— Reprinted from: *Theory and Applications of Monotone Operators.*
Proceedings of a NATO Advanced Study Institute
held in Venice, Italy, June 17-30, 1968.

CONVEX FUNCTIONS, MONOTONE OPERATORS AND VARIATIONAL INEQUALITIES (*)

R. TYRRELL ROCKAFELLAR, SEATTLE (WASH., USA)

The theory of extended-real-valued convex functions has been the subject of much development in the last few years; for expositions of the results, see Moreau [20] (infinite-dimensional case) and Rockafellar [31] (finite-dimensional case). Our aim here is to explain some of the connections between this theory and the theory of monotone operators, with special emphasis on applications to extremum problems and variational inequalities.

The basic connection between the two theories is the fact that the subdifferential of a proper convex function f is a monotone operator, indeed often a maximal monotone operator. Subdifferential mappings ∂f are defined in § 1, and a number of useful examples are given. Among these is the subdifferential of the indicator $\delta(\cdot \mid K)$ of a convex set K. This mapping, which assigns to each $x \in K$ the normal cone to K at x, is very important in studying extremum problems.

In § 2 we show that the solutions x to the problem of minimizing a proper convex function f over a non-empty convex set K usually can be described as the points x such that

$$0 \in \partial f(x) + \partial \delta(x \mid K).$$

In general, the condition

$$0 \in T(x) + \partial \delta(x \mid K)$$

(*) Supported in part by the Air Force Office of Scientific Research under grant no. AF-AFOSR-1202-67.

is known as the *variational inequality* for T and K. Variational inequalities do not necessarily arise from true extremum problems, but, as we explain in § 4, there are cases where a variational inequality for a maximal monotone operator T corresponds to an extremum problem (to be specific, a minimax problem) even though T is not the gradient (or generalized gradient) of any function.

A proof is given in § 4 of the fact that ∂f is a maximal monotone operator when f is a lower semi-continuous proper convex function (and in particular that $\partial\delta(\cdot|K)$ is a maximal monotone operator when K is a non-empty closed convex set).

The theory of Lagrange multipliers, developed in § 2, makes it possible to analyze general variational inequalities by decomposing $\partial\delta(x\,|\,K)$, when K is given by some finite or infinite system of linear or convex inequalities. In § 5, however, we pursue a question at the opposite extreme from decomposition. When is the monotone operator

$$T + \partial\delta(\cdot\,|\,K)$$

again maximal (assuming that T is maximal and K is closed), so that the existence theory for variational inequalities can be reduced to the fundamental existence theory for solutions x to relations of the simpler form $0 \in S(x)$, where S is a maximal monotone operator. Theorems are presented which allow this reduction in most cases of interest.

In passing from a given variational inequality to the study of a single maximal monotone operator S, one is led to many questions about the nature of the effective domain $D(S)$ and range $R(S)$ of S, as well as conditions for membership in these sets. A number of results which bear on such questions are discussed in § 6.

1. Subdifferentials of convex functions.

In everything that follows, X denotes a real Banach space, and X^* denotes the dual of X. For $x \in X$ and $x^* \in X^*$, we write $\langle x, x^* \rangle$ instead of $x^*(x)$.

A function f on X with values in $R \cup \}+\infty\{$ (where R denotes the real line) is said to be *convex* if

$$(1) \qquad f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y), \qquad 0 < \lambda < 1,$$

for any $x \in X$ and $y \in X$. This condition is equivalent to the condition that the set

$$\text{(2)} \qquad \operatorname{epi} f = \{(x, \lambda) \in X \oplus R \mid \mu \geq f(x)\},$$

which is called the *epigraph* of f, be a convex subset of the space $X \oplus R$. By a *proper* convex function on X, we shall mean a convex function with values in $R \cup \{+\infty\}$ which is not merely the constant function $+\infty$.

If f is any proper convex function on X, then the set

$$\text{(3)} \qquad \operatorname{dom} f = \{x \in X \mid f(x) < +\infty\},$$

called the *effective domain* of f, is a non-empty convex subset of X on which f is real-valued. Conversely, if K is a non-empty convex subset of X and f is a real-valued function on K which is convex (i. e. satisfies (1) when $x \in K$ and $y \in K$), then one can obtain a proper convex function on X by setting $f(x) = +\infty$ for every $x \notin K$.

A very useful example of a proper convex function is the *indicator* $\delta(\cdot \mid K)$ of a non-empty convex set K, which is defined by

$$\text{(4)} \qquad \delta(x \mid K) = \begin{cases} 0 & \text{if } x \in K, \\ +\infty & \text{if } x \notin K. \end{cases}$$

Indicator functions play a role in the theory of convex functions similar to that played by the characteristic functions of sets in some other areas of mathematics.

Let f be a proper convex function on X, and let x be a point of X. An element $x^* \in X^*$ is said to be a *subgradient* of f at x if

$$\text{(5)} \qquad f(y) \geq f(x) + \langle y - x, x^* \rangle, \quad \forall\, y \in X.$$

Geometrically, this condition means that the graph of the affine function

$$y \longrightarrow f(x) + \langle y - x, x^* \rangle$$

is a supporting hyperplane to the convex set epi f in $X \oplus R$ at the point $(x, f(x))$. The set of all subgradients x^* of f at x is denoted by $\partial f(x)$. The multivalued mapping

$$\partial f : x \longrightarrow \partial f(x) \subset X^*$$

is called the *subdifferential* of f.

It is obvious from (5) that $\partial f(x)$ is a weak* closed convex set in X^*, since the set of all $x^* \in X^*$ satisfying an inequality of the form

$$\langle z, x^* \rangle \leq \alpha$$

is always a weak* closed convex set. If $x \notin \operatorname{dom} f$, $\partial f(x)$ is trivially empty. If $x \in \operatorname{dom} f$ and f is continuous at x, then $\partial f(x)$ is necessarily non-empty and weak* compact [19]. If $x \in \operatorname{dom} f$ and f is differentiable at x in the sense of Gâteaux (or Fréchet), then $\partial f(x)$ consists of a unique element of X^*, namely the *gradient* of f at x, denoted by $\nabla f(x)$.

EXAMPLE 1. Let f be a real-valued convex function on X which is everywhere Gâteaux (or Fréchet) differentiable. Then ∂f reduces to the single-valued gradient mapping

$$\nabla f : X \to X^*.$$

(For the continuity properties of such mappings, see [3] aud [20]).

EXAMPLE 2. Let $j(x) = (1/2)\,\|x\|^2$. Then j is a continuous convex function and, for each $x \in X$, $\partial j(x)$ is the set of all $x^* \in X^*$ such that

$$(6) \qquad \langle x, x^* \rangle = \|x\| \cdot \|x^*\| \quad \text{and} \quad \|x^*\| = \|x\|\,.$$

The multivalued mapping ∂j is called the *extended spherical mapping* from X to X^*. When X is a Hilbert space, ∂j reduces to the canonical isomorphism between X and X^*. In general, a mapping of the form ∂f, where $f(x) = \Phi(\|x\|)$ and Φ is a non-negative real-valued (strictly) increasing strictly convex function on $[0, +\infty)$, is called a *duality mapping* from X to X^* (see [2]).

EXAMPLE 3. Let K be a non-empty convex set in X. For the indicator $\delta(\cdot \mid K)$, $x^* \in \partial\delta(x \mid K)$ if and only if

$$(7) \qquad x \in K \quad \text{and} \quad \langle y - x, x^* \rangle \leq 0, \qquad \forall\, y \in K.$$

An $x^* \in X^*$ is said to be a *normal* to K at x if (7) holds. The set of all normals to K at x is a certain weak* closed convex cone, non-empty (because it contains the zero element of X^*) when $x \in K$, but empty by definition when $x \notin K$. The subdifferential mapping $\partial\delta(\cdot \mid K)$ thus associates with each x the *normal cone* to K at x.

EXAMPLE 4. As a special case of Example 3, let K be a subspace of X. Then

$$\partial\delta(x \mid K) = \begin{cases} K^{\perp} & \text{if } x \in K, \\ \varnothing & \text{if } x \notin K, \end{cases}$$

where $K^{\perp}$ is the subspace of X^* « orthogonal » to K, i.e.

$$K^{\perp} = \{x^* \in X^* \mid \langle x, x^* \rangle = 0, \quad \forall\, x \in K\}.$$

EXAMPLE 5. As another specialization of Example 3, suppose that K is of the form $\{x \mid f(x) \leq 0\}$, where f is a continuous real-valued convex function on X such that $f(x) < 0$ for at least one x. In this case, it can be shown that $\partial\delta(x \mid K)$ is the convex cone in X^* generated by the (weak* compact convex) set $\partial f(x)$ when $f(x)=0$, whereas $\partial\delta(x \mid K) = \{0\}$ when $f(x) < 0$ and $\partial\delta(x \mid K) = \varnothing$ when $f(x) > 0$. (If f is Gâteaux differentiable at x, the convex cone generated by $\partial f(x)$ reduces to the set of all non-negative scalar multiples of the gradient $\nabla f(x)$).

EXAMPLE 6. Let K^* be any non-empty weak* closed convex subset of X^*, and let f be the *support function* of K^* on X, i.e.

$$f(x) = \sup \{\langle x, x^* \rangle \mid x^* \in K^*\}.$$

Then f is a lower semi-continuous proper convex function, and it can be shown that, for each $x \in X$, $\partial f(x)$ is the subset of K^* consisting of the points (if any) where the linear function $\langle x, \cdot \rangle$ attains its maximum. Note as a special case here that, if K^* is the unit ball of X^*, f is the norm on X.

2. Conditions for a minimum.

If f is a proper convex function on X, then, trivially, the (global) minimum of f occurs at the point x if and only if $0 \in \partial f(x)$. This condition may be regarded as an analogue of the familiar condition $\nabla f(x) = 0$ for extrema of differentiable functions. To make non-trivial use of it, however, we need to have some means of computing subgradients in given cases.

Consider, for instance, the proper convex function g defined by

$$g(x) = f(x) + \delta(x \mid K) = \begin{cases} f(x) & \text{if } x \in K, \\ +\infty & \text{if } x \notin K, \end{cases} \tag{8}$$

where f is a real-valued (continuous) Gâteaux differentiable convex function on X and K is a non-empty convex subset of X. Minimizing g over X is equivalent to minimizing f over K. Thus the minimum of f over K is attained at the point $x \in K$ if and only if $0 \in \partial g(x)$. In order to analyze this condition, one needs to analyze ∂g in terms of the subdifferentials of the convex functions f and $\delta(\cdot K)$, which are known (to some extent at least — see also below) from Examples 1 and 2 of § 1. The following theorem may be applied.

THEOREM 1. *Let f_1 and f_2 be proper convex functions on X. Suppose there exists a point of* $\operatorname{dom} f_1 \cap \operatorname{dom} f_2$ *at which one of the two functions, say f_1, is continuous. Then, for every $x \in X$,*

$$\begin{aligned} \partial(f_1 + f_2)(x) &= \partial f_1(x) + \partial f_2(x) \\ &= \{x_1^* + x_2^* \mid x_1^* \in \partial f_1(x),\, x_2^* \in \partial f_2(x)\}. \end{aligned} \tag{9}$$

This result was first proved in the finite-dimensional case in [22] and later extended to the infinite-dimensional case in [19] and [23]. For a further extension, see [15].

Proof. If $x_1^* \in \partial f_1(x)$ and $x_2^* \in \partial f_2(x)$, we have

$$f_1(y) \geq f_1(x) + \langle y - x, x_1^* \rangle, \quad \forall\, y \in X,$$

$$f_2(y) \geq f_2(x) + \langle y - x, x_2^* \rangle, \quad \forall\, y \in X,$$

by definition. Adding these inequalities, we obtain

$$(f_1 + f_2)(y) \geq (f_1 + f_2)(x) + \langle y - x, x_1^* + x_2^* \rangle, \quad \forall\, y \in X,$$

or in other words

$$x_1^* + x_2^* \in \partial(f_1 + f_2)(x).$$

This shows that

$$\partial(f_1 + f_2)(x) \supset \partial f_1(x) + \partial f_2(x).$$

These conditions can be expressed, respectively, as

$$f_1(x) \geq f_1(0) + \langle x - 0, x^* \rangle, \qquad \forall x \in X,$$

$$f_2(x) \geq f_2(0) + \langle x - 0, -x^* \rangle, \qquad \forall x \in X,$$

or in other words

$$x^* \in \partial f_1(0), \qquad -x^* \in \partial f_2(0).$$

It follows from this that

$$0 \in \partial f_1(0) + \partial f_2(0),$$

and the proof of Theorem 1 is complete.

To apply Theorem 1, let us return to the example given at the beginning of this section. There, taking $f_1 = f$ and $f_2 = \delta(\cdot \mid X)$, the hypothesis of the theorem is satisfied, so we have

$$\partial g(x) = \partial f(x) + \partial \delta(x \mid K), \qquad \forall x \in X. \tag{11}$$

Also, $\partial f(x)$ reduces to $\nabla f(x)$ by the differentiability assumption on f. Thus

$$0 \in \partial g(x) \Longleftrightarrow -\nabla f(x) \in \partial \delta(x \mid K),$$

i. e. f attains its minimum over K at x if and only if ($x \in K$ and) $-\nabla f(x)$ is a normal to K at x. More generally, we may state

COROLLARY 1. *Let f be a proper convex function on X, and let K be a convex subset of X. Suppose there exists a point of $K \cap \operatorname{dom} f$ which is interior to K, or which is a point of continuity of f. Then f attains its minimum over K at x if and only if ($x \in K$ and) there exists an $x^* \in \partial f(x)$ such that $-x^*$ is a normal to K at x.*

Proof. The hypothesis of Theorem 1 is satisfied either for $f_1 = \delta(\cdot \mid K)$ and $f_2 = f$, or for $f_1 = f$ and $f_2 = \delta(\cdot \mid K)$. Hence (11) holds for $g = f + \delta(\cdot \mid K)$, and the condition $0 \in \partial g(x)$ for a minimum of g can be analyzed as we have just done.

Theorem 1 may also be employed to analyze further the condition that a given vector be a normal to K at x.

COROLLARY 2. *Let K_1 and K_2 be convex subsets of X, and let $K = K_1 \cap K_2$. Suppose there is a point of K which is interior to* *either* K_1 *or* K_2. *Then*

$$(12)\qquad \partial\delta(x \mid K) = \partial\delta(x \mid K_1) + \partial\delta(x \mid K_2), \qquad \forall x.$$

In other words, x^* *is a normal to* K *at* x *if and only if* ($x \in K$ *and*) *there exist elements* x_1^* *and* x_2^* *of* X^*, *such that* x_1^* *is a normal to* K_1 *at* x, x_2^* *is a normal to* K_2 *at* x, *and* $x_1^* + x_2^* = x^*$.

Proof. Apply Theorem 1 to

$$\delta(\cdot \mid K) = \delta(\cdot \mid K_1) + \delta(\cdot \mid K_2).$$

COROLLARY 3. *Let* $f_1, \dots, f_m$ *be continuous real-valued convex functions on* X, *and let*

$$(13)\qquad K = \{x \mid f_i(x) \leq 0,\ i = 1, \dots, m\}.$$

Suppose that

$$(14)\qquad \exists\, x \quad \text{such that } f_i(x) < 0,\ i = 1, \dots, m.$$

Then x^* *is a normal to* K *at a point* x *if and only if there exist real numbers* $\lambda_1, \dots, \lambda_m$ *such that*

$$x^* \in \lambda_1\, \partial f_1(x) + \dots + \lambda_m\, \partial f_m(x),$$

$$\lambda_i \geq 0, \qquad f_i(x) \leq 0, \qquad \lambda_i f_i(x) = 0, \qquad i = 1, \dots, m.$$

Proof. We have $K = K_1 \cap \dots \cap K_m$, where

$$K_i = \{x \mid f_i(x) \leq 0\}.$$

Moreover, the hypothesis implies that

$$\operatorname{int} K_1 \cap \dots \cap \operatorname{int} K_m \neq \emptyset.$$

Apply Corollary 2 and the description of $\partial\delta(\cdot \mid K_i)$ given in Example 5 of § 1.

To combine some of these results, consider the problem of minimizing f over the set (13), where $f, f_1, \dots, f_m$ are all real-valued (continuous) Gâteaux differentiable functions on X. By Corollary 1 and Corollary 3, the minimum of f over K is attained at x if and only if the conditions

$$\nabla f(x) + \lambda_1 \nabla f_1(x) + \dots + \lambda_m \nabla f_m(x) = 0, \tag{15}$$

$$\lambda_i \geq 0, \qquad f_i(x) \leq 0, \qquad \lambda_i f_i(x) = 0, \qquad i = 1, \dots, m, \tag{16}$$

are satisfied for certain real numbers λ_i. Here λ_i is called the *Lagrange multiplier* associated with the constraint $f_i(x) \leq 0$. Conditions (15) and (16) are called the *Kuhn-Tucker* conditions; the above derivation of them is taken from [23]. Note that (15) implies in particular that the convex function

$$f + \lambda_1 f_1 + \dots + \lambda_m f_m$$

attains its (unconstrained) minimum on X at the point x.

It may be mentioned in connection with Corollary 3 that a linear constraint of the from

$$\langle x, a^* \rangle = \alpha, \qquad a^* \in X^*, \qquad \alpha \in R$$

can always be expressed as a pair of convex inequality constraint

$$f_1(x) \leq 0 \quad \text{and} \quad f_2(x) \leq 0,$$

where

$$f_1(x) = \langle x, a^* \rangle - \alpha, \qquad f_2(x) = \alpha - \langle x, a^* \rangle.$$

In the presence of such constraints, hypothesis (14) of Corollary 3, as stated, cannot be satisfied. However, it may be proved that the conclusion of Corollary 3 is still valid if the condition $f_i(x) < 0$ in (14) is weakened to $f_i(x) \leq 0$ for each i such that f_i is affine (i. e. linear-plus-a-constant) — see [23].

An analogue of Corollary 3 may also be proved for systems of infinitely many constraints.

THEOREM 2. *Let I be a compact Hausdorff space, and for each $i \in I$ let f_i be a real-valued (continuous) Gâteaux differentiable convex function on X. Let*

$$K = \{x \in X \mid f_i(x) \leq 0, \ \forall\, i \in I\}.$$

Suppose that $f_i(x)$ is a continuous function of $i \in I$ for each $x \in X$, and that

$$\exists x, \ \exists \varepsilon > 0, \quad \forall\, i \in I, f_i(x) \leq -\varepsilon. \tag{17}$$

Then x^* is a normal to K at a point x if and only if there exists a finite Borel measure λ on I such that

$$x^* = \int_I \nabla f_i(x)\, d\lambda(i) \tag{18}$$

$$x \in K, \qquad \lambda \geq 0, \qquad \int_I f_i(x)\, d\lambda(i) = 0. \tag{19}$$

Here (18) is interpreted to mean that x^* is the unique element of X^* such that

$$\langle u, x^* \rangle = \int_I \langle u, \nabla f_i(x) \rangle\, d\lambda_i(x), \qquad \forall\, u \in X.$$

Of course, (19) says that x satisfies $f_i(x) \leq 0$ for every $i \in I$, and that λ is a non-negative measure whose support is contained in the (closed) set of indices i such that actually $f_i(x) = 0$.

Proof of Theorem 2. Suppose that (18) and (19) are satisfied, and let y be any point of K. Then

$$0 \geq f_i(y) \geq f_i(x) + \langle y - x, \nabla f_i(x) \rangle, \qquad \forall\, i \in I,$$

so that

$$0 \geq \int_I f_i(y)\, d\lambda(i) \geq \int_I [f_i(x) + \langle y - x, \nabla f_i(x) \rangle]\, d\lambda(i) \tag{20}$$

$$= \langle y - x, \int_I \nabla f_i(x)\, d\lambda(i) \rangle = \langle y - x, x^* \rangle.$$

This shows that x^* is a normal to K at x.

Conversely, suppose that x^* is a normal to K at x. Then, by definition, the linear function $\langle \cdot, x^* \rangle$ attains its maximum over K at x.

Let $C(I)$ be the Banach space of all continuous real-valued functions on I and, for each $u \in C(I)$, let $p(u)$ be the infimum of $-\langle \cdot, x^* \rangle$ over the set

$$K_u = \{ y \mid f_i(x) \leq u(i) \}, \qquad \forall\, i \in I.$$

It is easily verified that the epigraph of p is a convex subset of $C(I) \oplus R$. Moreover

$$p(0) = -\langle x, x^* \rangle > -\infty$$

and by (17) p is bounded above on a certain neighborhood of the origin It follows from these facts that p is a proper convex function on $C(I)$. which is continuous on a certain neighborhood of the origin. Hence $\partial p(0) \neq \emptyset$. Let u^* be an element of the dual space $C(I)^*$ such that $-u^* \in \partial p(0)$. Of course, u^* corresponds by the Riesz representation theorem to some finite Borel measure λ on I, and for this λ we have

$$\begin{aligned} p(u) &\geq p(0) + \langle u - 0, -u^* \rangle \\ &= p(0) - \int_I u(i)\, d\lambda(i), \qquad \forall\, u \in C(I). \end{aligned}$$

In view of the definition of p, the latter condition means that

$$(21) \qquad -\langle y, x^* \rangle \geq -\langle x, x^* \rangle - \int_I [f_i(y) + w(i)]\, d\lambda(i)$$

for every $y \in X$ and every $w \in C(I)$ such that

$$w(i) \geq 0, \qquad \forall\, i \in I.$$

This implies in particular that λ is a non-negative measure, for otherwise there would exist a non-negative $v \in C(I)$ such that

$$\int_I v(i)\, d\lambda(i) < 0,$$

and (21) could be contradicted for any given y by taking $w = \alpha v$ for a sufficiently large constant $\alpha > 0$. Setting $y = x$ and $w(i) \equiv 0$ in (20), we see that

$$\int_I f_i(x)\, d\lambda(i) \geq 0.$$

Since $f_i(x) \leq 0$ and $\lambda \geq 0$, however, this integral is also ≤ 0. Therefore (19) holds. It follows then further from (21) that, for every $y \in X$,

$$\begin{aligned}\langle y - x, x^* \rangle &\leq \int_I f_i(y)\, d\lambda(i) \\ &= \int_I f_i(y)\, d\lambda(i) - \int_I f_i(x)\, d\lambda(i).\end{aligned}$$

Setting $y = x + \tau u$, where $\tau > 0$, we can write this as

$$\langle u, x^* \rangle \leq \int_I \tau^{-1} [f_i(x + \tau u) - f_i(x)]\, d\lambda(i). \tag{22}$$

The difference quotient in the integral in (22) is, by the convexity of f_i, a non-decreasing function of τ for each u, and

$$\lim_{\tau \downarrow 0} \tau^{-1} [f_i(x + \tau u) - f_i(x)] = \langle u, \nabla f_i(x) \rangle. \tag{23}$$

Therefore (23) implies that

$$\begin{aligned}\langle u, x^* \rangle &\leq \int_I \langle u, \nabla f_i(x) \rangle\, d\lambda(i), \\ &= \langle u, \int_I \nabla f_i(x)\, d\lambda(i) \rangle, \qquad \forall\, u \in X,\end{aligned} \tag{24}$$

and, since this is equivalent to (18), our proof is finished. (The measurability of the functions of the form

$$i \to \langle u, \nabla f_i(x) \rangle,$$

which was used in (20) and (24), follows from (23) and the assumed continuity of $f_i(x)$ in i.)

3. The maximal monotonicity of subdifferentials.

If f be a proper convex function on X. Given $x_0^* \in \partial f(x_0)$ and $x_1^* \in \partial f(x_1)$, we have

$$f(x_1) \geq f(x_0) + \langle x_1 - x_0, x_0^* \rangle,$$

$$f(x_0) \geq f(x_1) + \langle x_0 - x_1, x_1^* \rangle,$$

and adding these inequalities, we obtain

$$\langle x_1 - x_0 , x_0^* \rangle + \langle x_0 - x_1 , x_1^* \rangle \leq 0.$$

Therefore

$$\langle x_1 - x_0 , x_1^* - x_0^* \rangle \geq 0 \text{ whenever } x_i^* \in \partial f(x_i), \quad i = 0, 1.$$

This condition means by definition that the multivalued mapping $\partial f : X \longrightarrow X^*$ is a *monotone operator.*

More generally, if $x_i^* \in \partial f(x_i)$ for $i = 0, 1, \dots, n$,

$$\begin{aligned} f(x_1) &\geq f(x_0) + \langle x_1 - x_0 , x_0^* \rangle, \\ f(x_2) &\geq f(x_1) + \langle x_2 - x_1 , x_0^* \rangle, \\ &\cdots\cdots\cdots\cdots \\ f(x_n) &\geq f(x_{n-1}) + \langle x_n - x_{n-1} , x_{n-1}^* \rangle, \\ f(x_0) &\geq f(x_n) + \langle x_0 - x_n , x_n^* \rangle, \end{aligned}$$

and consequently

$$\langle x_1 - x_0 , x_0^* \rangle + \dots + \langle x_0 - x_n , x_n^* \rangle \leq 0. \tag{25}$$

A multivalued mapping $T : X \longrightarrow X^*$ which satisfies (25) whenever

$$x_i^* \in T(x_i) \text{ for } i = 0, 1, \dots, n \ (n \text{ arbitrary})$$

is said to be a *cyclically monotone operator.* Thus $\partial f : Y \longrightarrow X^*$ is a cyclically monotone operator.

The importance of cyclic monotonicity is apparent from the following result.

THEOREM 3 [25]. *Let $T : X \longrightarrow X^*$ be a multivalued mapping. In order that there exist a proper convex function f on X such that*

$$T(x) \subset \partial f(x), \qquad \forall\, x \in X, \tag{26}$$

it is necessary and sufficient that T be a cyclically monotone operator. (The function f in (26) can, without loss of generality be taken to be lower semi-continuous).

The proof of Theorem 3 will not be repeated here. We shall, be concerned instead with the question of the maximality of a subdifferential ∂f. By definition, ∂f is a *maximal* monotone operator (resp. maximal cyclically monotone operator) if there does not exist a monotone (resp. cyclically monotone) operator T whose graph

$$G(T) = \{(x, x^*) \in X \times X^* \mid x^* \in T(x)\}$$

properly includes the graph of ∂f.

THEOREM 4. *If f is a lower semi-continuous proper convex function on X, then $\partial f : X \longrightarrow X^*$ is a maximal monotone operator.*

This result was stated as Theorem 4 of [25]. However H. Brézis has brought to our attention the fact that an oversight occurs in the proof given in [25]. (The penultimate sentence of the proof ignores the fact that x^* depends on ε as well as on x, so that $\| x^* \|$ might conceivably increase without bound as ε decreases to 0). Fortunately it is possible to give alternative proofs of Theorem 4 which avoid this difficulty. One of these proofs will be published in [30]. A somewhat different proof will be sketched here.

PROOF OF THEOREM 4. We already know that ∂f is a monotone operator. Thus, given u and u^* such that $u^* \notin \partial f(u)$, we must show that there exist x and x^* such that $x^* \in \partial f(x)$ and

$$\langle x - u, x^* - u^* \rangle < 0.$$

Replacing f by the lower semi-continuous proper convex function

$$g(x) = f(x + u) - \langle x, u^* \rangle$$

if necessary, we can reduce the argument to the case where $u = 0$ and $u^* = 0$. We assume therefore that $0 \notin \partial f(0)$, and we argue towards the conclusion that there exist x and x^* such that $x^* \in \partial f(x)$ and $\langle x, x^* \rangle < 0$.

It is instructive to consider first the case where X is reflexive, since the argument there is much simpler. Let $j(x) = (1/2) \| x \|^2$. Then $f + j$ is a lower semi-continuous proper convex function on X. It is not a difficult exercise to show that all the (convex) level sets of the form

$$\{x \in X \mid (f + j)(x) \leq \alpha\}, \qquad \alpha \in R,$$

are bounded in X, and consequently by the lower semi-continuity of $f + j$ and the fact that X is reflexive) weakly compact. This implies that $f+j$ attains its minimum over X at a certain x. We then have $0 \in \partial(f+j)(x)$, so that by Theorem 1

$$0 \in \partial f(x) + \partial j(x),$$

i. e. there exists an $x^* \in \partial f(x)$ such that $-x^* \in \partial j(x)$. In view of the nature of $\partial j(x)$, as described in Example 2 of § 1, the latter means that

$$(27) \qquad -\langle x, x^* \rangle = \| x \| \cdot \| x^* \|, \quad \| x^* \| = \| x \|.$$

If $x = 0$, we would have $x^* = 0$ by (27), contrary to $0 \notin \partial f(0)$. Therefore

$$-\langle x, x^* \rangle = \| x \|^2 > 0$$

by (27), and we are done.

To prove Theorem 4 in the case where X is not reflexive, we make use of weak* compactness in X^* by applying the argument just given to the conjugate function f^* defined by

$$f^*(x^*) = \sup \{ \langle x, x^* \rangle - f(x) \mid x \in X \}.$$

This is a proper convex function on X^* which is actually lower semi-continuous in the weak* topology (see [20]). Thus there exist $x^* \in X^*$ and $x^{**} \in X^{**}$ such that

$$(28) \qquad x^{**} \in \partial f^*(x^*) \text{ and } \langle x^{**}, x^* \rangle < 0,$$

The proof is completed by invoking the following general fact, whose proof (not elementary) is detailed in [30]: if $x^{**} \in \partial f^*(x^*)$, there exists a bounded net $(x_i \mid i \in I)$ in X converging to x^{**} in the weak** topology of X^{**}, as well as a net $(x_i^* \mid i \in I)$ (with the same partially ordered index set I) converging to x^* in the strong topology of X^*, such that $x_i^* \in \partial f(x_i)$ for every $i \in I$. (Here X is identified in the canonical way with a subspace of X^{**} which is dense in the weak** topology, i. e. the weak topology induced on X^{**} by X^*). For such nets we have

$$x_i^* \in \partial f(x_i) \text{ and } \langle x_i, x_i^* \rangle < 0$$

for some i by (28), as desired.

Any maximal monotone operator which is cyclically monotone is, of course, in particular a maximal cyclically monotone operator. On the other hand, any cyclically monotone operator can be imbedded in a maximal cyclically monotone operator by Zorn's lemma. Thus from Theorem 3 and Theorem 4 we have :

THEOREM 3′. *Let $T: X \to X^*$ be a multivalued mapping. In order that there exist a lower semi-continuous proper convex function f on X such that $T = \partial f$, it is necessary and sufficient that T be a maximal cyclically monotone operator.*

It can be shown, incidentally, that the f in Theorem 3′ is determined by T uniquely up to an additive constant; see [30].

From Theorem 4 and Theorem 3′, we may conclude that the multivalued mappings described in the six examples at the end of § 1 are maximal monotone operators and at the same time maximal cyclically monotone operators, provided that K is closed in Example 3 and Example 4.

4. Monotone operators associated with minimax problems.

If a monotone operator $T: X \to X^*$ happens to be the (single-valued) gradient operator ∇f for some real-valued Gâteaux differentiable function f on X, this f must be convex. Indeed, the monotonicity inequality

$$\langle x_1 - x_0 , \nabla f(x_1) - \nabla f(x_0) \rangle \geq 0$$

implies that

$$\langle x_1 - x_0 , \nabla f(x_0) \rangle \leq \langle x_1 - x_0 , \nabla f(x_1) \rangle$$

for every x_0 and x_1, and hence that

$$\langle u, \nabla f(x + \tau_1 u) \rangle \leq \langle u, \nabla f(x + \tau_2 u) \rangle$$

whenever $x \in X$, $u \in X$, $\tau_1 < \tau_2$. In other words, for any x and u the function

$$g(\tau) = f(x + \tau u), \; \tau \in R,$$

has a non-decreasing derivative

$$g'(\tau) = \langle u, \nabla f(x + \tau u) \rangle ,$$

implying that g is a convex function on R. The restriction of f to each line in X being convex, f is itself convex.

Theorem 3 may therefore be interpreted as saying that a monotone operator is a « generalized gradient operator » if and only if it is actually cyclically monotone. Cyclic monotonicity is thus an analogue of the classical condition that a continuously differentiable mapping $T: R^m \longrightarrow R^m$ is the gradient of some function if and only if its Jacobian matrix of first derivatives is symmetric at each point.

In the case where $T: X \longrightarrow X^*$ is a monotone operator of the form ∂f, the relation $0 \in T(x)$, which is a primary object of study in the theory of monotone operators, reduces to $0 \in \partial f(x)$ and thus describes the solutions x to a certain variational problem, namely that of minimizing the proper convex function f over X. One might get the impression from the facts described above, however, that, when T is not of the form ∂f, the points x such that $0 \in T(x)$ do not correspond suitably to any extremum and therefore cannot be characterized in terms of any variational principle. Our purpose here is to point out that this is not necessarily true. There exist maximal monotone operators T, which are not the subdifferentials of convex functions, and yet for which the points x satisfying $0 \in T(x)$ are, in a natural way, the solutions to certain variational problems.

The variational problems in question are *minimax problems*. Suppose that the Banach space X is of the form $Y \oplus Z$, where Y and Z are Banach spaces with duals Y^* and Z^*, respectively. (Then X^* can be identified with $Y^* \oplus Z^*$, so that

$$\langle x, x^* \rangle = \langle y, y^* \rangle + \langle z, z^* \rangle$$

for $x = (y, z)$ and $x^* = (y^*, z^*)$). Let $L(y, z)$ be a real-valued function of $y \in Y$ and $z \in Z$. A point $x = (y, z)$ is said to be a *saddle-point* of L (with respect to maximizing over Y and minimizing over Z) if the maximum of $L(\cdot, z)$ over Y is achieved at y, and at the same time the minimum of $L(y, \cdot)$ over Z is achieved at z. It can be demonstrated that in this event

$$L(y, z) = \sup_{u \in Y} \inf_{v \in Z} L(u, v) = \inf_{v \in Z} \sup_{u \in Y} L(u, v).$$

This extremum (when it exists unambiguously) is called the *minimax* of L. Saddle-points (if they exist) are regarded as the points where the minimax of L is attained, i. e. the solutions to the minimax problem for L.

If L is Gâteaux differentiable in y and z, the saddle points of L can be characterized in terms of gradients. Let $\nabla_1 L(y, z) \in Y^*$ denote the gradient of $L(\cdot, z)$ at y as a function on Y, and let $\nabla_2 L(y, z) \in Z^*$ denote the gradient of $L(y, \cdot)$ at z as a function on Z. A necessary condition for $x = (y, z)$ to be a saddle-point is that

$$\nabla_1 L(y, z) = 0 \quad \text{and} \quad \nabla_2 L(y, z) = 0, \tag{29}$$

or in other words, in terms of the gradient

$$\nabla L(y, z) = (\nabla_1 L(y, z), \nabla_2 L(y, z)) \in X^*,$$

simply

$$\nabla L(y, z) = 0.$$

If $L(y, z)$ is convex as a function of z for each y and concave as a function of y for each z (i. e. $-L(y, z)$ is convex as a function of y for each z), these conditions are not only necessary but sufficient.

Of course L itself is not a convex function on X in the latter case, so $\nabla L : X \longrightarrow X^*$ is not a monotone operator.

But consider the mapping $T : X \longrightarrow X^*$ defined by

$$T(x) = (-\nabla_1 L(y, z), \ \nabla_2 L(y, z)), \quad x = (y, z). \tag{30}$$

The saddle-point condition (29) can just as well be written as

$$T(x) = 0.$$

It turns out that (when L is concave-convex as described above) T *is a maximal monotone operator*. The proof of this is given in [27], where it is also shown how to generalize the result to concave-convex functions L which are not differentiable, and which may even have the values $+\infty$ and $-\infty$. Here we shall content ourselves merely with showing why T is monotone.

The monotonicity of a single-valued mapping T means that

$$\langle x_1 - x_0, T(x_1) - T(x_0) \rangle \geq 0, \ \forall \ x_0, x_1 .$$

Thus, for the T defined by (30), monotonicity means that

$$\begin{aligned} \langle y_1 - y_0, -\nabla_1 L(y_1, z_1) + \nabla_1 L(y_0, z_0) \rangle \\ + \langle z_1 - z_0, \nabla_2 L(y_1, z_1) - \nabla_2 L(y_0, z_0) \rangle \geq 0 \end{aligned} \tag{31}$$

for all y_0, y_1, z_0, z_1. To prove (31), we need only observe that the inequalities

$$L(y_0, z_1) \geq L(y_0, z_0) + \langle z_1 - z_0, \nabla_2 L(y_0, z_0) \rangle,$$

$$L(y_1, z_0) \geq L(y_1, z_1) + \langle z_0 - z_1, \nabla_2 L(y_1, z_1) \rangle,$$

hold by the convexity of the functions $L(y_0, \cdot)$ and $L(y_1, \cdot)$ on Z, whereas the inequalities

$$-L(y_1, z_0) \geq -L(y_0, z_0) + \langle y_1 - y_0, -\nabla_1 L(y_0, z_0) \rangle,$$

$$-L(y_0, z_1) \geq -L(y_1, z_1) + \langle y_0 - y_1, -\nabla_1 L(y_1, z_1) \rangle,$$

hold by the convexity of the functions $-L(\cdot, z_0)$ and $-L(\cdot, z_1)$ on Y. Adding these four inequalities, we get (31).

Minimax problems may also be considered in which the extremum is taken over a certain subset of $X = Y \oplus Z$, rather than over the whole space. Let C and D be non-empty subsets of Y and Z respectively, and let

$$K = C \oplus D = \{x = (y, z) \mid y \in C, z \in D\}.$$

A point $x = (y, z)$ is said to be a *saddle-point of L relative to K* if the maximum of $L(\cdot, z)$ over C is achieved at y, while the minimum of $L(y, \cdot)$ over D is achieved at z. Suppose that C and D are convex, and that L is concave-convex and differentiable as above. The conditions for a maximum or minimum can then be analyzed in terms of subgradients as in § 2. One sees in this way that $x = (y, z)$ is a saddle-point of L relative to K if and only if $\nabla_1 L(y, z)$ is a normal to C at y and $-\nabla_2 L(y, z)$ is a normal to D at z. Obviously an element $x^* = (y^*, z^*)$ is normal to K at $x = (y, z)$ if and only if y^* is normal to C at y and z^* is normal to D at z. *It follows that $x = (y, z)$ is a saddle-point of L relative to K if and only if $-T(x)$ is a normal to K at x, where T is given by* (30). Put another way: the saddle points of L relative to K are the points x satisfying

$$0 \in T(x) + \partial\delta(x \mid K).$$

5. Variational inequalities and sums of monotone operators.

If T_1 and T_2 are monotone operators from X to X^*, the sum $T_1 + T_2$ is defined by

$$(T_1 + T_2)(x) = T_1(x) + T_2(x)$$
$$= \{x_1^* + x_2^* \mid x_1^* \in T_1(x), x_2^* \in T_2(x)\}.$$

It is easy to see that $T_1 + T_2$ is again a monotone operator. Namely, if

$$x^* \in (T_1 + T_2)(x) \quad \text{and} \quad y^* \in (T_1 + T_2)(y),$$

there exist elements

$$x_1^* \in T_1(x), x_2^* \in T_2(x), \; y_1^* \in T_1(y), y_2^* \in T_2(y),$$

such that

$$x_1^* + x_2^* = x^* \quad \text{and} \quad y_1^* + y_2^* = y^*.$$

We then have

$$\langle x - y, x^* - y^* \rangle = \langle x - y, x_1^* - y_1^* \rangle + \langle x - y, x_2^* - y_2^* \rangle \geq 0$$

by the monotonicity of T_1 and T_2. Of course

$$D(T_1 + T_2) = D(T_1) \cap D(T_2), \tag{32}$$

where, for a multivalued mapping $T : X \to X^*$, $D(T)$ denotes the set of all x such that $T(x) \neq \emptyset$.

A deeper question is this: if T_1 and T_2 are maximal monotone operators, is $T_1 + T_2$ again a *maximal* monotone operator? The answer in general has to be no since, for example, if $D(T_1)$ does not meet $D(T_1)$, the graph of $T_1 + T_2$ is empty and hence certainly not maximal. At the very least, some kind of condition about the way $D(T_1)$ and $D(T_2)$ overlap is needed, if maximality is to be preserved when $T_1 + T_2$ is formed.

The preservation of maximality under addition is of interest in the study of variational inequalities, among other things. Let $T : X \to X^*$ be a monotone operator, and let $K \subset X$ be a non-empty convex set. The condition

$$x^* \in T(x) \text{ and } -x^* \text{ is a normal to } K \text{ at } x \tag{33}$$

is called the *variational inequality* for T and K, since, in the case where T is single-valued, (33) can be written as

$$x \in K \text{ and } \langle y - x, T(x) \rangle \geq 0, \; \forall \, y \in K. \tag{34}$$

Variational inequalities have been investigated by Browder, Lions, Hartmann and Stampacchia because of important applications to problems of partial differential equations (see [12] for some examples).

However, variational inequalities for operators of the form $T = \partial f$ were studied earlier in [22] in the spirit of § 2; see also [23].

When $T = \partial f$, the variational inequality for T and K is the condition for a minimum of f relative to K (Corollary 1 of Theorem 1). In certain other cases, as seen at the end of § 4, a variational inequality can be the condition for a saddle-point of a concave-convex function L relative to K, even though T is not actually the gradient of L.

The variational inequality for T and K can also be written equivalently as the condition that

$$(35) \qquad 0 \in S(x), \quad S = T + \partial\delta(\cdot \mid K).$$

If K is closed, the indicator $\delta(\cdot \mid K)$ of K is a lower semicontinuous proper convex function on X, and consequently by Theorem 4 the multivalued mapping $\partial\delta(\cdot \mid K)$ is a maximal monotone operator from X to X^*. If the monotone operator T is likewise maximal, one may hope that the sum S too will be maximal. If it is, then the study of the variational inequality for T and K is reduced to the more fundamental study of the condition $0 \in S(x)$ for a maximal monotone operator S.

The main advantage of the reduction just described, when it is possible, is that it leads to a unified theory of existence of solutions. As far as characterization of solutions and regularity of solutions are concerned, one is interested, not so much in lumping T and $\partial\delta(\cdot \mid K)$ together, as in decomposing $\partial\delta(\cdot \mid K)$ further in the manner of Theorem 2 or Corollaries 2 and 3 of Theorem 1.

Observe incidentally that, even if T is single-valued and everywhere defined on X, and $K(\neq X)$ has a very regular boundary, the monotone operator S in (35) will be multivalued, and $D(S)$ will be a proper subset of X. This is one of the principle motivations for developing the theory of *multivalued* maximal monotone operators.

In order to see what theorems might be possible concerning the maximality of a sum of monotone operators, it is helpful first to investigate the case where the operators are the subdifferentials of convex functions, since knowledge and intuition are so much greater there. (This is a good heuristic method to keep in mind).

Some remarks about the relationship between dom f and $D(\partial f)$, for a lower semi-continuous proper convex function f on X, are necessary at the outset. As mentioned in § 1, $\partial f(x) = \emptyset$ when $x \in \operatorname{dom} f$ and f is continuous at x. It turns out that lower semi-continuity of f

implies that f is continuous at every interior point of dom f [24]. Furthermore, it has been established in [4] that, even if the interior of the convex set dom f is empty, $D(\partial f)$ is dense in dom f. Thus in general

$$\text{int}\ (\text{dom}\ f) \subset D(\partial f) \subset \text{dom}\ f \subset \text{cl}\ D(\partial f). \tag{36}$$

Suppose now that f_1 and f_2 are lower semi-continuous proper convex functions on X. The hypothesis of Theorem 1 can then be translated into a condition on $D(\partial f_1)$ and $D(\partial f_2)$. The points $x \in \text{dom}\ f_1$ at which f_1 is continuous are simply the interior points of dom f_1, and we have

$$\text{int}\ (\text{dom}\ f_1) = \text{int}\ D(\partial f_1)$$

by (36). Moreover, the latter set meets dom f_2 if and only if it meets $D(\partial f_2)$, because $D(\partial f_2)$ is dense in dom f_2. In view of Theorem 3′ and the fact that $f_1 + f_2$ is another lower semi-continuous convex function, we may draw the following conclusion from Theorem 1.

THEOREM 5. *If T_1 and T_2 are maximal cyclically monotone operators from X to X^* such that*

$$D(T_2) \cap \text{int}\ D(T_1) \neq \emptyset,$$

then $T_1 + T_2$ is again a maximal monotone operator.

This result leads one to conjecture that a similar theorem might be true for arbitrary maximal monotone operators. The conjecture turns out to be true, at least in the reflexive case:

THEOREM 5′. *If T_1 and T_2 are maximal monotone operators from X to X^* such that*

$$D(T_2) \cap \text{int}\ D(T_1) \neq \emptyset,$$

and if X is reflexive, then $T_1 + T_2$ is again a maximal monotone operator.

COROLLARY. *Let $T: X \to X^*$ be a maximal monotone operator, and let $K \subset X$ be a non-empty closed convex set. Suppose that X is reflexive, and that*

$$D(T) \cap \text{int}\ K \neq \emptyset \quad \text{or} \quad K \cap \text{int}\ D(T) \neq \emptyset. \tag{37}$$

Then the monotone operator

$$S = T + \partial\delta(\cdot \mid K)$$

is maximal.

Theorem 5′, whose proof is given in [29], is a substantial generalization of earlier results of Lescarret and Browder; see [10] for a discussion of the literature on the subject. For more details on the applications of the corollary to the existence theory for variational inequalities, see [29].

The preceding corollary covers many important cases of variational inequalities, but there are certain cases studied by Browder and Stampacchia in which (37) does not necessarily hold. For these, one needs a special maximality result, whose proof is easy enough to be stated here in its entirety.

THEOREM 6 [29]. *Let $K \subset X$ be a non-empty closed convex set, and let $T: X \longrightarrow X^*$ be a monotone operator (not necessarily maximal) such that $D(T) \supset K$ and T is single-valued and hemi-continuous on K (i. e. continuous from line segments in K to the weak* topology of X^*). Then*

$$S = T + \partial\delta(\cdot \mid K)$$

is a maximal monotone operator.

Proof. Let $y \in X$ and $y^* \in X^*$ be such that

$$\langle y - x, y^* - x^* \rangle \geq 0 \quad \text{whenever } x^* \in S(x). \tag{38}$$

We must show that $y^* \in S(y)$. In view of the definition of S, (38) means that

$$0 \leq \langle y - x, y^* - T(x) \rangle - \langle y - x, u^* \rangle \quad \text{whenever } u^* \in \partial\delta(x \mid K). \tag{39}$$

Now $\partial\delta(x \mid K)$ is a convex cone, so if (39) holds for x and a given u^* it must also hold for λu^* in place of u^*, for any $\lambda \geq 0$. Therefore (39) implies that

$$\langle u - x, u^* \rangle \leq 0 \leq \langle y - x, y^* - T(x) \rangle \quad \text{whenever } u^* \in \partial\delta(x \mid K). \tag{40}$$

The left side of (40) can be written as

$$\langle y - x, 0 - u^* \rangle \geq 0.$$

Since this holds for every x and u^* such that $u^* \in \partial\delta(x \mid K)$, and since $\partial\delta(\cdot \mid K)$ is a maximal monotone operator by Theorem 4, we must have

$$0 \in \partial\delta(y \mid K), \quad \text{i.e. } y \in K.$$

Consider next the right side of (40), which holds for every $x \in K$. Fix any $x \in K$, and let

$$x_\lambda = (1-\lambda)y + \lambda x \in K, \quad 0 < \lambda \leq 1.$$

We have

$$0 \leq \langle y - x_\lambda, y^* - T(x) \rangle = \lambda \langle y - x, y^* - T(x_\lambda) \rangle$$

by (40). Dividing through by λ and taking the limit as λ goes to 0, which is possible by the hemicontinuity of T, we get

$$0 \leq \langle y - x, y^* - T(y) \rangle.$$

This has been verified for arbitrary $x \in K$, and we may therefore conclude that $T(y) - y^*$ is normal to K at y. Thus

$$y^* \in T(y) + \partial\delta(y \mid K),$$

and we are through.

6. Domains and ranges of maximal monotone operators.

We have seen above that, in most cases of interest, the solutions to a variational inequality can be described as the points $x \in X$ such that $0 \in S(x)$, where S is a certain maximal monotone operator. When does such a point x exist?

In general, given a maximal monotone operator $T: X \to X^*$, we may ask: for which choices of $x^* \in X^*$ does there exist an $x \in X$ such that $x^* \in T(x)$? By definition, to answer this question, we have to explore various properties of the range of T, i.e. the

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

Among the things we would like to know are: what conditions on T ensure that $0 \in R(T)$, or that $R(T) = X^*$? We might also be interested in whether a given point belongs to int $R(T)$, or whether $R(T)$ is dense in X^* in some topology. We are led thus to a broad investigation of the geometric nature of $R(T)$.

The investigation of the geometric nature of the effective domain $D(T)$ of a maximal monotone operator $T: X \to X^*$ is likewise worth undertaking, for example in connection with the domain condition in the hypothesis of Theorem 5′.

It is useful to observe that the theory of ranges and the theory of domains are equivalent in the reflexive case. If X is reflexive and T is a maximal monotone operator from X to X^*, then the operator T^{-1} defined by

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

is obviously a maximal monotone operator from X^* to $X = X^{**}$. Furthermore,

$$D(T^{-1}) = R(T) \quad \text{and} \quad R(T^{-1}) = D(T).$$

Some known facts about $D(T)$ and $R(T)$ will now be mentioned without proof. We begin with recent results about convexity properties.

THEOREM 7 [26]. *Let $T: X \to X^*$ be a maximal monotone operator. If X is reflexive, then* $\operatorname{cl} D(T)$ *and* $\operatorname{cl} R(T)$ *are convex sets.*

Here cl denotes closure with respect to the *norm* topology.

Suppose, for the sake of illustration, that T is a single-valued hemicontinuous monotone operator with $D(T) = X$. Then T is maximal, as has been proved by Browder [6] (the maximality also follows from Theorem 6 above with $K = X$). Then $\operatorname{cl} R(T)$ is convex by Theorem 7. Thus, either there exist elements $x^* \in X^*$ of arbitrarily small norm for which the equation $T(x) = x^*$ has a solution x, or 0 can be strictly separated from $R(T)$ by some closed hyperplane, i. e. there exists a $y^* \in X^*$ and an $\varepsilon > 0$ such that

$$\langle T(x), y^* \rangle \leq -\varepsilon, \quad \forall x \in X.$$

Theorem 7 can be improved in the case where X satisfies the following condition: there exists an equivalent norm on X which is Fréchet differentiable except at the origin and whose polar norm on X^* is Fréchet differentiable at the origin. It is well known that this condition implies X is reflexive. The Banach spaces satisfying this condition will be called *strongly reflexive.*

The class of strongly reflexive spaces obviously includes all L^p spaces for $1 < p < \infty$, in particular all Hilbert spaces. It also includes all separable reflexive Banach spaces; this has been shown by Asplund [1] by means of a theorem of Kadec.

(We would like to point out that the condition above is satisfied for *every* reflexive Banach space, if Fréchet differentiability is weakened to Gâteaux differentiability; see [1]. Equivalently, any reflexive Banach space X can be renormed in such a way that the unit balls of X and X^* are strictly convex. This fact enters into the proof of Theorem 7).

THEOREM 7′ [26]. *Let $T: X \longrightarrow X^*$ be a maximal monotone operator. If X is strongly reflexive, then $D(T)$ and $R(T)$ are virtually convex sets.*

A set C is by definition *virtually convex* if, given any relatively (strongly) compact subset A of the convex hullof C (for example, A could be taken to be the convex hull of any finite subset of C), and given any $\varepsilon > 0$, there exists a (strongly) continuous mapping $p: A \longrightarrow C$ such that

$$\| p(u) - u \| < \varepsilon, \qquad \forall\, u \in A.$$

For example, let X be a separable reflexive Banach space, let K^* be a non-empty closed (but not necessarily bounded) subset of X^*, and let C be the set of all $x \in X$ such that the linear function $\langle x, \cdot \rangle$ is bounded above on K^* and attains its maximum. Then C is virtually convex, because $C = D(\partial f)$ for the f given by Example 6 of § 1, and ∂f is a maximal monotone operator by Theorem 4.

As a more special case, to get some intuitive idea of what a virtually convex set may be like, let $X = L^2(G)$, where G is a bounded region of R^n, and let K^* be the set of all non-negative functions in $X^* = L^2(G)$ whose integral is 1 (density functions for probability distributions). Then C consists of all the *flat-topped* functions in $L^2(G)$, i.e. functions which attain their essential supremum on a set of positive measure. This C is therefore virtually convex, but it is clearly not convex, since the « average » of two flat-topped functions need not be flat-topped.

Here is another convexity result which does *not* depend on reflexivity.

THEOREM 8 [28]. *Let $T: X \longrightarrow X^*$ be a maximal monotone operator. If the convex hull of $D(T)$ has a non-empty interior, then* int $D(T)$ *is a non-empty open convex set whose closure contains all of* $D(T)$.

Corollary 1. *Let* $T : X \longrightarrow X^*$ *be a maximal monotone operator. If* X *is reflexive and the convex hull of* $R(T)$ *has a non-empty interior, then* int $R(T)$ *is a non empty open convex set whose closure contains all of* $R(T)$.

Theorem 8 is closely related to results about local boundedness, T being *locally bounded* at a point x if there exists a neighborhood of x such that

$$T(U) = \bigcup \{ T(u) \mid u \in U \}$$

is a bounded subset of X^*. It is proved in [28] that, if X is reflexive, a maximal monotone operator $T : X \longrightarrow X^*$ is locally bounded at a given point x if and only if x is not a boundary point of $D(T)$. This fact, in conjunction with Corollary 1, yields

Corollary 2 [28]. *Let* $T : X \longrightarrow X^*$ *be a maximal monotone operator. Suppose that* X *is reflexive. In order that* $R(T) = X^*$, *it is necessary and sufficient that, whenever* $x_1, x_2, \ldots,$ *is an unbounded sequence in* $D(T)$ *and* $x_i^* \in T(x_i)$ *for every* i, *then the sequence* $x_1^*, x_2^*, \ldots,$ *has no strongly convergent subsequence.*

In the finite-dimensional case, the necessary and sufficient condition in Corollary 2 is equivalent to a sufficient condition previously established by Browder [6]: whenever $x_1, x_2, \ldots,$ is an unbounded sequence in $D(T)$ and $x_i^* \in T(x_i)$ for every i, then the sequence $x_1^*, x_2^*, \ldots,$ is unbounded.

Another condition for $R(T)$ to be all of X^* can be stated in terms of coercivity. A multivalued mapping $T : X \longrightarrow X^*$ is said to be *coercive* if

$$\lim_{\alpha \to +\infty} \inf_{x} \{ \langle x, x^* \rangle \mid x^* \in T(x), \| x \| \geq \alpha \} = +\infty .$$

(Here we use the convention that the infimum of the empty set is $+\infty$.)

Theorem 9. *Let* : $X \longrightarrow X^*$ *be a maximal monotone operator. If* X *is reflexive and* T *is coercive, then* $R(T) = X^*$.

This result was originally developed independently by Minty [18] and Browder [5] in the case of single-valued T with $D(T) = X$, and after various improvements it was extended by Browder [10] to the case of general multivalued T with $0 \in D(T)$ and X and X^* strictly convex. The minor step of removing the latter restrictions was carried out in [29].

Finally we state a condition for simple membership in $R(T)$. Of course, a given x^* belongs to $R(T)$ if and only if 0 belongs to $R(T')$, where

$$T'(x) = T(x) - x^*, \qquad \forall\, x \in X.$$

If T is maximal monotone, then so is T'. Thus it suffices to consider conditions granteeing that 0 belongs to $R(T)$.

THEOREM 10. *Let $T: X \longrightarrow X^*$ be a maximal monotone operator. Suppose that X is reflexive. If there exists some $a \in D(T)$ and $\alpha > 0$ such that*

$$\langle x - a, x^* \rangle \geq 0 \text{ whenever } x^* \in T(x),\ \| x \| \geq \alpha,$$

then $0 \in R(T)$.

Theorem 10 is an easy extension of some results of Browder [6,10], a proof is given in [29]. Theorem 9 is an immediate corollary of Theorem 10.

Existence theorems for variational inequalities can be deduced from Theorems 9 and 10 as a simple exercise using the results of § 5; see [29].

REFERENCES

[1] E. ASPLUND, *Averaged norms*, Israel J. Math. 5 (1967), 227-233.

[2] E. ASPLUND, *Positivity of duality mappings*, Bull. Amer. Math. Soc. 73 (1967), 200-203.

[3] E. ASPLUND, and R. T. ROCKAFELLAR, *Gradients of convex functions*, Trans Amer. Math. Soc. 139, (May, 1969).

[4] A. BRØNDSTED and R. T. ROCKAFELLAR, *On the subdifferentiability of convex functions*, Proc. Amer Math. Soc. 16 (1965), 405-411.

[5] F. E. BROWDER, *Nonlinear elliptic boundary value problems*, Bull. Amer. Math. Soc. 69 (1963), 862-874.

[6] F. E. BROWDER, *Multivalued monotone nonlinear mappings and duality mappings in Banach spaces*, Trans. Amer. Math. Soc. 118 (1965), 338-351.

[7] F. E. BROWDER, *Nonlinear monotone operators and convex sets in Banach spaces*, Bull. Amer. Math. Soc. 71 (1965), 780-785.

[8] F. E. BROWDER, *On the unification of the calculus of variations and the theory of monotone operators*, Proc. Nat. Acad. Sci. U.S. 56 (1966), 419-425.

[9] F. E. BROWDER, *Existence and approximation of solutions of nonlinear variational inequalities*, Proc. Nat. Acad. Sci. U. S. 56 (1966), 1080-1086.

[10] F. E. BROWDER, *Nonlinear maximal monotone operators in Banach space*, Math. Annalen 175 (1968), 89-113.

[11] F. E. BROWDER, *Nonlinear variational inequalities and maximal monotone mappings in Banach spaces*. Math. Annalen, to appear.

[12] P. HARTMAN and G. STAMPACCHIA, *On some nonlinear elliptic functional-differential equations*, Acta. Math. 115 (1966), 271-310.

[13] R. I. KACUROVSKI, *On monotone operators and convex functionals*, Usp. Mat. Nauk. 15 (1960), 213-215.

[14] C. LESCARRET, *Cas d'addition des applications maximales dans un espace de Hilbert*, C. R. Acad. Sci Paris 261 (1965), 1160-1163.

[15] C. LESCARRET, *Sur la sous-différentiabilité d'une somme de fonctionnelles convexes semi-continues inférieurment*, C. R. Acad. Sci. Paris 262 (1966), 443-446.

[16] J. L. LIONS and G. STAMPACCHIA, *Inéquations variationelles noncoercives*, C. R. Acad. Sci. Paris. 261 (1965), 25-27.

[17] J. L. LIONS and G. STAMPACCHIA, *Variational inéqualities*, Comm. Pure Appl. Math. XX (1967), 493-519.

[18] G. J. MINTY, *On a «monotonicity» method for the solution of nonlinear equations in Banach spaces*, Proc. Nat. Acad. Sci. U. S. 50 (1963), 1038-1041.

[19] J. J. MOREAU, *Sur la fonction polaire d'une fonction semicontinue supérieurment*, C. R. Acad. Sci. Paris 258 (1964), 1128-1131.

[20] J. J. MOREAU, *Fonctionelles convexes*, mimeographed lecture notes, Collège de France, 1967.

[21] U. MOSCO, *A remark on a theorem of F. E. Browder*, J. Math. Anal. Appl. 20 (1967), 90-93.

[22] R. T. ROCKAFELLAR, *Convex functions and dual extremum problems,* thesis, Harvard, 1963.

[23] R. T. ROCKAFELLAR, *An extension of Fenchel's duality theorem for convex functions,* Duke Math. J. 33 (1966). 81-90.

[24] R. T. ROCKAFELLAR, *Level sets and continuity of conjugate convex functions,* Trans Amer. Math. Soc. 123 (1966), 46-63.

[25] R. T. ROCKAFELLAR, *Characterization of the subdifferentials of convex functions,* Pacific J. Math. **17** (1966), 497-510.

[26] R. T. ROCKAFELLAR, *On the virtual convexity of the domain and range of a nonlinear maximal monotone operator,* Math. Annal., to appear.

[27] R. T. ROCKAFELLAR, *Monotone operators associated with saddle-functions and minimax problems,* in Nonlinear Functional Analysis, *Proceedings of Symposia in Pure Mathematics,* Am. Math. Soc., 1969.

[28] R. T. ROCKAFELLAR, *Local boundedness of nonlinear monotone operators,* Michigan Math. J., to appear.

[29] R. T. ROCKAFELLAR, *On the maximality of sums of nonlinear monotone operators,* Trans. Amer. Math. Soc., to appear.

[30] R. T. ROCKAFELLAR, *On the maximal monotonicity of subdifferential mappings,* Michigan Math. J., to appear.

[31] R. T. ROCKAFELLAR, *Convex Analysis,* Princeton Univ. Press, 1969.

[32] G. STAMPACCHIA, *Forms bilineares coercitives sur les ensembles convexes,* C. R. Acad. Sci. Paris 258 (1964) 4413-4416.