Augmented Lagrangian methods for equilibrium problems

Alfredo N. Iusem * and Mostafa Nasri †

November 16, 2007

Abstract

We introduce Augmented Lagrangian methods for solving finite dimensional equilibrium 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 equilibrium 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 solutions, Proximal point method.

1 Introduction

Let K be a non-empty, closed and convex subset of $\mathbb{R}^n$. Given $f : K \times K \to \mathbb{R}$ such that

P1: $f(x, x) = 0$ for all $x \in K$,

P2: $f(x, \cdot) : K \to \mathbb{R}$ is convex and lower semicontinuous for all $x \in K$,

P3: $f(\cdot, y) : K \to \mathbb{R}$ is upper semicontinuous for all $y \in K$,

the equilibrium problem $\mathrm{EP}(f, K)$ consists of finding $x^* \in K$ such that $f(x^*, y) \geq 0$ for all $y \in K$. The set of solutions of $\mathrm{EP}(f, K)$ will be denoted by $S(f, K)$.

The equilibrium problem encompasses, among its particular cases, convex minimization problems, fixed point problems, complementarity problems, Nash equilibrium problems, variational inequality problems, and vector minimization problems (see, e.g., [4], [14]).

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

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

The equilibrium problem has been rather widely studied, but most of the work on the issue deals with conditions for the existence of solutions (see, e.g., [2], [3], [5], [7], [12], and [13]).

In terms of computational methods for equilibrium problems, only a few references can be found in the literature. Among those of interest, we mention the algorithms introduced in [14], [15], [16], and [20].

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

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

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 equilibrium problem and then λ^j is obtained through a closed formula. We comment next on Augmented Lagrangian methods.

The augmented Lagrangian method for equality constrained optimization problems (non-convex, in general) was introduced in [8] and [21]. Its extension to inequality constrained problems started with [6] and was continued in [1], [17], [22], [23], and [24].

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

$$\min h_0(x) \tag{1}$$

$$\text{s.t. } \ h_i(x) \leq 0 \ (1 \leq i \leq m), \tag{2}$$

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

The Lagrangian for (1)–(2) is the function $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ given by

$$L(x, \lambda) = h_0(x) + \sum_{i=1}^{m} \lambda_i h_i(x), \tag{3}$$

and the dual problem associated to (1)–(2) is the convex minimization problem given by

$$\min -\psi(y) \ \text{ s.t. } y \in \mathbb{R}^m_+, \tag{4}$$

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

$$\psi(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda). \tag{5}$$

The Augmented Lagrangian associated to the problem given by (1)–(2) is the function $\bar{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_{++} \to \mathbb{R}$ defined as

$$\bar{L}(x, \lambda, \gamma) = h_0(x) + \gamma \sum_{i=1}^{m} \left[\left(\max\left\{0, \lambda_i + \frac{h_i(x)}{2\gamma}\right\}\right)^2 - \lambda_i^2 \right], \tag{6}$$

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

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

Assuming that both the primal problem (1)–(2) and the dual problem (4) 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 (4) and that the cluster points of the sequence $\{x^j\}$ (if any) solve the primal problem (1)-(2) (see, e.g., [10] or [24]).

Another augmented Lagrangian method for the same problem, with better convergence properties, is the proximal Augmented Lagrangian method (see [24]; this method is called "doubly Augmented Lagrangian" in [10]). In this case, $\bar{L}$ is replaced by $\bar{\bar{L}} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_{++} \times \mathbb{R}^n \to \mathbb{R}$, defined as

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

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

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

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., [10] or [24]).

The main tool used in [24] for establishing the above mentioned convergence results is the proximal point algorithm, whose origins can be traced back to [18] and [19]. It attained its basic formulation in the work of Rockafellar [25], where it is presented as an algorithm for finding zeroes of a maximal monotone point-to-set operator $T : \mathbb{R}^p \to \mathcal{P}(\mathbb{R}^p)$, 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 $T_j : \mathbb{R}^p \to \mathcal{P}(\mathbb{R}^p)$ defined as $T_j(z) = T(z) - \gamma_j(z - z^j)$. It has been proved in [24] 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., [11], [26], and [27]).

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 $T : \mathbb{R}^m \to \mathcal{P}(\mathbb{R}^m)$ defined as $T = \partial(-\psi)$, with ψ as in (5). 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.,[10] or [24]). Hence, the convergence of $\{\lambda^j\}$ to some solution of the dual problem (4) 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 $\widehat{T} : \mathbb{R}^n \times \mathbb{R}^m \to \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^m)$ 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 (3) 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., [10] or [24]).

The convergence analysis of the Augmented Lagrangian methods for equilibrium problems to be introduced here invokes the proximal point method for equilibrium problems, presented in [15]. At iteration j of this method, given $x^j \in \mathbb{R}^n$, one solves $\mathrm{EP}(\bar{f}_j, K)$, where the regularized function $\bar{f}_j$ is defined as

$$\bar{f}_j(x, y) = f(x, y) + \gamma_j \langle x - x^j, y - x \rangle. \tag{7}$$

Two inexact versions of this method in Banach spaces have been recently proposed in [16]. In finite dimensional spaces, the first one can be described as follows: at iteration j, problem $\mathrm{EP}(f_j^e, K)$ is solved, where f_j^e is defined as:

$$f_j^e(x, y) = f(x, y) + \gamma_j \langle x - x^j, y - x \rangle - \langle e^j, y - x \rangle. \tag{8}$$

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 $\mathrm{EP}(f_j^e, K)$ 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 [16] that the sequences $\{x^j\}$ generated by these methods converge to a solution of $\mathrm{EP}(f,K)$ under appropriate assumptions on f, when $\mathrm{EP}(f,K)$ has solutions.

The outline of this paper is as follows. In Section 2 we introduce Algorithm IALEM (*Inexact Augmented Lagrangian-Extragradient Method*) for solving $\mathrm{EP}(f,K)$. In Section 3 we establish the convergence properties of Algorithm IALEM through the construction of an appropriate proximal point method for a certain equilibrium 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 equilibrium problems

We will assume that the function f can be extended to $\mathbb{R}^n \times \mathbb{R}^n$, while preserving P1–P3. In addition, we assume that the closed convex set K in $\mathrm{EP}(f,K)$ is defined as

$$K = \{x \in \mathbb{R}^n : h_i(x) \le 0 \ \ (1 \le i \le m)\}, \tag{9}$$

where $h_i : \mathbb{R}^n \to \mathbb{R}$ is convex $(1 \le i \le m)$. 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 $h_i(w) \le 0$ for $i \in I$, and $h_i(w) < 0$ for $i \notin I$.

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

$$\mathcal{L}((x,\lambda),(y,\mu)) = f(x,y) + \sum_{i=1}^m \lambda_i h_i(y) - \sum_{i=1}^m \mu_i h_i(x). \tag{10}$$

It is worthwhile to mention that when we consider the optimization problem (1)–(2) as a particular case of $\mathrm{EP}(f,K)$ by taking $f(x,y) = h_0(y) - h_0(x)$, (10) 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 (3). We introduce now the proximal Augmented Lagrangian for $\mathrm{EP}(f,K)$.

Define $s_i : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_{++} (1 \le i \le m)$, $\widetilde{\mathcal{L}} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}$ as

$$s_i(x,y,\lambda,\gamma) = \frac{\gamma}{2}\left[\left(\max\left\{0, \lambda_i + \frac{h_i(y)}{\gamma}\right\}\right)^2 - \left(\max\left\{0, \lambda_i + \frac{h_i(x)}{\gamma}\right\}\right)^2\right], \tag{11}$$

$$\widetilde{\mathcal{L}}(x,y,\lambda,z,\gamma) = f(x,y) + \gamma\langle x - z, y - x\rangle + \gamma \sum_{i=1}^m s_i(x,y,\lambda,\gamma). \tag{12}$$

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

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

$$\widetilde{\mathcal{L}}_j(x,y) = \widetilde{\mathcal{L}}(x,y,\lambda^j,x^j,\gamma_j) = f(x,y) + \gamma_j\langle x - x^j, y - x\rangle + \sum_{i=1}^{m} s_i(x,y,\lambda^j,\gamma_j). \tag{13}$$

Then, the dual variables are updated as

$$\lambda_i^{j+1} = \max\left\{0, \lambda_i^j + \frac{h_i(x^{j+1})}{\gamma_j}\right\} \quad (1 \le i \le m). \tag{14}$$

We introduce now our inexact Augmented Lagrangian method for solving EP(f, K).

Algorithm IALEM: Inexact Augmented Lagrangian-Extragradient Method for 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 EP$(\widetilde{\mathcal{L}}^e_j, \mathbb{R}^n)$, where $\widetilde{\mathcal{L}}^e_j$ is defined as

$$\widetilde{\mathcal{L}}^e_j(x,y) := f(x,y) + \gamma_j\langle x - x^j, y - x\rangle + \sum_{i=1}^{m} s_i(x,y,\lambda^j,\gamma_j) - \langle e^j, y - x\rangle, \tag{15}$$

with s_i as given by (11), and e^j satisfies

$$\left\|e^j\right\| \le \sigma\gamma_j \left\|\tilde{x}^j - x^j\right\|. \tag{16}$$

3. Define λ^{j+1} as

$$\lambda_i^{j+1} = \max\left\{0, \lambda_i^j + \frac{h_i(\tilde{x}^j)}{\gamma_j}\right\} \quad (1 \le i \le m). \tag{17}$$

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

$$x^{j+1} = \tilde{x}^j - \frac{1}{\gamma_j} e^j. \tag{18}$$

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 EP(f, K), to be called IPPEM, introduced in [16]. We will use it as an auxiliary tool in the convergence analysis of IALEM.

Algorithm IPPEM: Inexact Proximal Point-Extragradient Method for EP(f, K)

1. Consider an exogenuous 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 EP(f_j^e, K) with

$$f_j^e(x,y) = f(x,y) + \gamma_j \langle x - x^j, y - x\rangle - \langle e^j, y - x\rangle, \tag{19}$$

and

$$\left\|e^j\right\| \leq \sigma\gamma_j \left\|\hat{x}^j - x^j\right\|. \tag{20}$$

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

$$x^{j+1} = \hat{x}^j - \gamma_j^{-1} e^j. \tag{21}$$

Some monotonicity properties of f are required in the convergence analysis of IPPEM. We recall that f is said to be monotone if $f(x,y) + f(y,x) \leq 0$ for all $x, y \in \mathbb{R}^n$. In the prototypical example of an equilibrium problem, for which $f(x,y) = \langle T(x), y - x\rangle$ for some continuous $T : \mathbb{R}^n \to \mathbb{R}^n$, in which case EP$(f, K)$ is equivalent to the variational inequality problem VIP(T, K), the above property of f is equivalent to monotonicity of T, i.e., $\langle T(x) - T(y), x - y\rangle \geq 0$ for all $x, y \in \mathbb{R}^n$. We follow the notation in [16] for naming these properties.

P4: f is θ-undermonotone, i.e. there exists $\theta \geq 0$ such that

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

P4": For all $x^1, \ldots, x^q \in \mathbb{R}^n$ and all $t_1, \ldots, t_q \in \mathbb{R}_+$ such that $\sum_{\ell=1}^q t_\ell = 1$, it holds that

$$\sum_{\ell=1}^q t_\ell f\left(x^\ell, \sum_{k=1}^q t_k x^k\right) \leq 0.$$

Both P4 and P4" are weaker than monotonicity of f. Consider for instance, the function $f(x,y) = -\alpha\|x\|^2 + \beta\|y\|^2 + (\alpha - \beta)\langle x, y\rangle$, with $\beta > \alpha > 0$. It is easy to check that

$$\sum_{\ell=1}^q t_\ell f\left(x^\ell, \sum_{k=1}^q t_k x^k\right) = \alpha\left[\left\|\sum_{k=1}^q t_k x^k\right\|^2 - \sum_{k=1}^q t_k \left\|x^k\right\|^2\right] \leq 0.$$

Also $f(x,y) + f(y,x) = (\alpha - \beta)\|x-y\|^2$. So f satisfies P4" and it is θ-undermonotone with $\theta = \beta - \alpha$, but it is not monotone, since $\beta - \alpha > 0$. Note also that under P1, concavity of $f(\cdot, y)$ for all $y \in \mathbb{R}^n$ is sufficient for P4" to hold. We state next the convergence theorem for IPPEM.

Theorem 3.1. *Consider* EP(f,K) *satisfying P1–P4 and P4". Take an exogenous sequence* $\{\gamma_j\} \subset (\theta, \bar{\gamma}]$ *for some* $\bar{\gamma} > \theta$*, where* θ *is the undermonotonicity constant in P4. Let* $\{x^j\}$ *be the sequence generated by Algorithm* IPPEM*. If* EP(f,K) *has solutions, then* $\{x^j\}$ *converges to some solution* x^* *of* EP(f,K).

Proof. See Theorem 5.8 of [16], 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. □

We will apply IPPEM for solving problem EP$(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$, with $\mathcal{L}$ as in (10), for which we must check that this equilibrium problem satisfies P1–P4 and P4".

Proposition 3.2. *Assume that* f *satisfies P1–P4 and P4" on* $\mathbb{R}^n \times \mathbb{R}^n$*, and that* K *is given by* (9)*. Then* $\mathcal{L}$*, as defined in* (10)*, satisfies P1–P4 and P4"on* $(\mathbb{R}^n \times \mathbb{R}^m_+) \times (\mathbb{R}^n \times \mathbb{R}^m_+)$.

Proof. It follows easily from (10) that EP$(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$ inherits P1–P3 from EP(f,K). Furthermore, (10) implies that

$$\mathcal{L}((x,\lambda),(y,\mu)) + \mathcal{L}((y,\mu),(x,\lambda)) = f(x,y) + f(y,x) \le \theta \|x-y\|^2,$$

using the fact that f satisfies P4 on $\mathbb{R}^n \times \mathbb{R}^n$. We have shown that P4 holds for $\mathcal{L}$ with the same undermonotonicity constant θ valid for f. We prove next that $\mathcal{L}$ satisfies P4" on $(\mathbb{R}^n \times \mathbb{R}^m_+) \times (\mathbb{R}^n \times \mathbb{R}^m_+)$. Take $x^1, \ldots, x^q \in \mathbb{R}^n$, $\lambda^1, \ldots, \lambda^q \in \mathbb{R}^m_+$ and $t_1, \ldots, t_q \ge 0$ such that $\sum_{\ell=1}^q t_\ell = 1$. Then

$$\mathcal{L}\left((x^\ell,\lambda^\ell),\left(\sum_{k=1}^q t_k x^k, \sum_{k=1}^q t_k \lambda^k\right)\right) = f\left(x^\ell, \sum_{k=1}^q t_k x^k\right) + \sum_{i=1}^m \lambda_i^\ell h_i\left(\sum_{k=1}^q t_k x^k\right) - \sum_{i=1}^m \sum_{k=1}^q t_k \lambda_i^k h_i(x^\ell)$$

$$\le f\left(x^\ell, \sum_{k=1}^q t_k x^k\right) + \sum_{i=1}^m \sum_{k=1}^q \lambda_i^\ell t_k h_i(x^k) - \sum_{i=1}^m \sum_{k=1}^q t_k \lambda_i^k h_i(x^\ell), \tag{22}$$

using the convexity of the h_i's in the inequality. Multiplying the leftmost and rightmost expressions of (22) by t_ℓ and then summing with $1 \le \ell \le q$, we get

$$\sum_{\ell=1}^q t_\ell \mathcal{L}\left((x^\ell,\lambda^\ell),\left(\sum_{k=1}^q t_k x^k, \sum_{k=1}^q t_k \lambda^k\right)\right) \le$$

$$\sum_{\ell=1}^{q} t_\ell f\left(x^\ell, \sum_{k=1}^{q} t_k x^k\right) + \sum_{\ell=1}^{q}\sum_{i=1}^{m}\sum_{k=1}^{q} t_\ell t_k \lambda_i^\ell h_i(x^k) - \sum_{\ell=1}^{q}\sum_{i=1}^{m}\sum_{k=1}^{q} t_\ell t_k \lambda_i^k h_i(x^\ell). \tag{23}$$

The first term in the right hand site of (23) is non-positive because f satisfies P4" on the whole space $\mathbb{R}^n \times \mathbb{R}^n$, and the sum of the remaining terms vanishes. Thus $\mathcal{L}$ satisfies P4". ☐

Now we can apply Algorithm IPPEM for solving $\mathrm{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$. In view of (19), the regularized function at iteration j is given by

$$\widehat{\mathcal{L}}_j^e((x,\lambda),(y,\mu)) = \mathcal{L}((x,\lambda),(y,\mu)) + \gamma_j\langle x - x^j, y - x\rangle + \gamma_j\langle \lambda - \lambda^j, \mu - \lambda\rangle - \langle e^j, y - x\rangle =$$

$$f(x,y) + \sum_{i=1}^{m} \lambda_i h_i(y) - \sum_{i=1}^{m} \mu_i h_i(x) + \gamma_j\langle x - x^j, y - x\rangle + \gamma_j\langle \lambda - \lambda^j, \mu - \lambda\rangle - \langle e^j, y - x\rangle, \tag{24}$$

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 $\mathrm{EP}(\widehat{\mathcal{L}}_j^e, \mathbb{R}^n \times \mathbb{R}^m_+)$ with $\widehat{\mathcal{L}}_j^e$ as defined in (24), and the iterative formulae (20)-(21) take the form:

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

$$x^{j+1} = \hat{x}^j - \gamma_j^{-1} e^j, \tag{25}$$

$$\lambda^{j+1} = \hat{\lambda}^j. \tag{26}$$

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* $\mathrm{EP}(f, K)$ *with* K *given by* (9) *and* f *satisfying P1–P4 and P4" on* $\mathbb{R}^n \times \mathbb{R}^n$. *Take* $\{\gamma_j\} \subset (\theta, \bar{\gamma}]$ *for some* $\bar{\gamma} > \theta$, *where* θ *is the undermonotonicity constant of* f. *Let* $\{(x^j, \lambda^j)\}$ *be the sequence generated by Algorithm* IPPEM *applied to* $\mathrm{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$. *If the problem* $\mathrm{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$ *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. ☐

For each $x \in \mathbb{R}^n$, we define $F_x : \mathbb{R}^n \to \mathbb{R}$ as

$$F_x(y) := f(x,y). \tag{27}$$

F_x is convex for all x by P2. We will use this function for establishing the relation between $S(f, K)$ and $S(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$. We start with an elementary result.

Proposition 3.4. *Consider* EP(f, K). *The following two statements are equivalent.*

i) $x^* \in S(f, K)$.

ii) x^* *minimizes* F_{x^*} *over* K, *with* F_{x^*} *as in* (27).

Proof. Assume that $x^* \in S(f, K)$. By (27) and P1 we have that

$$F_{x^*}(y) = f(x^*, y) \geq 0 = f(x^*, x^*) = F_{x^*}(x^*)$$

for all $y \in K$, establishing (ii). Now assume that (ii) is satisfied. Using again P1 and (27), we get

$$f(x^*, y) = F_{x^*}(y) \geq F_{x^*}(x^*) = f(x^*, x^*) = 0$$

for all $y \in K$, which gives the desired result. ☐

Now we introduce the concept of *optimal pair for* EP(f, K).

Definition 3.5. *We say* $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m$ *is an optimal pair for* EP(f, K) *if*

$$0 \in \partial F_{x^*}(x^*) + \sum_{i=1}^{m} \lambda_i^* \partial h_i(x^*), \tag{28}$$

$$\lambda_i^* \geq 0 \ (1 \leq i \leq m), \tag{29}$$

$$h_i(x^*) \leq 0 \ (1 \leq i \leq m), \tag{30}$$

$$\lambda_i^* h_i(x^*) = 0 \ (1 \leq i \leq m). \tag{31}$$

The sets $\partial F_{x^*}(x^*)$, $\partial h_i(x^*)$ denote the subdifferentials of the convex functions F_{x^*} and h_i, respectively, at the point x^*. Note that (28)–(31) are the KKT conditions associated to the problem of minimizing $F_{x^*}(x)$ subject to $x \in K$. However, a KKT pair for this problem is not in general an optimal pair for EP(f, K); the point x^* must be a minimizer of F_x over K precisely for $x = x^*$. On the other hand, if x^* does minimize F_{x^*} on K, then any vector λ^* of KKT multipliers for this problem will make, together with x^*, an optimal pair for EP(f, K).

The next two propositions and corollary establish the relations between solutions of EP(f, K), solutions of EP$(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$ and optimal pairs for EP(f, K). We mention that the next proposition does not require a constraint qualification for the feasible set K, while Proposition 3.7 does.

Proposition 3.6. *Consider* EP(f, K) *and assume that* f *satisfies P1–P3 on* $\mathbb{R}^n \times \mathbb{R}^n$. *Then the following two statements are equivalent.*

i) (x^*, λ^*) *is an optimal pair for* EP(f, K).

ii) $(x^*, \lambda^*) \in S(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$.

Proof.

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

$$\min \mathcal{F}_{(x^*,\lambda^*)}(x, \lambda) \tag{32}$$

$$\text{s.t. } (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m_+. \tag{33}$$

Since EP$(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$ satisfies P1–P3 by Proposition 3.2, we conclude from Proposition 3.4 that the pair (x^*, λ^*) solves (32)–(33). 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. Theorem 2.3.2 in Chapter VII of [9], which deals with the non-smooth case), there exists a vector of KKT multipliers $\eta^* \in \mathbb{R}^m$ such that

$$0 \in \partial F_{x^*}(x^*) + \sum_{i=1}^{m} \lambda_i^* \partial h_i(x^*), \tag{34}$$

$$h_i(x^*) + \eta_i^* = 0 \ \ (1 \leq i \leq m), \tag{35}$$

$$\lambda^* \geq 0, \tag{36}$$

$$\eta^* \geq 0, \tag{37}$$

$$\lambda_i^* \eta_i^* = 0 \ \ (1 \leq i \leq m). \tag{38}$$

Note that (34) and (36) coincide with (28) and (29) respectively. Since $\eta_i = -h_i(x^*)$ by (35), we get (30) and (31) from (37) and (38) respectively.

i)⇒ ii) Now we assume that the pair (x^*, λ^*) satisfies (28)–(31). Taking $\eta_i^* = -h_i(x^*)$, we get (34)–(38). Since problem (32)–(33) is convex, the KKT conditions are sufficient for optimality, so that the pair (x^*, λ^*) solves this problem. In view of Proposition 3.4, this pair solves EP$(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$.

□

Proposition 3.7. *Consider* $\mathrm{EP}(f,K)$ *and assume that* f *satisfies P1–P3 on* $\mathbb{R}^n \times \mathbb{R}^n$. *If* x^* *minimizes* F_{x^*} *over* K, *with* F_{x^*} *as in* (27), *and the constraint qualification* CQ *in Section 2 holds for the functions* h_i *which define the feasible set* K, *then there exists* $\lambda^* \in \mathbb{R}^m_+$ *such that* (x^*, λ^*) *is an optimal pair for* $\mathrm{EP}(f,K)$. *Conversely, if* (x^*, λ^*) *is an optimal pair for* $\mathrm{EP}(f,K)$ *then* x^* *minimizes* F_{x^*} *over* K, *with* F_{x^*} *as in* (27).

Proof. For the first statement, since CQ holds, we invoke again e.g. Theorem 2.3.2 in Chapter VII of [9] to conclude that there exists a vector $\lambda^* \in \mathbb{R}^m$ such that (28)–(31) hold (we mention that, since we are assuming that both F_{x^*} and the h_i's are finite on the whole $\mathbb{R}^n$, there is no difficulty with the non-smooth Lagrangian condition (28)). It follows from Definition 3.5 that (x^*, λ^*) is an optimal pair for $\mathrm{EP}(f,K)$. Reciprocally, if (x^*, λ^*) is an optimal pair for $\mathrm{EP}(f,K)$, then (28)–(31) 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^*} and K, and hence x^* solves this problem. □

Corollary 3.8. *Consider* $\mathrm{EP}(f,K)$ *and assume that* f *satisfies P1–P3 on* $\mathbb{R}^n \times \mathbb{R}^n$. *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 3.4, 3.6 and 3.7. □

Corollary 3.8 shows that solving $\mathrm{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m)$ is enough for solving $\mathrm{EP}(f,K)$. Next we will prove that the sequence generated by IPPEM for solving the latter problem coincides with the sequence generated by IALEM for solving the former. We need first a technical result.

Proposition 3.9. *Consider* f *satisfying P1–P4. Fix* $e, z \in \mathbb{R}^n$ *and* $\gamma > \theta$, *where* θ *is the undermonotonicity constant of* f *introduced in P4. If* $\tilde{f} : K \times K \to \mathbb{R}$ *is defined as*

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

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

Proof. See Proposition 3.1 in [16]. □

P4 and the condition $\gamma > \theta$ are essential for the validity of Proposition 3.9, whose proof is based upon an existence result for $\mathrm{EP}(f,K)$, established in [13] and extended in [12]. Now we prove the equivalence between IALEM and IPPEM.

Theorem 3.10. *Assume that* $\mathrm{EP}(f,K)$ *satisfies P1–P4 on* $\mathbb{R}^n \times \mathbb{R}^n$. *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* $\mathrm{EP}(f,K)$, *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* $\mathrm{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$, *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 $\mathrm{EP}(\widehat{\mathcal{L}}_j^e, \mathbb{R}^n \times \mathbb{R}_+^m)$, with $\widehat{\mathcal{L}}_j^e$ as in (24), which has a unique solution by Proposition 3.9. Let $(\hat{x}^j, \hat{\lambda}^j)$ be the solution of this problem. By Proposition 3.6 $(\hat{x}^j, \hat{\lambda}^j)$ solves the convex minimization problem defined as

$$\min \widehat{\mathcal{F}}_{(\hat{x}^j, \hat{\lambda}^j)}(x, \lambda) \tag{39}$$

$$\text{s.t. } (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}_+^m, \tag{40}$$

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

$$\gamma_j[\bar{x}^j - \hat{x}^j] + e^j \in \partial F_{\hat{x}^j}(\hat{x}^j) + \sum_{i=1}^m \hat{\lambda}_i^j \partial h_i(\hat{x}^j), \tag{41}$$

$$-h_i(\hat{x}^j) + \gamma_j[\hat{\lambda}_i^j - \bar{\lambda}_i^j] = \eta_i^j \ \ (1 \le i \le m), \tag{42}$$

$$\hat{\lambda}^j \ge 0, \tag{43}$$

$$\eta^j \ge 0, \tag{44}$$

$$\hat{\lambda}_i^j \eta_i^j = 0 \ \ (1 \le i \le m). \tag{45}$$

Using (42) to eliminate η^j, (41)–(45) can be rewritten, after some elementary calculations, as

$$\gamma_j[\bar{x}^j - \hat{x}^j] + e^j \in \partial F_{\hat{x}^j}(\hat{x}^j) + \sum_{i=1}^m \hat{\lambda}_i^j \partial h_i(\hat{x}^j), \tag{46}$$

$$\hat{\lambda}_i^j = \max\left\{0, \bar{\lambda}_i^j + \frac{h_i(\hat{x}^j)}{\gamma_j}\right\} \ \ (1 \le i \le m). \tag{47}$$

Replacing (47) in (46) we get

$$\gamma_j[\bar{x}^j - \hat{x}^j] + e^j \in \partial F_{\hat{x}^j}(\hat{x}^j) + \sum_{i=1}^m \max\left\{0, \bar{\lambda}_i^j + \frac{h_i(\hat{x}^j)}{\gamma_j}\right\} \partial h_i(\hat{x}^j) \ \ (1 \le i \le m). \tag{48}$$

Now we look at Step 2 of Algorithm IALEM, which demands the solution $\tilde{x}^j$ of $\mathrm{EP}(\widetilde{\mathcal{L}}^e_j, \mathbb{R}^n)$. Applying now Proposition 3.4 to this problem, we obtain that $\tilde{x}^k$ belongs to $S(\widetilde{\mathcal{L}}^e_j, \mathbb{R}^n)$ if and only if

$$\gamma_j[x^j - \tilde{x}^j] + e^j \in \partial F_{\tilde{x}^j}(\tilde{x}^j) + \sum_{i=1}^m \max\left\{0, \lambda_i^j + \frac{h_i(\tilde{x}^j)}{\gamma_j}\right\} \partial h_i(\tilde{x}^j). \tag{49}$$

Since $x^j = \bar{x}^j$, $\lambda^j = \bar{\lambda}^j$ by inductive hypothesis, we get from (48) that (49) holds with $\hat{x}^j$ substituting for $\tilde{x}^j$, and hence $\hat{x}^j$ also solves $\mathrm{EP}(\widetilde{\mathcal{L}}^e_j, \mathbb{R}^n)$. Since this problem has a unique solution by Proposition 3.9, we conclude that

$$\hat{x}^j = \tilde{x}^j. \tag{50}$$

Taking now into account on the one hand (18) in Step 3 of IALEM, and on the other hand (25) in Step 3 of IPPEM we conclude, using again the inductive hypothesis and (50), that $x^{j+1} = \bar{x}^{j+1}$. Now we look at the updating of the dual variables. In view of (24), (26) and (47), for IPPEM we have

$$\bar{\lambda}_i^{j+1} = \hat{\lambda}_i^j = \max\left\{0, \bar{\lambda}_i^j + \frac{h_i(\hat{x}^j)}{\gamma_j}\right\}. \tag{51}$$

Comparing now (51) with (17) and taking into account (50) 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. □

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

Proposition 3.11. *Suppose that Algorithm* IALEM *stops at iteration* j. *Then the vector* $\tilde{x}^j$ *generated by the algorithm is a solution of* $\mathrm{EP}(f, K)$.

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 (16) and the fact that $x^j = \tilde{x}^j$, we get $e^j = 0$. For $x \in \mathbb{R}^n$, define the function $\breve{F}_x : \mathbb{R}^n \to \mathbb{R}$ as

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

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

$$\breve{F}_{\tilde{x}^j}(y) = f(\tilde{x}^j, y) + \sum_{i=1}^m s_i(\tilde{x}^j, y, \lambda^j, \gamma_j). \tag{52}$$

By Proposition 3.4, $\tilde{x}^j$ is an unconstrained minimizer of $\breve{F}_{\tilde{x}^j}$. Thus, in view of (52),

$$0 \in \partial \breve{F}_{\tilde{x}^j}(\tilde{x}^j) = \partial F_{\tilde{x}^j}(\tilde{x}^j) + \sum_{i=1}^{m} \max\left\{0, \lambda_i^j + \frac{h_i(\tilde{x}^j)}{\gamma_j}\right\} \partial h_i(\tilde{x}^j) = \partial F_{\tilde{x}^j}(\tilde{x}^j) + \sum_{i=1}^{m} \lambda_i^j \partial h_i(\tilde{x}^j), \tag{53}$$

with $F_{\tilde{x}^j}$ as in (27), using (17) and the fact that $\lambda^j = \lambda^{j+1}$, which also gives

$$\lambda_i^{j+1} = \lambda_i^j = \max\left\{0, \lambda_i^j + \frac{h_i(\tilde{x}^j)}{\gamma_j}\right\} \quad (1 \le i \le m). \tag{54}$$

It follows easily from (54) that

$$\lambda_i^j \ge 0, \quad \lambda_i^j h_i(\tilde{x}^j) = 0, \qquad h_i(\tilde{x}^j) \le 0 \qquad (1 \le i \le m). \tag{55}$$

In view of (53) and (55), $(\tilde{x}^j, \lambda^j)$ is an optimal pair for $\mathrm{EP}(f, K)$ and we conclude from Corollary 3.8 that $\tilde{x}^j \in S(f, K)$. □

Now we use Theorem 3.10 for completing the convergence analysis of Algorithm IALEM.

Theorem 3.12. *Consider* $\mathrm{EP}(f, K)$. *Assume that*

i) f satisfies P1–P4 and P4" on $\mathbb{R}^n \times \mathbb{R}^n$,

ii) K is given by (9),

iii) the constraint qualification CQ *stated in Section 2 holds for K,*

iv) $\{\gamma_j\} \subset (\theta, \bar{\gamma}]$ *for some $\bar{\gamma} > \theta$, where θ is the undermonotonicity constant of f in P4.*

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

Proof. By Theorem 3.10 the sequence $\{(x^j, \lambda^j)\}$ coincides with the sequence generated by IPPEM applied to $\mathrm{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$. Since $\mathrm{EP}(f, K)$ has solutions and CQ holds, Corollary 3.8 implies that $\mathrm{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$ has solutions. By Theorem 3.1, the sequence $\{(x^j, \lambda^j)\}$ converges to a solution (x^*, λ^*) of $\mathrm{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$. By Proposition 3.6, (x^*, λ^*) is an optimal pair for $\mathrm{EP}(f, K)$. By Corollary 3.8 again, x^* belongs to $S(f, K)$. □

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 equilibrium 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 (16). To fix ideas, consider the smooth case, i.e., assume that both f and 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 (49) we have

$$e^j = \nabla F_{\tilde{x}^{j,k}}(\tilde{x}^{j,k}) + \sum_{i=1}^{m} \max\left\{0, \lambda_i^j + \frac{h_i(\tilde{x}^{j,k})}{\gamma_j}\right\} \nabla h_i(\tilde{x}^{j,k}) + \gamma_j[\tilde{x}^{j,k} - x^j]. \tag{56}$$

If $\tilde{x}^{j,k}$ were the exact solution of the j-th subproblem, then the right hand side of (56) would vanish. If $\tilde{x}^{j,k}$ is just an approximation of this solution, then the right-hand side of (56) 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 (16), 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 (15)–(16), 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 EP(f, K) and assume that f is continuously differentiable. We look at Algorithm IPPEM as described in (19)–(21). Let $\breve{x}^j$ be the exact solution of the j-th subproblem, i.e. the solution of EP(f_j^e, K) with f_j^e as in (19) and $e^j = 0$. It has been proved in Theorem 6.11 of [16] that if $\breve{x}^j$ belongs to the interior of K then there exists $\delta > 0$ such that any vector $x \in B(\breve{x}^j, \delta)$ will be accepted as $\tilde{x}^j$ by the algorithm, or, in other words, for all $x \in B(\breve{x}^j, \delta)$ there exists $e \in \mathbb{R}^n$ such that (19) and (20) are satisfied with x, e substituting for $\tilde{x}^j, e^j$ respectively.

Observe now that the j-th IALEM subproblem, namely EP$(\widetilde{\mathcal{L}}_j^e, \mathbb{R}^n)$, is unconstrained, i.e. $K = \mathbb{R}^n$, so that the condition $\breve{x}^j \in \text{int}(K)$ is automatically satisfied. Regarding the continuous differentiability of $\widetilde{\mathcal{L}}_j^e$, it follows from (11) and (15) that if the h_i's are continuously differentiable, and the same holds for f, then $\widetilde{\mathcal{L}}_j^e$ is continuously differentiable (it is worthwhile to mention that $\widetilde{\mathcal{L}}_j^e$ is never twice continuously differentiable, due to the two maxima in the definition of s_i; see (11)). Thus the above result from [16] can be rephrased for the case of IALEM as follows.

Corollary 3.13. *Consider* EP(f, K). *Assume that f satisfies P1–P4 and P4" on $\mathbb{R}^n \times \mathbb{R}^n$, f is continuously differentiable and h_i is differentiable ($1 \le i \le m$). Let $\{(x^j, \lambda^j)\}$ be the sequence generated by Algorithm* IALEM. *Assume that x^j is not a solution of* EP(f, K) *and let $\breve{x}^j$ be the unique solution of* EP$(\widetilde{\mathcal{L}}_j^e, \mathbb{R}^n)$, *as defined in* (15), *with $e^j = 0$. Then there exists $\delta_j > 0$ such that any $x \in B(\breve{x}^j, \delta_j)$ solves the subproblem* (15)–(16).

In view of Corollary 3.13, 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 both f and all the h_i's are continuously differentiable. We will use extensively the notation $F_x(y) = f(x,y)$, and in particular the gradient of F_x, denoted as $\nabla F_x : \mathbb{R}^n \to \mathbb{R}^n$.

If we linearize the Lagrangian given by (10) as a function of y around $y = x$, we obtain the function $\bar{\mathcal{L}} : (\mathbb{R}^n \times \mathbb{R}^m) \times (\mathbb{R}^n \times \mathbb{R}^m) \to \mathbb{R}$ defined as

$$\bar{\mathcal{L}}((x,\lambda),(y,\mu)) = \langle \nabla F_x(x), y - x\rangle + \sum_{i=1}^{m} \lambda_i \langle \nabla h_i(x), y - x\rangle + \sum_{i=1}^{m} (\lambda_i - \mu_i) h_i(x). \tag{57}$$

We will denote $\bar{\mathcal{L}}$ as the *Linearized Lagrangian* for $\mathrm{EP}(f,K)$. 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 (15) we obtain a variant of IALEM, to be called LIALEM, which we describe next.

Algorithm LIALEM: Linearized Inexact Augmented Lagrangian-Extragradient Method for $\mathrm{EP}(f,K)$

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 $\bar{s}_i : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}_{++} \to \mathbb{R}$ as

$$\bar{s}_i(x,y,\lambda,\gamma) = \max\left\{0, \lambda_i + \frac{h_i(x)}{\gamma}\right\} \langle \nabla h_i(x), y - x\rangle \;\; (1 \le i \le m), \tag{58}$$

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

$$\bar{\mathcal{L}}^e_j(x,y) = \langle \nabla F_x(x), y - x\rangle + \gamma_j \langle x - x^j, y - x\rangle + \gamma_j \sum_{i=1}^{m} \bar{s}_i(x,y,\lambda^j,\gamma_j) - \langle e^j, y - x\rangle, \tag{59}$$

with $\bar{s}_i$ as in (58), and e^j satisfies

$$\left\| e^j \right\| \le \sigma \gamma_j \left\| \tilde{x}^j - x^j \right\|. \tag{60}$$

3. Define λ^{j+1} as

$$\lambda_i^{j+1} = \max\left\{0, \lambda_i^j + \frac{h_i(\tilde{x}^j)}{\gamma_j}\right\} \quad (1 \le i \le m). \tag{61}$$

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

$$x^{j+1} = \tilde{x}^j - \frac{1}{\gamma_j} e^j. \tag{62}$$

Observe that the only difference difference between Algorithm IALEM and Algorithm LIALEM appears in the bifunction defining the unconstrained equilibrium subproblem. In fact, in iteration j of Algorithm LIALEM one solves $\mathrm{EP}(\bar{\mathcal{L}}_j^e, \mathbb{R}^n)$ with $\bar{\mathcal{L}}$ as in (59), while in the j-th iteration of Algorithm IALEM one solves $\mathrm{EP}(\widetilde{\mathcal{L}}_j^e, \mathbb{R}^n)$ with $\widetilde{\mathcal{L}}_j^e$ as in (15).

We show next that $\mathrm{EP}(\bar{\mathcal{L}}, \mathbb{R}^n \times \mathbb{R}^m_+)$ satisfies P1–P4 and P4", so that, in view of Theorem 3.1, the sequence generated by Algorithm IPPEM applied to $\mathrm{EP}(\bar{\mathcal{L}}, \mathbb{R}^n \times \mathbb{R}^m_+)$ will converge to a solution of $\mathrm{EP}(\bar{\mathcal{L}}, \mathbb{R}^n \times \mathbb{R}^m_+)$.

Proposition 4.1. *Consider* $\mathrm{EP}(f, K)$. *Assume that f satisfies P1–P4 and P4" on $\mathbb{R}^n \times \mathbb{R}^n$. Then, $\bar{\mathcal{L}}$ satisfies P1–P4 and P4" on $(\mathbb{R}^n \times \mathbb{R}^m_+) \times (\mathbb{R}^n \times \mathbb{R}^m_+)$, with $\bar{\mathcal{L}}$ as given by* (57).

Proof. The fact that $\mathrm{EP}(\bar{\mathcal{L}}, \mathbb{R}^n \times \mathbb{R}^m_+)$ inherits P1–P3 from $\mathrm{EP}(f, K)$ is immediate. We prove next that P4 holds. Using (57), we get

$$\bar{\mathcal{L}}((x,\lambda),(y,\mu)) + \bar{\mathcal{L}}((y,\mu),(x,\lambda)) = \langle \nabla F_x(x), y - x\rangle + \langle \nabla F_y(y), x - y\rangle$$

$$+ \sum_{i=1}^m \lambda_i [h_i(x) + \langle \nabla h_i(x), y - x\rangle - h_i(y)] + \sum_{i=1}^m \mu_i [h_i(y) + \langle \nabla h_i(y), x - y\rangle - h_i(x)]$$

$$\le f(x,y) - f(x,x) + f(y,x) - f(y,y) = f(x,y) + f(y,x) \le \theta \left\| x - y \right\|^2, \tag{63}$$

using (57) in the first equality, the convexity of F_x and F_y resulting from P2, and also of the h_i's, in the first inequality, property P1 in the second equality, and the fact that f satisfies P4 in the second inequality. We have shown that $\bar{\mathcal{L}}$ satisfies P4 with the same undermonotonicity constant as f, namely θ.

In order to show that $\bar{\mathcal{L}}$ satisfies P4" on $(\mathbb{R}^n \times \mathbb{R}^m_+) \times (\mathbb{R}^n \times \mathbb{R}^m_+)$, take $x^1, \dots, x^q \in \mathbb{R}^n$, $\lambda^1, \dots, \lambda^q \in \mathbb{R}^m_+$ and $t_1, \dots, t_q \ge 0$ such that $\sum_{\ell=1}^q t_\ell = 1$. Then

$$\bar{\mathcal{L}}\left((x^\ell, \lambda^\ell), \left(\sum_{k=1}^q t_k x^k, \sum_{k=1}^q t_k \lambda^k\right)\right) = \left\langle \nabla F_{x^\ell}(x^\ell), \sum_{k=1}^q t_k x^k - x^\ell \right\rangle$$

$$+\sum_{i=1}^{m}\lambda_i^{\ell}\left[h_i(x^{\ell})+\left\langle\nabla h_i(x^{\ell}),\sum_{k=1}^{q}t_kx^k-x^{\ell}\right\rangle\right]-\sum_{i=1}^{m}\sum_{k=1}^{q}t_k\lambda_i^k h_i(x^{\ell})$$

$$\le f\left(x^{\ell},\sum_{k=1}^{q}t_kx^k\right)-f(x^{\ell},x^{\ell})+\sum_{i=1}^{m}\lambda_i^{\ell}h_i\left(\sum_{k=1}^{q}t_kx^k\right)-\sum_{i=1}^{m}\sum_{k=1}^{q}t_k\lambda_i^k h_i(x^{\ell})\le$$

$$f\left(x^{\ell},\sum_{k=1}^{q}t_kx^k\right)+\sum_{i=1}^{m}\sum_{k=1}^{q}t_k\lambda_i^{\ell}h_i(x^k)-\sum_{i=1}^{m}\sum_{k=1}^{q}t_k\lambda_i^k h_i(x^{\ell}),\tag{64}$$

using convexity of $F_{x^{\ell}}$ resulting from P2, and of the h_i's, in the first inequality, and convexity of the h_i's and P1 in the second one. The fact that $\bar{\mathcal{L}}$ satisfies P4" can be obtained from (64) using the same argument as in the proof of Proposition 3.2 after (22). □

It is easy to check that Propositions 3.4, 3.6, and 3.7 remain true with $\mathrm{EP}(\bar{\mathcal{L}},\mathbb{R}^n\times\mathbb{R}^m_+)$ substituting for $\mathrm{EP}(\mathcal{L},\mathbb{R}^n\times\mathbb{R}^m_+)$. The only difference is that due to the smoothness of F_x and the h_i's, the Lagrangian condition (28) takes the form

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

It is a matter of routine to check that the proofs of Theorem 3.10, Theorem 3.12 and Corollary 3.13 also remain valid for LIALEM, resulting in the following convergence theorem.

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

i) f satisfies P1–P4 and P4" on $\mathbb{R}^n\times\mathbb{R}^n$,

ii) f is continuously differentiable,

iii) h_i is differentiable ($1\le i\le m$),

iv) the constraint qualification CQ *of Section 2 holds for the feasible set K.*

Take an exogenous sequence $\{\gamma_j\}\subset(\theta,\bar{\gamma}]$, for some $\bar{\gamma}>\theta$, where θ is the undermonotonicity constant of f resulting from P4, and a relative error tolerance $\sigma\in(0,1)$. Let $\{(x^j,\lambda^j)\}$ be the sequence generated by Algorithm LIALEM *applied to* $\mathrm{EP}(f,K)$. *If* $\mathrm{EP}(f,K)$ *has solutions then $\{(x^j,\lambda^j)\}$ converges to an optimal pair (x^*,λ^*) for* $\mathrm{EP}(f,K)$, *so that x^* belongs to $S(f,K)$. Additionally, if x^j is not a solution of* $\mathrm{EP}(f,K)$ *and $\breve{x}^j$ is the unique solution of* $\mathrm{EP}(\bar{\mathcal{L}}_j^e,\mathbb{R}^n)$ *with $e^j=0$, then there exists $\delta_j>0$ such that any $x\in B(\breve{x}^j,\delta_j)$ 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 equilibrium problem by a sequence of unconstrained ones. It is worthwhile to comment on the advantages of such a substitution in the equilibrium 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 $\mathrm{EP}(f, K)$ proposed in [14]. 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 [14] will be enough for satisfying our error criteria, as discussed in Section 3.

Another option consists of solving the system of equations resulting from (49) in the case of IALEM, namely

$$0 = \gamma_j(x - x^j) + \nabla F_x(x) + \sum_{i=1}^{m} \max\left\{0, \lambda_i^j + \frac{h_i(x)}{\gamma_j}\right\} \nabla h_i(x). \tag{65}$$

We observe that the right hand side of (65) 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 $\mathrm{EP}(f, K)$ was presented in [16], 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 [16]. Both an inexact Augmented Lagrangian method for $\mathrm{EP}(f, K)$ and its linearized version can be developed from Algorithm I in [16]. 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.

Acknowledgment

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] Bertsekas, D.P. On penalty and multiplier methods for constrained optimization problems. *SIAM Journal on Control and Optimization* **14** (1976) 216-235.

[2] Bianchi, M., Pini, R. Coercivity conditions for equilibrium problems. *Journal of Optimization Theory and Applications* **124** (2005) 79-92.

[3] Bianchi, M., Schaible, S. Generalized monotone bifunctions and equilibrium problems. *Journal of Optimization Theory and Applications* **90** (1996) 31-43.

[4] Blum, E. Oettli, W. From optimization and variational inequalities to equilibrium problems. *The Mathematics Student* **63** (1994) 123-145.

[5] Brezis, H., Nirenberg, L., Stampacchia, S. A remark on Ky Fan minimax principle. *Bolletino della Unione Matematica Italiana* **6** (1972) 293-300.

[6] Buys, J.D. *Dual algorithms for constrained optimization problems.* Ph.D. Thesis. University of Leiden, Leiden, The Netherlands, (1972).

[7] Flores-Bazán, F. Existence theorems for generalized noncoercive equilibrium problems: quasi-convex case. *SIAM Journal on Optimization* **11** (2000) 675-790.

[8] Hestenes, M.R. Multiplier and gradient methods. *Journal of Optimization Theory and Applications* **4** (1969) 303-320.

[9] Hiriart-Urruty, J.-B., Lemaréchal, C. *Convex Analysis and Minimization Algorithms*. Springer, Berlin (1993).

[10] Iusem, A.N. Augmented Lagrangian methods and proximal point methods for convex optimization. *Investigación Operativa* **8** (1999) 11-49.

[11] Iusem, A.N., Gárciga Otero, R. Inexact versions of proximal point and augmented Lagrangian algorithms in Banach spaces. *Numerical Functional Analysis and Optimization* **22** (2001) 609-640.

[12] Iusem, A.N., Kassay, G., Sosa, W. On certain conditions for the existence of solutions of equilibrium problems (to be published in *Mathematical Programming*).

[13] Iusem, A.N., Sosa, W. New existence results for equilibrium problems. *Nonlinear Analysis* **52** (2003) 621-635.

[14] Iusem, A.N., Sosa, W. Iterative algorithms for equilibrium problems. *Optimization* **52** (2003) 301-316.

[15] Iusem, A.N., Sosa, W. A proximal point method for equilibrium problems in Hilbert spaces (to be published).

[16] Iusem, A.N., Nasri, M. Inexact proximal point methods for equilibrium problems in Banach spaces (to be published in *Numerical Functional Analysis and Optimization*).

[17] Kort, B.W., Bertsekas, D.P. Combined primal-dual and penalty methods for convex programming. *SIAM Journal on Control and Optimization* **14** (1976) 268-294.

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

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

[20] Muu, L.D., Oettli, W. Convergence of an adaptive penalty scheme for finding constraint equilibria. *Nonlinear Analysis* **18** (1992) 1159-1166.

[21] Powell, M.J.D. Method for nonlinear constraints in minimization problems. In *Optimization* (R. Fletcher, editor). Academic Press, London (1969).

[22] Rockafellar, R.T. A dual approach to solving nonlinear programming problems by unconstrained optimization. *Mathematical Programming* **5** (1973) 354-373.

[23] Rockafellar, R.T. The multiplier method of Hestenes and Powell applied to convex programming. *Journal of Optimization Theory and Applications* **12** (1973) 555-562.

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

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

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

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