Inexact variants of the proximal point algorithm without monotonicity

A. N. Iusem[*] T. Pennanen[†] B. F. Svaiter[‡]

January 4, 2002

Abstract

This paper studies convergence properties of inexact variants of the proximal point algorithm when applied to a certain class of nonmonotone mappings. The presented algorithms allow for constant relative errors, in the line of the recently proposed hybrid proximal-extragradient algorithm. The main convergence result extends a recent work of the second author, where exact solutions for the proximal subproblems were required. We also show that the linear convergence property is preserved in the case when the inverse of the operator is locally Lipschitz continuous near the origin. As an application, we give a convergence analysis for an inexact version of the proximal method of multipliers for a rather general family of problems which includes variational inequalities and constrained optimization problems.

Key words: proximal point algorithms, inexact iterates, hybrid proximal-extragradient algorithms, hypomonotone operators, multiplier methods.

[*]Instituto de Matemática Pura e Aplicada (IMPA), Estrada Dona Castorina 110, Rio de Janeiro, RJ, CEP 22460-320, Brazil (**iusp@impa.br**). The work of this author was partially supported by CNPq grant no. 301280/86.

[†]Helsinki School of Economics and Business Administration, Department of Management Sciences, PL 1210, 00101, Helsinki, Finland (**pennanen@hkkk.fi**)

[‡]Instituto de Matemática Pura e Aplicada (IMPA), Estrada Dona Castorina 110, Rio de Janeiro, RJ, CEP 22460-320, Brazil (**benar@impa.br**).

1. Introduction

We deal in this paper with methods for finding zeroes of point-to-set operators in Hilbert spaces, i.e., given a Hilbert space H and an operator $T : H \to \mathcal{P}(H)$, we intend to find some $x^* \in H$ such that $0 \in T(x^*)$.

The proximal point algorithm, whose origins can be traced back to [9], was born in the 60's (see e.g. [12], [10]) and attained its current formulation in the works of Rockafellar ([14], [15]), where its connection with the augmented Lagrangian method for constrained nonlinear optimization was established. Basically, given a sequence $\{\gamma_n\}$ of positive real numbers bounded away from zero, the algorithm generates a sequence $\{x^n\} \subset H$, starting from some $x^0 \in H$, through the iteration

$$x^{n+1} \in (I + \gamma_n T)^{-1}(x^n). \tag{1}$$

When T is monotone, i.e.

$$\langle x - y, u - v \rangle \geq 0 \tag{2}$$

for all $x, y \in H$, all $u \in T(x)$ and all $v \in T(y)$, and furthermore maximal monotone, i.e. $T = T'$ whenever $T' : H \to \mathcal{P}(H)$ is monotone and $T(x) \subset T'(x)$ for all $x \in H$, it follows from Minty's theorem (see [11]) that $I + \gamma T$ is onto and $(I + \gamma T)^{-1}$ is single valued for all positive $\gamma \in \mathbb{R}$, so that the sequence defined by (1) is well defined. It has been proved in [14] that maximal monotonicity of T also ensures the weak convergence of the sequence $\{x^n\}$ defined by (1) to a zero of T when T has zeroes, and its unboundedness otherwise. Such weak convergence is global, i.e. the result just announced holds in fact for any $x^0 \in H$.

The situation becomes considerably more complicated when T fails to be monotone. A survey of results on convergence of the proximal algorithm without monotonicity up to 1997 can be found in [8]. A new approach to the subject was taken in [13], which deals with a class of nonmonotone operators that, when restricted to a neighborhood of the solution set, are not far from being monotone. More precisely, it was assumed that, for some $\rho > 0$, the mapping $T^{-1} + \rho I$ is monotone when restricted to a neighborhood of $\hat{S}^* \times \{0\}$, where $\hat{S}^*$ is a nonempty connected component of the solution set $S^* = T^{-1}(0)$. When this happens, the main convergence result of [13] states that a "localized" version of (1) generates a sequence that converges to a point in $\hat{S}^*$, provided x^0 is close enough to $\hat{S}^*$ and $\inf \gamma_n > \rho$.

The issue of convergence of the algorithm under inexact computation of the iterates, i.e. when (1), or equivalently the inclusion

$$x^n - x^{n+1} \in \gamma_n T(x^{n+1}), \tag{3}$$

is solved only approximately, comes up immediately when dealing with the proximal algorithm, for at least two reasons. First, it is generally imposible to find an exact value for x^{n+1} given by (1), or (3), particularly when T is nonlinear; secondly, it is clearly inefficient to spend too much effort in the computation of a given iterate x^n, when only the limit of the sequence has the desired properties. Thus, the issue was dealt with even in the early treatment of the subject, e.g. [14], but always, as far as we know, for the case of a monotone T. For instance, it has been proved in [14] that convergence is preserved when an error e^n is committed when performing the iteration given by (3), i.e. when (3) is replaced by

$$e^n + x^n - x^{n+1} \in \gamma_n T(x^{n+1}), \tag{4}$$

as long as

$$\sum_{n=0}^{\infty} \|e^n\| < \infty. \tag{5}$$

Other related conditions, but including always summability of some measure of the error, can be found, e.g. in [14], [5]. These criteria are somewhat undesirable, because they impose increasing precision along the iterative process.

Recently, new related procedures have been presented in [17] and [18] which allow for constant relative error, in the sense, e.g., that the norm of the error e^n in (4) must be smaller than a given fraction of the distance from the current iterate to the previous one. The price to be paid for this less stringent error criterion is that the resulting point (i.e. x^{n+1} in (4)) is not the next iterate, but rather an intermediate point which determines a hyperplane separating x^n from the solution set, and thus a direction pointing from x^n to this set; the actual next iterate is obtained by taking a certain step from x^n in this direction. More precisely, taking y^n as the intermediate point, and defining

$$\gamma_n v^n = e^n + x^n - y^n, \tag{6}$$

inclusion (4) becomes

$$\gamma_n v^n \in T(y^n) \tag{7}$$

and the error criterion is

$$\|e^n\| \leq \sigma \max\{\gamma_n \|v^n\|, \|y^n - x^n\|\}, \tag{8}$$

with $\sigma \in [0,1)$. Indeed, the vector v^n gives the desired direction, so that

$$x^{n+1} = x^n - \eta_n v^n \tag{9}$$

for some appropriate $\eta_n > 0$ (e.g. $\eta_n = \langle v^n, x^n - y^n \rangle / [\|v^n\|^2])$. It is important to emphasize that the additional cost of computing x^{n+1} once v^n and y^n have been determined (i.e. the cost of (9)) is negligeable as compared to the solution, even in an inexact way, of the inclusion (4) (or the pair

(6)–(7)). Algorithms of this kind have been called "hybrid" due to the presence of the additional step (9), in addition to the proximal step given by (6)–(7). We also remark that (8) can be seen as a sort of stopping criterion in an internal iterative procedure for the solution of (3): given candidate points (y^n, v^n) computed by such a procedure, if e^n as given by (6) satisfies (8) then v^n is accepted and x^{n+1} is computed according to (9); otherwise the procedure must continue, generating a new pair. In this sense, we can say that error criteria like (8) are particularly appropriate for computer implementation. In [17] it has been proved that the sequence generated by (6)–(9) is globally convergent to a zero of T in the weak topology, under the only assumptions of existence of zeroes and monotonicity of T (besides boundedness away from 0 of $\{\gamma_n\}$). Other related error criteria for the proximal point algorithm, allowing also for constant relative error, can be found in [19] ,[3] and [4].

In this paper we will consider the following inexact procedure for finding zeroes of an operator $T : H \to \mathcal{P}(H)$ whose inverse is maximal ρ-hypomonotone on a set $U \times V \subset H \times H$ (see Definition 1 below). Given $x^n \in H$, find $(y^n, v^n) \in U \times V$ such that

$$v^n \in T(y^n), \tag{10}$$

$$\gamma_n v^n + y^n - x^n = e^n, \tag{11}$$

where the error term e^n satisfies either

$$\|e^n\| \leq \sigma \left(\frac{\hat{\gamma}}{2} - \rho \right) \|v^n\|, \tag{12}$$

or

$$\|e^n\| \leq \nu \|y^n - x^n\|, \tag{13}$$

with

$$\nu = \frac{\sqrt{\sigma + (1-\sigma)\left(2\rho/\hat{\gamma}\right)^2} - 2\rho/\hat{\gamma}}{1 + 2\rho/\hat{\gamma}}, \tag{14}$$

where $\sigma \in [0, 1)$, $\hat{\gamma} = \inf\{\gamma_n\}$, and ρ is the hypomonotonicity constant of T^{-1}. Then, under any of our two error criteria, the next iterate x^{n+1} is given by

$$x^{n+1} = x^n - \gamma_n v_n. \tag{15}$$

From now on, Algorithm 1 refers to the algorithm given by (10)–(12) and (15), and Algorithm 2 to the one given by (10), (11) and (13)–(15). We will prove that, when ρ-hypomonotonicity of T^{-1} holds on the whole space (i.e. $U = V = H$), both Algorithm 1 and Algorithm 2 generate sequences which are weakly convergent to a zero of T, starting from any $x^0 \in H$, under the assumptions of existence of zeroes of T and $2\rho < \hat{\gamma} = \inf\{\gamma_n\}$. For the case in which the set $U \times V$ where T^{-1} is ρ-hypomonotone is an appropriate neighborhood of $\hat{S}^* \times \{0\} \subset H \times H$, where $\hat{S}^*$ is a connected component of the set S^* of zeroes of T, we still get a local convergence result, meaning weak convergence of $\{x^n\}$ to a zero of T, but requiring additionally that x^0 be sufficiently close to $\hat{S}^* \cap U$, in a sense which is presented in a precise way in section 4.

We remark that when the tolerance σ vanishes, we get $e^n = 0$ from either (12) or (13)–(14), and then $x^n - y^n = \gamma_n v_n$ from (11), so that $x^{n+1} = y^n$ from (15). Thus, with $\sigma = 0$ our algorithm reduces to the exact algorithm in [13]. When comparing our analysis for this exact case with those in [13] it is worthwhile to point out the following difference: in [13] the proposed algorithm is studied by assuming first that T is locally monotone, and convergence is proved by showing that the resulting sequence coincides with the one generated by the algorithm applied to some maximal monotone operator. Then, the algorithm applied to a mapping whose inverse is locally ρ-hypomonotone is showed to be equivalent to an overrelaxed version of the proximal algorithm applied to the locally monotone operator $(T^{-1} + \rho I)^{-1}$, with a different sequence of regularization parameters, and convergence is finally obtained by invoking results in [5] on the convergence of such over-relaxed variant of the proximal point algorithm. Our approach is less convoluted: we prove directly the Fejér monotonicity properties of $\{x^n\}$, which have as a consequence the weak convergence of $\{x^n\}$ to a zero of T. The issue of over-relaxation of the proximal point applied to the Yoshida regularization $(T^{-1} + \rho I)^{-1}$ of T is confined to a lemma, also proved from scratch (up to an invocation of Minty's theorem), on the issue of existence of the iterates. Thus, our proof is (almost) self-contained and, in the exact case, it can be seen as a streamlined version of the analysis in [13].

2. Hypomonotone operators

From now on we will identify, in a set theoretic fashion, a point-to-set operator $T : H \to \mathcal{P}(H)$ with its graph, i.e. with $\{(x, v) \in H \times H : v \in T(x)\}$. Thus, $(x, v) \in T$ has the same meaning as $v \in T(x)$.

Definition 1. *Given a positive $\rho \in \mathbb{R}$ and a subset W of $H \times H$, an operator $T : H \to \mathcal{P}(H)$ is said to be*

(a) ρ-hypomonotone if and only if $\langle x - y, u - v\rangle \geq -\rho \|x - y\|^2$ for all $(x, u), (y, v) \in T$,

(b) maximal ρ-hypomonotone if and only if T is ρ-hypomonotone and additionally $T = T'$ whenever $T' \subset H \times H$ is ρ-hypomonotone and $T \subset T'$.

(c) ρ-hypomonotone in W if and only if $T \cap W$ is ρ-hypomonotone.

(d) maximal ρ-hypomonotone in W if and only if T is ρ-hypomonotone in W and aditionally $T \cap W = T' \cap W$ whenever $T' \in H \times H$ is ρ-hypomonotone and $T \cap W \subset T' \cap W$.

It follows from 13.33 and 13.36 of [16] that if a function $f : H \to \mathbb{R} \cup \{\infty\}$ can be written as $g - h$ in a neighborhood of a point $x \in H$, where g is finite and h is $\mathcal{C}^2$, then the subdifferential ∂f of f is ρ-hypomonotone for some $\rho > 0$ in a neighborhood of any point $(x, v) \in H \times H$ with $v \in \partial f(x)$. It is also easy to check that a locally Lipschitz continuous mapping is hypomonotone for every ρ greater than the Lipschitz constant. In particular, if H is finite dimensional and $T : H \to \mathcal{P}(H)$ is such that T^{-1} is point-to-point and differentiable in a neighborhood of some $v \in H$, then T is ρ-hypomonotone in a neighborhood of (x, v) for any x such that $v \in T(x)$, and for any ρ larger than the absolute value of the most negative eigenvalue of $J + J^t$, where J is the Jacobian matrix of T^{-1} at v. In other words, local ρ-hypomonotonicity for some $\rho > 0$ is to be expected of any T which is not too badly behaved.

Note that a mapping T is ρ-hypomonotone if and only if $T + \rho I$ is monotone. We also have the following.

Proposition 1. *If $T : H \to \mathcal{P}(H)$ is ρ-hypomonotone, then there exists a maximal ρ-hypomonotone $\hat{T} : H \to \mathcal{P}(H)$ such that $T \subset \hat{T}$.*

Proof. The proof is a routine application of Zorn's Lemma, with exactly the same argument as the one used to prove that any monotone operator is contained in a maximal monotone one. □

Next we introduce in a slightly different way the Yosida regularization of an operator. For $\rho > 0$, define $Y_\rho : H \times H \to H \times H$ (Y for Yosida) as

$$Y_\rho(x, v) = (x + \rho v, v). \tag{16}$$

Observe that Y_ρ is a bijection, and $(Y_\rho)^{-1}(y, u) = (y - \rho u, u)$.

Proposition 2. *Take $\rho > 0$, $T : H \to \mathcal{P}(H)$ and Y_ρ as in (16). Then*

(i) T^{-1} is ρ-hypomonotone if and only if $Y_\rho(T)$ is monotone,

(ii) T^{-1} is maximal ρ-hypomonotone if and only if $Y_\rho(T)$ is maximal monotone.

Proof. (i) Monotonicity of the Yosida regularization means that $(T^{-1} + \rho I)^{-1}$ is monotone, which is equivalent to monotonicity of $T^{-1} + \rho I$.

(ii) Assume that T^{-1} is maximal ρ-hypomonotone. We prove maximal monotonicity ot $Y_\rho(T)$. The monotonicity follows from item (a). Assume that $Y_\rho(T) \subset Q$ for some monotone $Q \subset H \times H$. Note that $Q = Y_\rho(T')$ for T' defined as $T'(x, v) = (x - \rho v, v)$. It follows easily from Definition 1(a) and the monotonicity of Q that $(T')^{-1}$ is ρ-hypomonotone. By Definition 1(b), $T = T'$, so that $Q = Y_\rho(T') = Y_\rho(T)$, proving that $Y_\rho(T)$ is maximal monotone. The converse statement is proved with a similar argument.

□

We continue with an elementary result on the Yosida regularization $Y_\rho(T)$.

Proposition 3. *For all $T : H \to \mathcal{P}(H)$ and all $\rho > 0$, $0 \in T(x)$ if and only if $0 \in [Y_\rho(T)](x)$.*

Proof. The result follows immediately from (16). $\square$

Remark 1. It is well known that the set of zeroes of a monotone operator is closed and convex. In view of Propositions 2 and 3, the same holds for mappings whose inverses are ρ-hypomonotone. Thus, though reasonably well behaved operators can be expected to be locally ρ-hypomonotone for some $\rho > 0$, as discussed above, global ρ-hypomonotonicity is not at all generic; looking for instance at point-to-point operators in $\mathbb{R}$, we observe that polynomials with more than one real root, or analytic functions like $T(x) = \sin x$, are not ρ-hypomonotone for any $\rho > 0$.

Next we establish local demi-closedness of maximal locally ρ-hypomonotone operators, with a proof which mirrors the one on demi-closedness of maximal monotone operators.

Proposition 4. *Assume that $T^{-1} : H \to \mathcal{P}(H)$ is maximal ρ-hypomonotone in $W \subset H \times H$, and consider a sequence $\{(x^n, v^n)\} \subset T \cap W$. If $\{v^n\}$ is strongly convergent to $\bar{v}$, $\{x^n\}$ is weakly convergent to $\bar{x}$, and $(\bar{x}, \bar{v}) \in W$, then $\bar{v} \in T(\bar{x})$.*

Proof. Define $T' : H \to \mathcal{P}(H)$ as $T' = T \cup \{(\bar{x}, \bar{v})\}$. We claim that $(T')^{-1}$ is ρ-hypomonotone in W. Since T^{-1} is ρ-hypomonotone in W, clearly it suffices to prove that

$$-\rho \|\bar{v} - v\|^2 \leq \langle \bar{x} - x, \bar{v} - v \rangle \tag{17}$$

for all $(x, v) \in T \cap W$. Observe that, for all $(x, v) \in T \cap W$,

$$-\rho \|v^n - v\|^2 \leq \langle x^n - x, v^n - v \rangle = \langle x^n - x, v^n - \bar{v} \rangle + \langle x^n - x, \bar{v} - v \rangle \leq$$

$$\|x^n - x\| \, \|v^n - \bar{v}\| + \langle x^n - x, \bar{v} - v \rangle. \tag{18}$$

Now we take limits with $n \to \infty$ in the extreme expressions of (18). The leftmost one converges to the right hand side of (17) because $\{v^n\}$ is strongly convergent to $\bar{v}$, and the second term in the rightmost one converges to the right hand side of (17) because $\{x^n\}$ is weakly convergent to $\bar{x}$. Finally, the first term in the rightmost expression of (18) converges to zero, because $\lim_{n\to\infty} \|v^n - \bar{v}\| = 0$ by strong convergence of $\{v^n\}$ to $\bar{v}$, and $\{\|x^n - x\|\}$ is bounded, since $\{x^n\}$ is weakly convergent (see Proposition III.12(iii) in [2]). It follows that (17) holds and the claim is established. Since $T \subset T'$, $(T')^{-1}$ is ρ-hypomonotone in W and T^{-1} is maximal ρ-hypomonotone in W, we have that $T \cap W = T' \cap W$, by Definition 1(d). Since $\bar{v} \in T'(\bar{x})$ and $(\bar{x}, \bar{v}) \in W$, we conclude that $\bar{v} \in T(\bar{x})$. $\square$

We close this section with a result on convexity and weak closedness of some sets related to the set of zeroes of operators whose inverses are ρ-hypomonotone. We use the usual notation for sums of sets, i.e., for $A, B \subset H$, $A + B \subset H$ is defined as $A + B = \{x + y : x \in A, y \in B\}$. Also, for $x \in H$ and $\delta > 0$, $B(x, \delta)$ will denote the closed ball of radius δ centered at x.

Proposition 5. *Assume that $T^{-1} : H \to \mathcal{P}(H)$ is maximal ρ-hypomonotone in a subset $U \times V \subset H \times H$, where U is convex and $0 \in V$. Let $S^* \subset H$ be the set of zeroes of T. Then*

(i) $S^* \cap U$ is convex.

(ii) If $S^* \cap U$ is closed, then $(S^* \cap U) + B(0, \delta)$ is weakly closed for all $\delta \geq 0$.

Proof. (i) By Proposition 1, $T \cap (U \times V) \subset \hat{T}$ for some $\hat{T} : H \to \mathcal{P}(H)$ such that $\hat{T}^{-1}$ is maximal ρ-hypomonotone . Let $\hat{S}^*$ be the set of zeroes of $\hat{T}$. By Proposition 3, $\hat{S}^*$ is also the set of zeroes of $Y_\rho(\hat{T})$, which is maximal monotone by Proposition 2(ii). Since the set of zeroes of a maximal monotone operator is convex (e.g. 12.8(a) and (c) in [16]), we conclude that $\hat{S}^*$ is convex, and therefore, $\hat{S}^* \cap U$ is convex, because U is convex. Since T^{-1} is maximal ρ-hypomonotone in $U \times V$, $\hat{T}^{-1}$ is ρ-hypomonotone and $T \subset T'$, we have that $\hat{T} \cap (U \times V) = T \cap (U \times V)$, and then, since $0 \in V$, it follows easily that $\hat{S}^* \cap U = S^* \cap U$. The result follows.

(ii) First we claim that if $A, B \subset H$ are weakly closed and B is bounded then $A + B$ is weakly closed. It suffices to prove that the weak limit $\bar{x}$ of a sequence $\{x^n\} \subset A + B$ belongs to $A + B$. Since $\{x^n\}$ is weakly convergent, it is bounded. Writing $x^n = a^n + b^n$, with $a^n \in A$, $b^n \in B$, we get that $\{b^n\}$ is bounded by boundedness of B, so that $\{a^n\}$ is also bounded. In view of Bourbaki-Alaoglu's Theorem (see e.g. Theorem III.8 in [2]) there exists a subsequence $\{(a^{k_n}, b^{k_n})\}$ of $\{(a^n, b^n)\}$, which is weakly convergent, say to (a, b), which belongs to $A \times B$ because A and B are weakly closed. Thus, by uniqueness of the weak limit, $\bar{x} = a + b \in A + B$. The claim is established, and we apply the result with $A = S^* \cap U$, $B = B(0, \delta)$. A is closed by assumption and convex by item (i), and B is obviously closed and convex. Hence, both A and B are weakly closed (see Theorem III.7 in [2]). Since B is bounded, the result follows. □

3. Existence results

The issue of existence of iterates for proximal algorithms applied to nonmonotone operators is delicate. The main tool used for establishing existence in the monotone case, namely Minty's Theorem, does not work iwithout monotonicity. Overcoming this obstacle requires some technicalities, where the notion of ρ-hypomonotonicity becomes crucial.

Note that if a pair (y^n, v^n) satisfies (10)–(12), or (10)–(11) together with (13)–(14), with $\sigma = 0$ (which in the second case implies $\nu = 0$), then such pair satisfies those conditions with any $\sigma > 0$. Since $\sigma = 0$ also implies that the error term e^n vanishes, existence of exact iterates is enough to settle the existence issue for our inexact schemes. Now, we already mentioned that our scheme reduces, in the absence of errors, to the algorithm studied in [13], and therefore we could refer to the existence results in this reference without further discussion. But since we dressed our setting somewhat differently from the one in [13] (e.g. the definition of local ρ-hypomonotonicity), we prefer to offer a full proof, which also contributes to make this paper more self-contained. The technicalities will be encapsulated in the following lemma, where, for $x \in H$ and $A \subset H$, $d(x, A)$ will denote the distance from x to A, i.e. $d(x, A) = \inf_{y \in A} \|x - y\|$.

Lemma 1. *Let $T : H \to \mathcal{P}(H)$ be an operator such that T^{-1} is maximal ρ-hypomonotone in a subset $U \times V$ of $H \times H$. Assume that T has a nonempty set of zeroes S^*, and that*

(i) $S^* \cap U$ is nonempty and closed,

(ii) there exists $\beta > 0$ such that $B(0, \beta) \subset V$,

(iii) there exists $\delta > 0$ such that $(S^* \cap U) + B(0, \delta) \subset U$.

Take any $\gamma > 2\rho$ and a define $\varepsilon = \min\{\delta, \beta\gamma/2\}$. If $x \in H$ is such that $d(x, S^* \cap U) \leq \varepsilon$, then there exists $y \in H$ such that $\gamma^{-1}(x - y) \in T(y)$ and $d(y, S^* \cap U) \leq \varepsilon$.

Proof. By Definition 1(c) and (d), $T \cap (U \times V)$ is ρ-hypomonotone. By Proposition 1, there exists a maximal ρ-hypomonotone $\hat{T} \subset H \times H$ such that

$$[T \cap (U \times V)] \subset \hat{T}. \tag{19}$$

By Proposition 2(ii), $Y_\rho(\hat{T})$ is maximal monotone, with Y_ρ as defined in (16). Let $\hat{\gamma} = \gamma - \rho$. Since $\hat{\gamma} > 0$ by assumption, it follows from Minty's Theorem (see [11]) that the operator $\left[I + \hat{\gamma} Y_\rho(\hat{T})\right]^{-1}$ is onto (and also one-to-one, but this does not concern us), so that there exists $z \in H$ such that $x \in \left[I + \hat{\gamma} Y_\rho(\hat{T})\right]^{-1}(z)$, or equivalently

$$\hat{\gamma}^{-1}(x - z) \in \left[Y_\rho(\hat{T})\right](z). \tag{20}$$

Letting

$$v := \hat{\gamma}^{-1}(x - z), \tag{21}$$

we can rewrite (20) as $(z, v) \in Y_\rho(\hat{T})$, which is equivalent, in view of (16), to

$$(z - \rho v, v) \in \hat{T}. \tag{22}$$

Let now $y = z - \rho v$. In view of (21) and the definition of $\hat{\gamma}$, (22) is in turn equivalent to

$$(y, \hat{\gamma}^{-1}(x - z)) \in \hat{T}. \tag{23}$$

It follows easily from (21) and the definitions of y and $\hat{\gamma}$ that

$$\hat{\gamma}^{-1}(x - z) = (\gamma - \rho)^{-1}(x - z) = \gamma^{-1}(x - y) = \rho^{-1}(z - y). \tag{24}$$

We conclude from (23) and (24) that

$$\gamma^{-1}(x-y) \in \hat{T}(y). \tag{25}$$

Note that (25) looks pretty much like the statement of the lemma, excepting that we have $\hat{T}$ instead of T. The operators T and $\hat{T}$ do coincide on $U \times V$, as we will see, but in order to use this fact we must first establish that $(y, \gamma^{-1}(x-y))$ belongs indeed to $U \times V$, which will result from the assumption on $d(x, S^* \cap U)$. The analysis in the following paragraph is tantamount to establishing Fejér monotonicity of the iterates of an over-relaxed proximal algorithm applied to a maximal monotone operator, which can be found in [5].

Take any $\bar{x} \in S^* \cap U$, nonempty by condition (i), and z as in (20). Note that $\bar{x}$ is a zero of $T \cap (U \times V)$, because it belongs to $S^* \cap U$ and $0 \in V$ by condition (ii). Thus $\bar{x}$ is a zero of $\hat{T}$, which contains $T \cap (U \times V)$. By Proposition 3, $\bar{x}$ is a zero of $Y_\rho(T)$. Then

$$\|x-\bar{x}\|^2 = \|x-z\|^2 + \|z-\bar{x}\|^2 + 2\langle x-z, z-\bar{x}\rangle =$$

$$\|x-z\|^2 + \|z-\bar{x}\|^2 + 2\hat{\gamma}\langle \hat{\gamma}^{-1}(x-z) - 0, z-\bar{x}\rangle \geq \|x-z\|^2 + \|z-\bar{x}\|^2, \tag{26}$$

using (20), monotonicity of $Y_\rho(\hat{T})$, nonnegativity of $\hat{\gamma}$ and the fact that $\bar{x}$ is a zero of $Y_\rho(\hat{T})$ in the inequality. Take now y as defined after (22). Then

$$\|y-\bar{x}\|^2 = \|y-z\|^2 + \|z-\bar{x}\|^2 - 2\langle y-z, \bar{x}-z\rangle =$$

$$\left(\frac{\rho}{\gamma-\rho}\right)^2 \|z-x\|^2 + \|z-\bar{x}\|^2 - \frac{2\rho}{\gamma-\rho}\langle z-x, \bar{x}-z\rangle \leq$$

$$\|x-\bar{x}\|^2 - \left[1 - \left(\frac{\rho}{\gamma-\rho}\right)^2\right] \|z-x\|^2 - 2\rho\langle 0 - (\gamma-\rho)^{-1}(x-z), \bar{x}-z\rangle \leq$$

$$\|x-\bar{x}\|^2 - \left[1 - \left(\frac{\rho}{\gamma-\rho}\right)^2\right] \|z-x\|^2 = \|x-\bar{x}\|^2 - \frac{\gamma(\gamma-2\rho)}{(\gamma-\rho)^2}\|z-x\|^2 \leq \|x-\bar{x}\|^2, \tag{27}$$

using (24) in the first equality, (26) in the first inequality, (20) and monotonicity of $Y_\rho(\hat{T})$ in the second inequality, and the assumption that $\gamma > 2\rho$ in the third inequality. It follows from (27) that $\|y-\bar{x}\| \leq \|x-\bar{x}\|$ for all $\bar{x} \in S^* \cap U$, in particular when $\bar{x}$ is the orthogonal projection of x onto $S^* \cap U$, which exists because $S^* \cap U$ is closed by condition (i). For this choice of $\bar{x}$ we have that

$$\|y-\bar{x}\| \leq \|x-\bar{x}\| = d(x, S^* \cap U) \leq \varepsilon = \min\{\delta, \beta\gamma/2\} \leq \delta, \tag{28}$$

where the second inequality holds by then assumption on x. Since $\bar{x}$ belongs to $S^* \cap U$, we get from (28) that

$$y \in (S^* \cap U) + B(0, \delta) \subset U, \tag{29}$$

using condition (iii) in the inclusion.

Observe now that, with the same choice of $\bar{x}$,

$$\gamma^{-1} \|x - y\| \leq \gamma^{-1}(\|x - \bar{x}\| + \|y - \bar{x}\|) \leq 2\varepsilon\gamma^{-1} \leq \beta, \tag{30}$$

using (28) and the assumption on x in the second inequality, and the fact that $\varepsilon = \min\{\delta, \beta\gamma/2\}$ in the third one. It follows from (29), (30) and condition (ii) that

$$(y, \gamma^{-1}(x - y)) \in U \times V. \tag{31}$$

Since T is maximal ρ-hypomonotone in $U \times V$ and $\hat{T}$ is ρ-hypomonotone, it follows from (19) and Definition 1(d) that $T \cap (U \times V) = \hat{T} \cap (U \times V)$. In view of (25), we conclude from (31) that $\gamma^{-1}(x - y) \in T(y)$. Finally, using (28), $d(y, S^* \cap U) \leq \|y - \bar{x}\| \leq \varepsilon$, completing the proof. □

Remark 2. As we mentioned along the proof of Lemma 1, the vector z satisfying (20) is unique by Minty's Theorem, and thus it is easy to check that the vector y in the statement of the lemma is also unique. This is not too relevant for our inexact algorithms: the iterates are unique for $\sigma = 0$ (i.e. in the exact case) but hopefully not so for other values of σ. We emphasize that uniqueness of the iterates is no blessing for inexact algorithms; it is rather pretty catastrophic.

Corollary 1. *Consider either Algorithm 1 or Algorithm 2 applied to an operator $T : H \to \mathcal{P}(H)$ which is ρ-hypomonotone on a subset $U \times V$ of $H \times H$ satisfying conditions (i)–(iii) of Lemma 1. If $d(x^n, S^* \cap U) \leq \varepsilon$, with ε as in the statement of Lemma 1, and $\gamma_n > 2\rho$, then there exists a pair $(y^n, v^n) \in U \times V$ satisfying* (10) *and* (11)*, and consequently a vector x^{n+1} satisfying* (15)*.*

Proof. Apply Lemma 1 with $x = x^n$, $\gamma = \gamma_n$. Take y^n as the vector y whose existence is ensured by the lemma and $v^n = \gamma_n^{-1}(x^n - y^n)$. Then y^n and v^n satisfy (10) and (11) with $e^n = 0$, so that (12) or (13)–(14) hold for any $\sigma \geq 0$. Once a pair (y^n, v^n) exists, the conclusion about x^{n+1} is obvious, since (15) raises no existence issues. □

In order to ensure existence of the iterates, we still have to prove, in view of Corollary 1, that the whole sequence $\{x^n\}$ is contained in $B(\bar{x}, \varepsilon)$, where $\bar{x}$ is the orthogonal projection of x^0 onto $S^* \cap U$ and ε is as in Lemma 1. This will be a consequence of the Fejér monotonicity properties of $\{x^n\}$, which we will establish in the following section.

4. Convergence analysis

The next lemma establishes the Fejér monotonicity property of sequences generated by our inexact algorithm. We have not proved yet existence of such sequences, but the lemma is phrased so as to circumvent the existencial issue, for the time being.

Lemma 2. *Let $\{x^n\} \subset H$ be a sequence generated by either Algorithm 1 or Algorithm 2 applied to an operator $T : H \to \mathcal{P}(H)$ which is ρ-hypomonotone in a subset W of $H \times H$, and take x^* in the set S^* of zeroes of T. If $2\rho < \hat{\gamma} = \inf\{\gamma_n\}$ and both $(x^*, 0)$ and (y^n, v^n) belong to W, then*

(i)

$$\left\|x^{n+1} - x^*\right\|^2 \leq \|x^n - x^*\|^2 - (1-\sigma)\gamma_n(\hat{\gamma} - 2\rho)\|v^n\|^2$$

for Algorithm 1, and

(ii)

$$\left\|x^* - x^{n+1}\right\|^2 \leq \|x^* - x^n\|^2 - (1-\sigma)\left(1 - \frac{2\rho}{\hat{\gamma}}\right)\|y^n - x^n\|^2,$$

for Algorithm 2.

Proof. We start with the following elementary algebraic equality:

$$\|x^* - x^n\|^2 - \left\|x^* - x^{n+1}\right\|^2 - \|y^n - x^n\|^2 + \left\|y^n - x^{n+1}\right\|^2 = 2\langle x^* - y^n, x^{n+1} - x^n\rangle. \tag{32}$$

Using first (15) in the right hand side of (32), and then ρ-hypomonotonicity of T in W, together with the fact that both $(x^*, 0)$ and (y^n, v^n) belong to $T \cap W$, by (10) and the assumptions of the lemma, we get

$$\|x^* - x^n\|^2 - \left\|x^* - x^{n+1}\right\|^2 - \|y^n - x^n\|^2 + \left\|y^n - x^{n+1}\right\|^2 =$$

$$-2\gamma_n\langle x^* - y^n, 0 - v^n\rangle \geq -2\rho\gamma_n\|v^n\|^2. \tag{33}$$

¿From this point the computation differ according to the error criterion. We start with the one given by (12). It follows from (11) and (15) that $y^n - x^n = e^n - \gamma_n v^n$ and $y^n - x^{n+1} = e^n$. Substituting these two equalities in (33) we get

$$\|x^* - x^n\|^2 - \left\|x^* - x^{n+1}\right\|^2 \geq \gamma_n^2\|v^n\|^2 - 2\gamma_n\langle v^n, e^n\rangle - 2\rho\gamma_n\|v^n\|^2 \geq$$

$$\gamma_n^2\|v^n\|^2 - 2\gamma_n\|v^n\|\,\|e^n\| - 2\rho\gamma_n\|v^n\|^2 = \gamma_n\|v^n\|\left[(\gamma_n - 2\rho)\|v^n\| - 2\|e^n\|\right] \geq$$

$$\gamma_n\|v^n\|\left[(\gamma_n - 2\rho)\|v^n\| - \sigma(\hat{\gamma} - 2\rho)\|v^n\|\right] \geq \gamma_n\|v^n\|\left[(1-\sigma)(\hat{\gamma} - 2\rho)\|v^n\|\right] =$$

$$(1-\sigma)\gamma_n(\hat{\gamma} - 2\rho)\|v^n\|^2, \tag{34}$$

using (12) in the second inequality and the definition of $\hat{\gamma}$ in the third inequality. The results follows immediately from (34).

Now we look at the error criterion given by (13)–(14). Using again (11) and (15), we can replace $y^n - x^{n+1}$ by e^n and v^n by $\gamma_n^{-1}(y^n - x^n - e^n)$ in (33), obtaining

$$\|x^* - x^n\|^2 - \left\|x^* - x^{n+1}\right\|^2 \geq \|y^n - x^n\|^2 - \left(\|e^n\|^2 + 2\rho\gamma_n^{-1}\|y^n - x^n - e^n\|^2\right) \geq$$

$$\|y^n - x^n\|^2 - \left[\|e^n\|^2 + 2\rho\gamma_n^{-1}\left(\|y^n - x^n\| + \|e^n\|^2\right)\right]. \tag{35}$$

Using now (13) in (35) we get

$$\|x^* - x^n\|^2 - \left\|x^* - x^{n+1}\right\|^2 \geq \left[1 - \nu^2 - \frac{2\rho}{\gamma_n}(1+\nu^2)\right]\|y^n - x^n\|^2 \geq$$

$$\left[1 - \nu^2 - \frac{2\rho}{\hat{\gamma}}(1+\nu^2)\right]\|y^n - x^n\|^2. \tag{36}$$

It follows from (14), after some elementary algebra, that

$$\left[1 - \nu^2 - \frac{2\rho}{\hat{\gamma}}(1+\nu^2)\right] = (1-\sigma)\left(1 - \frac{2\rho}{\hat{\gamma}}\right). \tag{37}$$

Replacing (37) in (36), we obtain

$$\|x^* - x^n\|^2 - \left\|x^* - x^{n+1}\right\|^2 \geq (1-\sigma)\left(1 - \frac{2\rho}{\hat{\gamma}}\right)\|y^n - x^n\|^2, \tag{38}$$

and the results follows immediately from (38) □

Next we combine the results of Lemmas 1 and 2 in order to obtain our convergence theorem.

Theorem 1. *Let $T : H \to \mathcal{P}(H)$ be maximal ρ-hypomonotone in a subset $U \times V$ of $H \times H$ satisfying*

(i) $S^* \cap U$ is nonempty and closed,

(ii) there exists $\beta > 0$ such that $B(0,\beta) \subset V$,

(iii) there exists $\delta > 0$ such that $(S^* \cap U) + B(0,\delta) \subset U$,

(iv) U is convex,

where S^* is the set of zeroes of T. Take a sequence $\{\gamma_n\}$ of positive real numbers such that $2\rho < \hat{\gamma} = \inf\{\gamma_n\}$. Define $\varepsilon = \min\{\delta, \beta\hat{\gamma}/2\}$. If $d(x^0, S^* \cap U) \leq \varepsilon$, then, both for Algorithm 1 and Algorithm 2,

(a) for all n there exist $y^n, v^n, e^n, x^{n+1} \in H$ satisfying (10)–(12) and (15), in the case of Algorithm 1, and (10)–(11) and (13)–(15), in the case of Algorithm 2, and such that $(y^n, v^n) \in U \times V$, $d(x^{n+1}, S^* \cap U) \leq \varepsilon$,

(b) For any sequence as in (a), we have that (x^n) converges weakly to a point in $S^* \cap U$.

Proof. (a) We proceed by induction. Take any $n \geq 0$. We have that

$$d(x^n, S^* \cap U) \leq \varepsilon, \tag{39}$$

by inductive hypothesis, if $n \geq 1$, and by assumption, if $n = 0$. We are within the hypotheses of Corollary 1, which indicates that the desired vectors exist and that $(y^n, v^n) \in U \times V$. It remains to establish that $d(x^{n+1}, S^* \cap U) \leq \varepsilon$. Let $\bar{x}$ be the orthogonal projection of x^n onto $S^* \cap U$, which exists by condition (i). Note that both $(\bar{x}, 0)$ and (y^n, v^n) belong to $U \times V$. Thus we are within the hypotheses of Lemma 2, with $W = U \times V$, and both for Algorithm 1 and Algorithm 2 we get from either Lemma 2(i) or Lemma 2(ii) that

$$\left\|x^* - x^{n+1}\right\| \leq \|x^* - x^n\| \tag{40}$$

for all $x^* \in S^* \cap U$. By (40) with $\bar{x}$ instead of x^*,

$$d(x^{n+1}, S^* \cap U) \leq \left\|\bar{x} - x^{n+1}\right\| \leq \|\bar{x} - x^n\| = d(x^n, S^* \cap U) \leq \varepsilon,$$

using (39) in the last inequality.

(b) We follow here with minor variations the standard convergence proof for the proximal point algorithm, see e.g. [14]. In view of (40), for all $x^* \in S^* \cap U$ the sequence $\{\|x^n - x^*\|\}$ is non-increasing, and certainly nonnegative, hence convergent. Also, since $\|x^n - x^*\| \leq \left\|x^0 - x^*\right\|$ for all n, we get that $\{x^n\}$ is bounded.

Now we consider separately both algorithms. In the case of Algorithm 1, we get from Lemma 2(i),

$$(1-\sigma)(\hat{\gamma} - 2\rho)\gamma_n \|v^n\|^2 \leq \|x^n - x^*\|^2 - \left\|x^{n+1} - x^*\right\|^2, \tag{41}$$

Since the right hand side of (41) converges to 0, we conclude that $\lim_{n\to\infty} \gamma_n \|v^n\| = 0$, and therefore, since $\gamma_n \geq \hat{\gamma} > 0$ for all n

$$\lim_{n\to\infty} v^n = 0, \tag{42}$$

which implies, in view of (12), that $\lim_{n\to\infty} e^n = 0$, and therefore, by (11),

$$\lim_{n\to\infty} (y^n - x^n) = 0. \tag{43}$$

In the case of Algorithm 2, we get from Lemma 2(ii),

$$(1-\sigma)\left(1 - \frac{2\rho}{\hat{\gamma}}\right) \|y^n - x^n\|^2 \leq \|x^* - x^n\|^2 - \left\|x^* - x^{n+1}\right\|^2, \tag{44}$$

Again, the right hand side of (44) converges to 0, and thus (43) also holds in this case, so that, in view of (13), $\lim_{n\to\infty} e^n = 0$, which gives, in view of (43) and (11), $\lim_{n\to\infty} \gamma_n v^n = 0$, so that in this case we also have (42). We have proved that (42) and (43) hold both for Algorithm 1 and Algorithm 2, and we proceed from now on with an argument which holds for both algorithms.

Since $\{x^n\}$ is bounded, it has weak cluster points. Let $\tilde{x}$ be any weak cluster point of $\{x^n\}$, i.e. $\tilde{x}$ is the weak limit of a subsequence $\{x^{k_n}\}$ of $\{x^n\}$. By (43), $\tilde{x}$ is also the weak limit of $\{y^{k_n}\}$. We claim that $(\tilde{x}, 0)$ belongs to $U \times V$. In view of condition (ii), it suffices to check that $\tilde{x} \in U$. Note that $\{x^n\} \subset (S^* \cap U) + B(0, \varepsilon)$ by item (a). Since U is convex by condition (iv) and $S^* \cap U$ is closed by condition (i), we can apply Proposition 5(ii) to conclude that $(S^* \cap U) + B(0, \varepsilon)$ is weakly closed. Thus, the weak limit $\tilde{x}$ of $\{x^{n_k}\}$ belongs to $(S^* \cap U) + B(0, \varepsilon)$, and henceforth to U, in view of condition (ii) and the fact that $\varepsilon \leq \delta$. The claim holds, and we are within the hypotheses of Proposition 4: $\{v^{k_n}\}$ is strongly convergent to 0 by (42), $\{x^{k_n}\}$ is weakly convergent to $\tilde{x}$ and $(\tilde{x}, 0)$ belongs to $U \times V$, where T is maximal ρ-hypomonotone. Then $0 \in T(\tilde{x})$, i.e. $\tilde{x} \in S^* \cap U$.

Finally we establish uniqueness of the weak cluster point of $\{x^n\}$, with the standard argument (e.g. [14]) which we include just in order to keep our self-containment policy. Let $\tilde{x}$, $\hat{x}$ be two weak cluster points of $\{x^n\}$, say the weak limits of $\{x^{k_n}\}$, $\{x^{j_n}\}$ respectively. We have just proved that both $\tilde{x}$ and $\hat{x}$ belong to $S^* \times U$, and thus, by (40), both $\{\|\hat{x} - x^n\|\}$ and $\{\|\tilde{x} - x^n\|\}$ are nonincreasing, hence convergent, say to $\hat{\alpha} \geq 0$ and to $\tilde{\alpha} \geq 0$ respectively. Now,

$$\|\hat{x} - x^n\|^2 = \|\hat{x} - \tilde{x}\|^2 + \|\tilde{x} - x^n\|^2 + 2\langle \hat{x} - \tilde{x}, \tilde{x} - x^n \rangle. \tag{45}$$

Taking limits in (45) as $n \to \infty$ along the subsequence $\{x^{k_n}\}$, we get

$$\|\hat{x} - \tilde{x}\|^2 = \hat{\alpha}^2 - \tilde{\alpha}^2. \tag{46}$$

Reversing now the roles of $\tilde{x}$, $\hat{x}$ in (45), and taking limits along the subsequence $\{x^{j_n}\}$, we get

$$\|\hat{x} - \tilde{x}\|^2 = \tilde{\alpha}^2 - \hat{\alpha}^2. \tag{47}$$

It follows from (46) and (47) that $\tilde{x} = \hat{x}$, and thus the whole sequence $\{x^n\}$ has a weak limit which is a zero of T and belongs to U.

□

The next corollary states the global result for the case in which T is ρ-hypomonotone in the whole $H \times H$.

Corollary 2. *Let $T : H \to \mathcal{P}(H)$ be maximal ρ-hypomonotone with nonempty set of zeroes S^*. Take a sequence $\{\gamma_n\}$ of positive real numbers such that $2\rho < \hat{\gamma} = \inf\{\gamma_n\}$. Then, both for Algorithm 1 and Algorithm 2, given any $x^0 \in H$,* *(a) for all n there exist $y^n, v^n, e^n, x^{n+1} \in H$ satisfying* (10)–(12) *and* (15)*, in the case of Algorithm 1, and* (10)–(11) *and* (13)–(15)*, in the case of Algorithm 2,*

(b) any sequence generated by either Algorithm 1 or Algorithm 2 is weakly convergent to a point in S^*.

Proof. This is just Theorem 1 for the case of $U = V = H$. In this case all the assumptions above hold trivially. Regarding condition (i), note that S^* is closed because, by Proposition 3, it is also the set of zeroes of the maximal monotone operator $Y_\rho(T)$, which is closed (see e.g. 12.8(a) and (c) in [16]). Conditions (ii) and (iii) hold for any $\beta, \delta > 0$, so that the result will hold for any $\varepsilon > 0$, in particular for $\varepsilon > d(x^0, S^*)$. □

We close this section with a restatement of Theorem 1, which is needed in our section on multiplier methods.

Corollary 3. *Let $Z \subset H$ be a linear subspace, and consider Algorithms 1 and 2, with the additional requirement that e^n, besides satisfying* (12) *or* (13)-(14)*, belong to Z. Then, the results of Theorem 1 still hold.*

Proof. The inductive step in the proof of Theorem 1(a), based on Lemma 1, esentially consists of establishing, for all n, the existence of exact iterates, i.e. with $e^n = 0$, which certainly belongs to Z, so that item (a) does hold with the additional requirement. The proof of item (b), depending on the results of Lemma 2, is valid for all sequence $\{x^n\}$ as in item (a), hence in particular for those sequences such that $e^n \in Z$. □

5. Convergence rate results

We prove in this section that our inexact algorithm still enjoys the convergence rate results which are already classical for proximal point algorithms, namely at least a linear convergence rate when T^{-1} is locally Lipschitz at 0 (see [14] for the monotone case and [13] for the nonmonotone one with exact iterates). We will say that $Q : H \to \mathcal{P}(H)$ is Lipschitz continuous at $W \subset H$ if there exists a constant $\lambda \geq 0$ such that $\|v - v'\| \leq \lambda \|x - x'\|$ for all $x, x' \in W$, all $v \in Q(x)$ and all $v' \in Q(x')$. Our convergence rate result is the following.

Theorem 2. *Under all the assumptions of Theorem 1, suppose furthermore that T^{-1} is Lipschitz continuous, with constant λ, in a neighborhood $W \subset H$ of 0. Let x^* be the weak limit point of the sequence $\{x^n\}$. Then there exists n_0 such that the following inequalities hold for all $n \geq n_0$:*

(i)

$$\left\|x^{n+1} - x^*\right\| \leq \frac{\lambda + \mu}{\sqrt{(\lambda + \mu)^2 + \theta_n}} \|x^n - x^*\| \tag{48}$$

for Algorithm 1, where

$$\mu = \sigma\left(\frac{\hat{\gamma}}{2} - \rho\right), \quad \theta_n = (1-\sigma)\gamma_n(\hat{\gamma} - 2\rho), \tag{49}$$

(ii)

$$\left\|x^{n+1} - x^*\right\| \leq \frac{\omega_n}{\sqrt{\omega_n^2 + \xi}} \left\|x^n - x^*\right\|, \tag{50}$$

for Algorithm 2, where

$$\xi = (1-\sigma)\left(1 - \frac{2\rho}{\hat{\gamma}}\right), \quad \omega_n = \nu + (1+\nu)\frac{\lambda}{\gamma_n}, \tag{51}$$

with ν *as in* (13).

Proof. By (10), $y^n \in T^{-1}(v^n)$. By Theorem 1(b), $x^* \in T^{-1}(0)$. By (42), $\lim_{n\to\infty} v^n = 0$, and so there exists n_0 such that $v^n \in W$ for $n \geq n_0$. By Lipschitz continuity of T^{-1} in W, for $n \geq n_0$,

$$\left\|y^n - x^*\right\| \leq \lambda \left\|v^n\right\|. \tag{52}$$

By (11) and (15), $x^{n+1} = y^n - e^n$. Thus,

$$\left\|x^{n+1} - x^*\right\| = \left\|y^n - x^* - e^n\right\| \leq \left\|y^n - x^*\right\| + \left\|e^n\right\| \leq \lambda\left\|v^n\right\| + \left\|e^n\right\|, \tag{53}$$

using (52) in the second inequality.

Now we consider separately both algorithms.

(i) Combining Lemmma 2(ii), (49), (53) and (12) we get,

$$\left\|x^{n+1} - x^*\right\|^2 \leq \left\|x^n - x^*\right\|^2 - \theta_n\left\|v^n\right\|^2 \leq \left\|x^n - x^*\right\|^2 - \frac{\theta_n}{(\lambda+\mu)^2}\left\|x^{n+1} - x^*\right\|^2, \tag{54}$$

and the conclusion of item (i) follows directly from (54).

(ii) By (11), $v^n = \gamma_n^{-1}(x^n - y^n + e^n)$. Thus

$$\lambda\left\|v^n\right\| \leq \lambda\gamma_n^{-1}\left\|x^n - y^n + e^n\right\| \leq \lambda\gamma_n^{-1}\left(\left\|x^n - y^n\right\| + \left\|e^n\right\|\right). \tag{55}$$

Combining (53) and (55)

$$\left\|x^{n+1} - x^*\right\| \leq \lambda\gamma_n^{-1}\left(\left\|x^n - y^n\right\| + \left\|e^n\right\|\right) + \left\|e^n\right\| \leq$$

$$\left[\nu + (1+\nu)\frac{\lambda}{\gamma_n}\right]\left\|x^n - y^n\right\| = \omega_n\left\|x^n - y^n\right\|, \tag{56}$$

using (13) in the last inequality and (51) in the equality. By Lemma 2(ii), (51) and (56),

$$\left\|x^{n+1}-x^*\right\|^2 \le \|x^n-x^*\|^2-\xi\,\|x^n-y^n\|^2 \le \|x^n-x^*\|^2-\frac{\xi}{\omega_n^2}\left\|x^{n+1}-x^*\right\|^2, \tag{57}$$

and the conclusion of item (ii) follows immediately from (57).

□

Corollary 4. *Under the assumptions of Theorem 2, the sequences generated by Algorithms 1 and 2 converge at least linearly, when $\bar\gamma=\sup\{\gamma_n\}<\infty$, with asymptotic error constants given by $\frac{\lambda+\mu}{\sqrt{(\lambda+\mu)^2+\bar\theta}}$, $\frac{\bar\omega}{\sqrt{\bar\omega^2+\xi}}$ respectively, where $\bar\theta=(1-\sigma)\bar\gamma(\hat\gamma-2\rho)$, $\bar\omega=\nu+(1+\nu)\frac{\lambda}{\hat\gamma}$, and superlinearly, when $\lim_{n\to\infty}\gamma_n=\infty$.*

Proof. Note that $\omega_n\le\bar\omega$, $\theta_n\le\bar\theta$ for all n. Thus (54), and consequently (48), hold with $\bar\theta$ instead of θ_n. By the same token, (57), and consequently (50), hold with $\bar\omega$ instead of ω_n, establishing the asymptotic error constants. The statement on superlinear convergence follows directly from Theorem 2, observing that $\lim_{n\to\infty}\gamma_n=\infty$ implies that $\lim_{n\to\infty}\theta_n=\infty$ and $\lim_{n\to\infty}\omega_n=0$. □

Of course, a caveat is in order in connection with the result on superlinear convergence in Corollary 4, as is the case with similar results for other variants of the proximal point algorithm (e.g. [7]). Proximal procedures, in general, replace the inversion of T for a sequence of subproblems, each one of which demands inversion of $I+\gamma_nT$, or equivalently, of $\gamma_n^{-1}I+T$. On one hand, when γ_n becomes very large, $\gamma_n^{-1}I+T$ gets very close to T, and thus an arbitrarily high convergence rate can be achieved by making γ_n go fast enough to ∞ (in the limit, for $\gamma_n=\infty$ the algorithm would find a zero of T in one iteration). But this high convergence rate is deceiving, in the following sense. Replacement of the inversion of T by a sequence of subproblems is recommended basically when the inversion of T is hard, i.e. when T is somewhat ill-conditioned. The properties of T imply that $\gamma_n^{-1}I+T$ is instead theoretically well conditioned (for all $\gamma_n>0$ when T is monotone, for $\gamma_n>2\rho$ when T is ρ-hypomonotone), but, for very large γ_n, $\gamma_n^{-1}I+T$ becomes numerically almost as ill-conditioned as T, in which case the regularization properties of the algorithm are lost (think, e.g. of the case of linear and singular T). In fact, one of the main advantages of the proximal point algorithm with respect to other regularization schemes is that it works without requiring that the regulatrization coefficients go to ∞, i.e. for constant γ_n, for instance. In other words, when γ_n increases, a trade off takes place between the improvement in the convergence rate and the deterioration of the numerical stability.

6. Inexact proximal method of multipliers

Let X and Y be Hilbert spaces. For an arbitrary $S:X\to\mathcal{P}(X)$, a maximal monotone $T:Y\to\mathcal{P}(Y)$ and a $\mathcal{C}^2$ function $F:X\to Y$, we consider the problem of finding a solution to the inclusion

$$(P) \qquad S(x) + \nabla F(x)^* T(F(x)) \ni 0,$$

where $\nabla F(x)^*$ is the adjoint of the Jacobian of F at a point $x \in X$. This provides a flexible model for various applications, and it has an associated duality theory that can be seen as a generalization of the traditional convex programming duality theory. Combining dualization with the proximal point algorithm leads to multiplier methods for a wide class of problems much like in Rockafellar [15] where convex programs were treated. In [13], this approach was extended to problems of the form (P), and multiplier methods for variational inequalities and nonlinear convex programs were obtained as special cases. The purpose of this section is to derive an inexact version of the proximal method of multipliers for (P).

We reproduce here those parts of the duality theory which are needed in the sequel (see [13] for a full exposition). Denote

$$F_0(x) = S(x) + \nabla F(x)^* T(F(x))$$

so that (P) can be written as $F_0(x) \ni 0$. Define the *Lagrangian* $L : X \times Y \to \mathcal{P}(X \times Y)$ by

$$L(x, y) = (\nabla F(x)^* y, -F(x)) + S(x) \times T^{-1}(y),$$

and consider the *primal-dual problem*

$$(PD) \qquad L(x, y) \ni (0, 0).$$

The mapping F_0 is related to L by

$$\begin{aligned} F_0(x) &= \{ S(x) + \nabla F(x)^* y \mid \exists y \in Y : \ y \in T(F(x)) \} \\ &= \left\{ S(x) + \nabla F(x)^* y \;\middle|\; \exists y \in Y : \ 0 \in -F(x) + T^{-1}(y) \right\} \\ &= \{ v \in X \mid \exists y \in Y : \ (v, 0) \in L(x, y) \} \end{aligned} \tag{58}$$

The following is immediate.

Lemma 3. *We have $F_0(x) \ni e$ if and only if there exists a y such that $L(x, y) \ni (e, 0)$.*

We will also need to reformulate Algorithm 2. Eliminating v^n, and denoting y^n by $\tilde{z}^n$ and x^n by z^n, we can write it as follows.

Method 1.

1. Given z^n, find a $\tilde{z}^n$ such that

$$\gamma_n T(\tilde{z}^n) + \tilde{z}^n - z^n \ni e^n$$

for some $e^n \in Z$ satisfying

$$\|e^n\| \leq \nu \|\tilde{z}^n - z^n\|.$$

2. Set

$$z^{n+1} = \tilde{z}^n - e^n,$$

$n = n + 1$, and go to 1.

In step 1, Z denotes a subspace of H. Corollary 3 can now be stated in the following equivalent form.

Theorem 3. *Under the assumptions of Theorem 1,*

(a) There exists an infinite sequence $\{z^n\} \subset H$, conforming to Method 1, and satisfying

$$\begin{aligned} \tilde{z}^n &\in U \\ e^n - \tilde{z}^n + z^n &\in \gamma_n V \\ d(x^{n+1}, S^* \cap U) &\le \varepsilon; \end{aligned}$$

(b) Any sequence as in (a) converges weakly to a point in $S^* \cap U$.

If we apply Method 1 with $Z = X \times \{0\}$ to (PD), we get the following.

Method 2.

1. Given $(x^n, y^n) \in X \times Y$, find a $(\tilde{x}^n, \tilde{y}^n) \in X \times Y$ such that

$$\gamma_n L(\tilde{x}^n, \tilde{y}^n) + [(x, y) - (x^n, y^n)] \ni (e^n, 0),$$

for some e^n satisfying

$$\|e^n\| \le \nu \|(\tilde{x}^n, \tilde{y}^n) - (x^n, y^n)\|.$$

2. Set

$$\begin{aligned} x^{n+1} &= \tilde{x}^n - e^n, \\ y^{n+1} &= \tilde{y}^n, \end{aligned}$$

$n = n + 1$, and go to 1.

The inclusion in step 1 can be written as

$$L_n(\tilde{x}^n, \tilde{y}^n) \ni \gamma_n^{-1}(e^n, 0), \tag{59}$$

where

$$
\begin{aligned}
L_n(x,y) &= L(x,y) + \gamma_n^{-1}(x - x^n, y - y^n) \\
&= (\nabla F(x)^* y, -F(x)) + [S(x) + \gamma_n^{-1}(x - x^n)] \times [T^{-1}(y) + \gamma_n^{-1}(y - y^n)] \\
&= (\nabla F(x)^* y, -F(x)) + S_n(x) \times T_n^{-1}(y),
\end{aligned}
$$

with $S_n(x) = S(x) + \gamma_n^{-1}(x - x^n)$, and $T_n(u) = (I + \gamma_n T^{-1})^{-1}(y^n + \gamma_n u)$. We thus get from (59) that $\tilde{y}^n \in T_n(F(\tilde{x}^n))$. But since T_n is single-valued by the maximal monotonicity of T, we see that the value of $\tilde{x}^n$ determines the value of $\tilde{y}^n$ uniquely by

$$\tilde{y}^n = T_n(F(\tilde{x}^n)).$$

Since L_n is in the format of the general duality framework, we have by Lemma 3 that $(\tilde{x}^n, \tilde{y}^n)$ satisfies (59) if and only if

$$
\begin{aligned}
S(\tilde{x}^n) + \gamma_n^{-1}(\tilde{x}^n - x^n) + \nabla F(\tilde{x}^n)^* T_n(F(\tilde{x}^n)) &\ni \gamma_n^{-1} e^n \\
\tilde{y}^n &= T_n(F(\tilde{x}^n)).
\end{aligned}
$$

Method 2 can thus be written as

Method 3 (Proximal method of multipliers).

1. Given $(x^n, y^n) \in X \times Y$, find an $\tilde{x}^n \in X$ such that

$$S(\tilde{x}^n) + \gamma_n^{-1}(\tilde{x}^n - x^n) + \nabla F(\tilde{x}^n)^* T_n(F(\tilde{x}^n)) \ni \gamma_n^{-1} e^n,$$

for some e^n satisfying

$$\|e^n\| \le \nu \|(\tilde{x}^n, T_n(F(\tilde{x}^n))) - (x^n, y^n)\|.$$

2. Set

$$
\begin{aligned}
x^{n+1} &= \tilde{x}^n - e^n, \\
y^{n+1} &= T_n(F(\tilde{x}^n)),
\end{aligned}
$$

$n = n + 1$ and go to 1.

Theorem 3 can be herefore restated in the following form.

Theorem 4. *Assume that L^{-1} is maximal ρ-hypomonotone in a subset $V \times U$, where $U \subset X \times Y$ and $V \subset X \times Y$ satisfy* *(i)* $S^* \cap U$ *is nonempty and closed,*

(ii) there exists $\beta > 0$ *such that* $B(0, \beta) \subset V$,

(iii) there exists $\delta > 0$ *such that* $(S^* \cap U) + B(0, \delta) \subset U$,

(iv) U *is convex,*

where S^* *is the set of zeroes of* L. *Take a sequence* $\{\gamma_n\}$ *of positive real numbers such that* $2\rho < \hat{\gamma} = \inf\{\gamma_n\}$. *Define* $\varepsilon = \min\{\delta, \beta\gamma/2\}$. *If* $d(x^0, y^0, S^* \cap U) \leq \varepsilon$, *then*

(a) There exists an infinite sequence $\{(x^n, y^n)\} \subset X \times Y$, *conforming to Method 3, such that for all* n

$$\begin{aligned} (\tilde{x}^n, T_n(F(\tilde{x}^n))) &\in U, \\ (e^n, 0) - (\tilde{x}^n, T_n(F(\tilde{x}^n))) + (x^n, y^n) &\in \gamma_n V, \\ d((x^{n+1}, y^{n+1}), S^* \cap U) &\leq \varepsilon. \end{aligned}$$

(b) Any sequence as in (a), converges to a point in $S^* \cap U$.

References

[1] Bertsekas, D.P. *Constrained Optimization and Lagrange Multipliers*. Academic Press, New York (1982).

[2] Brezis, H. *Analyse Fonctionelle, Théorie et Applications*. Masson, Paris (1983).

[3] Burachik, R.S., Scheimberg, S., Svaiter, B.F. Robustness of the hybrid extragradient proximal point algorithm. *Journal of Optimization Theory and Applications* **111** (2001).

[4] Burachik, R.S., Svaiter, B.F. A relative error tolerance for a family of generalized proximal point methods (to be published).

[5] Eckstein, J., Bertsekas, D. On the Douglas-Ratford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming* **55** (1992) 293-318.

[6] Eckstein, J., Ferris, M. Smooth methods of multipliers for complementarity problems. *Mathematical Programming* **86** (1999) 65-90.

[7] Iusem, A.N., Teboulle, M. On the convergence rate of entropic proximal optimization algorithms. *Computational and Applied Mathematics* **12** (1993) 153-168.

[8] Kaplan, A., Tichatschke, R. Proximal point methods and nonconvex optimization. *Journal of Global Optimization* **13** (1998) 389-406.

[9] Krasnoselskii, M.A. Two observations about the method of succesive approximations. *Uspekhi Matematicheskikh Nauk* **10** (1955) 123-127.

[10] Martinet, B. Régularisation d'inéquations variationelles par approximations succesives. *Revue Française de Informatique et Recherche Opérationelle* **2** (1970) 154-159.

[11] Minty, G. A theorem on monotone sets in Hilbert spaces. *Journal of Mathematical Analysis and Applications* **11** (1967) 434-439.

[12] Moreau, J. Proximité et dualité dans un espace hilbertien. *Bulletin de la Societé Mathématique de France* **93** (1965).

[13] Pennanen, T. Local convergence of the proximal point algorithm and multiplier methods without monotonicity, *Mathematics of Operations Research*, to appear.

[14] Rockafellar, R.T. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization* **14** (1976) 877-898.

[15] Rockafellar, R.T. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research* **1** (1976) 97-116.

[16] Rockafellar, R.T., Wets, R.J.-B. *Variational Analysis*. Springer, Berlin (1998).

[17] Solodov, M.V., Svaiter, B.F. A hybrid projection–proximal point algorithm. *Journal of Convex Analysis* **6** (1999) 59-70.

[18] Solodov, M.V., Svaiter, B.F. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Analysis* **7** (1999) 323-345.

[19] Solodov, M.V., Svaiter, B.F. An inexact hybrid generalized proximal point algorithms and some new results on the theory of Bregman functions. *Mathematics of Operations Research* **51** (2000) 479-494.

[20] Spingarn, J.E. Submonotone mappings and the proximal point algorithm. *Numerical Functional Analysis and Optimization* **4** (1981) 123-150.