A new version of the Hahn–Banach theorem (continued)

S. Simons

1. Introduction

This is an extract from a paper titled "Hahn–Banach theorems and maximal monotonicity" that will appear in the volume "Variational analysis and Applications" edited by F. Giannessi and A. Maugeri. In it, we discuss new versions of the Hahn–Banach theorem that have a number of applications in different fields of analysis. We shall give applications to linear and nonlinear functional analysis, and convex analysis. All vector spaces in this paper will be *real*.

The main result appears in Theorem 2.8, which is bootstrapped from the special case contained in Lemma 2.4.

In Section 3, we sketch how Theorem 2.8 can be used to give the main existence theorems for linear functionals in functional analysis, and also how it gives a result that leads to a minimax theorem. We also discuss three applications of Theorem 2.8 to convex analysis, pointing the reader to [17] for further details in two of these cases. One noteworthy property of proofs using Theorem 2.8 is that they allow us to avoid the problem of the "vertical hyperplane".

In Section 4, we show how Theorem 2.8 can be used to obtain considerable insight on the existence of Lagrange multipliers for constrained convex minimization problems. The usual *sufficient* condition for the existence of such multipliers is normally found using the Eidelheit separation theorem. In Theorem 4.5, we use Theorem 2.8 to derive this sufficient condition, with the added bonus that we obtain a bound on the norm of the multiplier. Here again, the proof using Theorem 2.8 allows us to avoid the problem of the "vertical hyperplane". More to the point, the results leading up to Theorem 4.5, namely Lemma 4.1 and Theorem 4.2, use Theorem 2.8 to obtain a *necessary and sufficient* condition for the existence of Lagrange multipliers, with a *sharp lower bound* on the norm of the multiplier.

Section 5 is motivated by the theory of monotone multifunctions. Theorem 5.1 is an existence theorem without any *a priori* scalar bounds in normed spaces that has proved very useful in the investigation of these multifunctions. A new feature of the result as presented here is a sharp lower bound on the norm of the linear functional obtained.

In the final section, we return to our consideration of abstract Hahn–Banach theorems. Noting a certain formal similarity between the statements of Theorem 5.1 and Theorem 2.8, we ask the question whether these two results can be unified. Indeed, they have a common generalization, which is given in Theorem 6.1.

2. The main result

Theorem 2.8 contains the new version of the Hahn-Banach theorem that forms the main topic of this paper. Theorem 2.8 is proved by bootstrapping from the special case contained in Lemma 2.4 — most of the work is actually done in Lemma 2.3.

We start by recalling in Lemma 2.2 the classical Hahn–Banach theorem for sublinear functionals.

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

$$x, y \in E \quad \Longrightarrow \quad S(x+y) \le S(x) + S(y)$$

and

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

Lemma 2.2. *Let E be a nontrivial vector space and $S\colon\ E \mapsto \mathbb{R}$ be sublinear. Then there exists a linear functional L on E such that $L \le S$ on E.*

Proof. See Kelly–Namioka, [6, 3.4, p. 21] for a proof using cones, Rudin, [12, Theorem 3.2, p. 56–57] for a proof using an extension by subspaces argument, and König, [7] and Simons, [13] for a proof using an ordering on sublinear functionals. ▌

Lemma 2.3. *Let E be a nontrivial vector space and $S\colon\ E \mapsto \mathbb{R}$ be sublinear. Let D be a nonempty convex subset of a vector space, $a\colon\ D \mapsto E$ be affine and $\beta := \inf_D S \circ a \in \mathbb{R}$. For all $x \in E$, let*

$$T(x) := \inf_{d \in D,\ \lambda > 0} \big[S\big(x + \lambda a(d)\big) - \lambda\beta\big]. \tag{2.3.1}$$

Then $T\colon\ E \mapsto \mathbb{R}$, T is sublinear, $T \le S$ on E and, for all $d \in D$, $-T\big(-a(d)\big) \ge \beta$.

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

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

Taking the infimum over $d \in D$ and $\lambda > 0$, $T(x) \ge -S(-x) > -\infty$. Thus $T\colon\ E \mapsto \mathbb{R}$. It is now easy to check that T is positively homogeneous, 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 $\mu_1 a(d_1) + \mu_2 a(d_2) = a(d)$,

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

Taking the infimum over d_1, d_2, λ_1 and λ_2 gives $T(x_1) + T(x_2) \ge T(x_1 + x_2)$. Thus T is subadditive, and consequently, sublinear. Fix $d \in D$. Let x be an arbitrary element of E. Then, for all $\lambda > 0$, $T(x) \le S(x) + \lambda\big[S\big(a(d)\big) - \beta\big]$. Letting $\lambda \to 0$, $T(x) \le S(x)$. Thus $T \le S$ on E. Finally, let d be an arbitrary element of D. Then, taking $\lambda = 1$ in (2.3.1),

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

hence $-T\big(-a(d)\big) \ge \beta$, which completes the proof of Lemma 2.3. ▌

Lemma 2.4. *Let E be a nontrivial vector space and $S\colon E \mapsto \mathbb{R}$ be sublinear. Let D be a nonempty convex subset of a vector space and $a\colon D \mapsto E$ 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 2.2 (take any linear functional L on E such that $L \le S$ on E). So we can suppose that $\beta \in \mathbb{R}$. Define T as in Lemma 2.3. From Lemma 2.2, 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

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

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, $\inf_D L \circ a \le \inf_D S \circ a$. ▮

Definition 2.5. Let C be a nonempty convex subset of a vector space and $\mathcal{PC}(C)$ stand for the set of all convex functions $k\colon C \mapsto (-\infty, \infty]$ such that $\operatorname{dom} k \ne \emptyset$, where $\operatorname{dom} k$, the *effective domain* of k, is defined by

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

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

Definition 2.6. Let E be a nontrivial vector space and $S\colon E \mapsto \mathbb{R}$ be sublinear. Let C be a nonempty convex subset of a vector space and $j\colon C \mapsto E$. We say that j is *S–convex* if

$$x_1, x_2 \in C,\ \mu_1, \mu_2 > 0 \text{ and } \mu_1 + \mu_2 = 1 \quad\Longrightarrow\quad S\big(j(\mu_1 x_1 + \mu_2 x_2) - \mu_1 j(x_1) - \mu_2 j(x_2)\big) \le 0.$$

Note that if we define an ordering "$\le_S$" on E by declaring that $y \le_S z$ if $S(y - z) \le 0$ then j is S–convex if, and only if,

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

An affine function is clearly S–convex.

Remark 2.7. Suppose that C_S is the level set $\{y \in E\colon\ S(y) \le 0\}$. It is clear that the ordering $\le_S$ on E is determined solely by C_S (though the proof of Theorem 2.8 depends on the other values of S). Now let us consider the special case when $E = \mathbb{R}$. Since C_S is a convex cone vertex the origin, there are exactly four possibilities for C_S, namely $\{0\}$, $(-\infty, 0]$, $[0, \infty)$ and $\mathbb{R}$. These can be realized by $S(y) := |y|$, $S(y) := y$, $S(y) := -y$ and $S(y) := 0$, respectively. In these four cases, "S–convex" means "affine", "convex", "concave" and "arbitrary", repectively. In general, when $E \ne \mathbb{R}$, there is no analog of convex or concave function from C into E, and it makes sense to ask the question when a function $j\colon C \mapsto E$ is S–convex with respect to some nontrivial sublinear functional S on E. A solution to this problem has been provided by Giandomenico Mastroeni (personal communication).

Theorem 2.8. *Let E be a nontrivial vector space and $S\colon E \mapsto \mathbb{R}$ be sublinear. Let C be a nonempty convex subset of a vector space, $k \in \mathcal{PC}(C)$ and $j\colon C \mapsto E$ be S–convex. Then there exists a linear functional L on E such that $L \le S$ on E and*

$$\inf_C \big[L \circ j + k\big] = \inf_C \big[S \circ j + k\big]. \tag{2.8.1}$$

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

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

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

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

and $a\colon D \mapsto \widetilde{E}$ be defined by

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

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

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

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

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

The result follows since, by direct computation,

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

3. Applications to functional analysis and minimax theorems

In this section, we mention without proof a number of applications of Theorem 2.8 that were discussed in [17]. We then state and prove in Theorem 3.5 a (necessary and sufficient) criterion for the Fenchel duality condition to hold.

Theorem 3.1 is the *sandwich theorem* (see [7, Theorem 1.7, p. 112]). It follows immediately from Theorem 2.8 with $C := E$ and $j(x) := x$.

Theorem 3.1. *Let E be a nontrivial vector space, $S\colon E \mapsto \mathbb{R}$ 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.*

Theorem 3.1 implies in turn two other well known existence results: the *extension form of the Hahn–Banach theorem*, Corollary 3.2, (see [7, Corollary 1.8, p. 112]) and the *Mazur–Orlicz theorem*, Corollary 3.3, (see [7, Theorem 1.9, p. 112]).

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

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

Theorem 3.4 below was essentially proved by Fan–Glicksberg–Hoffman (see [5, Theorem 1, p. 618]), and leads to a short proof of the minimax theorem proved by Fan in [4] (see [14, Theorem 3.1, p. 17] for details of this). Theorem 3.4 follows easily from Theorem 2.8 with $E := \mathbb{R}^m$, $S(\mu_1, \ldots, \mu_m) := \mu_1 \vee \cdots \vee \mu_m$, $j(c) := \big(f_1(c), \ldots, f_m(c)\big)$ and $k(c) := 0$.

Theorem 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 $\lambda_1, \ldots, \lambda_m \ge 0$ such that $\lambda_1 + \cdots + \lambda_m = 1$ and*

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

Let E be a nontrivial Hausdorff locally convex space with dual E^*. If $f \in \mathcal{PC}(E)$, the *Fenchel conjugate*, f^*, of f is the function from E^* into $(-\infty, \infty]$ defined by

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

It follows easily from the definitions above that, for all $y \in E$,

$$f(y) \ge \sup_{E^*} (y - f^*). \tag{3.4.1}$$

It was proved by Moreau in [9, Section 5–6, p. 26–39] that if f is lower semicontinuous on E then, for all $y \in E$, we have equality in (3.4.1). If f is lower semicontinuous at $y \in E$ but not on E then it does *not* follow that equality holds in (3.4.1) $\big($see [17, Remark 3.1]$\big)$. On the other hand, Theorem 2.8 can be used to find a necessary and sufficient condition for equality to hold in (3.4.1) for a given $y \in E$ $\big($see [17, Theorem 3.2]$\big)$. This provides a proof of Moreau's original result with the advantage that we do not have to deal with the elimination of the "vertical hyperplane".

We now show how Theorem 2.8 leads to a version of the Fenchel duality theorem.

Theorem 3.5. *Let E be a nontrivial Hausdorff locally convex space with dual E^*, and $f, g \in \mathcal{PC}(E)$. Then*

$$\text{there exists } z^* \in E^* \text{ such that} \qquad f^*(-z^*) + g^*(z^*) \le 0 \tag{3.5.1}$$

if, and only if, writing $\mathcal{S}(E)$ for the family of continuous seminorms on E,

$$\text{there exists } S \in \mathcal{S}(E) \text{ such that} \qquad x,\, y \in E \Longrightarrow f(x) + g(y) + S(x-y) \geq 0. \tag{3.5.2}$$

Proof. Suppose first that (3.5.1) is satisfied. Then, for all $x,\, y \in E$,

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

consequently,

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

and (3.5.2) follows with $S := |z^*|$. Suppose, conversely, that (3.5.2) is satisfied. Then we apply Theorem 2.8 with $C := E \times E$, $j(x,y) := x - y$ and $k(x,y) := f(x) + g(y)$ and obtain a linear functional L on E such that $L \leq S$ and

$$x,\, y \in E \Longrightarrow f(x) + g(y) + L(x-y) \geq 0,$$

or equivalently,

$$x,\, y \in E \Longrightarrow (-L)(x) - f(x) + L(y) - g(y) \leq 0.$$

(3.5.1) now follows (with $z^* = L$) by taking the supremum over x and y. ∎

In the normed case, Theorem 3.5 takes the following form:

Corollary 3.6. *Let E be a nontrivial normed space with dual E^*, and $f,\, g \in \mathcal{PC}(E)$. Then*

$$\text{there exists } z^* \in E^* \text{ such that} \qquad f^*(-z^*) + g^*(z^*) \leq 0 \tag{3.5.1}$$

if, and only if,

$$\text{there exists } M \geq 0 \text{ such that} \qquad x,\, y \in E \Longrightarrow f(x) + g(y) + M\|x - y\| \geq 0.$$

Corollary 3.6 leads easily to proofs of the versions of the Fenchel duality theorem and the formula for the subdifferential of a sum due to Moreau–Rockafellar (see [10, Theorem 3, p. 85]) and Attouch–Brézis (see [1, Theorem 1.1, p. 126–127] and [1, Corollary 2.1, p. 130–131]). Yet again, we do not have to deal with the elimination of the "vertical hyperplane". We emphasize that Theorem 3.5 and Corollary 3.6 give a *necessary and sufficient* condition for the existence of the linear functional, and not merely *sufficient* conditions.

In [11], Rockafellar develops a theory of dual problems and Lagrangians that gives a very large number of results in convex analysis. It was shown in [17, Theorem 3.6] how Theorem 2.8 can be used to give an efficient proof of [11, Theorem 17(a), p. 41], one of the main existence results in [11].

4. A sharp result on the existence of Lagrange multipliers

This section is about Lagrange multipliers for the constrained convex optimization problem outlined below. The main result is Theorem 4.2 which, combined with Lemma 4.1, 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 4.5 how Theorem 4.2 implies the classical result, with an upper bound on the norm as a bonus. The analysis in this section depends only on Theorem 2.8 — it does not depend on Section 3 in any way.

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

$$x_1, x_2 \in C,\ \mu_1, \mu_2 > 0 \text{ and } \mu_1 + \mu_2 = 1 \implies j(\mu_1 x_1 + \mu_2 x_2) \preceq \mu_1 j(x_1) + \mu_2 j(x_2) \quad (4.0.1)$$

(i.e., j is *convex* with respect to $\preceq$), and

$$\inf_{j^{-1}N} k = \inf\big\{k(x)\colon\ x \in C,\ j(x) \preceq 0\big\} = \mu_0 \in \mathbb{R}. \quad (4.0.2)$$

A *Lagrange multiplier* for the problem is an element z_0^* of E^* such that

$$\sup_N z_0^* \le 0 \quad (4.0.3)$$

(i.e., z_0^* is *positive* with respect to $\preceq$), and

$$\inf_{x \in C} \big[\langle j(x), z_0^*\rangle + k(x)\big] = \mu_0. \quad (4.0.4)$$

Clearly 0 is a Lagrange multiplier $\iff$ $\inf_C k \ge \mu_0$. In order to exclude this trivial case, we shall suppose that $\inf_C k < \mu_0$. Let

$$A := \big\{x \in C\colon\ k(x) < \mu_0\big\} \quad\text{and}\quad B := \big\{v \in C\colon\ j(v) \prec 0\big\}, \quad (4.0.5)$$

where we write $j(v) \prec 0$ to mean that $j(v) \in \operatorname{int} N$. The above conditions imply that $A \ne \emptyset$. We start off with a simple consequence of the existence of a Lagrange multiplier.

Lemma 4.1. *Let z_0^* be a Lagrange multiplier, and A be as in (4.0.5). Then*

$$0 < \sup_{x \in A} \frac{\mu_0 - k(x)}{\operatorname{dist}\big(j(x), N\big)} \le \|z_0^*\| < \infty.$$

Proof. Let $x \in A$, and u be an arbitrary element of N. Then, from (4.0.3) and (4.0.4),

$$\|j(x) - u\|\,\|z_0^*\| \ge \langle j(x), z_0^*\rangle - \langle u, z_0^*\rangle \ge \langle j(x), z_0^*\rangle \ge \mu_0 - k(x) > 0.$$

Taking the infimum over $u \in N$,

$$\operatorname{dist}\big(j(x), N\big)\|z_0^*\| \ge \mu_0 - k(x) > 0.$$

The result follows on division by $\operatorname{dist}\big(j(x), N\big)$ and then taking the supremum over $x \in A$. ∎

The main result of this section is the following partial converse to Lemma 4.1.

Theorem 4.2. *Suppose that* $0 < M := \sup_{x\in A} \dfrac{\mu_0 - k(x)}{\operatorname{dist}\big(j(x), N\big)} < \infty$. *Then there exists a Lagrange multiplier* z_0^* *such that* $\|z_0^*\| \le M$. *It then follows from Lemma 4.1 that* $M = \min\big\{\|z_0^*\|\colon\ z_0^*$ *is a Lagrange multiplier*$\big\}$.

Proof. Let $S\colon\ E \mapsto [0,\infty)$ be defined by $S(y) := \operatorname{dist}(y, N) = \inf_{u\in N} \|y-u\| \quad (y \in E)$. It is easily checked from this definition that

$$S \text{ is sublinear,} \tag{4.2.1}$$

$$S \le \|\cdot\| \text{ on } E, \tag{4.2.2}$$

and

$$y \in N \quad\Longrightarrow\quad S(y) = 0. \tag{4.2.3}$$

The definition of M gives

$$x \in A \Longrightarrow MS\circ j(x) + k(x) \ge \mu_0.$$

Since $k \ge \mu_0$ on $C \setminus A$ and $S \ge 0$ on E, in fact

$$x \in C \Longrightarrow MS\circ j(x) + k(x) \ge \mu_0,$$

that is to say

$$\inf_C \big[MS\circ j + k\big] \ge \mu_0.$$

Let $x_1, x_2 \in C,\ \mu_1, \mu_2 > 0$ and $\mu_1 + \mu_2 = 1$. Then it follows from (4.0.1) that

$$j(\mu_1 x_1 + \mu_2 x_2) - \mu_1 j(x_1) - \mu_2 j(x_2) \in N,$$

and so (4.2.3) implies that j is MS–convex. Thus (4.2.1) and Theorem 2.8 give a linear functional L on E such that $L \le MS$ on E and

$$\inf_C \big[L\circ j + k\big] = \inf_C \big[MS\circ j + k\big] \ge \mu_0. \tag{4.2.4}$$

We now derive from (4.2.2) and (4.2.3) that $L \in E^*$, $\|L\| \le M$ and $\sup_N L \le 0$. Since $x \in j^{-1}(N) \Longrightarrow j(x) \in N \Longrightarrow L\circ j(x) \le 0$, (4.0.2) now gives

$$\mu_0 = \inf_{j^{-1}N} k \ge \inf_{j^{-1}N} \big[L\circ j + k\big] \ge \inf_C \big[L\circ j + k\big].$$

Thus we have equality in (4.2.4), which gives the required result (with $z_0^* = L$). ∎

Remark 4.3. At this point, we make some comments about the formulation of the preceding analysis in terms of *Lagrangians*. Let $\mathcal{P} := \{z^* \in E^*\colon\ \sup_N z^* \le 0\}$, and define $L\colon\ C \times \mathcal{P} \mapsto \mathbb{R}$ by $L(x, z^*) := \big\langle j(x), z^*\big\rangle + k(x)$. Then z_0^* is a Lagrange multiplier exactly when $\inf_{x\in C} L(x, z_0^*) = \mu_0$. Arguing as in the final few lines of Theorem 4.2, if $z^* \in \mathcal{P}$ then $\inf_{x\in C} L(x, z^*) \le \mu_0$, so in fact

$$\sup_{z^*\in\mathcal{P}} \inf_{x\in C} L(x, z^*) = \inf_{x\in C} L(x, z_0^*) = \mu_0.$$

In the event that there exists $x_0 \in j^{-1}N$ such that $k(x_0) = \mu_0$ then (x_0, z_0^*) is a saddle point of L. See [8, Corollary 8.3.1, p. 219] for details of the argument.

We recall from (4.0.5) that $B := \{v \in C\colon\ j(v) \prec 0\}$. The classical sufficient condition for the existence of Lagrange multipliers is that $B \neq \emptyset$. (See [8, Theorem 8.3.1, p. 217–218].) This will be improved in Theorem 4.5. We first give a preliminary lemma.

Lemma 4.4.
(a) Let $x \in A$, $u \in N$, $v \in B$, $0 < \eta < \operatorname{dist}\big(j(v), E \setminus N\big)$ and $\alpha := \|j(x) - u\|$. Then

$$j\Big(\frac{\eta x + \alpha v}{\eta + \alpha}\Big) \preceq 0.$$

(b) Let $x \in A$ and $v \in B$. Then

$$\operatorname{dist}\big(j(x), N\big)\big(k(v) - \mu_0\big) \geq \operatorname{dist}\big(j(v), E \setminus N\big)\big(\mu_0 - k(x)\big) > 0.$$

Proof. (a) If $\alpha = 0$ then $j(x) = u$ and so

$$j\Big(\frac{\eta x + \alpha v}{\eta + \alpha}\Big) = j(x) = u \preceq 0,$$

which gives the required result. If $\alpha > 0$ then

$$\Big\|\frac{\eta}{\alpha}\big(j(x) - u\big)\Big\| = \eta < \operatorname{dist}\big(j(v), E \setminus N\big)$$

and so

$$\frac{\eta}{\alpha}\big(j(x) - u\big) + j(v) \in N,$$

from which

$$\eta j(x) + \alpha j(v) \preceq \eta u \preceq 0.$$

(4.0.1) now gives

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

which completes the proof of (a).
(b) Let $u \in N$ and α and η be as in (a). Using (a), the convexity of k and (4.0.2), we obtain

$$\frac{\eta k(x) + \alpha k(v)}{\eta + \alpha} \geq k\Big(\frac{\eta x + \alpha v}{\eta + \alpha}\Big) \geq \mu_0,$$

from which $\alpha\big(k(v) - \mu_0\big) \geq \eta\big(\mu_0 - k(x)\big)$. If we now let $\eta \to \operatorname{dist}\big(j(v), E \setminus N\big)$ and then take the infimum over $u \in N$, we obtain that

$$\operatorname{dist}\big(j(x), N\big)\big(k(v) - \mu_0\big) \geq \operatorname{dist}\big(j(v), E \setminus N\big)\big(\mu_0 - k(x)\big),$$

and (b) follows from (4.0.5). ▌

Theorem 4.5. *Suppose that $B \neq \emptyset$. Then there exists a Lagrange multiplier z_0^* such that*

$$\|z_0^*\| \le \inf_{v \in B} \frac{k(v) - \mu_0}{\operatorname{dist}\bigl(j(v), E \setminus N\bigr)}.$$

Proof. Let $x \in A$ and $v \in B$. From Lemma 4.4(b), $\operatorname{dist}\bigl(j(x), N\bigr) > 0$ and

$$\frac{\mu_0 - k(x)}{\operatorname{dist}\bigl(j(x), N\bigr)} \le \frac{k(v) - \mu_0}{\operatorname{dist}\bigl(j(v), E \setminus N\bigr)}.$$

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

$$\sup_{x \in A} \frac{\mu_0 - k(x)}{\operatorname{dist}\bigl(j(x), N\bigr)} \le \inf_{v \in B} \frac{k(v) - \mu_0}{\operatorname{dist}\bigl(j(v), E \setminus N\bigr)}.$$

The result now follows from Theorem 4.2. ▌

5. Existence theorems without *a priori* scalar bounds for normed spaces

The main result is this section is Theorem 5.1. The equivalence of (5.1.1) and (5.1.2) actually first appeared in [14, Theorem 7.2, p. 27–28], and was used in [14] to obtain a number of criteria for a monotone multifunction on a reflexive Banach space to be maximal monotone (including Rockafellar's "surjectivity theorem", to obtain conditions for the sum of maximal monotone multifunctions on a reflexive Banach space to be maximal monotone, and to obtain some results on maximal monotone multifunctions of Gossez's type (D) on an arbitrary Banach space. For more information, see the introductions to Sections 5 and 6 of [17]. This equivalence was also used in [16] to prove other results on maximal monotonicity.

The proof of the equivalence of (5.1.1) and (5.1.2) given in [14, Theorem 7.2] was quite nonconstructive, and a more constructive proof was given in [17, Theorem 5.1], together with the bound $\inf_{c \in C} \Bigl[\|j(c)\| + \sqrt{k(c) + \|j(c)\|^2}\Bigr]$ on the norm of $\|y^*\|$ (see [17, Remark 5.6]). We now give a new proof of this equivalence, which relies on the direct Dedekind section argument (5.1.6)–(5.1.7) and is much simpler than the proofs given in [14] and [17]. Furthermore, as is clear from (5.1.4), the bound $\sup_{c \in C} \Bigl[\|j(c)\| - \sqrt{k(c) + \|j(c)\|^2}\Bigr] \vee 0$ on the norm of $\|y^*\|$ found in Theorem 5.1 is sharp. The analysis in this section depends only on Theorem 2.8 — it does not depend on Sections 3–4 in any way.

Theorem 5.1. *Let C be a nonempty convex subset of a vector space, F be a nontrivial normed space, $j\colon C \mapsto F$ be affine and $k \in \mathcal{PC}(C)$. Then*

$$c \in C \quad \Longrightarrow \quad k(c) + \|j(c)\|^2 \ge 0 \tag{5.1.1}$$

if, and only if,

$$\text{there exists } y^* \in F^* \text{ such that} \qquad c \in C \implies k(c) - 2\langle j(c), y^*\rangle \ge \|y^*\|^2. \tag{5.1.2}$$

Furthermore, if

$$M := \sup_{c\in C}\Big[\|j(c)\| - \sqrt{k(c)+\|j(c)\|^2}\Big] \vee 0 \tag{5.1.3}$$

then

$$\min\big\{\|y^*\|\colon\ y^* \text{ is as in } (5.1.2)\big\} = M. \tag{5.1.4}$$

Proof. Since the values of c in $C \setminus \operatorname{dom} k$ have no impact on (5.1.1), (5.1.2) or the definition of M, we can and will suppose that $k\colon C \mapsto \mathbb{R}$. We first prove the implication (5.1.2)$\Longrightarrow$(5.1.1). Suppose that y^* is as in (5.1.2). Then

$$\begin{aligned} c\in C \quad &\Longrightarrow \quad k(c) \ge 2\langle j(c), y^*\rangle + \|y^*\|^2 \\ &\Longrightarrow \quad k(c) + \|j(c)\|^2 \ge \|j(c)\|^2 + 2\langle j(c), y^*\rangle + \|y^*\|^2 \\ &\Longrightarrow \quad k(c) + \|j(c)\|^2 \ge \|j(c)\|^2 - 2\|j(c)\|\|y^*\| + \|y^*\|^2 \\ &\Longrightarrow \quad k(c) + \|j(c)\|^2 \ge \big(\|j(c)\| - \|y^*\|\big)^2 \ge 0 \\ &\Longrightarrow \quad \sqrt{k(c)+\|j(c)\|^2} \ge \|j(c)\| - \|y^*\| \\ &\Longrightarrow \quad \|y^*\| \ge \|j(c)\| - \sqrt{k(c)+\|j(c)\|^2}. \end{aligned} \tag{5.1.5}$$

(5.1.5) gives (5.1.1) and, since $\|y^*\| \ge 0$, this also establishes that $\|y^*\| \ge M$. We now prove the implication (5.1.1)$\Longrightarrow$(5.1.2). So suppose that (5.1.1) is satisfied. We first show that

$$a, b \in C \quad \Longrightarrow \quad \|j(b)\| - \sqrt{k(b)+\|j(b)\|^2} \le \|j(a)\| + \sqrt{k(a)+\|j(a)\|^2}. \tag{5.1.6}$$

To this end, let $a, b \in C$, $\lambda > \sqrt{k(a)+\|j(a)\|^2} \ge 0$ and $\mu > \sqrt{k(b)+\|j(b)\|^2} \ge 0$. Write $\alpha := \|j(a)\| + \lambda$ and $\beta := \|j(b)\| - \mu$. Then, since j is affine,

$$0 \le \Big\|j\Big(\frac{\mu a + \lambda b}{\mu+\lambda}\Big)\Big\| = \Big\|\frac{\mu j(a) + \lambda j(b)}{\mu+\lambda}\Big\| \le \frac{\mu\|j(a)\| + \lambda\|j(b)\|}{\mu+\lambda} = \frac{\mu\alpha+\lambda\beta}{\mu+\lambda}.$$

Thus, from (5.1.1) applied to $c = \dfrac{\mu a + \lambda b}{\mu+\lambda} \in C$, and the convexity of k and $(\cdot)^2$,

$$0 \le k\Big(\frac{\mu a + \lambda b}{\mu+\lambda}\Big) + \Big(\frac{\mu\alpha+\lambda\beta}{\mu+\lambda}\Big)^2 \le \frac{\mu k(a) + \lambda k(b) + \mu\alpha^2 + \lambda\beta^2}{\mu+\lambda}.$$

Multiplying by $\mu+\lambda$ gives

$$\begin{aligned} 0 &\le \mu k(a) + \lambda k(b) + \mu\alpha^2 + \lambda\beta^2 \\ &= \mu\big(k(a)+\alpha^2\big) + \lambda\big(k(b)+\beta^2\big) \\ &= \mu\big(k(a) + \|j(a)\|^2 + 2\lambda\|j(a)\| + \lambda^2\big) + \lambda\big(k(b) + \|j(b)\|^2 - 2\mu\|j(b)\| + \mu^2\big) \\ &< \mu\big(2\lambda^2 + 2\lambda\|j(a)\|\big) + \lambda\big(2\mu^2 - 2\mu\|j(b)\|\big) = 2\mu\lambda\big(\lambda + \|j(a)\| + \mu - \|j(b)\|\big). \end{aligned}$$

On dividing by $2\mu\lambda$, we obtain $\|j(b)\| - \mu < \|j(a)\| + \lambda$, and (5.1.6) follows by letting $\mu \to \sqrt{k(b) + \|j(b)\|^2}$ and $\lambda \to \sqrt{k(a) + \|j(a)\|^2}$. Now (5.1.3) and (5.1.6) imply that, for all $c \in C$,

$$\|j(c)\| - \sqrt{k(c) + \|j(c)\|^2} \le M \quad \text{and} \quad M \le \|j(c)\| + \sqrt{k(c) + \|j(c)\|^2}, \tag{5.1.7}$$

from which

$$\begin{aligned} c \in C \quad &\Longrightarrow \quad \big|\|j(c)\| - M\big| \le \sqrt{k(c) + \|j(c)\|^2} \\ &\Longrightarrow \quad \big(\|j(c)\| - M\big)^2 \le k(c) + \|j(c)\|^2 \\ &\Longrightarrow \quad k(c) + 2M\|j(c)\| \ge M^2. \end{aligned}$$

It now follows from Theorem 2.8 that there exists $L \in F^*$ such that $\|L\| \le 2M$ and

$$k + L \circ j \ge M^2 \text{ on } C.$$

Thus (5.1.2) is satisfied with $y^* := -L/2$. This completes the proof of (5.1.2), and also shows that we can find y^* satisfying (5.1.2) with $\|y^*\| \le M$, establishing (5.1.4). ▮

Remark 5.2. We note that $y^* = 0$ satisfies (5.1.2) exactly when $k \ge 0$ on C and, in this case, $M = 0$. In all other cases, M is given by the simpler formula

$$\sup_{c \in C} \Big[\|j(c)\| - \sqrt{k(c) + \|j(c)\|^2}\Big].$$

6. An existence theorem without *a priori* scalar bounds for sublinear functionals

We note that (5.1.1) can be written $\inf_C \big[k + \psi \circ S \circ j\big] \ge 0$, where $\psi\colon \mathbb{R} \mapsto \mathbb{R}$ is defined by $\psi := (\cdot)^2$ and $S := \|\cdot\|$, and $\inf_C \big[S \circ j + k\big]$ in (2.8.1) can be written $\inf_C \big[k + \psi \circ S \circ j\big]$ where $\psi\colon \mathbb{R} \mapsto \mathbb{R}$ is defined by $\psi := (\cdot)$. Thus it is natural to ask whether there is a result that simultaneously generalizes Theorem 2.8 and Theorem 5.1. Theorem 6.1, which is such a result, is the topic of this section. The equivalence of (6.1.3) and (6.1.4) was first proved in [17, Theorem 5.4] using a rather technical product space argument and giving a weaker bound on N than that given here. We give here a new proof of this equivalence, which relies on the much simpler Dedekind section argument (6.1.7)–(6.1.11). Furthermore, as is clear from (6.1.6), the bound on N found in Theorem 6.1 is sharp. We refer the reader to [17, Remarks 5.5 and 5.6] for the details of how Theorem 6.1 implies Theorem 2.8 and Theorem 5.1.

We first discuss the conditions (6.1.1) and (6.1.2) on the function ψ. (6.1.1) is to ensure that the quantity M defined in (6.1.5) is finite, while (6.1.2) is needed in (6.1.8). Of course, (6.1.1) is automatically true if ψ is real–valued, as is the case with the two examples mentioned above. As for (6.1.2), if $\psi := (\cdot)$, ψ is increasing on $\mathbb{R}$ and so (6.1.2) is automatic while, if $\psi := (\cdot)^2$ and $S := \|\cdot\|$, (6.1.2) is true since $S \circ j(c) \le \gamma \Longrightarrow S \circ j(c), \gamma \in [0, \infty)$ and ψ is increasing on $[0, \infty)$. (We note that (6.1.1) was described in [17] by saying that ψ is "S, j–compatible".)

Theorem 6.1. *Let C be a nonempty convex subset of a vector space, E be a nontrivial vector space, S: $E \mapsto \mathbb{R}$ be sublinear, j: $C \mapsto E$ be S–convex and $k \in \mathcal{PC}(C)$. Let $\psi \in \mathcal{PC}(\mathbb{R})$ satisfy*

$$\big(S \circ j(\operatorname{dom} k) + (0,\infty)\big) \cap \operatorname{dom} \psi \ne \emptyset \tag{6.1.1}$$

and

$$c \in C \text{ and } S \circ j(c) \le \gamma \quad\Longrightarrow\quad \psi \circ S \circ j(c) \le \psi(\gamma). \tag{6.1.2}$$

Then

$$k + \psi \circ S \circ j \ge 0 \text{ on } C \tag{6.1.3}$$

if, and only if,

$$\left.\begin{array}{c}\text{there exist } N \ge 0 \text{ and a linear functional } L \text{ on } E \text{ such that}\\ L \le NS \text{ on } E \quad\text{and}\quad k + L \circ j \ge \psi^*(N) \text{ on } C.\end{array}\right\} \tag{6.1.4}$$

Furthermore, if

$$M := \sup_{c \in C,\ \mu < 0} \frac{k(c) + \psi\big(S \circ j(c) + \mu\big)}{\mu} \vee 0 \tag{6.1.5}$$

then

$$\min\big\{N\colon\ N \text{ is as in } (6.1.4)\big\} = M. \tag{6.1.6}$$

Proof. Suppose first that (6.1.4) is satisfied, from which $\psi^*(N) \in \mathbb{R}$. Then, for all $c \in C$ and $\nu \in \mathbb{R}$,

$$\begin{aligned} k(c) + \psi\big(S \circ j(c) + \nu\big) &\ge k(c) + N\big(S \circ j(c) + \nu\big) - \psi^*(N)\\ &= k(c) + NS \circ j(c) - \psi^*(N) + N\nu\\ &\ge k(c) + L \circ j(c) - \psi^*(N) + N\nu \ge N\nu. \end{aligned}$$

If we put $\nu = 0$ in this, we obtain (6.1.3). On the other hand, we also derive that

$$c \in C \text{ and } \mu < 0 \quad\Longrightarrow\quad \frac{k(c) + \psi\big(S \circ j(c) + \mu\big)}{\mu} \le N$$

and, since $N \ge 0$, this also shows that $N \ge M$. Suppose, conversely, that (6.1.3) is satisfied. We first show that

$$a, b \in C \text{ and } \mu < 0 < \lambda \Longrightarrow \frac{k(b) + \psi\big(S \circ j(b) + \mu\big)}{\mu} \le \frac{k(a) + \psi\big(S \circ j(a) + \lambda\big)}{\lambda}. \tag{6.1.7}$$

To this end, let $a, b \in C$ and $\mu < 0 < \lambda$. Write $\alpha := S \circ j(a) + \lambda$ and $\beta := S \circ j(b) + \mu$. Then, from the S–convexity of j and the sublinearity of S,

$$S \circ j\Big(\frac{\lambda b - \mu a}{\lambda - \mu}\Big) \le S\Big(\frac{\lambda j(b) - \mu j(a)}{\lambda - \mu}\Big) \le \frac{\lambda S \circ j(b) - \mu S \circ j(a)}{\lambda - \mu} = \frac{\lambda\beta - \mu\alpha}{\lambda - \mu}.$$

Thus, using (6.1.2) with $c := (\lambda b - \mu a)/(\lambda - \mu)$ and $\gamma := (\lambda\beta - \mu\alpha)/(\lambda - \mu)$, (6.1.3) and the convexity of k and ψ,

$$0 \le k\Big(\frac{\lambda b - \mu a}{\lambda - \mu}\Big) + \psi\Big(\frac{\lambda\beta - \mu\alpha}{\lambda - \mu}\Big) \le \frac{\lambda k(b) - \mu k(a) + \lambda\psi(\beta) - \mu\psi(\alpha)}{\lambda - \mu}, \tag{6.1.8}$$

and (6.1.7) follows on multiplication by $\lambda - \mu > 0$ and substituting in the values of α and β. From (6.1.2) and (6.1.3), for all $c \in C$ and $\lambda > 0$,

$$\frac{k(c) + \psi\big(S \circ j(c) + \lambda\big)}{\lambda} \ge \frac{k(c) + \psi \circ S \circ j(c)}{\lambda} \ge 0, \tag{6.1.9}$$

and (6.1.1) provides $a \in \operatorname{dom} k$ and $\lambda > 0$ such that $S \circ j(a) + \lambda \in \operatorname{dom}\psi$, from which

$$\frac{k(a) + \psi\big(S \circ j(a) + \lambda\big)}{\lambda} < \infty. \tag{6.1.10}$$

(6.1.7) and (6.1.10) imply that $M \in [0, \infty)$, and (6.1.7) and (6.1.9) that, for all $c \in C$ and $\mu < 0 < \lambda$,

$$\frac{k(c) + \psi\big(S \circ j(c) + \mu\big)}{\mu} \le M \le \frac{k(c) + \psi\big(S \circ j(c) + \lambda\big)}{\lambda}. \tag{6.1.11}$$

Combining this with (6.1.3), we obtain

$$\begin{aligned} c \in C \text{ and } \nu \in \mathbb{R} \quad &\Longrightarrow \quad k(c) + \psi\big(S \circ j(c) + \nu\big) \ge M\nu \\ &\Longleftrightarrow \quad k(c) + MS \circ j(c) \ge M\big(S \circ j(c) + \nu\big) - \psi\big(S \circ j(c) + \nu\big). \end{aligned}$$

Taking the supremum of the right–hand side over $\nu \in \mathbb{R}$ shows that

$$c \in C \quad \Longrightarrow \quad k(c) + MS \circ j(c) \ge \psi^*(M)$$

and (6.1.4) (with N replaced by M) now follows from Theorem 2.8. This completes the proof of Theorem 6.1. ▮

References

[1] H. Attouch and H. Brézis, *Duality for the sum of convex funtions in general Banach spaces*, Aspects of Mathematics and its Applications, J. A. Barroso, ed., Elsevier Science Publishers (1986), 125–133.

[2] F. E. Browder, *Nonlinear maximal monotone operators in Banach spaces*, Math. Annalen **175** (1968), 89–113.

[3] R. S. Burachik and B. F. Svaiter, *Maximal monotonicity, conjugation and the duality product*, IMPA Preprint 129/2002, February 28, 2002.

[4] K. Fan, *Minimax theorems*, Proc. Nat. Acad. Sci. U.S.A. **39** (1953), 42–47.

[5] K. Fan, I. Glicksberg and A. J. Hoffman, *Systems of inequalities involving convex functions*, Proc. Amer. Math. Soc. **8** (1957), 617–622.

[6] J. L. Kelley, I. Namioka, and co-authors, *Linear Topological Spaces*, D. Van Nostrand Co., Inc., Princeton – Toronto – London – Melbourne (1963).

[7] H. König, *Some Basic Theorems in Convex Analysis*, in "Optimization and operations research", edited by B. Korte, North-Holland (1982).

[8] D. L. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, Inc, New York – Chichester – Brisbane – Toronto – Singapore (1969).

[9] J.–J. Moreau, *Fonctionelles convexes*, Séminaire sur les équations aux derivées partielles, Lecture notes, Collège de France, Paris 1966.

[10] R. T. Rockafellar, *Extension of Fenchel's duality theorem for convex functions*, Duke Math. J. **33** (1966), 81–89.

[11] —–, *Conjugate duality and optimization*, Conference Board of the Mathematical Sciences **16**(1974), SIAM publications.

[12] W. Rudin, *Functional analysis*, McGraw-Hill, New York (1973).

[13] S. Simons, *Minimal sublinear functionals*, Studia Math. **37** (1970), 37–56.

[14] —–, *Minimax and monotonicity*, Lecture Notes in Mathematics **1693** (1998), Springer–Verlag.

[15] —–, *Maximal monotone multifunctions of Brøndsted–Rockafellar type*, Set–Valued Anal. **7** (1999), 255–294.

[16] —–, *Five kinds of maximal monotonicity*, Set–Valued Anal. **9** (2001), 391–409.

[17] —–, *A new version of the Hahn–Banach theorem*, Archiv der Mathematik, in press.

[18] S. Simons and C. Zălinescu, *A New Proof for Rockafellar's Characterization of Maximal Monotone Operators*, Proc. Amer. Math. Soc., in press.

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