

AUGMENTED LAGRANGIAN METHODS FOR VARIATIONAL INEQUALITY PROBLEMS

ALFREDO N. IUSEM¹ AND MOSTAFA NASRI¹

Abstract. We introduce augmented Lagrangian methods for solving finite dimensional variational inequality problems whose feasible sets are defined by convex inequalities, generalizing the proximal augmented Lagrangian method for constrained optimization. At each iteration, primal variables are updated by solving an unconstrained variational inequality problem, and then dual variables are updated through a closed formula. A full convergence analysis is provided, allowing for inexact solution of the subproblems.

Keywords. Augmented Lagrangian method, equilibrium problem, inexact solution, proximal point method, variational inequality problem.

Mathematics Subject Classification. 90C47, 49J35.

1. INTRODUCTION

Let $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a continuous operator and K be a nonempty, closed and convex subset of \mathbb{R}^n . The variational inequality problem, denoted by $VIP(F, K)$, consists of finding $x^* \in K$ such that

$$\langle F(x^*), y - x^* \rangle \geq 0 \quad \forall y \in K. \quad (1.1)$$

The set of solutions of $VIP(F, K)$ will be denoted by $S(F, K)$.

Received April 22, 2008. Accepted October 16, 2009.

¹ Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro, RJ, CEP 22460-320, Brazil; iusp@impa.br; mostafa@impa.br

In this paper we will assume the monotonicity of F , *i.e.* we assume that

$$\langle F(x) - F(y), x - y \rangle \geq 0 \quad \forall x, y \in \mathbb{R}^n.$$

We recall that for a monotone F , continuity is equivalent to maximal monotonicity when seen as a set-valued operator, *i.e.* as an operator $G : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^n)$. In such a setting, G is monotone if

$$\langle u - v, x - y \rangle \geq 0$$

for all $x, y \in \mathbb{R}^n$, all $u \in G(x)$ and all $v \in G(y)$, and maximal monotone when $G = G'$ whenever $\text{Graph}(G) \subset \text{Graph}(G')$, where $\text{Graph}(G) = \{(x, u) \in \mathbb{R}^n \times \mathbb{R}^n : u \in G(x)\}$.

Another problem, closely related to the variational inequality problem, is the equilibrium problem, consisting of finding an $x^* \in \hat{K}$ such that

$$\hat{f}(x^*, y) \geq 0 \quad \forall y \in \hat{K},$$

where $\hat{f} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ satisfies certain conditions (see (a)–(e) below), and \hat{K} is a nonempty, closed and convex subset of \mathbb{R}^n . The above equilibrium problem and its set of solutions are denoted by $\text{EP}(\hat{f}, \hat{K})$ and $S(\hat{f}, \hat{K})$, respectively.

The bifunction \hat{f} is said to be monotone if

$$\hat{f}(x, y) + \hat{f}(y, x) \leq 0 \quad \forall x, y \in \mathbb{R}^n. \quad (1.2)$$

In this paper, we will use the proximal point method for solving $\text{EP}(\hat{f}, \hat{K})$, developed in [20], as an essential tool in our convergence analysis. In [20] the bifunction \hat{f} defining the equilibrium problem is assumed to satisfy the following conditions, which ensure the convergence of the proximal point method for $\text{EP}(\hat{f}, \hat{K})$ to a solution of the problem, whenever such solution exists.

- (a) $\hat{f}(x, x) = 0$ for all $x \in \mathbb{R}^n$;
- (b) $\hat{f}(x, \cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex and lower semicontinuous for all $x \in \mathbb{R}^n$;
- (c) $\hat{f}(\cdot, y) : \mathbb{R}^n \rightarrow \mathbb{R}$ is upper semicontinuous for all $y \in \mathbb{R}^n$;
- (d) there exists $\theta \geq 0$ such that

$$\hat{f}(x, y) + \hat{f}(y, x) \leq \theta \|x - y\|^2 \quad \forall x, y \in \mathbb{R}^n; \quad (1.3)$$

- (e) there exists an $x^* \in S(\hat{f}, \hat{K})$ such that $\hat{f}(y, x^*) \leq 0$ for all $y \in \hat{K}$.

Now assume that $\text{VIP}(F, K)$ is given and has solutions, and that F is monotone and continuous. Define $f : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ as

$$f(x, y) = \langle F(x), y - x \rangle \quad \forall x, y \in K. \quad (1.4)$$

We observe that $\text{EP}(f, K)$ satisfies the properties (a)–(e): (a) follows immediately from (1.4), (b) from the fact that f is affine as a function of y , (c) from the continuity of F , and (d) from the monotonicity of F , which easily entails monotonicity

of f , in the sense of (1.2), so that (1.3) holds with $\theta = 0$. In connection with (e), observe first that the solution sets of $\text{EP}(f, K)$ and $\text{VIP}(F, K)$ coincide, as a consequence of (1.4), and that the inequality in (e) is valid for all solution of $\text{EP}(f, K)$ as a consequence of the already established monotonicity of f . So, existence of x^* is assured because $\text{VIP}(F, K)$ has solutions by hypothesis, and all of them are solutions of $\text{EP}(f, K)$.

Since, under our assumptions on F , $\text{VIP}(F, K)$ and $\text{EP}(f, K)$ (with f as in (1.4)) have the same solution set, it is just a matter of notation to describe our method in terms of f or F . Since we will refer to [20] in the sequel, we will use from now on the notation in this reference (namely the equilibrium one), with f rather than F . We emphasize that this is just a notational issue, with no substantial consequence whatsoever.

The variational inequality problem encompasses, among its particular cases, convex minimization problems, fixed point problems, complementarity problems, Nash equilibrium problems, and vector minimization problems (see, *e.g.*, [7,22]). For recent developments in the realm of variational inequality problems, we refer the readers to [10,11], and [14].

The variational inequality problem has been extensively studied in recent years, with emphasis on existence results (see, *e.g.*, [5,6,8,13,19,21] and [33]). In terms of computational methods for variational inequality problems, several references can be found in the literature. Among those of interest, we mention the algorithms introduced in [12,20,23,24,29–31,34] and [35] which are proximal-like methods, as well as the ones proposed in [22] which are projection-like methods. Methods based on a gap function approach can be found in [27]. Furthermore, Newton-like methods to solve the same problem has been introduced in [2] and penalty-like methods in [32].

To our knowledge, the closest approach to the one contributed here can be found in [1], where the feasible set is assumed to be of the form given in (1.5), and primal-dual methods are proposed. However, no Lagrangian function as in (2.2), or augmented Lagrangian as in (2.4), appear in this reference, so that from an algorithmical point of view or approach is completely unrelated to the one in [1].

In the current paper we introduce exact and inexact versions of augmented Lagrangian methods for solving $\text{EP}(f, K)$ in \mathbb{R}^n , for the case in which the feasible set K is of the form

$$K = \{x \in \mathbb{R}^n : h_i(x) \leq 0 \ (1 \leq i \leq m)\}, \quad (1.5)$$

where all the h_i 's are convex. These methods generate a sequence $\{(x^j, \lambda^j)\} \subseteq \mathbb{R}^n \times \mathbb{R}_+^m$ such that at iteration j , x^j is the unique solution of an unconstrained variational inequality problem and then λ^j is obtained through a closed formula. We comment next on augmented Lagrangian methods.

We remark that the most significant novelty in this paper is the introduction of our Lagrangian functions for variational inequality problems (the exact Lagrangian \mathcal{L} in (2.2), the augmented Lagrangian $\widehat{\mathcal{L}}$ in (2.4), and the Linearized Augmented Lagrangian $\widetilde{\mathcal{L}}$ in (4.1)), which are significantly different from their optimization

counterparts defined in (1.8), (1.11) and (1.12), and which are the basic ingredient of the algorithms introduced here.

The augmented Lagrangian method for equality constrained optimization problems (non-convex, in general) was introduced in [15] and [36]. Its extension to inequality constrained problems started with [9] and was continued in [4,25,37,38], and [39].

We describe next the augmented Lagrangian method for convex optimization, which is the departure point for the methods in this paper. Consider the problem

$$\min h_0(x) \tag{1.6}$$

$$\text{s.t. } h_i(x) \leq 0 \quad (1 \leq i \leq m), \tag{1.7}$$

where $h_i : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex ($0 \leq i \leq m$).

The Lagrangian for (1.6)–(1.7) is the function $L : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ given by

$$L(x, \lambda) = h_0(x) + \sum_{i=1}^m \lambda_i h_i(x), \tag{1.8}$$

and the dual problem associated to (1.6)–(1.7) is the convex minimization problem given by

$$\min -\psi(y) \quad \text{s.t. } y \in \mathbb{R}_+^m, \tag{1.9}$$

where $\psi : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{-\infty\}$ is defined as

$$\psi(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda). \tag{1.10}$$

The augmented Lagrangian associated to the problem given by (1.6)–(1.7) is the function $\bar{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ defined as

$$\bar{L}(x, \lambda, \gamma) = h_0(x) + \gamma \sum_{i=1}^m \left[\left(\max \left\{ 0, \lambda_i + \frac{h_i(x)}{2\gamma} \right\} \right)^2 - \lambda_i^2 \right], \tag{1.11}$$

where \mathbb{R}_{++} is the set of positive real numbers. The augmented Lagrangian method requires an exogenous sequence of regularization parameters $\{\gamma_j\} \subset \mathbb{R}_{++}$. The method starts with some $\lambda^0 \in \mathbb{R}_+^m$, and, given $x^j \in \mathbb{R}^n$ and $\lambda^j \in \mathbb{R}_+^m$, the algorithm first determines $x^{j+1} \in \mathbb{R}^n$ as any unconstrained minimizer of $\bar{L}(x, \lambda^j, \gamma_j)$ and then it updates λ^j as

$$\lambda_i^{j+1} = \max \left\{ 0, \lambda_i^j + \frac{h_i(x^{j+1})}{2\gamma_j} \right\} \quad (1 \leq i \leq m).$$

Assuming that both the primal problem (1.6)–(1.7) and the dual problem (1.9) have solutions, and that the sequence $\{x^j\}$ is well defined, in the sense that all the unconstrained minimization subproblems are solvable, it has been proved that the sequence $\{\lambda^j\}$ converges to a solution of the dual problem (1.9) and that the

cluster points of the sequence $\{x^j\}$ (if any) solve the primal problem (1.6)–(1.7) (see, *e.g.*, [17] or [39]).

Another augmented Lagrangian method for the same problem, with better convergence properties, is the proximal augmented Lagrangian method (see [39]; this method is called “doubly augmented Lagrangian” in [17]). In this case, \bar{L} is replaced by $\bar{\bar{L}} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_{++} \times \mathbb{R}^n \rightarrow \mathbb{R}$, defined as

$$\begin{aligned} \bar{\bar{L}}(x, \lambda, \gamma, z) &= \bar{L}(x, \lambda, \gamma) + \gamma \|x - z\|^2 \\ &= h_0(x) + \gamma \sum_{i=1}^m \left[\left(\max \left\{ 0, \lambda_i + \frac{h_i(x)}{2\gamma} \right\} \right)^2 - \lambda_i^2 \right] + \gamma \|x - z\|^2. \end{aligned} \quad (1.12)$$

The method uses an exogenous sequence $\{\gamma_j\} \subset \mathbb{R}_{++}$ as before, and it starts with $x^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}_+^m$. Given x^j, λ^j , the next primal iterate x^{j+1} is the unique unconstrained minimizer of $\bar{\bar{L}}(x, \lambda^j, \gamma_j, x^j)$ and the next dual iterate is

$$\lambda_i^{j+1} = \max \left\{ 0, \lambda_i^j + \frac{h_i(x^{j+1})}{2\gamma_j} \right\} \quad (1 \leq i \leq m).$$

In this case, the primal unconstrained subproblem always has a unique solution, due to the presence of the quadratic term $\|x - z\|^2$ in $\bar{\bar{L}}$, and assuming that both the primal and the dual problem are solvable, the sequences $\{x^j\}, \{\lambda^j\}$ converge to a primal and a dual solution respectively (see, *e.g.*, [17] or [39]). Augmented Lagrangian methods for variational inequality problems have been studied in [3].

The main tool used in [39] for establishing the above mentioned convergence results is the proximal point algorithm, whose origins can be traced back to [26] and [28]. It attained its basic formulation in the work of Rockafellar [40], where it is presented as an algorithm for finding zeroes of a maximal monotone point-to-set operator $T : \mathbb{R}^p \rightarrow \mathcal{P}(\mathbb{R}^p)$, *i.e.*, for finding $z \in \mathbb{R}^p$ such that $0 \in T(z)$.

Given an exogenous sequence of regularization parameters $\{\gamma_j\} \subset \mathbb{R}_{++}$ and an initial $z^0 \in \mathbb{R}^p$, the proximal point method generates a sequence $\{z^j\} \subset \mathbb{R}^p$ in the following way: given the j -th iterate z^j , the next iterate z^{j+1} is the unique zero of the operator $T_j : \mathbb{R}^p \rightarrow \mathcal{P}(\mathbb{R}^p)$ defined as $T_j(z) = T(z) - \gamma_j(z - z^j)$. It has been proved in [39] that if T has zeroes then $\{z^j\}$ converges to a zero of T .

Inexact versions of the method are also available; instead of requiring $\gamma_j(z^j - z^{j+1}) \in T(z^{j+1})$, they compute an auxiliary vector \tilde{z}^j satisfying $e^j + \gamma_j(z^j - \tilde{z}^j) \in T(\tilde{z}^j)$, where $e^j \in \mathbb{R}^p$ is an error vector, whose norm is small enough. The auxiliary vector \tilde{z}^j defines a hyperplane H_j which separates z^j from the set of zeroes of T . The next iterate z^{j+1} is then obtained by projecting orthogonally z^j onto H_j , or by taking a step from x^j in the direction of H_j (see, *e.g.*, [18,41], and [42]).

The connection between the augmented Lagrangian method for convex optimization and the proximal point method can be described as follows. Let $\{x^j\}, \{\lambda^j\}$ be the sequences generated by the augmented Lagrangian method. Consider the maximal monotone operator $T : \mathbb{R}^m \rightarrow \mathcal{P}(\mathbb{R}^m)$ defined as $T = \partial(-\psi)$, with ψ as in (1.10). The sequence $\{z^j\}$ generated by the proximal point for finding zeroes

of T coincides with $\{\lambda^j\}$, assuming that $\lambda^0 = z^0$, and that the same sequence $\{\gamma_j\}$ is used for both methods (see, *e.g.*, [17] or [39]). Hence, the convergence of $\{\lambda^j\}$ to some solution of the dual problem (1.9) follows from the convergence of the sequence $\{z^j\}$, generated by the proximal point method, to a zero of T .

The convergence analysis of the proximal augmented Lagrangian method proceeds in a similar way. In this case, the proximal point method is used for finding zeroes of $\widehat{T} : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^m)$ defined as

$$\widehat{T}(z) = (\partial_x L(z), -\partial_\lambda L(z)) + N_{\mathbb{R}_+^m}(z),$$

with $z = (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m$, where L is as in (1.8) and $N_{\mathbb{R}_+^m}$ is the normalizing operator of the non-negative orthant of \mathbb{R}^m . In this case, the sequence $\{z^j\}$ generated by the proximal point method coincides with the sequence $\{(x^j, \lambda^j)\}$ generated by the proximal augmented Lagrangian method, assuming again that $z^0 = (x^0, \lambda^0)$, and that the same regularization sequence $\{\gamma_j\}$ is used in both algorithms (see, *e.g.*, [17] or [39]).

The convergence analysis of the augmented Lagrangian methods for variational inequality problems to be introduced here invokes the proximal point method, presented in [23]. At iteration j of this method, given $x^j \in \mathbb{R}^n$, one solves $\text{EP}(\bar{f}_j, K)$, where the regularized function \bar{f}_j is defined as

$$\bar{f}_j(x, y) = f(x, y) + \gamma_j \langle x - x^j, y - x \rangle. \quad (1.13)$$

Two inexact versions of this method in Banach spaces have been recently proposed in [20]. In finite dimensional spaces, the first one can be described as follows: at iteration j , problem $\text{EP}(f_j^e, K)$ is solved, where f_j^e is defined as:

$$f_j^e(x, y) = f(x, y) + \gamma_j \langle x - x^j, y - x \rangle - \langle e^j, y - x \rangle. \quad (1.14)$$

Here, $e^j \in \mathbb{R}^n$ is an error vector, whose norm is small, in a sense to be defined below. The solution \tilde{x}^j of $\text{EP}(f_j^e, K)$ makes it possible to construct a hyperplane separating x^j from $S(f, K)$. A step is then taken from x^j in the direction of the separating hyperplane, generating the next iterate x^{j+1} . In the second version, x^{j+1} is the orthogonal projection of x^j onto the separating hyperplane.

It has been proved in [20] that the sequences $\{x^j\}$ generated by these methods converge to a solution of $\text{EP}(f, K)$ under appropriate assumptions on f , when $\text{EP}(f, K)$ has solutions.

The outline of this paper is as follows. In Section 2 we introduce Algorithm IALEM (*Inexact Augmented Lagrangian-Extragradient Method*) for solving $\text{EP}(f, K)$. In Section 3 we establish the convergence properties of Algorithm IALEM through the construction of an appropriate proximal point method for a certain variational inequality problem. In Section 4 we construct and analyze a variant of IALEM, called LIALEM (*Linearized Inexact Augmented Lagrangian-Extragradient Method*). Section 5 contains some final remarks.

2. AUGMENTED LAGRANGIAN METHODS FOR VARIATIONAL INEQUALITY PROBLEMS

We will assume that the closed convex set K in $\text{EP}(f, K)$ is defined as

$$K = \{x \in \mathbb{R}^n : h_i(x) \leq 0 \ (1 \leq i \leq m)\}, \quad (2.1)$$

where $h_i : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex ($1 \leq i \leq m$). We will also assume that this set of constraints satisfies any standard constraint qualification, for instance the following Slater's condition.

CQ: If I is the (possibly empty) set of indices i such that the function h_i is affine, then there exists $w \in \mathbb{R}^n$ such that $h_i(w) \leq 0$ for $i \in I$, and $h_i(w) < 0$ for $i \notin I$.

We define next our Lagrangian bifunction for $\text{EP}(f, K)$, $\mathcal{L} : (\mathbb{R}^n \times \mathbb{R}^m) \times (\mathbb{R}^n \times \mathbb{R}^m) \rightarrow \mathbb{R}$ as

$$\mathcal{L}((x, \lambda), (y, \mu)) = f(x, y) + \sum_{i=1}^m \lambda_i h_i(y) - \sum_{i=1}^m \mu_i h_i(x). \quad (2.2)$$

It is worthwhile to mention that when we consider the optimization problem (1.6)–(1.7) as a particular case of $\text{EP}(f, K)$ by taking $f(x, y) = h_0(y) - h_0(x)$, (2.2) reduces to

$$\mathcal{L}((x, \lambda), (y, \mu)) = h_0(y) - h_0(x) + \sum_{i=1}^m \lambda_i h_i(y) - \sum_{i=1}^m \mu_i h_i(x) = L(y, \lambda) - L(x, \mu),$$

where L is the usual Lagrangian for optimization problems, defined in (1.8). We introduce now the proximal augmented Lagrangian for $\text{EP}(f, K)$.

Define $s_i : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_{++} (1 \leq i \leq m)$, $\tilde{\mathcal{L}} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ as

$$s_i(x, y, \lambda, \gamma) = \frac{\gamma}{2} \left[\left(\max \left\{ 0, \lambda_i + \frac{h_i(y)}{\gamma} \right\} \right)^2 - \left(\max \left\{ 0, \lambda_i + \frac{h_i(x)}{\gamma} \right\} \right)^2 \right], \quad (2.3)$$

$$\tilde{\mathcal{L}}(x, y, \lambda, z, \gamma) = f(x, y) + \gamma \langle x - z, y - x \rangle + \gamma \sum_{i=1}^m s_i(x, y, \lambda, \gamma). \quad (2.4)$$

Now we present Algorithm EALM (*Exact Augmented Lagrangian Method*) for $\text{EP}(f, K)$. Take a bounded sequence $\{\gamma_j\} \subset \mathbb{R}_{++}$. The algorithm is initialized with a pair $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}_+^m$.

At iteration j , x^{j+1} is computed as the unique solution of the unconstrained regularized variational inequality problem $\text{EP}(\tilde{\mathcal{L}}_j, \mathbb{R}^n)$ with $\tilde{\mathcal{L}}_j$ given by

$$\tilde{\mathcal{L}}_j(x, y) = \tilde{\mathcal{L}}(x, y, \lambda^j, x^j, \gamma_j) = f(x, y) + \gamma_j \langle x - x^j, y - x \rangle + \sum_{i=1}^m s_i(x, y, \lambda^j, \gamma_j). \quad (2.5)$$

Then, the dual variables are updated as

$$\lambda_i^{j+1} = \max \left\{ 0, \lambda_i^j + \frac{h_i(x^{j+1})}{\gamma_j} \right\} \quad (1 \leq i \leq m). \quad (2.6)$$

We introduce now our inexact augmented Lagrangian method for solving $\text{EP}(f, K)$.

Algorithm IALEM: Inexact augmented Lagrangian-extragradient method for $\text{EP}(f, K)$

1. Take an exogenous bounded sequence $\{\gamma_j\} \subset \mathbb{R}_{++}$ and a relative error tolerance $\sigma \in (0, 1)$. Initialize the algorithm with $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}_+^m$.
2. Given (x^j, λ^j) , find a pair $(\tilde{x}^j, e^j) \in \mathbb{R}^n$ such that \tilde{x}^j solves $\text{EP}(\tilde{\mathcal{L}}_j^e, \mathbb{R}^n)$, where $\tilde{\mathcal{L}}_j^e$ is defined as

$$\tilde{\mathcal{L}}_j^e(x, y) := f(x, y) + \gamma_j \langle x - x^j, y - x \rangle + \sum_{i=1}^m s_i(x, y, \lambda^j, \gamma_j) - \langle e^j, y - x \rangle, \quad (2.7)$$

with s_i as given by (2.3), and e^j satisfies

$$\|e^j\| \leq \sigma \gamma_j \|(\tilde{x}^j - x^j, \lambda^{j+1} - \lambda^j)\|, \quad (2.8)$$

where $\lambda^{j+1} = (\lambda_1^{j+1}, \dots, \lambda_m^{j+1})$ is introduced in next the step.

3. Define λ^{j+1} as

$$\lambda_i^{j+1} = \max \left\{ 0, \lambda_i^j + \frac{h_i(\tilde{x}^j)}{\gamma_j} \right\} \quad (1 \leq i \leq m). \quad (2.9)$$

4. If $(x^j, \lambda^j) = (\tilde{x}^j, \lambda^{j+1})$, then stop. Otherwise,

$$x^{j+1} = \tilde{x}^j - \frac{1}{\gamma_j} e^j. \quad (2.10)$$

We mention that EALM can be realized as a particular instance of IALEM by taking $e^j = 0$ for all $j \in \mathbb{N}$.

3. CONVERGENCE ANALYSIS OF IALEM

We start this section by presenting an inexact proximal point-extragradient method for solving $\text{EP}(f, K)$, to be called IPPEM, introduced in [20]. We will use it as an auxiliary tool in the convergence analysis of IALEM.

Algorithm IPPEM. Inexact proximal point-extragradient method for $\text{EP}(f, K)$

1. Consider an exogenous bounded sequence of regularization parameters $\{\gamma_j\} \subset \mathbb{R}_{++}$ and a relative error tolerance $\sigma \in (0, 1)$. Initialize the algorithm with $x^0 \in K$.

2. Given x^j , find a pair $(\hat{x}^j, e^j) \in \mathbb{R}^n \times \mathbb{R}^n$ such that \hat{x}^j solves $\text{EP}(f_j^e, K)$ with

$$f_j^e(x, y) = f(x, y) + \gamma_j \langle x - x^j, y - x \rangle - \langle e^j, y - x \rangle, \quad (3.1)$$

and

$$\|e^j\| \leq \sigma \gamma_j \|\hat{x}^j - x^j\|. \quad (3.2)$$

3. If $\hat{x}^j = x^j$, then stop. Otherwise,

$$x^{j+1} = \hat{x}^j - \frac{1}{\gamma_j} e^j. \quad (3.3)$$

We emphasize here some features of IPPEM, which are shared by its exact counterpart (*e.g.* [23]), and by the proximal point method for finding zeroes of maximal monotone operators. The proximal point method is not an implementable algorithm, but rather a conceptual or theoretical scheme, where a certain problem is replaced by a sequence of problems of the same kind (in our case, variational inequality problems), which are in general better conditioned than the original one. However, in terms of actual implementation, some specific procedure is needed for solving *e.g.* $\text{EP}(f_j^e, K)$ in Step 2 of IPPEM, and in principle such a procedure could also be used for solving $\text{EP}(f, K)$. On the other hand, IALEM is indeed devised as an implementable method: it replaces a constrained variational inequality problem by a sequence of unconstrained one, which represents a big computational advantage in term of most effective methods for solving variational inequality problems (*e.g.*, the methods studied in [22]). Thus, we want to make it clear that we do not propose here to effectively implement IPPEM. Rather, we will use the convergence analysis for IPPEM, developed in [20], in order to obtain convergence results for IALEM. We remark that though we will prove that IALEM (applied to $\text{EP}(f, K)$), and IPPEM (applied to a very specific instance of the variational inequality problem, namely $\text{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$), generate the same sequence, both algorithms are of a rather different nature: IPPEM can be applied to a rather large class of variational inequality problems besides the above mentioned specific instance; IALEM, on the other hand, can in principle be used for solving variational inequality problems lacking any monotonicity property, in the same way as the augmented Lagrangian method for optimization problems is of interest also in the nonconvex case. Indeed, we do not analyze in this paper the convergence behavior of IALEM in the absence of the monotonicity, but it is certainly an issue which deserves further study. IPPEM also makes sense when f lacks monotonicity-like properties, but the connection between both methods breaks down in such a situation. In this paper, IPPEM is just an ancillary procedure to be used just for the sake of proving the convergence properties of IALEM.

We state next the convergence theorem for IPPEM.

Theorem 3.1. *Consider $\text{EP}(f, K)$ such that f is monotone (*i.e.*, (1.2) is satisfied). Take an exogenous sequence $\{\gamma_j\} \subset (0, \bar{\gamma}]$, for some $\bar{\gamma} > 0$. Let $\{x^j\}$ be the sequence generated by Algorithm IPPEM. If $\text{EP}(f, K)$ has solutions, then $\{x^j\}$ converges to some solution x^* of $\text{EP}(f, K)$.*

Proof. See Theorem 5.8 of [20], and the comments following its proof, establishing that some technical hypotheses required for the validity of this theorem hold automatically in the finite dimensional case, which is the one of interest here. \square

We will apply IPPEM for solving problem $\text{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$, with \mathcal{L} as in (2.2), for which we must check that this variational inequality is monotone.

Proposition 3.2. *Assume that f is monotone (i.e., (1.2) is satisfied) and that K is given by (2.1). Then \mathcal{L} , as defined in (2.2), is monotone.*

Proof. It follows easily from (2.2) and the monotonicity of f that

$$\mathcal{L}((x, \lambda), (y, \mu)) + \mathcal{L}((y, \mu), (x, \lambda)) = f(x, y) + f(y, x) \leq 0. \quad \square$$

Now we can apply Algorithm IPPEM for solving $\text{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$. In view of (3.1), the regularized function at iteration j is given by

$$\begin{aligned} \widehat{\mathcal{L}}_j^e((x, \lambda), (y, \mu)) &= \mathcal{L}((x, \lambda), (y, \mu)) + \gamma_j \langle x - x^j, y - x \rangle + \gamma_j \langle \lambda - \lambda^j, \mu - \lambda \rangle - \langle e^j, y - x \rangle \\ &= f(x, y) + \sum_{i=1}^m \lambda_i h_i(y) - \sum_{i=1}^m \mu_i h_i(x) + \gamma_j \langle x - x^j, y - x \rangle \\ &\quad + \gamma_j \langle \lambda - \lambda^j, \mu - \lambda \rangle - \langle e^j, y - x \rangle, \end{aligned} \quad (3.4)$$

so that at iteration j we must find a pair $(\hat{x}^j, \hat{\lambda}^j), (e^j, 0) \in \mathbb{R}^n \times \mathbb{R}^m$ such that $(\hat{x}^j, \hat{\lambda}^j)$ solves the problem $\text{EP}(\widehat{\mathcal{L}}_j^e, \mathbb{R}^n \times \mathbb{R}_+^m)$ with $\widehat{\mathcal{L}}_j^e$ as defined in (3.4), and the iterative formulae (3.2)–(3.3) take the form:

$$\|(e^j, 0)\| = \|e^j\| \leq \sigma \gamma_j \left\| (\hat{x}^j - x^j, \hat{\lambda}^j - \lambda^j) \right\|,$$

$$x^{j+1} = \hat{x}^j - \gamma_j^{-1} e^j, \quad (3.5)$$

$$\lambda^{j+1} = \hat{\lambda}^j. \quad (3.6)$$

Note that we do not use an error vector associated with the λ and μ arguments of \mathcal{L} . This is related to the fact that in Step 3 of Algorithm IALEM the λ_i^j 's are updated through a closed formula, so that we can assume that such an updating is performed in an exact way.

We state next the convergence result for this particular instance of IPPEM.

Corollary 3.3. *Consider $\text{EP}(f, K)$ with K given by (2.1) and f monotone (i.e., (1.2) is satisfied). Take $\{\gamma_j\} \subset (0, \bar{\gamma}]$ for some $\bar{\gamma} > 0$. Let $\{(x^j, \lambda^j)\}$ be the sequence generated by Algorithm IPPEM applied to $\text{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$. If the problem $\text{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$ has solutions, then $\{(x^j, \lambda^j)\}$ converges to some pair $(x^*, \lambda^*) \in S(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$.*

Proof. It follows from Theorem 3.1 and Proposition 3.2. \square

Now we introduce the concept of *optimal pair* for $\text{EP}(f, K)$.

Definition 3.4. We say $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m$ is an optimal pair for $\text{EP}(f, K)$ if

$$0 \in F(x^*) + \sum_{i=1}^m \lambda_i^* \partial h_i(x^*), \quad (3.7)$$

$$\lambda_i^* \geq 0 \quad (1 \leq i \leq m), \quad (3.8)$$

$$h_i(x^*) \leq 0 \quad (1 \leq i \leq m), \quad (3.9)$$

$$\lambda_i^* h_i(x^*) = 0 \quad (1 \leq i \leq m), \quad (3.10)$$

where the set $\partial h_i(x^*)$ denotes the subdifferential of the convex function h_i at the point x^* and F is defined as (1.4).

The next two propositions and corollary establish the relations between solutions of $\text{EP}(f, K)$, solutions of $\text{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$ and optimal pairs for $\text{EP}(f, K)$. We mention that the next proposition does not require a constraint qualification for the feasible set K , while Proposition 3.6 does.

Proposition 3.5. Consider $\text{EP}(f, K)$. Then the following two statements are equivalent.

- (i) (x^*, λ^*) is an optimal pair for $\text{EP}(f, K)$.
- (ii) $(x^*, \lambda^*) \in S(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$.

Proof.

(ii) \Rightarrow (i) Define $\mathcal{F}_{(x^*, \lambda^*)}(x, \lambda) = \mathcal{L}((x^*, \lambda^*), (x, \lambda))$ and consider the problem

$$\min \mathcal{F}_{(x^*, \lambda^*)}(x, \lambda) \quad (3.11)$$

$$\text{s.t. } (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}_+^m. \quad (3.12)$$

Note that (x^*, λ^*) solves (3.11)–(3.12) since $\mathcal{F}_{(x^*, \lambda^*)}(x^*, \lambda^*) = \mathcal{L}((x^*, \lambda^*), (x^*, \lambda^*)) = 0$ and that $(x^*, \lambda^*) \in S(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$. Since the constraints of this problem are affine, the constraint qualification CQ of Section 2 holds for this problem and, invoking a classical result (*e.g.* Thm. 2.3.2 in Chap. VII of [16], which deals with the non-smooth case), there exists a vector of KKT multipliers $\eta^* \in \mathbb{R}^m$ such that

$$0 \in F(x^*) + \sum_{i=1}^m \lambda_i^* \partial h_i(x^*), \quad (3.13)$$

$$h_i(x^*) + \eta_i^* = 0 \quad (1 \leq i \leq m), \quad (3.14)$$

$$\lambda^* \geq 0, \quad (3.15)$$

$$\eta^* \geq 0, \quad (3.16)$$

$$\lambda_i^* \eta_i^* = 0 \quad (1 \leq i \leq m), \quad (3.17)$$

where F is given by (1.4). Note that (3.13) and (3.15) coincide with (3.7) and (3.8) respectively. Since $\eta_i = -h_i(x^*)$ by (3.14), we get (3.9) and (3.10) from (3.16) and (3.17) respectively.

(i) \Rightarrow (ii) Now we assume that the pair (x^*, λ^*) satisfies (3.7)–(3.10). Taking $\eta_i^* = -h_i(x^*)$, we get (3.13)–(3.17). Since problem (3.11)–(3.12) is convex, the KKT conditions are sufficient for optimality, so that the pair (x^*, λ^*) solves this problem. Consequently, this pair must solve $\text{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$. \square

Proposition 3.6. *Consider $\text{EP}(f, K)$. If $x^* \in S(f, K)$ and the constraint qualification CQ in Section 2 holds for the functions h_i 's which define the feasible set K , then there exists $\lambda^* \in \mathbb{R}_+^m$ such that (x^*, λ^*) is an optimal pair for $\text{EP}(f, K)$. Conversely, if (x^*, λ^*) is an optimal pair for $\text{EP}(f, K)$ then $x^* \in S(f, K)$.*

Proof. For the first statement, since CQ holds, we invoke again *e.g.* Theorem 2.3.2 in Chapter VII of [16] to conclude that there exists a vector $\lambda^* \in \mathbb{R}^m$ such that (3.7)–(3.10) hold (we mention that, since we are assuming that both f and the h_i 's are finite on the whole $\mathbb{R}^n \times \mathbb{R}^n$ and \mathbb{R}^n respectively, there is no difficulty with the non-smooth Lagrangian condition (3.7)). It follows from Definition 3.4 that (x^*, λ^*) is an optimal pair for $\text{EP}(f, K)$. Reciprocally, if (x^*, λ^*) is an optimal pair for $\text{EP}(f, K)$, then (3.7)–(3.10) hold, but these are the KKT conditions for the problem of minimizing $f(x^*, x)$ subject to $x \in K$, which are sufficient by convexity of $f(x^*, \cdot)$ and K , and hence x^* solves this problem. \square

Corollary 3.7. *Consider $\text{EP}(f, K)$. If $(x^*, \lambda^*) \in S(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$, then $x^* \in S(f, K)$. Conversely, if $x^* \in S(f, K)$ and the constraint qualification CQ in Section 2 holds, then there exists $\lambda^* \in \mathbb{R}_+^m$ such that $(x^*, \lambda^*) \in S(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$.*

Proof. It follows from Propositions 3.5 and 3.6. \square

Corollary 3.7 shows that solving $\text{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$ is enough for solving $\text{EP}(f, K)$. Next we will prove that the sequence generated by IALEM for solving the latter problem coincides with the sequence generated by IPPEM for solving the former. We need first a technical result.

Proposition 3.8. *Consider $\text{EP}(f, K)$. Assume that f is monotone (i.e., (1.2) is satisfied). Fix $e, z \in \mathbb{R}^n$ and $\gamma > 0$. If $\tilde{f} : K \times K \rightarrow \mathbb{R}$ is defined as*

$$\tilde{f}(x, y) = f(x, y) + \gamma \langle x - z, y - x \rangle - \langle e, y - x \rangle,$$

then $\text{EP}(\tilde{f}, K)$ has a unique solution.

Proof. See Proposition 3.1 in [20]. \square

The monotonicity of f and the condition $\gamma > 0$ are essential for the validity of Proposition 3.8, whose proof is based upon an existence result for $\text{EP}(f, K)$, established in [21] and extended in [19].

Theorem 3.9. *Consider $\text{EP}(f, K)$. Assume that f is monotone (i.e., (1.2) is satisfied). Fix a sequence $\{\gamma_j\} \subset \mathbb{R}_{++}$ and a relative error tolerance $\sigma \in (0, 1)$. Let $\{(x^j, \lambda^j)\}$ be the sequence generated by Algorithm IALEM applied to $\text{EP}(f, K)$, with associated error vector $e^j \in \mathbb{R}^n$, and $\{(\bar{x}^j, \bar{\lambda}^j)\}$ the sequence generated by Algorithm IPPEM applied to $\text{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$, with associated error vector $(e^j, 0) \in$*

$\mathbb{R}^n \times \mathbb{R}^m$, using the same γ_j 's and σ . If $(x^0, \lambda^0) = (\bar{x}^0, \bar{\lambda}^0)$ then $(x^j, \lambda^j) = (\bar{x}^j, \bar{\lambda}^j)$ for all j .

Proof. We proceed by induction on j . The result holds for $j = 0$ by assumption. Assume that $(x^j, \lambda^j) = (\bar{x}^j, \bar{\lambda}^j)$. In view of Step 2 of algorithm IPPEM, we must solve $\text{EP}(\widehat{\mathcal{L}}_j^e, \mathbb{R}^n \times \mathbb{R}_+^m)$, with $\widehat{\mathcal{L}}_j^e$ as in (3.4), which has a unique solution by Proposition 3.8. Let $(\hat{x}^j, \hat{\lambda}^j)$ be the solution of this problem. By Proposition 3.5, $(\hat{x}^j, \hat{\lambda}^j)$ solves the convex minimization problem defined as

$$\min \widehat{\mathcal{F}}_{(\hat{x}^j, \hat{\lambda}^j)}(x, \lambda) \quad (3.18)$$

$$\text{s.t. } (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}_+^m, \quad (3.19)$$

with $\widehat{\mathcal{F}}_{(\hat{x}^j, \hat{\lambda}^j)}(x, \lambda) = \widehat{\mathcal{L}}_j^e((\hat{x}^j, \hat{\lambda}^j), (x, \lambda))$. The constraints of this problem are affine, so that CQ holds and therefore there exists a KKT vector $\eta^j \in \mathbb{R}^m$ such that

$$\gamma_j[\bar{x}^j - \hat{x}^j] + e^j \in F(\hat{x}^j) + \sum_{i=1}^m \hat{\lambda}_i^j \partial h_i(\hat{x}^j), \quad (3.20)$$

$$-h_i(\hat{x}^j) + \gamma_j[\hat{\lambda}_i^j - \bar{\lambda}_i^j] = \eta_i^j \quad (1 \leq i \leq m), \quad (3.21)$$

$$\hat{\lambda}^j \geq 0, \quad (3.22)$$

$$\eta^j \geq 0, \quad (3.23)$$

$$\hat{\lambda}_i^j \eta_i^j = 0 \quad (1 \leq i \leq m), \quad (3.24)$$

where F is given by (1.4). Using (3.21) to eliminate η^j , (3.20)–(3.24) can be rewritten, after some elementary calculations, as

$$\gamma_j[\bar{x}^j - \hat{x}^j] + e^j \in F(\hat{x}^j) + \sum_{i=1}^m \hat{\lambda}_i^j \partial h_i(\hat{x}^j), \quad (3.25)$$

$$\hat{\lambda}_i^j = \max \left\{ 0, \bar{\lambda}_i^j + \frac{h_i(\hat{x}^j)}{\gamma_j} \right\} \quad (1 \leq i \leq m). \quad (3.26)$$

Replacing (3.26) in (3.25) we get

$$\gamma_j[\bar{x}^j - \hat{x}^j] + e^j \in F(\hat{x}^j) + \sum_{i=1}^m \max \left\{ 0, \bar{\lambda}_i^j + \frac{h_i(\hat{x}^j)}{\gamma_j} \right\} \partial h_i(\hat{x}^j) \quad (1 \leq i \leq m). \quad (3.27)$$

Now we look at Step 2 of Algorithm IALEM, which demands the solution \tilde{x}^j of $\text{EP}(\tilde{\mathcal{L}}_j^e, \mathbb{R}^n)$. This problem is equivalent to saying that \tilde{x}^j is the unconstrained minimizer of the convex function $\tilde{\mathcal{L}}_j^e(\tilde{x}^j, \cdot)$ over \mathbb{R}^n since $\tilde{\mathcal{L}}_j^e(\tilde{x}^j, \tilde{x}^j) = 0$. That is, \tilde{x}^j belongs to $S(\tilde{\mathcal{L}}_j^e, \mathbb{R}^n)$ if and only if

$$\gamma_j[x^j - \tilde{x}^j] + e^j \in F(\tilde{x}^j) + \sum_{i=1}^m \max \left\{ 0, \lambda_i^j + \frac{h_i(\tilde{x}^j)}{\gamma_j} \right\} \partial h_i(\tilde{x}^j). \quad (3.28)$$

Since $x^j = \bar{x}^j$, $\lambda^j = \bar{\lambda}^j$ by inductive hypothesis, we get from (3.27) that (3.28) holds with \hat{x}^j substituting for \tilde{x}^j , and hence \hat{x}^j also solves $\text{EP}(\tilde{\mathcal{L}}_j^e, \mathbb{R}^n)$. Since this problem has a unique solution by Proposition 3.8, we conclude that

$$\hat{x}^j = \tilde{x}^j. \quad (3.29)$$

Taking now into account on the one hand (2.10) in Step 3 of IALEM, and on the other hand (3.5) in Step 3 of IPPEM we conclude, using again the inductive hypothesis and (3.29), that $x^{j+1} = \bar{x}^{j+1}$. Now we look at the updating of the dual variables. In view of (3.4), (3.6) and (3.26), for IPPEM we have

$$\bar{\lambda}_i^{j+1} = \hat{\lambda}_i^j = \max \left\{ 0, \bar{\lambda}_i^j + \frac{h_i(\hat{x}^j)}{\gamma_j} \right\}. \quad (3.30)$$

Comparing now (3.30) with (2.9) and taking into account (3.29) and the fact that $\bar{\lambda}^j = \lambda^j$ by the inductive hypothesis, we conclude that $\bar{\lambda}^{j+1} = \lambda^{j+1}$, completing the inductive step and the proof. \square

Now we settle the issue of finite termination of Algorithm IALEM.

Proposition 3.10. *Suppose that Algorithm IALEM stops at iteration j . Then the vector \tilde{x}^j generated by the algorithm is a solution of $\text{EP}(f, K)$.*

Proof. If Algorithm IALEM stops at the j th iteration, then, in view of Step 4, $(x^j, \lambda^j) = (\tilde{x}^j, \lambda^{j+1})$. Using (2.8) and the fact that $x^j = \tilde{x}^j$, we get $e^j = 0$. For $x \in \mathbb{R}^n$, define the function $\check{F}_x : \mathbb{R}^n \rightarrow \mathbb{R}$ as

$$\check{F}_x(y) = f(x, y) + \gamma_j \langle x - x^j, y - x \rangle + \sum_{i=1}^m s_i(x, y, \lambda^j, \gamma_j) = \check{\mathcal{L}}_j^e(x, y),$$

where the second equality holds because $e^j = 0$. Since $\tilde{x}^j = x^j$, we get

$$\check{F}_{\tilde{x}^j}(y) = f(\tilde{x}^j, y) + \sum_{i=1}^m s_i(\tilde{x}^j, y, \lambda^j, \gamma_j). \quad (3.31)$$

Note that \tilde{x}^j is an unconstrained minimizer of $\check{F}_{\tilde{x}^j}$. Thus, in view of (1.4) and (3.31),

$$0 \in \partial \check{F}_{\tilde{x}^j}(\tilde{x}^j) = F(\tilde{x}^j) + \sum_{i=1}^m \max \left\{ 0, \lambda_i^j + \frac{h_i(\tilde{x}^j)}{\gamma_j} \right\} \partial h_i(\tilde{x}^j) = F(\tilde{x}^j) + \sum_{i=1}^m \lambda_i^j \partial h_i(\tilde{x}^j), \quad (3.32)$$

using (2.9) and the fact that $\lambda^j = \lambda^{j+1}$, which also gives

$$\lambda_i^{j+1} = \lambda_i^j = \max \left\{ 0, \lambda_i^j + \frac{h_i(\tilde{x}^j)}{\gamma_j} \right\} \quad (1 \leq i \leq m). \quad (3.33)$$

It follows easily from (3.33) that

$$\lambda_i^j \geq 0, \quad \lambda_i^j h_i(\tilde{x}^j) = 0, \quad h_i(\tilde{x}^j) \leq 0 \quad (1 \leq i \leq m). \quad (3.34)$$

In view of (3.32) and (3.34), (\tilde{x}^j, λ^j) is an optimal pair for $\text{EP}(f, K)$ and we conclude from Proposition 3.6 that $\tilde{x}^j \in S(f, K)$. \square

Now we use Theorem 3.9 for completing the convergence analysis of Algorithm IALEM.

Theorem 3.11. *Consider $\text{EP}(f, K)$. Assume that*

- (i) *f is monotone (i.e., (1.2) is satisfied);*
- (ii) *K is given by (2.1);*
- (iii) *the constraint qualification CQ stated in Section 2 holds for K ;*
- (iv) *$\{\gamma_j\} \subset (0, \bar{\gamma}]$ for some $\bar{\gamma} > 0$.*

Let $\{(x^j, \lambda^j)\}$ be the sequence generated by Algorithm IALEM for solving $\text{EP}(f, K)$. If $\text{EP}(f, K)$ has solutions then the sequence $\{(x^j, \lambda^j)\}$ converges to some optimal pair (x^, λ^*) for $\text{EP}(f, K)$, and consequently $x^* \in S(f, K)$.*

Proof. By Theorem 3.9 the sequence $\{(x^j, \lambda^j)\}$ coincides with the sequence generated by IPPEM applied to $\text{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$. Since $\text{EP}(f, K)$ has solutions and CQ holds, Corollary 3.7 implies that $\text{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$ has solutions. By Corollary 3.3, the sequence $\{(x^j, \lambda^j)\}$ converges to a solution (x^*, λ^*) of $\text{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$. By Corollary 3.7 again, x^* belongs to $S(f, K)$. \square

We comment now on the real meaning of the error vector e^j appearing in Algorithms IALEM and IPPEM. These algorithms define the vector \tilde{x}^j as the exact solution of an variational inequality problem involving e^j . Though this is convenient for the sake of the presentation (and also frequent in the analysis of inexact algorithms), in actual implementations one does not consider the vector e^j “*a priori*”. Rather some auxiliary subroutine is used for solving the exact j th subproblem (i.e. the subproblem with $e^j = 0$), generating approximate solutions $\tilde{x}^{j,k}$ ($k = 1, 2, \dots$), which are offered as “candidates” for the \tilde{x}^j of the method, each of which giving rise to an associated error vector e^j , which may pass or fail the test of (2.8). To fix ideas, consider the smooth case, i.e., assume that the h_i ’s are differentiable. If $x^{j,k}$ is proposed by the subroutine as a solution of the j -th subproblem, in view of (3.28) we have

$$e^j = F(\tilde{x}^{j,k}) + \sum_{i=1}^m \max \left\{ 0, \lambda_i^j + \frac{h_i(\tilde{x}^{j,k})}{\gamma_j} \right\} \nabla h_i(\tilde{x}^{j,k}) + \gamma_j [\tilde{x}^{j,k} - x^j] \quad (3.35)$$

where F is defined as (1.4). If $\tilde{x}^{j,k}$ were the exact solution of the j -th subproblem, then the right hand side of (3.35) would vanish. If $\tilde{x}^{j,k}$ is just an approximation of this solution, then the right-hand side of (3.35) is non-zero, and we call it e^j . Then we perform the test in Step 2 of the algorithm. If e^j satisfies the inequality in (2.8), with $x^{j,k}$ substituting for \tilde{x}^j , then $\tilde{x}^{j,k}$ is accepted as \tilde{x}^j and the algorithm proceeds

to Step 3. Otherwise, the proposed $\tilde{x}^{j,k}$ is not good enough, and an additional step of the auxiliary subroutine is needed, after which the test will be repeated with $x^{j,k+1}$. It is thus important to give conditions under which any candidate vector x close enough to the exact solution of the j -th subproblem will pass the test of (2.7)–(2.8), and thus will be accepted as \tilde{x}^j . It happens to be the case that smoothness of the data functions is enough, as we explain next.

Consider $\text{EP}(f, K)$ and assume that F is continuous. We look at Algorithm IPPEM as described in (3.1)–(3.3). Let \check{x}^j be the exact solution of the j -th subproblem, *i.e.* the solution of $\text{EP}(f_j^e, K)$ with f_j^e as in (3.1) and $e^j = 0$. It has been proved in Theorem 6.11 of [20] that if \check{x}^j belongs to the interior of K then there exists $\delta > 0$ such that any vector $x \in B(\check{x}^j, \delta)$ will be accepted as \tilde{x}^j by the algorithm, or, in other words, for all $x \in B(\check{x}^j, \delta)$ there exists $e \in \mathbb{R}^n$ such that (3.1) and (3.2) are satisfied with x, e substituting for \tilde{x}^j, e^j respectively.

Observe now that the j th IALEM subproblem, namely $\text{EP}(\tilde{\mathcal{L}}_j^e, \mathbb{R}^n)$, is unconstrained, *i.e.* $K = \mathbb{R}^n$, so that the condition $\check{x}^j \in \text{int}(K)$ is automatically satisfied. Regarding the continuous differentiability of $\tilde{\mathcal{L}}_j^e$, it follows from (2.3) and (2.7) that if the h_i 's are continuously differentiable and F is continuous, then $\tilde{\mathcal{L}}_j^e$ is continuously differentiable (it is worthwhile to mention that $\tilde{\mathcal{L}}_j^e$ is never twice continuously differentiable, due to the two maxima in the definition of s_i ; see (2.3)). Thus the above result from [20] can be rephrased for the case of IALEM as follows.

Corollary 3.12. *Consider $\text{EP}(f, K)$. Assume that f is monotone (*i.e.*, (1.2) is satisfied), h_i is differentiable ($1 \leq i \leq m$), and that F , defined as in (1.4), is continuous. Let $\{(x^j, \lambda^j)\}$ be the sequence generated by Algorithm IALEM. Assume that x^j is not a solution of $\text{EP}(f, K)$ and let \check{x}^j be the unique solution of $\text{EP}(\tilde{\mathcal{L}}_j^e, \mathbb{R}^n)$, as defined in (2.7), with $e^j = 0$. Then there exists $\delta_j > 0$ such that any $x \in B(\check{x}^j, \delta_j)$ solves the subproblem (2.7)–(2.8).*

In view of Corollary 3.12, if the subproblems of IALEM are solved with a procedure guaranteed to converge to the exact solution, in the smooth case a finite number of iterations of this inner loop will suffice for generating a pair (\tilde{x}^j, e^j) satisfying the error criterium of IALEM.

4. LINEARIZED AUGMENTED LAGRANGIAN

An interesting feature of Algorithm ALEM is that its convergence properties are not altered if the Lagrangian is replaced by its first order approximation as a function of the second argument. This linearization gives rise to a variant of ALEM and IALEM which might be more suitable for actual computation. In order to perform this linearization we assume that all the h_i 's are continuously differentiable.

If we linearize the Lagrangian given by (2.2) as a function of y around $y = x$, we obtain the function $\bar{\mathcal{L}} : (\mathbb{R}^n \times \mathbb{R}^m) \times (\mathbb{R}^n \times \mathbb{R}^m) \rightarrow \mathbb{R}$ defined as

$$\bar{\mathcal{L}}((x, \lambda), (y, \mu)) = \langle F(x), y - x \rangle + \sum_{i=1}^m \lambda_i \langle \nabla h_i(x), y - x \rangle + \sum_{i=1}^m (\lambda_i - \mu_i) h_i(x), \quad (4.1)$$

where F is given by (1.4). We will denote $\bar{\mathcal{L}}$ as the *Linearized Lagrangian* for EP(f, K). Note that there is no need to linearize in the second variable of the second argument, namely μ , because \mathcal{L} is already affine as a function of μ .

Performing the same linearization on the augmented Lagrangian given by (2.7) we obtain a variant of IALEM, to be called LIALEM, which we describe next.

Algorithm LIALEM: Linearized inexact augmented Lagrangian-extragradient method for newline EP(f, K).

1. Take an exogenous bounded sequence $\{\gamma_j\} \subset \mathbb{R}_{++}$ and a relative error tolerance $\sigma \in (0, 1)$. Initialize the algorithm with $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}_+^m$.
2. Given (x^j, λ^j) , define $\bar{s}_i : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ as

$$\bar{s}_i(x, y, \lambda, \gamma) = \max \left\{ 0, \lambda_i + \frac{h_i(x)}{\gamma} \right\} \langle \nabla h_i(x), y - x \rangle \quad (1 \leq i \leq m), \quad (4.2)$$

and find a pair $(\tilde{x}^j, e^j) \in \mathbb{R}^n \times \mathbb{R}^n$ such that \tilde{x}^j solves EP($\bar{\mathcal{L}}_j^e, \mathbb{R}^n$), where $\bar{\mathcal{L}}_j^e : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ is defined as

$$\bar{\mathcal{L}}_j^e(x, y) = \langle F(x), y - x \rangle + \gamma_j \langle x - x^j, y - x \rangle + \gamma_j \sum_{i=1}^m \bar{s}_i(x, y, \lambda^j, \gamma_j) - \langle e^j, y - x \rangle, \quad (4.3)$$

with F as in (1.4) and \bar{s}_i as in (4.2), and e^j satisfies

$$\|e^j\| \leq \sigma \gamma_j \|(\tilde{x}^j - x^j, \lambda^{j+1} - \lambda^j)\|, \quad (4.4)$$

where $\lambda^{j+1} = (\lambda_1^{j+1}, \dots, \lambda_m^{j+1})$ is introduced in the next step.

3. Define λ^{j+1} as

$$\lambda_i^{j+1} = \max \left\{ 0, \lambda_i^j + \frac{h_i(\tilde{x}^j)}{\gamma_j} \right\} \quad (1 \leq i \leq m). \quad (4.5)$$

4. If $(x^j, \lambda^j) = (\tilde{x}^j, \lambda^{j+1})$, then stop. Otherwise,

$$x^{j+1} = \tilde{x}^j - \frac{1}{\gamma_j} e^j. \quad (4.6)$$

Observe that the only difference between Algorithm IALEM and Algorithm LIALEM appears in the bifunction defining the unconstrained variational inequality subproblem. In fact, in iteration j of Algorithm LIALEM one solves EP($\bar{\mathcal{L}}_j^e, \mathbb{R}^n$)

with $\bar{\mathcal{L}}$ as in (4.3), while in the j -th iteration of Algorithm IALEM one solves $\text{EP}(\bar{\mathcal{L}}_j^e, \mathbb{R}^n)$ with $\bar{\mathcal{L}}_j^e$ as in (2.7).

We show next that $\bar{\mathcal{L}}$ is monotone, so that, in view of Theorem 3.1, the sequence generated by Algorithm IPPEM applied to $\text{EP}(\bar{\mathcal{L}}, \mathbb{R}^n \times \mathbb{R}_+^m)$ will converge to a solution of $\text{EP}(\bar{\mathcal{L}}, \mathbb{R}^n \times \mathbb{R}_+^m)$.

Proposition 4.1. *Consider $\text{EP}(f, K)$. Assume that f is monotone (i.e., (1.2) is satisfied). Then, $\bar{\mathcal{L}}$ is monotone, with $\bar{\mathcal{L}}$ as given by (4.1).*

Proof. We have that

$$\begin{aligned} \bar{\mathcal{L}}((x, \lambda), (y, \mu)) + \bar{\mathcal{L}}((y, \mu), (x, \lambda)) &= \langle F(x), y - x \rangle + \langle F(y), x - y \rangle \\ + \sum_{i=1}^m \lambda_i [h_i(x) + \langle \nabla h_i(x), y - x \rangle - h_i(y)] &+ \sum_{i=1}^m \mu_i [h_i(y) + \langle \nabla h_i(y), x - y \rangle - h_i(x)] \leq 0, \end{aligned} \quad (4.7)$$

using (4.1) in the equality, and the monotonicity of F and the convexity of h_i 's in the inequality. \square

It is easy to check that Propositions 3.5, and 3.6 remain true with $\text{EP}(\bar{\mathcal{L}}, \mathbb{R}^n \times \mathbb{R}_+^m)$ substituting for $\text{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}_+^m)$. The only difference is that due to the smoothness h_i 's, the Lagrangian condition (3.7) takes the form

$$0 = F(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*),$$

where F is defined as (1.4). It is a matter of routine to check that the proofs of Theorem 3.9, Theorem 3.11 and Corollary 3.12 also remain valid for LIALEM, resulting in the following convergence theorem.

Theorem 4.2. *Consider $\text{EP}(f, K)$. Assume that*

- (i) f is monotone (i.e., (1.2) is satisfied);
- (ii) F , defined as in (1.4), is continuous;
- (iii) h_i is differentiable ($1 \leq i \leq m$);
- (iv) the constraint qualification CQ of Section 2 holds for the feasible set K .

Take an exogenous sequence $\{\gamma_j\} \subset (0, \bar{\gamma}]$, for some $\bar{\gamma} > 0$, and a relative error tolerance $\sigma \in (0, 1)$. Let $\{(x^j, \lambda^j)\}$ be the sequence generated by Algorithm LIALEM applied to $\text{EP}(f, K)$. If $\text{EP}(f, K)$ has solutions then $\{(x^j, \lambda^j)\}$ converges to an optimal pair (x^*, λ^*) for $\text{EP}(f, K)$, so that x^* belongs to $S(f, K)$. Additionally, if x^j is not a solution of $\text{EP}(f, K)$ and \tilde{x}^j is the unique solution of $\text{EP}(\bar{\mathcal{L}}_j^e, \mathbb{R}^n)$ with $e^j = 0$, then there exists $\delta_j > 0$ such that any $x \in B(\tilde{x}^j, \delta_j)$ solves the j -th subproblem of Algorithm LIALEM.

5. FINAL REMARKS

In the case of the augmented Lagrangian methods for optimization, a constrained optimization problem is replaced by a sequence of unconstrained ones. This procedure makes sense because a wide variety of fast solvers (*e.g.* quasi-Newton methods) are available for unconstrained optimization. The methods introduced in this paper (IALEM, LIALEM, etc.), in a similar fashion, replace a constrained variational inequality problem by a sequence of unconstrained ones. It is worthwhile to comment on the advantages of such a substitution in the variational inequality context, namely on the available options for solving the unconstrained subproblems. In order to avoid technicalities, we restrict our comments to the smooth case.

One interesting possibility is the projection method for solving $\text{EP}(f, K)$ proposed in [22]. At iteration j , the method requires approximate maximization of $f(\cdot, y^j)$ on the intersection of K with a ball centered at 0, followed by a projection onto a hyperplane, whose computational cost is negligible. If this procedure is applied to the unconstrained subproblems of the methods discussed here, the computationally heavy task reduces to maximization of a continuous function on a ball, which is relatively easy, as compared to the same maximization with the additional constraints $h_i(x) \leq 0$, which would be the case if the same algorithm is applied to the original problem.

We remind also that our convergence analysis, allowing for inexact solution of the subproblems, ensures that a finite number of steps of the projection method in [22] will be enough for satisfying our error criteria, as discussed in Section 3.

Another option consists of solving the system of equations resulting from (3.28) in the case of IALEM, namely

$$0 = \gamma_j(x - x^j) + F(x) + \sum_{i=1}^m \max \left\{ 0, \lambda_i^j + \frac{h_i(x)}{\gamma_j} \right\} \nabla h_i(x) \quad (5.1)$$

with F as in (1.4). We observe that the right hand side of (5.1) is continuous but not differentiable, due to the presence of the maximum. However, there is a substantial choice of efficient methods for non-smooth equations which can be used in this case.

We also mention that another inexact Proximal Point method for $\text{EP}(f, K)$ was presented in [20], where it is called Algorithm I. In this case, instead of Step 3 of IPPEM, the solution \hat{x}^j of the subproblem is used for constructing a hyperplane H_j which separates x^j from $S(f, K)$, and the next iterate x^{j+1} is the so called Bregman projection of x^j onto H_j . In our current finite dimensional context, such a Bregman projection is just the orthogonal projection. The convergence analysis of the algorithm can be found in Theorem 5.5 of [20]. Both an inexact augmented Lagrangian method for $\text{EP}(f, K)$ and its linearized version can be developed from Algorithm I in [20]. We omit the explicit development of these methods for the sake of conciseness.

The actual computational implementation of the methods introduced here is left for future research. We expect to have some results in this direction within a short period.

Acknowledgements. Research for this paper by the first author was partially supported by CNPq grant N^o 301280-86. The second author acknowledges his scholarship for his doctoral studies, granted jointly by CNPq and TWAS.

REFERENCES

- [1] A.S. Antipin, Equilibrium programming: proximal methods. *Comput. Math. Math. Phys.* **37** (1997) 1285–1296.
- [2] A.S. Antipin, F.P. Vasilev and A.S. Stukalov, A regularized Newton method for solving equilibrium programming problems with an inexact specified set. *Comput. Math. Math. Phys.* **47** (2007) 19–31.
- [3] A. Auslender and M. Teboulle, Lagrangian duality and related multiplier methods for variational inequality problems. *SIAM J. Optim.* **10** (2000) 1097–1115.
- [4] D.P. Bertsekas, On penalty and multiplier methods for constrained optimization problems. *SIAM J. Control Optim.* **14** (1976) 216–235.
- [5] M. Bianchi and R. Pini, Coercivity conditions for equilibrium problems. *J. Optim. Theory Appl.* **124** (2005) 79–92.
- [6] M. Bianchi and S. Schaible, Generalized monotone bifunctions and equilibrium problems. *J. Optim. Theory Appl.* **90** (1996) 31–43.
- [7] E. Blum and W. Oettli, From optimization and variational inequalities to equilibrium problems. *The Mathematics Student* **63** (1994) 123–145.
- [8] H. Brezis, L. Nirenberg and S. Stampacchia, A remark on Ky Fan minimax principle. *Bolletino della Unione Matematica Italiana* **6** (1972) 293–300.
- [9] J.D. Buys, *Dual algorithms for constrained optimization problems*, Ph.D. thesis, University of Leiden, The Netherlands (1972).
- [10] F. Facchinei and J.S. Pang, *Finite-dimensional Variational Inequalities and Complementarity Problems*. Springer, Berlin (2003).
- [11] M.C. Ferris and J.S. Pang, Engineering and economic applications of complementarity problems. *SIAM Rev.* **39** (1997) 669–713.
- [12] S.D. Flåm and A.S. Antipin, Equilibrium programming using proximal-like algorithms. *Math. Prog.* **78** (1997) 29–41.
- [13] F. Flores-Bazán, Existence theorems for generalized noncoercive equilibrium problems: quasiconvex case. *SIAM J. Optim.* **11** (2000) 675–790.
- [14] P.T. Harker and J.S. Pang, Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Math. Prog.* **48** (1990) 161–220.
- [15] M.R. Hestenes, Multiplier and gradient methods. *J. Optim. Theory Appl.* **4** (1969) 303–320.
- [16] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*. Springer, Berlin (1993).
- [17] A.N. Iusem, Augmented Lagrangian methods and proximal point methods for convex optimization. *Investigación Operativa* **8** (1999) 11–49.
- [18] A.N. Iusem and R. Gárciga Otero, Inexact versions of proximal point and augmented Lagrangian algorithms in Banach spaces. *Numer. Funct. Anal. Optim.* **22** (2001) 609–640.
- [19] A.N. Iusem, G. Kassay and W. Sosa, On certain conditions for the existence of solutions of equilibrium problems. *Math. Prog.* **116** (2009) 259–273.
- [20] A.N. Iusem and M. Nasri, Inexact proximal point methods for equilibrium problems in Banach spaces. *Numer. Funct. Anal. Optim.* **28** (2007) 1279–1308.
- [21] A.N. Iusem and W. Sosa, New existence results for equilibrium problems. *Nonlinear Anal.* **52** (2003) 621–635.

- [22] A.N. Iusem and W. Sosa, Iterative algorithms for equilibrium problems. *Optimization* **52** (2003) 301–316.
- [23] A.N. Iusem and W. Sosa, On the proximal point method for equilibrium problems in Hilbert spaces in appear *Optimization*.
- [24] I.V. Konnov, Application of the proximal point method to nonmonotone equilibrium problems. *J. Optim. Theory Appl.* **119** (2003) 317–333.
- [25] B.W. Kort and D.P. Bertsekas, Combined primal-dual and penalty methods for convex programming. *SIAM J. Control Optim.* **14** (1976) 268–294.
- [26] M.A. Krasnoselskii, Two observations about the method of successive approximations. *Uspekhi Matematicheskikh Nauk* **10** (1955) 123–127.
- [27] G. Mastroeni, Gap functions for equilibrium problems. *J. Glob. Optim.* **27** (2003) 411–426.
- [28] J. Moreau, Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France* **93** (1965).
- [29] A. Moudafi, Proximal point methods extended to equilibrium problems. *Journal of Natural Geometry* **15** (1999) 91–100.
- [30] A. Moudafi, Second-order differential proximal methods for equilibrium problems. *Journal of Inequalities in Pure and Applied Mathematics* **4** (2003) Article no. 18.
- [31] A. Moudafi and M. Théra, Proximal and dynamical approaches to equilibrium problems, in *Ill-posed Variational Problems and Regularization Techniques, Lect. Notes in Economics and Mathematical Systems* **477**, Springer, Berlin (1999) 187–201.
- [32] L.D. Muu and W. Oettli, Convergence of an adaptive penalty scheme for finding constraint equilibria. *Nonlinear Anal.* **18** (1992) 1159–1166.
- [33] M. Nasri and W. Sosa, Generalized Nash games and equilibrium problems (submitted).
- [34] M.A. Noor, Auxiliary principle technique for equilibrium problems. *J. Optim. Theory Appl.* **122** (2004) 371–386.
- [35] M.A. Noor and T.M. Rassias, On nonconvex equilibrium problems. *J. Math. Analysis Appl.* **212** (2005) 289–299.
- [36] M.J.D. Powell, Method for nonlinear constraints in minimization problems, in *Optimization*, edited by R. Fletcher, Academic Press, London (1969).
- [37] R.T. Rockafellar, A dual approach to solving nonlinear programming problems by unconstrained optimization. *Math. Prog.* **5** (1973) 354–373.
- [38] R.T. Rockafellar, The multiplier method of Hestenes and Powell applied to convex programming. *J. Optim. Theory Appl.* **12** (1973) 555–562.
- [39] R.T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.* **1** (1976) 97–116.
- [40] R.T. Rockafellar, Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **14** (1976) 877–898.
- [41] M.V. Solodov and B.F. Svaiter, A hybrid projection-proximal point algorithm. *Journal of Convex Analysis* **6** (1999) 59–70.
- [42] M.V. Solodov and B.F. Svaiter, An inexact hybrid extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Analysis* **7** (1999) 323–345.