Hahn–Banach theorems and maximal monotonicity

S. Simons

0. Introduction

In this paper, 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, convex analysis, and the theory of monotone multifunctions. All vector spaces in this paper will be *real*.

The main result appears in Theorem 1.7, which is bootstrapped from the special case contained in Lemma 1.4.

In Section 2, we sketch how Theorem 1.7 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 1.7 to convex analysis, pointing the reader to [26] for further details in two of these cases. One noteworthy property of proofs using Theorem 1.7 is that they allow us to avoid the problem of the "vertical hyperplane".

In Section 3, we show how Theorem 1.7 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 3.5, we use Theorem 1.7 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 1.7 allows us to avoid the problem of the "vertical hyperplane". More to the point, the results leading up to Theorem 3.5, namely Lemma 3.1 and Theorem 3.2, use Theorem 1.7 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 4 is motivated by the theory of monotone multifunctions. Theorem 4.1 is an existence theorem without any *a priori* scalar bounds in normed spaces that has proved very useful in the investigation of these multifunctions, and will be used in Theorem 5.5. A new feature of the result as presented here is a sharp lower bound on the norm of the linear functional obtained. Theorem 4.3 is a two–stage result obtained by combining Theorems 4.1 and 1.7, and will be used in the proof of Theorem 6.4.

In Section 5, we discuss the *free convexification* technique, which has many applications to the theory of monotone multifunctions. We list several of these without proof. We also use Theorem 4.1 to derive Rockafellar's surjectivity theorem for general (i.e., not renormed) reflexive spaces, with a sharp lower bound for solutions of the problem. Apart from its intrinsic interest, we have given this result here to introduce the techniques that are used in the more difficult problem treated in Section 6.

Maximal monotone multifunctions of "type (D)" were introduced by Gossez in order to generalize to nonreflexive spaces some of the results previously known for reflexive spaces (see Gossez, [10, Lemme 2.1, p. 375] and Phelps, [15, Section 3] for an exposition). Maximal monotone multifunctions of "type (FP)" were introduced by Fitzpatrick–Phelps in [8, Section 3] under the name of "locally maximal monotone" multifunctions. The motivation for their introduction was as follows. If E is reflexive then every maximal monotone multifunction on E can be approximated by "nicer" maximal monotone multifunctions using the Moreau-Yosida approximation. If E is nonreflexive then every subdifferential can also be approximated by "nicer" subdifferentials by using the operation of inf–convolution. So the question arises whether a general maximal monotone multifunction on a nonreflexive space can also be approximated by "nicer" maximal monotone multifunctions in some appropriate sense. Fitzpatrick–Phelps defined an appropriate sense of approximation in [8], and showed that the multifunctions of type (FP) can be approximated by "nicer" maximal monotone multifunctions in their sense. There has been considerable speculation for the past several years about the relationship between multifunctions of type (D) and multifunctions of type (FP). The main result of Section 6 (in Theorem 6.4) is then that that *every maximal monotone multifunction of type (D) is necessarily of type (FP).*

In the final section, we return to our consideration of abstract Hahn–Banach theorems. Noting a certain formal similarity between the statements of Theorem 4.1 and Theorem 1.7, we ask the question whether these two results can be unified. Indeed, they have a common generalization, which is given in Theorem 7.1.

1. The main result

Theorem 1.7 contains the new version of the Hahn-Banach theorem that forms the main topic of this paper. Theorem 1.7 is proved by bootstrapping from the special case contained in Lemma 1.4 — most of the work is actually done in Lemma 1.3.

We start by recalling in Lemma 1.2 the classical Hahn–Banach theorem for sublinear functionals.

Definition 1.1. 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 1.2. *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, [11, 3.4, p. 21] for a proof using cones, Rudin, [20, Theorem 3.2, p. 56–57] for a proof using an extension by subspaces argument, and König, [12] and Simons, [21] for a proof using an ordering on sublinear functionals. ▮

Lemma 1.3. *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*

$$T(x) := \inf_{d \in D,\ \lambda > 0} \big[S\big(x + \lambda a(d)\big) - \lambda\beta\big]. \tag{1.3.1}$$

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 (1.3.1),

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

hence $-T\big(-a(d)\big) \ge \beta$, which completes the proof of Lemma 1.3. ▌

Lemma 1.4. *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 1.2 (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 1.3. From Lemma 1.2, 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 1.5. 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 1.6. 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.

Theorem 1.7. *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*

$$\inf_C \big[L \circ j + k\big] = \inf_C \big[S \circ j + k\big]. \tag{1.7.1}$$

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 1.4 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,

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

2. Applications to functional analysis and minimax theorems

In this section, we mention without proof a number of applications of Theorem 1.7 that were discussed in [26]. We then state and prove in Theorem 2.5 a (necessary and sufficient) criterion for the Fenchel duality condition to hold.

Theorem 2.1 is the *sandwich theorem* (see [12, Theorem 1.7, p. 112]). It follows immediately from Theorem 1.7 with $C := E$ and $j(x) := x$.

Theorem 2.1. *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 2.1 implies in turn two other well known existence results: the *extension form of the Hahn–Banach theorem*, Corollary 2.2, (see [12, Corollary 1.8, p. 112]) and the *Mazur–Orlicz theorem*, Corollary 2.3, (see [12, Theorem 1.9, p. 112]).

Corollary 2.2. *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 2.3. *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 2.4 below was essentially proved by Fan–Glicksberg–Hoffman (see [6, Theorem 1, p. 618]), and leads to a short proof of the minimax theorem proved by Fan in [5] (see [23, Theorem 3.1, p. 17] for details of this). Theorem 2.4 follows easily from Theorem 1.7 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 2.4. *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 $y \in E$,

$$f(y) \ge \sup_{E^*} (y - f^*). \tag{2.4.1}$$

It was proved by Moreau in [14, Section 5–6, p. 26–39] that if f is lower semicontinuous on E then, for all $y \in E$, we have equality in (2.4.1). If f is lower semicontinuous at $y \in E$ but not on E then it does *not* follow that equality holds in (2.4.1) $\big($see [26, Remark 3.1]$\big)$. On the other hand, Theorem 1.7 can be used to find a necessary and sufficient condition for equality to hold in (2.4.1) for a given $y \in E$ $\big($see [26, 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 1.7 leads to a version of the Fenchel duality theorem.

Theorem 2.5. *Let E be a nontrivial Hausdorff locally convex space with dual E^*, and $f,\ g \in \mathcal{PC}(E)$. Then*

$$\text{there exists } z^* \in E^* \text{ such that} \qquad f^*(-z^*) + g^*(z^*) \le 0 \tag{2.5.1}$$

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

$$\text{there exists } S \in \mathcal{S}(E) \text{ such that} \qquad x,\ y \in E \Longrightarrow f(x) + g(y) + S(x-y) \ge 0. \tag{2.5.2}$$

Proof. Suppose first that (2.5.1) 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,

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

and (2.5.2) follows with $S := |z^*|$. Suppose, conversely, that (2.5.2) is satisfied. Then we apply Theorem 1.7 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 $L \le S$ and

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

or equivalently,

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

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

In the normed case, Theorem 2.5 takes the following form:

Corollary 2.6. *Let E be a nontrivial normed space with dual E^*, and $f,\ g \in \mathcal{PC}(E)$. Then*

$$\text{there exists } z^* \in E^* \text{ such that} \qquad f^*(-z^*) + g^*(z^*) \le 0 \tag{2.5.1}$$

if, and only if,

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

Corollary 2.6 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 [17, 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 2.5 and Corollary 2.6 give a *necessary and sufficient* condition for the existence of the linear functional, and not merely *sufficient* conditions.

In [19], Rockafellar develops a theory of dual problems and Lagrangians that gives a very large number of results in convex analysis. It was shown in [26, Theorem 3.6] how Theorem 1.7 can be used to give an efficient proof of [19, Theorem 17(a), p. 41], one of the main existence results in [19].

3. 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 3.2 which, combined with Lemma 3.1, 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 3.5 how Theorem 3.2 implies the classical result, with an upper bound on the norm as a bonus. The analysis in this section depends only on Theorem 1.7 — it does not depend on Section 2 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

$$x_1, x_2 \in C,\ \mu_1, \mu_2 > 0 \text{ and } \mu_1 + \mu_2 = 1 \implies j(\mu_1 x_1 + \mu_2 x_2) \preceq \mu_1 j(x_1) + \mu_2 j(x_2) \quad (3.0.1)$$

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

$$\inf_{j^{-1}N} k = \inf\big\{k(x)\colon\ x \in C,\ j(x) \preceq 0\big\} = \mu_0 \in \mathbb{R}. \quad (3.0.2)$$

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

$$\sup_N z_0^* \le 0 \quad (3.0.3)$$

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

$$\inf_{x \in C} \big[\langle j(x), z_0^*\rangle + k(x)\big] = \mu_0. \quad (3.0.4)$$

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

$$A := \big\{x \in C\colon\ k(x) < \mu_0\big\} \quad \text{and} \quad B := \big\{v \in C\colon\ j(v) \prec 0\big\}, \quad (3.0.5)$$

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 3.1. *Let z_0^* be a Lagrange multiplier, and A be as in (3.0.5). 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 (3.0.3) and (3.0.4),

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

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 3.1.

Theorem 3.2. *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 3.1 that* $M = \min\big\{\|z_0^*\|\colon\ z_0^* \text{ is a Lagrange multiplier}\big\}$.

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

$$S \text{ is sublinear,} \tag{3.2.1}$$

$$S \le \|\cdot\| \text{ on } E, \tag{3.2.2}$$

and

$$y \in N \quad \Longrightarrow \quad S(y) = 0. \tag{3.2.3}$$

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 (3.0.1) that

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

and so (3.2.3) implies that j is MS–convex. Thus (3.2.1) and Theorem 1.7 give a linear functional L on E such that $L \le MS$ on E and

$$\inf_C \big[L \circ j + k\big] = \inf_C \big[MS \circ j + k\big] \ge \mu_0. \tag{3.2.4}$$

We now derive from (3.2.2) and (3.2.3) that $L \in E^*$, $\|L\| \le M$ and $\sup_N L \le 0$. Since $x \in j^{-1}(N) \Longrightarrow j(x) \in N \Longrightarrow L \circ j(x) \le 0$, (3.0.2) 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 (3.2.4), which gives the required result (with $z_0^* = L$). ▌

Remark 3.3. 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 3.2, 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 [13, Corollary 8.3.1, p. 219] for details of the argument.

We recall from (3.0.5) 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 [13, Theorem 8.3.1, p. 217–218].) This will be improved in Theorem 3.5. We first give a preliminary lemma.

Lemma 3.4.
(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

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

(3.0.1) 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 (3.0.2), 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 (3.0.5). ▌

Theorem 3.5. *Suppose that $B \neq \emptyset$. Then there exists a Lagrange multiplier z_0^* such that*

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

Proof. Let $x \in A$ and $v \in B$. From Lemma 3.4(b), $\operatorname{dist}\big(j(x), N\big) > 0$ and

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

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

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

The result now follows from Theorem 3.2. ▌

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

The main result is this section is Theorem 4.1. The equivalence of (4.1.1) and (4.1.2) actually first appeared in [23, Theorem 7.2, p. 27–28], and was used in [23] to obtain a number of criteria for a monotone multifunction on a reflexive Banach space to be maximal monotone (including Rockafellar's "surjectivity theorem", which we revisit in Theorem 5.5), 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 [26]. This equivalence was also used in [25] to prove other results on maximal monotonicity. We will revisit the least technical of these in Theorem 6.4, but this time using Theorem 4.3, obtained by combining Theorems 4.1 and 1.7.

The proof of the equivalence of (4.1.1) and (4.1.2) given in [23, Theorem 7.2] was quite nonconstructive, and a more constructive proof was given in [26, Theorem 5.1], together with the bound $\inf_{c \in C}\Big[\|j(c)\| + \sqrt{k(c) + \|j(c)\|^2}\Big]$ on the norm of $\|y^*\|$ (see [26, Remark 5.6]). We now give a new proof of this equivalence, which relies on the direct Dedekind section argument (4.1.6)–(4.1.7) and is much simpler than the proofs given in [23] and [26]. Furthermore, as is clear from (4.1.4), the bound $\sup_{c \in C}\Big[\|j(c)\| - \sqrt{k(c) + \|j(c)\|^2}\Big] \vee 0$ on the norm of $\|y^*\|$ found in Theorem 4.1 is sharp. The analysis in this section depends only on Theorem 1.7 — it does not depend on Sections 2–3 in any way.

Theorem 4.1. *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*

$$c \in C \quad \Longrightarrow \quad k(c) + \|j(c)\|^2 \ge 0 \tag{4.1.1}$$

if, and only if,

$$\textit{there exists } y^* \in F^* \textit{ such that} \qquad c \in C \implies k(c) - 2\langle j(c), y^*\rangle \ge \|y^*\|^2. \tag{4.1.2}$$

Furthermore, if

$$M := \sup_{c \in C}\Big[\|j(c)\| - \sqrt{k(c) + \|j(c)\|^2}\Big] \vee 0 \tag{4.1.3}$$

then

$$\min\big\{\|y^*\|\colon\ y^* \textit{ is as in } (4.1.2)\big\} = M. \tag{4.1.4}$$

Proof. Since the values of c in $C \setminus \operatorname{dom} k$ have no impact on (4.1.1), (4.1.2) or the definition of M, we can and will suppose that $k\colon C \mapsto \mathbb{R}$. We first prove the implication (4.1.2)$\implies$(4.1.1). Suppose that y^* is as in (4.1.2). Then

$$\begin{aligned} c \in C \quad &\implies \quad k(c) \ge 2\langle j(c), y^*\rangle + \|y^*\|^2 \\ &\implies \quad k(c) + \|j(c)\|^2 \ge \|j(c)\|^2 + 2\langle j(c), y^*\rangle + \|y^*\|^2 \\ &\implies \quad k(c) + \|j(c)\|^2 \ge \|j(c)\|^2 - 2\|j(c)\|\|y^*\| + \|y^*\|^2 \\ &\implies \quad k(c) + \|j(c)\|^2 \ge \big(\|j(c)\| - \|y^*\|\big)^2 \ge 0 \\ &\implies \quad \sqrt{k(c) + \|j(c)\|^2} \ge \|j(c)\| - \|y^*\| \\ &\implies \quad \|y^*\| \ge \|j(c)\| - \sqrt{k(c) + \|j(c)\|^2}. \end{aligned} \tag{4.1.5}$$

(4.1.5) gives (4.1.1) and, since $\|y^*\| \ge 0$, this also establishes that $\|y^*\| \ge M$. We now prove the implication (4.1.1)$\implies$(4.1.2). So suppose that (4.1.1) is satisfied. We first show that

$$a, b \in C \quad \implies \quad \|j(b)\| - \sqrt{k(b) + \|j(b)\|^2} \le \|j(a)\| + \sqrt{k(a) + \|j(a)\|^2}. \tag{4.1.6}$$

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 (4.1.1) 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 (4.1.6) follows by letting $\mu \to \sqrt{k(b) + \|j(b)\|^2}$ and $\lambda \to \sqrt{k(a) + \|j(a)\|^2}$. Now (4.1.3) and (4.1.6) imply that, for all $c \in C$,

$$\|j(c)\| - \sqrt{k(c) + \|j(c)\|^2} \le M \quad \text{and} \quad M \le \|j(c)\| + \sqrt{k(c) + \|j(c)\|^2}, \tag{4.1.7}$$

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 1.7 that there exists $L \in F^*$ such that $\|L\| \le 2M$ and

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

Thus (4.1.2) is satisfied with $y^* := -L/2$. This completes the proof of (4.1.2), and also shows that we can find y^* satisfying (4.1.2) with $\|y^*\| \le M$, establishing (4.1.4). ∎

Remark 4.2. We note that $y^* = 0$ satisfies (4.1.2) exactly when $k \ge 0$ on C and, in this case, $M = 0$. In all other cases, M is given by the simpler formula

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

Theorem 4.3. *Let C be a nonempty convex subset of a vector space, F be a nontrivial normed space, Q: $F \mapsto \mathbb{R}$ be sublinear, h: $C \mapsto F$ be Q–convex, j: $C \mapsto F$ be affine, $k \in \mathcal{PC}(C)$ and*

$$c \in C \quad \Longrightarrow \quad k(c) + Q \circ h(c) + \|j(c)\|^2 \ge 0. \tag{4.3.1}$$

Then there exist a linear functional Λ on F such that $\Lambda \le Q$ on F, and $y^* \in F^*$ such that

$$c \in C \quad \Longrightarrow \quad k(c) - 2\langle j(c), y^*\rangle + \Lambda \circ h(c) \ge \|y^*\|^2. \tag{4.3.2}$$

Proof. Since $k + Q \circ h$ is convex, we first apply Theorem 4.1, with k replaced by $k + Q \circ h$ and obtain $y^* \in F^*$ such that

$$\begin{aligned} c \in C \quad &\Longrightarrow \quad k(c) + Q \circ h(c) - 2\langle j(c), y^*\rangle \ge \|y^*\|^2 \\ &\Longrightarrow \quad k(c) - 2\langle j(c), y^*\rangle + Q \circ h(c) \ge \|y^*\|^2. \end{aligned}$$

The result now follows from Theorem 1.7, with E, S, k, and j replaced by F, Q, $k - 2y^* \circ j$, and h, respectively. ∎

5. The free convexification technique

The main idea introduced in this section is a technique, the *free convexification* technique, which we will discuss in Definitions 5.1 and 5.2, Lemma 5.3 and Corollary 5.4. If we combine this technique with Theorem 1.7, Theorem 4.1 or Theorem 4.3, we can obtain a large number of results on (or related to) monotone multifunctions on a Banach space. (Specifically, Lemma 11.1, p. 41, Lemma 18.1, p. 65–66, Lemma 20.1, p. 77-78, Corollary 29.2, p. 114, Lemma 36.1, p. 141–142, Theorem 38.2, p. 146–147 and Theorem 38.3, p. 147–149 of [23] fall into this category, as well as some of the results of [24] and [25].) These results had been obtained previously using the minimax theorem of Fan referred to before Theorem 2.4. Since Theorem 1.7, Theorem 4.1 and Theorem 4.3 use the sublinear functional (nearly always, a scalar multiple of the norm) directly, this alternative method of proof is not only shorter, but it also avoids the need for the Banach–Alaoglu theorem, required to establish the compactness needed for the minimax theorem. As an illustration, we gave in [26, Theorem 4.1] a proof using Theorem 1.7 that a maximal monotone multifunction on a normed space with bounded range necessarily has full domain. This result can also be established using the Debrunner–Flor extension theorem (which depends on Brouwer's fixed–point theorem, see Phelps, [15, Lemma 1.7, p. 4] and the comments preceding), or the Farkas Lemma (see Fitzpatrick–Phelps, [8, Lemma 2.4, p. 580–581]). In Theorem 5.5 of this section, we will show how Theorem 4.1 leads to a proof of Rockafellar's surjectivity theorem for reflexive Banach spaces, with a sharp lower bound on the norm of solutions. For more details, see the discussion preceding Theorem 5.5. In Theorem 6.4 of the next section, we will show how Theorem 4.3 leads to a proof of a more recent result on maximal monotone multifunctions on nonreflexive Banach spaces. The analysis in this section does not depend on Sections 2–3 in any way.

Definition 5.1. Let $X \ne \emptyset$ and $\mathbb{R}^{(X)}$ be the direct sum of X copies of $\mathbb{R}$, the vector space of functions $\mu\colon X \mapsto \mathbb{R}$ such that

$$\big\{x \in X\colon\ \mu(x) \ne 0\big\} \quad \text{is finite.}$$

Define the injection $\delta_X\colon X \mapsto \mathbb{R}^{(X)}$ by

$$\delta_X(x)(y) := \begin{cases} 1, & (y = x); \\ 0, & (y \ne x). \end{cases}$$

$\big(\mathbb{R}^{(X)}, \delta_X\big)$ is the *free vector space over* X. Since $\delta_X(X)$ is a Hamel basis of $\mathbb{R}^{(X)}$, if V is any vector space and $f\colon X \mapsto V$ is any function whatsoever then there exists a linear map $g\colon \mathbb{R}^{(X)} \mapsto V$ such that $g \circ \delta_X = f$. We define $\mathcal{CO}(X) := \operatorname{co}\delta_X(X)$, the convex hull of $\delta_X(X)$ in $\mathbb{R}^{(X)}$. If $h = g|_{\mathcal{CO}(X)}$ then h is affine and $h\circ\delta_X = f$. We call $(\mathcal{CO}(X), \delta_X)$ the *free convexification of* X. We can give the following explicit description of h: if $c \in \mathcal{CO}(X)$ then there exist uniquely determined $\alpha_1, \ldots, \alpha_m \ge 0$ and $x_1, \ldots, x_m \in X$ such that $\sum_i \alpha_i = 1$ and $c = \sum_i \alpha_i\delta_X(x_i)$. In this case, $h(c) = \sum_i \alpha_i f(x_i)$.

Definition 5.2. Let E be a nontrivial Banach space with dual E^*, and $T\colon E \rightrightarrows E^*$ be a multifunction with

$$G(T) := \big\{(t, t^*)\colon\ t \in E,\ t^* \in Tt\big\} \ne \emptyset.$$

We say that (C,δ,p,q,r) is an $(E,E^*,\mathbb{R})$*–convexification of* T if C is a convex subset of a vector space, $p\colon C\mapsto E$, $q\colon C\mapsto E^*$ and $r\colon C\mapsto\mathbb{R}$ are affine, $\delta\colon G(T)\mapsto C$ with

$$C=\operatorname{co}\delta\big(G(T)\big)\tag{5.2.1}$$

and

$$(t,t^*)\in G(T)\quad\Longrightarrow\quad p\circ\delta(t,t^*)=t,\ q\circ\delta(t,t^*)=t^*\text{ and }r\circ\delta(t,t^*)=\langle t,t^*\rangle.\tag{5.2.2}$$

It is clear from Definition 5.1, applied with $V=E$, $V=E^*$ and $V=\mathbb{R}$ in turn, that there always exist $(E,E^*,\mathbb{R})$–convexifications of T.

Lemma 5.3. *Let E be a nontrivial Banach space, $T\colon E\rightrightarrows E^*$ be a multifunction with $G(T)\ne\emptyset$, and (C,δ,p,q,r) be an $(E,E^*,\mathbb{R})$–convexification of T. Then T is monotone if, and only if*

$$c\in C\quad\Longrightarrow\quad r(c)\ge\big\langle p(c),q(c)\big\rangle.\tag{5.3.1}$$

Proof. ($\Longrightarrow$) Let $c\in C$. From (5.2.1), $c=\sum_i\alpha_i\delta(t_i,t_i^*)$, where $\alpha_1,\dots,\alpha_m\ge 0$, $\sum_i\alpha_i=1$, and $(t_1,t_1^*),\dots,(t_m,t_m^*)\in G(T)$. Then

$$\begin{aligned}r(c)-\langle p(c),q(c)\rangle&=\textstyle\sum_i\alpha_i\langle t_i,t_i^*\rangle-\langle\sum_i\alpha_it_i,\sum_i\alpha_it_i^*\rangle\\&=\textstyle\sum_{i,j}\alpha_i\alpha_j\langle t_i,t_i^*\rangle-\sum_{i,j}\alpha_i\alpha_j\langle t_i,t_j^*\rangle\\&=\textstyle\sum_{i,j}\alpha_i\alpha_j\langle t_i,t_i^*-t_j^*\rangle\\&=\textstyle\sum_{i<j}\alpha_i\alpha_j\langle t_i,t_i^*-t_j^*\rangle+\sum_{j<i}\alpha_i\alpha_j\langle t_i,t_i^*-t_j^*\rangle\\&=\textstyle\sum_{i<j}\alpha_i\alpha_j\langle t_i,t_i^*-t_j^*\rangle+\sum_{i<j}\alpha_i\alpha_j\langle t_j,t_j^*-t_i^*\rangle\\&=\textstyle\sum_{i<j}\alpha_i\alpha_j\langle t_i-t_j,t_i^*-t_j^*\rangle\ge 0,\end{aligned}$$

where the final inequality follows from the monotonicity of T.

($\Longleftarrow$) Let (x,x^*), $(y,y^*)\in G(T)$. Then $\frac12\delta(x,x^*)+\frac12\delta(y,y^*)\in C$ and so, from (5.2.2) and (5.3.1),

$$\begin{aligned}2\langle x,x^*\rangle+2\langle y,y^*\rangle&=2r\circ\delta(x,x^*)+2r\circ\delta(y,y^*)\\&=4r\big(\tfrac12\delta(x,x^*)+\tfrac12\delta(y,y^*)\big)\\&\ge 4\big\langle p\big(\tfrac12\delta(x,x^*)+\tfrac12\delta(y,y^*)\big),q\big(\tfrac12\delta(x,x^*)+\tfrac12\delta(y,y^*)\big)\big\rangle\\&=4\big\langle\tfrac12p\circ\delta(x,x^*)+\tfrac12p\circ\delta(y,y^*),\tfrac12q\circ\delta(x,x^*)+\tfrac12q\circ\delta(y,y^*)\big\rangle\\&=4\big\langle\tfrac12x+\tfrac12y,\tfrac12x^*+\tfrac12y^*\big\rangle=\langle x+y,x^*+y^*\rangle.\end{aligned}$$

It follows from this that T is monotone. ∎

Corollary 5.4. *Let E be a nontrivial Banach space, $T\colon E\rightrightarrows E^*$ be a monotone multifunction with $G(T)\ne\emptyset$, and (C,δ,p,q,r) be an $(E,E^*,\mathbb{R})$–convexification of T. Then:*

(a) $c\in C\quad\Longrightarrow\quad 4r(c)+\big(\|p(c)\|+\|q(c)\|\big)^2\ge 0.$

(b) $(\tau,\tau^*)\in G(T)$ *and* $c\in C\quad\Longrightarrow\quad r(c)\ge\big\langle p(c),\tau^*\big\rangle+\big\langle\tau,q(c)\big\rangle-\langle\tau,\tau^*\rangle.$

Proof. (a) Let $c \in C$. From Lemma 5.3,

$$\begin{aligned} 4r(c) + \big(\|p(c)\| + \|q(c)\|\big)^2 &\ge \big(\|p(c)\| + \|q(c)\|\big)^2 + 4\big\langle p(c), q(c)\big\rangle \\ &\ge \big(\|p(c)\| + \|q(c)\|\big)^2 - 4\|p(c)\|\|q(c)\| \\ &= \big(\|p(c)\| - \|q(c)\|\big)^2 \ge 0. \end{aligned}$$

(b) Let $(\tau, \tau^*) \in G(T)$ and $c \in C$. From (5.2.2) and the monotonicity of T, for all $(t, t^*) \in G(T)$,

$$\begin{aligned} r\big(\delta(t,t^*)\big) - \big\langle p\big(\delta(t,t^*)\big), \tau^*\big\rangle - \big\langle \tau, q\big(\delta(t,t^*)\big)\big\rangle + \langle \tau, \tau^*\rangle &= \langle t, t^*\rangle - \langle t, \tau^*\rangle - \langle \tau, t^*\rangle + \langle \tau, \tau^*\rangle \\ &= \langle t - \tau, t^* - \tau^*\rangle \ge 0 \end{aligned}$$

Thus, from (5.2.1) and the affineness of p, q, and r, $r(c) - \big\langle p(c), \tau^*\big\rangle - \big\langle \tau, q(c)\big\rangle + \langle \tau, \tau^*\rangle \ge 0$. ∎

Rockafellar proved in [18, Proposition 1, p. 77-78] that if E is a non–trivial reflexive Banach space with dual E^* and duality map J: $E \rightrightarrows E^*$, J and J^{-1} are single–valued and T: $E \rightrightarrows E^*$ is a monotone multifunction then T is maximal monotone $\iff$ $T + J$ is surjective. Now ($\Longleftarrow$) of the above statement fails if J or J^{-1} is not single–valued (see [23, Remark 10.8, p. 39] for a discussion of this), while ($\Longrightarrow$) remains true (see [23, Theorem 10.7, p. 38]). It follows from a simple translation argument that, in order to prove that $T + J$ is surjective, it suffices to prove that there exists $y \in E$ such that $Ty + Jy \ni 0$. In Theorem 5.5 below, we give a proof of this result with a sharp lower bound on $\|y\|$ obtained from Theorem 4.1. (See also [23, Theorem 10.3, Corollary 10.4 and Theorem 10.6 p. 36–37] for characterizations of maximal monotonicity that are valid in general reflexive spaces with no restriction on J.) We mention parenthetically that the result of Rockafellar mentioned above depends on results of Browder, [3], which depend, in turn, on Brouwer's fixed–point theorem. We note for future reference that

$$G(J) = \big\{(x, x^*) \in E \times E^*\colon\ \|x\|^2 \vee \|x^*\|^2 = \langle x, x^*\rangle\big\}.$$

Theorem 5.5. *Let E be a non–trivial reflexive Banach space, T: $E \rightrightarrows E^*$ be a maximal monotone multifunction, (C, δ, p, q, r) be an $(E, E^*, \mathbb{R})$–convexification of T and*

$$M := \tfrac{1}{2} \sup_{c \in C} \Big[\|p(c)\| + \|q(c)\| - \sqrt{4r(c) + \big(\|p(c)\| + \|q(c)\|\big)^2}\,\Big] \vee 0. \tag{5.5.1}$$

Then there exists $x \in E$ such that $Tx + Jx \ni 0$, and

$$M = \min\big\{\|x\|\colon\ x \in E,\ Tx + Jx \ni 0\big\}. \tag{5.5.2}$$

Proof. Write $F := E \times E^*$ with $\|(x, x^*)\| := \|x\| + \|x^*\|$ and, for all $c \in C$, $k(c) := 4r(c)$ and $j(c) := \big(p(c), q(c)\big)$. It follows from Corollary 5.4(a) and Theorem 4.1 that there exists $y^* \in F^*$ such that

$$\|y^*\| = \sup_{c \in C} \Big[\|p(c)\| + \|q(c)\| - \sqrt{4r(c) + \big(\|p(c)\| + \|q(c)\|\big)^2}\,\Big] \vee 0 = 2M \tag{5.5.3}$$

and

$$c \in C \quad \Longrightarrow \quad 4r(c) - 2\big\langle j(c), y^* \big\rangle \geq \|y^*\|^2. \tag{5.5.4}$$

Now we can write $y^* = (2x^*, 2x)$ for some $(x, x^*) \in E \times E^*$, and $\|y^*\| = 2\|x\| \vee 2\|x^*\|$. (This is where we use the reflexivity of E.) Dividing (5.5.4) by 4, we obtain

$$c \in C \quad \Longrightarrow \quad r(c) - \big\langle p(c), x^* \big\rangle - \big\langle x, q(c) \big\rangle \geq \|x\|^2 \vee \|x^*\|^2.$$

If now $(t, t^*) \in G(T)$ and we substitute $c = \delta(t, t^*)$, we obtain from (5.2.2) that

$$\begin{aligned} (t, t^*) \in G(T) \quad &\Longrightarrow \quad \langle t, t^* \rangle - \langle t, x^* \rangle - \langle x, t^* \rangle \geq \|x\|^2 \vee \|x^*\|^2, \\ &\Longrightarrow \quad \langle t - x, t^* - x^* \rangle \geq \|x\|^2 \vee \|x^*\|^2 + \langle x, x^* \rangle \end{aligned}$$

Now $\|x\|^2 \vee \|x^*\|^2 + \langle x, x^* \rangle \geq \|x\|^2 \vee \|x^*\|^2 - \|x\|\|x^*\| \geq 0$, and so the maximal monotonicity of T implies that $(x, x^*) \in G(T)$. Substituting $(t, t^*) = (x, x^*)$ yields $\|x\|^2 \vee \|x^*\|^2 + \langle x, x^* \rangle \leq 0$, from which $-x^* \in Jx$. Since $0 = x^* + (-x^*)$, it is now immediate that $Tx + Jx \ni 0$, and it follows from (5.5.3) that $\|x\| = M$.

Suppose, conversely, that $x \in E$ and $Tx + Jx \ni 0$. Then there exists $x^* \in Tx$ such that $\|x\|^2 \vee \|x^*\|^2 + \langle x, x^* \rangle = 0$. Since T is monotone, using (5.2.2),

$$\begin{aligned} (t, t^*) \in G(T) \quad &\Longrightarrow \quad \langle t - x, t^* - x^* \rangle \geq \|x\|^2 \vee \|x^*\|^2 + \langle x, x^* \rangle \\ &\Longrightarrow \quad \langle t, t^* \rangle - \langle t, x^* \rangle - \langle x, t^* \rangle \geq \|x\|^2 \vee \|x^*\|^2, \\ &\Longrightarrow \quad r\big(\delta(t, t^*)\big) - \big\langle p\big(\delta(t, t^*)\big), x^* \big\rangle - \big\langle x, q\big(\delta(t, t^*)\big) \big\rangle \geq \|x\|^2 \vee \|x^*\|^2. \end{aligned}$$

It follows from (5.2.1), the affineness of p, q and r on C and the fact that $\|x^*\| = \|x\|$ that

$$\begin{aligned} c \in C \quad &\Longrightarrow \quad r(c) - \big\langle p(c), x^* \big\rangle - \big\langle x, q(c) \big\rangle \geq \|x\|^2 \vee \|x^*\|^2 \\ &\Longrightarrow \quad r(c) + \|p(c)\|\|x^*\| + \|x\|\|q(c)\| \geq \|x\|^2 \vee \|x^*\|^2 \\ &\Longrightarrow \quad r(c) + \big(\|p(c)\| + \|q(c)\|\big)\|x\| \geq \|x\|^2. \end{aligned}$$

On completing the square, we obtain that

$$c \in C \quad \Longrightarrow \quad \|x\| \geq \tfrac{1}{2}\Big[\|p(c)\| + \|q(c)\| - \sqrt{4r(c) + \big(\|p(c)\| + \|q(c)\|\big)^2}\,\Big].$$

Since $\|x\| \geq 0$, it is immediate from this that $\|x\| \geq M$, completing the proof of Theorem 5.5. ∎

Remark 5.6. It was shown by Zălinescu that the existence of $x \in E$ such that $Tx + Jx \ni 0$ in Theorem 5.5 can also be established by an argument using the Fitzpatrick function on $E \times E^*$ (see [7]), a technique due to Burachik and Svaiter (see [4]), and the Moreau–Rockafellar formula for the subdifferential of a sum, though it is not clear that this argument leads easily to a sharp lower bound on $\|x\|$. See [27] for details.

6. Type (D) implies type (FP)

In Theorem 6.4 of this section, we show how Theorem 4.3 and the free convexification technique introduced in Section 5 lead to a proof that *every maximal monotone multifunction of type (D) on a (possibly nonreflexive) Banach space is of type (FP) (i.e. locally maximal monotone)*, thus settling a question that has been open for some time. Yet again, the analysis in this section does not depend on Sections 2–3 in any way.

We now proceed to the definitions of the terms introduced above. In order to define maximal monotone multifunctions of types (D), we must introduce a concept due to Gossez: if $W\colon E \rightrightarrows E^*$, we define the multifunction $\overline{W}\colon E^{**} \rightrightarrows E^*$ by:

$$x^* \in \overline{W}x^{**} \quad \iff \quad \inf_{(w,w^*)\in G(W)} \langle w^* - x^*, \widehat{w} - x^{**}\rangle \geq 0,$$

where $\widehat{w}$ is the canonical image of w in E^{**}. In what follows, $R(W) := \bigcup_{x\in E} Wx$.

Definition 6.1. Let $W\colon E \rightrightarrows E^*$ be maximal monotone. W is said to be *of type (D)* if, for all $(x^{**}, x^*) \in G(\overline{W})$, there exists a bounded net $\{(w_\gamma, w_\gamma^*)\}$ of elements of $G(W)$ such that $(\widehat{w_\gamma}, w_\gamma^*) \to (x^{**}, x^*)$ in $w(E^{**}, E^*) \times \mathcal{T}_{\| \, \|}(E^*)$, where $\mathcal{T}_{\| \, \|}(E^*)$ is the norm topology of E^*. Clearly
• *if E is reflexive then every maximal monotone multifunction $W\colon E \rightrightarrows E^*$ is of type (D).*
It was essentially proved by Gossez in [10] (see Phelps, [15, Theorem 3.8, p. 221] for an exposition) that
• *if W is maximal monotone of type (D) then $\overline{R(W)}$ is convex.*
It was proved by Gossez in [10, Théorème 3.1, p. 376–378] that
• *if $f\colon E \mapsto (-\infty, \infty]$ is proper, convex and lower semicontinuous then $\partial f\colon E \rightrightarrows E^*$ is maximal monotone of type (D).*

Definition 6.2. A monotone multifunction $W\colon E \rightrightarrows E^*$ is said to be *of type (FP)* or *locally maximal monotone* provided the following holds: for any open convex subset U of E^* such that $U \cap R(W) \neq \emptyset$, if $(v, v^*) \in E \times U$ is such that

$$(w, w^*) \in G(W) \text{ and } w^* \in U \quad \Longrightarrow \quad \langle w - v, w^* - v^*\rangle \geq 0$$

then $(v, v^*) \in G(W)$. (If we take $U = E^*$, we see that every multifunction of type (FP) is maximal monotone.) It was proved by Fitzpatrick and Phelps in [8, Proposition 3.3, p. 585] that
• *if E is reflexive then every maximal monotone multifunction $W\colon E \rightrightarrows E^*$ is of type (FP).*
It was proved by Fitzpatrick and Phelps in [8, Theorem 3.5, p. 585] that
• *if W is maximal monotone of type (FP) then $\overline{R(W)}$ is convex.*
It was proved in [22, Main theorem, p. 470] and [23, Theorem 30.3, p. 120] that
• *if $f\colon E \mapsto (-\infty, \infty]$ is proper, convex and lower semicontinuous then $\partial f\colon E \rightrightarrows E^*$ is maximal monotone of type (FP).*
Finally, it was proved by Fitzpatrick and Phelps in [9, Theorem 3.7, p. 67] that
• *if W is maximal monotone and $R(W) = E^*$ then W is of type (FP).*

Most of the work for Theorem 6.4 will be done in the rather technical Lemma 6.3, below.

Lemma 6.3. *Let E be a nontrivial Banach space, $S\colon E \mapsto \mathbb{R}$ be sublinear, $T\colon E \rightrightarrows E^*$ be monotone and such that, for some $\tau^* \in R(T)$ and $\varepsilon > 0$,*

$$x \in E \quad\Longrightarrow\quad S(x) \ge \langle x, \tau^*\rangle + \varepsilon\|x\|. \tag{6.3.1}$$

Write $B := \{x^* \in E^*\colon\ x^* \le S \text{ on } E\}$. Then there exists $(z^*, z^{**}, y^{**}) \in B \times E^{**} \times E^{**}$ such that

$$\inf_{(t,t^*)\in G(T)} \langle t^* - z^*, \widehat{t} - z^{**} + y^{**}\rangle \ge \|z^*\|^2 \vee \|z^{**}\|^2 + \langle z^*, z^{**}\rangle + \sup\langle B - z^*, y^{**}\rangle \ge 0. \tag{6.3.2}$$

Proof. We fix $\tau \in T^{-1}\tau^*$, and then write $M := \|\tau\| \vee \|\tau^*\|$ and $N := 3M^2/\varepsilon$. Now let (C, δ, p, q, r) be a $(E, E^*, \mathbb{R})$–convexification of T and $D := C \times E \times B$. We first prove that, for all $(c, x, x^*) \in D$,

$$r(c) + S(x) + N\|q(c) - x^*\| + \tfrac14\big(\|p(c) + x\| + \|q(c)\|\big)^2 \ge 0. \tag{6.3.3}$$

So let us suppose that $(c, x, x^*) \in D$. If $\|x\| \le N$ then, from Lemma 5.3 and the definition of D,

$$\begin{aligned} r(c) + S(x) + N\|q(c) - x^*\| &+ \tfrac14\big(\|p(c) + x\| + \|q(c)\|\big)^2 \\ &\ge r(c) + \langle x, x^*\rangle + N\|q(c) - x^*\| + \|p(c) + x\|\|q(c)\| \\ &\ge r(c) + \langle x, x^*\rangle + \big\langle x, q(c) - x^*\big\rangle - \big\langle p(c) + x, q(c)\big\rangle \\ &\ge r(c) - \big\langle p(c), q(c)\big\rangle \ge 0, \end{aligned}$$

which gives (6.3.3). Suppose, on the other hand, that $\|x\| > N$. Then, from (6.3.1),

$$S(x) \ge \langle x, \tau^*\rangle + \varepsilon\|x\| \ge \langle x, \tau^*\rangle + 3M^2, \tag{6.3.4}$$

and, from Corollary 5.4(b),

$$r(c) \ge \big\langle p(c), \tau^*\big\rangle + \big\langle \tau, q(c)\big\rangle - \langle \tau, \tau^*\rangle \ge \big\langle p(c), \tau^*\big\rangle - M\|q(c)\| - M^2. \tag{6.3.5}$$

Using (6.3.4) and (6.3.5), we have

$$\begin{aligned} &r(c) + S(x) + \tfrac14\big(\|p(c) + x\| + \|q(c)\|\big)^2 \\ &\ge \big[\big\langle p(c), \tau^*\big\rangle - M\|q(c)\| - M^2\big] + \big[\langle x, \tau^*\rangle + 3M^2\big] + \tfrac14\big[\|p(c) + x\|^2 + \|q(c)\|^2\big] \\ &= 2M^2 + \big\langle p(c) + x, \tau^*\big\rangle - M\|q(c)\| + \tfrac14\|p(c) + x\|^2 + \tfrac14\|q(c)\|^2 \\ &\ge 2M^2 - M\|p(c) + x\| - M\|q(c)\| + \tfrac14\|p(c) + x\|^2 + \tfrac14\|q(c)\|^2 \\ &= \big(\tfrac12\|p(c) + x\| - M\big)^2 + \big(\tfrac12\|q(c)\| - M\big)^2 \ge 0, \end{aligned}$$

and (6.3.3) follows, since $N\|q(c)-x^*\|\ge 0$. Write $F:=E\times E^*$, normed by

$$\|(x,x^*)\|:=\|x\|+\|x^*\|,$$

and define $Q\colon\ F\mapsto\mathbb{R}$, $h\colon\ D\mapsto F$, $j\colon\ D\mapsto F$ and $k\colon\ D\mapsto\mathbb{R}$ by

$$Q(x,x^*):=S(x)+N\|x^*\|, \qquad \big((x,x^*)\in F\big)$$
$$h(c,x,x^*):=\big(x,q(c)-x^*\big), \qquad \big((c,x,x^*)\in D\big)$$
$$j(c,x,x^*):=\tfrac{1}{2}\big(p(c)+x,q(c)\big), \qquad \big((c,x,x^*)\in D\big)$$

and

$$k(c,x,x^*):=r(c). \qquad \big((c,x,x^*)\in D\big)$$

We note then that (6.3.3) can be written in the form

$$(c,x,x^*)\in D \quad\Longrightarrow\quad k(c,x,x^*)+Q\circ h(c,x,x^*)+\|j(c,x,x^*)\|^2\ge 0.$$

We now apply Theorem 4.3, with C replaced by D, and obtain a linear functional Λ on F such that $\Lambda\le Q$ on F, and $y^*\in F^*$, such that

$$(c,x,x^*)\in D \quad\Longrightarrow\quad r(c)-\big\langle\big(p(c)+x,q(c)\big),y^*\big\rangle+\Lambda\big(x,q(c)-x^*\big)\ge\|y^*\|^2.$$

Now there exists $(z^*,z^{**})\in E^*\times E^{**}$ such that $y^*=(z^*,z^{**})$. Furthermore, the form of Q implies that there exist a linear functional L on E such that $L\le S$ on E, and $y^{**}\in E^{**}$ with $\|y^{**}\|\le N$ such that $\Lambda=(L,y^{**})$. Consequently,

$$\begin{aligned}
&(c,x,x^*)\in D\\
&\quad\Longrightarrow\quad r(c)-\big\langle p(c)+x,z^*\big\rangle-\big\langle q(c),z^{**}\big\rangle+L(x)+\big\langle q(c)-x^*,y^{**}\big\rangle\ge\|y^*\|^2\\
&\quad\Longleftrightarrow\quad (L-z^*)(x)+r(c)-\big\langle p(c),z^*\big\rangle-\big\langle q(c),z^{**}-y^{**}\big\rangle-\langle x^*,y^{**}\rangle\ge\|y^*\|^2
\end{aligned}$$

thus, adding $\langle x^*,y^{**}\rangle$ to both sides of the above, and then taking the supremum over $x^*\in B$,

$$\begin{aligned}
&(c,x)\in C\times E\\
&\quad\Longrightarrow (L-z^*)(x)+r(c)-\big\langle p(c),z^*\big\rangle-\big\langle q(c),z^{**}-y^{**}\big\rangle\ge\|y^*\|^2+\sup\langle B,y^{**}\rangle.
\end{aligned}$$

For the moment fix c. It follows by taking the infimum over $x\in E$ that $L=z^*$, thus $z^*\le S$ on E, and so $z^*\in B$, as required. Substituting $L=z^*$ in the above, we have

$$c\in C\quad\Longrightarrow\quad r(c)-\big\langle p(c),z^*\big\rangle-\big\langle q(c),z^{**}-y^{**}\big\rangle\ge\|y^*\|^2+\sup\langle B,y^{**}\rangle.$$

It follows by taking $c=\delta(t,t^*)$ and using (5.2.2) that

$$\begin{aligned}
(t,t^*)\in G(T)\Longrightarrow\quad &\langle t,t^*\rangle-\langle t,z^*\rangle-\langle t^*,z^{**}-y^{**}\rangle\ge\|y^*\|^2+\sup\langle B,y^{**}\rangle\\
\Longleftrightarrow\quad &\langle t^*-z^*,\widehat{t}+y^{**}\rangle-\langle t^*,z^{**}\rangle\ge\|y^*\|^2+\sup\langle B-z^*,y^{**}\rangle.
\end{aligned}$$

We obtain the first inequality in (6.3.2) by adding $\langle z^*,z^{**}\rangle$ to both sides of the above and observing that $\|y^*\|^2=\|z^*\|^2\vee\|z^{**}\|^2$, and the second inequality follows since

$$\|z^*\|^2\vee\|z^{**}\|^2+\langle z^*,z^{**}\rangle\ge\|z^*\|^2\vee\|z^{**}\|^2-\|z^*\|\|z^{**}\|\ge 0 \tag{6.3.6}$$

and

$$\sup\langle B-z^*,y^{**}\rangle\ge\langle z^*-z^*,y^{**}\rangle\ge 0. \tag{6.3.7}$$

This completes the proof of Lemma 6.3. ▌

Theorem 6.4. *Let $W\colon\ E \rightrightarrows E^*$ be maximal monotone of type (D). Then W is of type (FP).*

Proof. Let U be an open convex subset of E^* such that $U \cap R(W) \neq \emptyset$ and $(v, v^*) \in E \times U$ be such that

$$(w, w^*) \in G(W) \text{ and } w^* \in U \quad \Longrightarrow \quad \langle w - v, w^* - v^* \rangle \geq 0.$$

We want to prove that $(v, v^*) \in G(W)$. Now define $T\colon\ E \rightrightarrows E^*$ by $G(T) := G(W) - (v, v^*)$. Further, writing $V := U - v^*$, we have that V is an open convex subset of E^* such that $V \ni 0$, $V \cap R(T) \neq \emptyset$ and

$$(t, t^*) \in G(T) \text{ and } t^* \in V \quad \Longrightarrow \quad \langle t, t^* \rangle \geq 0 \tag{6.4.1}$$

and now what we must prove is that

$$(0, 0) \in G(T). \tag{6.4.2}$$

We first find $\tau^* \in V \cap R(T)$ and choose $\varepsilon > 0$ so that

$$[0, \tau^*] + \big\{x^* \in E^*\colon\ \|x^*\| \leq \varepsilon\big\} \subset V.$$

We define the sublinear functional $S\colon\ E \mapsto \mathbb{R}$ by $S(x) := \langle x, \tau^* \rangle \vee 0 + \varepsilon\|x\|$, and B as in Lemma 6.3. It is then easy to see that $B = [0, \tau^*] + \big\{x^* \in E^*\colon\ \|x^*\| \leq \varepsilon\big\} \subset V$. Lemma 6.3 then gives us $(z^*, z^{**}, y^{**}) \in B \times E^{**} \times E^{**}$, such that

$$\inf_{(t,t^*) \in G(T)} \langle t^* - z^*, \widehat{t} - z^{**} + y^{**} \rangle \geq \|z^*\|^2 \vee \|z^{**}\|^2 + \langle z^*, z^{**} \rangle + \sup\langle B - z^*, y^{**} \rangle \geq 0. \tag{6.3.2}$$

It follows from this that $(z^{**} - y^{**}, z^*) \in G(\overline{T})$. Since W is of type (D), the same is true of T and so there exists a bounded net $\big\{\big(t_\gamma, t_\gamma^*\big)\big\}$ of elements of $G(T)$ such that $\big(\widehat{t_\gamma}, t_\gamma^*\big) \to (z^{**} - y^{**}, z^*)$ in $w(E^{**}, E^*) \times \mathcal{T}_{\|\ \|}(E^*)$. This implies that $\langle t_\gamma^* - z^*, \widehat{t_\gamma} - z^{**} + y^{**} \rangle \to 0$, and so putting $(t, t^*) = \big(t_\gamma, t_\gamma^*\big)$ and passing to the limit in (6.3.2),

$$0 \geq \|z^*\|^2 \vee \|z^{**}\|^2 + \langle z^*, z^{**} \rangle + \sup\langle B - z^*, y^{**} \rangle.$$

(6.3.6) now implies that

$$0 \geq \sup\langle B - z^*, y^{**} \rangle, \tag{6.4.3}$$

and (6.3.7) that

$$0 \geq \|z^*\|^2 \vee \|z^{**}\|^2 + \langle z^*, z^{**} \rangle. \tag{6.4.4}$$

Since $B \supset \big\{x^* \in E^*\colon\ \|x^*\| \leq \varepsilon\big\}$, (6.4.3) gives

$$\langle z^*, y^{**} \rangle \geq \sup\langle B, y^{**} \rangle \geq \varepsilon\|y^{**}\|. \tag{6.4.5}$$

Now $z^* \in B \subset V$ and $t_\gamma^* \to z^*$ in $\mathcal{T}_{\|\ \|}(E^*)$, so by truncating the net $\big\{\big(t_\gamma, t_\gamma^*\big)\big\}$ if necessary, we may suppose that, for all γ, $t_\gamma^* \in V$. Using (6.4.1), we now derive that, for all γ, $\big\langle t_\gamma^*, \widehat{t_\gamma} \big\rangle = \big\langle t_\gamma, t_\gamma^* \big\rangle \geq 0$. Passing to the limit in this, $\langle z^*, z^{**} - y^{**} \rangle \geq 0$ and, combining with (6.4.5), we obtain $\langle z^*, z^{**} \rangle \geq \varepsilon\|y^{**}\|$. If we now substitute this into (6.4.4), we obtain $0 \geq \|z^*\|^2 \vee \|z^{**}\|^2 + \varepsilon\|y^{**}\|$, hence $z^* = 0$ and $z^{**} = y^{**} = 0$. Substituting back into (6.3.2) yields $\inf_{(t,t^*) \in G(T)} \langle t^* - 0, \widehat{t} - 0 \rangle \geq 0$, that is to say,

$$\inf_{(t,t^*) \in G(T)} \langle t - 0, t^* - 0 \rangle \geq 0.$$

Since T is maximal monotone, this gives (6.4.2) and completes the proof of Theorem 6.4. ∎

We note that it was proved in Bauschke and Borwein [2, Theorem 4.1] (see also [16, Theorem 8.1, p. 327]) that every *continuous single–valued linear* maximal monotone multifunction of type (FP) is necessarily of type (D). However, we do not know the solution to the following problem:

Problem 6.5. Is every maximal monotone multifunction of type (FP) necessarily of type (D)?

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

We note that (4.1.1) 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 (1.7.1) 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 1.7 and Theorem 4.1. Theorem 7.1, which is such a result, is the topic of this section. The equivalence of (7.1.3) and (7.1.4) was first proved in [26, 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 (7.1.7)–(7.1.11). Furthermore, as is clear from (7.1.6), the bound on N found in Theorem 7.1 is sharp. We refer the reader to [26, Remarks 5.5 and 5.6] for the details of how Theorem 7.1 implies Theorem 1.7 and Theorem 4.1.

We first discuss the conditions (7.1.1) and (7.1.2) on the function ψ. (7.1.1) is to ensure that the quantity M defined in (7.1.5) is finite, while (7.1.2) is needed in (7.1.8). Of course, (7.1.1) is automatically true if ψ is real–valued, as is the case with the two examples mentioned above. As for (7.1.2), if $\psi := (\cdot)$, ψ is increasing on $\mathbb{R}$ and so (7.1.2) is automatic while, if $\psi := (\cdot)^2$ and $S := \|\cdot\|$, (7.1.2) is true since $\ S\circ j(c)\le\gamma \Longrightarrow S\circ j(c),\gamma\in[0,\infty)$ and ψ is increasing on $[0,\infty)$. $\big($We note that (7.1.1) was described in [26] by saying that ψ is "S, j–compatible".$\big)$

Theorem 7.1. *Let C be a nonempty convex subset of a vector space, E be a nontrivial vector space, $S\colon\ E\mapsto\mathbb{R}$ be sublinear, $j\colon\ C\mapsto E$ be S–convex and $k\in\mathcal{PC}(C)$. Let $\psi\in\mathcal{PC}(\mathbb{R})$ satisfy*

$$\big(S\circ j(\operatorname{dom} k)+(0,\infty)\big)\cap\operatorname{dom}\psi\ne\emptyset \tag{7.1.1}$$

and

$$c\in C \text{ and } S\circ j(c)\le\gamma \quad\Longrightarrow\quad \psi\circ S\circ j(c)\le\psi(\gamma). \tag{7.1.2}$$

Then

$$k+\psi\circ S\circ j\ge 0 \text{ on } C \tag{7.1.3}$$

if, and only if,

$$\left.\begin{array}{c}\text{there exist } N\ge 0 \text{ and a linear functional } L \text{ on } E \text{ such that}\\ L\le NS \text{ on } E \quad\text{and}\quad k+L\circ j\ge\psi^*(N) \text{ on } C.\end{array}\right\} \tag{7.1.4}$$

Furthermore, if

$$M := \sup_{c\in C,\ \mu<0} \frac{k(c)+\psi\big(S\circ j(c)+\mu\big)}{\mu} \vee 0 \tag{7.1.5}$$

then

$$\min\big\{N\colon\ N \textit{ is as in } (7.1.4)\big\} = M. \tag{7.1.6}$$

Proof. Suppose first that (7.1.4) 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 (7.1.3). 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 (7.1.3) is satisfied. We first show that

$$a,b\in C \text{ and } \mu<0<\lambda \Longrightarrow \frac{k(b)+\psi\big(S\circ j(b)+\mu\big)}{\mu}\le\frac{k(a)+\psi\big(S\circ j(a)+\lambda\big)}{\lambda}. \tag{7.1.7}$$

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 (7.1.2) with $c:=(\lambda b-\mu a)/(\lambda-\mu)$ and $\gamma:=(\lambda\beta-\mu\alpha)/(\lambda-\mu)$, (7.1.3) and the convexity of k and ψ,

$$0\le k\Big(\frac{\lambda b-\mu a}{\lambda-\mu}\Big)+\psi\Big(\frac{\lambda\beta-\mu\alpha}{\lambda-\mu}\Big)\le\frac{\lambda k(b)-\mu k(a)+\lambda\psi(\beta)-\mu\psi(\alpha)}{\lambda-\mu}, \tag{7.1.8}$$

and (7.1.7) follows on multiplication by $\lambda-\mu>0$ and substituting in the values of α and β. From (7.1.2) and (7.1.3), for all $c\in C$ and $\lambda>0$,

$$\frac{k(c)+\psi\big(S\circ j(c)+\lambda\big)}{\lambda}\ge\frac{k(c)+\psi\circ S\circ j(c)}{\lambda}\ge 0, \tag{7.1.9}$$

and (7.1.1) provides $a\in\operatorname{dom}k$ and $\lambda>0$ such that $S\circ j(a)+\lambda\in\operatorname{dom}\psi$, from which

$$\frac{k(a)+\psi\big(S\circ j(a)+\lambda\big)}{\lambda}<\infty. \tag{7.1.10}$$

(7.1.7) and (7.1.10) imply that $M \in [0, \infty)$, and (7.1.7) and (7.1.9) that, for all $c \in C$ and $\mu < 0 < \lambda$,

$$\frac{k(c) + \psi\big(S \circ j(c) + \mu\big)}{\mu} \le M \le \frac{k(c) + \psi\big(S \circ j(c) + \lambda\big)}{\lambda}. \tag{7.1.11}$$

Combining this with (7.1.3), 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 (7.1.4) (with N replaced by M) now follows from Theorem 1.7. This completes the proof of Theorem 7.1. ▌

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] H. H. Bauschke and J. M. Borwein, *Maximal monotonicity of dense type, local maximal monotonicity, and monotonicity of the conjugate are all the same for continuous linear operators*, Pacific J. Math. **189** (1999), 1–20.

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

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

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

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

[7] S. Fitzpatrick, *Representing monotone operators by convex functions*, Workshop/ Miniconference on Functional Analysis and Optimization (Canberra, 1988), 59–65, Proc. Centre Math. Anal. Austral. Nat. Univ., **20**, Austral. Nat. Univ., Canberra, 1988.

[8] S. P. Fitzpatrick and R. R. Phelps, *Bounded approximants to monotone operators on Banach spaces*, Ann. Inst. Henri Poincaré, Analyse non linéaire **9** (1992), 573–595.

[9] —–, *Some properties of maximal monotone operators on nonreflexive Banach spaces*, Set–Valued Analysis **3**(1995), 51–69.

[10] J.- P. Gossez, *Opérateurs monotones non linéaires dans les espaces de Banach non réflexifs*, J. Math. Anal. Appl. **34** (1971), 371–395.

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

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

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

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

[15] R. R. Phelps, *Lectures on Maximal Monotone Operators*, Extracta Mathematicae **12** (1997), 193–230.

[16] R. R. Phelps and S. Simons, *Unbounded linear monotone operators on nonreflexive Banach spaces*, J. Convex Analysis, **5** (1998), 303–328.

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

[18] —–, *On the Maximality of Sums of Nonlinear Monotone Operators*, Trans. Amer. Math. Soc. **149**(1970),75-88.

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

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

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

[22] —–, *Subdifferentials are locally maximal monotone*, Bull. Australian Math. Soc. **47** (1993), 465–471.

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

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

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

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

[27] 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