September, 1981 LIDS-P-1142

ASYMPTOTIC CONVERGENCE ANALYSIS OF THE PROXIMAL POINT ALGORITHM*

by

Fernando Javier Luque

Laboratory for Information and Decision Systems
and
Operations Research Center
Massachusetts Insitute of Technology
Cambridge, Massachusetts 02139

*Work supported by Grant NSF 79-20834 and ITP Foundation of Madrid.

To appear in SIAM Journal on Control and Optimization

ABSTRACT

The asymptotic convergence of the proximal point algorithm (PPA), for the solution of equations of type $0 \in Tz$, where T is a multivalued maximal monotone operator in a real Hilbert space is analyzed. When $0 \in Tz$ has a nonempty solution set $\bar{Z}$, convergence rates are shown to depend on how rapidly T^{-1} grows away from $\bar{Z}$ in a neighbourhood of 0. When this growth is bounded by a power function with exponent s, then for a sequence $\{z^k\}$ generated by the PPA, $\{|z^k - \bar{Z}|\}$ converges to zero, like $O(k^{-s/2})$, linearly, superlinearly, or in a finite number of steps according to whether $s \in (0,1)$, $s = 1$, $s \in (1, +\infty)$, or $s = +\infty$.

1. Introduction

Let H be a real Hilbert space with inner product $\langle\cdot,\cdot\rangle$, and induced norm $|\cdot|$, where for all $z \in H$, $|z| = \langle z,z\rangle^{\frac{1}{2}}$. Let us consider a multi-valued mapping $T: H \to 2^H$. Its domain $D(T)$ is defined by

$$D(T) = \{z \in H : Tz \neq \emptyset\} ,$$

its range by

$$R(T) = \cup\{Tz : z \in H\}$$

and its graph by

$$G(T) = \{(z,w) \in H \times H: w \in Tz\} .$$

The inverse point to set mapping T^{-1} is defined by $T^{-1}w = \{z \in H : w \in Tz\}$ if $w \in R(T)$ and $T^{-1}w = \emptyset$ otherwise. It is an elementary fact that $D(T^{-1}) = R(T)$, $R(T^{-1}) = D(T)$, and $G(T^{-1}) = \{(w,z) \in H \times H: (z,w) \in G(T)\}$

Such a mapping T is said to be a monotone operator, if and only if,

$$\forall z, z' \in D(T), \forall w \in Tz, \forall w' \in Tz' \quad \langle z - z', w-w'\rangle \geq 0.$$

If in addition, its graph, is not properly contained in the graph of any other monotone operator, then T is maximal monotone. For a detailed treatment of the theory and applications of such mappings, the reader may consult the works by Brézis (1973), Browder (1976), Pascali and Sburlan (1978), and the references cited therein.

A fundamental problem is to find a vector $z \in H$ such that $0 \in Tz$. Some of the most important problems in the area of convex programming and related fields can be cast into this general framework.

If f is a closed proper convex function, then $T = \partial f$ is maximal monotone (Moreau 1965). Thus solving the equation $0 \in Tz$ is equivalent to minimizing the convex function f, since f attains its minimum at $\bar{z}$ if and only if $0 \in \partial f(\bar{z})$.

Let H_1, H_2 be real Hilbert spaces and let K be a closed proper saddle function on $H = H_1 \times H_2$, which is convex in the first argument and concave in the second, and let the subdifferential of K at $(x,y) \in H_1 \times H_2$, $\partial K(x,y)$, be defined as the set of vectors $(u,v) \in H_1 \times H_2$ satisfying

$$\forall (x',y') \in H_1 \times H_2 \quad K(x',y) - \langle x'-x, u\rangle \geq K(x,y) \geq K(x,y') - \langle y'-y, v\rangle ,$$

then the multifunction

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

is a maximal monotone operator (Rockafellar 1970a). The solutions $(\bar{x},\bar{y})$ of the equation $(0,0) \in T(x,y)$ are the saddle points of K.

A variational inequality problem is to find a vector $\bar{z} \in C$ satisfying

$$\exists \bar{w} \in A\bar{z} : \quad \forall v \in C \quad \langle a - \bar{w}, \bar{z} - v\rangle \geq 0 ,$$

where $C \subseteq H$ is a nonempty closed convex set, $A : H \to 2^H$ is a multivalued monotone mapping with $D(A) = C$, and a is a given vector in H. Equivalently, it can be expressed by: find a vector $\bar{z} \in C$ such that

$$a \in A\bar{z} + N_C(\bar{z}) ,$$

where $N_C(z)$ is the normal cone to C at z. Its expression valid for all $u \in H$ is (Rockafellar 1970b, p. 15)

$$N_C(u) = \{x \in H : \langle x, u-v \rangle \geq 0 \quad \text{for all } v \in C\} \; .$$

When C is a cone and C° denotes its polar, the variational inequality problem above is reduced to the complementarity problem of finding a vector $\bar{z} \in C$ such that

$$\exists \bar{w} \in A\bar{z} : a - \bar{w} \in C^\circ, \quad \langle a - \bar{w}, \bar{z} \rangle = 0 \; .$$

These last two problems can be reduced to solving $0 \in Tz$ for the operator T defined by (Rockafellar 1976a)

$$Tz = \begin{cases} -a + Az + N_C(z) & z \in C \\ \emptyset & z \notin C \end{cases}$$

Conditions for the maximal monotonicity of such operators T, were given by Rockafellar (1970c, Th. 5). Further results are contained in papers by Rockafellar (1978, 1980) and McLinden (1980).

We will now introduce the Proximal Point Algorithm (PPA). Most of the notation has been borrowed from Rockafellar (1976a).

Minty (1962) proved that if T is a maximal monotone operator and c is a positive constant, for any $u \in H$ there is a unique z, such that $u \in (I+cT)z$. The operator $P = (I+cT)^{-1}$ (the proximal mapping associated with cT in the terminology of Moreau (1965)), is thus single-valued from all of H into H. The monotonicity of T is a necessary and sufficient condition for the nonexpansiveness of P wherever P is defined (Brézis 1973, p. 21, Prop. 2.1). Thus

$$\forall z, z' \in H \qquad |Pz - Pz'| \leq |z - z'|$$

The PPA generates for any starting point $z^0 \in H$, 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 for the approximate computation of z^{k+1} used in this analysis will be

$$(A_r) \qquad |z^{k+1} - P_k z^k| \le \varepsilon_k \min\{1, |z^{k+1} - z^k|^r\}$$

where r and ε_k satisfy

$$r \ge 0, \quad \forall k \quad \varepsilon_k \ge 0, \quad \sum_{k=0}^{\infty} \varepsilon_k < +\infty \quad .$$

It has been shown by Rockafellar (1976a, Th. 1) that when $0 \in Tz$ has at least one solution, the condition $|z^{k+1} - P_k z^k| \le \varepsilon_k$ is a sufficient condition for $|z^{k+1} - z^k| \to 0$. Therefore when the PPA is implemented with criterion (A_r) $r \ge 1$, there exists some $k' \in \mathbb{Z}_+$ such that for all $k \ge k'$, $|z^{k+1} - z^k|^r \le |z^{k+1} - z^k| < 1$, and thus the larger r is, the more accurate the computation of z^{k+1} will be. In previous papers dealing with the PPA (Rockafellar 1976a, 1978, 1980), the value of r was always taken equal to 1, but as will be shown below one takes r strictly greater than one in order to achieve superlinear convergence of order greater than one.

As shown by Rockafellar (1976a, Prop. 3), the estimate $|z^{k+1} - P_k z^k| \le c_k \text{ dist}(0, S_k z^{k+1})$ holds for all k, where $S_k z = Tz + c_k^{-1} (z-z^k)$. Therefore criterion (A_r) is implied by

$$(A_r') \qquad \text{dist}(0, S_k z^{k+1}) \le \frac{\varepsilon_k}{c_k} \min\{1, |z^{k+1} - z^k|^r\} \quad .$$

The set of solutions (possibly empty) of the equation $0 \in Tz$, will be denoted by $\bar{Z} = \{z \in H : 0 \in Tz\}$. When T is maximal monotone, for every

$u \in H$, $T^{-1}u = \{z \in H : u \in Tz\}$ is a, possibly empty, closed, convex set (Minty 1964, Th. 1). Therefore $\bar{Z} = T^{-1}0$ is closed and convex. If $\bar{Z}$ is nonempty, the vector in $\bar{Z}$ closest to z will be denoted by $\bar{z}$. We will use the notation

$$|z - \bar{Z}| = \min\{|z - z'| : z' \in \bar{Z}\} = |z - \bar{z}|.$$

Our analysis will focus on the convergence properties of the sequence $\{|z^k - \bar{Z}|\}$ corresponding to any sequence $\{z^k\}$ generated by the PPA.

In addition to the proximal mappings $P_k = (I+c_kT)^{-1}$ where $c_k > 0$ and T is a maximal monotone operator, use will also be made of the mappings Q_k defined by

$$Q_k = I - P_k \quad .$$

Clearly $0 \in Tz \iff P_kz = z \iff Q_kz = 0$.

Rockafellar (1976a, Prop. 1) proved the following facts

$$\forall k \geq 0, \quad \forall z \in H \qquad c_k^{-1}Q_kz^k \in TP_kz^k \tag{1.1}$$

$$\forall k \geq 0, \quad \forall z,z' \in H \quad |P_kz - P_kz'|^2 + |Q_kz - Q_kz'|^2 \leq |z-z'|^2. \tag{1.2}$$

In the same paper, the following theorem was also proved.

Theorem 1.1. (Rockafellar 1976a, Th. 1). Let $Z = T^{-1}0 \neq \emptyset$, and let $\{z^k\}$ be any sequence generated by the PPA with stopping criterion (A_r), $r \geq 0$, and a sequence of positive numbers $\{c_k\}$, such that $\liminf_{k\to\infty} c_k > 0$. Then $\{z^k\}$ is bounded, and converges in the weak topology to a unique point $z^\infty \in \bar{Z}$. Also

$$0 = \lim_{k\to\infty} |c_k^{-1} Q_k z^k| = \lim_{k\to\infty} |Q_k z^k| = \lim_{k\to\infty} |z^{k+1} - z^k| . \tag{1.3}$$

This paper addresses the issue of the speed of convergence of the PPA, in both its exact and approximate versions. Under various hypotheses, we show linear convergence, superlinear convergence, convergence in a finite number of steps, and convergence in one step. A condition that implies sub-linear convergence is given, and an estimate of the convergence rate in this case is also provided. The results previously available on the speed of convergence of the PPA have been reported by Rockafellar (1976a) for the case in which $\bar{Z} = \{\bar{z}\}$. If T^{-1} is Lipschitz continuous at 0 with modulus $a\geq 0$, then the approximate algorithm with $\{c_k\}$ nondecreasing, converges linearly at a rate bounded by $a/(a^2+c_\infty^2)^{1/2}$, which becomes superlinear if $\{c_k\}$ is unbounded (ibid., p. 883, th. 2). If $0 \in \text{int } T\bar{z}$, the exact algorithm with $\liminf_{k\to\infty} c_k > 0$ converges in a finite number of steps, while the approximate one with $\{c_k\}$ nondecreasing, achieves superlinear convergence without requiring that $\{c_k\}$ be unbounded (ibid., p.888, th.3).

The hypotheses used are of a geometric nature and concern the growth properties of the multivalued mapping T^{-1} in a neighbourhood of 0, away from the solution set. The general form of these growth conditions is

$$\exists \delta > 0 : \forall w \in B(0,\delta), \quad \forall z \in T^{-1}w \qquad |z-\bar{z}| \leq f(|w|),$$

where $B(0,\delta) = \{x \in H: |x| < \delta\}$, and $f: [0, +\infty) \to [0, +\infty)$ is a continuous function such that $f(0) = 0$. This type of assumption was suggested to the author by Professor D. Bertsekas of MIT. The characterizations of Rockafellar (1976a) discussed above, are special cases of this one.

When f is linear with slope $a>0$, we are able to guarantee linear convergence at a rate bounded by $a/(a^2+c_\infty^2)^{\frac{1}{2}}$. This is valid for both the exact and the approximate versions of the PPA with $r\geq 1$ in criterion (A_r). By means of an example we show that this bound is tight. The extension to general solution sets $\bar{Z}$, and the proof of tightness of the bound are new.

As shown by Rockafellar (1976b), the quadratic method of multipliers for convex programming is a realization of the PPA in which $T = -\partial g$, g being the essential objective function of the ordinary dual program. Taking this into account our Theorem 2.1, allows some extensions of the circumstances under which the quadratic method of multipliers achieves linear convergence as reported in Kort and Bertsekas (1973, 1976), and Rockafellar (1976b).

When f is a power function with exponent $s \geq 1$, we show in Theorem 3.1, that superlinear convergence of order at least s is obtained for the exact algorithm. For the approximate implementation, with criterion (A_r) and $r \geq 1$, the order of convergence is at least $\min\{r,s\}$. This result is entirely new. A comparison is made with results on the superlinear convergence of the quadratic method of multipliers reported by Kort and Bertsekas (1973, 1976).

If f is flat in some neighbourhood of 0 in $[0, +\infty)$, i.e. there is some $\delta>0$ such that $f(x) = 0$ for all $x \in [0,\delta)$, then the exact algorithm converges in a finite number of steps. The approximate version of the algorithm, with stopping criterion (A_r), $r \geq 1$, achieves superlinear convergence of order r at least. A sufficient condition for the convergence of the exact algorithm in a single step is also given.

When T^{-1} is such that its growth exceeds any linear bounding in any neighbourhood, however small, of zero then it is proved in Theorem 4.1, that the PPA will converge sublinearly when the penalty sequence $\{c_k\}$ remains bounded. This result is valid for both the exact and approximate versions. To the best of our knowledge this is the first result dealing with sublinear convergence of the PPA.

Finally, if f is a power function with exponent $s \in (0,1)$, we give a conservative estimate of the speed of convergence. It is shown that $|z^k - \bar{Z}|$ decreases to zero faster than $k^{-s/2}$. This result is also new.

This section ends with a proposition on the global convergence of the PPA which will be used repeatedly in what follows.

<u>Proposition 1.2:</u> Let $\bar{Z} \neq \emptyset$, and let $\{z^k\}$ be any sequence generated by the PPA with stopping criterion (A_r), $r \geq 0$, and a sequence of positive numbers $\{c_k\}$ such that $\liminf_{k\to\infty} c_k > 0$. Let us also assume that

$$\exists \delta > 0: \forall w \in B(0,\delta),\ \forall z \in T^{-1}w \quad |z-\bar{Z}| \leq f(|w|), \tag{1.4}$$

where $B(0,\delta) = \{z \in H : |z| < \delta\}$, and $f : [0, +\infty) \to [0, +\infty)$, is such that $f(0) = 0$, and upper semicontinuous at 0 (in this case equivalent to continuity at 0). Then $|z^k - \bar{Z}| \to 0$.

<u>Proof</u>: By Theorem 1.1, equation (1.3), we have $|c_k^{-1}Q_k z^k| \to 0$. There exists then, some $k_1 \in \mathbb{Z}_+$ such that $|c_k^{-1}Q_k z^k| < \delta$ for all $k \geq k_1$. By equation (1.1) and assumption (1.4)

$$\forall k \geq k_1 \qquad |P_k z^k - \bar{Z}| \leq f(|c_k^{-1}Q_k z^k|) \quad .$$

Using the continuity of f

$$\limsup_{k\to\infty} |P_k z^k - \bar{Z}| \leq \lim_{k\to\infty} f(|c_k^{-1}Q_k z^k|) = 0,$$

from which it follows that $|P_k z^k - \bar{Z}| \to 0$.

Also the definition of $|z - \bar{Z}|$, the triangle inequality, and criterion (A_r) yield for all k

$$|z^{k+1} - \bar{z}| \le |z^{k+1} - \overline{P_k z^k}| \le |z^{k+1} - P_k z^k| + |P_k z^k - \overline{P_k z^k}|$$

$$\le \varepsilon_k \min\{1, |z^{k+1} - z^k|^r\} + |P_k z^k - \bar{z}| \quad .$$

Since $\varepsilon_k \to 0$, the result follows.

Remark. Condition (1.4) is not necessary. To see this consider in $\ell^2(\mathbb{N})$ the quadratic function $q(x) = \sum_{i=0}^{\infty} \frac{x_i^2}{i+1}$. The PPA for $T=\partial q$ converges strongly to the unique solution x=0 (Kryanev 1973). Nonetheless, the graph of T is as flat as we may want in any neighborhood of zero, and such an f does not exist.

2. Linear Convergence

In this section a sufficient condition for the linear convergence of $\{|z^k - \bar{z}|\}$ to zero when the PPA is operated in an approximate manner is provided. The upper bound on the rate of convergence is shown to be tight. Implications for the quadratic method of multipliers are pointed out and comparisons with previous results are made. The main result of this section is embodied in the following theorem.

Theorem 2.1. Let $\bar{Z} \neq \emptyset$, and let $\{z^k\}$ be any sequence generated by the PPA with stopping criterion (A_r), $r \geq 1$, and a nondecreasing sequence of positive numbers $\{c_k\}$, such that $0 < c_k \uparrow c_\infty \leq +\infty$. Let us also assume that

$$\exists a > 0, \quad \exists \delta > 0 : \forall w \in B(0,\delta), \quad \forall z \in T^{-1}w \qquad |z - \bar{Z}| \leq a|w|. \tag{2.1}$$

Then $|z^k - \bar{Z}| \to 0$ linearly with a rate bounded from above by $a/(a^2+c_\infty^2)^{1/2} < 1$. If $c_\infty = +\infty$, the convergence is superlinear.

Proof. The hypothesis implies the one of Theorem 1.1, thus its conclusion is in force. By (1.3) there exists some $k_1 \in \mathbb{Z}_+$ such that $|c_k^{-1}Q_k z^k| < \delta$ for all $k \geq k_1$. Using formula (1.1) and assumption (2.1)

$$\forall k \geq k_1 \qquad |P_k z^k - \bar{Z}| \leq \frac{a}{c_k} |Q_k z^k| . \tag{2.2}$$

Equation (1.2) with $z = z^k$, $z' = \overline{z^k} \in \bar{Z}$ (thus $P_k z' = z'$, $Q_k z' = 0$), together with the flat $|P_k z^k - \overline{z^k}| \geq |P_k z^k - \bar{Z}|$, yields

$$\forall k \geq 0 \qquad |Q_k z^k|^2 \leq |z^k - \bar{Z}|^2 - |P_k z^k - \bar{Z}|^2 . \tag{2.3}$$

Using (2.3) to eliminate $|Q_k z^k|$ in (2.2)

$$\forall k \geq k_1 \qquad |P_k z^k - \bar{z}|^2 \frac{c_k^2+a^2}{a^2} \leq |z^k - \bar{z}|^2 .$$

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

$$\forall k \geq k_1 \qquad |P_k z^k - \bar{z}| \leq \mu_k |z^k - \bar{z}| . \tag{2.4}$$

From equation (2.3) we have

$$\forall k \geq 0 \qquad |Q_k z^k| \leq |z^k - \bar{z}| . \tag{2.5}$$

The triangle inequality gives

$$\forall k \geq 0 \qquad |z^k - \overline{P_k z^k}| \leq |z^k - \overline{z^k}| + |\overline{z^k} - \overline{P_k z^k}| .$$

Projection onto a nonempty closed convex set ($\bar{Z}$ in our case) is a non-expansive operation (in fact it is a proximal mapping, see Moreau 1965, p.279), thus $|\overline{z^k} - \overline{P_k z^k}| \leq |z^k - P_k z^k|$, and using (2.5)

$$\forall k \geq 0 \qquad |z^k - \overline{P_k z^k}| \leq 2|z^k - \bar{z}| . \tag{2.6}$$

By Theorem 1.1, $|z^{k+1} - z^k| \to 0$, and therefore there exists an index $k_2 \in \mathbb{Z}_+$, such that for all $k \geq k_2$ $|z^{k+1} - z^k| < 1$. If $r \geq 1$, then for all $k \geq k_2$ $|z^{k+1} - z^k|^r \leq |z^{k+1} - z^k|$, and criterion (A_r) can be used to obtain the estimate

$$\begin{aligned} \forall k \geq k_2 \qquad |z^{k+1} - \overline{P_k z^k}| &\leq |z^{k+1} - P_k z^k| + |P_k z^k - \overline{P_k z^k}| \\ &\leq \varepsilon_k |z^{k+1} - z^k| + |P_k z^k - \bar{z}| \\ &\leq \varepsilon_k |z^{k+1} - \overline{P_k z^k}| + \varepsilon_k |z^k - \overline{P_k z^k}| + |P_k z^k - \bar{z}| . \end{aligned}$$

But $|z^{k+1} - \overline{P_k z^k}| \geq |z^{k+1} - \bar{z}|$. By criterion (A_r), $\varepsilon_k \to 0$, thus there is some $k_3 \in \mathbb{Z}_+$ such that $\varepsilon_k < 1$ for all $k \geq k_3$. Let $\tilde{k} = \max\{k_2, k_3\}$. Using equation (2.6) and rearranging

$$\forall k \geq \tilde{k} \qquad |P_k z^k - \bar{z}| \geq (1 - \varepsilon_k)\, |z^{k+1} - \bar{z}| - 2\varepsilon_k |z^k - \bar{z}| \tag{2.7}$$

Let $\bar{k} = \max\{k_1, \tilde{k}\}$. Combining (2.4) and (2.7)

$$\forall k \geq \bar{k} \qquad |z^{k+1} - \bar{z}| \leq \frac{\mu_k + 2\varepsilon_k}{1-\varepsilon_k} |z^k - \bar{z}| \quad .$$

Thus the rate of linear convergence β, satisfies

$$\beta \leq \limsup_{k\to\infty} \frac{|z^{k+1} - \bar{z}|}{|z^k - \bar{z}|} \leq \lim_{k\to\infty} \frac{\mu_k + 2\varepsilon_k}{1 - \varepsilon_k} = \frac{a}{(a^2+c_\infty^2)^{\frac{1}{2}}} < 1 \quad . \tag{2.8}$$

<u>Example</u>. We will show by means of an example that the bound for the rate of linear convergence obtained in Theorem 2.1 is achieved.

Let us consider in $H = \mathbb{R}^2$ the linear transformation $A: \mathbb{R}^2 \to \mathbb{R}^2$, where A is given by the matrix

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

(Its effect is to rotate vectors counterclockwise by an angle of $\pi/2$).

Let us consider the quadratic form $\langle z, Az\rangle$ It is obvious that $\langle z, Az\rangle = 0$ for all $z \in \mathbb{R}^2$. The mapping $T: \mathbb{R}^2 \to 2^{\mathbb{R}^2}$, given by $Tz = \{Az\}$ is monotone because

$$\forall z, z' \in \mathbb{R}^2 \qquad \langle z-z', Az - Az'\rangle = \langle z-z', A(z-z')\rangle = 0.$$

Clearly T is single-valued and is continuous in $\mathbb{R}^2$. Therefore it is maximal monotone (Pascali & Sburlan 1978, Cor. 2.3, p. 106). T is not the subdifferential mapping of any proper lower semicontinuous convex function $f : \mathbb{R}^2 \to (-\infty, +\infty]$, as A is not self-adjoint (Rockafellar 1970b, p. 240).

It is easy to see that $|Az| = |z|$ for all $z \in \mathbb{R}^2$. Therefore $\bar{Z} = \{0\}$, the constants a, δ appearing in assumption (2.4) are $a = 1$, $\delta = +\infty$, and the inequality $|z-\bar{Z}| \le a|w|$ becomes $|z| = |w|$, valid for all $w \in \mathbb{R}^2$ and all $z \in T^{-1}w$, i.e., $z = A^{-1}w$.

When the PPA is implemented in its exact form, i.e., $\varepsilon_k \equiv 0$, it becomes $z^{k+1} = (I+c_kT)^{-1}z^k$, and in our case we obtain $|z^{k+1}| = |(I+c_kA)^{-1}z^k|$.

Elementary computations show that

$$(I+c_kA)^{-1} = \frac{1}{1+c_k^2}\begin{pmatrix} 1 & c_k \\ -c_k & 1 \end{pmatrix} ,$$

and also that

$$|z^{k+1}| = |(I+c_kA)^{-1}z^k| = \frac{|z^k|}{\sqrt{1+c_k^2}} .$$

Therefore the convergence rate is

$$\beta = \lim_{k\to\infty} \frac{|z^{k+1}|}{|z^k|} = \lim_{k\to\infty} \frac{1}{\sqrt{1+c_k^2}} = \frac{1}{\sqrt{1+c_\infty^2}} .$$

Since in this example $a = 1$, we have $\beta = a/(a^2+c^2)^{1/2}$, and the bound is achieved.

Let us consider the following convex programming problem

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0 \qquad i = 1,2,\ldots, m \\ & x \in C \end{aligned} \tag{2.9}$$

where C is a nonempty closed convex subset of $\mathbb{R}^n$ and $f_i : C \to \mathbb{R}$, is a lower semicontinuous convex function for $i = 0,1,\ldots,m$.

Its ordinary dual problem is

$$\begin{aligned} \max \quad & g_0(y) \\ \text{s.t.} \quad & y \ge 0 \end{aligned} \tag{2.10}$$

where $g_0 : \mathbb{R}^m_+ \to \overline{\mathbb{R}}$, is the concave function defined by

$$g_0(y) = \inf_{x \in C} \{f_0(x) + y_1 f_1(x) + \ldots + y_m f_m(x)\} \quad .$$

The quadratic method of multipliers involves a sequence of minimizations of the Augmented Lagrangian function

$$L(x,y,c) = \begin{cases} f_0(x) + \frac{1}{2c} \sum_{i=1}^{m} [(\max\{0, y_i + cf_i(x)\})^2 - y_i^2] & c>0 \quad x \in C \\ +\infty & \qquad\quad x \notin C \end{cases} \quad .$$

The method of multipliers can thus be expressed as

$$\begin{aligned} x^k &= \arg\min_x L(x, y^k, c_k) \\ y_i^{k+1} &= \max\{0, y_i^k + c_k f_i(x^k)\} \quad i = 1,2,\ldots,m. \end{aligned} \tag{2.11}$$

Rockafellar (1976b) has shown that the quadratic method of multipliers (2.11) for the solution of (2.9), is a realization of the PPA, in which $T = -\partial g$, where g is the essential objective function of the dual problem

(2.10) defined by

$$g(y) = \begin{cases} g_0(y) & \text{if } y \in \mathbb{R}^m_+ \\ -\infty & \text{if } y \notin \mathbb{R}^m_+ , \end{cases}$$

assuming that g is proper, i.e. sup $g > -\infty$, so that T is maximal monotone. He assumes that T^{-1} is Lipschitz continuous at the origin - which implies that there is only one Lagrange multiplier vector $\bar{y}$- and that the minimization to determine x^k in (2.11) is carried out only approximately with stopping criterion

$$L(x^{k+1}, y^k, c_k) - \inf_x L(x, y^k, c_k) \le (\varepsilon_k^2/2c_k)|y^{k+1}-y^k|^2 ,$$

where y^{k+1} is given as in (2.11), $\varepsilon_k \ge 0$ for all k, and $\sum_{k=0}^{\infty} \varepsilon_k < +\infty$. He concludes that the sequences $\{y^k\}$ generated by the algorithm converge towards $\bar{y}$ linearly, at a rate bounded as in (2.8). He also shows that

$$|y^{k+1} - P_k y^k|^2/2c_k \le L(x^{k+1}, y^k, c_k) - \inf_x L(x, y^k, c_k) ,$$

and thus the stopping criterion implies (A_r) with $r = 1$.

Kort and Bertsekas (1973, 1976) have also studied the convergence for this method of multipliers. In their analysis it is assumed

(i) Problem (2.9) has a nonempty compact solution set $\bar{X}$, and a nonempty compact set of Lagrange multipliers $\bar{Y}$.

(ii) f_0 is strongly convex with modulus $\mu > 0$. This implies that $\bar{X} = \{\bar{x}\}$.

(iii) The following growth condition on the dual function

$$\exists b > 0, \quad \exists \delta > 0: \quad \forall y \in B(\bar{Y},\delta) \qquad g(y) \le \bar{g} - \frac{1}{b}|y - \bar{Y}|^2,$$

where

$$\bar{g} = \max_y g, \quad B(\bar{Y},\delta) = \{y \in \mathbb{R}^m : |y - \bar{Y}| < \delta\},$$

and

$$|y - \bar{Y}| = \max\{|y-y'| : y' \in \bar{Y}\}$$

which is well defined as $\bar{Y}$ is closed and convex.

The strong convexity assumption (ii) allows them to develop an implementable criterion which implies the following one

$$|y^{k+1} - P_k y^k|^2 \le \frac{\eta_k}{2\mu}|y^{k+1} - y^k|^2, \tag{2.12}$$

where η_k is a prespecified sequence such that $\eta_k < 2\mu$ for all k large enough. When minimization of the Augmented Lagrangian is carried out exactly (ii) need not be assumed. The growth condition (iii) implies (2.1) with $a = b$. Linear convergenge is guaranteed if $\bar{\eta} < 4\mu\bar{c}/b$ where $\bar{\eta} = \limsup_{k\to\infty} \eta_k$, $\bar{c} = \limsup_{k\to\infty} c_k$. If $\eta_k \to 0$, the rate of linear convergence is bounded (as in the case of exact implementation of (2.11)) by $a/(a+\bar{c})$.

When interpreted in the framework of the method of multipliers, Theorem 2.1, gives a sufficient condition for its linear convergence under still weaker assumptions than those discussed above. First, both $\bar{X}$ and $\bar{Y}$ are required to be only nonempty (they will always be closed and convex by the lower semicontinuity and convexity of the functions $f_i, i = 0,1,\ldots, m$), and no assumption is made on their compactness. Secondly, the strong convexity assumption on f_0 is not made.

3. Superlinear Convergence and Convergence in a Finite Number of Iterations

The (Q-) order of convergence (Ortega & Rheinboldt 1970) of $\{|z^k-\bar{Z}|\}$, assuming that $|z^k - \bar{Z}| \neq 0$ for all k, is the supremum t of the numbers $\tau \geq 1$ such that

$$\limsup_{k\to\infty} \frac{|z^{k+1} - \bar{Z}|}{|z^k - \bar{Z}|^\tau} < +\infty \quad .$$

Theorem 3.1. Let $\bar{Z} \neq \emptyset$, and let $\{z^k\}$ be any sequence generated by the PPA with stopping criterion (A_r), $r \geq 1$, and a nondecreasing sequence of positive numbers $\{c_k\}$, such that $0 < c_k \uparrow c_\infty \leq +\infty$. Let us also assume that

$$\exists a > 0,\ \exists s \geq 1,\ \exists \delta > 0:\ \forall\, w \in B(0,\delta),\ \forall z \in T^{-1}w \quad |z-\bar{Z}| \leq a|w|^s. \tag{3.1}$$

Then $|z^k-\bar{Z}| \to 0$, and its (Q-) order of convergence satisfies $t \geq \min\{r,s\}$.

Proof. The hypothesis of the theorem subsumes that of Theorem 1.1, and therefore $|c_k^{-1}Q_k z^k| \to 0$. By equation (1.1) and assumption (3.1), there is some $k_1 \in \mathbb{Z}_+$ such that

$$\forall k \geq k_1 \qquad |P_k z^k - \bar{Z}| \leq \frac{a}{c_k^s} |Q_k z^k|^s \quad . \tag{3.2}$$

Using (2.3) to eliminate $|Q_k z^k|$ in (3.2)

$$\forall k \geq k_1 \qquad |P_k z^k - \bar{Z}|^2 + \frac{c_k^2}{a^{2/s}} |P_k z^k - \bar{Z}|^{2/s} \leq |z^k - \bar{Z}|^2 \ ,$$

from which

$$\forall k \geq k_1 \qquad |P_k z^k - \bar{z}| \leq \frac{a|z^k - \bar{z}|^s}{(c_k^2 + a^{2/s}|P_k z^k - \bar{z}|^{2(s-1)/s})^{s/2}} . \tag{3.3}$$

The triangle inequality and criterion (A_r) yield the following estimate for all k

$$|z^{k+1} - \overline{P_k z^k}| \leq |z^{k+1} - P_k z^k| + |P_k z^k - \overline{P_k z^k}|$$

$$\leq \varepsilon_k |z^{k+1} - z^k|^r + |P_k z^k - \bar{z}|$$

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

Rearranging and using the fact $|z^k - \overline{P_k z^k}| \leq 2|z^k - \bar{z}|$ (cf. (2.6))

$$\forall k \geq 0 \qquad |P_k z^k - \bar{z}| \geq (1 - \varepsilon_k |z^{k+1} - z^k|^{r-1})|z^{k+1} - \overline{P_k z^k}|$$

$$- 2\varepsilon_k |z^{k+1} - z^k|^{r-1}|z^k - \bar{z}| \quad .$$

By Theorem 1.1, $|z^{k+1} - z^k| \to 0$, and therefore there is some $k_2 \in \mathbb{Z}_+$ such that $|z^{k+1} - z^k| < 1$, and $|z^{k+1} - z^k|^{r-1} \leq 1$ for all $k \geq k_2$ as $r \geq 1$. Also, by criterion (A_r), $\varepsilon_k \to 0$, and thus there is some $k_3 \in \mathbb{Z}_+$ such that $\varepsilon_k < 1$ for all $k \geq k_3$. Hence for all $k \geq \tilde{k} = \max\{k_2, k_3\}$, $1-\varepsilon_k|z^{k+1}-z^k|^r \geq 1-\varepsilon_k > 0$, and being $|z^{k+1} - \overline{P_k z^k}| \geq |z^{k+1} - \bar{z}|$

$$\forall k \geq \tilde{k} \qquad |P_k z^k - \bar{z}| \geq (1 - \varepsilon_k)|z^{k+1} - \bar{z}|$$

$$-2\varepsilon_k |z^{k+1} - z^k|^{r-1}|z^k - \bar{z}| \quad .$$

From equation (2.5), the triangle inequality, and the fact $r \geq 1$

$$|z^k-\bar{z}| \geq |Q_k z^k| = |z^k - P_k z^k| \geq |z^k-z^{k+1}| - |z^{k+1} - P_k z^k|$$

$$\geq |z^k - z^{k+1}|(1-\varepsilon_k|z^k - z^{k+1}|^{r-1}),$$

which can be transformed into the following estimate valid for all $k \geq \tilde{k}$

$$|z^k - z^{k+1}| \leq \frac{|z^k - \bar{z}|}{1-\varepsilon_k} . \tag{3.5}$$

Combining (3.4) and (3.5)

$$\forall k \geq \tilde{k} \qquad |P_k z^k - \bar{z}| \geq (1-\varepsilon_k)|z^{k+1} - \bar{z}| - \frac{2\varepsilon_k}{(1-\varepsilon_k)^{r-1}} |z^k - \bar{z}|^r . \tag{3.6}$$

Let $\bar{k} = \max\{k_1, \tilde{k}\} = \max\{k_1, k_2, k_3\}$. The combination of (3.3) and (3.6) yields for all $k \geq \bar{k}$

$$|z^{k+1} - \bar{z}| \leq \frac{a|z^k - z|^s}{(1-\varepsilon_k)(c_k^2 + a^{2/s}|P_k z^k-\bar{z}|^{2(s-1)/s})^{s/2}} + \frac{2\varepsilon_k}{(1-\varepsilon_k)^r} |z^k-\bar{z}|^r. \tag{3.7}$$

Assumption (3.1) implies the hypothesis of Proposition (1.2) with $f(|\cdot|) = a|\cdot|^s$, and thus $|z^k - \bar{z}| \to 0$. Also, by criterion (A_r), $\varepsilon_k \to 0$, and therefore from (3.7) it clearly follows that the (Q-) order of convergence of $\{|z^k - \bar{z}|\}$ is at least $\min\{r,s\} \geq 1$.

<u>Remark</u>. An alternative proof can be obtained by using (2.5) instead of (2.3) to eliminate $|Q_k z^k|$ in (3.2), and then (3.6) to obtain

$$\forall k \geq \bar{k} \qquad |z^{k+1}-\bar{Z}| \leq \frac{a}{(1-\varepsilon_k)c_k^s}|z^k - \bar{Z}|^s + \frac{2\varepsilon_k}{(1-\varepsilon_k)^r}|z^k - \bar{Z}|^r .$$

The proof chosen has the advantage of clearly showing the connection in equation (3.7) with the linear convergence case (take s=1 to obtain (2.8)).

In the context of the quadratic method of multipliers, Kort and Bertsekas (1973, 1976) have also specified conditions for the superlinear convergence to zero of the sequences $\{|y^k - \bar{Y}|\}$. The assumptions under which this result is obtained include (i) and (ii) as in § 2 above for the case of inexact minimization of the Augmented Lagrangian, while (iii) takes the form

$$\exists b > 0, \quad \exists q \in (1,2), \quad \exists \delta > 0: \quad \forall y \in B(Y,\delta) \qquad g(y) \leq \bar{g} - \frac{1}{b}|\bar{y} - \bar{Y}|^q.$$

With the help of the subgradient inequality for the concave function g

$$\forall y^* \in \partial g(y), \forall \bar{y} \in \bar{Y} \qquad \bar{g} \leq g(y) + \langle \bar{y} - y, y^* \rangle \leq g(y) + |y-y||y^*|,$$

the assumption above becomes

$$\exists b > 0, \exists q \in (1,2), \exists \delta > 0: \quad \forall y \in B(\bar{Y},\delta), \forall y^* \in \partial g(y) \quad |y-\bar{Y}| \leq b|y^*|^{1/(q-1)}.$$

Clearly when $q \in (1,2)$, $s = 1/(q-1) \in (1, +\infty)$, thus obtaining a growth condition on ∂g^{-1} analogous to the assumption (3.1) used in the proof of our theorem. When the algorithm is implemented in exact form (the strong convexity of f_0 is not needed in this case), the (Q-) order of convergence is at least $1/(q-1)$ which coincides with our result (3.7). When the algorithm is implemented only approximately (see (2.12)), the (Q-) order of convergence obtained is $2/q$ (Kort & Bertsekas 1976, Prop. 7, p. 286). Taking into account that $1/(q-1) = s$, this order becomes $2s/(1+s)$ in our notation and satisfies

$$\forall s > 1 \qquad 1 < \frac{2s}{1+s} < s \ .$$

In order to achieve the same order of convergence as with the exact algorithm, the sequence η_k in (2.12) has to be replaced by $\min\{\hat{\eta}_k,\ c|y^{k+1} - y^k|^2\}$, where $\hat{\eta}_k \to 0$, $c > 0$, and $a \geq s-1$ (Kort & Bertsekas 1976, Cor. 7.1, p.288). With this modification the actual criterion for the approximate implementation implies

$$|y^{k+1} - P_k y^k|^2 \leq c|y^{k+1} - y^k|^{s+1} \ .$$

This is less stringent than criterion (A_r) with $r \geq s$ which implies that for all k large enough (after $|y^{k+1} - y^k| < 1$)

$$|z^{k+1} - P_k z^k| \leq \varepsilon_k |z^{k+1} - z^k|^s, \qquad \sum_{k=0}^{\infty} \varepsilon_k < +\infty \ .$$

The difference in orders of convergence might be accounted for by the following facts

a) The presence in the method of multipliers of subgradient inequalities which are not available for a general monotone operator.

b) The assumptions made on f_0, $\overline{X}$, and $\overline{Y}$.

We analyze now the conditions under which finite convergence is obtained.

<u>Theorem 3.2.</u> Let $\overline{Z} \neq \emptyset$, and let $\{z^k\}$ be any sequence generated by the PPA either in exact form ($\varepsilon_k \equiv 0$), or with stopping criterion

(A_r), with $r = 0$ or $r \geq 1$, and a sequence of positive numbers $\{c_k\}$, such that $\liminf_{k\to\infty} c_k > 0$. Let us also assume that

$$\exists \delta > 0: \ \forall w \in B(0,\delta),\ \forall z \in T^{-1}w \quad z \in \bar{Z} \tag{3.8}$$

Then for all k large enough

$$|z^{k+1} - \bar{z}| \leq \frac{\varepsilon_k}{(1-\varepsilon_k)^r} |z^k - \bar{z}|^r \tag{3.9}$$

If the PPA is operated in exact form ($\varepsilon_k \equiv 0$), convergence is achieved in a finite number of iterations. Otherwise, if $r \geq 1$, superlinear convergence of order at least r, is guaranteed.

Proof. Theorem 1.1 applies, and by (1.3), $|c_k^{-1}Q_k z^k| \to 0$, so there is some $k_1 \in \mathbb{Z}_+$ such that $|c_k^{-1}Q_k z^k| < \delta$ for all $k \geq k_1$. By equation (1.1) and assumption (3.8)

$$\forall k \geq k_1 \qquad |P_k z^k - \bar{Z}| = 0 \ . \tag{3.10}$$

Equation (1.3) implies that $|z^{k+1} - z^k| \to 0$, so there is some $k_2 \in \mathbb{Z}_+$ such that $|z^{k+1} - z^k| \leq 1$ for all $k \geq k_2$, and the inequality $\min\{1, |z^{k+1} - z^k|^r\} \leq |z^{k+1} - z^k|^r$ is valid for all $k \geq k_2$ for $r = 0$ or $r \geq 1$. Letting $\tilde{k} = \max\{k_1, k_2\}$, the triangle inequality, criterion (A_r), and (3.10) yield

$$\forall k \geq \tilde{k} \qquad |z^{k+1} - \bar{z}| \leq |z^{k+1} - \overline{P_k z^k}| \leq |z^{k+1} - P_k z^k| + |P_k z^k - \overline{P_k z^k}|$$
$$\leq \varepsilon_k \min\{1, |z^{k+1} - z^k|^r\} \leq \varepsilon_k |z^{k+1} - z^k|^r \ , \tag{3.11}$$

valid for r=0 or $r \geq 1$.

By criterion (A_r), $\varepsilon_k \downarrow 0$, thus there is some $k_3 \in \mathbb{Z}_+$, such that $\varepsilon_k < 1$ for all $k \geq k_3$. When $r \geq 1$ and $k \geq \bar{k} = \max\{k_1, k_2, k_3\} \geq \tilde{k}$, (3.5) holds, and (3.11) can be transformed into

$$\forall k \geq \bar{k} \qquad |z^{k+1} - \bar{Z}| \leq \frac{\varepsilon_k}{(1-\varepsilon_k)^r} |z^k - \bar{Z}|^r . \tag{3.12}$$

The theorem follows because it is clear that (3.10), (3.11), and (3.12) are equivalent to (3.9) when the PPA is implemented in exact form ($\varepsilon_k \equiv 0$), with (A_r) and $r = 0$, and with (A_r) and $r \geq 1$ respectively.

Remark. A condition for the convergence of the exact PPA in a single step can be easily obtained as follows. By (1.1) $c_0^{-1}Q_0z^0 \in TP_0z^0$, so if $|c_0^{-1}Q_0z^0| < \delta$ then $z^1 = P_0z^0 \in \bar{Z}$. Q_0 is the proximal mapping for the maximal monotone operator $(c_0T)^{-1}$, and thus it is non-expansive. Hence for any z, $z' \in H$, $|Q_0z - Q_0z'| \leq |z-z'|$. We know that if $z' \in \bar{Z}$, then $Q_0z' = 0$. Let us choose $z = z^0$, $z' = \overline{z^0} \in \bar{Z}$, then the estimate $|Q_0z^0| \leq |z^0 - \bar{Z}|$ is obtained. Thus a sufficient condition for $|c_0^{-1}Q_0z^0| < \delta$ is $c_0 > |z^0 - \bar{Z}|/\delta$. A condition of this type appeared for the first time in Bertsekas (1975).

Rockafellar (1976a, Th. 3, p.888) showed the finite convergence of the PPA under the assumption that $0 \in \text{int } T\bar{z}$ for some $\bar{z} \in H$. This assumption implies that $\bar{z}$ is the unique solution of $0 \in Tz$. On the other hand, our result applies in the general case in which $\bar{Z}$ need not be a singleton or even compact.

Viewed in the context of the quadratic method of multipliers, Theorem 3.2, guarantees finite convergence without the need of making compactness assumptions on $\bar{X}$ (Bertsekas 1975) or uniqueness of the Lagrange multipliers, i.e. $\overline{Y} = \{\overline{y}\}$ (Rockafellar 1976b).

The generalization of Rockafellar's criterion $0 \in \text{int } T\overline{z}$ for some $\overline{z} \in H$, to a general nonempty $\overline{Z}$ would be

$$\exists \delta > 0: \quad B(0,\delta) \subseteq T\overline{Z} \quad . \tag{3.13}$$

Instead we have used (cf. (3.8))

$$\exists \delta > 0: \quad T^{-1}B(0,\delta) \subseteq \overline{Z} \ , \tag{3.14}$$

which is the obvious limiting case of (3.1) when $s \to \infty$ and $\delta < 1$ (this last condition can be arranged by taking some $\delta' < \min\{1,\delta\}$).

It is interesting to explore the relationship between (3.13) and (3.14). From our analysis (see Prop. 3.4 below), it follows that (3.13) implies not only (3.14) but also that $\overline{Z}$ is bounded. On the other hand there are instances in which (3.14) holds but (3.13) does not. For example, if $\overline{Z}$ is unbounded, as it happens for $H = \mathbb{R}$, when the graph of T is given by $G(T) = \mathbb{R}_- \times \{0\} \cup \{0\} \times [0,1] \cup \mathbb{R}_+ \times \{1\}$.

To show this relationship we will first prove two technical lemmas.

Lemma 3.2. Let T be a maximal monotone operator such that $\bar{Z} = T^{-1}0$ is nonempty.

Then $Tz \subseteq N_{\bar{Z}}(z)$ for all $z \in H$, where $N_{\bar{Z}}(z)$ denotes the normal cone to $\bar{Z}$ at z. In particular if $z \in \text{int}\,\bar{Z}$, the interior of $\bar{Z}$ in the strong topology of H, $Tz = \{0\}$.

Proof. For all $z \in H$, the cone normal to $\bar{Z}$ at z is given by (Rockafellar 1970b, p.15).

$$N_{\bar{Z}}(z) = \{x \in H: \ \forall u \in \bar{Z} \quad \langle z-u, x\rangle \geq 0\} \ . \tag{3.15}$$

If $z \notin D(T)$, then $Tz = \emptyset$ and the inclusion $Tz \subseteq N_{\bar{Z}}(z)$ is trivial. Let $z \in D(T)$, and $w \in Tz$. Then the monotonicity of T implies

$$\forall z' \in \bar{Z} \qquad \langle z-z', w\rangle \geq 0$$

and it follows that $w \in N_{\bar{Z}}(z)$. If $z \in \text{int}\,\bar{Z}$ then $N_{\bar{Z}}(z) = \{0\} \supseteq Tz \neq \emptyset$, so $Tz = \{0\}$.

Lemma 3.3. Let C be a nonempty closed convex and bounded subset of a real Hilbert space H. Let $N_C(z)$ denote the normal cone to C at z.

If $z \notin C$, then $N_C(z)$ has a nonempty interior in the strong topology which is also a convex cone.

Proof: Since C is nonempty closed and convex, for any $z \in H$ there is a unique vector $\bar{z} \in C$ which is closest to z. This vector is characterized by (Luenberger 1969, Th. 1, p.69) $\langle z-\bar{z}, \bar{z}-u\rangle \geq 0$ for all $u \in C$. But

$<z-\bar{z}, \bar{z}-u> = <z-\bar{z}, z-u> - |z-\bar{z}|^2$ which clearly shows that for all $u \in C$ $<z-\bar{z}, z-u> \geq |z-\bar{z}|^2 \geq 0$, and therefore $z - \bar{z} \in N_C(z)$ (see (3.15)).

It will now be shown that if C is bounded and $z \notin C$, then $z-\bar{z} \in \text{int } N_C(z)$. Let us suppose that $z-\bar{z} \notin \text{int } N_C(z)$, then for any $\delta > 0$, there is some $v \in B(0,\delta)$ such that $z - \bar{z} + v \notin N_C(z)$. By the definition of $N_C(z)$, this implies that there is some vector $p \in C$ such that $<z-p, z-\bar{z} + v> < 0$, or $<z-p, z-\bar{z}> < <p-z, v>$. But $<z-p, z-\bar{z}> = <z-\bar{z}, z-\bar{z}> + <\bar{z}-p, z-\bar{z}> \geq |z-\bar{z}|^2 > 0$, because $p \in C$ and $\bar{z}$ is the projection of z onto C.

Using successively the Cauchy-Bunyakovsky and triangle inequalities, and boundedness of C (i.e., $\exists M \in \mathbb{R}$: $C \subseteq B(0,M)$)

$$0 < |z-\bar{z}|^2 < <p-z, v> \leq |p-z||v|$$

$$\leq (|p| + |z|)|v| \leq (M+|z|)|v| \; .$$

Thus $|v| > |z-\bar{z}|^2/(M+|z|) > 0$. Let us choose $0 < \delta < |z-\bar{z}|^2/(M+|z|)$ to obtain a contradiction with $v \in B(0,\delta)$, and therefore $z-\bar{z} \in \text{int } N_C(z)$. It is easy to prove that if K is a convex cone so is int K, and therefore int $N_C(z)$ is a convex cone.

Proposition 3.4: Let T and $\bar{Z}$ be as above. Then $0 \in \text{int } T\bar{Z}$ implies that $\bar{Z}$ is bounded. Moreover, there is some $\delta>0$ such that for all $w \in B(0,\delta)$, $z \in T^{-1}w \Rightarrow z \in \bar{Z}$. In particular, if $w \in B(0,\delta)\backslash\{0\}$ and $z \in T^{-1}w$ then $z \in \partial\bar{Z} = \bar{Z}\backslash\text{int } \bar{Z}$, or more suggestively $T^{-1}(B(0,\delta)\backslash\{0\}) \subseteq \partial\bar{Z}$.

Proof. $\bar{Z}$ is closed, therefore it contains its boundary $\partial\bar{Z}$, and $\bar{Z} = \text{int } \bar{Z} \cup \partial\bar{Z}$. We also have by Lemma 3.2

$$T\bar{Z} = \cup\{Tz\colon z \in \bar{Z}\} = T \text{ int } \bar{Z} \cup T\partial\bar{Z}$$

$$= \{0\} \cup T\partial\bar{Z} = T\partial\bar{Z} .$$

By the hypothesis, there is some $\delta>0$ such that $B(0,\delta) \subseteq T\bar{Z}$, and thus $B(0,\delta) \subseteq T\partial\bar{Z}$.

Let us denote by $N_{\bar{Z}}(z)$ the cone normal to $\bar{Z}$ at some given vector $z \in H$. $N_{\bar{Z}}(z)$ is convex and its expression is given by (3.15). Using Lemma 3.2, and the hypothesis

$$B(0,\delta) \subseteq \bigcup\{Tz\colon z \in \partial\bar{Z}\} \subseteq \bigcup\{N_{\bar{Z}}(z)\colon z \in \partial\bar{Z}\} , \tag{3.16}$$

and this implies that $\bar{Z}$ is bounded. Suppose that $\bar{Z}$ were unbounded, then since $\bar{Z}$ is convex, there is some $z \in \bar{Z}$ and $u \in H$ with $|u| = 1$ such that for all $\lambda \geq 0$, $z + \lambda u \in \bar{Z}$ (Rockafellar 1970b, p.61). For any $\tau \in (0,\delta)$ $\tau u \in B(0,\delta)$ which implies, by (3.16), that there is some $z' \in \partial\bar{Z}$ such that $\tau u \in Tz'$. By the monotonicity of T and the fact $z + \lambda u \in \bar{Z}$.

$$\forall \lambda \geq 0 \qquad \langle z' - (z+\lambda u), \tau u\rangle \geq 0 ,$$

or since $\tau>0$ and $|u| = 1$,

$$\forall\lambda \geq 0 \qquad \lambda \leq \langle z'-z, u\rangle \leq |z'-z|$$

whic is clearly a contradiction. Thus such u does not exist for any $z \in \bar{Z}$, and $\bar{Z}$ is bounded.

To prove the second part, let us assume that for some $z \in D(T)\setminus\bar{Z}$, there is some $w \in Tz$ such that $|w| < \delta$. Since $\bar{Z}$ is convex and bounded, by Lemma 3.3, for any $z \notin \bar{Z}$, the interior of $N_{\bar{Z}}(z)$ is a nonempty convex cone. Let $p \in \text{int } N_{\bar{Z}}(z) \cap B(0, \delta-|w|) \neq \emptyset$. Clearly, $0 < |p| < \delta - |w|$, and

$\forall u \in \bar{Z} \qquad \langle z-u, p+w\rangle = \langle z-u, p\rangle + \langle z-u,w\rangle \geq 0,$

because $p \in \text{int } N_{\bar{Z}}(z) \subseteq N_{\bar{Z}}(z)$, and $w \in Tz \subseteq N_{\bar{Z}}(z)$. The triangle inequality yields $|p+w| \leq |p| + |w| < \delta - |w| + |w| = \delta$, and $p+w \in B(0,\delta)$. By (3.16), there is some $z' \in \partial\bar{Z}$ such that $p+w \in Tz' \subseteq N_{\bar{Z}}(z')$. The monotonicity of T implies $0 \leq \langle z-z', w-(p+w)\rangle = -\langle z-z', p\rangle$. But $p \in N_{\bar{Z}}(z)$ and $z' \in \bar{Z}$ imply that $\langle z-z', p\rangle \geq 0$, thus $\langle z-z', p\rangle = 0$. As $p \in \text{int } N_{\bar{Z}}(z)$, there is some $\tau>0$ such that $B(p,\tau) \subseteq N_{\bar{Z}}(z)$. For any $\nu \in (0,\tau)$, $p + \nu(z'-z)/|z'-z| \in N_{\bar{Z}}(z)$. By the definition of $N_{\bar{Z}}(z)$ given in (3.15), this implies that

$$\langle z-z', p + \nu \frac{z'-z}{|z'-z|}\rangle \geq 0 .$$

Since $\nu>0$ and $\langle z-z', p\rangle = 0$ we obtain $0 \leq \langle z-z', z'-z\rangle < 0$ a contradiction. Therefore we cannot assume that for some $z \in D(T)\setminus\bar{Z}$ there exists some $w \in Tz$ with $|w| < \delta$. It follows that $|w| < \delta$ implies $z \in \bar{Z}$.

4. Sublinear Convergence

This section starts with a partial converse to Theorem 2.1.

Theorem 4.1. Let $\bar{Z} \neq \emptyset$, and let $\{z^k\}$ be any sequence generated by the PPA with stopping criterion (A_r), with $r \geq 1$, and a nondecreasing sequence of positive numbers $\{c_k\}$, such that $0 < c_k \uparrow c_\infty < +\infty$. Let us also assume that

$$\forall a > 0,\ \exists\ \delta>0:\quad \forall w \in B(0,\delta),\ \forall z \in T^{-1}w \qquad |z-\bar{z}| \geq a|w|. \tag{4.1}$$

Then if $\{z^k\}$ does not converge to $\bar{Z}$ in a finite number of steps (i.e., $z^k \notin \bar{Z}$ for all k)

$$\liminf_{k\to\infty} \frac{|z^{k+1}-\bar{z}|}{|z^k-\bar{z}|} = 1,$$

and $\{|z^k - \bar{z}|\}$ cannot converge to zero faster than sublinearly.

Proof. Let us choose some fixed $a>0$. Theorem 1.1 applies and by (1.3) $|c_k^{-1}Q_k z^k| \to 0$. Therefore there is some $k_1 \in \mathbb{Z}_+$ such that $|c_k^{-1}Q_k z^k| < \delta$ for all $k \geq k_1$. By equation (1.1) and assumption (4.1):

$$\forall k \geq k_1 \qquad |P_k z^k-\bar{z}| \geq \frac{a}{c_k} |Q_k z^k| . \tag{4.2}$$

By the triangle inequality

$$|Q_k z^k| = |z^k-P_k z^k| \geq |z^k-z^{k+1}| - |z^{k+1}-P_k z^k| . \tag{4.3}$$

The triangle inequality, and the fact that projection onto a nonempty closed convex set is a nonexpansive mapping (Moreau 1965, p.279) yield

$$|P_k z^k - \bar{z}| \leq |P_k z^k - z^{k+1}| + |z^{k+1} - \overline{z^{k+1}}| + |\overline{z^{k+1}} - \overline{P_k z^k}|$$

$$\leq 2|P_k z^k - z^{k+1}| + |z^{k+1} - \bar{z}| \ . \tag{4.4}$$

Using (4.3) and (4.4) we see that (4.2) can be transformed into

$$\forall k \geq k_1 \qquad (2c_k + a)|P_k z^k - z^{k+1}| + c_k |z^{k+1} - \bar{z}| \geq a|z^k - z^{k+1}| \ . \tag{4.5}$$

By equation (1.3) there is some $k_2 \in \mathbb{Z}_+$ such that $|z^{k+1} - z^k|^r \leq |z^{k+1} - z^k| < 1$ for all $k \geq k_2$. Hence criterion (A_r) yields

$$\forall k \geq k_2 \qquad |P_k z^k - z^{k+1}| \leq \varepsilon_k \min\{1, |z^{k+1} - z^k|^r\} \leq \varepsilon_k |z^{k+1} - z^k| . \tag{4.6}$$

We also have that

$$|z^k - z^{k+1}| \geq |z^k - \overline{z^{k+1}}| - |z^{k+1} - \overline{z^{k+1}}| \geq |z^k - \bar{z}| - |z^{k+1} - \bar{z}| . \tag{4.7}$$

By combining (4.5)-(4.7) we obtain for all $k \geq \tilde{k} = \max\{k_1, k_2\}$

$$(c_k + a)|z^{k+1} - \bar{z}| \geq a|z^k - \bar{z}| - (2c_k + a)\varepsilon_k |z^{k+1} - z^k| .$$

Criterion (A_r) implies that $\varepsilon_k \to 0$. Thus there is some $k_3 \in \mathbb{Z}_+$ such that $\varepsilon_k < 1$ for all $k \geq k_3$. If $k \geq \bar{k} = \max\{k_3, \tilde{k}\}$, the above inequality and (3.5) are valid and using the latter to substitute for $|z^{k+1} - z^k|$ in the former

$$\forall k \geq \bar{k} \qquad (c_k + a)|z^{k+1} - \bar{z}| \geq a|z^k - \bar{z}| - \frac{(2c_k + a)\varepsilon_k}{1 - \varepsilon_k} |z^k - \bar{z}| \ .$$

From it, taking into account that $\varepsilon_k \to 0$,

$$\liminf_{k\to\infty} \frac{|z^{k+1} - \bar{z}|}{|z^k - \bar{z}|} \geq \lim_{k\to\infty} \frac{a}{a + c_k} = \frac{a}{a + c_\infty} \ .$$

Since a can be arbitrarily large, the theorem follows.

The preceding theorem provides an essentially negative result. In the next theorem we try to quantify the speed of convergence. Since the convergence maybe sublinear, we will have to look for an estimate of the form $|z^k - \bar{z}| = O(k^{-\sigma})$ for some $\sigma > 0$.

Theorem 4.2. Let $\bar{Z} \neq \emptyset$, and let $\{z^k\}$ be any sequence generated by the PPA in exact form with a nondecreasing sequence of positive numbers $\{c_k\}$, such that $0 < c_k \uparrow c_\infty < +\infty$. Let us also assume that

$$\exists a > 0, \exists s \in (0,1), \exists \delta > 0: \quad \forall w \in B(0,\delta), \quad \forall z \in T^{-1}w \quad |z - \bar{z}| \leq a|w|^s. \tag{4.8}$$

Then $|z^k - \bar{z}| \to 0$ as $o(k^{-s/2})$, i.e., $\lim_{k\to\infty} |z^k - \bar{z}|^{2/s} k = 0$.

Proof. By Theorem 1.1, $|c_k^{-1} Q_k z^k| \to 0$, thus there exists some $k_1 \in \mathbb{Z}_+$ such that $|c_k^{-1} Q_k z^k| < \delta$ for all $k \geq k_1$. Also by (1.1) $c_k^{-1} Q_k z^k \in TP_k z^k = Tz^{k+1}$. Using these facts and assumption (4.8)

$$\forall k \geq k_1 \qquad |z^{k+1} - \bar{z}| \leq \frac{a}{c_k^s} |Q_k z^k|^s .$$

Using (2.3) to eliminate $|Q_k z^k|$, and rearranging

$$\forall k \geq k_1 \quad |z^{k+1} - \bar{z}|^2 + \frac{c_k^2}{a^{2/s}} |z^{k+1} - \bar{z}|^{2/s} \leq |z^k - \bar{z}|^2$$

From which we obtain the following inequality for all $n \geq k_1$

$$\sum_{k=k_1}^{n} |z^{k+1} - \bar{z}|^2 + \sum_{k=k_1}^{n} \frac{c_k^2}{a^{2/s}} |z^{k+1} - \bar{z}|^{2/s} \leq \sum_{k=k_1}^{n} |z^k - \bar{z}|^2,$$

which reduces to

$$\forall n \geq k_1 \quad |z^{n+1}-\bar{z}|^2 + \sum_{k=k_1}^{n} \frac{c_k^2}{a^{2/s}} |z^{k+1} - \bar{z}|^{2/s} \leq |z^{k_1} - \bar{z}|^2 .$$

Taking the limit as $n \to \infty$, $|z^{n+1} - \bar{z}| \to 0$ by Proposition 1.2

$$\sum_{k=k_1}^{\infty} \frac{c_k^2}{a^{2/s}} |z^{k+1} - \bar{z}|^{2/s} \leq |z^{k_1} - \bar{z}|^2 < +\infty .$$

For the series to converge, its terms have to decrease to zero faster than the terms of the harmonic series, thus

$$\limsup_{k\to\infty} c_k^2 |z^{k+1} - \bar{z}|^{2/s} k = 0$$

Obviously, any speed of decrease can be obtained by making $c_k \uparrow \infty$ fast enough. If $c_k \uparrow c_\infty < +\infty$ then it follows that $|z^k - \bar{z}| = o(k^{-s/2})$.

<u>Remark</u>. This estimate seems conservative at least when $s \uparrow 1$ because for $s = 1$ linear convergence is achieved and then $|z^k - \bar{z}| = o(k^{-\sigma})$ for all $\sigma > 0$ (Ortega & Rheinboldt 1970).

References

1. Bertsekas, D.P. (1975), Necessary and Sufficient Conditions for a Penalty Method to be Exact. Mathematical Programming, 9, 87-99.

2. Brézis, H. (1973), Operateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert. North-Holland, Amsterdam.

3. Browder, F.E. (1976), Nonlinear Operators and Nonlinear Equations of Evolution in Banach Spaces. Proceedings of Symposia in Pure Mathematics, Vol. XVIII, Pt. 2, AMS, Providence.

4. Kort, B.W. & Bertsekas, D.P. (1973), Multiplier Methods for Convex Programming. Proceedings 1973 IEEE Conference on Decision and Control, San Diego, December 1973, 428-432.

5. Kort, B.W. & Bertsekas, D.P. (1976), Combined Primal-Dual and Penalty Methods for Convex Programming. SIAM J. Control and Optimization, 14, 2, 268-294.

6. Kryanev, A.V. (1973), The Solution of Incorrectly Posed Problems by Methods of Successive Approximations. Soviet Math. Dokl., 14, 673-676.

7. Luenberger, D.G. (1969), Optimization by Vector Space Methods. John Wiley, New York.

8. McLinden, L. (1980), The Complementarity Problem for Maximal Monotone Multifunctions. Variational Inequalities and Complentarity Problems: Theory and Applications, Cottle, R.W., Giannessi, F. & Lions J.-L., eds., Chapter 17, 251-270, John Wiley, Chichester, UK.

9. Minty, G.J. (1962), Monotone (Nonlinear) Operators in Hilbert Space. Duke Math. J., 29, 341-362.

10. Minty, G.J. (1964), On the Solvability of Nonlinear Functional Equations of 'Monotonic" Type. Pac. J. of Math., 14, 249-255.

11. Moreau, J.-J. (1965), Proximité et Dualité dans un Espace Hilbertien Bull. Soc. Math. France, 93, 3, 273-299.

12. Ortega, J.M. & Rheinboldt, W.C. (1970), Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York.

13. Pascali, D. & Sburlan, S. (1978), Nonlinear Mappings of Monotone Type. Sijthoff & Noordhoff International Publishers, Alphen aan den Rijn, Th Netherlands.

14. Rockafellar, R.T. (1970a), Monotone Operators Associated with Saddle Functions and Minimax Problems. Nonlinear Functional Analysis, Browder, F.E., ed., Proceedings of Symposia in Pure Mathematics, Vol. XVIII, pt. 1, 241-250, AMS, Providence.

15. Rockafellar, R.T. (1970b), Convex Analysis. Princeton University Press, Princeton.

16. Rockafellar, R.T. (1970c), On the Maximality of Sums of Nonlinear Monotone Operators. Transactions of the AMS, 149, 75-88.

17. Rockafellar, R.T. (1976a), Monotone Operators and the Proximal Point Algorithm. SIAM J. Control and Optimization, 4, 5, 877-898.

18. Rockafellar, R.T. (1976b), Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming. Mathematics of Operations Research, 1, 2, 97-116.

19. Rockafellar, R.T. (1978), Monotone Operators and Augmented Lagrangian Methods in Nonlinear Programming. Nonlinear Programming 3, Mangasarian, O.L., Meyer R.R. & Robinson, S.M., eds., 1-25, Academic Press, New York.

20. Rockafellar, R.T. (1980), Lagrange Multipliers and Variational Inequalities. Variational Inequalities and Complementarity Problems: Theory and Applications, Cottle, R.W., Giannessi, F. & Lions, J.-L., eds., Chapter 20, 303-322, John Wiley, Chichester, UK.

21. Rockafellar, R.T. (1966), Level Sets and Continuity of Conjugate Convex Functions. Trans. Amer. Math. Soc. 123, 46-63.