MITLibraries
Document Services

Room 14-0551
77 Massachusetts Avenue
Cambridge, MA 02139
Ph: 617.253.5668 Fax: 617.253.1690
Email: docs@mit.edu
http://libraries.mit.edu/docs

DISCLAIMER OF QUALITY

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available. If you are dissatisfied with this product and find it unusable, please contact Document Services as soon as possible.

Thank you.

Due to the poor quality of the original document, there is some spotting or background shading in this document.

August 1986 LIDS-P-1598

THE NONLINEAR PROXIMAL POINT ALGORITHM

JAVIER LUQUE

Department of Mathematical Sciences
University of North Florida
4567 St. Johns Bluff Rd. So.
Jacksonville, FL 32216

August 12, 1986

Submitted to SIAM Journal on Control and Optimization

THE NONLINEAR PROXIMAL POINT ALGORITHM*

Abstract. A new Algorithm, the nonlinear proximal point algorithm (NPA) is introduced. Let T be a maximal monotone map on a real Hilbert space. The Proximal Point Algorithm (PPA) for the solution of $0 \in Tz$, is the iteration $z^{k+1} = P(c_k T, I) z^k$, where $P(c_k T, I)$ is the proximal map of $c_k T$ with respect to the identity I, and $\{c_k\}$ is a sequence of positive real numbers. In the NPA a monotone map S is substituted for I yielding $z^{k+1} \in P(c_k T, S) z^k$. The object is to control the speed of convergence throuhg S. The set of maps S is identified by requiring that the NPA be globally weakly convergent. The growth properties of S and T^{-1} in a neighborhood of zero are used to characterize the asymptotic convergence for both exact and approximate versions. If those growths are bounded by power functions with exponents $s, t > 0$, respectively, with $st \geq 1$, the convergence is linear, superlinear, or in finitely many steps --which can be reduced to one-- depending on whether $st = 1$, $st > 1$, or $t = \infty$. If $st = 1$, and $\lim_k c_k = \infty$, superlinear convergence obtains. Sufficient conditions for sublinear convergence are also given. It is shown how thw criterion for approximate computation can be implemented when T is strongly monotone. Both versions of the NPA are applied to minimizing convex functions, and finding saddle points of convex-concave saddle functions. The spped with which minima and saddle values are approached is determined.

*This research is part of the author's dissertation written under the supervision of Prof. Dimitri P. Bertsekas of The Massachusetts Institute of Technology. It was supported by the ITP Foundation, Madrid, Spain, and the National Science Foundation under grant ECS 8217668.

Key words. monotone maps, algorithms, global convergence, asymptotic convergence, convex functions, saddle functions

Classifications. AMS (MOS) 39, 46, 47, 49, 52, 90.

1. Introduction. Let H denote a real Hilbert space with inner product $(\cdot,\cdot)$: $H \times H \to \mathbb{R}$, and induced norm $|x| = (x,x)^{\frac{1}{2}}$. $T: H \to 2^H$ is a monotone map, if and only if, $y_i \in Tx_i$ $(i = 1,2)$ implies that $(y_2 - y_1, x_2 - x_1) \geqq 0$. T is maximal monotone iff it is monotone and its graph, gph $T = \{(x,y) \in H \times H | y \in Tx\}$, is not properly contained in the graph of any other monotone map. Other concepts associated with T are its effective domain, dom $T = \{x \in H | Tx \neq \emptyset\}$, its range, ran $T = \cup\{Tx | x \in H\}$, and its inverse, which is defined by $u \in T^{-1}v$ iff $v \in Tu$. Elementarily, $(T^{-1})^{-1} = T$, dom T = ran T^{-1}, and gph $T^{-1} = \{(x,y) \in H \times H | (y,x) \in \text{gph } T\}$. $\bar{M}(H)$ will denote the set of maximal monotone maps from H into 2^H.

Let f be a proper closed convex function defined on a real Hilbert space. Its subdifferential map ∂f is maximal monotone (Moreau 1965, p 296, prop 12.b). $\bar{x} \in H$ is a global minimum of f iff $0 \in \partial f(\bar{x})$. Thus $\bar{x}$ is a solution of $0 \in Tx$, where $T = \partial f \in \bar{M}(H)$. Many other problems in convex analysis such as finding saddle points, and solving variational inequalities, can be expressed abstractly as the problem of finding a root of a maximal monotone map defined suitably.

In 1962, Minty proved by methods of convex analysis the existence of solutions of nonlinear functional equations of monotone type. Since then an intensive development has taken place. General references are Brézis (1973), Deimling (1985), Joshi and Bose (1985), and Pascali and Sburlan (1978).

A fundamental algorithm for solving $0 \in Tz$ is the Proximal Point Algorithm (PPA). The PPA is based on a theorem of Minty (1962, p 344, cor, see also Brézis 1973, p 21, prop 2.1), stating that $T \in \bar{M}(H)$ iff for all $c > 0$, $(I + cT)^{-1}$ is defined on all of H and is nonexpansive, thus single-valued. A map $P: H \to 2^H$ is nonexpansive iff whenever $y \in Px$, $y' \in Px'$, $|y - y'| \leq |x - x'|$.

The proximal map of S with respect to T, P(S,T), where $S, T: H \to 2^H$, was introduced (in greater generality) by the author in Luque (1984a), see also Luque (1986a,b). It is given by

$$\forall x \in H, \quad P(S,T)x = \{u \in H | Su \cap T(x - u) \neq \emptyset\}.$$

One can show (ibid.) that $P(S,T) = (I + T^{-1} \circ S)^{-1}$. Thus $(I + cT)^{-1}$ is the proximal map of cT with respect to I.

From any starting point $z^0 \in H$, the PPA generates a sequence z^k according to the rule $z^{k+1} \cong P_k z^k$, where $P_k = (I + c_k T)^{-1}$ and $\{c_k\}$ is some sequence of positive real numbers. The criterion used for the approximate computation of z^{k+1} is

$$(A_r) \quad \begin{aligned} &|z^{k+1} - P_k z^k| \leqq \varepsilon_k \min \{1, |z^{k+1} - z^k|^r\}, \\ &r \geqq 1, \ \varepsilon_k \geqq 0, \ \Sigma_{k=0}^{\infty} \varepsilon_k < \infty. \end{aligned}$$

This criterion was introduced by Rockafellar (1976a) with r = 1. Luque (1984) used $r \geqq 1$ in order to show superlinear convergence. Rockafellar (1976a, th 1) has shown that if $T^{-1}(0) \neq \emptyset$, $|z^{k+1} - P_k z^k| \leqq \varepsilon_k$ is sufficient for $z^{k+1} - z^k \to 0$. Therefore, the larger r is, the more accurate the computation of z^{k+1} will be.

In the review of the known main results -which will be obtained in this paper as special cases- on the convergence and the speed of convergence of the PPA that follows, it will assumed that $T^{-1}(0) = \bar{Z} \neq \emptyset$. $\{z^k\}$ will denote any sequence generated by the algorithm with criterion (A_r), and a sequence of positive real numbers $\{c_k\}$ bounded away from zero. Rockafellar (1976a, p 883, th 1) proved that any such sequence $\{z^k\}$ is bounded and converges weakly to a unique $z^\infty \in \bar{Z}$.

The solution set $\bar{Z} = T^{-1}(0)$ is a closed convex subset of H (Minty 1964, th 1). If $\bar{Z} \neq \emptyset$, for any $z \in H$, the vector in $\bar{Z}$ closest to z will be denoted by $\bar{z}$. The distance from z to $\bar{Z}$ satisfies

$$d(z,\bar{Z}) = \min \{|z - z'|:z' \in \bar{Z}\} = |z - \bar{z}|.$$

When studying the asymptotic convergence of the PPA, attention is focused on the sequence $d(z^k,\bar{Z})$ corresponding to any sequence $\{z^k\}$ as above. The most useful hypotheses concern the growth properties of T^{-1}, in a neighborhood of zero, away from the solution set. The general form of these growth conditions is

$$\exists \delta > 0,\ \forall w \in \delta B,\ \forall z \in T^{-1}w,\ d(z,\bar{Z}) \leqq \tau(|w|), \tag{1.1}$$

where $B = \{x \in H|\ |x| \leqq 1\}$, and $\tau:[0,\infty) \to [0,\infty)$ is such that $\tau(0) = 0$ and is continuous at zero. This asssumption was first introduced in Luque (1984b, p 280). It is almost equivalent, when T is a subdifferential map, to a growth condition on concave functions used by Kort and Bertsekas (1976, p 278, A6), in fact, Bertsekas suggested it to me. It must also be noted that Rockafellar (1976a, p 885, th 2, p 888, th 3) used growth conditions which are particular cases of (1.1) (Luque 1984b, p 289, prop 3.5). Luque (1984b, p 281, prop 1.2) proved that when T satisfies (1.1), $d(z^k,\bar{Z}) \to 0$. If $\bar{Z}$ is a singleton, it follows that z^k converges strongly to the unique solution of $0 \in Tz$. But (1.1) is not necessary for the strong convergence of the PPA (ibid., p 281).

Let $\tau(x) = ax^t$ with $a,t,x \in \mathbb{R}_+$. If $a > 0$, $t = 1$, $\{c_k\}$ is nondecreasing, and $\bar{Z}$ is a singleton, Rockafellar (1976a, p 885, th 2) has proved that $d(z^k,\bar{Z}) \to 0$ linearly at a rate bounded by $a/(a^2 + c^2)^{\frac{1}{2}}$, where $\lim_{k\to\infty} c_k = c_\infty \leqq \infty$, and superlinearly if $c_\infty = \infty$. Luque (1984b, p 281, th 2.1) extended this result to general $\bar{Z} \neq \emptyset$, and showed that the bound is tight (ibid. p 282, example.

If $a > 0$, $t \in [1,\infty)$, and $\{c_k\}$ is nondecreasing, Luque (1984b, p 285, th 3.1) has proved that the (Q-) order of convergence is at least min $\{r,t\}$. To operate the algorithm exactly, i.e., $z^{k+1} = P_k z^k$, is equivalent to $r = \infty$, in that case the (Q-) order of convergence is at least t.

If $a = 0$ and $\bar{Z}$ is a singleton, Rockafellar (1976a, p 888, th 3) has proved that the approximate algorithm converges superlinearly even when $c_\infty < \infty$, and the exact one does so in finitely many steps. Luque (1984b, p 287, th 3.2) extended this result to general $\bar{Z} \neq \emptyset$, showing that the approximate algorithm has a (Q-) order of convergence of at least r, while the exact one converges in finitely many steps, giving also a sufficient condition for the convergence in one step (ibid., p 288).

If T is such that

$$\forall a > 0, \exists \delta > 0, \forall w \in \delta B, \forall z \in T^{-1}w, d(z,\bar{Z}) \geqq a|w|,$$

and $\{c_k\}$ is nondecreasing with $c_\infty < \infty$, the convergence cannot be faster than sublinear (Luque 1984b, p 290, th 4.1). If (1.1) is satisfied for some $a > 0$ and $t \in (0,1)$, $\{c_k\}$ is nondecreasing, and the algorithm is exact, then the convergence is faster than $k^{-t/2}$, i.e., $o(k^{-t/2})$ (ibid., p 291, th 4.2)

As seen above, at each step, the PPA computes the next iterate according to the rule

$$z^{k+1} \cong P_k z^k = (I + c_k T)^{-1} z^k,$$

or

$$z^{k+1} \cong P_k z^k, \ \{P_k z^k\} = P(c_k T, I) z^k = P(T, c_k^{-1} I) z^k,$$

where we have used the fact that (Luque 1984a, 1986b)

$$(I + c_k T)^{-1} = P(c_k T, I) = P(T, c_k^{-1} I)$$

where (ibid.)

$$P(c_k T, I) z^k = \{x \in H | c_k Tx \cap I(z^k - x) \neq \emptyset\},$$

and that $(I + c_k T)^{-1}$ is single-valued (Minty 1962, p334, cor).

In this chapter we generalize the PPA by substituting a maximal monotone map $S \in \bar{M}(H)$ for I. Then $P(c_k T, S)$ is no longer single-valued, and the expression of the Nonlinear Proximal Point Algorithm (NPA) becomes

$$z^{k+1} \cong x^{k+1}, \ x^{k+1} \in P(c_k T, S) z^k = P(T, c_k^{-1} S) z^k = (I + S^{-1} c_k T)^{-1} z^k.$$

The modification of criterion (A_r) is straightforward, one requires that at each step, x^{k+1} satisfy

$$|z^{k+1} - x^{k+1}| \leqq \varepsilon_k \min\{1, |z^{k+1} - z^k|^r\},$$

where r and $\{\varepsilon_k\}$ are as originally.

The motivation for such an algorithm comes from applications of the PPA to convex programming. Rockafellar (1973 , p 560) noted the connection between the method of multipliers of Hestenes and Powell, and the theory of proximal maps of Moreau(1965). Later Rockafellar (1976b) showed that the method of multipliers applied to a convex program with a quadratically augmented Lagrangian is a realization of the PPA in which T $= -\partial g$, g being the essential objective function of the ordinary dual program. There he showed that the Lipschitz continuity at 0 of $(-\partial g)^{-1}$ implies the linear convergence of such an algorithm, superlinear if $c_k \to \infty$. Kort and Bertsekas (1972, 1973, 1976) generalized the method of multipliers by introducing a much wider class of Augmented Lagrangians in which the augmenting terms are not quadratic. They showed that the (Q-) order of convergence depends not only on the problem at hand, i.e., on the growth properties of g, but also on the growth of the penalty function used in augmenting the Lagrangian. Bertsekas suggested me that the same would happen in the PPA if the identity map I (which is the subdifferential of $\frac{1}{2}|\cdot|^2$), were replaced by some other appropriate maximal monotone map.

Section 2 starts by describing the Nonlinear Proximal Point Algorithm, and a list of desirable characteristics that the NPA is to possess is given. These include that the algorithm be globally convergent. These characteristics are then used to specify the class of maps S to be used in the NPA.

The main characteristic of the maps S to be used is that $y \in Sx$ implies that $y = \lambda x$ for some $\lambda > 0$ (definition 2.11). This assumption might seem too strong, but we show that when it is not satisfied it is possible to construct easy examples in $\mathbb{R}^2$ for which the NPA diverges.

The section concludes by showing the weak global convergence of the approximate NPA towards some root of T (theorem 2.12). When T satisfies condition 1.1, $d(z^k,\bar{Z}) \to 0$ (theorem 2.13).

Section 3 is devoted to the study of the asymptotic convergence of the NPA. Under the assumption that the growths of T^{-1} and S are bounded by power functions of the form ax^t and bx^s respectively with $st \geqq 1$, theorem 3.1 proves the following facts. If $st = 1$, the exact and approximate versions of the NPA converge linearly at a rate bounded above by $a/(a^2 + (\underline{c}/b)^{2t})^{\frac{1}{2}}$ with $\underline{c} = \liminf_{k \to \infty} c_k$, thus superlinearly if $\underline{c} = \infty$. The (Q-) order of convergence is at least min $\{r,st\}$. Thus without requiring that $\underline{c} = \infty$ any (Q-) order can be achieved for s large enough.

When the function τ appearing in (1.1) is flat in a neighborhood of zero, the NPA converges superlinearly if $r = 1$ and in any case the (Q-) order is at least r. The exact algorithm converges in finitely many steps which can be reduced to one. This generalizes a result of Bertsekas (1975).

Upper bounds on the speed of convergence are given in theorem 3.3. In particular a condition implying sublinear convergence is given.

In section 4 it is shown how criterion (A_r) can be implemented when T is strongly monotone.

Section 5 deals with the application of the NPA to convex and saddle functions. In these cases it is possible to estimate the speed at which the infimum, or the saddle values are approached (cf Rockafellar 1976a, p 896, §§4,5).

2. The Nonlinear Proximal Point Algorithm. The Nonlinear Proximal Point Algorithm (NPA) in its exact form can be partially specified as follows: Given the equation $0 \in Tz$ where $T \in \bar{M}(H)$, the set of all maximal monotone maps from H into 2^H, select S from a suitable subset of $\bar{M}(H)$ and

(1) Pick $z^0 \in H$, $c_0 > 0$.

(2) Given $z^k \in H$, $c_k > 0$, find

$$z^{k+1} \in P(c_k T,S)z^k = \{x \in H | c_k Tx \cap S(z^k - x) \neq \emptyset\}.$$

(3) If $z^{k+1} = z^k$ stop, otherwise set $k = k+1$, pick $c_{k+1} > 0$, and go to (2).

To complete the specification of the NPA it is necessary to select a subset of $\bar{M}(H)$ such that for any S belonging to it, the NPA will produce sequences $\{z^k\}$ which converge towards the solution set in some sense, and such that the termination criterion implicit in step (3) is valid. In addition there might be other desirable properties that we want the NPA to satisfy.

Concretely the set of suitable maps S can be determined by requiring that the corresponding NPA satisfy several of the following conditions. Throughout $T \in \bar{M}(H)$ and $c > 0$.

(C1) For any $z^0 \in H$, the NPA generates a sequence $\{z^k\}$. Equivalently, for any $z \in H$, the NPA produces a nonempty set $P(cT,S)z$, from which to choose $(z)^{+1}$, the next iterate. Thus $\text{dom}\, P(cT,S) = H$.

(C2) Same as (C1) but restricted to those maps $T \in \bar{M}(H)$ such that $T^{-1}(0) \neq \emptyset$.

(C3) The fixed points of $P(cT,S)$ are precisely the elements of $\bar{Z} =$

$T^{-1}(0)$, i.e., $z \in P(cT,S)z$ iff $0 \in Tz$.

(C4) Once the solution set is reached, it is never left, i.e., $P(cT,S)\bar{Z} \subseteq \bar{Z}$.

(C5) The stopping creiterion implicit in step (3) is valid for any element of $T^{-1}(0)$, i.e., $P(cT,S)z = \{z\}$ iff $0 \in Tz$.

(C6) The map $P(cT,S)$ is single-valued on H.

(C7) If $\bar{Z} = T^{-1}(0) \neq \emptyset$, for any $z^0 \in H$, the sequence $\{z^k\}$ generated by the NPA should converge, in some sense, towards $\bar{Z}$.

Clearly not all of these conditions are independent. (C1) implies (C2), (C5) implies (C3) and (C4). The minimum set of conditions on S so that the algorithm be as specified in (1)-(3) is (C1) (or (C2) if it is known that $T^{-1}(0) \neq \emptyset$), (C5), and (C7). (C6) losses its meaning for the approximate version because then z^{k+1} is selected from a certain neighborhood of $P(c_kT,S)z^k$.

We now turn to the propositions that are sufficient for each of (C1)-(C7) to hold. Note that for any $c > 0$, $(cT)^{-1}(0) = T^{-1}(0)$, and that $T \in \bar{M}(H)$ iff $(cT) \in \bar{M}(H)$. Thus we can consider the constant c included in T, and will always do so whenever it is possible.

There are several approaches to satisfying (C1),(C2). If all that is known is that $T \in \bar{M}(H)$, then a sufficient condition on S is sought, so that for every $T \in \bar{M}(H)$, dom $P(T,S) = H$. If more is known about T, it is possible to select some subset of $\bar{M}(H)$ to which T belongs to. Then a sufficient condition on S is sought, so that for every T in the above subset of $\bar{M}(H)$, dom $P(T,S) = H$. Finally, it is possible to search for a sufficient condition on both S and T, in order that dom $P(T,S) = H$.

The first approach corresponds to (C1). The second one includes (C2), just consider the set of $T \in \bar{M}(H)$ such that $0 \in \text{ran } T$. The third one puts S and T on equal footing, which is not very appropriate for our problem as we are trying to solve $0 \in Tz$ for, in principle, arbitrary maps $T \in \bar{M}(H)$. In order to proceed we need the following

Definition 2.1. (See Pascali and Sburlan 1978, p 247.) $T \in \bar{M}(H)$ satisfies condition (*) iff

$$(*) \quad \forall x \in \text{dom } T,\ \forall u \in \text{ran } T,\ \inf \{(z-x, w-u) \mid w \in Tz\} > -\infty.$$

Clearly T satisfies (*) iff so does T^{-1}. The next proposition implies (C1).

Proposition 2.2. Let $S \in \bar{M}(H)$ be such that dom S = ran S = H, and S satisfies (*). Then for all $T \in \bar{M}(H)$, dom $P(T,S) = H$.

Proof. dom $P(T,S) = \text{ran } (T^{-1} + S^{-1})$ see Luque (1984a, ch II, §2, and 1986b, §2) . As $\text{int}(\text{dom } S^{-1}) = H$ and $\text{dom } T^{-1} \neq \emptyset$, $T^{-1} + S^{-1} \in \bar{M}(H)$. Also $\text{dom } T^{-1} \subseteq \text{dom } S^{-1}$, (thus see Pascali and Sburlan 1978, p 249, th 3.2)

$$\text{int}(\text{ran}(T^{-1} + S^{-1})) = \text{int}(\text{ran } T^{-1} + \text{ran } S^{-1}) = H. \qquad \text{QED}$$

Remark. $\text{ran } (T^{-1} + S^{-1}) \subseteq \text{ran } T^{-1} + \text{ran } S^{-1}$. Thus $\text{ran } (T^{-1} + S^{-1}) = H$ implies that $\text{ran } S^{-1} = H$. Furthermore, as $\text{dom } (T^{-1} + S^{-1}) = \text{dom } T^{-1} \cap \text{dom } S^{-1}$ has to be nonempty for all $T \in \bar{M}(H)$, it follows that $\text{dom } S^{-1} = H$. This shows the necessity of assuming that dom S = ran S = H. However this latter condition is not sufficient for dom $P(T,S) = H$ for every $T \in \bar{M}(H)$. One can show with the following example (ibid., p 245). Let $H = \mathbb{R}^2$, T be counterclockwise rotation by an angle of $\pi/2$, and S clockwise rotation by the same angle. Both T and S are bijective, but dom $P(T,S) =$

ran $(T^{-1} + S^{-1}) = \{0\}$, as $S^{-1} = (-T)^{-1} = -T^{-1}$. Clearly S does not satisfy (*). However subdifferential maps always do (ibid., p 247, ex 1), and the following has been proved.

Corollary 2.3. Let $g \in \Gamma_o(H)$. dom g = dom g* = H iff for all $T \in \bar{M}(H)$, dom $P(T,\partial g) = H$.

If S satisfies the hypotheses of proposition 2.2, for any $T \in \bar{M}(H)$, dom T + int(dom S) = H, and (Luque 1984a, 1986b) $P(T,S)z$ will be a nonempty closed convex set for any $z \in H$. (A_r) will then be satisfied iff

$$d(z^{k+1},P(c_kT,S)z^k) \leqq \varepsilon_k \min\{1,|z^k - z^{k+1}|^r\}$$

with

$$d(z^{k+1},P(c_kT,S)z^k) = \min\{|z^k - x| : x \in P(c_kT,S)z^k\},$$

where the use of min is justified by the fact that $P(c_kT,S)z^k$ is closed convex and nonempty. Henceforth, x^{k+1} will denote the orthogonal projection of z^{k+1} onto $P(c_kT,S)z^k$, then one has

$$d(z^{k+1},P(c_kT,S)z^k) = |z^{k+1} - x^{k+1}|,$$

$$c_kTx^{k+1} \cap S(z^k - x^{k+1}) \neq \emptyset.$$

The next proposition gives a necessary and sufficient condition for condition (C3) to hold.

Proposition 2.4. Let $S \in \bar{M}(H)$, then

(1) $S(0) \subseteq \{0\} \iff [\forall T \in \bar{M}(H),\ z \in P(T,S)z \Rightarrow 0 \in Tz]$,

(2) $\{0\} \subseteq S(0) \iff [\forall T \in M(H),\ 0 \in Tz \Rightarrow z \in P(T,S)z]$.

Proof. The forward direction of (1) and (2) is proved using the equivalence (Luque 1984a, ch II, §2, 1986b, §2)

$$z \in P(T,S)z \iff Tz \cap S(0) \neq \emptyset \iff z \in T^{-1}(S(0)).$$

To prove the backward implications, the argument in Luque (1984a, 1986b) has to be modified as T is maximal monotone and not just an arbitrary map from H into 2^H. Moreover it is more interesting to prove them even if we restrict ourselves to maps T such that $T^{-1}(0) \neq \emptyset$. Let I be the identity map. Let $w \in S(0)\setminus\{0\}$, select any $z \in H$ and set $T = I(\cdot - z) + w$. Clearly $0 \in Tz$, but $w \in Tz \cap S(z - z)$, thus $z \in P(T,S)z$, and the proof of (1) is concluded. If $0 \notin S(0)$, pick any $z \in H$ and set $T = I(\cdot - z)$, then $0 \in Tz$, but $z \notin P(T,S)z$, otherwise $0 \in S(0)$. QED

Remark. This proposition is valid in the context of maps from a real Banach space X into the power set of its dual X*. One has to substitute the normalized duality map corresponding to the norm on X, J for I.

The second part of the next proposition gives a sufficient condition for (C4).

Proposition 2.5. Let $S \in \bar{M}(H)$, then

(1) $0 \in S(0) \iff [\forall T \in \bar{M}(H),\ \bar{Z} \subseteq P(T,S)\bar{Z}]$,

(2) If S is such that $w \in Sz$ and $z \neq 0 \neq w$ imply $(z,w) > 0$ (the maximality of S implies $0 \in S(0)$), then for all $T \in M(H)$, $P(T,S)\bar{Z} \subseteq \bar{Z}$.

Proof. The direct implication of (1) is a straightforward consequence of proposition 2.4(2) via

$$\bar{Z} = \{z | 0 \in Tz\} \subseteq \bigcup \{P(T,S)z | 0 \in Tz\} = P(T,S) \circ T^{-1}(0).$$

When $0 \in S(0)$ we saw that for any $z \in H$, $T = I(\cdot - z)$ is such that $0 \in Tz$ and $z \in P(T,S)z$. But $T^{-1}(0) = \{z\}$, thus $\bar{Z} \not\subseteq P(T,S)\bar{Z}$.

To prove (2), one has

$$z \in P(T,S) \circ T^{-1}(0) \Longleftrightarrow [\exists x \in T^{-1}(0), \exists w \in Tz \cap S(x - z)].$$

The monotonicity of T implies $(x - z, 0 - w) \geqq 0$, but $w \in S(x - z)$ and as $0 \in S(0)$, $(x - z - 0, w - 0) \geqq 0$. Thus $(x - z, w) = 0$, by the assumption on S, either $x = z$ and/or $w = 0$, in either case $0 \in Tz$. QED

<u>Remark.</u> This proposition is valid in the context of maps from a real Banach space X into the parts of its dual X*. One substitutes J, the normalized duality map of the norm of X, for the identity map I, in the proofs.

The next proposition gives a sufficient condition for the validity of the stopping criterion of step (3) of the algorithm as explained in condition (C5).

<u>Proposition 2.6.</u> Let $S \in \bar{M}(H)$, then

(1) $S(0) \subseteq \{0\} \Rightarrow [\forall T \in \bar{M}(H),\ P(T,S)z = \{z\} \Rightarrow 0 \in Tz]$.

(2) $S^{-1}(0) \subseteq \{0\}$, and $w \in Sz$ with $w \neq 0 \neq z$ implies that $(z,w) > 0$ (by the maximality of S, it follows that $0 \in S(0)$), then for all $T \in \bar{M}(H)$, $0 \in Tz$ implies that $P(T,S)z = \{z\}$.

<u>Proof.</u> (1) is immediate from proposition 2.4. To prove (2), let $T \in \bar{M}(H)$, $0 \in Tz$, as $0 \in S(0)$, clearly $z \in P(T,S)z$. Let $x \in P(T,S)z$, then

there is a vector $v \in Tx \quad S(z - x)$. The monotonicity of T implies that $(z - x, 0 - v) \geqq 0$, while the monotonicity of S implies $(z - x - 0, v - 0) \geqq 0$. Thus $(z - x, v) = 0$ and $z = x$ or $v = 0$. If $v = 0$, then $z - x \in S^{-1}(0) \subseteqq \{0\}$, and $z = x$. QED

Remark. A monotone map is strictly monotone at $x_0 \in \text{dom } S$ iff for all $y_0 \in Sx_0$, and all $(x,y) \in \text{gph } S$

$$(x - x_0, y - y_0) = 0 \implies x = x_0.$$

In particular, if S is such that $0 \in S(0)$, then its strict monotonicity at 0 implies that for all $(x,y) \in \text{gph } S$, $x \neq 0 \implies (x,y) > 0$. This condition implies in turn that $S^{-1}(0) = \{0\}$, otherwise by picking $x \in S^{-1}(0)$ $\{0\}$ one would contradict it. Thus strict monotonicity at 0 is stronger than the hypotheses on S of both propositions 2.5(2) and 2.6(2). Analogously one shows that strict monotonicity of S^{-1} at 0 implies the hypothesis on S of proposition 2.6(2).

The following example shows how the conclusions of propositions 2.5(2), 2.6(2) fail if S does not satisfy that whenever $w \in Sz$ with $w \neq 0 \neq z$, then $(z,w) > 0$.

Example. Let $H = \mathbb{R}^2$, and let T,S be rotations by $\pi/2$ and $-\pi/2$ respectively. For all $x \in \mathbb{R}^2$, $Tx = -Sx = S(-x)$. Also $T(0) = T^{-1}(0) = \{0\} = S(0) = S^{-1}(0)$. Clearly, whenever $y \in Sx$, $(y,x) = 0$, so S does not satisfy the hypotheses of either 2.5(2) or 2.6(2). Then

$$P(T,S)(0) = \{x \in \mathbb{R}^2 \mid Tx \cap S(-x) \neq \emptyset\} = \{x \in \mathbb{R}^2 \mid Tx = Tx\} = \mathbb{R}^2.$$

But $\{0\} = T^{-1}(0) = \bar{Z}$, thus we have shown that $\bar{Z} = \{0\} \neq \mathbb{R}^2 = P(T,S)\bar{Z}$ (cf. proposition 2.5(2)), and that $0 \in T(0)$ but $P(\ddot{T},S)(0) = \mathbb{R}^2$ (cf. proposition 2.6(2)).

In some applications it is desirable that the maps $P(c_k T,S)$ be everywhere single-valued, not just on $\bar{Z}$. In the next proposition, a necessary and sufficient condition for (C6) is given.

Proposition 2.7. Let $S \in \bar{M}(H)$. S is strictly monotone iff for all $T \in \bar{M}(H)$, $P(T,S)$ is single-valued.

Proof. The sufficiency follows from proposition II.4.3. To show the necessity, let $w_i \in Sz_i$ $(i = 1,2)$ and $(z_1 - z_2, w_1 - w_2) = 0$. Consider the monotone map defined as follows. For all $z \in H$

$$Tz = \begin{cases} \{(1-t)w_1 + tw_2\}, & \text{if } z = (1-t)(-z_1) + t(-z_2) \text{ for some } t \in [0,1], \\ \emptyset, & \text{otherwise.} \end{cases}$$

Zorn's lemma implies the existence of an extension of T, $\bar{T} \in \bar{M}(H)$ such that

$$T|[-z_1,-z_2] \subseteq \bar{T}|[-z_1,-z_2]$$

with dom $\bar{T} \subseteq [-z_1,-z_2]$, actually equal as dom $T = [-z_1,-z_2]$ (see Pascali and Sburlan 1978, p 123, th 2.12). Then

$$w_i \in \bar{T}(-z_i) \cap S(0 - (-z_i)) \qquad (i = 1,2),$$

and $-z_1, -z_2 \in P(\bar{T},S)(0)$. QED

The basis of the ordinary Proximal Point Algorithm is the fact that

$$P_k = (I+c_kT)^{-1} = P(c_kT,I)$$

is both everywhere defined and nonexpansive. Rockafellar (1976a) used these facts, together with $z^{k+1} - P_kz^k \in c_kTP_kz^k$, and $T \in \bar{M}(H)$ to prove his fundamental proposition 1 (ibid, p. 881) on which much of his proof of the global convergence (ibid, p. 883, th. 1) rests.

In the present case, $x \in P(c_kT,S)z^k = (I + S^{-1}(c_kT))^{-1}z^k$, and to require that $P(c_kT,S)$ be everywhere defined and nonexpansive is equivalent to requiring that $S^{-1}c_kT \in \bar{M}(H)$ which essentially would reduce the algorithm introduced here to the usual Proximal Point Algorithm.

However it is easy to prove that if $T \in M(H)$ is such that $T^{-1}(O)$ $\emptyset$, then $\forall z \in H$, $Tz \subseteq N(z;T^{-1}(O))$, where $N(z;T^{-1}(O))$ is the normal cone to $T^{-1}(O)$ at z. As $T^{-1}(O) = (c_kT)^{-1}O$, the maps P_k in the ordinary Proximal Point Algorithm satisfy $z^k - P_kz^k \in N(P_kz^k;T^{-1}0)$. The surprising fact is that any sequence $\{z^k\}$ such that for all k, $z^k - z^{k+1} \in N(z^{k+1};C)$, where C is a nonempty closed convex subset of H, will have most of the convergence characteristics of the Proximal Point Algorithm.

Proposition 2.8: Let H be a real Hilbert space, and C a nonempty closed convex subset of H. Let $\{z^n\}$ be some sequence in H such that $\forall n \geq 0$, $z^n - z^{n+1} \in N(z^{n+1};C)$. Then

(1) $z^m \in C \Rightarrow \forall n \geq m$, $z^n = z^m$.

(2) $\forall z \in C$, $\forall n \geq 0$, $|z^n-z|^2 \geq |z^n-z^{n+1}|^2 + |z^{n+1}-z|^2$. Thus $\{z^n\}$ is both bounded and $z^{n+1}-z^n \to 0$.

(3) $\forall z \in C$, $|z^n-z|$ monotonically decreases to $\mu(z)$, where μ is a nonnegative Lipschitz continuous (with modulus 1) function on

C which attains its minimum on C, $\bar{\mu}$, uniquely.

(4) Let $d(z^n,C) = \min\{|z^n-z| : z \in C\}$. Then $d(z^n,C)$ decreases monotonically to d^∞, and $\mu \geq d^\infty$ on C.

(5) $\{z^n\}$ possesses a weakly convergent subsequence. Let $\{z^n\}_{n\in J} \xrightarrow{w} z_J$, then

(5.1) $\lim_{n\in J} |z^n-z_J| = \mu_J < \infty$

(5.2) $\forall z \in C,\ (\mu(z))^2 = \mu_J^2 + |z-z_j|^2$

(5.3) $\bar{z}_J = \bar{z}$, where $\bar{z}_J$ is the orthogonal projection of z_J onto C, and $\bar{z}$ minimizes μ. Thus if all the weak cluster points of $\{z^n\}$ lie on C, $\{z^n\}$ converges weakly to $\bar{z}$.

Proof. (1) As $z^m - z^{m+1} \in N(z^{m+1};C)$, for all $v \in C$, $(z^m-z^{m+1}, z^{m+1}-v) \geq 0$. Setting $v = z^m$ the result follows.

(2) As $z^n - z^{n+1} \in N(z^{n+1};C)$ and $z \in C$,

$$\begin{aligned}|z^n-z|^2 &= |z^n-z^{n+1}+z^{n+1}-z|^2 \\ &= |z^n-z^{n+1}|^2 + |z^{n+1}-z|^2 + 2(z^n-z^{n+1}, z^{n+1}-z) \\ &\geq |z^n-z^{n+1}|^2 + |z^{n+1}-z|^2.\end{aligned}$$

Being $\{|z^n-z|\}$ bounded and decreasing, it converges. Taking limits in both sides above, $z^{n+1} - z^n \to 0$ follows.

(3) In (2) we have seen that $\lim_{n\to\infty} |z^n-z|$ exists. Let $z,z' \in C$, $t \in [0,1]$, then

$$|z^n-(1-t)z-tz'| \leq (1-t)|z^n-z| + t|z^n-z'|.$$

Taking limits, which exist as $(1-t)z + tz' \in C$ by convexity of C, the convexity of μ follows. Let $z, z' \in C$

$$|z^n - z'| \leq |z^n - z| + |z - z'|$$

from which $\mu(z') - \mu(z) \leq |z-z'|$. Interchanging the roles of z and z' one gets $-|z-z'| \leq \mu(z') - \mu(z)$, and thus $|\mu(z) - \mu(z')| \leq |z - z'|$. Let $C_\rho = \{z \in C | \mu(z) \leq \rho\}$, $\rho \geq 0$. For all $z \in C$, $n \geq 0$, $\mu(z) \leq |z^n - z|$, thus $z \in C_\rho$ whenever $|z^n - z| \leq \rho$ for some $n \geq 0$. Let $z_1, z_2 \in C_\rho$,

$$\lim_{n\to\infty} |z^n - z_i| = \mu(z_i) \leq \rho, \quad (i=1,2).$$

Given $\varepsilon > 0$ there are numbers n_1, n_2 such that $n \geq n_i$ implies $|z^n - z_i| \leq \rho + \varepsilon$ $(i=1,2)$. Choosing $n_0 = \max \{n_1, n_2\}$

$$|z_1 - z_2| \leq |z_1 - z^{n_0}| + |z^{n_0} - z_2| \leq \mu(z_1) + \mu(z_2) + 2\varepsilon \leq 2(\rho + \varepsilon).$$

$\varepsilon > 0$ begin arbitrary it follows that $|z_1 - z_2| \leq 2\rho$, and C_ρ is weakly compact. As μ is convex and continuous Weierstrass' Theorem implies that it attains its infimum on C. Let $z_1, z_2 \in C$ and let $z = (z_1+z_2)/2$, then

$$|z^n - z|^2 = \left|\frac{z^n - z_1}{2} + \frac{z^n - z_2}{2}\right|^2$$

$$= \frac{1}{4}|z^n - z_1|^2 + \frac{1}{4}|z^n - z_2| + \frac{1}{2}(z^n - z_1, z^n - z_2).$$

But

$$(z^n - z_1, z^n - z_2) = (z^n - z + \frac{z_1 - z_2}{2}, z^n - z - \frac{z_1 - z_2}{2})$$

$$= |z^n - z|^2 - \frac{1}{4}|z_1 - z_2|^2.$$

Using this equality above and rearranging

$$\frac{1}{2}|z^n-z|^2 = \frac{1}{4}|z^n-z_1|^2 + \frac{1}{4}|z^n-z_2|^2 - \frac{1}{8}|z_1-z_2|^2,$$

taking limits as $n \to \infty$

$$\frac{1}{2}(\mu(z))^2 = \frac{1}{4}(\mu(z_1))^2 + \frac{1}{4}(\mu(z_2))^2 - \frac{1}{8}|z_1-z_2|^2$$

If $\mu(z_1) = \mu(z_2) = \overline{\mu} = \inf_C \mu$, the convexity of μ and C implies that $\mu(z) = \overline{\mu}$, and it follows that $z_1 = z_2$.

(4) Denote by $\overline{z^n}$ the orthogonal projection of z^n onto C, by (2)

$$d(z^n,C) = |z^n-\overline{z^n}| \geq |z^{n+1}-\overline{z^n}| \geq d(z^{n+1},C).$$

Thus $\{d(z^n,C)\}$ is nonincreasing and bounded below thus convergent, let d^∞ be its limit. Also, for all $z \in C$, $d(z^n,C) \leq |z^n-z|$, and taking limits on both sides it follows that $d^\infty \leq \mu(z)$

(5) The boundedness of $\{z^n\}$ implies that it possesses some weakly convergent subsequence. If $\{z^n\}_{n\in J} \overset{w}{\to} z_J$, for all $z \in C$

$$|z^n-z|^2 = |z^n-z_J|^2 + |z_J-z|^2 + 2(z^n-z_J,z_J-z).$$

Taking limits as $n \to \infty$ along J on both sides

$$\forall z \in C,\ (\mu(z))^2 = \mu_J^2 + |z-z_J|^2.$$

Setting $z = \overline{z}$ the unique minimizer of μ on C, and being $\overline{z}_J$ the orthogonal projection of z_J onto C

$$(\mu(\overline{z}_J))^2 = \mu_j^2 + |\overline{z}_J-z_J|^2 \leq \mu_J^2 + |\overline{z}-z_J|^2 = \mu^2,$$

from which $\bar{z}_J = \bar{z}$. If the weak cluster points of $\{z^n\}$ lie on C, they are equal to $\bar{z}$ and then, it is easy to see that $\{z^n\} \overset{w}{\to} \bar{z}$. QED

Remark. $\{z^n\}$ may converge to a limit not in C. Consider in $\mathbb{R}$, $C = (-\infty, -1]$, and $\{z^n\}$ a decreasing sequence of positive numbers, then $\lim_{n\to\infty} z^n$ exists and does not belong to C. The sequences of this type generated by the Proximal Point Algorithm studied here are such that their weak cluster points always belong to C, therefore they are weakly convergent. Whether $\{z^n\}$ converges weakly in general is an open question.

We have shown (Luque 1984a, ch II, §2, 1986b, §2) that

$$I-P(c_kT,S) = P(S,c_kT) = (I+(c_kT)^{-1}S)^{-1} = (I+(S^{-1}c_kT)^{-1})^{-1}.$$

If in addition $S(0) = \{0\}$, from proposition 2.4 and the fact that $c_k > 0$

$$0 \in Tz \iff 0 \in c_kTz \iff z \in P(c_kT,S)z \iff 0 \in P(S,c_kT)z$$

(cf. Rockafellar 1976a, p. 881, eqns (2.1),(2.2)).

In order to show the global convergence of the NPA, I will require that S be such that the sequences generated by the exact version of the algorithm be of the type studied in the above proposition.

Proposition 2.9. Let $S \in \bar{M}(H)$ be such that

$$y \in Sx \Rightarrow (\exists \lambda > 0,\ y = \lambda x).$$

Then $\forall T \in \bar{M}(H)$ with $T^{-1}(0) \neq \emptyset$, $\forall z \in H$, $\forall x \in P(T,S)z$

(1) $z-x \in P(S,T)z \cap N(x;T^{-1}(0))$

(2) $\forall \bar{z} \in T^{-1}(0)$, $(z-x, x-\bar{z}) \geq 0$

(3) $\forall \bar{z} \in T^{-1}(0)$, $|z-\bar{z}|^2 \geq |z-x|^2 + |x-\bar{z}|^2$.

Proof. If $x \in P(T,S)z$, then $Tx \cap S(z-x) \neq \emptyset$, and for all $v \in T^{-1}(0)$, $w \in Tx \cap S(z-x)$, the monotonicity of T implies $(x-v,w) \geq 0$. As $w \in X(z-x)$, $w = \lambda(z-x)$ for some $\lambda > 0$, thus for all $v \in T^{-1}(0)$, $\lambda(z-x,x-v) \geq 0$, as $\lambda > 0$, this is equivalent to $z-x \in N(x;T^{-1}(0))$. This and $P(S,T) = I-P(T,S)$ conclude the proof of (1); while (2) is an immediate consequence. Part (3) follows from (1), (2) by expanding $|z-\bar{z}|^2 = |z-x+x-\bar{z}|^2$. QED

Remark. This proposition generalizes proposition 1 of Rockafellar (1976a, p. 881). The generally multivoque maps $P(c_kT,S)$, $P(S,c_kT)$ play the role of $P_k = P(c_kT,I)$, $Q_k = I-P_k = P(I,c_kT)$ respectively. In part (1) the difference between z and any of its one step iterates x need not belong to the image , under T, of x, but only to the convex cone generated by Tx. Therefore, the sequences generated by one exact version of the algorithm are of the type studied in proposition 3.2.8. Parts (2) and (3) are only stated for the case $z' \in T^{-1}(0)$ (cf. ibid.), which is all which is needed to prove the global convergence of the algorithm.

The assumption of proposition 2.9 may seem too strong at first sight. The following example shows how if it is not satisfied, it is possible to find maps S,T in $\mathbb{R}^2$ such that the exact NPA exhibits radically different convergence behaviours ranging from convergence to divergence towards infinite.

Example. Let $T \in \bar{M}(H)$ and define its moduli of Lipschitz continuity and strong monotonicity, respectively, by

$$\lambda_T = \sup \left\{ \frac{|y_1-y_2|}{|x_1-x_2|} : x_1 \neq x_2,\ y_i \in Tx_i \right\}$$

$$\mu_T = \inf \left\{ \frac{(x_1-x_2, y_1-y_2)}{|x_1-x_2|^2} : x_1 \neq x_2,\ y_i \in Tx_i \right\}$$

Let $\lambda_{\bar{T}}$, $\mu_{\bar{T}}$ respectively denote the modulus of Lipschitz continuity and strong monotonicity of T^{-1}.

Let $0 \in T\bar{z}$, from the above definitions, one has

$$\mu_T|z-\bar{z}|^2 \leq (z-\bar{z}, w-0) = |z-\bar{z}|\,|w| \cos \widehat{(z-\bar{z}, w)},$$

$$|x-0| \leq \lambda_T|z-\bar{z}|,$$

thus $\cos \widehat{(z-\bar{z},w)} \geq \mu_T/\lambda_T$. The same argument for T^{-1} yields $\cos \widehat{(z-\bar{z},w)} \geq \mu_{\bar{T}}/\lambda_{\bar{T}}$. Let $c_T = \max\{\mu_T/\lambda_T, \mu_{\bar{T}}/\lambda_{\bar{T}}\}$, then $1 \geq \cos \widehat{(z-\bar{z},w)} \geq c_T \geq 0$ and if $0 \leq \theta_T = \cos^{-1} c_T$, one has $0 \leq \widehat{(z-\bar{z},w)} \leq \theta_T \leq \frac{\pi}{2}$. Let S be well behaved enough so that $P(T,S)z \neq \emptyset$ and let $x \in P(T,S)z$, then $\exists w \in Tx \cap S(z-x)$. One can apply the above reasoning to the points $(\bar{z},0)$, $(x,w) \in$ graph of T, and to $(0,0)$, $(z-x,w) \in$ graph of S, obtaining

$$0 \leq \widehat{(x-\bar{z},w)} \leq \theta_T \leq \frac{\pi}{2}$$

$$0 \leq \widehat{(z-x-0,w-0)} \leq \theta_S \leq \frac{\pi}{2}$$

So w has to belong to two cones of semiaperture θ_T and θ_S, and axes $x-\bar{z}$ and $z-x$ respectively. From figure 1, it follows that that is possible only if $\phi \leq \theta_S + \theta_T$. Thus one arrives to the following condition on ϕ: $0 \leq \phi \leq \theta_S + \theta_T \leq \pi$. As cos (•) is monotonically decreasing

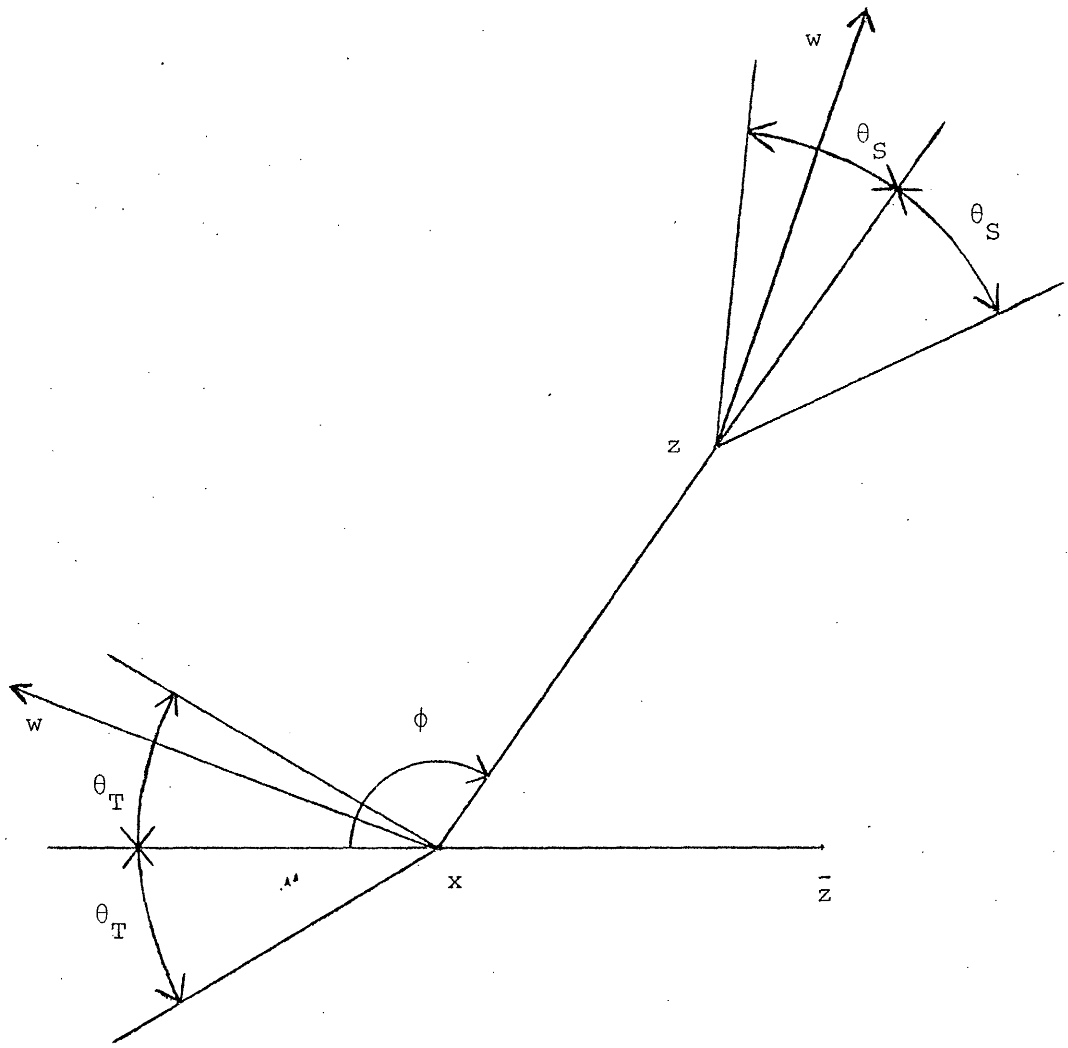

figure 1.

in $[0,\pi]$ we get $1 \geq \cos\phi \geq \cos(\theta_S + \theta_T) \geq -1$,

If one assumes that $\theta_S + \theta_T \leq \pi/2$ then $\phi \leq \pi/2$ and $z-x \in N(x;\{\bar{z}\})$. If $\{\bar{z}\} = T^{-1}(0)$ then we again obtain $z-x \in N(x;T^{-1}(0))$. Otherwise the argument may be easily extended to the general case. That convergence occurs in the former case can be quickly shown as follows.

$$
\begin{aligned}
|z-\bar{z}|^2 &= |z-x|^2 + |x-\bar{z}|^2 + 2(z-x,x-\bar{z}) \\
&= |z-x|^2 + |x-\bar{z}|^2 + 2|z-x||x-\bar{z}|\cos(z-x,x-\bar{z}) \\
&= |z-x|^2 + |x-\bar{z}|^2 + 2|z-x||x-\bar{z}|\cos\phi \\
&\geq |z-x|^2 + |x-\bar{z}|^2,
\end{aligned}
$$

as $\cos\phi \geq 0$.

$$
\begin{aligned}
&w \in S(z-x),\ 0 \in S(0) \Rightarrow |z-x| \geq |w|/\lambda_S, \\
&w \in Tx, \qquad 0 \in T\bar{z} \Rightarrow |x-z| \leq \lambda_{\bar{T}}|w|,
\end{aligned}
$$

using these estimates, it follows that

$$\frac{|x-z|}{|z-\bar{z}|} \leq \frac{1}{\sqrt{1+(\lambda_S\lambda_{\bar{T}})^{-2}}} \leq 1$$

Thus if $(\lambda_S\lambda_{\bar{T}})^{-2} > 0$, linear convergence at the rate $1/\sqrt{1+(\lambda_S\lambda_{\bar{T}})^{-2}}$ occurs.

In the case of the ordinary Proximal Point Algorithm, at each iteration $\lambda_S = c_k^{-1}$. Assuming that T^{-1} is Lipschitz continuous at zero with modulus a, we have

$$\frac{|z^{k+1}-\bar{z}|}{|z^k-\bar{z}|} \leq \frac{a}{\sqrt{a^2+c_k^2}}$$

which coincides with the result reported by Rockafellar (1976a, p 885, th. 2).

To see how anything can happen when one does not assume $\theta_S + \theta_T \leq \pi/2$, let us consider in $\mathbb{R}^2$ the linear operator corresponding to

$$A(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \quad , \qquad \alpha \in [-\frac{\pi}{2}, \frac{\pi}{2}].$$

$A(\alpha)$ notates vectors conterclockwise by an angle of α. This operator is strongly monotone with modulus $\cos\alpha$, and Lipschitz continuous with constant 1. If $a > 0$, $aA(\alpha)$ is strongly monotone with modulus $a\cos\alpha$ and Lipschitz continuous with modulus a. Its inverse is $a^{-1}A(-\alpha)$ with modulus of strong monotonicity $a^{-1}\cos(-\alpha) = a^{-1}\cos\alpha$ and Lipschitz continuous with modulus a^{-1}.

Let us consider $T = tA(\alpha)$, $S = sA(-\beta)$, $s,t > 0$, $\alpha,\beta \in [-\pi/2,\pi/2]$. Being both S, T bijective one can easily see that $x \in P(T,S)z \iff x = (I+S^{-1}T)^{-1}z$. Elementary computations yield

$$(I+S^{-1}T)^{-1} = (I + \frac{t}{s}A(\alpha+\beta))^{-1}$$

$$= \frac{1}{1+\frac{t}{s}\cos(\alpha+\beta)+\frac{t}{s}^2} [I + \frac{t}{s}A(-\beta-\alpha)].$$

If $x = (I+S^{-1}T)z$, then

$$\frac{|x|}{|z|} = \sqrt{1 + 2\frac{t}{s}\cos(\alpha+\beta) + \frac{t^2}{s^2}}$$

As $\mu_T = t\cos\alpha$, $\lambda_T = t$, $\mu_{\bar{T}} = t^{-1}\cos\alpha$, $\lambda_{\bar{T}} = t^{-1}$, $c_T = \{\mu_T/\lambda_T, \mu_{\bar{T}}/\lambda_{\bar{T}}\}$ $= \cos\alpha$ and $\theta_T = \alpha$. Similarly $\theta_S = \beta$.

If $\cos(\theta_S + \theta_T) \geq -1/2\, \lambda_S \lambda_{\overline{T}}$, it follows that $\cos(\alpha+\beta) \geq -t/2s$, and $|z| \geq |x|$. Thus the algorithm does not get further away from the solution set $T^{-1}(0) = \{0\}$.

On the other hand by an appropriate choice of s,t,α,β one can get that $|z| = |x|$, and the sequences generated by the algorithm remain on a circumference of radius $|z|$ about the solution set, or $|z| < |x|$ and the sequences become unbounded with $|z^k| \to \infty$. Setting $\cos(\alpha+\beta) = -t/2s$ one gets $|z| = |x|$. Setting $\cos(\alpha+\beta) > -t/2s$ one gets $|z| < |x|$. In the former case $\{z^k\}$ remains on the circle of radius r about zero, in the second $|z^k| \to \infty$.

We now continue studying the properties of the NPA, using a maximal monotone map S satisfying the hypothesis of proposition 2.9. The next proposition deals with the fixed point properties of the maps $P(T,S)$, where T is the maximal monotone map for which we wish to solve $0 \in Tz$.

Proposition 2.10. Let $S \in \overline{M}(H)$ be such that $y \in Sx$ implies that $y = \lambda x$ for some $\lambda > 0$. Then for all $T \in \overline{M}(H)$

(1) $0 \in Tz \iff P(T,S)z = \{z\}$.

(2) $\overline{Z} = P(T,S)\overline{Z}$.

Proof. Whenever $y \in Sx$, $y = \lambda x$, thus $(x,y) = \lambda|x|^2 \geq 0$, and the maximality of S implies $0 \in S(0)$. If $y \in S(0)$, then for some $\lambda > 0$, $y = \lambda \cdot 0 = 0$, thus $S(0) = \{0\}$. If $0 \in Sx$, for some $\lambda > 0$, $0 = \lambda x$, thus $x = 0$ and $S^{-1}(0) = \{0\}$. If $y \in Sx$, $(x,y) = \lambda|x|^2 > 0$ if $x \neq 0$. Using proposition 2.6, (1) is proved. (2) follows from proposition 2.5, or directly from (1). QED

If S satisfies the hypothesis of the above proposition, it does not follow that dom S = H or ran S = H. Consider in $\mathbb{R}$, S such that

$$\text{gph } S = \{(x,-1): x \leq -1\} \bigcup \{(x,x): -1 \leq x \leq 1\} \bigcup \{(1,x): 1 \leq x\}.$$

It is not clear either whether this type of map is necessarily a subdifferential map or more generally satisfies condition (*).

In view of the above two propositions we now specify the class of maps $S \in \overline{M}(H)$, suitable for use in the NPA.

Definition 2.11. A map $S \in \overline{M}(H)$ is in $\mathcal{S} \subseteq \overline{M}(H)$ iff

(1) S satisfies condition (*).

(2) dom S = ran S = H.

(3) $y \in Sx$ implies $y = \lambda x$ for some $\lambda > 0$.

Among others, $\mathcal{S}$ includes the subdifferential maps of functions of the form $\Phi \circ |\cdot|$, where $\Phi(\rho) = \int_0^\rho \phi$ and $\phi: [0,\infty) \to [0,\infty)$ is any continuous monotonically increasing function with $\phi(0) = 0$. For these type of maps $Sx = \phi(|x|) \cdot \text{sgn } x$, where

$$\text{sgn } x = \begin{cases} x/|x| & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Being subdifferential maps, they satisfy conditon (*). Also, it is clear that their domains and ranges equal H.

The iteration step of the NPA can be written as (see Luque 1984a, ch II, §2, 1986b, §2)

$$z^{k+1} \in P(c_k T, S) z^k = P(S^{-1} \circ (c_k T), I) z^k$$

$$= (I + S^{-1} \circ (c_k T))^{-1} z^k.$$

This is equivalent to the ordinary Proximal Point Algorithm applied to the map $S^{-1} \circ (c_k T)$. If $S^{-1} \circ (c_k T)$ were monotone, the NPA would reduce to the ordinary Proximal Point Algorithm. Even when $S \in \mathcal{S}$ (see definition 2.11) that is not the case, as shown in the next

Example. Let $\alpha, \beta \in (1,\infty)$ satisfy $\alpha^{-1}+\beta^{-1} = 1$, and let $\phi = (\cdot)^{\alpha-1}$. Thus for all $x \in H$

$$Sx = |x|^{\alpha-2} x \ , \ S^{-1} x = |x|^{\beta-2} x.$$

In $H = \mathbb{R}^2$, let T be the map that rotates vectors counterclockwise by an angle of $\pi/2$, and let $u,v \in \mathbb{R}^2$ be of the form

$$u = a\begin{pmatrix}1\\0\end{pmatrix}, \ v = \frac{b}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \ a,b \geq 0.$$

After some algebra one can obtain

$$(S^{-1}Tu - S^{-1}Tv)(u-v) = \frac{ab}{\sqrt{2}} (b^{\alpha-2} - a^{\beta-2}).$$

If $\alpha \neq 2$, one can choose positive values of a,b so that the above expression is negative. This proves that the NPA introduced here is a strict generalization of the ordinary Proximal Point Algorithm, as P(T,S) is not nonexpansive by the aforementioned theorem of Minty.

The global convergence properties of the NPA are given by the next

Theorem 2.12. Let $S \in \mathcal{S}$ and $\sigma : [0,\infty) \to [0,\infty)$ be such that $\sigma(0) = 0$, is continuous at 0 and $\exists \delta > 0, \forall x \in \delta B, \forall y \in Sx, |y| \leq \sigma(|x|)$.

Let $\{z^k\}$ be any sequence generated by the NPA under criterion (A_r) for some $r \geq 1$, when applied to $T \in \overline{M}(H)$ with a sequence of positive real numbers $\{c_k\}$ bounded away from zero. Suppose $\{z^k\}$ is bounded (this holds iff $\overline{Z} = T^{-1}0 \neq \emptyset$). Then, $\{z^k\}$ converges weakly to $z^\infty \in \overline{Z}$.

Proof. Proposition 2.2 and $S \in \mathcal{S}$ imply that dom $P(c_kT,s) = H$ for all k. Let $\bar{z} \in \overline{Z}$ and x^{k+1} be the orthogonal projection of z^{k+1} onto $P(c_kT,S)z^k$. Criterion (A_r) and proposition 2.9 imply

$$|z^{k+1}-\bar{z}| \leq |z^{k+1}-x^{k+1}| + |x^{k+1}-\bar{z}| \leq \varepsilon_k + |z^k-\bar{z}|, \qquad (2.1)$$

from which it follows that $\{z^k\}$ is bounded as

$$|z^{k+1}-\bar{z}| \leq |z^0-\bar{z}| + \sum_{m=0}^{\infty} \varepsilon_m < \infty.$$

Without assuming $\overline{Z} \neq \emptyset$, which will be proved below, let $\{z^k\}$ be any bounded sequence satisfying (A_r) for some $r \geq 1$. Let $c > 0$ be such that

$$\forall k \geq 0, \; |z^k| \leq c, \; \varepsilon_k < c,$$

thus $\{z^k\}$ has at least one weak cluster point $z^\infty \in cB$. Consider the closed ball 2cB, and let h be its convex indicator function. Let $T' = T + \partial h$, $T' = T$ on int $2cB = 2cU$, where U is the open unit ball. For any $k \geq 0$, (A_r) yields

$$|x^{k+1}| \leq |x^{k+1}-z^{k+1}| + |z^{k+1}| \leq \varepsilon_k + |z^{k+1}| < 2c,$$

therefore

$$x^{k+1} \in \text{dom } T \cap 2cU = \text{dom } T \cap \text{int } (\text{dom } \partial h),$$

thus dom T $\cap$ int (dom ∂h) $\neq \emptyset$ and $T' \in \bar{M}(H)$ (Rockafellar 1970a, p. 76, th. 1). Furthermore as $\{x^k\} \subset 2cU$

$$c_k T x^{k+1} \cap S(z^k - x^{k+1}) = c_k T' x^{k+1} \cap S(z^k - x^{k+1})$$

and $x^{k+1} \in P(c_k T', S) z^k$. If y^{k+1} denotes the orthogonal projection of z^{k+1} onto $P(c_k T', S) z^k$, $|z^{k+1} - y^{k+1}| \leq |z^{k+1} - x^{k+1}| \leq \varepsilon_k$, thus $\{z^k\}$ can be considered as generated by the algorithm when applied to T'. Clearly dom T' is bounded, therefore T' is surjective and $(T')^{-1} 0 \neq \emptyset$. Since $z^\infty \in cB \subset 2cU$, T' can be substituted for T when verifying that $0 \in Tz^\infty$.

Writing T instead of T' for simplicity, let $\bar{z} \in T^{-1} 0$ which is non-empty by above. Proposition 2.9 implies for any $k \geq 0$

$$|z^k - \bar{z}|^2 \geq |z^k - x^{k+1}|^2 + |x^{k+1} - \bar{z}|^2,$$

hence

$$\begin{aligned} &|z^k - x^{k+1}|^2 - |z^k - \bar{z}|^2 + |z^{k+1} - \bar{z}|^2 \\ &\quad \leq |z^{k+1} - \bar{z}|^2 - |x^{k+1} - \bar{z}|^2 \\ &\quad = (z^{k+1} - x^{k+1}), (z^{k+1} - \bar{z} + x^{k+1} - \bar{z}) \\ &\quad \leq |z^{k+1} - x^{k+1}| \, (|z^{k+1} - \bar{z}| + |x^{k+1} - \bar{z}|), \end{aligned}$$

rearranging

$$|z^k - x^{k+1}|^2 \leq |z^k - \bar{z}| - |z^{k+1} - \bar{z}|^2 + 2\varepsilon_k (c + |\bar{z}|); \qquad (2.2)$$

Equation 2.1, the boundedness of $\{z^k\}$and $\Sigma \varepsilon_k < \infty$, imply that

$$\limsup_{k \to \infty} |z^k - \bar{z}| = \liminf_{k \to \infty} |z^k - \bar{z}| = \mu < \infty, \qquad (2.3)$$

Taking lim sup in both sides of 2.2, it follows that $z^k - x^{k+1} \to 0$. But

$|z^{k+1}-z^k| \leq |z^{k+1}-x^{k+1}| + |x^{k+1}-z^k|$, since $|z^{k+1}-x^{k+1}| \leq \varepsilon_k \to 0$, it follows that $z^{k+1}-z^k \to 0$, i.e., $\{z^k\}$ is asymptotically regular. Let

$$w^k \in c_k T x^{k+1} \cap S(z^k - x^{k+1}).$$

For all k large enough $|z^k - x^{k+1}| \leq \delta$, thus assumption (c) implies $|w^k| \leq \sigma(|z^k-x^{k+1}|)$ for all k large enough. Taking lim sup in the preceding inequality and using the continuity of σ at zero, it follows that $w^k \to 0$, and $c_k^{-1} w^k \to 0$ as $\{c_k\}$ is bounded away from zero. Since z^∞ is a weak cluster point of $\{z^k\}$ and $z^k - x^{k+1} \to 0$, it is also a weak cluster point of $\{x^k\}$, and for some subsequence $\{x^k\}_K$, $x^k \overset{w}{\to} z^\infty$. Then

$$\limsup_{k \in K} x^{k+1}(c_k^{-1} w^k) \leq \limsup_{k \in K} |x^{k+1}||c_k^{-1} w^k| = 0 = (z^\infty, 0).$$

But $c_k^{-1} w^k \in T x^{k+1}$, thus $0 \in Tz^\infty$ (see Brézis 1973, p. 27, prop. 2.5).

Let $z_1^\infty, z_2^\infty \in T^{-1}0$ be two weak cluster points of $\{z^k\}$.

By (2.3)

$$\lim_{k\to\infty} |z^k - z_i^\infty| = \mu_i < \infty \qquad (i=1,2)$$

One has

$$|z^k - z_2^\infty|^2 = |z^k - z_1^\infty| + 2(z^k - z_1^\infty, z_1^\infty - z_2^\infty) + |z_1^\infty - z_2^\infty|^2$$

and

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

Thus the limit on the left side exists and it is the same for any subsequence of $\{z^k\}$, in particular if $\{z^k\}_{K_1} \overset{w}{\to} z_1^\infty$, said limit is zero.

Reversing the roles of z_1^∞ and z_2^∞, one obtains

$$\mu_2^2 - \mu_1^2 = |z_1^\infty - z_2^\infty|^2 = \mu_1^2 - \mu_2^2,$$

from which $z_1^\infty = z_2^\infty$. Suppose now that $z^k \overset{w}{\not\to} z^\infty$, then there are a weak neighborhood V of z^∞ and a subsequence $\{z^k\}_K \subset H \setminus V$. Being $\{z^k\}$ bounded, so is $\{z^k\}_K$ and z^∞ has to be its only weak cluster point in contradiction with $\{z^k\}_K \subset H\setminus V$. It follows that $z^k \overset{w}{\to} z^\infty$. QED

<u>Remark.</u> The proof given follows Rockafellar (1976a, p. 883, th. 1) with the necessary modifications as implied by proposition (2.9).

Assuming a growth condtion of the type (1.1) one is able to prove that the sequences $\{d(z^k,\overline{Z})\}$ corresponding to any sequence $\{z^k\}$ generated by the NPA coverge to zero.

<u>Theorem 2.13.</u> Under the hypothesis of the above theorem, let $\overline{Z} = T^{-1}0 \neq \emptyset$, and let $\tau:[0,\infty) \to [0,\infty)$ be such that $\tau(0) = 0$, τ is continuous at 0, and

$$\exists \eta > 0,\ \forall w \in \eta B,\ \forall z \in T^{-1}w,\ \ d(z,\overline{Z}) \leq \tau(|w|).$$

Then $d(z^k,\overline{Z}) \to 0$.

<u>Proof.</u> By the above theorem, $c_k^{-1}w^k \to 0$, thus for all k large enough $c_k^{-1}w^k \in \eta B$. As $w^k \in c_k Tx^{k+1} \cap S(z^k - x^{k+1})$, $x^{k+1} \in T^{-1}c_k^{-1}w^k$, and the properties of τ imply that $d(x^{k+1},\overline{Z}) \leq \tau(|c_k^{-1}w^k|)$. Taking lim sup in both sides of this inequality, it follows that $d(x^{k+1},\overline{Z}) \to 0$. Let $\overline{x^{k+1}}$ denote the orthogonal projection of x^{k+1} onto $\overline{Z}$ and analogously for z^{k+1}. By (A_r)

$$|z^{k+1}-\overline{z^{k+1}}| \le |z^{k+1}-\overline{x^{k+1}}| \le |z^{k+1}-x^{k+1}| + |x^{k+1}-\overline{x^{k+1}}|$$
$$\le \varepsilon_k + d(x^{k+1},\overline{Z}),$$

Taking lim sup on both sides the result follows. QED

3. Asymptotic convergence. The (Q-) order of convergence (Ortega and Rheinboldt 1970) of $\{d(z^k,\bar{Z})\}$, assuming that $d(z^k,\bar{Z}) \neq 0$ for all k, is the supremum of the numbers $\alpha \geqq 1$ such that

$$\limsup_{k \to \infty} \frac{d(z^{k+1},\bar{Z})}{d(z^k,\bar{Z})^\alpha} < \infty.$$

Theorem 3.1. Under the hypotheses of both theorems 2.12 and 2.13, let us assume the following forms of σ and τ

$$\forall x \in \mathbb{R}_+,\ \tau(x) = ax^t,\ \sigma(x) = bx^s,$$

where a,b,s,t are positive real numbers with $st \geqq 1$.

If st = 1, the NPA in both the exact and approximate versions, converges linearly at a rate bounded above by $a/(a^2 + (\underline{c}/b)^{2t})^{\frac{1}{2}}$, where $\underline{c} = \liminf_{k \to \infty} c_k$, and thus superlinearly if $\underline{c} = \infty$.

In any case, the (Q-) order of convergence of the approximate algorithm is at least $\min\{r,st\} \geqq 1$, and at least $st \geqq 1$ for the exact version. If $st > r = 1$, superlinear convergence is attained without needing $\underline{c} = \infty$.

Proof. Let $x^{k+1} \in P(c_kT,S)z^k$ and $w^k \in c_kTx^{k+1} \cap S(z^k - x^{k+1})$. Theorem 2.12 implies $z^k - x^{k+1} \to 0$, thus for all k large enough $z^k - x^{k+1} \in \delta B$ and by the assumption on S of the same theorem and the above form of σ

$$|w^k| \leqq b|z^k - x^{k+1}|^s.$$

In theorem 2.12 it was also proved that $c_k^{-1}w^k \to 0$, thus for all k large enough, $c_k^{-1}w^k \in \eta B$. As $x^{k+1} \in T^{-1}(c_k^{-1}w^k)$, by the assumption on T of theorem 2.13 and the above form of τ one can conclude

$$d(x^{k+1},\bar{Z}) \leqq a|c_k^{-1}w^k|^t.$$

From these inequalities, for all k large enough

$$d(x^{k+1},\bar{Z}) \leqq \frac{ab^t}{c_k^t}|z^k - x^{k+1}|^{st}.$$

For $\bar{z} = \overline{z^k}$ the orthogonal projection of z^k onto $\bar{Z}$, proposition 2.8(2) yields for all k

$$\begin{aligned} d(z^k,\bar{Z})^2 &\geqq |z^k - x^{k+1}|^2 + |x^{k+1} - \overline{z^k}|^2 \\ &\geqq |z^k - x^{k+1}|^2 + d(x^{k+1},\bar{Z})^2. \end{aligned} \tag{3.1}$$

Eliminating $|z^k - x^{k+1}|$ between the above two inequalities and rearranging one obtains the following estimate valid for all k large enough

$$d(x^{k+1},\bar{Z}) \leqq \frac{d(z^k,\bar{Z})^{st}}{[c_k' + d(x^{k+1},\bar{Z})^{2(st-1)/st}]^{st/2}}, \tag{3.2}$$

where $c_k' = (c_k/b)^{2/s}a^{-2/st}$.

From the triangle inequality, (A_r), and $r \geqq 1$, for all k

$$\begin{aligned} |z^{k+1} - \overline{x^{k+1}}| &\leqq |z^{k+1} - x^{k+1}| + |x^{k+1} - \overline{x^{k+1}}| \\ &\leqq \varepsilon_k|z^{k+1} - z^k|^r + d(x^{k+1},\bar{Z}) \\ &\leqq \varepsilon_k|z^{k+1} - z^k|^{r-1}(|z^{k+1} - x^{k+1}| + |z^k - x^{k+1}|) + d(x^{k+1},\bar{Z}). \end{aligned}$$

Also

$$|z^k - \overline{x^{k+1}}| \leqq |z^k - \overline{z^k}| + |\overline{z^k} - \overline{x^{k+1}}|$$

$$\leqq d(z^k,\bar{Z}) + |z^k - x^{k+1}| \leqq 2d(z^k,\bar{Z}),$$

Where we have used the fact that projection onto a nonempty closed convex subset of H is a nonexpansive map, and that $|z^k - x^{k+1}| \leqq d(z^k,\bar{Z})$ by (3.1). Eliminating $|z^k - \overline{x^{k+1}}|$ between the last two inequalities and rearranging

$$d(x^{k+1},\bar{Z}) \leqq |z^{k+1} - \overline{x^{k+1}}|(1 - \varepsilon_k|z^{k+1} - z^k|^{r-1})$$
$$- 2\varepsilon_k|z^{k+1} - z^k|^{r-1} d(z^k,\bar{Z}).$$

By theorem 2.12 , $|z^{k+1} - z^k| \to 0$, by (A_r), $r \geqq 1$ and $\varepsilon_k \to 0$, thus for all k large enough $\varepsilon_k|z^{k+1} - z^k|^{r-1} \leqq \varepsilon_k < 1$. Being $|z^{k+1} - \overline{x^{k+1}}| \geqq d(z^{k+1},\bar{Z})$, the following estimate is valid for all k large enough

$$d(x^{k+1},\bar{Z}) \geqq (1 - \varepsilon_k)d(z^{k+1},\bar{Z}) - 2\varepsilon_k d(z^k,\bar{Z})|z^{k+1} - z^k|^{r-1}.$$

Inequality (3.1) and (A_r) yield, by the same argument as above, for all k large enough

$$\begin{aligned} d(x^{k+1},\bar{Z}) &\geqq |z^k - x^{k+1}| \geqq |z^k - z^{k+1}| - |z^{k+1} - x^{k+1}| \\ &\geqq |z^k - z^{k+1}|(1 - \varepsilon_k|z^k - z^{k+1}|^{r-1}) \\ &\geqq (1 - \varepsilon_k)|z^k - z^{k+1}|. \end{aligned} \tag{3.3}$$

Eliminating $|z^k - z^{k+1}|$ between the last two inequalities, for all k large enough

$$d(x^{k+1},\bar{Z}) \geqq (1 - \varepsilon_k)d(z^{k+1},\bar{Z}) - \frac{2\varepsilon_k}{(1 - \varepsilon_k)^r} d(z^k,\bar{Z})^r.$$

Combining this inequality with (3.2) one can eliminate $d(x^{k+1},\bar{Z})$ obtaining for all k large enough

$$d(z^{k+1},\bar{Z}) \leqq \frac{d(z^k,\bar{Z})^{st}/(1-\varepsilon_k)}{[c_k' + d(x^{k+1},\bar{Z})^{2(st-1)/st}]^{st/2}} + \frac{2\varepsilon_k d(z^k,\bar{Z})^r}{(1-\varepsilon_k)^r}, \qquad (3.4)$$

where one should remember that $d(x^{k+1},\bar{Z}) \to 0$ by theorem 2.13. Suppose now that $d(z^k,\bar{Z}) \neq 0$ for all k. If st = 1, (3.4) implies that $d(z^k,\bar{Z}) \to 0$ linearly at a rate bounded above by

$$\frac{a}{(a^2 + (\underline{c}/b)^{2t})^{\frac{1}{2}}} = \frac{ab^t}{((ab^t)^2 + \underline{c}^{2t})^{\frac{1}{2}}},$$

where $\underline{c} = \liminf_{k\to\infty} c_k$, and thus superlinearly if $\underline{c} = \infty$. This result is valid for both the exact and approximate versions of the algorithm. In any case the (Q-) order of convergence is at least $\min\{r,st\} \geqq 1$ for the approximate version or $st \geqq 1$ for the exact one, as then $\varepsilon_k \equiv 0$. QED

Remark. Without requiring $\underline{c} = \infty$, any (Q-) order of convergence can be achieved if either $\min\{r,st\}$ for the approximate version, or st for the exact version is large enough.

The following result gives a sufficient condition for the convergence in a finite number of steps of the exact ($\varepsilon_k \equiv 0$) algorithm. This generalizes a result of Bertsekas (1975).

Theorem 3.2. Let $S \in \mathcal{S}$, and let δ, b, s be three positive real numbers such that $x \in \delta B$ and $y \in Sx$ implies that $|y| \leqq b|x|^s$. Let $T: H \to 2^H$

a maximal monotone map be such that $\bar{Z} = T^{-1}(0) \neq \emptyset$ and satisfy

$$\exists \eta > 0,\ \forall w \in \eta B,\ \forall z \in T^{-1}(w),\ d(z,\bar{Z}) = 0.$$

Then for the approximate version of the Nonlinear Proximal Point Algorithm operated under (A_r), $d(z^k,\bar{Z}) \to 0$, superlinearly if $r = 1$, and with (Q-) order of convergence at least r for all $r \geqq 1$. The exact version of the algorithm converges in finitely many steps which can be reduced to one if $\delta \geqq d(z^0,\bar{Z})$ and $c_0 \geqq bd(z^0,\bar{Z})^s/\eta$.

<u>Proof.</u> T satisfies

$$\forall w \in \eta B,\ \forall z \in T^{-1}(w),\ d(z,\bar{Z}) \leqq a|w|^t$$

for any $a,t \geqq 0$, and in particular T satisfies the hypothesis of the preceding theorem for all $a > 0$, $t \geqq 1/s$. Thus the (Q-) order of convergence is at least $\min\{r,st\}$ for all $t \geqq 1/s$, i.e., r. If $r = 1$, said theorem also implies that the convergence is superlinear.

If the algorithm is operated exactly, for all k, there is an $w^k \in c_k Tz^{k+1} \cap S(z^k - z^{k+1})$. By theorem 2.12 $c_k^{-1} w^k \to 0$ thus it lies in ηB for all k large enough, and as $z^{k+1} \in T^{-1}(c_k^{-1} w^k)$, it follows that $z^{k+1} \in \bar{Z}$. Alternatively, the preceding theorem guarantees an order of convergence of at least st for all $t \geqq 1/s$ by the above argument. Thus the (Q-) order is ∞, i.e., $d(z^k,\bar{Z}) = 0$ for all but finitely many k's.

Let $w^0 \in c_0 Tz^1 \cap S(z^0 - z^1)$. If $c_0^{-1} w^0 \in \eta B$, then $z^1 \in \bar{Z}$, thus c_0 should be larger than $|w^0|/\eta$. By proposition 2.8(2), $|z^0 - \bar{z}| \geqq |z^0 - z^1|$ and choosing $\bar{z} = \overline{z^0}$, $|z^0 - z^1| \leqq d(z^0,\bar{Z})$, thus $z^0 - z^1 \in \delta B$ if $\delta \geqq d(z^0,\bar{Z})$. As $w^0 \in S(z^0 - z^1)$, it follows that $|w^0| \leqq b|z^0 - z^1|^s \leqq bd(z^0,\bar{Z})^s/\eta$. QED

The sublinear convergence of the algorithm is explored in the next result.

Theorem 3.3. Let $S \in \mathcal{S}$ be such that for positive numbers b, s, η, and a function $\sigma: \mathbb{R}_+ \to \mathbb{R}_+$ continuous at zero and with $\sigma(0) = 0$, it satisfies

$$(x \in \eta B,\ y \in Sx) \implies b|x|^s \leqq |y| \leqq \sigma(|x|).$$

Let $T: H \to 2^H$ be maximal monotone with $\bar{Z} = T^{-1}(0) \neq \emptyset$, and for positive numbers a, t, δ, and a function $\tau: \mathbb{R}_+ \to \mathbb{R}_+$, continuous at zero and with $\tau(0) = 0$, it satisfies

$$(w \in \delta B,\ z \in T^{-1}(w)) \implies a|w|^t \leqq d(z, \bar{Z}) \leqq \tau(|w|).$$

Let the NPA be operated under criterion (A_r) with $r \geqq 1$ and a bounded sequence $\{c_k\}$. If $st > 1$, then

$$\liminf_{k \to \infty} \frac{d(z^{k+1}, \bar{Z})}{d(z^k, \bar{Z})} = 1$$

and the convergence cannot be faster than sublinear. If $st = 1$, then

$$\liminf_{k \to \infty} \frac{d(z^{k+1}, \bar{Z})}{d(z^k, \bar{Z})} \geqq \frac{ab^t}{ab^t + \bar{c}}$$

where $\bar{c} = \limsup_{k \to \infty} c_k$.

Proof. The global convergence theorem applies and if $w^k \in c_k T x^{k+1} \cap S(z^k - x^{k+1})$, $c_k^{-1} w^k \to 0$ and eventually lies in δB. As $x^{k+1} \in T^{-1}(c_k^{-1} w^k)$, it follows that for all k large enough

$$c_k d(z^{k+1},\bar{Z}) \geqq a|w^k|^t.$$

Also, by global convergence, $z^k - x^{k+1} \to 0$ thus eventually it lies in ηB and as $w^k \in S(z^k - x^{k+1})$, for all k large enough

$$b|z^k - x^{k+1}|^s \leqq |w^k|.$$

Combining both inequalities

$$c_k d(x^{k+1},\bar{Z}) \geqq ab^t|z^k - x^{k+1}|^{st} .$$

Projection onto $\bar{Z}$ is a nonexpansive map , thus

$$\begin{aligned} d(x^{k+1},\bar{Z}) &\leqq |x^{k+1} - z^{k+1}| + |z^{k+1} - \overline{z^{k+1}}| + |\overline{z^{k+1}} - \overline{x^{k+1}}| \\ &\leqq 2|x^{k+1} - z^{k+1}| + d(z^{k+1},\bar{Z}), \end{aligned}$$

which substituted above yields

$$c_k d(z^{k+1},\bar{Z}) + 2c_k|x^{k+1} - z^{k+1}| \geqq ab^t|z^k - x^{k+1}|^{st}.$$

Criterion (A_r) implies the following inequalities

$$\begin{aligned} |x^{k+1} - z^{k+1}| &\leqq \varepsilon_k|z^k - z^{k+1}|^r, \\ |z^k - x^{k+1}| &\geqq |z^k - z^{k+1}| - |z^{k+1} - x^{k+1}| \\ &\geqq |z^k - z^{k+1}| - \varepsilon_k|z^k - z^{k+1}|^r \\ &= |z^k - z^{k+1}|(1 - \varepsilon_k|z^k - z^{k+1}|^{r-1}) \\ &\geqq |z^k - z^{k+1}|(1 - \varepsilon_k). \end{aligned}$$

The last inequality is only valid for k large enough. Being $z^k - z^{k+1} \to 0$ by the global convergence theorem, eventually $|z^k - z^{k+1}| < 1$ and $|z^k - z^{k+1}|^{r-1} \leqq 1$. Using this two inequalities to eliminate $|x^{k+1} - z^{k+1}|$ in the preceding one

$$\begin{aligned} c_k d(z^{k+1},\bar{Z}) &\geqq ab^t|z^k - z^{k+1}|^{st}(1 - \varepsilon_k)^{st} - 2c_k\varepsilon_k|z^k - z^{k+1}|^r \\ &= |z^k - z^{k+1}|^{st}[ab^t(1-\varepsilon_k)^{st} - 2c_k\varepsilon_k|z^k - z^{k+1}|^{r-st}]. \end{aligned}$$

Being $\{c_k\}$ bounded $\varepsilon_k \to 0$, $z^k - z^{k+1} \to 0$, $r - st \geqq 0$, the second member inside the brackets tends to zero so that for all k large enough the bracket is nonnegative. Using this fact and

$$\begin{aligned} |z^k - z^{k+1}| &\geqq |z^k - \overline{z^{k+1}}| - |z^{k+1} - \overline{z^{k+1}}| \\ &\geqq d(z^k,\bar{Z}) - d(z^{k+1},\bar{Z}), \end{aligned}$$

the above expression can be transformed into

$$\begin{aligned} &[c_k d(z^{k+1},\bar{Z})]^{1/st} + [ab^t(1-\varepsilon_k)^{st} - 2c_k\varepsilon_k|z^k - z^{k+1}|^{r-st}]^{1/st} d(z^{k+1},\bar{Z}) \\ &\geqq [ab^t(1 - \varepsilon_k)^{st} - 2c_k\varepsilon_k|z^k - z^{k+1}|^{r-st}]^{1/st} d(z^k,\bar{Z}). \end{aligned}$$

If st = 1, without assuming the existence of τ, it is easy to obtain

$$\begin{aligned} &\liminf_{k\to\infty} \frac{d(z^{k+1},\bar{Z})}{d(z^k,\bar{Z})} \geqq \\ &\liminf_{k\to\infty} \frac{[ab^t(1-\varepsilon_k) - 2c_k\varepsilon_k|z^k - z^{k+1}|^{r-1}]}{[ab^t(1-\varepsilon_k) - 2c_k\varepsilon_k|z^k - z^{k+1}|^{r-1}] + c_k} = \frac{ab^t}{ab^t + \bar{c}}. \end{aligned}$$

If st < 1, the existence of τ implies that $d(z^{k+1},\bar{Z}) \to 0$, and so does

$d(z^{k+1},\bar{Z})^{(1-st)/st}$ It follows then that

$$\liminf_{k \to \infty} \frac{d(z^{k+1},\bar{Z})}{d(z^{k},\bar{Z})} = 1.$$

QED

4. Implementation. Criterion (A_r) requires that the current estimate of a root of T, z^k and the new one z^{k+1} satisfy

$$|z^{k+1} - x^{k+1}| \leqq \varepsilon_k \min\{1, |z^k - z^{k+1}|^r\},$$

where $x^{k+1} \in P(c_k T, S) z^k$. Thus x^{k+1} is a new estimate, computed exactly from the current one z^k. By the definition of $P(c_k T, S) z^k$, it is clear that x^{k+1} satisfies

$$0 \in c_k T x^{k+1} - S(z^k - x^{k+1})$$

or $0 \in R_k x^{k+1}$, where $R_k = c_k T - S(z^k - \cdot)$ is maximal monotone. The problem with (A_r) is that x^{k+1} is not known and as it stands (A_r) is not implementable. Rockafellar (1976a, p 882, prop 3) showed a sufficient condition for (A_r) which does not require the knowledge of x^{k+1}. In the present case, the fact that S is no longer the identity map complicates matters. The approach followed was suggested by Kort and Bertsekas (1976).

Proposition 4.1. Let $T : H \to 2^H$ be a maximal monotone map which is strongly monotone with modulus $\alpha > 0$. Then criterion (A_r) is implied by

$$(A_r') \quad d(0, R_k z^{k+1}) \leqq \alpha c_k \varepsilon_k \min\{1, |z^k - z^{k+1}|^r\}.$$

Proof. R_k is strongly monotone with modulus $\alpha c_k > 0$. Let $w \in R_k z^{k+1}$, as $0 \in R_k x^{k+1}$

$$(z^{k+1} - x^{k+1}, w) \geqq \alpha c_k |z^{k+1} - z^k|^2.$$

Using the Cauchy-Buniakovskii inequality and selecting as w^k the least norm element of $R_k z^{k+1}$

$$d(0,R_k z^{k+1}) \geqq \alpha c_k |z^{k+1} - x^{k+1}|. \qquad \text{QED}$$

The estimate obtained by Rockafellar (ibid.) is, in our notation

$$d(0,R_k z^{k+1}) \geqq |z^{k+1} - x^{k+1}|.$$

The fact that $P(c_k T,S)$ is nonexpansive for $S = I$ was used there. This is no longer the case if $S \neq I$ as we have seen in the example after definition 2.11.

Note that if T is strongly monotone, then $\bar{\bar{Z}} = \{z^\infty\}$, and $w \in Tz$ implies that $d(z,z^\infty) \leqq \alpha^{-1}|w|$. The strong convergence of the algorithm in its two versions, to the unique solution is guaranteeed.

5. Application to convex and saddle functions. When T is the subdifferential map of a proper closed convex function, or the "twisted" subdifferential of a proper closed saddle function, it is possible to estimate the speed of convergence of the NPA towards the minimum value, or the saddle value, respectively. This speed is given in terms of the sequence $\{d(z^k,\bar{Z})\}$ whose decrease towards zero has been quantified in section 3.

Let H_1, H_2 be real Hilbert spaces. Let $\langle \cdot,\cdot \rangle$ and $|\cdot|$ denote the inner product and corresponding induced norm of both H_1 and H_2. In $H_1 \times H_2$ we can define an inner product as follows. For all (x,y), $(x',y') \in H_1 \times H_2$

$$\langle (x,y),(x',y') \rangle = \langle x,x' \rangle + \langle y,y' \rangle$$

The norm induced on $H_1 \times H_2$ by this inner product is such that for all $(x,y) \in H_1 \times H_2$

$$|(x,y)|^2 = \langle (x,y),(x,y) \rangle = \langle x,x \rangle + \langle y,y \rangle = |x|^2 + |y|^2.$$

Let $L: H_1 \times H_2 \to \bar{\mathbb{R}}$ be a closed proper convex-concave saddle function. The subdifferential map of L, ∂L, is defined as follows. For all $(x,y) \in H_1 \times H_2$, $\partial L(x,y)$ is the set of vectors $(u,v) \in H_1 \times H_2$ such that for all $(x',y') \in H_1 \times H_2$,

$$L(x',y') - \langle x'-x,u \rangle \geqq L(x,y) \geqq L(x,y') - \langle y'-y,v \rangle.$$

The "twisted" subdifferential of L, T, is such that for all $(x,y) \in H_1 \times H_2$,

$$T(x,y) = \{(u,-v) \in H_1 \times H_2 \mid (u,v) \in \partial L(x,y)\},$$

Under the above assumptions on L, T is maximal monotone (Rockafellar 1970b). Clearly $(0,0) \in T(\bar{x},\bar{y})$ iff $(\bar{x},\bar{y})$ is a saddle point of L.

Theorem 5.1. Let $\{(x^k,y^k)\}$ be any sequence generated by the NPA when applied, under the hypothesis of theorem 3.1, to T as above. Then $\{(x^k,y^k)\}$ converges weakly to (x^∞,y^∞), an element of $\bar{Z} = T^{-1}(0,0)$, and $d[(x^k,y^k),\bar{Z}] \to 0$. Let us also assume that $(x^k,y^k) \notin \bar{Z}$ for all k.

(1) If the algorithm is implemented exactly, then for all k large enough

$$\frac{|L(x^\infty,y^\infty) - L(x^{k+1},y^{k+1})|}{|(x^\infty,y^\infty) - (x^{k+1},y^{k+1})|} \leqq c_k^{-1} b\, d[(x^k,y^k),\bar{Z}]^s.$$

(2) Let S = I, and let the algorithm be implemented approximately, requiring at each step that the following condition (sufficient for (A_r), see Rockafellar 1976a, p. 889, th. 4) be satisfied for all k

$$d[(x^k - x^{k+1}, y^k - y^{k+1}), c_k T(x^{k+1},y^{k+1})] \leqq$$
$$\leqq \varepsilon_k \min\{1, |(x^k,y^k) - (x^{k+1},y^{k+1})|^r\}.$$

Then for all k large enough

$$\frac{|L(x^\infty,y^\infty) - L(x^{k+1},y^{k+1})|}{|(x^\infty,y^\infty) - (x^{k+1},y^{k+1})|} \leqq$$

$$\leqq \frac{\varepsilon_k}{c_k(1-\varepsilon_k)^r}\, d[(x^k,y^k),\bar{Z}]^r + \frac{1}{c_k(1-\varepsilon_k)}\, d[(x^k,y^k),\bar{Z}].$$

(3) If T is strongly monotone with modulus $\alpha > 0$, then τ with $t = 1$, $a = \alpha^{-1}$ is globally valid, and $\bar{Z} = \{(x^{\infty}, y^{\infty})\}$. Let the NPA be operated approximately using criterion (A_r') (see section 4), i.e., for all k

$$d[S(x^k - x^{k+1}, y^k - y^{k+1}), c_k T(x^{k+1}, y^{k+1})] \leqq$$
$$\leqq \alpha c_k \varepsilon_k \min\{1, |(x^k, y^k) - (x^{k+1}, y^{k+1})|^r\}.$$

Then $(x^k, y^k) \to (x^{\infty}, y^{\infty})$ strongly, and for all k large enough

$$\frac{|L(x^{\infty}, y^{\infty}) - L(x^{k+1}, y^{k+1})|}{|(x^{\infty}, y^{\infty}) - (x^{k+1}, y^{k+1})|} \leqq$$

$$\leqq \frac{\alpha \varepsilon_k}{(1 - \varepsilon_k)^r} |(x^k, y^k) - (x^{\infty}, y^{\infty})|^r + \frac{b}{c_k (1 - \varepsilon_k)^s} |(x^k, y^k) - (x^{\infty}, y^{\infty})|^s.$$

<u>Proof.</u> The hypothesis of theorem 3.1 imples that of theorem 2.12 from which $d[(x^k, y^k), \bar{Z}] \to 0$ follows.

(1) Let $(w_x, -w_y) \in c_k T(x^{k+1}, y^{k+1}) - S(x^k - x^{k+1}, y^k - y^{k+1})$, and let $(v_x, -v_y) \in S(x^k - x^{k+1}, y^k - y^{k+1})$.

Then

$$(v_x + w_x, -v_y - w_y) \in c_k T(x^{k+1}, y^{k+1})$$

and

$$c_k^{-1}(v_x + w_x) \in \partial_x L(x^{k+1}, y^{k+1}),$$
$$c_k^{-1}(v_y + w_y) \in \partial_y L(x^{k+1}, y^{k+1}).$$

The subgradient inequalities yield

$$L(x^{\infty},y^{k+1}) \geqq L(x^{k+1},y^{k+1}) + c_k^{-1}\langle v_x + w_x, x^{\infty} - x^{k+1}\rangle,$$
$$L(x^{k+1},y^{\infty}) \leqq L(x^{k+1},y^{k+1}) + c_k^{-1}\langle v_y + w_y, y^{\infty} - y^{k+1}\rangle.$$

From the saddle point inequality for (x^{∞},y^{∞})

$$L(x^{\infty},y^{k+1}) \leqq L(x^{\infty},y^{\infty}) \leqq L(x^{k+1},y^{\infty}),$$

one obtains

$$c_k^{-1}\langle v_x + w_x, x^{\infty} - x^{k+1}\rangle \leqq L(x^{\infty},y^{\infty}) - L(x^{k+1},y^{k+1}) \leqq$$
$$= c_k^{-1}\langle v_y + w_y, y^{\infty} - y^{k+1}\rangle.$$

Using the Cauchy-Buniakovskii and triangle inequalities, and dividing throughout by $|(x^{\infty},y^{\infty}) - (x^{k+1},y^{k+1})|$ the following estimate is obtained

$$-c_k^{-1}(|v_x| + |w_x|) \leqq \frac{|L(x^{\infty},y^{\infty}) - L(x^{k+1},y^{k+1})|}{|(x^{\infty},y^{\infty}) - (x^{k+1},y^{k+1})|} \leqq$$
$$\leqq c_k^{-1}(|v_y| + |w_y|).$$

By theorem 2.12 $(x^k,y^k) - (x^{k+1},y^{k+1}) \to 0$, thus the growth condition on S eventually holds, and using the form of σ chosen in theorem 3.1,

$$|(v_x,-v_y)| \leqq b|(x^k,y^k) - (x^{k+1},y^{k+1})|^s.$$

But for all k large enough, by equation (3.3),

$$|(x^k,y^k) - (x^{k+1},y^{k+1})| \leqq d[(x^k,y^k),\bar{Z}]/(1-\varepsilon_k).$$

Also $|v_x|,|v_y| \leqq |(v_x,-v_y)|$, thus for all k large enough

$$-c_k^{-1}|w_x| - \frac{b}{c_k(1-\varepsilon_k)^s} d[(x^k,y^k),\bar{Z}]^s \leqq$$

$$\leqq \frac{|L(x^\infty,y^\infty) - L(x^{k+1},y^{k+1})|}{|(x^\infty,y^\infty) - (x^{k+1},y^{k+1})|} \leqq \tag{5.1}$$

$$\leqq c_k^{-1}|w_y| + \frac{b}{c_k(1-\varepsilon_k)^s} d[(x^k,y^k),\bar{Z}]^s.$$

When the algorithm is implemented exactly, i.e., $\varepsilon_k \equiv 0$, one can choose $w_x = 0$, $w_y = 0$, and the above inequality yields the desired estimate.

(2) Let $S = I$, and let the algorithm be implemented with the criterion given in the statement. Choosing as $(w_x,-w_y)$ the minimum norm element of

$$c_k T(x^{k+1},y^{k+1}) - S(x^k - x^{k+1}, y^k - y^{k+1}).$$

By the same argument as in proving (1)

$$|w_x|,|w_y| \leqq |(w_x,-w_y)| \leqq$$

$$\leqq \varepsilon_k|(x^k,y^k) - (x^{k+1},y^{k+1})|^r \leqq$$

$$\leqq \frac{\varepsilon_k}{(1-\varepsilon_k)^r} d[(x^k,y^k),\bar{Z}]^r,$$

which yields via (5.1) the desired estimate.

(3) The strong monotonicity of T implies that $\bar{Z}$ is a singleton, and $(x^k,y^k) \to (x^\infty,y^\infty)$ strongly follows from $d[(x^k,y^k),\bar{Z}] \to 0$. Choosing as $(w_x,-w_y)$ the minimum norm element of

$$c_k^{-1}T(x^{k+1},y^{k+1}) - S(x^k - x^{k+1}, y^k - y^{k+1}),$$

criterion (A_r') yields, as above,

$$|w_x|,|w_y| \leqq |(w_x,-w_y)| \leqq$$

$$\leqq \alpha c_k \varepsilon_k |(x^k,y^k) - (x^{k+1},y^{k+1})|^r$$

$$\leqq \frac{\alpha c_k \varepsilon_k}{(1-\varepsilon_k)^r} |(x^k,y^k) - (x^\infty,y^\infty)|^r,$$

combining this inequality with (5.1) one concludes the proof of (3)

QED.

Let H be a real Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and induced norm $|\cdot|$. Let $f:H \to (-\infty,\infty]$ be a closed proper convex function and $T = \partial f$. Then $0 \in T\bar{z}$ iff f achieves its global minimum at $\bar{z}$.

Theorem 5.2. Let $\{z^k\}$ be any sequence generated by the NPA when applied, under the hypothesis of theorem 3.1, to T as above. Then $\{z^k\}$ converges weakly to z^∞ an element of $\bar{Z} = T^{-1}(0)$, and $d(z^k,\bar{Z}) \to 0$. Let us also assume that $z^k \notin \bar{Z}$ for all k.

(1) If the algorithm is implemented exactly, then for all k large enough

$$\frac{f(z^{k+1}) - f(z^\infty)}{|z^{k+1} - z^\infty|} \leqq c_k^{-1} b d(z^k,\bar{Z})^s.$$

(2) Let S = I, and let the algorithm be implemented approximately, requiring at each step that the following condition (see theorem 5.1(2))

be satisfied for all k

$$d(z^k - z^{k+1}, c_k T z^{k+1}) \leqq \varepsilon_k \min\{1, |z^k - z^{k+1}|^r\}.$$

Then for all k large enough

$$\frac{f(z^{k+1}) - f(z^\infty)}{|z^{k+1} - z^\infty|} \leqq \frac{\varepsilon_k}{c_k(1-\varepsilon_k)^r} d(z^k, \bar{Z})^r + \frac{1}{c_k(1-\varepsilon_k)} d(z^k, \bar{Z}).$$

(3) If T is strongly monotone with modulus α, then τ with $t = 1$, $a = \alpha^{-1}$ is globally valid and $\bar{Z} = \{z^\infty\}$. Let the NPA be implemented approximately using criterion (A_r') (see section 4). Then $z^k \to z^\infty$ strongly, and for all k large enough

$$\frac{f(z^{k+1}) - f(z^\infty)}{|z^{k+1} - z^\infty|} \leqq \frac{\alpha\varepsilon_k}{(1-\varepsilon_k)^r} |z^k - z^\infty|^r + \frac{b}{c_k(1-\varepsilon_k)^s} |z^k - z^\infty|^s.$$

<u>Proof.</u> f can be considered as a saddle function on $H \times H_1$ where $H_1 = \{0\}$. QED.

REFERENCES

[1] D.P. Bertsekas (1975), Necessary and sufficient conditions for a penalty method to be exact. Math. Programming, 9, 87-99.

[2] H. Brézis (1973) Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert. North-Holland, Amsterdam.

[3] K. Deimling (1985), Nonlinear functional analysis. Springer-Verlag, Berlin.

[4] M.C. Joshi, and R.K. Bose (1985), Some topics in nonlinear functional analysis. Wiley, New Delhi.

[5] B.W. Kort, and D.P. Bertsekas (1972), A new penalty function method for constrained optimization. Proceedings IEEE Conference on Decision and Control, New Orleans, LA, 162-166.

[6] ----- (1973), Multiplier methods for convex programming. Proceedings IEEE Conference on Decision and Control, San Diego, CA, 428-432.

[7] ----- (1976), Combined primal-dual and penalty methods for convex programming. SIAM J. Control and Optimization 14: 2, 268-294.

[8] J. Luque (1984a), Nonlinear Proximal Point Algorithms. Dissertation, Operations Research Center, Massachusetts Institute of Technology, May 1984.

[9] ----- (1984b), Asymptotic convergence analysis of the proximal point algorithm. SIAM J. Control and Optimization 22: 2, 277-293.

[10] ----- (1986a), Convolutions of maximal monotone mappings, abstract. Abstracts of papers presented to the Amer. Math. Soc. 7: 1, 108, no. *825-49-541.

[11] ----- (1986b), Convolution of maximal monotone mappings. Submitted.

[12] G.J. Minty (1962), Monotone (nonlinear) operators in Hilbert space. Duke Math. J. 29, 341-362.

[13] ----- (1964), On the solvability of nonlinear functional equations of "monotonic" type. Pacific J. Math. 14, 249-255.

[14] J.-J. Moreau (1965), Proximité et dualité dans un espace hilbertien. Bull. Soc. Math. France 93, 273-299.

[15] J.M. Ortega, and W.C. Rheinboldt (1970), Iterative solution of non-linear equations in several variables. Academic Press, New York.

[16] D. Pascali, and S. Sburlan (1978), Nonlinear mappings of monotone type. Sijthoff and Noordhoff, Alphen aan den Rijn, Holland.

[17] R.T. Rockafellar (1970a), On the maximality of sums of nonlinear monotone operators. Trans. Amer. Math. Soc. 149, 75-88.

[18] ----- (1970b), Monotone operators associated with saddle-functions and minimax problems. Proceedings of Symposia in Pure Mathematics, vol 18, part 1, 241-250, F.E. Browder, ed., AMS, Providence.

[19] ----- (1973), The multiplier method of Hestenes and Powell applied toconvex programming. J.O.T.A. 12: 6, 555-562.

[20] ----- (1976a), Monotone operators and the proximal point algorithm. SIAM J. Control and Optimization 14: 5, 877-898.

[21] -----(1976b), Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Operations Research 1: 2, 97-116.