On the Virtual Convexity of the Domain and Range of a Nonlinear Maximal Monotone Operator

R. T. ROCKAFELLAR

Let X be a real Banach space, and let X^* be the dual of X, with $\langle x, x^* \rangle$ written in place of $x^*(x)$ for $x \in X$ and $x^* \in X^*$. A multivalued mapping (operator) $T: X \to X^*$ is said to be *monotone* if

$$\langle x - y, x^* - y^* \rangle \ge 0$$
 whenever $x^* \in T(x), y^* \in T(y)$.

It is said to be a maximal monotone mapping if its graph

$$\{(x, x^*) | x \in X, x^* \in T(x)\} \subset X \times X^* \tag{1}$$

is not contained properly in the graph of any other monotone mapping $T: X \to X^*$. Monotone mappings have received considerable attention recently because of applications to the theory of nonlinear differential equations (see Browder [5] and the references given there). It is known that, in particular, the subdifferential ∂f of any lower semi-continuous proper convex function f on X is a maximal monotone mapping (Rockafellar [12]).

Let D(T) and R(T) denote the effective domain and range of T, i.e.

$$D(T) = \{ x \in X \mid T(x) \neq 0 \} .$$

$$R(T) = \bigcup \{ T(x) \mid x \in X \} .$$

Minty [10] has shown that, if T is a maximal monotone mapping and X is *finite-dimensional*, then D(T) is *almost convex* in the sense that D(T) contains the (non-empty) relative interior of $\operatorname{conv} D(T)$, where conv denotes convex hull. It follows in this case that R(T) is likewise almost convex , because $R(T) = D(T^{-1})$, T^{-1} being the maximal (multivalued) monotone mapping defined by

$$T^{-1}(x^*) = \{x \mid x^* \in T(x)\}\ .$$

The purpose of this note is to prove, by means of some recent theorems of Browder [5], a certain infinite-dimensional generalization of Minty's result which is valid for the class of all Banach spaces X such that

X has an equivalent norm which is everywhere Fréchet differentiable except at the origin and whose dual norm on X^* is everywhere Fréchet differentiable except at the origin. (2)

It is well-known that (2) implies the reflexivity of X. For convenience in this paper, we shall refer to the Banach spaces satisfying (2) as *smoothly reflexive*. Of course, all Banach spaces isomorphic to L^p spaces with 1 are smoothly reflexive—in particular all Hilbert spaces. Asplund [1, Theorem 4]

has shown recently by means of a result of Kadec [8] that the class of smoothly reflexive Banach spaces includes all separable reflexive Banach spaces. At present it is an open question whether or not every reflexive Banach space is in fact smoothly reflexive, but it is known (Asplund [1, Theorem 4], based on Lindenstrauss [9]) that every reflexive Banach space does satisfy the weaker form of (2) in which Fréchet differentiability is replaced by Gâteaux differentiability.

We shall call a subset C of X virtually convex if, given any relatively (strongly) compact subset K of conv C and any $\varepsilon > 0$, there exists a strongly continuous single-valued mapping φ of all of K into C such that

$$|\varphi(x) - x| \le \varepsilon$$
 for every $x \in K$.

This condition implies in particular that cl C is convex, where cl denotes strong closure. It also implies that C is *finitely convex*¹ in the sense of Halkin [7], i.e. that given any finite subset S of C and any $\varepsilon > 0$, there exists a strongly continuous single-valued mapping φ of conv S into C such that $||\varphi(x) - x|| \le \varepsilon$ for every x in conv S. Trivially, every convex set is virtually convex.

Our main result is the following.

Theorem 1. Let X be a smoothly reflexive Banach space, and let $T: X \to X^*$ be a maximal monotone mapping. Then D(T) and R(T) are virtually convex sets.

The result of Minty may be regarded as a special case of Theorem 1 according to the lemma below. In Theorem 2, we shall state a weaker convexity result which does not imply Minty's result, but which is true in *any* reflexive Banach space.

Lemma. When X is finite-dimensional, a subset C of X is virtually convex it and only if it is almost convex.

Proof. It is enough to consider the case where the affine hull of C is all of X, so that cl C (which is in either event a convex set) has a non-empty interior.

Suppose first that C is almost convex, and let K by any relatively compact subset of conv C. Let z be an interior point of cl C. Given any $\varepsilon > 0$, choose $\lambda \in (0, 1]$ small enough that $\lambda \|z - x\| \le \varepsilon$ for every $x \in K$, and define φ by

$$\varphi(x) = (1 - \lambda) x + \lambda z$$
.

This φ is a continuous mapping such that $|\varphi(x) - x| \le \varepsilon$ for every $x \in K$. Moreover, $\varphi(x) \in \operatorname{int}(\operatorname{cl} C)$ for every $x \in \operatorname{cl} C$. Since C is almost convex, one has

$$\operatorname{int}(\operatorname{cl} C) \subset C \subset \operatorname{conv} C \subset \operatorname{cl} C$$
.

and it follows that φ maps K into C. Thus C is virtually convex.

Conversely now, suppose that C is virtually convex and let z be any interior point of conv C. Choose $\varepsilon > 0$ so small that

$$K_{\varepsilon} \subset \operatorname{conv} C$$
,

¹ Since this paper was written, J. P. Gossez has communicated to the author a proof that, conversely, every finitely convex set is virtually convex.

where K_{ε} is the closed ball of radius ε around z. Since K_{ε} is compact, there exists a continuous mapping φ of K_{ε} into C such that $\|\varphi(x) - x\| \le \varepsilon$ for every $x \in K_{\varepsilon}$. The mapping ψ defined by

$$\psi(x) = z + x - \varphi(x)$$

is continuous from K_r into itself, and hence ψ has a fixed point \overline{x} . One has $\varphi(\overline{x}) = z$, and this implies that $z \in C$. Thus C includes the (non-empty) interior of conv C. (We are grateful to Prof. V. L. Klee for pointing out this simple fixed point argument.)

Proof of Theorem 1. Inasmuch as X^* also satisfies (2) and $D(T) = R(T^{-1})$, it suffices by symmetry to prove the virtual convexity of R(T). Replacing T if necessary by the maximal monotone mapping T' defined by

$$T'(x) = T(x + \overline{x}) - \overline{x}^*$$
,

where \overline{x} is some point of D(T) and $\overline{x}^* \in T(\overline{x})$, we can assume without loss of generality that $0 \in T(0)$ so that

$$\langle x, x^* \rangle \ge 0$$
 whenever $x^* \in T(x)$. (3)

(The range of T' would be just a translate of R(T), and translations preserve virtual convexity.)

We can assume that the given norm $\|\cdot\|$ on X already has the differentiability properties described in (2). The corresponding extended spherical mapping J (Cudia [6]) is then a strong homeomorphism of X onto X^* such that

$$J(\lambda x) = \lambda J(x)$$
 for every real λ , (4)

$$|x|^2 = \langle x, J(x) \rangle = ||J(x)||^2.$$
 (5)

The mapping J is, of course, the Fréchet gradient of the convex function

$$j(x) = (1/2) ||x||^2, x \in X,$$
(6)

whose conjugate function is

$$j^*(x^*) = (1/2) \|x^*\|^2, x^* \in X^*.$$
(7)

Thus J is a maximal monotone mapping such that

(1/2)
$$||x||^2 + (1/2) ||x^*||^2 \ge \langle x, x^* \rangle$$
,
with equality $\Leftrightarrow x^* = J(x)$. (8)

Other properties of J will be mentioned as needed.

According to Browder [5, Theorem 2], for any $\lambda > 0$, $T + \lambda J$ is a maximal monotone mapping from X to X^* . Moreover, for any $x \in D(T + \lambda J) = D(T)$ and any $x^* \in (T + \lambda J)(x)$ we have

$$\langle x, x^* \rangle = \langle x, x^* - \lambda J(x) \rangle + \langle x, \lambda J(x) \rangle \ge \lambda \langle x, J(x) \rangle = \lambda |x|^2$$
 (9)

by (3) and (5). This says that $T + \lambda J$ is coercive, so that

$$X^* = R(T + \lambda J) = D((T + \lambda J)^{-1})$$
(10)

1

by Browder [5, Theorem 3]. We show now by a slight modification of an argument of Browder [5, p. 107] that $(T + \lambda J)^{-1}$ is single-valued and strongly continuous. Let x_1^*, x_2^*, \ldots , be a sequence in X^* converging to a point x_0^* . For each index k let x_k be an element of $(T + \lambda J)^{-1}(x_k^*)$, and let

$$y_k^* = x_k^* - \lambda J(x_k) \in T(x_k).$$

It suffices to show that x_k converges strongly to x_0 . Since

$$\lambda \|x_k\|^2 \le \|x_k\| \cdot \|x_k^*\|$$

by (9), the sequence $x_1, x_2, ...,$ must be bounded. Therefore

$$\begin{split} 0 &= \lim_{k \to \infty} \left\langle x_k - x_0, x_k^* - x_0^* \right\rangle \\ &= \lim_{k \to \infty} \left[\left\langle x_k - x_0, y_k^* - y_0^* \right\rangle + \lambda \left\langle x_k - x_0, J(x_k) - J(x_0) \right\rangle \right], \end{split}$$

where the latter "inner products" are non-negative by the monotonicity of T and J. It follows that

$$0 = \lim_{k \to x} \left\langle x_k - x_0, J(x_k) - J(x_0) \right\rangle. \tag{11}$$

Since J is the gradient mapping of the convex function j, we have (cf. Rockafellar [12])

$$j(x_0) \ge j(x_k) + \langle x_0 - x_k, J(x_k) \rangle,$$

$$j(x_k) \ge j(x_0) + \langle x_k - x_0, J(x_0) \rangle,$$

and consequently

$$\langle x_k - x_0, J(x_k) - J(x_0) \rangle \ge j(x_k) - j(x_0) - \langle x_k - x_0, J(x_0) \rangle \ge 0$$
.

Therefore

$$0 = \lim_{k \to \infty} \left[j(x_k) - j(x_0) - \langle x_k - x_0, J(x_0) \rangle \right]. \tag{12}$$

The Fréchet differentiability of the conjugate function j^* at $J(x_0)$ corresponds by Asplund-Rockafellar [3. Theorem 1] to j being rotund in the norm topology at x_0 relative to $J(x_0)$, which means that (12) implies that x_k converges strongly to x_0 . (This argument differs from Browder's [5, p. 107] in that Browder concludes the convergence of x_k from (11) by applying a lemma which requires X to be uniformly convex. We avoid uniform convexity by appealing to the more general Asplund-Rockafellar result.)

For each $\lambda > 0$ and $x^* \in X^*$, let

$$Q_{\lambda}(x^*) = (T + \lambda J)^{-1}(x^*), \tag{13}$$

$$P_{\lambda}(x^*) = x^* - \lambda J(Q_{\lambda}(x^*)). \tag{14}$$

From what we have just seen, Q_{λ} is a strongly continuous single-valued mapping of all of X^* into X. Moreover

$$|Q_{\lambda}(x^*)| \leq \lambda^{-1} ||x^*||$$

by (9), so that Q_{λ} carries bounded sets into bounded sets. Therefore P_{λ} is a strongly continuous single-valued mapping of all of X^* into X^* carrying bounded sets into bounded sets. The range of P_{λ} is contained in R(T); in fact

$$P_{\lambda}(x^*) \in T(Q_{\lambda}(x^*)), \tag{15}$$

because

$$x^* \in (T + \lambda J) (Q_{\lambda}(x^*)) = T(Q_{\lambda}(x^*)) + \lambda J(Q_{\lambda}(x^*))$$

by the definition of Q_{λ} . We shall demonstrate that

$$\lim_{\lambda \downarrow 0} P_{\lambda}(x^*) = x^* , \quad \forall x^* \in \operatorname{cl}(\operatorname{conv} R(T)), \tag{16}$$

where the convergence is strong and uniform on compact sets. This will establish the virtual convexity of R(T). (The choice of P_{λ} is motivated by Moreau's theory of proximal mappings [11]: if $T = \partial f$ for a lower semi-continuous proper convex function f on X with conjugate f^* on X^* , $P_{\lambda}(x^*)$ turns out to be the unique point y^* for which

$$(1/2) \|y^* - x^*\|^2 + \lambda f^*(y^*) \tag{17}$$

attains its minimum.)

Let K be any non-empty compact subset of cl(conv R(T)), and let $\varepsilon > 0$ (with $\varepsilon < 1$). We have

$$K \subset \bigcup \{C_{\alpha}, |0 \le \alpha < \infty\}, \tag{18}$$

where

$$C_{\alpha,\varepsilon} = \{x^* \mid \exists y^* \in \operatorname{cl}(\operatorname{conv} R_{\alpha}), \|y^* - x^*\| < \varepsilon/2\},$$

$$R_{\alpha} = \{x^* \mid \|x^*\| \le \alpha \quad \text{and} \quad \exists x \in T^{-1}(x^*), \|x\| \le \alpha\}.$$
(19)

Since K is compact, while the sets $C_{\alpha,\varepsilon}$ are open and nested, K must actually be contained in $C_{\alpha,\varepsilon}$ for some sufficiently large $\alpha > 1$. We shall demonstrate the existence of a $\lambda_0 > 0$ such that

$$||P_{\lambda}(x^*) - x^*|| \le \varepsilon, \quad \forall x^* \in C_{\alpha, \varepsilon}, \quad 0 < \lambda \le \lambda_0,$$
 (20)

and this will finish the proof of the theorem. (It will show also, by the way, that P_{λ} converges uniformly to the identity on $cl(\operatorname{conv} R_{\alpha})$ for any α as $\lambda \downarrow 0$.)

We need to establish as a preliminary that, for any $\mu > 0$, the set

$$\{P_{\lambda}(x^*) | x^* \in C_{\alpha,\varepsilon}, 0 < \lambda \le \mu\}$$
 (21)

is bounded. By (8) we have

$$(1/2) \|J^{-1}(x^* - P_{\lambda}(x^*))\|^2 + (1/2) \|x^* - P_{\mu}(x^*)\|^2$$

$$\geq \langle J^{-1}(x^* - P_{\lambda}(x^*)), x^* - P_{\mu}(x^*) \rangle, \qquad (22)$$

while by (5) we have

$$(1/2) \|J^{-1}(x^* - P_{\lambda}(x^*))\|^2 + (1/2) \|x^* - P_{\lambda}(x^*)\|^2$$

$$= \langle J^{-1}(x^* - P_{\lambda}(x^*)), x^* - P_{\lambda}(x^*) \rangle.$$
(23)

Subtracting (23) from (22), we get

$$(1/2) ||x^* - P_{\mu}(x^*)||^2 - (1/2) ||x^* - P_{\lambda}(x^*)||^2 \ge \langle J^{-1}(x^* - P_{\lambda}(x^*)), P_{\lambda}(x^*) - P_{\mu}(x^*) \rangle.$$
 (24)

On the other hand, the monotonicity of T implies by (15) that

$$0 \leq \langle Q_{\lambda}(x^*) - Q_{\mu}(x^*), P_{\lambda}(x^*) - P_{\mu}(x^*) \rangle.$$

and hence that

$$\langle Q_{\lambda}(x^*), P_{\lambda}(x^*) - P_{\mu}(x^*) \rangle \ge \langle Q_{\mu}(x^*), P_{\lambda}(x^*) - P_{\mu}(x^*) \rangle$$
. (25)

Of course

$$Q_{\lambda}(x^*) = \lambda^{-1} J^{-1}(x^* - P_{\lambda}(x^*)), \qquad Q_{\mu}(x^*) = \mu^{-1} J^{-1}(x^* - P_{\mu}(x^*)),$$

so by (24) and (25)

(1.2)
$$|x^* - P_n(x^*)|^2 - (1/2) |x^* - P_{\lambda}(x^*)|^2$$

$$\geq (\lambda |\mu) \langle J^{-1}(x^* - P_n(x^*)), P_{\lambda}(x^*) - P_n(x^*) \rangle$$

$$= (\lambda |\mu) \langle J^{-1}(x^* - P_n(x^*)), x^* - P_n(x^*) \rangle$$

$$+ (\lambda |\mu) \langle J^{-1}(x^* - P_n(x^*)), P_{\lambda}(x^*) - x^* \rangle$$

$$\geq -(\lambda |\mu) \langle J^{-1}(x^* - P_n(x^*)), x^* - P_{\lambda}(x^*) \rangle.$$

It follows that

$$||x^* - P_{\lambda}(x^*)||^2 = ||x^* - P_{\mu}(x^*)||^2$$

$$\leq (2\lambda |\mu) \langle J^{-1}(x^* - P_{\mu}(x^*)), x^* - P_{\lambda}(x^*) \rangle$$

$$\leq (2\lambda |\mu) ||x^* - P_{\mu}(x^*)|| \cdot ||x^* - P_{\lambda}(x^*)||.$$
(26)

For any $\mu > 0$, the quantity

$$\beta_u = \sup\{ ||x^* - P_u(x^*)|| | |x^* \in C_{x,x} \}$$

is finite, since $C_{x,x}$ is bounded and P_{μ} maps bounded sets into bounded sets. For any $x^* \in C_{x,x}$ and $0 < \lambda \le \mu$ we have

$$|x^* - P_{\lambda}(x^*)|^2 \le \beta_{\mu}^2 + 2\beta_{\mu}|x^* - P_{\lambda}(x^*)|$$

in other words

$$(|x^* - P_{\lambda}(x^*)| - \beta_{\mu})^2 \le 2\beta_{\mu}^2$$
. (27)

This proves that the set in (21) is bounded as claimed.

Now, taking $\mu = 1$, we choose a real number β such that

$$P_{\lambda}(x^*) \leq \beta \quad \text{when} \quad x^* \in C_{\lambda, \epsilon}, 0 < \lambda \leq 1,$$
 (28)

and we set

$$\hat{\lambda}_0 = \varepsilon^2 / 4\alpha(\alpha + \beta) \,, \tag{29}$$

where $\lambda_0 < 1$ (by our assumption that $\varepsilon < 1$, $\alpha > 1$). Given any $y^* \in R_\alpha$, there exists by definition a y such that $y^* \in T(y)$ and $||y|| \le \alpha$. For this y and y^* we have, by (15) and the monotonicity of T,

$$0 \leq \langle y - Q_{\lambda}(x^*), y^* - P_{\lambda}(x^*) \rangle = \langle y - \lambda^{-1} J^{-1}(x^* - P_{\lambda}(x^*)), y^* - P_{\lambda}(x^*) \rangle,$$

and hence

$$\lambda \langle y, y^* - P_{\lambda}(x^*) \rangle \ge \langle J^{-1}(x^* - P_{\lambda}(x^*)), y^* - P_{\lambda}(x^*) \rangle$$

$$= \langle J^{-1}(x^* - P_{\lambda}(x^*)), y^* - x^* \rangle + \langle J^{-1}(x^* - P_{\lambda}(x^*)), x^* - P_{\lambda}(x^*) \rangle$$

$$= \langle J^{-1}(x^* - P_{\lambda}(x^*)), y^* - x^* \rangle + \|x^* - P_{\lambda}(x^*)\|^2.$$
(30)

When $x^* \in C_{\alpha, \varepsilon}$ and $0 < \lambda \le \lambda_0$, we have

$$\lambda \langle y, y^* - P_{\lambda}(x^*) \rangle \leq \lambda \|y\| (\|y^*\| + \|P_{\lambda}(x^*)\|)$$

$$\leq \lambda_0 \alpha(\alpha + \beta) = \varepsilon^2 / 4.$$
 (31)

It follows from (30) and (31) that

$$||x^* - P_1(x^*)||^2 \le \langle J^{-1}(x^* - P_1(x^*)), x^* - y^* \rangle + (\varepsilon^2/4)$$
(32)

when $x^* \in C_{\alpha,\varepsilon}$ and $0 < \lambda \le \lambda_0$. We have shown that (32) holds for any $y^* \in R_{\alpha}$. But, for each x^* and λ , (32) is a linear inequality in y^* , so if it is satisfied for every $y^* \in R_{\alpha}$ it must actually be satisfied for every $y^* \in \operatorname{cl}(\operatorname{conv} R_{\alpha})$. Thus, when $x^* \in C_{\alpha,\varepsilon}$ and $0 < \lambda \le \lambda_0$, we have

$$||x^* - P_{\lambda}(x^*)||^2 \le ||J^{-1}(x^* - P_{\lambda}(x^*))|| \cdot ||x^* - y^*|| + (\varepsilon^2/4)$$

$$= ||x^* - P_{\lambda}(x^*)|| \cdot ||x^* - y^*|| + (\varepsilon^2/4)$$
(33)

for every $y^* \in cl(conv R_a)$. Since

$$\inf\{\|x^* - y^*\| \mid y^* \in \operatorname{cl}(\operatorname{conv} R_{\alpha})\} < \varepsilon/2 \tag{34}$$

for any $x^* \in C_{\alpha, \varepsilon}$ by the definition of $C_{\alpha, \varepsilon}$, it follows that

$$||x^* - P_{\lambda}(x^*)||^2 < (\varepsilon/2) ||x^* - P_{\lambda}(x^*)|| + (\varepsilon^2/4)$$
 (35)

when $x^* \in C_{\alpha,\varepsilon}$ and $0 < \lambda \le \lambda_0$. It is easily verified that (35) implies $||x^* - P_{\lambda}(x^*)|| < \varepsilon$.

Remark. The proof actually shows that, given any bounded (not necessarily relatively compact) subset K of

$$\operatorname{cl}(\operatorname{conv} \cup \{T(x) | \|x\| \le \alpha\}) \tag{36}$$

(for any $\alpha > 0$), and given any $\varepsilon > 0$, there exists a strongly continuous (single-valued) mapping φ of K onto a bounded subset of R(T) such that $\|\varphi(x^*) - x^*\| \le \varepsilon$ for every $x^* \in K$. Likewise, given any bounded subset K of

$$\operatorname{cl}(\operatorname{conv}\{x \mid \exists x^* \in T(x), \|x^*\| \le \alpha\}) \tag{37}$$

and any $\varepsilon > 0$, there exists a strongly continuous (single-valued) mapping φ of K onto a bounded subset of D(T) such that $\|\varphi(x) - x\| \le \varepsilon$ for every $x \in K$.

Note added in proof. H. Brezis, M. G. Crandall and A. Pazy have now shown that P_{λ} is always continuous from the strong topology to the weak topology. (See the remarks after Lemma 1.3 of their forthcoming paper "Perturbations of nonlinear maximal monotone sets in Banach space.") Thus Theorem 2 can be strengthened to the following.

Theorem 2'. Let X be any reflexive Banach space, and let $T: X \to X^*$ be a maximal monotone mapping. Then D(T) and R(T) are "weakly" virtually convex, that is, they satisfy the definition of virtual convexity, except that φ is not necessarily continuous from the strong topology to the strong topology, but only from the strong topology to the weak topology.

References

- 1. Asplund, E.: Averaged norms. Israel J. Math. 5, 227-233 (1967).
- 2. Positivity of duality mappings. Bull. Amer. Math. Soc. 73, 200—203 (1967).
- 3. —, Rockafellar, R. T.: Gradients of convex functions. To appear in Trans. Amer. Math. Soc.
- 4. Brøndsted, A., Rockafellar, R. T.: On the subdifferentiability of convex functions. Proc. Amer. Math. Soc. 16, 405—411 (1965).
- 5. Browder, F. E.: Nonlinear maximal monotone operators in Banach space. Math. Ann. 175,* 89—113 (1968).
- Cudia, D. F.: The geometry of Banach spaces. Smoothness. Trans. Amer. Math. Soc. 110, 284—314 (1964).
- 7. Halkin, H.: Finite convexity in infinite-dimensional spaces. Proc. of the Colloquium on Convexity, Copenhagen, 1965, W. Fenchel (ed.), Copenhagen (1967), 126—131.
- 8. Kadec, M. I.: Spaces isomorphic to a locally uniformly convex space. Izv. Vyss. Ucebn. Zaved. Matematika 13, 51—57 (1959) (Russian).
- Lindenstrauss, J.: On nonseparable reflexive Banach spaces. Bull. Amer. Math. Soc. 72, 967—970 (1966).
- 10. Minty, G. J.: On the maximal domain of a "monotone" function. Michigan Math. J. 8, 135-137 (1961).
- Moreau, J. J.: Proximité et dualité dans un espace hilbertien. Bull. Soc. Math. France 93, 273—299 (1965).
- 12. Rockafellar, R. T.: Characterization of the subdifferentials of convex functions. Pacific J. Math. 17, 497—510 (1966). (A correction to the maximality proof in this paper is given in: On the maximal monotonicity of subdifferential mappings. Pacific J. Math. (1970).)

Professor R. T. Rockafellar Department of Mathematics University of Washington Seattle, Washington 98105, USA

(Received May 13, 1968)