Noname manuscript No.
(will be inserted by the editor)

A Variant of the Hybrid Proximal Extragradient Method for Solving Strongly Monotone Inclusions and its Complexity Analysis

M. Marques Alves · B. F. Svaiter

Received: date / Accepted: date

Abstract This paper presents and studies the iteration-complexity of a variant of the hybrid proximal extragradient method for solving inclusion problems with strongly (maximal) monotone operators. As applications, we propose and analyze two special cases: variants of the Tseng's forward-backward method for solving monotone inclusions with strongly monotone and Lipschitz continuous operators and of the Korpelevich extragradient method for solving (strongly monotone) variational inequalities.

Keywords Hybrid proximal extragradient method · Strongly monotone operators · Variational inequalities · Tseng's forward-backward method · Korpelevich extragradient method

Mathematics Subject Classification (2000) 47H05 · 47J20 · 90C060 · 90C33 · 65K10

1 Introduction

Monotone inclusion problems (MIPs) are inclusion problems for maximal monotone point-to-set operators and occur in different fields of applied mathematics, including optimization, equilibrium theory and variational inequalities. A classical scheme for solving MIPs is the proximal point method (PPM), proposed by Martinet [1] and further developed by Rockafellar [2], an iterative method where the current iterate is used to construct a regularized version of the original problem, namely, the proximal subproblem, whose approximate solution is taken as the next iterate. Convergence of the generated sequence requires summable errors, that is, the sum of the errors in the computation of the approximate solution of each proximal subproblem must be finite.

The hybrid proximal extragradient (HPE) method [3,4], another scheme for solving MIPs, proposed by Solodov and Svaiter, is a modification of the PPM which, instead of summable errors, requires for its convergence each proximal subproblem to be approximately solved within a relative error tolerance. An additional feature of the HPE is that it also allows the relaxation of the inclusion by means of the ε-enlargement of maximal monotone operators proposed in [5]. Complexity iteration of the HPE method was determined by Monteiro and Svaiter in [6], while the use of the HPE method as a framework for the design of new methods as well as for the complexity analysis of existing methods was presented in [6–14]. In particular, pointwise and ergodic iteration complexity of Tseng's modified forward-backward splitting (MFBS) method, Kopelevich's method and the alternating direction method of multipliers (ADMM) in [6,12].

M. Marques Alves (corresponding author)
Departamento de Matemática, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 88040-900. This work was done while this author was a postdoc at the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205.
Tel.: +55 48 37213678
Fax: +55 48 37216560
E-mail: maicon.alves@ufsc.br

B. F. Svaiter
IMPA, Estrada Dona Castorina 110, 22460-320 Rio de Janeiro, Brazil.
Tel.: +55 21 2529 5000
Fax: +55 21 2512 4115
E-mail: benar@impa.br

The aim of this work is to present a variant of the HPE specially adapted for solving MIPs for operators that can be decomposed as the sum of a strongly monotone and a monotone operator. MIPs are, in general, ill-posed; however, MIPs for strongly monotone operators are always well-posed. For this reason, problems in the first class have been approximated, in practical applications, by problems in the latter one by regularization procedures, a technique which goes back to the work of Tikhonov [15]. Hence, it is interesting to design numerical schemes for strongly monotone MIPs which are able to exploit the strong monotonicity of the problem under consideration. With this in mind, we modify the HPE's error criterion and iteration. Whereas the HPE method defines the new iterate as an extragradient step from the current iterate using the image of the (enlargement of the) operator at the approximate solution of the proximal (prox.) subproblem, the method proposed here defines the new iterate as a convex combination of the extragradient step and the approximate solution of the prox. subproblem. When the strong monotonicity parameter is zero, our algorithm reduces to the HPE method. Like in the HPE theory, we also make use of the notion of ε-enlargement of Burachik, Iusem and Svaiter [5] to relax the inclusion in each subproblem. Likewise, ε-enlargements are used, as in [6], to define the concept of approximate solution that we will consider in this paper. Moreover, based on a family of simple quadratic functions, defined by points in the graph of the strongly monotone component of our operator, we obtain several descent results (see Proposition 3.1) for the sequence generated by the method proposed here.

As applications, we consider two special MIPs in which the strongly monotone component is also (point-to-point) Lipschitz continuous, one in which the monotone component is an arbitrary maximal monotone operator and the other one in which the monotone component is a normal cone. For the first one, we propose and analyze a variant of the Tseng's MFBS method, which exploits the strong monotonicity assumption (in the case that the strong monotonicity parameter is zero, it reduces to the Tseng's method). We obtain rates of convergence and iteration-complexity results by showing that this method is a special case of our method. On the other hand, in the case of variational inequality problems (VIPs) for strongly monotone operators, we propose and analyze a variant of the Korpelevich extragradient method.

Nesterov and Scrimalli proposed and analyzed in [16] a method for solving strongly monotone variational inequalities with Lipschitz continuous operators. It worth to mention that, while the variant of the Tseng's MFBS method proposed here computes one resolvent/projection per iteration, their method computes two projections per iteration. Moreover, both methods have the same iteration-complexity.

This paper is organized as follows. In Section 2, we review some basic concepts on set-valued maps, maximal monotone operators, ε-enlargements, subdifferentials as well as some of their properties which we will use in this paper. In Section 3, we present a variant of the HPE method for strongly MIPs, namely Algorithm 1, and establish its convergence rates for general stepsizes as well as for stepsizes bounded away from zero. Section 4 is devoted to the presentation, convergence analysis, and iteration-complexity analysis of a variant of the Tseng's MFBS method for solving MIPs with strongly monotone operators. In Section 5, we present a variant of the Korpelevich extragradient method for solving strongly monotone VIPs.

2 Basic Concepts and Notation

Let X be a real Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and induced norm $\|\cdot\| := \sqrt{\langle\cdot,\cdot\rangle}$. Given a set-valued operator $S : X \rightrightarrows X$, its *graph* and *domain* are, respectively,

$$\mathrm{Gr}(S) := \{(x, v) \in X \times X \,:\, v \in S(x)\}, \quad \mathrm{Dom}(S) := \{x \in X \,:\, S(x) \neq \emptyset\}.$$

The *inverse* of $S : X \rightrightarrows X$ is $S^{-1} : X \rightrightarrows X$, $S^{-1}(v) = \{x \,:\, v \in S(x)\}$.

In this work, we are concerned with algorithms for solving strongly monotone inclusion problems, that is, inclusion problems for strongly (maximal) monotone operators, a special class of set-valued operators with several applications in applied mathematics and optimization. In what follows in this section, we will review some of the basic concepts of this class of operators.

An operator $A : X \rightrightarrows X$ is *η-strongly monotone* iff $\eta \geq 0$ and

$$\langle v - v', x - x'\rangle \geq \eta\|x - x'\|^2 \quad \forall (x, v), (x', v') \in \mathrm{Gr}(A). \tag{1}$$

If $\eta = 0$ in the above inequality, then A is said to be a *monotone* operator. Moreover, $A : X \rightrightarrows X$ is *maximal monotone* iff it is monotone and maximal in the following sense: if $B : X \rightrightarrows X$ is monotone and $\mathrm{Gr}(A) \subset \mathrm{Gr}(B)$, then $A = B$. The *resolvent* of a maximal monotone operator $A : X \rightrightarrows X$ with parameter $\lambda > 0$ is $(I + \lambda A)^{-1}$. It follows directly from this definition that $y = (I + \lambda A)^{-1}x$ if and only if $x - y \in \lambda A(y)$. The *sum* of two set-valued operators $S, S' : X \rightrightarrows X$ is defined by

$$S + S' : X \rightrightarrows X, \quad (S + S')(x) := \{s + s' \in X \, : \, s \in S(x), \; s' \in S(x)\}.$$

It is easy to see that, if $A : X \rightrightarrows X$ is η-strongly monotone and $B : X \rightrightarrows X$ is monotone, then the sum $A + B$ is also η-strongly monotone. In particular, the sum of two monotone operators is also a monotone operator.

Recall that the ε-*subdifferential* [17] of a closed convex function $f : X \to \overline{\mathbb{R}}$ is defined at $x \in X$ as

$$\partial_\varepsilon f(x) := \{v \in X \, : \, f(x') \geq f(x) + \langle v, x' - x\rangle - \varepsilon \;\; \forall x' \in X\}.$$

When $\varepsilon = 0$, then $\partial f_0(x)$ is denoted by $\partial f(x)$ and is called the *subdifferential* of f at x. The set-valued operator $\partial_\varepsilon f : X \rightrightarrows X$ with $\varepsilon \geq 0$ is an enlargement of ∂f (in the sense that $\partial f(x) \subset \partial f_\varepsilon(x)$ for every $x \in X$) which has better topological properties than ∂f (see [17]). The simplest example of subdifferential is given by considering indicator functions of closed convex sets. Given a closed convex set $C \subset X$, its *indicator function* is denoted by δ_C and is defined as $\delta_C(x) = 0$ if $x \in C$ and $\delta_C(x) = \infty$ otherwise, and its *normal cone* is defined as $N_C = \partial \delta_C$. We also denote the projection on C by P_C.

Using the above definitions, it is easy to check the following *transportation formula*:

$$v \in \partial f(x) \;\; \Rightarrow v \in \partial_\varepsilon f(x'), \qquad \varepsilon := f(x') - [f(x) + \langle v, x' - x\rangle]. \tag{2}$$

A generalization of the concept of ε-enlargement (of subgradients of convex functions) for arbitrary maximal monotone operators was introduced and studied in [5]. Given $B : X \rightrightarrows X$ maximal monotone and $\varepsilon \geq 0$, the ε-*enlargement* of B is defined by

$$B^{[\varepsilon]} : X \rightrightarrows X, \quad B^{[\varepsilon]}(x) := \{v \in X \, : \, \langle v - v', x - x'\rangle \geq -\varepsilon \;\; \forall (x', v') \in \mathrm{Gr}(B)\}. \tag{3}$$

The following summarizes some useful properties of $B^{[\varepsilon]}$.

Proposition 2.1 *Let $A, B : X \rightrightarrows X$ be maximal monotone operators. Then,*

(a) if $\varepsilon_1 \leq \varepsilon_2$, then $A^{[\varepsilon_1]}(x) \subseteq A^{[\varepsilon_2]}(x)$ for every $x \in X$;
(b) $A^{[\varepsilon']}(x) + (B)^{[\varepsilon]}(x) \subseteq (A + B)^{[\varepsilon' + \varepsilon]}(x)$ for every $x \in X$ and $\varepsilon, \varepsilon' \geq 0$;
(c) A is monotone if, and only if, $A \subseteq A^{[0]}$;
(d) A is maximal monotone if, and only if, $A = A^{[0]}$;
(e) if $f : X \to \overline{\mathbb{R}}$ is convex, proper and lower semicontinuous, then $\partial_\varepsilon f(x) \subset (\partial f)^{[\varepsilon]}(x)$ for any $\varepsilon \geq 0$ and $x \in X$.

3 Solving Inclusions with Strongly Monotone Operators

In this section, we consider the monotone inclusion problem (MIP)

$$0 \in A(x) + B(x), \tag{4}$$

where the following assumptions hold:

A.1) $A : X \rightrightarrows X$ is (maximal) η-strongly monotone, i.e, it is maximal monotone and there exists $\eta \geq 0$ satisfying the condition (1);
A.2) $B : X \rightrightarrows X$ is maximal monotone;
A.3) the solution set of (4), i.e., $(A + B)^{-1}(0)$, is nonempty.

Since by the assumptions A.1 and A.2 the sum $A + B$ is also η-strongly monotone, it follows that (4) is an inclusion for a strongly monotone operator, i.e., a *strongly MIP*.

The complexity results presented in this paper will consist in establishing bounds in the number of iterations to obtain a pair (y, v) and a scalar $\varepsilon \geq 0$ such that

$$v \in A(y) + B^{[\varepsilon]}(y), \quad \|v\| \leq \bar{\rho} \;\text{ and }\; \varepsilon \leq \bar{\varepsilon}, \tag{5}$$

for given precisions $\bar{\rho} > 0$ and $\bar{\varepsilon} > 0$. Such bounds will depend on the distance of the initial iterate (for the methods presented here) to the solution set of (4) and on the parameter η. Stronger bounds will be reached for the case that A is η-strongly monotone with parameter $\eta > 0$. On the other hand, in the case $\eta = 0$, (or in the "limit" case $\eta \to 0$) our results will be special cases of the ones obtained in [6].

An *exact* proximal point method (PPM) iteration for problem (4) is

$$x_k = (\lambda_k(A+B)+I)^{-1}x_{k-1}$$

where x_{k-1} is the current iterate, x_k the new iterate, and $\lambda_k > 0$. To compute $(\lambda(A+B)+I)^{-1}$ for $\lambda > 0$ is, in general, almost as hard (or as hard) as computing $(A+B)^{-1}$. So, the use of approximate solutions of this problem must be considered. To decouple the inclusion and the equality, note that the above equation is equivalent to

$$v_k \in (A+B)(x_k), \quad \lambda_k v_k + x_k - x_{k-1} = 0. \tag{6}$$

An HPE iteration for problem (4) is

$$\begin{aligned}&\text{find } \lambda_k > 0,\ y_k, v_k \text{ and } \varepsilon_k \text{ such that}\\ &\qquad v_k \in (A+B)^{[\varepsilon_k]}(y_k), \qquad \|\lambda_k v_k + y_k - x_{k-1}\|^2 + 2\lambda_k\varepsilon_k \le \sigma^2\|y_k - x_{k-1}\|^2,\\ &\text{define } x_k = x_{k-1} - \lambda_k v_k.\end{aligned}$$

where $0 \le \sigma < 1$ is a relative error tolerance. The scalar ε_k may be implicitly determined (see the analysis of Korpelevich's method in [6]). In this paper, we propose a variant of the HPE method, namely Algorithm 1, which takes into account the fact that the operator A in (4) is η-strongly monotone.

We will now state our method.

Algorithm 1: A HPE method for strongly monotone inclusions

(0) Let $x_0 \in X$ and $\sigma \in [0,1[$ be given and set $k = 1$;

(1) choose $\lambda_k > 0$ and find $y_k, v_k \in X$, $\sigma_k \in [0,\sigma]$, and $\varepsilon_k \ge 0$ such that

$$v_k \in A(y_k) + B^{[\varepsilon_k]}(y_k)\,, \quad \frac{\|\lambda_k v_k + y_k - x_{k-1}\|^2}{1+2\lambda_k\eta} + 2\lambda_k\varepsilon_k \le \sigma_k^2\|y_k - x_{k-1}\|^2; \tag{7}$$

(2) set

$$x_k = \frac{(x_{k-1} - \lambda_k v_k) + 2\lambda_k\eta y_k}{1+2\lambda_k\eta}, \tag{8}$$

let $k \leftarrow k+1$ and go to step 1.

end

Note that Algorithm 1 does not specify how to compute λ_k, σ_k, ε_k, y_k, and v_k satisfying (7). This indetermination adds generality to this method and we will be able, almost effortlessly, to define two new implementable algorithms for solving strongly monotone inclusion problems in Sections 4 and 5, namely, a Tseng's MFBS-like method for (4) when A is also Lipschitz continuous and the resolvent of B is easy to compute, and a Korpelevich's-like method for strongly variational inequalities with Lipschitz continuous strongly monotone operators.

As in the HPE method, the error criterion (7) relaxes the inclusion and the equation in (6) by allowing errors relative to the term $\|y_k - x_{k-1}\|$, However, the new iterate x_k is given in (8) as a convex combination of the extragradient step $x_{k-1} - \lambda_k v_k$ and the vector y_k. Since by Proposition 2.1(b,d) $A(y_k) + B^{[\varepsilon_k]}(y_k) \subset (A+B)^{[\varepsilon_k]}(y_k)$, if $\eta = 0$ then Algorithm 1 reduces to the HPE method applied to (4), for which convergence and complexity results were presented, respectively, in [3] and [6].

From now on in this section, $\{x_k\}$, $\{y_k\}$, $\{v_k\}$, $\{\lambda_k\}, \{\sigma_k\}$ and $\{\varepsilon_k\}$ are sequences generated by Algorithm 1. Define, for $k = 1, 2, \dots$

$$\gamma_k : X \to \mathbb{R}, \ \ \gamma_k(x') := \langle v_k, x' - y_k\rangle + \eta\|x' - y_k\|^2 - \varepsilon_k \quad \forall x' \in X. \tag{9}$$

Algorithm 1 is a "memory-less" method, in the sense that the unique information which is passed along the iterations is the current iterate. However, for analyzing it, we will define objects which gather aggregate data generated during the past iterations. Define also

$$\mu_0 := 1, \quad \mu_k := \prod_{i=1}^{k}(1+2\lambda_i\eta)\,. \tag{10}$$

In next proposition, we obtain some descent bounds for Algorithm 1. Special attention should be given for the item (d), which will be crucial for obtaining pointwise convergence rates.

Proposition 3.1 *Let $\gamma_k(\cdot)$ and μ_k be as in (9) and (10), respectively. For every $k \geq 1$:*

(a) $x_k = \arg\min \lambda_k \gamma_k(x') + \|x' - x_{k-1}\|^2/2$;
(b) $\min \lambda_k \gamma_k(x') + \|x' - x_{k-1}\|^2/2 \geq (1-\sigma^2)\|y_k - x_{k-1}\|^2/2$;
(c) $\gamma_k(x^*) \leq 0$ *for any* $x^* \in (A+B)^{-1}(0)$;
(d) for any $x^* \in (A+B)^{-1}(0)$,

$$(1+2\lambda_k\eta)\|x^* - x_k\|^2 + (1-\sigma^2)\|y_k - x_{k-1}\|^2 \leq \|x^* - x_{k-1}\|^2; \tag{11}$$

$$\mu_k\|x^* - x_k\|^2 + (1-\sigma^2)\sum_{j=1}^{k}\mu_{j-1}\|y_j - x_{j-1}\|^2 \leq \|x^* - x_0\|^2. \tag{12}$$

Proof (a) This statement follows trivially from (8) and (9).

(b) Letting $\alpha_k = 1/(1+2\lambda_k\eta)$ and using (8) we have that

$$x_k = \alpha_k(x_{k-1} - \lambda_k v_k) + (1-\alpha_k)y_k,$$

which together with the definition of $\|\cdot\|^2$ yields

$$\|x_k - x_{k-1}\|^2 = \alpha_k\|\lambda_k v_k\|^2 + (1-\alpha_k)\|y_k - x_{k-1}\|^2 - \alpha_k(1-\alpha_k)\|\lambda_k v_k + y_k - x_{k-1}\|^2\,.$$

Direct use of (a), condition (9), the latter identity and the fact that $\lambda_k\alpha_k\eta = (1-\alpha_k)/2$ give

$$\begin{aligned}\min \lambda_k\gamma_k(x') + \frac{1}{2}\|x' - x_{k-1}\|^2 &= \lambda_k\gamma_k(x_k) + \frac{1}{2}\|x_k - x_{k-1}\|^2 \\ &= -\frac{\alpha_k}{2}\left(\|\lambda_k v_k\|^2 + 2\langle \lambda_k v_k, y_k - x_{k-1}\rangle\right) + \frac{1-\alpha_k}{2}\|y_k - x_{k-1}\|^2 - \lambda_k\varepsilon_k \\ &= \frac{1}{2}\left(\|y_k - x_{k-1}\|^2 - \left[\frac{\|\lambda_k v_k + y_k - x_{k-1}\|^2}{1+2\lambda_k\eta} + 2\lambda_k\epsilon_k\right]\right),\end{aligned}$$

which, combined with the inequality in (7) proves (b).

(c) Taking $x^* \in (A+B)^{-1}(0)$ and using the inclusion in (7) we conclude that there exists $a_k, a^* \in X$ satisfying

$$a^* \in A(x^*), \quad a_k \in A(y_k), \quad v_k - a_k \in B^{[\varepsilon_k]}(y_k), \quad -a^* \in B(x^*).$$

Using the latter inclusions, assumption A.1 and (3) we find

$$\langle a^* - a_k, x^* - y_k\rangle \geq \eta\|x^* - y_k\|^2, \qquad \langle -(a^* - a_k) - v_k, x^* - y_k\rangle \geq -\varepsilon_k\,.$$

Adding the two above inequalities and using (9) we obtain the inequality in (c).

(d) Let $x^* \in (A+B)^{-1}(0)$. We will first show that (11) holds. In view of (9) and (a) and (b) we obtain, for all $x' \in X$,

$$\begin{aligned}\lambda_k\gamma_k(x') + \frac{1}{2}\|x' - x_{k-1}\|^2 &= \left(\min \lambda_k\gamma_k(x') + \frac{1}{2}\|x' - x_{k-1}\|^2\right) + \frac{1+2\lambda_k\eta}{2}\|x' - x_k\|^2 \\ &\geq \frac{1}{2}\Big((1-\sigma^2)\|y_k - x_{k-1}\|^2 + (1+2\lambda_k\eta)\|x' - x_k\|^2\Big)\,.\end{aligned}$$

To finish the proof of (11), use (c) and the latter inequality with $x' = x^*$.

We will now use a induction argument to show that (12) follows from (11). Indeed, using (10) and (11) for $k = 1$, we conclude that (12) holds for $k = 1$.

Assume that (12) holds for some $k = m \geq 1$, that is,

$$\mu_m\|x^* - x_m\|^2 + (1-\sigma^2)\sum_{j=1}^{m}\mu_{j-1}\|y_j - x_{j-1}\|^2 \leq \|x^* - x_0\|^2.$$

Multiplying (11) evaluated at $k = m+1$ by μ_m and using (10) we get

$$\mu_{m+1}\|x^* - x_{m+1}\|^2 + (1-\sigma^2)\mu_m\|y_{m+1} - x_m\|^2 \leq \mu_m\|x^* - x_m\|^2.$$

Adding the two displayed equations we conclude that (12) also holds $k = m+1$, which completes the induction proof. □

Lemma 3.1 *For $k = 1, 2, \ldots$*

$$\left(1 - \sigma_k\sqrt{1+2\lambda_k\eta}\right)\|y_k - x_{k-1}\| \le \|\lambda_k v_k\| \le \left(1 + \sigma_k\sqrt{1+2\lambda_k\eta}\right)\|y_k - x_{k-1}\|.$$

Proof Use the inequality in (7), triangle inequality and the fact that $\varepsilon_k \ge 0$. □

In the next theorem, we establish rates of convergence for the sequences $\{x_k\}$, $\{v_k\}$ and $\{\varepsilon_k\}$ generated by Algorithm 1.

Theorem 3.1 *Let $\{\mu_k\}$ be as in (10). Let also d_0 be the distance of x_0 to the solution set of (4) and let $x^* \in (A+B)^{-1}(0)$.*

Then, the following statements hold true:

(a) For every $k \ge 1$, $v_k \in A(y_k) + B^{[\varepsilon_k]}(y_k)$ and

$$\|v_k\| \le \frac{d_0\left(1+\sigma\sqrt{1+2\lambda_k\eta}\right)}{\lambda_k\sqrt{(1-\sigma^2)\mu_{k-1}}}\ ,\qquad \varepsilon_k \le \frac{\sigma^2 d_0^2}{2(1-\sigma^2)\lambda_k\mu_{k-1}}, \tag{13}$$

$$\|x^* - x_k\| \le \frac{1}{\sqrt{\mu_k}}\|x^* - x_0\|; \tag{14}$$

(b) for every $k \ge 1$, there exists $i \le k$ such that $v_i \in A(y_i) + B^{[\varepsilon_i]}(y_i)$ and

$$\|v_i\| \le \frac{d_0}{\sqrt{(1-\sigma^2)\sum_{j=1}^k \frac{\lambda_j^2\mu_{j-1}}{\left(1+\sigma\sqrt{1+2\lambda_j\eta}\right)^2}}},\qquad \varepsilon_i \le \frac{\sigma^2 d_0^2\lambda_i}{2(1-\sigma^2)\sum_{j=1}^k \lambda_j^2\mu_{j-1}}. \tag{15}$$

Proof Note that the inclusions in (a) and (b) follow directly from (7). To prove (a), first note that in view of (11) and (12), we obtain the following inequalities:

$$\|y_k - x_{k-1}\| \le \frac{1}{\sqrt{1-\sigma^2}}\|x^* - x_{k-1}\|,\quad \|x^* - x_{k-1}\| \le \frac{1}{\sqrt{\mu_{k-1}}}\|x^* - x_0\|,$$

which in turn imply that

$$\|y_k - x_{k-1}\| \le \frac{d_0}{\sqrt{(1-\sigma^2)\mu_{k-1}}}. \tag{16}$$

From the inequality in (7) and the fact that $\sigma_k \le \sigma$ we obtain

$$\varepsilon_k \le \frac{\sigma^2}{2\lambda_k}\|y_k - x_{k-1}\|^2. \tag{17}$$

Now, note that (13) follows from (16), (17) and the second inequality in Lemma 3.1. Moreover, (14) is a direct consequence of (12).

(b) Note first that from (12) and Lemma 3.1 we find

$$\sum_{j=1}^k \frac{\lambda_j^2\mu_{j-1}}{\left(1+\sigma\sqrt{1+2\lambda_j\eta}\right)^2}\|v_j\|^2 \le \frac{d_0^2}{1-\sigma^2},$$

and hence

$$\|v_i\|^2 := \min_{j=1,\ldots,k}\|v_j\|^2 \le \frac{d_0^2}{(1-\sigma^2)\sum_{j=1}^k \frac{\lambda_j^2\mu_{j-1}}{\left(1+\sigma\sqrt{1+2\lambda_j\eta}\right)^2}},$$

which, in turn, gives the first bound in (15). Likewise, using (17) and (12) we obtain

$$\sum_{j=1}^k \frac{2\lambda_j^2\mu_{j-1}}{\sigma^2}(\lambda_j^{-1}\varepsilon_j) = \sum_{j=1}^k \frac{2\lambda_j\mu_{j-1}}{\sigma^2}\varepsilon_j \le \frac{d_0^2}{1-\sigma^2},$$

and, consequently, the second inequality in (15). □

We observed in the paragraph following the statement of Algorithm 1 that if $\eta = 0$, then Algorithm 1 reduces to the HPE method [3] applied to the problem (4), for which a complexity analysis was established in [6]. Hence, it worths to note that in the case $\eta = 0$, Theorem 3.1(b) corresponds to Theorem 4.4(b) of [6].

In the next corollary, we assume that the sequence $\{\lambda_k\}$ is bounded away from zero.

Corollary 3.1 *In addition to the assumptions of Theorem 3.1, assume that* $\lambda_k \geq \underline{\lambda} > 0$ *for all* $k \geq 1$. *Then, the following statements hold for every* $k \geq 1$*:*

(a) $v_k \in A(y_k) + B^{[\varepsilon_k]}(y_k)$ *and*

$$\|v_k\| \leq d_0 \sqrt{\frac{1+\sigma}{1-\sigma}\left(\frac{\underline{\lambda}^{-2}}{(1+2\underline{\lambda}\eta)^{k-2}}\right)}\,, \quad \varepsilon_k \leq \frac{\sigma^2 d_0^2}{2(1-\sigma^2)\underline{\lambda}(1+2\underline{\lambda}\eta)^{k-1}}, \tag{18}$$

$$\|x^* - x_k\| \leq \sqrt{\frac{1}{(1+2\underline{\lambda}\eta)^k}}\|x^* - x_0\|; \tag{19}$$

(b) if $\eta > 0$ *in assumption A.1, then there exists* $i \leq k$ *such that* $v_i \in A(y_i) + B^{[\varepsilon_i]}(y_i)$ *and*

$$\|v_i\| \leq d_0 \sqrt{\frac{1+\sigma}{1-\sigma}\left(\frac{2\eta\underline{\lambda}^{-1}(1+2\underline{\lambda}\eta)}{(1+2\underline{\lambda}\eta)^k - 1}\right)}, \qquad \varepsilon_i \leq \frac{\eta\sigma^2 d_0^2}{(1-\sigma^2)\left[(1+2\underline{\lambda}\eta)^k - 1\right]}. \tag{20}$$

Proof We start by noting that the inclusions in (a) and (b) follows from the corresponding ones in itens (a) and (b) of Theorem 3.1. Moreover, using (10) and the assumption that $\lambda_k \geq \underline{\lambda}$ for all $k \geq 1$ we obtain

$$\mu_k \geq (1+2\underline{\lambda}\eta)^k \quad \forall k \geq 0. \tag{21}$$

(a) The scalar function $\psi : t \mapsto t^{-1}\left(1+\sigma\sqrt{1+2t\eta}\right)$ is nonincreasing in $]0,\infty[$ and hence $\psi(\lambda_k) \leq \psi(\underline{\lambda})$ for all $k \geq 1$, which in turn combined with (13), (14) and (21) gives (a).

(b) The first inequality in (20) follows from the first inequality in (15), the fact that the scalar function $1/\psi$ is nondecreasing in $(0,\infty)$, (21) and the assumptions that $\eta > 0$ and $\lambda_k \geq \underline{\lambda}$ for all $k \geq 1$.

To prove the second inequality in (20), note that from (7) and (12) we have

$$\sum_{j=1}^{k} 2\lambda_j \mu_{j-1}\varepsilon_j \leq \sigma^2 d_0^2/(1-\sigma^2),$$

and so that

$$\varepsilon_i := \min_{j=1,\ldots,k} \varepsilon_j \leq \frac{\sigma^2 d_0^2}{2(1-\sigma^2)\sum_{j=1}^{k}\lambda_j\mu_{j-1}}.$$

Hence, the desired result follows from the latter inequality, (21) and the assumption that $\eta > 0$. □

We note that in the "limit" case $\eta \to 0$ in Corollary 3.1(b), the inequalities in (20) are identical to the ones obtained in Theorem 4.4(a) of [6].

4 A Variant of the Tseng's Forward-Backward Method for Strongly MIPs

In this section, we are concerned with the MIP

$$0 \in F(x) + B(x), \tag{22}$$

where the following assumptions are assumed to hold:

B.1) $F : X \to X$ is a (single-valued) strongly monotone and L-Lipschitz continuous operator, i.e., there exist $\eta \geq 0$ and $L > 0$ such that

$$\langle F(x) - F(x'), x - x'\rangle \geq \eta\|x - x'\|^2, \quad \|F(x) - F(x')\| \leq L\|x - x'\| \quad \forall x, x' \in X; \tag{23}$$

B.2) $B : X \rightrightarrows X$ is maximal monotone;

B.3) the solution set of (22), $(F+B)^{-1}(0)$, is nonempty.

We observe that from the above assumptions, we have that problem (22) is a special case of (4) and, as a consequence of this observation, Algorithm 1 as well as its convergence and iteration-complexity analysis can be applied to (22). Moreover, we note that by combining (23) with the Cauchy-Scharwz inequality we obtain $L \geq \eta$.

The following algorithm is a variant of the Tseng's forward-backward method, which takes into consideration the fact that the operator F in (22) is strongly monotone. It worths to mention that in the case $\eta = 0$, Algorithm 2 reduces to the version of the Tseng's method presented in [6].

Algorithm 2: A variant of the Tseng's forward-backward method for strongly MIPs

(0) Let $x_0 \in X$ and $\sigma \in]0,1[$ be given and set $\lambda = \frac{\sigma}{L^2}\left(\sigma\eta + \sqrt{\sigma^2\eta^2 + L^2}\right)$; set $k = 1$;

(1) compute

$$y_k = (I + \lambda B)^{-1}(x_{k-1} - \lambda F(x_{k-1})), \quad x_k = y_k - \frac{\lambda}{1+2\lambda\eta}\left(F(y_k) - F(x_{k-1})\right); \tag{24}$$

(2) set $k = k + 1$ and go to step 1.

end

Now we have some comments about Algorithm 2. First, Nesterov and Scrimali proposed in [16] an algorithm for solving variational inequalities with strongly monotone operators. Their method, in contrast to Algorithm 2, is designed for solving (22) in the case that the operator B is the normal cone of a closed convex set. Moreover, while Algorithm 2 requires the computation of one resolvent per iteration, their method requires the computation of two projections/resolvents. We will show in Proposition 4.2 that Algorithm 2 preserves the same iteration-complexity of Nesterov-Scrimali's method. Second, using the definition of λ in Algorithm 2, we find

$$\frac{\lambda^2 L^2}{1+2\lambda\eta} = \sigma^2. \tag{25}$$

Therefore, defining

$$\omega := \frac{\sigma\eta + \sqrt{\sigma^2\eta^2 + L^2}}{L}, \tag{26}$$

it follows that

$$1 + \frac{\sigma\eta}{L} \le \omega \le 1 + \frac{2\sigma\eta}{L}. \tag{27}$$

Moreover, by (25) and the definition of λ (in Algorithm 2) we have

$$1 + 2\lambda\eta = \omega^2 \;\text{ and }\; \lambda = \frac{\sigma\omega}{L}. \tag{28}$$

By the second identity in (28), we have that if $\eta > 0$, then the stepsize $\lambda = \sigma/L$ that appears in the version of the Tseng's forward-backward method of [6] is increased in Algorithm 2 by a factor of ω.

In next lemma, we will prove that Algorithm 2 is a special case of Algorithm 1.

Lemma 4.1 *Let $\{x_k\}$ and $\{y_k\}$ be sequences generated by* Algorithm 2 *and define*

$$v_k = \frac{1}{\lambda}\big[x_{k-1} - y_k + \lambda\left(F(y_k) - F(x_{k-1})\right)\big]. \tag{29}$$

Then, the following statements hold for every $k \ge 1$:

$$v_k \in F(y_k) + B(y_k), \quad \frac{\|\lambda v_k + y_k - x_{k-1}\|^2}{1+2\lambda\eta} \le \sigma^2 \|y_k - x_{k-1}\|^2 \tag{30}$$

and

$$x_k = \frac{(x_{k-1} - \lambda v_k) + 2\lambda\eta y_k}{1+2\lambda\eta}. \tag{31}$$

As a consequence, it follows that Algorithm 2 *is a special case of* Algorithm 1 *with $\varepsilon_k = 0$ for all $k \ge 1$.*

Proof Using the definition of y_k and v_k in (24) and (29), respectively, we have

$$q_k := \frac{1}{\lambda}\left[x_{k-1} - \lambda F(x_{k-1}) - y_k\right] \in B(y_k), \quad v_k = F(y_k) + q_k \in F(y_k) + B(y_k),$$

which proves the inclusion in (30). From (29) we obtain $\lambda v_k + y_k - x_{k-1} = \lambda(F(y_k) - F(x_{k-1}))$, which in turn combined with the second inequality in (23) and (25) yields

$$\frac{\|\lambda v_k + y_k - x_{k-1}\|^2}{1+2\lambda\eta} = \frac{\|\lambda(F(y_k) - F(x_{k-1}))\|^2}{1+2\lambda\eta} \le \frac{\lambda^2 L^2}{1+2\lambda\eta}\|y_k - x_{k-1}\|^2 = \sigma^2 \|y_k - x_{k-1}\|^2.$$

Finally, (31) follows trivially from the definitions of x_k and v_k in (24) and (29), respectively. $\square$

In next proposition, we obtain rates of convergence for the sequence $\{v_k\}$ and $\{x_k\}$ generated by Algorithm 2.

Proposition 4.1 *Let $\{x_k\}$ and $\{y_k\}$ be generated by* Algorithm 2 *and let $\{v_k\}$ be as in* (29)*. Consider also ω as in* (26)*, d_0 be the distance of x_0 to the solution set of* (22) *and let $x^* \in (F+B)^{-1}(0)$. Then, the following statements hold for every $k \geq 1$:*

(a) $v_k \in F(y_k) + B(y_k)$ *and*

$$\|v_k\| \leq \frac{d_0 L}{\sigma}\sqrt{\frac{1+\sigma}{1-\sigma}}\frac{1}{\omega^{k-1}} \leq \frac{d_0 L}{\sigma}\sqrt{\frac{1+\sigma}{1-\sigma}}\frac{1}{\left(1+\frac{\sigma\eta}{L}\right)^{k-1}}, \tag{32}$$

$$\|x^* - x_k\| \leq \frac{1}{\omega^k}\|x^* - x_0\| \leq \frac{1}{\left(1+\frac{\sigma\eta}{L}\right)^k}\|x^* - x_0\|; \tag{33}$$

(b) if $\eta > 0$ in (23)*, then there exists $i \leq k$ such that $v_i \in F(y_i) + B(y_i)$ and*

$$\|v_i\| \leq \frac{d_0 L}{\sigma}\sqrt{\frac{1+\sigma}{1-\sigma}}\sqrt{\frac{\omega^2-1}{\omega^{2k}-1}} \leq \frac{d_0 L}{\sigma}\sqrt{\left(\frac{1+\sigma}{1-\sigma}\right)\frac{\left(1+\frac{2\sigma\eta}{L}\right)^2-1}{\left(1+\frac{\sigma\eta}{L}\right)^{2k}-1}}.$$

Proof Using Lemma 4.1, Corollary 3.1(a) (resp. Corollary 3.1(b)) (with $\underline{\lambda} = \lambda$), (28) and the first inequality in (27) we obtain (a) (resp. (b)). □

Next proposition is a direct consequence of Proposition 4.1. It gives iteration-complexity bounds for obtaining a pair (y, v) in the graph of $F + B$ satisfying (5) with $\bar{\varepsilon} = 0$. Moreover, it provides iteration-complexity bounds for the approximation $\{x_k\}$ with respect to any solution of (22).

Proposition 4.2 *Let $\{x_k\}$ and $\{y_k\}$ be generated* Algorithm 2*, d_0 be the distance of x_0 to the solution set of* (22)*, $x^* \in (F+B)^{-1}(0)$ and assume that $\eta > 0$ in* (23)*. Then, for every $\rho > 0$ the following statements hold:*

(a) There exists an index

$$k_0 = O\left(\frac{L}{\sigma\eta}\log\left(\frac{d_0 L}{\rho}\right)\right)$$

such that $v_k \in F(y_k) + B(y_k)$ and $\|v_k\| \leq \rho$ for all $k \geq k_0$;

(b) there exists an index

$$k_1 = O\left(\frac{L}{\sigma\eta}\log\left(\frac{\|x^* - x_0\|}{\rho}\right)\right)$$

such that $\|x^* - x_k\| \leq \rho$ for all $k \geq k_1$.

5 A Variant of the Korpelevich Extragradient Method for Strongly Monotone VIPs

In this section, we consider the problem (22) for the special case where the operator B is the normal cone of a closed convex set $C \subset X$, i.e,

$$0 \in F(x) + N_C(x). \tag{34}$$

This problem is trivially equivalent to the strongly monotone variational inequality problem

$$x \in C, \quad \langle x' - x, F(x)\rangle \geq 0 \;\; \forall x' \in C.$$

We also assume that the conditions B.1 and B.3 of the Section 4 hold true. In what follows, we present and analyze a variant of the Korpelevich extragradient method for computing approximate solutions of the variational inequality problem (VIP) (34). We will obtain iteration-complexity bounds by showing that the algorithm proposed in this section is a special case of the Algorithm 1 of Section 3.

Algorithm 3: A variant of the Korpelevich extragradient method for strongly monotone VIPs

(0) Let $x_0 \in X$ and $\sigma \in]0,1[$ be given and set $\lambda = \frac{\sigma}{L^2}\left(\sigma\eta + \sqrt{\sigma^2\eta^2 + L^2}\right)$; set $k = 1$;

(1) compute

$$y_k = P_C(x_{k-1} - \lambda F(x_{k-1})), \quad x_k = P_C\left(\frac{x_{k-1} - \lambda F(y_k) + 2\lambda\eta y_k}{1 + 2\lambda\eta}\right); \tag{35}$$

(2) set $k = k + 1$ and go to step 1.

end

We now make some remarks about Algorithm 3. First, it uses the same stepsize of Algorithm 2, but in contrast to the latter, requires two projections per iteration. Second, in the case $\eta = 0$, Algorithm 3 reduces to the Korpelevich extragradient method, for which iteration-complexity was first presented in [6].

In next lemma, we show that Algorithm 3 is a special case of Algorithm 1. This fact will allow us to obtain iteration-complexity bounds for Algorithm 3 as a direct consequence of the ones obtained for Algorithm 1 in Section 3.

Lemma 5.1 *Let* $\{x_k\}$ *and* $\{y_k\}$ *be sequences generated by* Algorithm 3 *and define*

$$q_k = \frac{x_{k-1} - \lambda F(y_k) + 2\lambda\eta y_k - (1 + 2\lambda\eta)x_k}{\lambda}, \tag{36}$$

$$v_k = F(y_k) + q_k\,, \tag{37}$$

$$\varepsilon_k = \langle q_k, x_k - y_k\rangle\,. \tag{38}$$

Then, $\varepsilon_k \geq 0$,

$$v_k \in F(y_k) + N_C^{\varepsilon_k}(y_k), \qquad \frac{\|\lambda v_k + y_k - x_{k-1}\|^2}{1 + 2\lambda\eta} + 2\lambda\varepsilon_k \leq \sigma^2\|y_k - x_{k-1}\|^2 \tag{39}$$

and

$$x_k = \frac{(x_{k-1} - \lambda v_k) + 2\lambda\eta y_k}{1 + 2\lambda\eta}\,. \tag{40}$$

As a consequence, it follows that Algorithm 3 *is a special case of* Algorithm 1.

Proof Using the second identity in (35) and the fact that $P_C = (I + N_C)^{-1}$ we obtain

$$\widetilde{q}_k := [x_{k-1} - \lambda F(y_k) + 2\lambda\eta y_k]/(1 + 2\lambda\eta) - x_k \in N_C(x_k), \tag{41}$$

which in turn combined with (36) (and the fact that $N_C(x_k)$ is a cone) yields

$$q_k = \frac{1 + 2\lambda\eta}{\lambda}\,\widetilde{q}_k \in N_C(x_k). \tag{42}$$

It follows from (42), (2) with $f = \delta_C$ and (38) that $\varepsilon_k \geq 0$ and $q_k \in \partial_{\varepsilon_k}\delta_C(y_k) \subset N_C^{\varepsilon_k}(y_k)$.

Using the latter inclusion and (37) we obtain the inclusion in (39). To prove the inequality in (39), note first that from the definition of y_k in (35) we have

$$p_k := \frac{1}{\lambda}\,(x_{k-1} - \lambda F(x_{k-1}) - y_k) \in N_C(y_k) \tag{43}$$

and thus $\langle p_k, x_k - y_k\rangle \leq 0$, because $x_k \in C$. By combining the latter inequality with (38) we find

$$\varepsilon_k = \langle q_k - p_k, x_k - y_k\rangle + \langle p_k, x_k - y_k\rangle \leq \langle q_k - p_k, x_k - y_k\rangle\,. \tag{44}$$

From (37), (42) and (41) we obtain the following identity $\lambda v_k + y_k - x_{k-1} = (1+2\lambda\eta)(y_k - x_k)$, which in turn combined with (44), (42), (43), the second inequality in (23) and the defintion of λ in Algorithm 3 yields

$$
\begin{aligned}
\frac{\|\lambda v_k + y_k - x_{k-1}\|^2}{1+2\lambda\eta} + 2\lambda\varepsilon_k &\le \frac{\|(1+2\lambda\eta)(x_k - y_k)\|^2 + 2\langle \lambda(q_k - p_k), (1+2\lambda\eta)x_k - y_k\rangle}{1+2\lambda\eta} \\
&\le \frac{\|(1+2\lambda\eta)(x_k - y_k) + \lambda(q_k - p_k)\|^2 - \|\lambda(q_k - p_k)\|^2}{1+2\lambda\eta} \\
&\le \frac{\|(1+2\lambda\eta)(x_k - y_k) + \lambda(q_k - p_k)\|^2}{1+2\lambda\eta} \\
&= \frac{\|\lambda(F(y_k) - F(x_{k-1}))\|^2}{1+2\lambda\eta} \\
&\le \frac{\lambda^2 L^2}{1+2\lambda\eta}\|y_k - x_{k-1}\|^2 \\
&= \sigma^2 \|y_k - x_{k-1}\|^2,
\end{aligned}
$$

where in the last inequality we also used the identity (25) (note that Algorithms 2 and 3 have the same stepsize λ). To finish the proof of the lemma, note that the identity in (40) follows directly from (37) and (36). ◻

The following two propositions are the analogues of Propositions 4.1 and 4.2 for the Algorithm 3.

Proposition 5.1 *Let $\{x_k\}$ and $\{y_k\}$ be generated by* Algorithm 3 *and let $\{v_k\}$ and $\{\varepsilon_k\}$ be as in* (37) *and* (38)*, respectively. Consider also ω as in* (26)*, let d_0 be the distance of x_0 to the solution set of* (34) *and let x^* be any solution of* (34)*. Then, the following statements hold for every $k \ge 1$:*

(a) $v_k \in F(y_k) + B^{[\varepsilon_k]}(y_k)$ and

$$
\|v_k\| \le \frac{d_0 L}{\sigma}\sqrt{\frac{1+\sigma}{1-\sigma}}\frac{1}{\omega^{k-1}} \le \frac{d_0 L}{\sigma}\sqrt{\frac{1+\sigma}{1-\sigma}}\frac{1}{\left(1+\frac{\sigma\eta}{L}\right)^{k-1}}, \tag{45}
$$

$$
\varepsilon_k \le \frac{d_0^2 \sigma L}{2(1-\sigma^2)\omega^{2k-1}} \le \frac{d_0^2 \sigma L}{2(1-\sigma^2)\left(1+\frac{\sigma\eta}{L}\right)^{2k-1}}, \tag{46}
$$

$$
\|x^* - x_k\| \le \frac{1}{\omega^k}\|x^* - x_0\| \le \frac{1}{\left(1+\frac{\sigma\eta}{L}\right)^k}\|x^* - x_0\|; \tag{47}
$$

(b) if $\eta > 0$ in (23)*, then there exists $i \le k$ such that $v_i \in F(y_i) + B^{[\varepsilon_i]}(y_i)$ and*

$$
\|v_i\| \le \frac{d_0 L}{\sigma}\sqrt{\frac{1+\sigma}{1-\sigma}}\sqrt{\frac{\omega^2 - 1}{\omega^{2k} - 1}} \le \frac{d_0 L}{\sigma}\sqrt{\left(\frac{1+\sigma}{1-\sigma}\right)\frac{\left(1+\frac{2\sigma\eta}{L}\right)^2 - 1}{\left(1+\frac{\sigma\eta}{L}\right)^{2k} - 1}},
$$

$$
\varepsilon_i \le \frac{d_0^2 \sigma L(\omega^2 - 1)}{2(1-\sigma^2)\omega(\omega^{2k} - 1)} \le \frac{d_0^2 \sigma L\left(\left(1+\frac{\sigma\eta}{L}\right)^2 - 1\right)}{2(1-\sigma^2)\left(1+\frac{\sigma\eta}{L}\right)\left(\left(1+\frac{\sigma\eta}{L}\right)^{2k} - 1\right)}.
$$

Proof Using Lemma 5.1, Corollary 3.1(a) (resp. Corollary 3.1(b)) (with $\underline{\lambda} = \lambda$), (28) and the first inequality in (27) (note that Algorithms 2 and 3 have the same stepsize λ) we obtain (a) (resp. (b)). ◻

Proposition 5.2 *Let $\{x_k\}$ and $\{y_k\}$ be generated by* Algorithm 3 *and assume that $\eta > 0$ in* (23)*. Let also d_0 denote the distance of x_0 to the solution set of* (34) *and x^* be any solution of* (34)*. Then, the following statements hold for every $\rho > 0$ and $\varepsilon > 0$:*

(a) There exists an index

$$
k_0 = O\left(\frac{L}{\sigma\eta}\max\left\{\log\left(\frac{d_0 L}{\rho}\right), \log\left(\frac{d_0^2 L}{\varepsilon}\right)\right\}\right)
$$

such that $v_k \in F(y_k) + B^{[\varepsilon_k]}(y_k)$, $\|v_k\| \le \rho$ and $\varepsilon_k \le \varepsilon$ for all $k \ge k_0$;

(b) there exists an index

$$
k_1 = O\left(\frac{L}{\sigma\eta}\log\left(\frac{\|x^* - x_0\|}{\rho}\right)\right)
$$

such that $\|x^* - x_k\| \le \rho$ for all $k \ge k_1$.

6 Conclusions

In this paper, we addressed the problem of solving monotone inclusions for operators that can be decomposed into two components, one which is strongly monotone and another one which is (maximal) monotone. We proposed a variant of the HPE method for these probelms and derived its iteration-complexity, which is stronger than the one of the HPE for general MIPs. The method we proposed is rather a *general framework* than an algorithm, because it did not specified how to compute an iteration. To ilustrate this fact, as applications, we have proposed variants of the Tseng's forward-backward and Korpelevich extragradient methods under the assumption of strong monotonicity. Moreover, we have proved that both methods have the same rate of convergence of a method previously proposed by Nesterov and Scrimali, and that the variant of the Tseng's method has the advantage of computing just one resolvent/projection per iteration. More precisely, we have shown that linear (global) rates of convergence can be reached for the method proposed in this paper as well as for its two variants, which improves previous sublinear rates obtained for the HPE method.

Acknowledgments

M. Marques Alves was partially supported by CNPq grants no. 406250/2013-8, 237068/2013-3 and 306317/2014-1. B. F. Svaiter was partially supported by CNPq grants no. 474996/2013-1, 302962/2011-5 and FAPERJ grant E-26/201.584/2014.

References

1. Martinet, B.: Régularisation d'inéquations variationnelles par approximations successives. Rev. Française Informat. Recherche Opérationnelle 4(Ser. R-3), 154–158 (1970).
2. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optimization 14(5), 877–898 (1976).
3. Solodov, M.V., Svaiter, B.F.: A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. Set-Valued Anal. 7(4), 323–345 (1999).
4. Solodov, M.V., Svaiter, B.F.: A hybrid projection-proximal point algorithm. J. Convex Anal. 6(1), 59–70 (1999).
5. Burachik, R.S., Iusem, A., Svaiter, B.F.: Enlargement of monotone operators with applications to variational inequalities. Set-Valued Anal. 5(2), 159–180 (1997).
6. Monteiro, R.D.C., Svaiter, B.F.: On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. SIAM J. Optim. 20(6), 2755–2787 (2010).
7. Monteiro, R.D.C., Ortiz, C., Svaiter, B.F.: A first-order block-decomposition method for solving two-easy-block structured semidefinite programs. Math. Program. Comput. 6(2), 103–150 (2014).
8. Monteiro, R.D.C., Ortiz, C., Svaiter, B.F.: Implementation of a block-decomposition algorithm for solving large-scale conic semidefinite programming problems. Comput. Optim. Appl. 57(1), 45–69 (2014).
9. Monteiro, R.D.C., Svaiter, B.F.: Complexity of variants of Tseng's modified F-B splitting and Korpelevich's methods for hemivariational inequalities with applications to saddle-point and convex optimization problems. SIAM J. Optim. 21(4), 1688–1720 (2011).
10. Monteiro, R.D.C., Svaiter, B.F.: Iteration-complexity of a Newton proximal extragradient method for monotone variational inequalities and inclusion problems. SIAM J. Optim. 22(3), 914–935 (2012).
11. Monteiro, R.D.C., Svaiter, B.F.: An accelerated hybrid proximal extragradient method for convex optimization and its implications to second-order methods. SIAM J. Optim. 23(2), 1092–1125 (2013).
12. Monteiro, R.D.C., Svaiter, B.F.: Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. SIAM J. Optim. 23(1), 475–507 (2013).
13. Solodov, M.V., Svaiter, B.F.: Forcing strong convergence of proximal point iterations in a Hilbert space. Math. Program. 87(1, Ser. A), 189–202 (2000).
14. Solodov, M.V., Svaiter, B.F.: A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem. SIAM J. Optim. 10(2), 605–625 (2000).
15. Tikhonov, A.N., Arsenin, V.Y.: Solutions of Ill-Posed Problems. Winston and Sons, Washington (1977).
16. Nesterov, Y., Scrimali, L.: Solving strongly monotone variational and quasi-variational inequalities. Discret Contin. Dyn. S. 31(4), 1383–1396 (2011).
17. Brøndsted, A., Rockafellar, R.T.: On the subdifferentiability of convex functions. Proc. Amer. Math. Soc. 16, 605–611 (1965).