MONOTONE OPERATORS AND THE PROXIMAL POINT ALGORITHM*

R. TYRRELL ROCKAFELLAR†

Abstract. For the problem of minimizing a lower semicontinuous proper convex function f on a Hilbert space, the proximal point algorithm in exact form generates a sequence $\{z^k\}$ by taking z^{k+1} to be the minimizer of $f(z) + (1/2c_k)\|z - z^k\|^2$, where $c_k > 0$. This algorithm is of interest for several reasons, but especially because of its role in certain computational methods based on duality, such as the Hestenes-Powell method of multipliers in nonlinear programming. It is investigated here in a more general form where the requirement for exact minimization at each iteration is weakened, and the subdifferential ∂f is replaced by an arbitrary maximal monotone operator T. Convergence is established under several criteria amenable to implementation. The rate of convergence is shown to be "typically" linear with an arbitrarily good modulus if c_k stays large enough, in fact superlinear if $c_k \to \infty$. The case of $T = \partial f$ is treated in extra detail. Application is also made to a related case corresponding to minimax problems.

1. Introduction. Let H be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$. A multifunction $T : H \to H$ is said to be a *monotone operator* if

$$(1.1) \qquad \langle z - z', w - w' \rangle \geqq 0 \quad \text{whenever} \quad w \in T(z), w' \in T(z').$$

It is said to be *maximal monotone* if, in addition, the graph

$$(1.2) \qquad G(T) = \{(z, w) \in H \times H \,|\, w \in T(z)\}$$

is not properly contained in the graph of any other monotone operator $T' : H \to H$.

Such operators have been studied extensively because of their role in convex analysis and certain partial differential equations. A fundamental problem is that of determining an element z such that $0 \in T(z)$.

For example, if T is the subdifferential ∂f of a lower semicontinuous convex function $f : H \to (-\infty, +\infty]$, $f \not\equiv +\infty$, then T is maximal monotone (see Minty [15] or Moreau [18]), and the relation $0 \in T(z)$ means that $f(z) = \min f$. The problem is then one of minimization subject to implicit constraints (the points where $f(z) = +\infty$ being effectively forbidden from the competition).

The basic case of *variational inequalities* corresponds to

$$(1.3) \qquad T(z) = \begin{cases} T_0(z) + N_D(z) & \text{if} \quad z \in D, \\ \varnothing & \text{if} \quad z \notin D, \end{cases}$$

where D is a nonempty closed convex subset of H, $T_0 : D \to H$ is single-valued, monotone and hemicontinuous (i.e. continuous along each line segment in H with respect to the weak topology), and $N_D(z)$ is the *normal cone* to D at z:

$$N_D(z) = \{w \in H \,|\, \langle z - u, w \rangle \geqq 0, \forall u \in D\}.$$

The maximal monotonicity of such a multifunction T was proved by Rockafellar [27]. The relation $0 \in T(z)$ reduces to $-T_0(z) \in N_D(z)$, or the so-called variational inequality:

(1.4) $z \in D$ and $\langle z - u, T_0(z) \rangle \leq 0$ for all $i \in D$.

If D is a cone, this condition can be written as

$$z \in D, \quad -T_0(z) \in D^\circ \quad \text{(the polar of } D\text{)}, \quad \text{and} \quad \langle z, T_0(z) \rangle = 0,$$

and the problem of finding such a z is an important instance of the well-known *complementarity problem* of mathematical programming.

Another example corresponds to minimax problems. Let H be a product of Hilbert spaces H_1 and H_2, and let $L : H \to [-\infty, +\infty]$ be such that $L(x, y)$ is convex in $x \in H_1$ and concave in $y \in H_2$. For each $z = (x, y)$, let $T_L(z)$ be the set of all $w = (v, u)$ such that

$$L(x', y) - \langle x', v \rangle + \langle y, u \rangle \geq L(x, y) - \langle x, v \rangle + \langle y, u \rangle$$

(1.5) $\geq L(x, y') - \langle x, v \rangle + \langle y', u \rangle$

$$\text{for all} \quad x' \in H_1, \quad y' \in H_2.$$

If L is "closed and proper" in a certain general sense, then T_L is maximal monotone; see Rockafellar [24]. The global saddle points of L (with respect to minimizing in x and maximizing in y) are the elements $z = (x, y)$ such that $0 \in T_L(z)$.

In this paper, we study a fundamental algorithm for solving $0 \in T(z)$ in the case of an arbitrary maximal monotone operator T. The algorithm is based on the fact (see Minty [14]) that for each $z \in H$ and $c > 0$ there is a unique $u \in H$ such that $z - u \in cT(u)$, or in other words,

$$z \in (I + cT)(u).$$

The operator $P = (I + cT)^{-1}$ is therefore single-valued from all of H into H. It is also *nonexpansive*:

(1.6) $\|P(z) - P(z')\| \leq \|z - z'\|,$

and one has $P(z) = z$ if and only if $0 \in T(z)$. P is called the *proximal mapping* associated with cT, following the terminology of Moreau [18] for the case of $T = \partial f$.

The *proximal point algorithm* generates for any starting point z^0 a sequence $\{z^k\}$ in H by the approximate rule

(1.7) $z^{k+1} \approx P_k(z^k)$,   where $P_k = (I + c_k T)^{-1}$.

Here $\{c_k\}$ is some sequence of positive real numbers. In the case of $T = \partial f$, this procedure reduces to

(1.8) $z^{k+1} \approx \arg\min_z \phi_k(z),$

where

(1.9) $\phi_k(z) = f(z) + \frac{1}{2c_k} \|z - z^k\|^2$

(see § 4). For T corresponding to a convex-concave function L, it becomes

(1.10) $$(x^{k+1}, y^{k+1}) \approx \arg \min_{x,y} \max \Lambda_k(x, y),$$

where

(1.11) $$\Lambda_k(x, y) = L(x, y) + \frac{1}{2c_k}\|x - x^k\|^2 - \frac{1}{2c_k}\|y - y^k\|^2$$

(see § 5).

Results on the convergence of the proximal point algorithm have already been obtained by Martinet for certain cases where $c_k \equiv c$. He showed in [12], [13] that if T is of the form (1.3) with D bounded, and if true equality is taken in (1.7), then z^k converges in the weak topology to a particular z^∞ such that $0 \in T(z^\infty)$. Similarly if $T = \partial f$ and the level sets

$$\{z \in H | f(z) \leq \alpha\}, \qquad \alpha \in R,$$

are all weakly compact, in which event it is also true that $f(z^k) \downarrow f(z^\infty) = \min f$.

These results of Martinet are based on a more general theorem concerning operators V with the property

(1.12) $$\|V(z) - V(z')\|^2 \leq \|z - z'\|^2 - \|(I - V)(z) - (I - V)(z')\|^2.$$

This class includes $(I + cT)^{-1}$ (cf. Proposition 1(c) below). If $V : C \to C$ satisfies (1.12), where C is a nonempty, closed, bounded, convex subset of H, then for any starting point $z^0 \in C$ the sequence $\{z^k\}$ generated by $z^{k+1} = V(z^k)$ converges weakly to some fixed point of V. This theorem is a corollary of one of Opial [32] treating iterates of $\lambda I + (1 - \lambda)U$ when U is nonexpansive, $0 < \lambda < 1$. In fact, V satisfies (1.12) if and only if $V = \frac{1}{2}(I + U)$, where U is nonexpansive. Genel and Lindenstrauss [33] have recently furnished an example of such a mapping V for which $\{z^k\}$ does not converge strongly. However, this V does not appear to be of the form $(I + cT)^{-1}$ for $c > 0$ and T maximal monotone.

The question of whether the weak convergence established by Martinet can be improved to strong convergence thus remains open. The answer is known to be affirmative if $T = \partial f$ with f quadratic. This follows from a theorem of Krasnoselskii [10], as has been noted by Kryanev [11]. In the quadratic case, ∂f reduces to a densely defined, single-valued mapping of the form $x \to A(x) - b$, where A is a nonnegative, closed, self-adjoint linear operator. Then the relation $0 \in T(z)$ is equivalent to $A(z) = b$.

Strong convergence of the algorithm in its exact form with $z^{k+1} = P_k(z^k)$ is also assured if c_k is bounded away from zero and T is *strongly monotone* (*with modulus $\alpha > 0$*), i.e., in place of (1.1) one has

(1.13) $\langle z - z', w - w' \rangle \geq \alpha\|z - z'\|^2$ whenever $w \in T(z)$, $w' \in T(z')$.

Indeed, the latter condition means that $T' = T - \alpha I$ is monotone, and hence the mapping $P'_k = (I + c'_k T')^{-1}$ is nonexpansive for any $c'_k > 0$; taking $c'_k = c_k(1 + \alpha c_k)^{-1}$ one has

$$P'_k[(1 - \alpha c_k(1 + \alpha c_k)^{-1})I + c_k(1 + \alpha c_k)^{-1}T]^{-1} = [(1 + \alpha c_k)^{-1}(I + c_k T)]^{-1}$$

or

$$P_k(z) = P'_k((1 + \alpha c_k)^{-1}z) \quad \text{for all} \quad z,$$

so that the nonexpansiveness of P'_k yields

(1.14) $\|P_k(z) - P_k(z')\| \leq (1 + \alpha c_k)^{-1} \|z - z'\|$ for all $z, z' \in H$.

In particular, this implies P_k has a unique fixed point, which must then be the unique point z^∞ satisfying $0 \in T(z^\infty)$. One has

(1.15) $\|z^{k+1} - z^\infty\| = \|P_k(z^k) - P_k(z^\infty)\| \leq (1 + \alpha c_k)^{-1} \|z^k - z^\infty\|$ for all k,

so if $c_k \geq c > 0$ for all k sufficiently large the sequence $\{z^k\}$ converges to the solution z^∞ of the problem, not only strongly, but at least as fast as the linear rate with coefficient $(1 + \alpha c)^{-1} < 1$. If $c_k \to \infty$, the convergence is *superlinear*:

$$\lim_{k \to \infty} \frac{\|z^{k+1} - z^\infty\|}{\|z^k - z^\infty\|} = 0.$$

Unfortunately, the assumption that T is strongly monotone excludes some of the most important applications, such as to typical problems of convex programming, and it is important therefore to study convergence under weaker assumptions. Of course, from a practical point of view it is also essential to replace the equation $z^{k+1} = P_k(z^k)$ by a looser relation which is computationally implementable for a wide variety of problems.

Two general criteria for the approximate calculation of $P_k(z^k)$ are treated here:

(A) $\|z^{k+1} - P_k(z^k)\| \leq \varepsilon_k,$ $\sum_{k=0}^{\infty} \varepsilon_k < \infty,$

(B) $\|z^{k+1} - P_k(z^k)\| \leq \delta_k \|z^{k+1} - z^k\|,$ $\sum_{k=0}^{\infty} \delta_k < \infty.$

It is shown (Proposition 3) that these are implied respectively by

(A') $\text{dist}(0, S_k(z^{k+1})) \leq \varepsilon_k / c_k,$ $\sum_{k=0}^{\infty} \varepsilon_k < \infty,$

and

(B') $\text{dist}(0, S_k(z^{k+1})) \leq (\delta_k / c_k) \|z^{k+1} - z^k\|,$ $\sum_{k=0}^{\infty} \delta_k < \infty,$

where

(1.16) $S_k(z) = T(z) + c_k^{-1}(z - z_k).$

(Note that these conditions are certainly satisfied if $z^{k+1} = P_k(z^k)$.)

We prove under very mild assumptions (Theorem 1) that (A) (or (A')) guarantees (for any starting point z^0) weak convergence of $\{z^k\}$ to a particular solution z^∞ to $0 \in T(z)$, even though there may be more than one solution. (In general, the set of all such points z forms a closed convex set in H, denoted by $T^{-1}(0)$.) The results of Martinet are thereby extended to a much larger class of problems, and with only $z^{k+1} \approx P_k(z^k)$.

When (B) (or (B')) is also satisfied and the multifunction T^{-1} happens to be "Lipschitz continuous at 0", we are able to show (Theorem 2) that the con-

vergence is at least at a linear rate, where the modulus can be brought arbitrarily close to zero by taking c_k large enough. If $c_k \to \infty$, one has superlinear convergence.

In other words, the same convergence properties noted above for the case of strong monotonicity are established under far weaker assumptions. A criterion for the convergence of the algorithm in a finite number of iterations is also furnished (Theorem 3).

The assumption of Lipschitz continuity of T^{-1} at 0 turns out to be very natural in applications to convex programming. It is fulfilled, for instance, under certain standard second-order conditions characterizing a "nice" optimal solution. Such applications, having many ramifications, will be discussed elsewhere [31].

There are actually three distinct types of applications of the proximal point algorithm in convex programming: (i) to $T = \partial f$, where f is the essential objective function in the problem, (ii) to $T = -\partial g$, where g is the concave objective function in the dual problem, and (iii) to the monotone operator T_L corresponding to the convex–concave Lagrangian function.

The second type of application corresponds to the "method of multipliers" of Hestenes [8] and Powell [21]. The relationship with the proximal point algorithm in this case has already been used by Rockafellar [29]. The third type of application yields a new form of the method of multipliers that seems superior, at least in some respects. Although the details will not be treated here, we mention these applications because of their role in motivating the present work.

Some implications of the theorems in this paper for the general cases of $T = \partial f$ or T corresponding to a convex–concave function L are nevertheless discussed in § 4.

Aside from the obvious case of strong monotonicity, or special results for the method of multipliers in convex programming (for a survey, see Bertsekas [5]), there are no rate-of-convergence results relating to the proximal point algorithm prior to those developed here.

For discussion of other methods for solving $0 \in T(z)$ in the case of a maximal monotone operator, we refer to Auslender [2] and Bakushinskii and Polyak [3].

2. Convergence of the general algorithm. Henceforth T is always maximal monotone. In addition to $P_k = (I + c_k T)^{-1}$, we shall make use of the mapping

$$(2.1) \qquad Q_k = I - P_k = (I + (c_k T)^{-1})^{-1}.$$

Thus

$$(2.2) \qquad 0 \in T(z) \Leftrightarrow P_k(z) = z \Leftrightarrow Q_k(z) = 0.$$

PROPOSITION 1.
(a) $z = P_k(z) + Q_k(z)$ and $c_k^{-1} Q_k(z) \in T(P_k(z))$ for all z.
(b) $\langle P_k(z) - P_k(z'), Q_k(z) - Q_k(z') \rangle \geq 0$ for all z, z'.
(c) $\|P_k(z) - P_k(z')\|^2 + \|Q_k(z) - Q_k(z')\|^2 \leq \|z - z'\|^2$ for all z, z'.

Proof. Part (a) is immediate from the definitions, while (b) is a consequence of (a) and the monotonicity of T. Part (c) follows from (a) and (b) by expanding

$$\|z - z'\|^2 = \|[P_k(z) - P_k(z')] + [Q_k(z) - Q_k(z')]\|^2.$$

Part (c) of Proposition 1 states that property (1.12) holds for P_k and Q_k. If $c_k \equiv c > 0$, the mappings P_k all reduce to a single V to which the Martinet's corollary of Opial's theorem (recalled in §1 after (1.12)) can be applied with respect to any nonempty closed bounded convex set C such that $V(C) \subset C$. Of course, if V is known to have at least one fixed point in H, then for arbitrary $z^0 \in H$ one can take C to be the closed ball of radius $\|z^0 - \bar{z}\|$ and center $\bar{z}$, where $\bar{z}$ is any fixed point.

In this way one obtains an immediate generalization of Martinet's results for the case of $T = \partial f$ or variational inequalities. *If there exists at least one z satisfying $0 \in T(z)$, then the proximal point algorithm in exact form ($z^{k+1} = P_k(z^k)$) with $c_k \equiv c$ converges weakly from any starting point z^0 to a particular z^∞ satisfying $0 \in T(z^\infty)$.* This should be compared with the still more general Theorem 1 below.

In connection with the existence of solutions to the problem we want to solve, it is worth mentioning the following result (Rockafellar [25, Prop. 2]; this is a generalization of Theorem 2.2 of Browder [7]).

PROPOSITION 2 (see [25]). *Suppose that for some $\bar{z} \in H$ and $\rho \geq 0$ one has*

$$(2.3) \qquad \langle z - \bar{z}, w \rangle \geq 0 \quad \text{for all} \quad z, w \quad \text{with} \quad w \in T(z), \|z - \bar{z}\| \geq \rho.$$

Then there exists at least one z satisfying $0 \in T(z)$. (This condition is not only sufficient for existence, but necessary.)

The condition in Proposition 2 holds trivially for example, if the effective domain

$$(2.4) \qquad\qquad\qquad D(T) = \{z \in H \mid T(z) \neq \varnothing\}$$

is a bounded set. A convenient, weaker condition, which is also sufficient for existence when $T = \partial f$, is the weak compactness of the level sets $\{z \in H \mid f(z) \leq \beta\}$, $\beta \in R$.

The relationship between the criteria (A) and (B) on the one hand and (A') and (B') on the other is laid out by the next of our preliminary results.

PROPOSITION 3. *The estimate*

$$\|z^{k+1} - P_k(z^k)\| \leq c_k \, \text{dist}(0, S_k(z^{k+1}))$$

holds, where S_k is given by (1.16). Thus (A') implies (A), and (B') implies (B).

Proof. For any $w \in S_k(z^{k+1})$ we have

$$c_k w + z^k \in (I + c_k T)(z^{k+1}),$$

and hence,

$$z^{k+1} = (I + c_k T)^{-1}(c_k w + z^k) = P_k(c_k w + z^k).$$

Then by virtue of the nonexpansiveness of P_k

$$\|z^{k+1} - P_k(z^k)\| = \|P_k(c_k w + z^k) - P_k(z^k)\| \leq c_k \|w\|.$$

Thus

$$\|z^{k+1} - P_k(z^k)\| \leq c_k \min \{\|w\| \mid w \in S_k(z^{k+1})\}$$

as claimed.

THEOREM 1. *Let $\{z^k\}$ be any sequence generated by the proximal point algorithm under criterion* (A) (*or* (A')) *with $\{c_k\}$ bounded away from zero. Suppose $\{z^k\}$ is bounded; this holds under the preceding assumption if and only if there exists at least one solution to $0 \in T(z)$.*

Then $\{z^k\}$ converges in the weak topology to a point z^∞ satisfying $0 \in T(z^\infty)$, and

$$(2.5) \qquad 0 = \lim_{k \to \infty} \|Q_k(z^k)\| = \lim_{k \to \infty} \|z^{k+1} - z^k\|.$$

Proof. First we verify the asserted sufficient condition for the boundedness of $\{z^k\}$. The necessity of the condition will follow from the last part of the theorem.

Suppose that $\bar{z}$ is a point satisfying $0 \in T(\bar{z})$. We have

$$(2.6) \qquad \|z^{k+1} - \bar{z}\| - \varepsilon_k \leq \|P_k(z^k) - \bar{z}\| = \|P_k(z^k) - P_k(\bar{z})\| \leq \|z^k - \bar{z}\|,$$

and this furnishes the bound

$$\|z^l - \bar{z}\| \leq \|z^0 - \bar{z}\| + \sum_{k=0}^{l-1} \varepsilon_k \leq \|z^0 - \bar{z}\| + \alpha \quad \text{for all} \quad l.$$

Thus $\{z^k\}$ must be bounded.

For the rest of the proof, we assume that $\{z^k\}$ is any bounded sequence satisfying (A). Let $s > 0$ be such that

$$(2.7) \qquad \|z^k\| \leq s \quad \text{and} \quad \varepsilon_k < s \quad \text{for all} \quad k.$$

Then $\{z^k\}$ has at least one weak cluster point z^∞, $\|z^\infty\| \leq s$.

Our next goal is to demonstrate that $0 \in T(z^\infty)$, but for this purpose it is helpful to show first that the argument can be reduced to the case where it is already known that $T^{-1}(0) \neq \varnothing$. Consider the multifunction T' defined by

$$T'(z) = T(z) + \partial h(z) \quad \text{for all} \quad z \in H,$$

where

$$h(z) = \begin{cases} 0 & \text{if} \quad \|z\| \leq 2s, \\ +\infty & \text{if} \quad \|z\| > 2s, \end{cases}$$

and consequently

$$\partial h(z) = \begin{cases} \{0\} & \text{if} \quad \|z\| < 2s, \\ \{\lambda z \,|\, \lambda \geq 0\} & \text{if} \quad \|z\| = 2s, \\ \varnothing & \text{if} \quad \|z\| > 2s. \end{cases}$$

Observe that ∂h is a maximal monotone operator, because h is a lower semicontinuous proper convex function; its effective domain is

$$D(\partial h) = \{z \,|\, \|z\| \leq 2s\}.$$

Furthermore,

$$(2.8) \qquad T'(z) = T(z) \quad \text{if} \quad \|z\| < 2s.$$

Since $\|P_k(z^k)\| < 2s$ for all k by (2.7) and condition (A), while

$$c_k^{-1}(z^k - P_k(z^k)) \in T(P_k(z^k))$$

by Proposition 1(a), we have

$$(2.9) \qquad P_k(z^k) \in D(T) \cap \text{int } D(\partial h) \quad \text{for all} \quad k,$$

$$(2.10) \qquad P_k(z^k) \in (I + c_k T')^{-1}(z^k) \quad \text{for all} \quad k.$$

Inasmuch as $D(T) \cap \text{int } D(\partial h) \neq \varnothing$ by (2.9), we know that T', as the sum of the maximal monotone operators T and ∂h, is itself maximal monotone (Rockafellar [27, Thm. 1]). Hence $P_k' = (I + c_k T')^{-1}$ is actually single-valued, and (2.10) implies

$$P_k(z^k) = P_k'(z^k) \quad \text{for all large} \quad k.$$

Thus the sequence $\{z^k\}$ can be regarded equally well as arising from the proximal point algorithm for T'. The advantage in this is that the effective domain $D(T')$ is bounded, so that $(T')^{-1}(0) \neq \varnothing$ by Proposition 2. Since $T'(z^\infty) = T(z^\infty)$ by (2.8), we could replace T by T' without loss of generality in verifying that $0 \in T(z^\infty)$.

We are therefore justified in assuming, from now on, the existence of a certain $\bar{z}$ such that $0 \in T(\bar{z})$. Applying Proposition 1(c) to $z = z^k$ and $z' = \bar{z}$, we get

$$(2.11) \qquad \|P_k(z^k) - \bar{z}\|^2 + \|Q_k(z^k)\|^2 \leq \|z^k - \bar{z}\|^2 \quad \text{for all} \quad k.$$

Hence,

$$\begin{aligned}
\|Q_k(z^k)\|^2 - \|z^k - \bar{z}\|^2 + \|z^{k+1} - \bar{z}\|^2 &\leq \|z^{k+1} - \bar{z}\|^2 - \|P_k(z^k) - \bar{z}\|^2 \\
&= \langle z^{k+1} - P_k(z^k), (z^{k+1} - \bar{z}) + (P_k(z^k) - \bar{z}) \rangle \\
&\leq \|z^{k+1} - P_k(z^k)\|(\|z^{k+1} - \bar{z}\| + \|z^k - \bar{z}\|),
\end{aligned}$$

and consequently,

$$(2.12) \qquad \|Q_k(z^k)\|^2 \leq \|z^k - \bar{z}\|^2 - \|z^{k+1} - \bar{z}\|^2 + 2\varepsilon_k(s + \|\bar{z}\|).$$

At the same time we have

$$\|z^{k+1} - \bar{z}\| \leq \|P_k(z^k) - \bar{z}\| + \varepsilon_k \leq \|z^k - \bar{z}\| + \varepsilon_k,$$

which because of $\sum_{k=0}^\infty \varepsilon_k < \infty$ implies the existence of

$$(2.13) \qquad \lim_{k \to \infty} \|z^k - \bar{z}\| = \mu < \infty.$$

We can therefore take the limit on both sides of (2.12), obtaining (2.5), because

$$\|Q_k(z^k)\| = \|(z^k - z^{k+1}) + (z^{k+1} - P_k(z^k))\| \geq \|z^{k+1} - z^k\| - \varepsilon_k.$$

It follows that

$$(2.14) \qquad c_k^{-1} Q_k(z^k) \to 0 \quad \text{strongly},$$

the numbers c_k being bounded away from zero.

Observe next that Proposition 1(a) entails

$$(2.15) \qquad 0 \leq \langle z - P_k(z^k), w - c_k^{-1} Q_k(z^k) \rangle \quad \text{for all} \quad k \quad \text{if} \quad w \in T(z).$$

Since z^∞ is a weak cluster point of $\{z^k\}$ and $\|z^{k+1} - P_k(z^k)\| \to 0$, it is also a weak cluster point of $\{P_k(z^k)\}$. Then (2.14) and (2.15) yield

$$0 \leq \langle z - z^\infty, w \rangle \quad \text{for all} \quad z, w \quad \text{satisfying} \quad w \in T(z).$$

This implies, in view of the maximality of T, that $0 \in T(z^\infty)$.

The next step is to show that there cannot be more than one weak cluster point of $\{z^k\}$. Suppose there were two: $z_1^\infty \neq z_2^\infty$. Then $0 \in T(z_i^\infty)$ for $i = 1, 2$, as just seen, so that each z_i^∞ can play the role of $\bar{z}$ in (2.13), and we get the existence of the limits

(2.16) $$\lim_{k \to \infty} \|z^k - z_i^\infty\| = \mu_i < \infty \quad \text{for} \quad i = 1, 2.$$

Writing

$$\|z^k - z_2^\infty\|^2 = \|z^k - z_1^\infty\|^2 + 2\langle z^k - z_1^\infty, z_1^\infty - z_2^\infty \rangle + \|z_1^\infty - z_2^\infty\|^2,$$

we see that the limit of $\langle z^k - z_1^\infty, z_1^\infty - z_2^\infty \rangle$ must also exist and

$$2 \lim_{k \to \infty} \langle z^k - z_1^\infty, z_1^\infty - z_2^\infty \rangle = \mu_2^2 - \mu_1^2 - \|z_1^\infty - z_2^\infty\|^2.$$

But this limit cannot be different from 0, because z_1^∞ is a weak cluster point $\{z^k\}$. Therefore

$$\mu_2^2 - \mu_1^2 = \|z_1^\infty - z_2\|^2 > 0.$$

However, the same argument works with z_1^∞ and z_2^∞ reversed, so that also $\mu_1^2 - \mu_2^2 > 0$. This is a contradiction which establishes the uniqueness of z^∞.

(The uniqueness argument just given closely follows the one of Martinet [12], and it was also suggested to the author by H. Brézis.)

Counterexample. The convergence of $\{z^k\}$ in Theorem 1 may fail if instead of $\sum_{k=0}^\infty \varepsilon_k < \infty$ one has only $\varepsilon_k \to 0$, even when H is one-dimensional. This can be seen for any maximal monotone T such that the set $T^{-1}(0) = \{z | 0 \in T(z)\}$, which is known always to be convex, contains more than one element. Then $T^{-1}(0)$ contains a nonconvergent sequence $\{z^k\}$ with

$$\|z^{k+1} - z^k\| \to 0$$

but

$$\sum_{k=0}^\infty \|z^{k+1} - z^k\| = \infty.$$

We have $P_k(z^k) = z^k$ and therefore a counterexample with $\varepsilon_k = \|z^{k+1} - z^k\|$. In particular, all this holds for $T = \partial f$ if the convex function f attains its minimum nonuniquely.

3. Rate of convergence. We shall say that T^{-1} is *Lipschitz continuous at* 0 (*with modulus* $a \geq 0$) if there is a unique solution $\bar{z}$ to $0 \in T(z)$ (i.e. $T^{-1}(0) = \{\bar{z}\}$), and for some $\tau > 0$ we have

(3.1) $$\|z - \bar{z}\| \leq a\|w\| \quad \text{whenever} \quad z \in T^{-1}(w) \quad \text{and} \quad \|w\| \leq \tau.$$

THEOREM 2. *Let* $\{z^k\}$ *be any sequence generated by the proximal point algorithm using criterion* (B) (*or* (B')) *with* $\{c_k\}$ *nondecreasing* ($c_k \uparrow c_\infty \leq \infty$).

Assume that $\{z^k\}$ is bounded (cf. Theorem 1) and that T^{-1} is Lipschitz continuous at 0 with modulus a; let

$$\mu_k = a/(a^2 + c_k^2)^{1/2} < 1.$$

Then $\{z^k\}$ converges strongly to $\bar{z}$, the unique solution to $0 \in T(z)$. Moreover, there is an index $\bar{k}$ such that

(3.2) $\|z^{k+1} - \bar{z}\| \leq \theta_k \|z^k - \bar{z}\|$ *for all* $k \geq \bar{k}$,

where

(3.3) $1 > \theta_k \equiv (\mu_k + \delta_k)/(1 - \delta_k) \geq 0$ *for all* $k \geq \bar{k}$,

(3.4) $\theta_k \rightarrow \mu_\infty$ (*where* $\mu_\infty = 0$ *if* $c_\infty = \infty$).

Proof. The sequence $\{z^k\}$, being bounded, also satisfies criterion (A) for $\varepsilon_k = \delta_k \|z^{k+1} - z^k\|$, so the conclusions of Theorem 1 are in force. We have

$$\|Q_k(z^k)\| = \|z^k - P_k(z^k)\| \leq \|z^k - z^{k+1}\| + \|z^{k+1} - P_k(z^k)\|,$$

so that

$$\|c_k^{-1} Q_k(z^k)\| \leq c_k^{-1}(1 + \delta_k)\|z^{k+1} - z^k\| \text{ for all }   k,$$

where $\|z^k - z^{k+1}\| \rightarrow 0$ (Theorem 1). Choose $\tilde{k}$ so that

(3.5) $c_k^{-1}(1 + \delta_k)\|z^{k+1} - z^k\| < \tau$ for all $k \geq \tilde{k}$.

Then $\|c_k^{-1} Q_k(z^k)\| \leq \tau$ for $k \geq \tilde{k}$. But $P_k(z^k) \in T^{-1}(c_k^{-1}Q_k(z^k))$ by Proposition 1(a). The Lipschitz condition (3.1) can therefore be invoked for $w = c_k^{-1}Q_k(z^k)$ and $z = P_k(z^k)$ if k is sufficiently large:

(3.6) $\|P_k(z^k) - \bar{z}\| \leq a\|c_k^{-1}Q_k(z^k)\|$ for all $k \geq \tilde{k}$.

We next apply (2.2) and Proposition 1(c) to $z = \bar{z}$ and $z' = z^k$ to obtain

$$\|\bar{z} - P_k(z^k)\|^2 + \|Q_k(z^k)\|^2 \leq \|\bar{z} - z^k\|^2,$$

which via (3.6) yields

$$\|P_k(z^k) - \bar{z}\|^2 \leq [(a/c_k)^2/(1 + (a/c_k)^2)]\|z^k - \bar{z}\|^2,$$

or in other words

(3.7) $\|P_k(z^k) - \bar{z}\| \leq \mu_k \|z^k - \bar{z}\|$ if $k \geq \bar{k}$.

But

$$\|z^{k+1} - \bar{z}\| \leq \|z^{k+1} - P_k(z^k)\| + \|P_k(z^k) - \bar{z}\|,$$

where under (B) we have

$$\|z^{k+1} - P_k(z^k)\| \leq \delta_k \|z^{k+1} - z^k\| \leq \delta_k \|z^{k+1} - \bar{z}\| + \delta_k \|z^k - \bar{z}\|.$$

Therefore by (3.7),

$$\|z^{k+1} - \bar{z}\| \leq \delta_k \|z^{k+1} - \bar{z}\| + \mu_k \|z^k - \bar{z}\| + \delta_k \|z^k - \bar{z}\| \text{ if }   k \geq \tilde{k}.$$

This inequality produces the one in (3.2) if $\bar{k} \geq \tilde{k}$ is taken so that (3.3) holds, as is possible since $1 > \mu_k \downarrow \mu_\infty$ and $\delta_k \rightarrow 0$.

Remark 1. The proof shows that the estimates in Theorem 2 are valid for any $\bar{k}$ such that (3.3) holds and, for some $\tilde{k} \leq \bar{k}$, also (3.5) holds. To cite a simple

specific case, let us suppose that

$$\delta_k \leqq \tfrac{1}{4} \quad \text{for all} \quad k,$$

and, as can easily be estimated explicitly for instance if the effective domain $D(T)$ is bounded, that for a certain $d > 0$,

$$\|z^{k+1} - z^k\| \leqq d \quad \text{for all} \quad k.$$

It may then be seen that the estimates in Theorem 2 are valid if $\bar{k}$ is such that

$$c_k \geqq 2 \max\{a, d/\tau\} \quad \text{for all} \quad k \geqq \bar{k}.$$

Remark 2. If we replace the condition on δ_k in (B) by the assumption that (A) is satisfied and

(3.8) $$\delta_k \to \delta_\infty < \tfrac{1}{2}(1 - \mu_\infty),$$

then all the conclusions of Theorem 2 hold, except that

$$\theta_k \to \theta_\infty = (\mu_\infty + \delta_\infty)/(1 - \delta_\infty) < 1.$$

Since

$$\mu_\infty = a/(a^2 + c_\infty^2)^{1/2},$$

the inequality (3.8) holds in particular if $\delta_\infty < \tfrac{1}{2}$ and $c_k \uparrow \infty$.

The next two results help illuminate the Lipschitz condition in Theorem 2.

We shall say that a multifunction $S : H \to H$ is *differentiable* at a point $\bar{w}$ if $S(\bar{w})$ consists of a single element $\bar{z}$ and there is a continuous linear transformation $A : H \to H$ such that, for some $\delta > 0$,

$$\varnothing \neq S(\bar{w} + w) - \bar{z} - Aw \subset o(\|w\|)B \quad \text{when} \quad \|w\| \leqq \delta,$$

where B is the closed unit ball and

$$\mathfrak{o}(\|w\|)/\|w\| \downarrow 0 \quad \text{as} \quad \|w\| \downarrow 0.$$

We then write $A = \nabla S(\bar{w})$. This coincides with the usual notion of differentiability (in the sense of Fréchet), if S is single-valued on a neighborhood of $\bar{w}$.

PROPOSITION 4. *The condition of Lipschitz continuity in Theorem 2 is satisfied if T^{-1} is differentiable at 0. In particular, it is satisfied if there is a $\bar{z}$ such that $0 \in T(\bar{z})$ and T is single-valued and continuously differentiable in a neighborhood of $\bar{z}$, with $\nabla T(\bar{z})$ invertible (i.e. having all of H as its range).*

Proof. If T^{-1} is differentiable at 0 and $A = \nabla T^{-1}(0)$, there is a unique $\bar{z}$ satisfying $0 \in T(\bar{z})$, and we have

$$T^{-1}(w) - \bar{z} - Aw \subset o(\|w\|)B \quad \text{when} \quad \|w\| \leqq \delta.$$

Thus there exist $a_0 \geqq 0$ and $\varepsilon > 0$ such that

$$\|z - \bar{z} - Aw\| \leqq a_0 \|w\| \quad \text{whenever} \quad z \in T^{-1}(w), \quad \|w\| \leqq \varepsilon.$$

It follows that

$$\|z - \bar{z}\| \leqq a_0 \|w\| + \|A\| \cdot \|w\| \quad \text{whenever} \quad w \in T(z), \quad \|w\| \leqq \varepsilon.$$

Thus (3.1) holds for $a = a_0 + \|A\|$. The second assertion then follows from the first

by way of the implicit function theorem [9]: under these assumptions T^{-1} is single-valued and continuously differentiable on a neighborhood of 0.

PROPOSITION 5. *Suppose T^{-1} is Lipschitz continuous globally, i.e. T^{-1} is everywhere single-valued and satisfies*

$$\|T^{-1}(w) - T^{-1}(w')\| \leqq a\|w - w'\| \quad \text{for all} \quad w, w',$$

where $a \geqq 0$; this is true in particular if T is strongly monotone with modulus $\alpha > 0 (a = \alpha^{-1})$. Then the explicit assumption that $\{z^k\}$ is bounded is superfluous for the conclusions of Theorem 2, and the estimate (3.2) is valid for any $\bar{k}$ large enough that (3.3) holds.

Proof. The proof of Theorem 2 works in this case with $\tilde{k} = 0$. If T is strongly monotone, we have (1.13) for some $\alpha > 0$. Then the operator $T' = T - \alpha I$ is monotone and hence $P = (I + \alpha^{-1}T')^{-1}$ is nonexpansive. But $T = \alpha P^{-1}$, so

$$T^{-1}(w) = P(\alpha^{-1}w) \quad \text{for all} \quad w,$$

and in particular from the nonexpansiveness of P:

(3.9) $$\|T^{-1}(w) - T^{-1}(w')\| \leqq \alpha^{-1}\|w - w'\| \quad \text{for all} \quad w, w'.$$

Finally, we describe a very special but noteworthy case where the algorithm can converge in finitely many iterations. This result was suggested by one of Bertsekas [4] for the method of multipliers in convex programming.

THEOREM 3. *Let $\{z^k\}$ be any sequence generated by the proximal point algorithm under any of the criteria (A), (A'), (B) or (B') with $\{c_k\}$ bounded away from zero. Suppose that $\{z^k\}$ is bounded (cf. the conditions in Theorem 1) and there exists $\bar{z}$ such that $0 \in \text{int } T(\bar{z})$. Then*

(3.10) $$z^\infty = \bar{z} = P_k(z^k) \quad \text{for all} \quad k \quad \text{sufficiently} \quad \text{large.}$$

Hence under (A) (or (A')) one has

$$\|z^k - \bar{z}\| \leqq \varepsilon_k \quad \text{for all} \quad k \quad \text{sufficiently} \quad \text{large,}$$

while under (B) (or (B')) with $c_k \uparrow c_\infty \leqq \infty$ one has the estimates (3.2) and (3.5) for

$$\theta_k = \delta_k/(1 - \delta_k) \to 0.$$

Thus in particular, the proximal point algorithm in its exact form (i.e. with $z^{k+1} = P_k(z^k)$) gives convergence to $\bar{z}$ in a finite number of iterations from any starting point z^0.

Proof. We demonstrate first that T^{-1} is single-valued and constant on a neighborhood of 0:

(3.11) $$T^{-1}(w) = \bar{z} \quad \text{if} \quad \|w\| < \varepsilon.$$

Let $\varepsilon > 0$ be chosen so that $\|w\| < \varepsilon$ implies $w \in \text{int } T(\bar{z})$. Taking any $z, w \in T(z)$, and w' with $\|w\| < \varepsilon$, we have

$$0 \leqq \langle z - \bar{z}, w - w' \rangle$$

by the monotonicity of T. Therefore

$$\sup \langle z - \bar{z}, w' \rangle \leqq \langle z - \bar{z}, w \rangle \quad \text{whenever} \quad w \in T(z), \quad \|w'\| < \varepsilon,$$

so that

$$\varepsilon \|z - \bar{z}\| \leq \|z - \bar{z}\| \cdot \|w\| \quad \text{whenever} \quad w \in T(z).$$

Thus if $z \neq \bar{z}$ we have $\|w\| \geq \varepsilon$ for all $w \in T(z)$. Stated another way, if $\|w\| < \varepsilon$ and $z \in T^{-1}(w)$, then $z = \bar{z}$, which is the same assertion as (3.11).

Our hypothesis subsumes that of Theorem 1, and hence we know as in Theorem 1 that $\|c_k^{-1} Q_k(z^k)\| \to 0$. However, $P_k(z^k) \in T^{-1}(c_k^{-1} Q_k(z^k))$ by Proposition 1(a). Therefore (3.11) implies (3.10), and everything else in Theorem 3 follows immediately, the Lipschitz condition in Theorem 2 being fulfilled with $a = 0$.

4. Application to minimization. Let $f : H \to (-\infty, +\infty]$ be a lower semicontinuous convex function which is not identically $+\infty$. Then, as noted in the introduction, the multifunction $T = \partial f$ is maximal monotone, where

(4.1)
$$w \in \partial f(z) \Leftrightarrow f(z') \geq f(z) + \langle z' - z, w \rangle \quad \text{for all} \quad z'$$
$$\Leftrightarrow z \in \arg\min (f - \langle \cdot, w \rangle).$$

Since in particular

$$0 \in \partial f(z) \Leftrightarrow z \in \arg\min f,$$

the proximal point algorithm for $T = \partial f$ is a method for minimizing f. We collect here some facts relevant to this special case.

Recall that a function $\phi : H \to (-\infty, +\infty]$ is said to be *strongly convex (with modulus α)* if $\alpha > 0$ and

(4.2)
$$\phi((1-\lambda)z + \lambda z') \leq (1-\lambda)\phi(z) + \lambda\phi(z') - \tfrac{1}{2}\alpha\lambda(1-\lambda)\|z - z'\|^2$$
$$\text{for all} \quad z, z' \quad \text{if} \quad 0 < \lambda < 1.$$

THEOREM 4. *Let $T = \partial f$. Then $S_k = \partial \phi_k$ in criteria (A') and (B'), where ϕ_k is the function defined by (1.9), and ϕ_k is lower semicontinuous and strongly convex with modulus $1/c_k$. Furthermore, if $\{z^k\}$ is any sequence generated by the proximal point algorithm under the hypothesis of Theorem 1 with criterion (A'), then $z^k \to z^\infty$ weakly, where $f(z^\infty) = \min f$ and*

(4.3)
$$f(z^{k+1}) - f(z^\infty) \leq c_k^{-1} \|z^{k+1} - z^\infty\|(\varepsilon_k + \|z^{k+1} - z^k\|) \to 0.$$

Proof. The strong convexity of ϕ follows directly from formula (1.9). Subdifferentiating both sides of this formula, we also get

$$\partial \phi_k(z) = \partial f(z) + c_k^{-1}(z - z^k) \equiv S_k(z) \quad \text{for all} \quad z.$$

(For the relevant rule of subdifferentiation, see Moreau [17] or Rockafellar [22, Thm. 3].) To establish (4.3), let w^k denote the unique element of $\partial \phi_k(z^{k+1})$ nearest the origin. (This exists, because $\partial \phi_k(z^{k+1})$ is a closed convex set which, since (A') is supposed to hold, is nonempty.) Then

$$w^k - c_k^{-1}(z^{k+1} - z^k) \in T(z^{k+1}) = \partial f(z^{k+1}),$$

where

(4.4)
$$\|w^k\| \leq \varepsilon_k / c_k \to 0.$$

Let z^∞ be the weak limit of $\{z^k\}$ (Theorem 1). Then $0 \in \partial f(z^\infty)$, and the defining inequality for subgradients yields

$$f(z^{k+1}) + \langle z^\infty - z^{k+1}, w^k - c_k^{-1}(z^{k+1} - z^k)\rangle \leq f(z^\infty) = \min f,$$

so that

$$f(z^{k+1}) - f(z^\infty) \leq \|z^{k+1} - z^\infty\|(\|w^k\| + c_k^{-1}\|z^{k+1} - z^k\|).$$

Applying (4.4) and (2.5), we reach the desired conclusion (4.3).

Remark 3. The quantity $\text{dist}(0, \partial\phi_k(z^{k+1}))$ occurring in criteria (A') and (B') for $T = \partial f$ is generally convenient as a measure of how near z^{k+1} is to being a minimizer of ϕ_k. Exact minimization corresponds, of course, to $\text{dist}(0, \partial\phi_k(z^{k+1})) = 0$. Many methods that might be used for minimizing ϕ_k depend on the calculation of gradients or subgradients, and one can use the estimate

$$\text{dist}(0, \partial\phi_k(z^{k+1})) \leq \|u\| \quad \text{for any} \quad u \in \partial\phi_k(z^{k+1}). \qquad \bullet$$

This is not the place to describe all of the possible structures of $\partial\phi_k$ corresponding to minimization problems of different types, but we nevertheless mention an important case. Suppose f is of the form

$$f(z) = \begin{cases} f_0(z) & \text{if} \quad z \in D, \\ +\infty & \text{if} \quad z \in D, \end{cases}$$

where D is a nonempty closed convex set and f_0 is a function which is convex on D and differentiable on a neighborhood of D. Then minimizing f on H is equivalent to minimizing f_0 on D, while minimizing ϕ_k on H is equivalent to minimizing

$$\phi_k^0(z) = f_0(z) + \tfrac{1}{2}\alpha\|z - z^k\|^2$$

on D. Furthermore,

$$\partial\phi_k(z) = \nabla\phi_k^0(z) + N_D(z),$$

where $N_D(z)$ is the normal cone to D at z, and hence $\text{dist}(0, \partial\phi_k(z^{k+1}))$ is the norm of the projection of $-\nabla\phi_k^0(z^{k+1})$ on the tangent cone to D at z^{k+1} (where $z^{k+1} \in D$).

In particular, if $D = H$, i.e., f itself is differentiable on all of H, we have

$$\text{dist}(0, S_k(z^{k+1})) = \|\nabla\phi_k(z^{k+1})\|$$

in (A') and (B').

It remains now to show how the various conditions in the hypotheses of Theorems 2 and 3 are realized in the case of $T = \partial f$.

Let f^* be the lower semicontinuous convex function conjugate to f. Thus $\partial f^* = T^{-1}$ for $T = \partial f$. (For the theory of conjugate functions, see [19], [30].)

PROPOSITION 6. *The following conditions are equivalent for $T = \partial f$:*

(a) *T is strongly monotone with modulus α,*

(b) *f is strongly convex with modulus α,*

(c) *whenever $w \in \partial f(z)$, one has for all $z' \in H$:*

$$f(z') \geq f(z) + \langle z' - z, w\rangle + \tfrac{1}{2}\alpha\|z' - z\|^2.$$

Proof. (b) $\Rightarrow$ (a). Suppose $w \in T(z)$ and $w' \in T(z')$, and let $0 < \lambda < 1$. Then

$$f((1-\lambda)z + \lambda z') \geqq f(z) + \langle [(1-\lambda)z + \lambda z'] - z, w \rangle$$
$$= f(z) + \lambda \langle z' - z, w \rangle,$$

and hence by (4.2) for f:

$$-\lambda f(z) + \lambda f(z') - \tfrac{1}{2}\alpha\lambda(1-\lambda)\|z - z'\|^2 \geqq \lambda\langle z' - z, w \rangle,$$

or equivalently,

$$\langle z - z', w \rangle \geqq \tfrac{1}{2}\alpha(1-\lambda)\|z - z'\|^2 + f(z) - f(z').$$

By symmetry it is also true that

$$\langle z' - z, w' \rangle \geqq \tfrac{1}{2}\alpha(1-\lambda)\|z' - z\|^2 + f(z') - f(z),$$

and in adding these two inequalities we obtain

$$\langle z - z', w - w' \rangle \geqq \alpha(1-\lambda)\|z' - z\|^2.$$

This holds for arbitrary $\lambda \in (0, 1)$, so it must also hold for $\lambda = 0$, which is the assertion of (a).

(a) $\Rightarrow$ (c). As observed in the proof of Proposition 5, the strong monotonicity implies that T^{-1} is single-valued and satisfies the global Lipschitz condition (3.9). But $T^{-1} = \partial f^*$. In particular, therefore, ∂f^* is single-valued and continuous everywhere, from which it follows that f^* is differentiable everywhere and ∇f^* reduces to the gradient mapping of f^* (see Asplund/Rockafellar [1, p. 461]). For any w and w', we have

$$\|\nabla f^*(w + t(w' - w)) - \nabla f^*(w)\| \leqq (t/\alpha)\|w' - w\| \quad \text{for} \quad t > 0,$$

so that

$$\langle \nabla f^*(w + t(w' - w)), w' - w \rangle \leqq \langle \nabla f^*(w), w' - w \rangle + (t/\alpha)\|w' - w\|^2 \quad \text{for} \quad t > 0.$$

From this we obtain

$$f^*(w') - f^*(w) = \int_0^1 \langle \nabla f^*(w + t(w' - w)), w' - w \rangle \, dt$$

$$\leqq \langle \nabla f^*(w), w' - w \rangle + \frac{1}{2\alpha}\|w' - w\|^2.$$

Fixing arbitrary z and w with $w \in \partial f(z)$, we have $z = \nabla f^*(w)$ and $f(z) + f^*(w) = \langle z, w \rangle$. Then for any z',

$$f(z') = f^{**}(z') = \sup_{w' \in H} \{\langle z', w' \rangle - f^*(w')\}$$

$$\geqq \sup_{w' \in H} \left\{ \langle z', w' \rangle - f^*(w) - \langle \nabla f^*(w), w' - w \rangle - \frac{1}{2\alpha}\|w' - w\|^2 \right\}$$

$$= \sup_{w' \in H} \left\{ f(z) + \langle z' - z, w' \rangle - \frac{1}{2\alpha}\|w' - w\|^2 \right\}$$

$$= f(z) + \langle z' - z, w \rangle + \tfrac{1}{2}\alpha\|z' - z\|^2.$$

Thus (c) holds.

(c) $\Rightarrow$ (b). Let $G = \{(z, w) | w \in \partial f(z)\}$, and for each $(z, w) \in G$ define the functions $g_{z,w}$ and $h_{z,w}$ by

$$g_{z,w}(z') = f(z) + \langle z' - z, w \rangle + \tfrac{1}{2}\alpha \|z' - z\|^2,$$

$$h_{z,w}(z') = f(z) + \langle z' - z, w \rangle.$$

Then $f \geqq g_{z,w} \geqq h_{z,w}$. It is a known fact, however, that

$$f(z') = \sup_{(z,w) \in G} h_{z,w}(z') \quad \text{for all} \quad z'$$

(Brønsted/Rockafellar [6, Thm. 2]). Hence

$$f(z') = \sup_{(z,w) \in G} g_{z,w}(z').$$

Each function $g_{z,w}$ is strongly convex with modulus α, and therefore f has this same property. This completes the proof of Proposition 6.

PROPOSITION 7. *The following conditions are equivalent for $T = \partial f$ and $\bar{z} \in H$.*
(a) *T^{-1} is Lipschitz continuous at 0, and $\bar{z}$ is the unique solution to $0 \in T(z)$.*
(b) *$\bar{z}$ is the unique minimizing point for f, and*

$$\liminf_{z \to \bar{z}} \frac{f(z) - f(\bar{z})}{\|z - \bar{z}\|^2} > 0.$$

(c) *$\bar{z}$ is the unique element of $\partial f^*(0)$, and*

$$\limsup_{u \to 0} [(f^*(u) - f^*(0) - \langle \bar{z}, u \rangle)/\|u\|^2] < \infty.$$

Proof. (a) $\Rightarrow$ (c). Since $T^{-1} = \partial f^*$, we have

(4.5) $\|z - \bar{z}\| \leqq a\|w\|$ whenever $z \in \partial f^*(w)$ and $\|w\| \leqq \varepsilon$.

This implies the boundedness of the set

(4.6) $\bigcup_{\|w\| \leqq \varepsilon} \partial f^*(w)$,

which contains $\bar{z}$; in other words, ∂f^* is locally bounded at 0, which is a point of the effective domain

(4.7) $D(\partial f^*) = \{w | \partial f^*(z) \neq \varnothing\}$.

But ∂f^* is a maximal monotone operator, so this property necessitates $0 \in \text{int } D(\partial f^*)$ (see Rockafellar [25, Thm. 1]). Since

(4.8) $D(\partial f^*) \subset \text{dom } f^* = \{w | f^*(w) < \infty\}$

it follows that f^* is finite on a neighborhood of 0. This implies in turn that f^* is continuous on a neighborhood of 0 [23, Cor. 7c] and hence that for all u in some neighborhood of 0, say for $\|u\| \leqq \delta (0 < \delta \leqq \varepsilon)$ we have $\partial f^*(u)$ nonempty weakly compact and

(4.9) $f^{*\prime}(w; u) = \max \{\langle z, u \rangle | z \in \partial f^*(w)\}$ for all $u \in H$,

where

$$f^{*\prime}(w; u) = \lim_{\lambda \downarrow 0} [f^*(w + \lambda u) - f^*(w)]/\lambda.$$

(Moreau, [17]). Moreover (4.5) and (4.9) give the estimate

(4.10) $f^{*\prime}(w; u) \leq \langle \bar{z}, u \rangle + a \|w\| \cdot \|u\|$ if $\|w\| \leq \delta.$

Observe next that if $\|u\| \leq \delta$ and $\zeta(t) = f^*(tu)$, then ζ is a finite continuous convex function on $[0, 1]$, and hence

$$\zeta(1) = \zeta(0) + \int_0^1 \zeta'_+(t) \, dt,$$

where ζ'_+ is the right derivative of ζ [26, Cor. 24.2.1]. This formula says that

$$f^*(u) = f^*(0) + \int_0^1 f^{*\prime}(tu; u) \, dt,$$

and hence by (4.10),

(4.11) $f^*(u) \leq f^*(0) + \langle \bar{z}, u \rangle + \frac{1}{2} a \|u\|^2$ if $\|u\| \leq \delta.$

Therefore (c) is valid.

(c) $\Rightarrow$ (b). Under (c), we have (4.11) for some $a > 0$ and $\delta > 0$. Let

$$\xi(s) = \begin{cases} \frac{1}{2} as^2 & \text{if } |s| \leq \delta, \\ +\infty & \text{if } |s| > \delta. \end{cases}$$

Then (4.11) can be expressed as

(4.12) $f^*(u) - f^*(0) - \langle \bar{z}, u \rangle \leq \xi(\|u\|)$ for all $u \in H,$

where $\xi(\|u\|)$ is convex in u. Taking conjugates on both sides, we obtain

$$f(\bar{z} + v) + f^*(0) \geq \xi^*(\|v\|)   \text{for all}   v \in H,$$

where

(4.13) $\xi^*(r) = \begin{cases} \frac{1}{2} a^{-1} r^2 & \text{if } |r| \leq a\delta, \\ \delta|r| - \frac{1}{2} a\delta^2 & \text{if } |r| \geq a\delta. \end{cases}$

But

(4.14) $f(\bar{z}) + f^*(0) = \langle \bar{z}, 0 \rangle,$

since $\bar{z} \in \partial f^*(0)$. Therefore

(4.15) $f(z) - f(\bar{z}) \geq \xi^*(\|z - \bar{z}\|)$ for all $z \in H,$

and in particular

$$f(z) - f(\bar{z}) \geq \frac{1}{2} a^{-1} \|z - \bar{z}\|^2   \text{if}   \|z - \bar{z}\| \leq a\delta.$$

Thus (b) holds.

(b) $\Rightarrow$ (c). The hypothesis means that for a certain $a > 0$ and $\delta > 0$ we have

(4.16) $f(z) - f(\bar{z}) \geq \frac{1}{2} a^{-1} \|z - \bar{z}\|^2$ whenever $\|z - \bar{z}\| \leq 2a\delta.$

We shall show first that this implies (4.15). Of course, since $\xi(s) \geq \frac{1}{2} as^2$ for all $s \in R$ we have (taking conjugates on both sides) that $\xi^*(r) \leq \frac{1}{2} a^{-1} r^2$ for all $r \in R$, and hence the inequality in (4.15) follows from the one in (4.16) if $\|z - \bar{z}\| \leq 2a\delta.$

and the inequality in (4.18) becomes

$$\tfrac{1}{2}a^{-1}\|z-\bar{z}\|^2 \leqq \|z-\bar{z}\| \cdot \|w\|.$$

Thus (4.18) gives us

$$\|z-\bar{z}\| \leqq 2a\|w\| \quad \text{whenever} \quad \|w\| \leqq \delta/2 \quad \text{and} \quad w \in T(z), \; .$$

and (a) is verified.

Remark 4. The proof of Proposition 7 shows that the infimum $\bar{a}$ of the numbers $a \geqq 0$ such that the Lipschitz condition in Theorem 2 holds (for $T = \partial f$) satisfies $\tfrac{1}{2}b^{-1} \leqq \bar{a} \leqq b^{-1}$, where

$$b = \liminf_{z \to \bar{z}} \frac{f(z)-f(\bar{z})}{\|z-\bar{z}\|^2} = \left[\limsup_{u \to 0} \frac{f^*(u)-f^*(0)-\langle \bar{z}, u\rangle}{\|u\|^2}\right]^{-1}$$

($\bar{z}$ being the unique minimizing point for f; $0^{-1} = \infty$ and $\infty^{-1} = 0$).

PROPOSITION 8. *Suppose that H is finite-dimensional and f is polyhedral convex (i.e. the epigraph of f is a polyhedral convex set). If f attains its minimum at a unique point $\bar{z}$, then $0 \in \operatorname{int} \partial f(\bar{z})$, so that Theorem 3 is applicable to $T = \partial f$. However, even if f does not attain its minimum at a unique point but merely is bounded below, the proximal point algorithm with exact minimization of ϕ_k at each step (and with c_k bounded away from zero) will converge to some minimizer of f in a finite number of iterations.*

Proof. The conjugate f^* is also polyhedral [26, p. 173]. If $\bar{z}$ is the unique minimizer of f, it is the sole element of $\partial f^*(0)$. Then f^* is differentiable at 0 [26, p. 242], hence actually affine in an open neighborhood W of 0 by polyhedral convexity, implying $\bar{z} = \nabla f^*(w)$ for all $w \in W$. Thus $w \in \partial f(\bar{z})$ for all $w \in W$.

More generally, if f is merely a polyhedral convex function which is bounded below, we still have $f^*(0) = -\inf f$ finite and attained [26, p. 268]. By Theorem 1, the proximal point algorithm with c_k bounded away from zero generates from any starting point z^0 a sequence $\{z^k\}$ such that $Q_k(z^k) \to 0$. We must show that in the case of exact minimization ($\varepsilon_k = 0$ in (A')) finite convergence is still obtained.

There is no loss of generality in supposing for convenience in the rest of the proof that $\min f = 0$, so that $f^*(0) = 0$. Let

$$M = \partial f^*(0) = \{z \mid f(z) = \min f\}$$

and

$$h(z) = \begin{cases} 0 & \text{if } z \in M, \\ +\infty & \text{if } z \notin M. \end{cases}$$

Then M is a polyhedral convex set, so that h is a polyhedral convex function. The conjugate h^* is then polyhedral too, and we have

$$h^*(w) = f^{*\prime}(0; w) = \lim_{\lambda \downarrow 0} [f^*(\lambda w) - f^{*\prime}(0)]/\lambda$$

[26, p. 216], since the polyhedral property of f^* implies that of $f^*(0; \cdot)$. It is clear from the latter formula that h^* coincides with f^* in some open neighborhood of 0. Moreover $c_k^{-1} Q_k(z^k)$ lies in this neighborhood for all k sufficiently large, since

$Q_k(z^k) \to 0$ and c_k is bounded away from 0. Thus

$$\partial h^*(c_k^{-1} Q_k(z^k)) = \partial f^*(c_k^{-1} Q_k(z^k)) \quad \text{for all large} \quad k.$$

Since $\partial f^* = T^{-1}$ for $T = \partial f$, we can conclude from Proposition 1(a) that

$$c_k^{-1} Q_k(z^k) \in (\partial h^*)^{-1}(P_k(z^k)) = \partial h(P_k(z^k))$$

for all k sufficiently large. This tells us that ultimately the algorithm acts on $\{z^k\}$ just as if the multifunction $T = \partial f$ were replaced by $T = \partial h$, or equivalently if f were replaced by h. But in that event $P_k(z^k)$ is just the point of M nearest to z^k.

Thus, as soon as we reach the stage where $c_k^{-1} Q_k(z^k)$ lies in the neighborhood where f^* coincides with $h^* = f^{*'}(0; \cdot)$ we have $z^{k+1} = P_k(z^k) \in M$. Since M consists of the fixed points of the mappings P_k, the sequence $\{z^k\}$ is constant thereafter.

Remark 5. In the case of Proposition 8, quadratic programming algorithms can be employed, at least in principle, to calculate the exact minimum of ϕ_k at each iteration. Then the exact form of the proximal point algorithm is reasonable, and according to Theorem 3 it will yield the unique minimizer $\bar{z}$ of f in a finite number of iterations. We shall show elsewhere [31] that this result, when applied to the dual of a linear programming problem, yields a fact proved by Polyak and Tretyakov [20]: when the "method of multipliers" is used on a linear programming problem with exact minimization of the augmented Lagrangian at each iteration, one has convergence to an optimal solution in a finite number of iterations.

5. Application to calculating saddle points. Let $L(x, y)$ be a convex–concave function on the Hilbert space $H_1 \times H_2$ which is closed and proper in the sense of [24], [28], and let T_L be the maximal monotone operator corresponding to L, as defined in the introduction. Then

$$(0, 0) \in T_L(x, y) \Leftrightarrow (x, y) = \arg \text{minimax } L.$$

The proximal point algorithm for $T = T_L$ is thus capable of computing saddle points of L, and some of the results in the preceding section have analogues for this case.

Let us say that a function $\Lambda : H_1 \times H_2 \to [-\infty, +\infty]$ is *strongly convex–concave (with modulus α)* if $\Lambda(x, y)$ is strongly convex in x and strongly concave in y, both with modulus α.

THEOREM 5. *Let $T = T_L$. Then one has $S_k = T_{\Lambda_k}$ in criteria (A') and (B'), where Λ_k is the function defined by (1.11), and Λ_k is closed, proper and strongly convex–concave with modulus $1/c_k$. Furthermore, if $\{z^k = (x^k, y^k)\}$ is any sequence generated by the proximal point algorithm under the hypothesis of Theorem 1 with criterion (A'), then $(x^k, y^k) \to (x^\infty, y^\infty)$ weakly, where (x^∞, y^∞) is a saddle point of L and*

$$(5.1) \qquad \lim_{k \to \infty} L(x^k, y^k) = L(x^\infty, y^\infty) = \text{minimax } L.$$

Proof. This is mostly an easy extension of the argument for $T = \partial f$ in Theorem 4, but the justification of (5.1) is trickier and deserves some attention. Since

(x^∞, y^∞) is a saddle point, we have

$$(5.2) \qquad L(x^{k+1}, x^\infty) \geqq L(x^\infty, y^\infty) \geqq L(x^\infty, y^{k+1}) \quad \text{for all} \quad k.$$

Let $w^k = (v^k, u^k)$ denote the element of $S_k(x^k, y^k)$ nearest the origin. Thus $(v^k, u^k) \to (0, 0)$ strongly and

$$(5.3) \qquad (v^k - c_k^{-1}(x^{k+1} - x^k), u^k - c_k^{-1}(y^{k+1} - y^k)) \in T_L(x^{k+1}, y^{k+1}).$$

The latter relation gives us

$$L(x^\infty, y^{k+1}) \geqq L(x^{k+1}, y^{k+1}) + \langle x^\infty - x^{k+1}, v^k - c_k^{-1}(x^{k+1} - x^k)\rangle,$$
$$L(x^{k+1}, y^\infty) \leqq L(x^{k+1}, y^{k+1}) - \langle y^\infty - y^{k+1}, u^k - c_k^{-1}(y^{k+1} - y^k)\rangle.$$

Combining these inequalities with (5.2), we obtain

$$-\langle y^\infty - y^{k+1}, u^k - c_k^{-1}(y^{k+1} - y^k)\rangle \geqq L(x^\infty, y^\infty) - L(x^{k+1}, y^{k+1})$$
$$\geqq \langle x^\infty - x^{k+1}, v^k - c_k^{-1}(x^{k+1} - x^k)\rangle,$$

where the outer expressions converge to 0 by virtue of the limits already mentioned and assertion (2.5) of Theorem 1.

The analogue of Proposition 6 is valid for $T = T_L$, but the other results in § 4 do not have obvious extensions to the minimax context. For Proposition 7, this is seen from the example of $L(x, y) = xy$ on $R \times R$, which has $T_L(x, y) = (y, -x)$ and therefore T_L^{-1} globally Lipschitz continuous with modulus 1.

REFERENCES

[1] E. ASPLUND, AND R. T. ROCKAFELLAR, *Gradients of convex functions*, Trans. Amer. Math. Soc., 139 (1969), pp. 443–467.

[2] A. AUSLENDER, *Problèmes de Minimax via l'Analyse Convexe et les Inégalités Variationelles: Théorie et algorithmes*, Lecture Notes in Econ. and Math. Systems, 77, Springer-Verlag, 1972.

[3] A. B. BAKUSHINSKII AND B. T. POLYAK, *On the solution of variational inequalities*, to appear.

[4] D. P. BERTSEKAS, *Necessary and sufficient conditions for a penalty method to be exact*, Math. Programming, to appear.

[5] ———, *Multiplier methods: a survey*, Automatica—J. IFAC., March (1976).

[6] A. BRØNSTED AND R. T. ROCKAFELLAR, *On the subdifferentiability of convex functions*, Proc. Amer. Math. Soc., 16 (1965), pp. 605–611.

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

[8] M. R. HESTENES, *Multiplier and gradient methods*, J. Optimization Theory and Appl., 4 (1969), pp. 303–320.

[9] L. V. KANTOROVICH AND G. P. AKILOV, *Functional Analysis in Normed Spaces*, (1950); English transl., Macmillan, New York, 1964.

[10] M. A. KRASNOSELSKII, *Solution of equations involving adjoint operators by successive approximations*, Uspekhi Mat. Nauk, 15 (1960), no. 3 (93), pp. 161–165.

[11] A. V. KRYANEV, *The solution of incorrectly posed problems by methods of successive approximations*. Dokl. Akad. Nauk SSSR, 210 (1973), pp. 20–22 = Soviet Math. Dokl., 14 (1973), pp. 673–676.

[12] B. MARTINET, *Regularisation d'inéquations variationelles par approximations successives*, Rev. Francaise Inf. Rech. Oper., (1970), pp. 154–159.

[13] ———, *Determination approchée d'un point fixe d'une application pseudo-contractante*, C.R. Acad. Sci. Paris, 274 (1972), pp. 163–165.

[14] G. J. MINTY, *Monotone (nonlinear) operators in Hilbert space*, Duke Math. J., 29 (1962), pp. 341–346.

[15] ———, *On the monotonicity of the gradient of a convex function*, Pacific J. Math., 14 (1964), pp. 243–247.

[16] J. J. MOREAU, *Fonctionelles sous-différentiables*, C.R. Acad. Sci. Paris, 257 (1963), pp. 4117–4119.

[17] ———, *Sur la fonction polaire d'une fonction sémi-continue supérieurement*, C.R. Acad. Sci. Paris, 258 (1964), pp. 1128–1131.

[18] ———, *Proximité et dualité dans un espace Hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.

[19] ———, *Fonctionelles Convexes*, lecture notes, Séminaire "Equations aux derivées partielles", Collège de France, Paris, 1966–67.

[20] B. T. POLYAK AND N. V. TRETYAKOV, *On an iterative method of linear programming and its economic interpretation*, Ekon. Mat. Met., 8 (1972), pp. 740–751.

[21] M. J. D. POWELL, *A method for nonlinear optimization in minimization problems*, Optimization, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298.

[22] R. T. ROCKAFELLAR, *Extension of Fenchel's duality theorem*, Duke Math. J., 33 (1966), pp. 81–89.

[23] ———, *Level sets and continuity of conjugate convex functions*, Trans. Amer. Math. Soc., 123 (1966), pp. 46–63.

[24] ———, *Monotone operators associated with saddle functions and minimax problems*, Nonlinear Functional Analysis, Part 1, F. E. Browder, ed., Symposia in Pure Math., vol. 18, Amer. Math. Soc., Providence, R.I., 1970, pp. 397–407.

[25] ———, *Local boundedness of nonlinear monotone operators*, Michigan Math. J., 16 (1969), pp. 397–407.

[26] ———, *Convex Analysis*, Princeton University Press, Princeton, N.J., 1970.

[27] ———, *On the maximality of sums of nonlinear monotone operators*, Trans. Amer. Math. Soc. 149 (1970), pp. 75–88.

[28] ———, *Saddle functions and convex analysis*, Differential Games and Related Topics, H. W. Kuhn and G. P. Szego, eds., North-Holland, Amsterdam, 1971, pp. 109–128.

[29] ———, *The multiplier method of Hestenes and Powell applied to convex programming*, J. Optimization Theory and Appl., 12 (1973), pp. 555–562.

[30] ———, *Conjugate Duality and Optimization*, Regional Conference Series in Applied Mathematics No. 16, Society for Industrial and Applied Mathematics, Philadelphia, 1974.

[31] ———, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. of Operations Research, 1976, to appear.

[32] Z. OPIAL, *Weak convergence of the successive approximations for nonexpansive mappings in Banach spaces*, Bull. Amer. Math. Soc., 73 (1967), pp. 591–597.

[33] A. GENEL AND L. LINDENSTRAUSS, *An example concerning fixed points*, Israel J. Math., 20 (1975).