

Teemu Pennanen · B. F. Svaiter

Solving monotone inclusions with linear multi-step methods

Received: April 2001 / Accepted: November 2002

Published online: February 14, 2003 – © Springer-Verlag 2003

Abstract. In this paper a new class of proximal-like algorithms for solving monotone inclusions of the form $T(x) \ni 0$ is derived. It is obtained by applying linear multi-step methods (LMM) of numerical integration in order to solve the differential inclusion $\dot{x}(t) \in -T(x(t))$, which can be viewed as a generalization of the steepest decent method for a convex function. It is proved that under suitable conditions on the parameters of the LMM, the generated sequence converges weakly to a point in the solution set $T^{-1}(0)$. The LMM is very similar to the classical proximal point algorithm in that both are based on approximately evaluating the resolvents of T . Consequently, LMM can be used to derive multi-step versions of many of the optimization methods based on the classical proximal point algorithm. The convergence analysis allows errors in the computation of the iterates, and two different error criteria are analyzed, namely, the classical scheme with summable errors, and a recently proposed more constructive criterion.

Key words. proximal point algorithm – monotone operator – numerical integration – strong stability – relative error criterion

1. Introduction

Consider the problem of finding a solution to the inclusion

$$T(x) \ni 0, \tag{1}$$

where $T : X \rightrightarrows X$ is a set-valued mapping on a real Hilbert space X . This is a very general format for problems of variational character, like those of minimization or maximization of a function, or variational inequalities. It often happens that the mapping T can be chosen so that it is *monotone* i.e.

$$v_1 \in T(x_1), v_2 \in T(x_2) \implies \langle x_1 - x_2, v_1 - v_2 \rangle \geq 0.$$

This is the case, for example, in convex programming and monotone variational inequalities. Recall that a monotone mapping is called *maximal monotone* if it does not have a proper monotone extension. Maximality is crucial when analyzing numerical algorithms, and fortunately, it is present in many problems arising in practice [17, 5, 28].

For monotone inclusions of the form (1) there are algorithms that can be used to generate a sequence of points $\{x^k\}$ converging to the solution set $T^{-1}(0)$ of (1). The most

T. Pennanen: Department of Management Science, Helsinki School of Economics, PL 1210 00101 Helsinki, Finland, e-mail: pennanen@hkkk.fi

Partially supported by Foundation for Economic Education and IMPA

B. F. Svaiter: Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro RJ, CEP 22460-320, Brazil, e-mail: benar@impa.br

Partially supported by CNPq Grant 301200/93-9 (RN) and by PRONEX-Optimization

Mathematics Subject Classification (1991): 20E28, 20G40, 20C20

famous of such general algorithms is the *proximal point algorithm* (PPA) of Rockafellar [27], which is very useful in developing numerical methods for variational problems. For example, the convergence properties of multiplier methods, bundle methods, and certain operator splitting methods can be most easily derived from the general convergence properties of the PPA; see for example [26, 10, 14]. The PPA has been generalized and improved in various ways by many authors over the years; see for example [13, 12, 37, 15, 29, 31, 3] and the references therein. Of special interest to us here is the original paper of Rockafellar [27] and the recent papers of Solodov and Svaiter [29, 31, 34] where hybrids extragradient-proximal and projection-proximal point algorithms are presented.

This paper derives a new class of methods that arises from the field of numerical integration of differential equations. It is shown that *linear multi-step methods* (LMM) can be implemented for general set-valued maximal monotone mappings, and under suitable conditions on the parameters, the generated sequence converges weakly to a solution of (1). These methods are closely related to proximal-type methods in that their implementation is based on the approximate computation of the resolvents of T . Consequently, just like the PPA, these methods can be used to derive methods similar to multiplier and splitting methods.

Let $T : X \rightrightarrows X$ be maximal monotone, let $x_0 \in X$, and consider the initial value problem

$$\begin{cases} \dot{x}(t) \in -T(x(t)) & \text{a.e. } t \geq 0, \\ x(0) = x_0. \end{cases} \quad (2)$$

By the famous result of Grandall and Pazy [11] (see also Brezis [5]) (2) has a unique absolutely continuous solution $x : [0, +\infty) \mapsto X$. Moreover, Bruck [6] showed that, for a wide class of mappings T that he called *demipositive*, the solution of (2) converges weakly to a solution of (1) as $t \rightarrow \infty$, whenever one exists. Higher-order systems with similar properties have been studied for example in [19, 1, 2, 4]. This suggests that one might be able to use methods of numerical integration to follow the solution of (2) to a solution of (1). Indeed, writing the proximal point algorithm in the form

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

we see that it is the implicit version of the classical Euler's method applied to (2). The implicit Euler's method is known for its stable behavior when applied to "stiff" problems like (2); see for example [18, 16]. In the numerical analysis literature, it is known that the implicit Euler's method is *A-stable*, which means that if it is applied to (2) with

$$T = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}, \quad (3)$$

where $a > 0$, the sequence it generates converges to zero. The main convergence result of [27] says more: for *any* maximal monotone mapping T (possibly nonlinear and set-valued), the generated sequence converges weakly to a point in $T^{-1}(0)$. Not even demipositivity of T is needed like in the continuous case. In particular, if the implicit Euler's method is applied to (3) it converges to zero even when $a = 0$, provided $b \neq 0$.

Note that in this case the exact solution of (2) doesn't converge, but oscillates at a fixed distance from the origin.

Euler's method and many other integration methods fall in the class of *linear multi-step methods* (LMM); see for example [18, 16]. In the literature, LMMs are usually applied to differential equations (for simplicity, we only consider the autonomous case)

$$\dot{x}(t) = f(x(t)) \quad (4)$$

for a continuous f . A linear m -step method with variable step size can then be written as

$$x_k = \sum_{j=1}^m \alpha_j x_{k-j} + \sum_{j=0}^m \gamma_j h_{k-j} f(x_{k-j}), \quad (5)$$

where h_k are positive step-size parameters, and α_j and γ_j are scalars to be chosen. The rule (5) means that having computed x_1, \dots, x_{k-1} , the next point x_k is chosen so that (5) is satisfied. If $\gamma_0 = 0$, an explicit expression is available for x^k . If $\gamma_0 \neq 0$, numerical computations are required to find x^k , and the method is called *implicit*. If a linear multi-step method is to be able to solve (1) for every maximal monotone T , then it has to be A-stable since the negative of the linear mapping (3) in the definition of A-stability is maximal monotone, with the unique root $(0, 0)$. From the theory of A-stability (see for example [18]) we know that all A-stable LMMs are implicit. In implicit LMMs an iterate x^k is obtained by solving the equation

$$(I - \gamma_0 h_k f)(x) = \sum_{j=1}^m [\alpha_j x^{k-j} + \gamma_j h_{k-j} f(x^{k-j})]. \quad (6)$$

In order to guarantee the solvability of (6) one needs to pose assumptions on f and the parameters. A simple way is to assume that $-f$ is maximal monotone, and γ_0 is positive. Then by the classical result of Minty [17], the mapping $(I - \gamma_0 h_k f)$ is surjective and (6) always has a unique solution.

Generalization of the implicit linear m -step method for the monotone differential inclusion (2) is now straightforward: we replace $-f$ by a general maximal monotone mapping T , and on the right hand side of (6), we replace $f(x^{k-j})$ by a single element $-v^{k-j} \in -T(x^{k-j})$. The equation (6) then becomes the inclusion

$$(I + c_k T)(x) \ni \sum_{j=1}^m (\alpha_j x^{k-j} - \beta_j c_{k-j} v^{k-j}), \quad (7)$$

where $c_k = \gamma_0 h_k$ and $\beta_j = \gamma_j / \gamma_0$. There remains the problem of finding $v^k \in T(x^k)$, that are needed on the right hand side of (7). Reordering (7), we see that if x^k solves (7) then

$$v^k := c_k^{-1} \left[\sum_{j=1}^m (\alpha_j x^{k-j} - \beta_j c_{k-j} v^{k-j}) - x^k \right] \in T(x^k).$$

The linear m -step method (in the exact form) for (2) can thus be written as follows.

Algorithm 1

0. Choose $c_0, \dots, c_{m-1} > 0$ and $\{x^0, \dots, x^{m-1}\}, \{v^0, \dots, v^{m-1}\} \subset X$;

1. Set

$$s^k = \sum_{j=1}^m (\alpha_j x^{k-j} - \beta_j c_{k-j} v^{k-j});$$

2. Choose $c_k > 0$ and solve

$$(I + c_k T)(x) \ni s^k$$

for x^k ;

3. Set

$$v^k = c_k^{-1}(s^k - x^k),$$

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

When T is maximal monotone, Algorithm 1 is well defined in the sense that for any starting points $\{x^0, \dots, x^{m-1}\}, \{v^0, \dots, v^{m-1}\} \subset X$, it generates infinite sequences $\{x^k\}$, $\{s^k\}$, and $\{v^k\}$. The main purpose of this paper is to show that the linear multi-step method has similar properties as the classical proximal point algorithm. More precisely, we give conditions on the parameters α_j and β_j that guarantee the weak convergence of the sequences $\{x^k\}$ and $\{s^k\}$ to a solution of (1), whenever T is maximal monotone and $0 \in \text{rge } T$. Linear multi-step methods can thus be used very much like the proximal point algorithm, for example, in derivation of multi-step versions of multiplier or splitting methods.

The biggest computational burden of Algorithm 1 is contained in step 2, where we are asked to solve an inclusion which can be as hard as the original problem (1). It is thus important that we can accept inexact solutions of these subproblems. In [27], Rockafellar proved the convergence of the proximal point algorithm while allowing iterates of the form $x^k = (I + c_k T)^{-1}(x^{k-1}) + e^k$, where the errors e^k form a summable sequence. Recently, Solodov and Svaiter proposed modifications to the classical PPA, which allow a considerable relaxation of the error criterion by adding an additional extragradient or a projection step to the PPA [29, 31, 34]. These methods provide a more constructive error criterion, and the extra step guarantees Fejér monotonicity and global convergence of the generated sequence, with a negligible cost in computation. We will show that both these approximation schemes can be applied also in the case of LMM, with similar convergence results as for the PPA.

Note that if we choose $m = 1, \alpha_1 = 1$ and $\beta_1 = 0$, Algorithm 1 becomes the classical proximal point algorithm. On the other hand, choosing $m = 1, \alpha_1 = 1$ and $\beta_1 = \sigma - 1$ for $\sigma \in (0, 2)$, and eliminating v^k from the equations of Algorithm 1, we obtain the following method

1. Set

$$s^k = \sigma x^{k-j} + (1 - \sigma)s^{k-1};$$

2. Choose $c_k > 0$, solve

$$(I + c_k T)(x) \ni s^k$$

for x^k , set $k = k + 1$ and go to 1,

which is the relaxed proximal point algorithm of Eckstein and Bertsekas [13] in the exact form, and with a fixed relaxation parameter. As shown in [14], over-relaxation can speed up convergence. This is important especially when variations in the step-size parameter c_k are not allowed, which is the case when deriving certain splitting methods [35, 13]. Combining Algorithm 1 with the ideas in [35, 13], one can derive splitting methods with arbitrarily many parameters that can be tuned for the particular problem at hand.

It is not possible to single out an LMM that would be best in general. To give a concrete example where there is an LMM that performs better than the classical PPA, consider the linear mapping $T : \mathbb{R} \rightarrow \mathbb{R}$ given by $Tx = \lambda x$, with $\lambda > 0$. The PPA with a fixed step-size c then generates a sequence satisfying

$$x^k = \frac{1}{1 + c\lambda} x^{k-1}, \quad (8)$$

whereas an LMM with $m = 1$ and $\alpha_1 = 1$, generates a sequence satisfying

$$x^k = \frac{1 - \beta_1 c\lambda}{1 + c\lambda} x^{k-1}. \quad (9)$$

The latter converges faster to $T^{-1}(0) = \{0\}$ whenever $\beta_1 \in (0, \frac{2}{c\lambda})$. With $\beta_1 = \frac{1}{c\lambda}$ the LMM solves the problem in a single iteration. More generally, one can analyze the behavior of LMMs for general linear mappings in terms of eigenvalues. For example, if we have a decomposition $T = X \text{diag}(\lambda_1, \dots, \lambda_n) X^{-1}$, an LMM with $m = 1$, $\alpha_1 = 1$ and $|1 - \beta_1 c\lambda_i| < 1 \forall i = 1, \dots, n$ has a better convergence rate than the PPA.

From numerical integration point of view, our convergence results can be viewed as stability results for LMM. In this sense, our aim is to describe a subset of the class of A-stable LMMs, which for any maximal monotone mapping T , generate sequences that converge to $T^{-1}(0)$. Since we are interested in solving problem (1) rather than (2), we will not be concerned with consistency or accuracy properties of LMM, which are central in the analysis of numerical integration methods. For a consistency analysis of LMMs in the case of (possibly discontinuous) monotone mappings see Nevanlinna [20]. Being aware of what happens for the exact solution of (2) with (3) and $a = 0$, we know that if we want to reach $T^{-1}(0)$ it is probably not a good idea to follow the solution of (2) too closely.

Section 2, provides a convergence analysis of the LMM for general monotone mappings under the classical condition that the errors in the computation of the resolvents be summable, and Section 3 studies the case of a relative error criterion with an extragradient step. In Section 4 we show that under the classical Lipschitz-type condition on the inverse of T , the LMM attains local linear convergence much like the standard proximal point algorithm [27].

2. Approximate iterates with summable errors

To make Algorithm 1 implementable, it is usually necessary to allow approximate solutions of the subproblems in step 2. In this section, we consider the linear m -step method in the following approximate form.

Algorithm 2

0. Choose $c_0, \dots, c_{m-1} > 0$ and $\{x^0, \dots, x^{m-1}\}, \{\tilde{v}^0, \dots, \tilde{v}^{m-1}\} \subset X$;

1. Set

$$s^k = \sum_{j=1}^m (\alpha_j x^{k-j} - \beta_j c_{k-j} \tilde{v}^{k-j});$$

2. Choose $c_k > 0$ and compute

$$x^k = (I + c_k T)^{-1}(s^k) + e^k;$$

3. Set

$$\tilde{v}^k = c_k^{-1}(s^k - x^k),$$

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

The vector e^k in step 2 is interpreted as an error allowed in the computation of the resolvent. It will be shown that if the error sequence is summable, and if the parameters satisfy suitable conditions, then this algorithm will generate sequences weakly converging to a point in $T^{-1}(0)$. Similar summability condition for the errors was used in [27] to prove the convergence of the classical proximal point algorithm.

Given an LMM, define the polynomial

$$p(z) = z^m - \sum_{j=1}^m \alpha_j z^{m-j},$$

and let A be its companion matrix:

$$A = \begin{bmatrix} \alpha_1 & \cdots & \alpha_m \\ 1 & & 0 \\ & \ddots & \vdots \\ & & 1 & 0 \end{bmatrix}. \tag{10}$$

Recall that the roots of p coincide with the eigenvalues of A , and the spectral radius of A (or p) is the number

$$\rho(A) = \max \{ |\lambda| \mid p(\lambda) = 0 \}.$$

Many numerical integration methods (Euler, Trapezoidal, . . .) share the following property; see for example [22, 16].

Definition 1. A linear m -step method is said to be strongly stable if $\rho(A) \leq 1$, and $|\lambda_i| = 1$ for exactly one eigenvalue λ_i of A .

Our convergence analysis is based on two lemmas.

Lemma 1. Consider a linear m -step method, and let $\{\xi_k\}$, $\{\omega_k\}$ and $\{v_k\}$ be sequences of nonnegative real numbers such that $\sum v_k < \infty$ and

$$\xi_k \leq \sum_{j=1}^m \alpha_j \xi_{k-j} - \omega_k + v_k \quad \forall k \geq m. \quad (11)$$

If $\alpha_j \geq 0$ and $\sum_{j=1}^m \alpha_j = 1$, then $\{\xi_k\}$ is bounded, and $\sum_{k=m}^{\infty} \omega_k < \infty$. If in addition, the method is strongly stable, then $\{\xi_k\}$ converges.

Proof. For $k \geq m$, define $\hat{\xi}_k = \max\{\xi_{k-1}, \dots, \xi_{k-m}\}$. Then (11) and our assumptions on the α_j 's imply

$$\hat{\xi}_{k+1} \leq \hat{\xi}_k + v_k,$$

so that by [23, Lemma 1, Sec. 2.2.1] the sequence $\{\hat{\xi}_k\}$ is bounded and convergent. This implies the boundedness of $\{\xi_k\}$.

Define for $l \geq m$,

$$\gamma_l = \sum_{k=m}^l \xi_k.$$

Then for $l \geq 2m$,

$$\begin{aligned} \gamma_l &\leq \sum_{k=m}^l \left(\sum_{j=1}^m \alpha_j \xi_{k-j} - \omega_k + v_k \right) = \sum_{j=1}^m \alpha_j \sum_{k=m}^l \xi_{k-j} - \sum_{k=m}^l \omega_k + \sum_{k=m}^l v_k \\ &= \sum_{j=1}^m \alpha_j \left(\gamma_{l-j} + \sum_{k=m-j}^{m-1} \xi_k \right) - \sum_{k=m}^l \omega_k + \sum_{k=m}^l v_k. \end{aligned}$$

Since γ_l is nondecreasing, we get

$$\gamma_l \leq \sum_{j=1}^m \alpha_j \gamma_{l-j} + \delta - \sum_{k=m}^l \omega_k + \sum_{k=m}^l v_k \leq \gamma_l + \delta - \sum_{k=m}^l \omega_k + \sum_{k=m}^l v_k,$$

so that

$$\sum_{k=m}^l \omega_k \leq \delta + \sum_{k=m}^l v_k,$$

where $\delta = \sum_{j=1}^m \alpha_j \sum_{k=m-j}^{m-1} \xi_k$ is a constant. Thus, $\sum_{k=m}^{\infty} \omega_k < \infty$.

Absorbing any possible slack in the inequality (11) into ω_k , we can assume without loss of generality that it holds as an equality, and then we can write it in the vector form

$$x^{k+1} = Ax^k + b^k, \quad (12)$$

where $x^k = (\xi^{k-1}, \dots, \xi^{k-m})$, $b^k = (v_k - \omega_k, 0, \dots, 0)$, and A is defined by (10). By the first part of the lemma the sequence $\{b^k\}$ is summable. With the Jordan decomposition $A = P^{-1}JP$ we can write (12) in the form

$$y^{k+1} = Jy^k + c^k,$$

where $y^k = Px^k$ and $c^k = Pb^k$. Since we are assuming that our method is strongly stable, J is of the form

$$J = \begin{bmatrix} \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & \tilde{J} & & \\ 0 & & & \end{bmatrix},$$

where $|\lambda| = 1$ and $\rho(\tilde{J}) < 1$. But the condition $\sum_{j=1}^m \alpha_j = 1$ implies $p(1) = 0$, so we must have $\lambda = 1$. Then, since the summability of $\{b^k\}$ implies that of $\{c^k\}$, it follows that $\{y^k\}$ converges, implying the convergence of $\{x^k\} = \{P^{-1}y^k\}$. \square

When $\sum_{j=1}^m \alpha_j = 1$ and $\alpha_j \geq 0$, we have a simple condition for strong stability.

Example 1. When $\sum_{j=1}^m \alpha_j = 1$ and $\alpha_j \geq 0$, the LMM is strongly stable if $\alpha_r, \alpha_{r+1} > 0$ for some $r = 1, \dots, m - 1$.

Proof. When $\sum_{j=1}^m \alpha_j = 1$ and $\alpha_j \geq 0$, we get from $p(\lambda) = 0$ that

$$|\lambda|^m = |\lambda^m| = \left| \sum_{j=1}^m \alpha_j \lambda^{m-j} \right| \leq \sum_{j=1}^m \alpha_j |\lambda|^{m-j},$$

which implies $|\lambda| \leq 1$. In other words, $\rho(A) \leq 1$. Also, from $\sum_{j=1}^m \alpha_j = 1$ we get $p(1) = 0$. Now assume that λ is a root of p with $|\lambda| = 1$. Then

$$\begin{aligned} 1 &= |\lambda|^m \leq \sum_{j \neq r, r+1} \alpha_j |\lambda|^{m-j} + |\alpha_r \lambda^{m-r} + \alpha_{r+1} \lambda^{m-r-1}| \\ &= \sum_{j \neq r, r+1} \alpha_j |\lambda|^{m-j} + |\alpha_r \lambda + \alpha_{r+1}| |\lambda|^{m-r-1} \\ &= \sum_{j \neq r, r+1} \alpha_j + |\alpha_r \lambda + \alpha_{r+1}|. \end{aligned}$$

Here $|\alpha_r \lambda + \alpha_{r+1}| \leq \alpha_r + \alpha_{r+1}$, with equality if and only if $\lambda = 1$. Thus, since $\sum_{j=1}^m \alpha_j = 1$, we must have $\lambda = 1$. To finish the proof it suffices to show that $\lambda = 1$ has multiplicity 1, or equivalently, that $p'(1) \neq 0$. This follows from the expression

$$p'(1) = m - \sum_{j=1}^{m-1} (m - j)\alpha_j,$$

and the fact that

$$\sum_{j=1}^{m-1} (m - j)\alpha_j \leq (m - 1) \sum_{j=1}^{m-1} \alpha_j \leq m - 1. \quad \square$$

The following is due to Opial [21].

Lemma 2. *Let $S \subset X$ be nonempty and let $\{u^k\} \subset X$ be an infinite sequence, such that the limit*

$$\lim_{k \rightarrow \infty} \|u^k - \bar{u}\|$$

exists for all $\bar{u} \in S$ and all the weak cluster points of $\{u^k\}$ belong to S . Then $\{u^k\}$ converges weakly to a point in S .

We can now prove the following.

Theorem 1. *Consider Algorithm 2 and assume that*

1. T is maximal monotone and $T^{-1}(0) \neq \emptyset$;
2. $\inf c_k > 0$;
3. $\alpha_j \geq 0 \forall j = 1, \dots, m$, $\sum_{j=1}^m \alpha_j = 1$, and the method is strongly stable;
4. $|\beta_j| \leq \alpha_j \forall j = 1, \dots, m$, with strict inequality for at least one j ;
5. $\sum \|e^k\| < \infty$.

Then for any starting points, the sequences $\{x^k\}$, and $\{s^k\}$ converge weakly to a point in $T^{-1}(0)$, and the sequence $\{c_k \tilde{v}^k\}$ converges strongly to zero.

Proof. For $k \geq m$, define

$$z^k := (I + c_k T)^{-1}(s^k), \tag{13}$$

$$u^k := c_k^{-1}(s^k - z^k), \tag{14}$$

and for $k \geq 2m$ define

$$\hat{s}^k = \sum_{j=1}^m (\alpha_j z^{k-j} - \beta_j c_{k-j} u^{k-j}). \tag{15}$$

Note that because of condition 4, we can restrict the summation above to the indices $\mathcal{J} = \{j \mid \alpha_j \neq 0\}$. Let $\bar{x} \in T^{-1}(0)$ and $k \geq 2m$. Since $z^i = s^i - c_i u^i$ for $i \geq m$, we get that

$$\hat{s}^k = \sum_{j \in \mathcal{J}} [\alpha_j s^{k-j} - (\alpha_j + \beta_j) c_{k-j} u^{k-j}],$$

and then,

$$\begin{aligned}
 \|\hat{s}^k - \bar{x}\|^2 &= \left\| \sum_{j \in \mathcal{J}} [\alpha_j (s^{k-j} - \bar{x}) - (\alpha_j + \beta_j) c_{k-j} u^{k-j}] \right\|^2 \\
 &= \left\| \sum_{j \in \mathcal{J}} \alpha_j \left[(s^{k-j} - \bar{x}) - \frac{\alpha_j + \beta_j}{\alpha_j} c_{k-j} u^{k-j} \right] \right\|^2 \\
 &\leq \sum_{j \in \mathcal{J}} \alpha_j \left\| (s^{k-j} - \bar{x}) - \frac{\alpha_j + \beta_j}{\alpha_j} c_{k-j} u^{k-j} \right\|^2 \\
 &= \sum_{j \in \mathcal{J}} \alpha_j \left[\|s^{k-j} - \bar{x}\|^2 \right. \\
 &\quad \left. + \left(\frac{\alpha_j + \beta_j}{\alpha_j} \right)^2 \|c_{k-j} u^{k-j}\|^2 - 2 \frac{\alpha_j + \beta_j}{\alpha_j} \langle s^{k-j} - \bar{x}, c_{k-j} u^{k-j} \rangle \right] \\
 &= \sum_{j \in \mathcal{J}} \alpha_j \|s^{k-j} - \bar{x}\|^2 \\
 &\quad + \sum_{j \in \mathcal{J}} (\alpha_j + \beta_j) \left[\frac{\alpha_j + \beta_j}{\alpha_j} \|c_{k-j} u^{k-j}\|^2 - 2 \langle s^{k-j} - \bar{x}, c_{k-j} u^{k-j} \rangle \right],
 \end{aligned} \tag{16}$$

where the inequality follows from the convexity of $\|\cdot\|^2$. Using (14), we get for every $i \geq m$

$$\langle s_i - \bar{x}, c_i u^i \rangle = \langle c_i u^i + z^i - \bar{x}, c_i u^i \rangle = \|c_i u^i\|^2 + c^i \langle z^i - \bar{x}, u^i - 0 \rangle \geq \|c_i u^i\|^2,$$

where the inequality holds since T is monotone and, by (13) and (14), $u^i \in T(z^i)$. Combining this with (16), we obtain

$$\|\hat{s}^k - \bar{x}\|^2 \leq \sum_{j \in \mathcal{J}} \alpha_j \|s^{k-j} - \bar{x}\|^2 - \sum_{j \in \mathcal{J}} \frac{\alpha_j^2 - \beta_j^2}{\alpha_j^2} \|c_{k-j} u^{k-j}\|^2. \tag{17}$$

By step 3, (13) and (14) we have $\tilde{v}^k = u^k - c_k^{-1} e^k$ for every $k \geq m$. This together with (15) gives

$$\begin{aligned}
 s^k &= \sum_{j \in \mathcal{J}} (\alpha_j x^{k-j} - \beta_j c_{k-j} \tilde{v}^{k-j}) \\
 &= \sum_{j \in \mathcal{J}} [\alpha_j z^{k-j} - \beta_j c_{k-j} u^{k-j} + (\alpha_j + \beta_j) e^{k-j}] \\
 &= \hat{s}^k + \sum_{j \in \mathcal{J}} (\alpha_j + \beta_j) e^{k-j} \quad \forall k \geq 2m,
 \end{aligned}$$

so that

$$\|s^k - \bar{x}\| \leq \|\hat{s}^k - \bar{x}\| + \sum_{j=1}^m (\alpha_j + \beta_j) \|e^{k-j}\|. \tag{18}$$

Now, if we define

$$\begin{aligned} \xi_k &= \|s^k - \bar{x}\|^2, \\ \zeta_k &= \|\hat{s}^k - \bar{x}\|^2, \\ \mu_k &= \sum_{j \in \mathcal{J}} \frac{\alpha_j^2 - \beta_j^2}{\alpha_j^2} \|c_{k-j} u^{k-j}\|^2, \\ \nu_k &= \sum_{j \in \mathcal{J}} (\alpha_j + \beta_j) \|e^{k-j}\|, \end{aligned}$$

we can write (17) and (18) as

$$\zeta_k \leq \sum_{j \in \mathcal{J}} \alpha_j \xi_{k-j} - \mu_k \tag{19}$$

$$\sqrt{\xi_k} \leq \sqrt{\zeta_k} + \nu_k, \tag{20}$$

Defining $\hat{\xi}_k = \max\{\xi_{k-1}, \dots, \xi_{k-m}\}$, we get from (19) that $\zeta_k \leq \hat{\xi}_k$ and then (20) implies $\sqrt{\xi_{k+1}} \leq \sqrt{\hat{\xi}_k} + \nu_k$, and thus, $\sqrt{\hat{\xi}_{k+1}} \leq \sqrt{\hat{\xi}_k} + \nu_k$, which by [23, Lemma 1, Sec. 2.2.1] guarantees that $\{\hat{\xi}_k\}$ converges, and so, both $\{\xi_k\}$ and $\{\zeta_k\}$ are bounded. Squaring (20) and using (19), we get

$$\xi_k \leq \sum_{j \in \mathcal{J}} \alpha_j \xi_{k-j} - \mu_k + \tilde{\nu}_k, \tag{21}$$

where $\tilde{\nu}_k = \nu_k^2 + 2\sqrt{\zeta_k} \nu_k$. The summability of $\{\|e^k\|\}$ implies that of $\{\nu_k\}$, so that $\sum \tilde{\nu}_k < \infty$ by the boundedness of $\{\zeta_k\}$. Thus, by Lemma 1 $\{\xi_k\} = \{\|s^k - \bar{x}\|^2\}$ converges, and

$$\sum_{k=m}^{\infty} \mu_k = \sum_{k=m}^{\infty} \sum_{j \in \mathcal{J}} \frac{\alpha_j^2 - \beta_j^2}{\alpha_j^2} \|c_{k-j} u^{k-j}\|^2 < \infty,$$

which can be written as

$$\left(\sum_{j \in \mathcal{J}} \frac{\alpha_j^2 - \beta_j^2}{\alpha_j^2} \right) \sum_{l=m}^{\infty} \|c_l u^l\|^2 < \infty.$$

Since by our assumptions on the β_j 's

$$\sum_{j \in \mathcal{J}} \frac{\alpha_j^2 - \beta_j^2}{\alpha_j^2} > 0,$$

we have $c_k u^k \rightarrow 0$. From (14) we then get also the convergence of $\|z^k - \bar{x}\|$. By (13) and (14), $u^k \in T(z^k)$, where $u^k \rightarrow 0$ by condition 2. Then by maximality of T , all the weak cluster points of $\{z^k\}$ are in $T^{-1}(0)$, and thus by Opial’s lemma $\{z^k\}$ converges weakly to a point in $T^{-1}(0)$. By (14) and the summability of $\{\|e^k\|\}$, the same is true of $\{s^k\}$ and $\{x^k\}$. The strong convergence of $c_k \tilde{v}^k$ to zero follows from that of $c_k u^k$ and from the fact that $c_k \tilde{v}^k = c_k u^k - e^k$ by step 3 and (14). \square

Note that choosing $m = 1, \alpha_1 = 1$ and $\beta_1 = 0$, we recover [27, Theorem 1].

3. Approximate iterates with an extragradient step

In this section we prove the convergence of a multi-step version of the hybrid extragradient-proximal point algorithm introduced in [31]. This method has the advantage of allowing a constructive error criterion for the subproblems of LMM. Instead of summability of an error sequence it uses a fixed parameter that measures the allowable error relative to the step size of an approximate proximal step. Convergence of the method is then guaranteed by performing a simple “extragradient step” after the usual (approximate) proximal one.

The general approximation scheme uses the ϵ -enlargement T^ϵ of T introduced by Burachik, Iusem and Svaiter in [7]. T^ϵ is defined, for $\epsilon \geq 0$ by

$$T^\epsilon(x) := \{v \in X \mid \langle v - u, x - y \rangle \geq -\epsilon, \forall (y, u) \in \text{gph } T\}. \tag{22}$$

See [7–9, 24, 25] and their references for a thorough study and further applications of this useful concept. Generalizations of this construct are studied in [36]. Note that if T is monotone, we have $T^\epsilon(x) \supseteq T(x)$ for all $x \in X, \epsilon \geq 0$, and if T is maximal monotone then $T^0(x) = T(x)$ for all $x \in X$.

Let x_k be the unique solution of the problem in step 2 of Algorithm 1, and consider the pair $(x_k, c_k^{-1}(s^k - x^k))$. This can be seen as a solution to the system

$$v \in T(x), \tag{23}$$

$$c_k v + x - s^k = 0. \tag{24}$$

In [31], the following definition of an approximate solution to (23)–(24) was introduced.

Definition 2. For $\sigma \in [0, 1)$, a pair (\tilde{x}, \tilde{v}) is called a σ -approximate solution of system (23)–(24) if there is an $\epsilon \geq 0$ such that

$$\begin{aligned} \tilde{v} &\in T^\epsilon(\tilde{x}), \\ \|c_k \tilde{v} + \tilde{x} - s^k\|^2 + 2c_k \epsilon &\leq \sigma \|\tilde{x} - s^k\|^2. \end{aligned}$$

If $\sigma_1 \leq \sigma_2$, then a σ_1 -approximate solution of (23)–(24) is also a σ_2 -approximate solution. In particular, 0-approximate solution, that is, the exact solution of system (23)–(24) is always a σ -approximate solution for any $\sigma \geq 0$. This implies that when T is maximal monotone and $c_k > 0$, then for any $\sigma \geq 0$, there is at least one σ -approximate solution of (23)–(24). For previous algorithmic applications of this definition and some of its theoretical properties see [30–33]

For a choice of parameters $\sigma \geq 0, \alpha_1, \dots, \alpha_m$, and β_1, \dots, β_m , the approximate LMM with an extragradient step generates infinite sequences $\{x^k\}, \{\tilde{x}^k\}, \{\tilde{v}^k\}, \{s^k\}$ as follows.

Algorithm 3

0. Choose $c_0, \dots, c_{m-1} > 0$ and $\{x^0, \dots, x^{m-1}\}, \{\tilde{v}^0, \dots, \tilde{v}^{m-1}\} \subset X$;

1. Set

$$s^k = \sum_{j=1}^m \alpha_j x^{k-j} - \beta_j c_{k-j} \tilde{v}^{k-j};$$

2. Choose $c_k > 0$ and find a σ -approximate solution $(\tilde{v}^k, \tilde{x}^k)$ of

$$\begin{aligned} v &\in T(x), \\ c_k v + x - s^k &= 0; \end{aligned}$$

3. Set

$$x^k = s^k - c_k \tilde{v}^k,$$

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

Note that step 2 can be implemented as

2. Choose $c_k > 0$ and find (\tilde{x}^k, e^k) such that

$$\begin{aligned} T^\epsilon(\tilde{x}^k) + c_k^{-1}(\tilde{x}^k - s^k) &\ni c_k^{-1}e^k \\ \|e^k\|^2 + 2c_k\epsilon &\leq \sigma \|\tilde{x}^k - s^k\|^2, \end{aligned}$$

$$\text{set } \tilde{v}^k = c_k^{-1}(s^k + e^k - \tilde{x}^k),$$

which shows its advantages over the classical approximation schemes where the error tolerance is driven to zero as the algorithm proceeds.

With error tolerance $\sigma = 0$, Algorithm 3 reduces to the exact LMM. Also, as noted above, when T is maximal monotone, there always exists a σ -approximate solution to the proximal system in step 2, so the method is well defined. To obtain convergence of the generated sequence $\{x^k\}$ to a solution of $T(x) \ni 0$ we also need to impose conditions on σ , and on the α_j s and β_j s.

Theorem 2. Consider Algorithm 3 and assume that

1. T is maximal monotone and $T^{-1}(0) \neq \emptyset$;
2. $\sigma \geq 0$, $\inf c_k > 0$;
3. $\alpha_j \geq 0 \forall j = 1, \dots, m$, $\sum_{j=1}^m \alpha_j = 1$, and the method is strongly stable;
4. $-\alpha_j \leq \beta_j \leq \frac{1-\sqrt{\sigma}}{1+\sqrt{\sigma}}\alpha_j \forall j = 1, \dots, m$, where the inequalities are strict for at least one j .

Then for any starting points, the sequences $\{x^k\}$, $\{\tilde{x}^k\}$, and $\{s^k\}$ converge weakly to a point in $T^{-1}(0)$, and the sequence $\{c_k \tilde{v}^k\}$ converges strongly to zero.

Proof. Let $\bar{x} \in T^{-1}(0)$ and $k \geq 2m$. Since $x^k = s^k - c_k \tilde{v}^k$ for $k \geq m$, we have from steps 1 and 2

$$s^k = \sum_{j \in \mathcal{J}} (\alpha_j x^{k-j} - \beta_j c_{k-j} \tilde{v}^{k-j}) = \sum_{j \in \mathcal{J}} [\alpha_j s^{k-j} - (\alpha_j + \beta_j) c_{k-j} \tilde{v}^{k-j}],$$

where again $\mathcal{J} = \{j \mid \alpha_j \neq 0\}$. Then just like in the proof of Theorem 1 we get

$$\begin{aligned} \|s^k - \bar{x}\|^2 &= \sum_{j \in \mathcal{J}} \alpha_j \|s^{k-j} - \bar{x}\|^2 \\ &\quad + \sum_{j \in \mathcal{J}} (\alpha_j + \beta_j) \left[\frac{\alpha_j + \beta_j}{\alpha_j} \|c_{k-j} \tilde{v}^{k-j}\|^2 - 2 \langle s^{k-j} - \bar{x}, c_{k-j} \tilde{v}^{k-j} \rangle \right]. \end{aligned} \tag{25}$$

To get an upper bound for the last term in (25), we first note that step 2 and Definition 2 imply for each $i \geq m$, the existence of an $\varepsilon_i \geq 0$ such that

$$\tilde{v}^i \in T^{\varepsilon_i}(\tilde{x}^i), \tag{26}$$

$$\|c_i \tilde{v}^i + \tilde{x}^i - s^i\|^2 + 2c_i \varepsilon_i \leq \sigma \|\tilde{x}^i - s^i\|^2. \tag{27}$$

The bound (27) can be written as

$$\|c_i \tilde{v}^i\|^2 + \|\tilde{x}^i - s^i\|^2 - 2 \left(\langle c_i \tilde{v}^i, s^i - \tilde{x}^i \rangle - c_i \varepsilon_i \right) \leq \sigma \|\tilde{x}^i - s^i\|^2,$$

and since $0 \in T(\bar{x})$ we get from (22) that

$$\begin{aligned} \langle s^i - \bar{x}, c_i \tilde{v}^i \rangle &= \langle s^i - \tilde{x}^i, c_i \tilde{v}^i \rangle + \langle \tilde{x}^i - \bar{x}, c_i \tilde{v}^i \rangle \\ &\geq \langle s^i - \tilde{x}^i, c_i \tilde{v}^i \rangle - c_i \varepsilon_i. \end{aligned}$$

Combining, we get

$$2 \langle s^i - \bar{x}, c_i \tilde{v}^i \rangle \geq (1 - \sigma) \|\tilde{x}^i - s^i\|^2 + \|c_i \tilde{v}^i\|^2 \quad \forall i \geq m. \tag{28}$$

By (27)

$$\|c_i \tilde{v}^i\| \leq (1 + \sqrt{\sigma}) \|\tilde{x}^i - s^i\|, \tag{29}$$

which can be used in (28) to get

$$2 \langle s^i - \bar{x}, c_i \tilde{v}^i \rangle \geq \frac{1 - \sigma}{(1 + \sqrt{\sigma})^2} \|c_i \tilde{v}^i\|^2 + \|c_i \tilde{v}^i\|^2 = \left(\frac{1 - \sqrt{\sigma}}{1 + \sqrt{\sigma}} + 1 \right) \|c_i \tilde{v}^i\|^2 \quad \forall i \geq m.$$

Since by condition 4, $\alpha_j + \beta_j \geq 0$, we can use this in (25) to obtain, after some manipulations

$$\|s^k - \bar{x}\|^2 \leq \sum_{j \in \mathcal{J}} \alpha_j \|s^{k-j} - \bar{x}\|^2 - \sum_{j \in \mathcal{J}} \mu_j \|c_{k-j} \tilde{v}^{k-j}\|^2, \tag{30}$$

where the constants

$$\mu_j = (\alpha_j + \beta_j) \left[\frac{1 - \sqrt{\sigma}}{1 + \sqrt{\sigma}} - \frac{\beta_j}{\alpha_j} \right]$$

are nonnegative by condition 4. Thus, by Lemma 1 the sequence $\{\|s^k - \bar{x}\|\}$ is convergent and

$$\sum_{k=m}^{\infty} \sum_{j \in \mathcal{J}} \mu_j \|c_{k-j} \tilde{v}^{k-j}\|^2 < \infty. \tag{31}$$

Reordering terms in (31) we get

$$\left(\sum_{j \in \mathcal{J}} \mu_j \right) \sum_{l=m}^{\infty} \|c_l \tilde{v}^l\|^2 < \infty,$$

where by our assumptions on β_j s

$$\sum_{j \in \mathcal{J}} \mu_j > 0,$$

so that

$$\|c_k \tilde{v}^k\| \rightarrow 0. \tag{32}$$

Since by (27) it holds for $k \geq m$ that $(1 - \sqrt{\sigma})\|\tilde{x}^k - s^k\| \leq \|c_k \tilde{v}^k\|$, we also get

$$\|s^k - \tilde{x}^k\| \rightarrow 0, \tag{33}$$

which together with the convergence of $\{\|s^k - \bar{x}\|\}$ implies the convergence of $\{\|\tilde{x}^k - \bar{x}\|\}$. To prove the weak convergence of $\{\tilde{x}^k\}$ to a point in $T^{-1}(0)$ it suffices by Lemma 2 to show that all its weak cluster points are in $T^{-1}(0)$.

Let \hat{x} be any weak cluster point of $\{\tilde{x}^k\}$, and let $\{\tilde{x}^{k_q}\}$ be a subsequence converging weakly to \hat{x} . By (26) $\tilde{v}^{k_q} \in T^{\varepsilon_{k_q}}(\tilde{x}^{k_q})$, where $\varepsilon_{k_q} \rightarrow 0$ by (27) and (33), and $\tilde{v}^{k_q} \rightarrow 0$ strongly by (32) and the fact that $c_k \geq c$ for all k . So from (22) we get

$$\langle 0 - u, \hat{x} - y \rangle \geq 0 \quad \forall (y, u) \in \text{gph } T,$$

which by maximality of T implies $0 \in T(\hat{x})$. Thus, $\{\tilde{x}^k\}$ converges weakly to a point in $T^{-1}(0)$. This holds also for $\{s^k\}$ by (33) and for $\{x^k\}$ by step 3 and (32). \square

Note that if we choose $m = 1$, $\alpha_1 = 1$ and $\beta_1 = 0$, Theorem 2 reduces to [31, Theorem 3.1].

4. A case of linear convergence

It was shown in [27] that under a local Lipschitz-type condition on the inverse of T the proximal point algorithm attains local linear convergence. The same holds for the hybrid extragradient-proximal point algorithm [31]. In this section we show that the Lipschitz condition implies the r -linear convergence of the LMM. We give the analysis only for the version of Section 3 under the additional simplification that $\epsilon_k = 0$ for all $k = 1, 2, \dots$. Using the same kind of analysis, similar results can be obtained also for the other versions of the LMM.

Proposition 1. *Consider Algorithm 3. In addition to the conditions of Theorem 2, assume that $\epsilon_k = 0$ for all k , that $T(x) \ni 0$ has a unique solution \bar{x} , and that for some neighborhood $U \ni 0$ and a constant $a \geq 0$*

$$y \in U, \quad x \in T^{-1}(y) \implies \|x - \bar{x}\| \leq a\|y\|. \tag{34}$$

Then $\{s^k\}$ converges r -linearly to \bar{x} .

Proof. By Definition 2, step 2 of Algorithm 3 implies that $\|\tilde{x}^k - s^k\| - \|c_k \tilde{v}^k\| \leq \sqrt{\sigma} \|\tilde{x}^k - s^k\|$ or

$$\|\tilde{x}^k - s^k\| \leq \frac{c_k}{1 - \sqrt{\sigma}} \|\tilde{v}^k\|.$$

Since $\tilde{v}^k \rightarrow 0$ by Theorem 2 and because we are assuming $\epsilon_k = 0$, condition (34) implies that

$$\|\tilde{x}^k - \bar{x}\| \leq a\|\tilde{v}^k\|,$$

for all k large enough. We thus get that for all k large enough

$$\|s^k - \bar{x}\| \leq \|\tilde{x}^k - s^k\| + \|\tilde{x}^k - \bar{x}\| \leq \left(\frac{c_k}{1 - \sqrt{\sigma}} + a \right) \|\tilde{v}^k\|,$$

or

$$\|c_k \tilde{v}^k\| \geq \left(\frac{1}{1 - \sqrt{\sigma}} + \frac{a}{c_k} \right)^{-1} \|s^k - \bar{x}\|.$$

Using this in (30) we get after some simplifications

$$\|s^k - \bar{x}\|^2 \leq \sum_{j \in \mathcal{J}} (1 - \theta_{k,j}) \alpha_j \|s^{k-j} - \bar{x}\|^2, \tag{35}$$

where again $\mathcal{J} = \{j \mid \alpha_j \neq 0\}$, and

$$\theta_{k,j} = \frac{\alpha_j + \beta_j}{\left(\frac{1}{1 - \sqrt{\sigma}} + \frac{a}{c_k}\right)^2 \alpha_j} \left(\frac{1 - \sqrt{\sigma}}{1 + \sqrt{\sigma}} - \frac{\beta_j}{\alpha_j} \right) \leq \frac{\alpha_j + \beta_j}{\alpha_j} \left(1 - \frac{\beta_j}{\alpha_j} \right) \leq 1.$$

If we define

$$\theta_j = \frac{\alpha_j + \beta_j}{\left(\frac{1}{1-\sqrt{\sigma}} + \frac{a}{\inf c_k}\right)^2 \alpha_j} \left(\frac{1 - \sqrt{\sigma}}{1 + \sqrt{\sigma}} - \frac{\beta_j}{\alpha_j} \right),$$

we have $0 \leq \theta_j \leq \theta_{k,j}$. Thus, letting A_θ be the companion matrix of the polynomial

$$p_\theta(z) = z^m - \sum_{j \in \mathcal{J}} (1 - \theta_j) \alpha_j z^{m-j},$$

we get from (35) that

$$d^{k+1} \leq A_\theta d^k$$

where $d^k = (\|s^{k-1} - \bar{x}\|^2, \dots, \|s^{k-m} - \bar{x}\|^2)$, and the inequality is interpreted componentwise. Since the elements of A and d^k are nonnegative this implies that for all $k \geq m$

$$d^{k+1} \leq A_\theta^{k-m} d^m,$$

so that

$$\|s^k - \bar{x}\|^2 \leq \|d^{k+1}\|_\infty \leq \|A_\theta^{k-m} d^m\|_\infty.$$

For each $\epsilon > 0$, there exists a norm $\|\cdot\|$ on \mathbb{R}^m (see for example [22, 1.3.6]) such that in the corresponding matrix norm

$$\|A_\theta\| \leq (\rho(A_\theta) + \epsilon).$$

By equivalence of norms on \mathbb{R}^m , we can then find a constant $C > 0$ such that

$$\|s^k - \bar{x}\|^2 \leq C \|A_\theta^{k-m} d^m\| \leq C \|A_\theta\|^{k-m} \|d^m\| \leq C(\rho(A_\theta) + \epsilon)^{k-m} \|d^m\|.$$

To finish the proof it suffices to show that $\rho(A_\theta) < 1$. So assume that λ is a complex number satisfying $p_\theta(\lambda) = 0$. Then

$$|\lambda|^m \leq \sum_{j \in \mathcal{J}} (1 - \theta_j) \alpha_j |\lambda|^{m-j}, \tag{36}$$

so $|\lambda| \geq 1$ would imply

$$|\lambda|^m \leq \sum_{j \in \mathcal{J}} (1 - \theta_j) \alpha_j |\lambda|^m \implies 1 \leq \sum_{j \in \mathcal{J}} (1 - \theta_j) \alpha_j,$$

which is a contradiction, since $\sum_{j \in \mathcal{J}} (1 - \theta_j) \alpha_j < 1$ by condition 4. □

References

1. Antipin, A.S.: Feedback-controlled second-order proximal differential systems. *Differential Equations* **29**, 1597–1607 (1993)
2. Antipin, A.S.: Second-order controlled differential gradient methods for solving equilibrium problems. *Differential Equations* **35**, 592–601 (1999)
3. Auslender, A., Teboulle, M.: Lagrangian duality and related multiplier methods for variational inequality problems. *SIAM J. Optim.* **10**, 1097–1115 (2000)
4. Alvarez, F., Attouch, H.: An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Wellposedness in optimization and related topics* (Gargnano, 1999). *Set-Valued Anal.* **9**, 3–11 (2001)
5. Brézis, H.: *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Mathematics Studies, No. 5. 1973
6. Bruck, R.E.: Asymptotic convergence of nonlinear contraction semigroups in Hilbert space. *J. Funct. Anal.* **18**, 15–26 (1975)
7. Burachik, R.S., Iusem, A.N., Svaiter, B.F.: Enlargements of maximal monotone operators with application to variational inequalities. *Set Valued Anal.* **5**, 159–180 (1997)
8. Burachik, R.S., Sagastizábal, C.A., Svaiter, B.F.: ϵ -enlargements of maximal monotone operators: theory and applications. In: M. Fukushima, L. Qi, (eds), *Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods* (Lausanne, 1997), vol. 22 of Applied Optimization, pp. 25–43. Kluwer Acad. Publ., Dordrecht, 1999
9. Burachik, R.S., Svaiter, B.F.: ϵ -enlargements of maximal monotone operators in Banach spaces, *Set-Valued Anal.* **7**(2), 117–132 (1999)
10. Correa, R., Lemaréchal, C.: Convergence of some algorithms for convex minimization. *Math. Program.* **62** Ser. B, 261–275 (1993)
11. Crandall, M.G., Pazy, A.: Semi-groups of nonlinear contractions and dissipative sets. *J. Funct. Anal.* **3**, 376–418 (1969)
12. Eckstein, J.: Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Math. Oper. Res.* **18**, 203–226 (1993)
13. Eckstein, J., Bertsekas, D.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.* **55**(3), Ser. A, 293–318 (1992)
14. Eckstein, J., Ferris, M.C.: Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control. *INFORMS J. Comput.* **10**(2), 218–235 (1998)
15. Iusem, A., Svaiter, B., Teboulle, M.: Entropy-like proximal methods in convex programming. *Math. Oper. Res.* **19**(4), 790–814 (1994)
16. Jain, M.K.: *Numerical Solution of Differential Equations*, Second edition. A Halsted Press Book. John Wiley & Sons, Inc., New York, 1984
17. Minty, G.J.: Monotone (nonlinear) operators in Hilbert space. *Duke Math. J.* **29**, 341–346 (1962)
18. Miranker, W.L.: Numerical methods for stiff equations and singular perturbation problems. *Mathematics and its Applications*, 5. D. Reidel Publishing Co., Dordrecht-Boston, Mass., 1981
19. Nedich, A.: A third-order continuous gradient projection method for minimization problems. *Differential Equations* **30**, 1767–1774 (1994)
20. Nevanlinna, O.: On the convergence of difference approximations to nonlinear contraction semigroups in Hilbert spaces. *Math. Comp.* **32**(142), 321–334 (1978)
21. Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc.* **73**, 591–597 (1967)
22. Ortega, J.M.: Numerical analysis. A second course. *Computer Science and Applied Mathematics*. Academic Press, New York-London, 1972
23. Polyak, B.T.: Introduction to optimization. Optimization Software Inc. Publications Division, New York, 1987. Translated from the Russian, With a foreword by Dimitri P. Bertsekas
24. Revalski, J.P., Théra, M.: Generalized sums of monotone operators. *C. R. Acad. Sci. Paris Sr. I Math.* **329**(11), 979–984 (1999)
25. Revalski, J.P., Théra, M.: Enlargements and sums of monotone operators. *Nonlinear Anal. Theory Meth. Appl.* **48**, 505–519 (2002)
26. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.* **1**, 97–116 (1976)
27. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **14**, 877–898 (1976)
28. Rockafellar, R.T., Wets, R.J.-B.: Variational analysis. *Grundlehren der Mathematischen Wissenschaften*, 317, Springer-Verlag, Berlin, 1998

29. Solodov, M.V., Svaiter, B.F.: Hybrid projection – proximal point algorithm. *J. Convex Anal.* **6**, 59–70 (1999)
30. Solodov, M.V., Svaiter, B.F.: A globally convergent inexact Newton method for systems of monotone equations. In: M. Fukushima, L. Qi (eds), *Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods (Lausanne, 1997)*, vol. 22 of Applied Optimization, pp. 25–43. Kluwer Acad. Publ., Dordrecht, 1999
31. Solodov, M.V., Svaiter, B.F.: An inexact hybrid extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.* **7**(4), 323–345 (1999)
32. Solodov, M.V., Svaiter, B.F.: Error bounds for proximal point subproblems and associated inexact proximal point algorithms. *Math. Program.* **7**(4), 323–345 (1999)
33. Solodov, M.V., Svaiter, B.F.: A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem. *SIAM J. Optim.* **10**, 605–625 (2000) (electronic)
34. Solodov, M.V., Svaiter, B.F.: A unified framework for some inexact proximal point algorithms. *Numer. Funct. Anal. Optim.* **22**(7–8), 1013–1035 (2001)
35. Spingarn, J.E.: Partial inverse of a monotone operator. *Appl. Math. Optim.* **10**, 247–265 (1983)
36. Svaiter, B.F.: A family of enlargements of maximal monotone operators. *Set-Valued Anal.* **8**, 11–328 (2000)
37. Tossings, P.: The perturbed proximal point algorithm and some of its applications. *Appl. Math. Optim.* **29**, 125–159 (1994)