p-1} = \sum_{n\geq 1} |x_n|^p = 1$$

and

$$\sup\langle C, y\rangle = \|y\|_q = \left(\sum_{n\geq 1} |x_n|^{q(p-1)}\right)^{1/q} = \left(\sum_{n\geq 1} |x_n|^p\right)^{1/q} = 1^{1/q} = 1,$$

so x is a support point of C. Since $2(p-1) < p$, there are plenty of extreme points of C that are not support points.

Problem 7.6 Let E, C, f, and g be as in Example 7.5, with x_0 an extreme nonsupport point of C. Write $\{0\}^c$ for $\ell_2\backslash\{0\}$. Then, is

$$\operatorname{epi} f^* + \operatorname{epi} g^* = (\{0\} \times [0, \infty[) \cup (\{0\}^c \times]0, \infty[)$$

The inclusion "$\subset$" is clear from the discussion in Example 7.5.

8. Some properties of $\frac{1}{2}\|\cdot\|^2$

In this section, we prove a sharp version of the Fenchel duality theorem that has been very useful in the investigation of monotone multifunctions. Let E be a nontrivial normed space with dual E^*. We define $g : E \mapsto \mathbb{R}$ by $g(z) := \frac{1}{2}\|z\|^2$. Then g is continuous and convex and, for all $z^* \in E^*$, $g^*(z^*) = \frac{1}{2}\|z^*\|^2$. We will use the following curious geometric property:

LEMMA 8.1 *Let E be a nontrivial normed space with dual E^*, $x \in E$, $\lambda \in \mathbb{R}$, and $\frac{1}{2}\|x\|^2 \geq \lambda$. Then*

$$\sup_{y \in E,\, y \neq x} \frac{\lambda - \frac{1}{2}\|y\|^2}{\|x-y\|} \vee 0 = \left[\|x\| - \sqrt{\|x\|^2 - 2\lambda}\right] \vee 0.$$

Proof Since $\frac{1}{2}\|\cdot\|^2$ is continuous on E, Corollary 6.4 implies that

$$\sup_{y \in E,\, y \neq x} \frac{\lambda - \frac{1}{2}\|y\|^2}{\|x-y\|} \vee 0 = \min\left\{\|z^*\| \colon z^* \in E^*,\ \frac{1}{2}\|z^*\|^2 - \langle x, z^*\rangle + \lambda \leq 0\right\} < \infty.$$

Let $N := \left[\|x\| - \sqrt{\|x\|^2 - 2\lambda}\right] \vee 0$. So what we have to prove is

$$\min\left\{\|z^*\| \colon z^* \in E^*,\ \tfrac{1}{2}\|z^*\|^2 - \langle x, z^*\rangle + \lambda \leq 0\right\} = N. \tag{8.1}$$

Suppose first that $z^* \in E^*$ and $\frac{1}{2}\|z^*\|^2 - \langle x, z^*\rangle + \lambda \leq 0$. Then $\|z^*\|^2 - 2\|x\|\|z^*\| \leq -2\lambda$, and so, by completing the square, we obtain $\|z^*\| \geq \|x\| - \sqrt{\|x\|^2 - 2\lambda}$. Since $\|z^*\| \geq 0$, it follows that $\|z^*\| \geq N$. On the other hand, $\|x\| + \sqrt{\|x\|^2 - 2\lambda} \geq \|x\| - \sqrt{\|x\|^2 - 2\lambda}$ and $\|x\| + \sqrt{\|x\|^2 - 2\lambda} \geq 0$, consequently $\|x\| + \sqrt{\|x\|^2 - 2\lambda} \geq N \geq \|x\| - \sqrt{\|x\|^2 - 2\lambda}$. It follows from this that $|\|x\| - N| \leq \sqrt{\|x\|^2 - 2\lambda}$, and so $\|x\|^2 - 2N\|x\| + N^2 \leq \|x\|^2 - 2\lambda$. Thus, $\frac{1}{2}N^2 - N\|x\| + \lambda \leq 0$. [Corollary 3.2 or Corollary 3.3 (the *one-dimensional Hahn–Banach theorem*) now implies that there exists $x^* \in E^*$ such that $\|x^*\| \leq N$ and $\langle x, x^*\rangle = N\|x\|$. But then $\frac{1}{2}\|x^*\|^2 - \langle x, x^*\rangle + \lambda \leq N^2 - N\|x\| + \lambda \leq 0$. Since $\|x^*\| \leq N$, this completes the proof of (8.1). ■

Remark 8.2 It is worthy of note that the axiom of choice is used in the proof of both directions of the inequality in Lemma 8.1. We do not know if this is necessary.

COROLLARY 8.3 *Let E be a nontrivial normed space with dual E^*, $x \in E$, $c \in [-\infty, \infty[$, and $\frac{1}{2}\|x^2\| \geq c$. Then*

$$\sup_{y \in E,\, y \neq x} \frac{c - \frac{n}{2}\|y\|^2}{\|x-y\|} \vee 0 = \left[\|x\| - \sqrt{\|x\|^2 - 2c}\right] \vee 0.$$

Proof If $c \in \mathbb{R}$, the result follows from Lemma 8.1 while if $c = -\infty$, both sides of the equality above have the value 0. ■

We end this section with a new proof of the sharp version of the Fenchel duality theorem first established in [16, Theorem 2.1, pp. 5–6], which has important applications to the theory of maximal monotone multifunctions.

THEOREM 8.4 *Let E be a nontrivial normed space, $f \in \mathcal{PC}(E)$ and*

$$x \in E \Rightarrow f(x) + \tfrac{1}{2}\|x\|^2 \geq 0.$$

Then

$$\min\left\{\|z^*\|: z^* \in E^*, f^*(z^*) + \tfrac{1}{2}\|z^*\|^2 \leq 0\right\} = \sup_{x \in E}\left[\|x\| - \sqrt{2f(x) + \|x\|^2}\right] \vee 0.$$

Proof We known already from Corollary 6.1(c) that

$$\min\left\{\|z^*\|: z^* \in E^*, f^*(z^*) + \tfrac{1}{2}\|z^*\|^2 \leq 0\right\} = \sup_{x, y \in E, x \neq y} \frac{-f(x) - \frac{1}{2}\|y\|^2}{\|x - y\|} \vee 0.$$

The result now follows from Corollary 8.3 by putting $c = -f(x)$ and then taking the supremum over $x \in E$. ∎

9. Lagrangians

In [10], Rockafellar develops a theory of dual problems and Lagrangians that gives a very large number of results in convex analysis. We will show in Theorem 9.1(c) how Theorem 2.9 can be used to give an efficient proof of [10, Theorem 17(a) p. 41], one of the main existence results in [10]. We note, in passing, that Theorem 9.1(b) gives a *necessary and sufficient* condition for (9.1.4). (Compare [10, Theorem 16, p. 40].) We refer the reader to [17, Theorem 2.7.1, pp. 113–115] for a more complete discussion of Lagrangians.

THEOREM 9.1 *Let V be a real vector space, E be a real locally convex space with dual E^*, and $F : V \times E \mapsto [-\infty, \infty]$ be convex. (P) is the "primal problem" of finding the value of $\beta := \inf_{x \in V} F(x, 0)$, and (D) is the "dual problem" of finding $\sup_{z^* \in E^*} h(z^*)$, where $h : E^* \mapsto [-\infty, \infty]$ is defined by*

$$h(z^*) := \inf_{(x, y) \in V \times E} [F(x, y) + \langle y, z^* \rangle].$$

Finally, let $\alpha \in [-\infty, \infty]$. We consider the four conditions:

$$\text{there exists } z^* \in E^* \text{ such that } \inf_{(x, y) \in V \times E}[F(x, y) + \langle y, z^* \rangle] \geq \alpha, \tag{9.1}$$

$$\text{there exists } S \in \mathcal{S}(E) \text{ such that } \inf_{(x, y) \in V \times E}[F(x, y) + S(y)] \geq \alpha, \tag{9.2}$$

$$\text{there exists } S \in \mathcal{S}(E) \text{ such that } \inf_{(x, y) \in V \times E}[F(x, y) + S(y)] \geq \beta, \tag{9.3}$$

and

$$\inf(P) = \sup(D) \text{ and there exists } z^* \in E^* \text{ solving } (D). \tag{9.4}$$

Then

(a) (9.1) ⇔ (9.2).
(b) (9.3) ⇔ (9.4).
(c) *Suppose that the optimal value function $y \mapsto \inf_{x \in V} F(x, y)$ is bounded above on a neighborhood of* 0. *Then* (9.4) *is satisfied.*

Proof (a) ($\Rightarrow$) This follows, because any element of E^* is dominated by an element of $\mathcal{S}(E)$. ($\Leftarrow$) Since (9.1) is automatic if $\alpha = -\infty$, we can and will suppose that $\alpha > -\infty$. (9.2) now implies that $F: V \times E \mapsto\ ]-\infty, \infty]$. If F is identically $+\infty$ then, again, (9.1) is automatic, and so we can suppose that $F \in \mathcal{PC}(V \times E)$. The result now follows from Theorem 2.9 with $C := V \times E$ and $j(x,y) := y$, since any linear functional on E dominated by an element of $\mathcal{S}(E)$ is continuous.

If $z^* \in E^*$ then $h(z^*) \le \inf_{x \in V}[F(x,0) + \langle 0, z^* \rangle] = \beta$, and so (b) is immediate from (a).

The assumption in (c) is that there exist $T \in \mathcal{S}(E)$ and $M \in \mathbb{R}$ such that

$$T(y) < 1 \Rightarrow \text{there exists } z \in V \text{ such that } F(z,y) \le M, \tag{9.5}$$

and we shall show that there exists $S \in \mathcal{S}(E)$ satisfying (9.3) – the result then follows from (b). It is clear by taking $y = 0$ in equation (9.5) that $\beta \le M$. If $\beta = -\infty$, then (9.3) is true with $S := 0$, and so we can and will suppose that $\beta \in \mathbb{R}$. Let $S := (M - \beta)T \in \mathcal{S}(E)$: we shall show that (9.3) is true for this value of S. So let $(x,y) \in V \times E$. Let $\lambda > T(y)$. Since $T(-y/\lambda) < 1$, (9.5) gives $z \in V$ such that $F(z, -y/\lambda) \le M$. But then, from the definition of β and the convexity of F,

$$\beta \le F\left(\frac{x + \lambda z}{1+\lambda}, 0\right) \le \frac{1}{1+\lambda} F(x,y) + \frac{\lambda}{1+\lambda} F\left(z, -\frac{y}{\lambda}\right) \le \frac{1}{1+\lambda} F(x,y) + \frac{\lambda}{1+\lambda} M.$$

Clearing of fractions, we obtain $F(x,y) + (M - \beta)\lambda \ge \beta$. If we now let $\lambda \to T(y)+$ in this, we derive $F(x,y) + (M - \beta)T(y) \ge \beta$, that is to say, $F(x,y) + S(y) \ge \beta$. Since this holds for all $(x,y) \in V \times E$, we obtain (9.3), completing the proof of (c). ■

10. A sharp result on the existence of Lagrange multipliers

This section is about Lagrange multipliers for the constrained convex optimization problem outlined in the following text. The main result is Theorem 10.1, which gives a necessary and sufficient condition for the existence of a Lagrange multiplier, with a sharp lower bound on its norm. We also show in Theorem 10.3 how Theorem 10.1 implies the classical "Slater condition" result, with a related upper bound on the norm as a bonus. The analysis in this section depends only on Theorem 2.9 – it does not depend on the intervening sections in any way.

Let $(E, \|\cdot\|)$ be a nontrivial normed space, C be a nonempty convex subset of a vector space, $k: C \mapsto \mathbb{R}$ be convex, $j: C \mapsto E$, and $\preceq$ be a partial ordering on E compatible with its vector space structure. Let N be the negative cone $\{y \in E: y \preceq 0\}$. Suppose that j is *convex with respect* to $\preceq$, that is to say,

$$x_1, x_2 \in C, \mu_1, \mu_2 > 0, \text{ and } \mu_1 + \mu_2 = 1 \Rightarrow j(\mu_1 x_1 + \mu_2 x_2) \preceq \mu_1 j(x_1) + \mu_2 j(x_2), \tag{10.1}$$

and

$$\inf_{j^{-1}N} k = \inf\{k(x): x \in C, j(x) \preceq 0\} = \mu \in \mathbb{R}. \tag{10.2}$$

A *Lagrange multiplier* for the problem is an element z^* of E^* such that z^* is *positive with respect to* $\preceq$, that is to say,

$$\sup_N z^* \le 0 \tag{10.3}$$

and

$$\inf_{x \in C} [\langle j(x), z^* \rangle + k(x)] = \mu. \tag{10.4}$$

Clearly 0 is a Lagrange multiplier $\Leftrightarrow \inf_C k \geq \mu$. In order to exclude this trivial case, we shall suppose that $\inf_C k < \mu$. Let $A := \{x \in C: k(x) < \mu\}$, so that $A \neq \emptyset$.

THEOREM 10.1

(a) *There exists a Lagrange multiplier if, and only if*

$$\text{there exists } M \geq 0 \text{ such that } \inf_{x \in C} [M \operatorname{dist}(j(x), N) + k(x)] \geq \mu. \tag{10.5}$$

(b) *If z^* is a Lagrange multiplier then $x \in A \Rightarrow \operatorname{dist}(j(x), N) > 0$ and*

$$0 < \sup_{x \in A} \frac{\mu - k(x)}{\operatorname{dist}(j(x), N)} \leq \|z^*\| < \infty.$$

(c) *If*

$$0 < M := \sup_{x \in A} \frac{\mu - k(x)}{\operatorname{dist}(j(x), N)} < \infty,$$

then

$$\min\left\{\|z^*\|: z^* \text{ is a Lagrange multiplier}\right\} = M.$$

Proof

(a) Suppose first that z^* is a Lagrange multiplier. Let $x \in A$, and u be an arbitrary element of N. Then, from (10.3) and (10.4),

$$\left\| j(x) - u \right\| \|z^*\| \geq \langle j(x), z^* \rangle - \langle u, z^* \rangle \geq \langle j(x), z^* \rangle \geq \mu - k(x).$$

Taking the infimum over $u \in N$,

$$\operatorname{dist}(j(x), N)\|z^*\| + k(x) \geq \mu.$$

Since $k \geq \mu$ on $C \backslash A$, this remains true if $x \in C$. Thus (10.5) is true with $M := \|z^*\|$.

Suppose, conversely, that (10.5) is satisfied. Let $S: E \mapsto [0, \infty[$ be defined by $S(y) := M\operatorname{dist}(y, N)$ $(y \in E)$. It is easily checked that S is sublinear,

$$S \leq M\|\cdot\| \text{ on } E \text{ and } y \in N \Rightarrow S(y) = 0. \tag{10.6}$$

From (10.5),

$$\inf_C [S \circ j + k] \geq \mu.$$

Furthermore, (10.6) implies that

$$y_1 \preceq y_2 \Leftrightarrow y_1 - y_2 \in N \Rightarrow S(y_1 - y_2) = 0 \Rightarrow y_1 \leq_S y_2.$$

Thus, from Definition 2.7 and (10.1), j is S-convex, and Theorem 2.9 gives a linear functional L on E such that $L \leq S$ on E and

$$\inf_C [L \circ j + k] \geq \mu. \tag{10.7}$$

We now derive from (10.6) that $L \in E^*$, $\|L\| \le M$, and $\sup_N L \le 0$. As a consequence of this,

$$x \in j^{-1}(N) \Rightarrow j(x) \in N \Rightarrow L \circ j(x) \le 0,$$

and (10.2) now gives

$$\mu = \inf_{j^- N} k \ge \inf_{j^{-1} N}[L \circ j + k] \ge \inf_C[L \circ j + k].$$

Thus, we have equality in (10.7), and L is a Lagrange multiplier, completing the proof of (a).

(b) Let z^* be a Lagrange multiplier. Then the proof of (a) ($\Rightarrow$) implies that, for all $x \in A$, $\text{dist}(j(x), N)\|z^*\| \ge \mu - k(x) > 0$. (b) follows by dividing by dist($j(x)$, N) and taking the supremum over $x \in A$.

(c) The conditions imply that $\inf_{x \in C}[M \text{ dist}(j(x), N) + k(x)] \ge \mu$ and then the proof of (a) ($\Leftarrow$) gives a Lagrange multiplier L such that $\|L\| \le M$. (c) follows by combining this with (b). ■

Remark 10.2 At this point, we make some comments about the formulation of the preceding analysis in terms of *Lagrangians*. Let $\mathcal{P} := \{x^* \in E^*\colon \sup_N x^* \le 0\}$, and define $F : C \times \mathcal{P} \mapsto \mathbb{R}$ by $F(x, x^*) := \langle j(x), x^* \rangle + k(x)$. Then z^* is a Lagrange multiplier exactly when $\inf_{x \in C} F(x, z^*) = \mu$. If this is the case, then, arguing as in the final few lines of Theorem 10.1(c), if $x^* \in \mathcal{P}$, then $\inf_{x \in C} F(x, x^*) \le \mu$, and so in fact

$$\sup_{x^* \in \mathcal{P}} \inf_{x \in C} F(x, x^*) = \inf_{x \in C} F(x, z^*) = \mu.$$

In the event that there exists $z \in j^{-1}N$ such that $k(z) = \mu$, then (z, z^*) is a saddle point of F. See [7, Corollary 8.3.1, p. 219] for details of the argument.

We now write $j(v) \prec 0$ to mean that $j(v) \in \text{int}\, N$, and $B := \{v \in C\colon j(v) \prec 0\}$. The classical sufficient condition for the existence of Lagrange multipliers is that $B \ne \emptyset$ [7, Theorem 8.3.1, p. 217–218]. This will be improved in Theorem 10.3.

THEOREM 10.3 *Suppose that* $B \ne \emptyset$. *Then there exists a Lagrange multiplier* z* *such that*

$$\|z^*\| \le \inf_{v \in B} \frac{k(v) - \mu}{\text{dist}(j(v), E \backslash N)}.$$

Proof Let $x \in A$, $u \in N$, $v \in B$, $0 < \eta < \text{dist}(j(v), E \backslash N)$ and $\alpha := \|j(x) - u\|$. We first prove that

$$j\left(\frac{\eta x + \alpha v}{\eta + \alpha}\right) \preceq 0. \tag{10.8}$$

If $\alpha = 0$ then $j(x) = u$, and so (10.8) is immediate since $j((\eta x + \alpha v)/(\eta + \alpha)) = j(x) = u \preceq 0$. If $\alpha > 0$, then

$$\left\|\frac{\eta}{\alpha}(j(x) - u)\right\| = \eta < \text{dist}(j(v), E \backslash N)$$

and so $(\eta/\alpha)(j(x)-u)+j(v)\in N,$ from which $\eta j(x)+\alpha j(v)\preceq \eta u\preceq 0$. (10.1) now gives

$$j\left(\frac{\eta x+\alpha v}{\eta+\alpha}\right)\preceq\frac{\eta j(x)+\alpha j(v)}{\eta+\alpha}\preceq 0,$$

which completes the proof of (10.8). It now follows from the convexity of k and equation (10.2) that

$$\frac{\eta k(x)+\alpha k(v)}{\eta+\alpha}\geq k\left(\frac{\eta x+\alpha v}{\eta+\alpha}\right)\geq\mu,$$

from which $\alpha(k(v)-\mu)\geq\eta(\mu-k(x))$. If we now let $\eta\to\operatorname{dist}(j(v),E\backslash N)$ and then take the infimum over $u\in N,$ we obtain that

$$\operatorname{dist}(j(x),N)(k(v)-\mu)\geq\operatorname{dist}(j(v),E\backslash N)(\mu-k(x))>0,$$

from which $\operatorname{dist}(j(x),N)>0$, $k(v)-\mu>0$ and

$$0<\frac{\mu-k(x)}{\operatorname{dist}(j(x),N)}\leq\frac{k(v)-\mu}{\operatorname{dist}(j(v),E\backslash N)}.$$

Taking the supremum over $x\in A$ and the infimum over $v\in B$,

$$0<\sup_{x\in A}\frac{\mu-k(x)}{\operatorname{dist}(j(x),N)}\leq\inf_{v\in B}\frac{k(v)-\mu}{\operatorname{dist}(j(v),E\backslash N)},$$

and the result now follows from Theorem 10.1(c). ■

References

[1] Attouch, H. and Brezis, H., 1986, Duality for the sum of convex functions in general Banach spaces. In: J.A. Barroso (Ed.) *Aspects of Mathematics and its Applications* (Amsterdam: Elsevier Science Publishers), pp. 125–133.
[2] Burachik, R.S., Jeyakumar, V. and Wu, Z.-Y., 2006, Necessary and sufficient conditions for stable conjugate duality. *Nonlinear Analysis*, **64**, 1998–2006.
[3] Fan, K., 1953, Minimax theorems. *Proceedings of the National Academy of Sciences of the United States of America*, **39**, 42–47.
[4] Fan, K., Glicksberg, I. and Hoffman, A.J., 1957, Systems of inequalities involving convex functions. *Proceedings of the American Mathematical Society*, **8**, 617–622.
[5] Kelley, J.L., Namioka, I. and coauthors, *Linear Topological Spaces* (Princeton–Toronto–London–Melbourne: D. Van Nostrand Co., Inc).
[6] König, H., 1982, Some basic theorems in convex analysis. In: B. Korte (Ed.) *Optimization and Operations Research* (Amsterdam, New York: North-Holland).
[7] Luenberger, D.L., 1969, *Optimization by Vector Space Methods* (New York–Chichester–Brisbane–Toronto–Singapore: John Wiley & Sons, Inc).
[8] Moreau, J.-J., 1966, Fonctionelles convexes. *Séminaire sur les Equations aux Dérivées Partielles*, Lecture notes (Paris: Collèges de France).
[9] Rockafellar, R.T., 1966, Extension of Fenchel's duality theorem for convex functions. *Duke Mathematical Journal*, **33**, 81–89.
[10] Rockafellar, R.T., 1974, *Conjugate Duality and Optimization*, Conference Board of the Mathematical Sciences, Vol. 16 (Philadelphia: SIAM publications).
[11] Rudin, W., 1973, *Fuctional Analysis* (New York: McGraw-Hill).
[12] Simons, S., 1970, Minimal sublinear functionals. *Studia Mathematica*, **37**, 37–56.

[13] Simons, S., 1998, *Minimax and Monotonicity*, Lecture Notes in Mathematics, Vol. 1693 (Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, Barcelona, Budapest: Springer-Verlag).
[14] Simons, S., 2003, A new version of the Hahn–Banach theorem. *Archives of Mathematics*, **80**, 630–646.
[15] Simons, S., 2004, Hahn–Banach theorems and maximal monotonicity. In: F. Giannessi and A. Maugeri (Eds) *Variational Analysis and Applications* (Dordrecht: Kluwer Acad Publ.).
[16] Simons, S. and Zălinescu, C., 2005, Fenchel duality, Fitzpatrick functions and maximal monotonicity. *Journal of Nonlinear and Convex Analysis*, **6**, 1–22.
[17] Zălinescu, C., 2002, *Convex Analysis in General Vector Spaces* (New Jersey, London, Singapore, Hong Kong: World Scientific).