Proximal Point Methods and Augmented Lagrangian Methods for Equilibrium Problems

Doctoral Thesis by
Mostafa Nasri

Supervised by
Alfredo Noel Iusem

IMPA - Instituto Nacional de Matemática Pura e Aplicada
Rio de Janeiro, Brazil
August 2008

Allah Has taught man that which he knew not

(From the Holy Quran, Surah: 96, AL-Alaq, Verse: 5)

Dedicated to my wife

Acknowledgments

I would like to express my gratitude to professor Alfredo Noel Iusem, my supervisor, for his constant supports, encouragements, and many fruitful suggestions during this project. I appreciate his vast knowledge and expertise in many areas. Many thanks to the committee members of my thesis, Gabor Kassay, Luis Mauricio Graña Drummond, Mikhail Solodov, Rolando Gárciga Otero, and Susana Scheimberg, for their comments to improve this manuscript. In addition, I acknowledge my scholarship for my doctorate studies, granted jointly by CNPq and TWAS.

I am thankful to my roommates, Alireza Mohebbi Ashtiani, Amin Esfahani, Azam Afzali, Emilio Ashton Brazil, Hernan Maycol Falla, Julio Daniel Machado, Meysam Nassiri, Renato Soares Siqueira, and Saeed Tafazolian, during my stay in Brazil. Thanks also goes out to my friends at IMPA, in particular Damián R. Fernández Ferreyra.

Finally I am grateful to my family for all their love, patience, and support through my entire life.

Abstract

In this work we propose two classes of methods for solving equilibrium problems. The first one consists of two inexact proximal-like algorithms for solving equilibrium problems in reflexive Banach spaces, generalizing two methods, called *Inexact Proximal Point+Bregman Projection Method* and *Inexact Proximal Point-extragradient Method*, proposed by Iusem and Gárciga for finding zeroes of maximal monotone operators in reflexive Banach spaces, to the context of equilibrium problems.

We state our two proximal point algorithms formally and their convergence properties, proving that the sequence generated by each one of them converges to a solution of the equilibrium problem under reasonable assumptions. In order to achieve this goal, we introduce a suitable regularization for equilibrium problem, and then we use this regularization in our two methods, named *Inexact Proximal Point+Bregman Projection Method* (Algorithm IPPBP) and *Inexact Proximal Point-Extragradient Method* (Algorithm IPPE).

These two methods are hybrid methods, which invoke the proximal point method at the beginning of each iteration for obtaining an auxiliary point. In the case of Algorithm IPPBP, this auxiliary point is used for constructing a hyperplane separating the current iterate from the solution set of the problem. The next iterate of Algorithm IPPBP is then obtained by projecting the current iterate onto the separating hyperplane. In the case of Algorithm IPPE, we use again the proximal point method in order to get an auxiliary point, from which an extragradient step is performed.

The second class of methods considered here consists of augmented Lagrangian methods for solving finite dimensional equilibrium problems whose feasible sets are defined by convex inequalities, extending the proximal augmented Lagrangian method for constrained optimization.

We propose two Lagrangian functions, and consequently two augmented Lagrangian functions for equilibrium problems. Using these functions we develop two augmented Lagrangian methods, called *Inexact Augmented Lagrangian-Extragradient Method* (Algorithm IALE) and a variant of Algorithm IALE, called *Linearized Inexact Augmented Lagrangian-Extragradient Method* (Algorithm LIALE). At each iteration of these methods, primal variables are updated by solving an unconstrained equilibrium problem, and then dual variables are updated through a closed formula. We provide a full convergence analysis, allowing for inexact solution of the subproblem of Algorithm IALE, applying the convergence theorem for Algorithm IPPE.

Along the same line, we propose two additional augmented Lagrangian methods for solving equilibrium problems, to be called *Inexact Augmented Lagrangian Projection Method* (Algorithm IALP) and *Linearized Inexact Augmented Lagrangian Projection Method* (Algorithm LIALP), whose convergence properties are proved using the convergence results for Algorithm IPPBP.

We also show that each equilibrium problem can be reformulated as a variational inequality problem. This approach reduces the augmented Lagrangian methods to implementable methods which substitute each subproblem of the augmented Lagrangian methods by a system of algebraic equations which admits a unique solution.

Keywords: Augmented Lagrangian method, Bregman distance, Convex minimization problem, Equilibrium problem, Inexact solution, Point-to-set operator, Proximal point method, Variational inequality problem.

Contents

Introduction

The equilibrium problem, in the sense that is used in this thesis, was defined formally for the first time in the work of Blum and Oettli in 1994. In [8] these authors studied the issue of existence of solutions of this problem and its relation with other well known problems in optimization. They showed that equilibrium problems encompass, among their particular cases, complementarity problems, fixed point problems, minimization problems, Nash equilibrium problems in noncooperative game, saddle point problems, and variational inequality problems (see [8]). Later on, it was shown that the equilibrium problem includes as a particular case vector minimization problems (see [50]) and also generalized Nash games (see [49]).

The equilibrium problem has been extensively studied in recent years, with emphasis on existence results (see, e.g., [6], [7], [10], [21], [27], [31], and [49]). In terms of computational methods for equilibrium problems which is the main purpose of this manuscript, several references can be found in the literature. Among those of interest, we mention the algorithms introduced in [20], [32], [35], [36], [37], [42], [43], [44], [45], [47], and [48] which are proximal-like methods, as well as the ones proposed in [30] which are projection-like methods. Methods based on a gap function approach can be found in [40]. Furthermore, Newton-like methods for solving the same problem has been introduced in [2] and penalty-like methods in [46].

The first type of methods, developed in Chapter 2, consists of proximal point algorithms, whose origins can be traced back to [39] and [41]. The proximal point method attained its basic formulation in the work of Rockafellar [59], where it is presented as an iterative algorithm for finding zeroes of a maximal monotone point-to-set operator in a Hilbert space. At each iteration, it requests finding a zero of a regularized maximal monotone point-to-set operator, which is the sum of the given operator and a shifted multiple of the identity operator. Proximal point methods for the same problem but in Banach spaces, can be traced back to [34], and were furtherly developed in [12], [13], [22] and [26]. In this set-up, one works with maximal monotone point-to-set operators which map the Banach space to the power set of its topological dual space. In this situation, the regularized operator at each iteration, instead of the shifted multiple of the identity, invokes the derivative of an auxiliary function satisfying certain conditions (see [26]), which reduces to the square of the norm in Hilbert spaces. Regarding proximal point method for equilibrium problems, an exact version of this method for nonmonotone and monotone equilibrium problems in a Hilbert space has been recently proposed in [32] and [42] respectively, besides the inexact versions in finite dimensional spaces which are available in [35], [36] and [37]. Other proximal-like methods for equilibrium problems can be found in [20], [43], [44], [45], [47] and [48] in Hilbert spaces.

Since finding exact solution of a regularized equilibrium problem at each iteration can be quite difficult our even impossible, in practice, the use of approximate solutions is essential for devising implementable algorithms. This issue was already dealt with in [37] for equilibrium problems, where it was assumed that the j-th subproblem was solved with an error bounded by a certain $\varepsilon_j > 0$, and the convergence results were preserved assuming that $\sum_{j=1}^{\infty} \varepsilon_j < \infty$. This summability condition is undesirable because it cannot be ensured in practice and also requires increasing accuracy along the iterative process. In the current work, we will build upon the results of [26] and [32], obtaining exact and inexact proximal point methods for equilibrium problems in Banach spaces with better error bounds than the one which requests the summability condition above.

The second type of methods, developed in Chapter 3, consists of augmented Lagrangian methods, or more generally multiplier methods, for equilibrium problems. The augmented Lagrangian method for equality constrained optimization problems (nonconvex, in general) was introduced in [23] and [54]. Its extension to inequality constrained problems started with [17] and was continued in [5], [38], [56], [57], and [58]. This kind of methods has been recognized as efficient strategies for solving constrained optimization problems whose feasible sets given by convex inequalities (see, e.g., [4]). In principle, augmented Lagrangian methods are implementable methods, while proximal point methods are not. Nevertheless, the convergence theorems of latter methods can be applied for proving convergence theorems for the former methods, as it has been already done in [58] for minimization problems. We will follow this path in the equilibrium problem setting. In fact, we will introduce proper exact and inexact proximal augmented Lagrangian methods for equilibrium problems in finite dimensional spaces, whose convergence theorems invoke the convergence results established in Chapter 2 for proximal point methods. To our knowledge, the closest approach to the Lagrangian method proposed here can be found in [1], where the feasible set is assumed to be similar to ours, i.e. 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 functions. In this reference, primal-dual methods are proposed. However, no Lagrangian function as in (3.2), or augmented Lagrangian function as in (3.4) appear in this reference, so that from an algorithmical point of view our approach is completely unrelated to the one in [1].

The outline of this thesis is the following. Chapter 1 is devoted to some preliminary materials used in the next chapters as well as a short review on proximal point and augmented Lagrangian methods for several problems. In Chapter 2 we associate to each given equilibrium problem a sequence of regularized equilibrium problems which is the starting point for proposing proximal point methods for equilibrium problems. In other words, given an equilibrium problem, we show that such a sequence of regularization problems exists and the sequence composed of the unique solutions of the regularized problems solved at each iteration converges to some solution of the original problem when certain conditions are met. We then formally state our two inexact proximal point algorithms, i.e., Algorithm IPPBP and Algorithm IPPE, and provide a full convergence analysis for both of them. Using a reformulation technique for equilibrium problems, we finish this chapter by proving that the solution set of an equilibrium problem coincides with the set of zeroes of a certain point-to-set operator, which allows us to refine our convergence results. Chapter 3 begins with the introduction of Algorithm IALE for solving equilibrium problem. Next, we establish the convergence properties of the algorithm through the construction of an appropriate proximal point method for a related equilibrium problem. We also develop linearized version of Algorithm IALE, called Algorithm LIALE, for the case in which the underlying equilibrium problem is smooth. Besides these two augmented Lagrangian methods, we construct two other variants of augmented Lagrangian methods for solving equilibrium problems, called Algorithm IALP and Algorithm LIALP, which are closely connected to Algorithm IPPBP.

We emphasize that all previously established results have been concentrated in Chapter 1, and that the results appearing in Chapters 2 and 3, taken from our papers [28] and [29], are, to our knowledge, substantially different from those which can be found in the previous literature on the subject.

Basic Notation and Terminology

Algorithm IPPBP: the Inexact Proximal Point+Bregman Projection Method,

Algorithm IPPE: the Inexact Proximal Point-Extragradient Method,

Algorithm IALE: the Inexact Augmented Lagrangian-Extragradient Method,

Algorithm LIALE: the Linearized Inexact Augmented Lagrangian-Extragradient Method,

Algorithm IALP: the Inexact Augmented Lagrangian Projection Method,

Algorithm LIALP: the Linearized Inexact Augmented Lagrangian Projection Method,

$\mathrm{EP}(f, K)$: the equilibrium problem whose objective function is f and whose feasible set is K,

$S_E(f, K)$: the solution set of $\mathrm{EP}(f, K)$,

$S^d(f, K)$: the solution set of the dual problem of $\mathrm{EP}(f, K)$,

$\mathrm{VIP}(T, C)$: the variational inequality problem whose objective operator is T and whose feasible set is C,

$S_V(T, C)$: the solution set of $\mathrm{VIP}(T, C)$,

CQ: the constraint qualification,

X: the real vector space,

X^*: the dual space of X,

X^{**}: the bidual space of X,

$\|\cdot\|$: the norm of space X,

$\|\cdot\|_*$: the norm of space X^*,

$\|\cdot\|_{**}$: the norm of space X^{**},

$\langle\cdot,\cdot\rangle$: the duality coupling in $X^* \times X$,

H: the Hilbert space,

B: the Banach space,

B^*: the topological dual of B,

$\mathbb{R}^n$: the n-dimensional Euclidean space,

$\mathbb{R}^m_+ = \{(x_1, \ldots, x_m) \in \mathbb{R}^m \mid x_i \geq 0 \ \ (1 \leq i \leq m)\}$: the nonnegative orthant in $\mathbb{R}^m$,

$\mathbb{R}_{++}$: the set of positive real numbers,

$H = \{y \in B : \langle v, y - \tilde{y} \rangle = 0\}$: the hyperplane in B associated to $v \in B^*$ and $\tilde{y} \in B$,

$H^- = \{y \in B : \langle v, y - \tilde{y} \rangle \leq 0\}$: the negative part of B associated to the hyperplane H,

$H^+ = \{y \in B : \langle v, y - \tilde{y} \rangle \geq 0\}$: the positive part of B associated to the hyperplane H,

$dom(h)$: the domain of a function h,

$int(dom(h))$: the interior of the domain of a function h,

$dom(T)$: the domain of an operator T,

I_C: the indicator function of a set C,

∂g: the subdifferential of a convex function g,

$N_C(x)$: the normal cone of a set C at a point x,

$\mathcal{P}(C)$: the power set of a set C,

$\mathcal{F}$: the family of functions which are strictly convex, lower semicontinuous, and Gâteaux differentiable in the interior of their domains,

$D_g(x, y)$: the Bregman distance between x and y with respect to a function g,

$\nu_g(x, \cdot)$: the modulus of total convexity of a function g at a point x,

$A(C)$: the affine hull of a set C,

$ri(C)$: the relative interior of a set C,

$cl(C)$: the closure of a set C,

∂C: the boundary of a set C,

$\ker(w)$: the kernel of a linear function w,

$B(x, \delta)$: the open ball with radius δ centered at x.

Chapter 1

Background Materials

In this chapter we collect some mathematical facts including definitions and theorems used in sequel. We also review proximal point methods for finding zeroes of operators, and augmented Lagrangian methods for solving constrained convex minimization problem, which are the motivations for our contributions in this thesis.

1.1 Elements of Topology

We begin our discussion with a few words about topology on a real vector space, say X. All the following materials can be found in any functional analysis book, like [3], [9], and [55].

Definition 1.1.1. *Let $h : X \to \mathbb{R} \cup \{-\infty, +\infty\}$ be a function on a real topological vector space (X, Γ).*

i) h is proper if $dom(h) \neq \emptyset$.

ii) h is lower semicontinuous if for each $x \in dom(h)$ and $\{x^j\}_{j=0}^{\infty} \subset dom(h)$ converging to x, it holds that $\liminf_{j\to\infty} h(x^j) \geq h(x)$.

iii) h is upper semicontinuous if for each $x \in dom(h)$ and $\{x^j\}_{j=0}^{\infty} \subset dom(h)$ converging to x, it holds that $\limsup_{j\to\infty} h(x^j) \leq h(x)$.

Definition 1.1.2. *Consider a real topological vector space (X, Γ). The topological dual space with respect to topology Γ on X, denoted by X^*, is the space of all linear functionals defined on X which are continuous with respect to Γ.*

Definition 1.1.3. *Consider a real topological vector space (X, Γ). The Γ-weak topology on X is defined as the coarsest topology (the topology with the fewest open sets) on X for which each element in X^* is continuous. In other words, X^* induces a topology on X which is contained in Γ and preserves the continuity of each element in X^*.*

Definition 1.1.4. *The function* $\|\cdot\| : X \to \mathbb{R}$ *is called a norm on* X *if the following assumptions are met.*

i) $\|x\| \geq 0, \ \forall x \in X$, *and that* $\|x\| = 0$ *if* $x = 0$.

ii) $\|\lambda x\| = |\lambda| \, \|x\|, \ \forall \lambda \in \mathbb{R}, \forall x \in X$.

iii) $\|x + y\| \leq \|x\| + \|y\|, \ \forall x, y \in X$.

Definition 1.1.5. *A normed vector space, denoted by a pair* $(X, \|\cdot\|)$*, is a vector space* X *with a norm* $\|\cdot\|$ *defined on* X.

Definition 1.1.6. *Consider a normed vector space* $(X, \|\cdot\|)$. *The topology induced on* X *by the norm is called the strong topology, and the topology induced on* X *by the dual space* X^* *is called the weak topology.*

Definition 1.1.7. *Consider the normed vector space* $(X, \|\cdot\|)$. *A sequence* $\{x^j\}_{j=0}^{\infty} \subset X$ *is a Cauchy sequence if*

$$\forall \epsilon > 0 \, \exists k \ \text{such that} \ \left\|x^i - x^l\right\| < \epsilon, \ \forall i, l \geq k.$$

Definition 1.1.8. *A real Banach space is a real vector space with a norm such that every Cauchy sequence converges in the strong topology.*

Definition 1.1.9. *The function* $\langle \cdot, \cdot \rangle : X^* \times X \to \mathbb{R}$ *defined as* $\langle \phi, x \rangle = \phi(x)$ *for all* $x \in X$ *and all* $\phi \in X^*$ *is called duality coupling.*

Definition 1.1.10. *Take* $x \in X$ *and define the functional* $\phi_x : X^* \to \mathbb{R}$ *as* $\phi_x(x^*) = \langle x^*, x \rangle$. *The coarsest topology on* X^* *for which* ϕ_x *is continuous for every* $x \in X$ *is called the weak* topology.*

In principle, the weak* topology is contained in the weak topology induced by the bidual space $X^{**} := (X^*)^*$, and this weak topology is contained in the strong topology induced by the norm defined as $\|\phi\|_* := \sup_{\|x\| \leq 1} \langle \phi, x \rangle, \ \forall \phi \in X^*$. Note that ϕ_x is continuous on X^* (i.e. continuous on the strong topology) since $|\phi_x(x^*)| \leq \|x^*\|_* \|x\|$ for every $x^* \in X^*$, and that ϕ_x is an element of the dual space of X^* (i.e. $\phi_x \in X^{**}$).

Consider the mapping $J : X \to X^{**}$ defined as $J(x) := \phi_x$ for every $x \in X$. That is, J maps x to the functional on X^* given by evaluation at x. As a consequence of the Hahn-Banach theorem, such a J is norm-preserving (i.e., $\|J(x)\|_{**} = \|x\|$, in which $\|\cdot\|_{**}$ is the norm of X^{**}), and hence injective, called the natural embedding of X into X^{**}.

Definition 1.1.11. *A Banach space* B *is called reflexive if the natural embedding* $J : B \to B^{**}$ *defined as* $J(x) := \phi_x$ *is surjective.*

Proposition 1.1.12. *Assume that* B *is a Banach space. The following statements are equivalent.*

i) B is reflexive.

ii) B^* is reflexive.

iii) The closed unit ball in B is compact in the topology induced by B^*.

iv) Every closed bounded subset in B is weakly (sequentially) compact.

Proof. See pages 44 and 45 of [9]. ■

Definition 1.1.13. *Given a real vector space X, a subset C of X is called convex if*

$$\forall x, y \in C,\ \forall t \in [0,1]: \quad tx + (1-t)y \in C.$$

Reflexive Banach spaces enjoy the following properties, which we will use frequently in Chapter 2.

Proposition 1.1.14. *Assume that B is a reflexive Banach space.*

i) If C is a closed convex bounded subset of B, then C is weakly compact in the topology induced by B^*.

ii) Every bounded sequence in B has a weakly convergent subsequence in the topology induced by B^*.

iii) If C is a convex subset of B, then C is weakly closed (in the topology induced by B^*) if and only if it is strongly closed.

iv) If $h : B \to \mathbb{R} \cup \{+\infty\}$ is a lower semicontinuous function, then h has at least one minimizer over each nonempty closed convex bounded subset of B.

Proof. See pages 38, 46 and 50 of [9]. ■

Now we present the concept of the relative interior of a set.

Definition 1.1.15. *Given $C \subset X$, the affine hull of C, denoted by $A(C)$, is the set of all affine combinations of elements of C, i.e.,*

$$A(C) = \left\{ \sum_{i=1}^{k} \alpha_i x^i \middle| x^i \in C,\ \alpha_i \in \mathbb{R},\ \sum_{i=1}^{k} \alpha_i = 1, k = 1, 2, \ldots \right\}.$$

Definition 1.1.16. *Given $C \subset X$, the relative interior of C, denoted by $ri(C)$, is defined as its interior within the affine hull of C. In other words,*

$$ri(C) = \{x \in C : \exists \delta > 0, B(x, \delta) \cap A(C) \subset C\}.$$

It is remarkable that in infinite dimensional spaces the relative interior of a set can be an empty set. In Chapter 2, through the convergence analysis of our algorithms in Banach spaces, we will assume that the relative interior of the feasible sets of equilibrium problems are nonempty.

Next we recall a Lemma from functional analysis which will be useful in Section 2.4.

Lemma 1.1.17. *Assume that* $\Lambda_1, \dots, \Lambda_n$ *and* Λ *are linear functionals on a vector space* X*. Let*

$$\Omega = \{x : \Lambda_1 = \cdots = \Lambda_n = 0\}.$$

The following three properties are equivalent.

i) There are scalars $\alpha_1, \dots, \alpha_n$ *such that*

$$\Lambda = \alpha_1 \Lambda_1 + \cdots + \alpha_n \Lambda_n.$$

ii) There exists $\gamma < \infty$ *such that*

$$|\Lambda x| \leq \gamma \max_{1 \leq i \leq n} |\Lambda_i x| \quad \forall x \in X.$$

iii) $\Lambda x = 0$ *for every* $x \in \Omega$*.*

Proof. See Lemma 3.9 of [60]. ■

1.2 Gâteaux and Fréchet Derivatives

The material in this section has been taken from [18] and [65].

Definition 1.2.1. *Consider* $h : B_1 \to B_2$*, where* B_1 *and* B_2 *are two real Banach spaces with norms* $\|\cdot\|_1$ *and* $\|\cdot\|_2$*, respectively.*

i) We say that h *has directional derivative at* x *in direction* $d \in B_1$*, denoted by* $h'(x, d)$*, if*

$$\lim_{t \to 0^+} \frac{h(x + td) - h(x)}{t} = h'(x, d).$$

ii) We say that h is Gâteaux differentiable at x, denoted by $h'(x)$, if there exists a linear and continuous operator (i.e., an element in $L(B_1, B_2)$) such that

$$\lim_{t\to 0} \frac{h(x+td)-h(x)}{t} = \langle h'(x), d\rangle,$$

for all $d \in B_1$.

iii) We say that h is Fréchet differentiable at x if the above limit exists uniformly on the unit sphere, i.e.,

$$\lim_{t\to 0} \sup_{\|d\|_1=1} \left\| \frac{h(x+td)-h(x)}{t} - \langle h'(x), d\rangle \right\|_2 = 0.$$

The relation between Gâteaux and Fréchet derivative is stated in the next proposition.

Proposition 1.2.2. *Consider $h : B_1 \to B_2$, where B_1 and B_2 are two real Banach spaces.*

i) If h is Fréchet differentiable at x, then it is Gâteaux differentiable at x.

ii) If h is Gâteaux differentiable on some neighborhood of x and its Gâteaux derivative is continuous at x, then h is Fréchet differentiable at x.

iii) If h is Fréchet differentiable at x, then h is continuous at x.

Proof. See Proposition 4.8 of [65]. ■

Definition 1.2.3. *Assume that $h(x) = \|x\|$, where $\|\cdot\|$ denotes the norm in the real Banach space B.*

i) B is called smooth if h is Gâteaux differentiable on $B \setminus \{0\}$.

ii) B is called locally uniformly smooth if h is Fréchet differentiable on $B \setminus \{0\}$.

iii) B is called uniformly smooth if h is uniformly Fréchet differentiable on the unit sphere, i.e., h is Fréchet differentiable and

$$\lim_{t\to 0} \sup_{\|x\|=\|y\|=1} \left\| \frac{h(x+ty)-h(x)}{t} - \langle h'(x), y\rangle \right\| = 0.$$

Definition 1.2.4. *Assume that B is a real Banach space.*

i) B is called strictly convex if for all $x, y \in B$ with $\|x\| = \|y\| = 1$ and $x \neq y$, it holds that $\|\lambda x + (1-\lambda)y\| < 1$ for all $\lambda \in (0,1)$.

ii) B is called locally uniformly convex if for each x in the unit sphere and each $\epsilon > 0$, there exists $\delta(x) > 0$, so that $\|y\| = 1$ and $\|x - y\| \geq \epsilon$ implies $\|x + y\| \leq 2[1 - \delta(x)]$.

iii) B is called uniformly convex if for all $\epsilon > 0$, there exists $\delta > 0$, so that $\|x\| = \|y\| = 1$ and $\|x - y\| \geq \epsilon$ implies $\|x + y\| \leq 2(1 - \delta)$.

Hilbert spaces, $\mathcal{L}^p(\Omega)$, and Sobolev spaces $W^{m,p}(\Omega)$ are well known examples of uniformly convex spaces.

Proposition 1.2.5. *Let B be a real reflexive Banach space. Then the following two statements are true.*

i) B is strictly convex (respectively smooth) if and only if B^* is smooth (respectively strictly convex).

ii) B is uniformly convex (respectively uniformly smooth) if and only if B^* is uniformly smooth (respectively uniformly convex).

Proof. See pages 43, 52, and 53 of [18]. ∎

1.3 Point-to-Set Operators

Here we state some properties of point-to-set or set-valued operators.

Definition 1.3.1. *Let B be a Banach space and B^* its topological dual space.*

i) A point-to-set operator is a map $T : B \to \mathcal{P}(B^*)$.

ii) The set $dom(T) = \{x \in B : T(x) \neq \emptyset\}$ is called the domain of T.

iii) The graph of T is the set $G(T) = \{(v, x) \in B^* \times B : v \in T(x)\}$.

iv) $x \in B$ is called a zero of T if $0 \in T(x)$.

v) The graph of T is demiclosed, if for all sequences $\{x^j\}_{j=0}^{\infty} \subset B$, weakly convergent (strongly convergent) to $x \in B$ and for all sequences $\{v^j\}_{j=0}^{\infty} \subset B^*$ strongly convergent (weakly convergent) to $v \in B^*$, with $(x^j, v^j) \in G(T)$ for all j, it holds that $v \in T(x)$.*

vi) T is monotone if $\langle w - w', x - x' \rangle \geq 0$ for all x, $x' \in B$ and all $w \in T(x)$, $w' \in T(x')$.

vii) *T is maximal monotone if it is monotone and its graph is not properly contained in the graph of any other monotone operator.*

viii) *T is pseudomonotone if the following property holds: if* $\langle v, x-y\rangle \geq 0$ *for some* $x, y \in dom(T)$ *and some* $v \in T(y)$ *then* $\langle u, x-y\rangle \geq 0$ *for all* $u \in T(x)$.

Subdifferential operators are prototypical examples of maximal monotone operators (see Proposition 1.4.7). It is also remarkable that graphs of maximal monotone operators are demiclosed, as stated in the next proposition.

Proposition 1.3.2. *Assume that B is a reflexive Banach space. If* $T : B \to \mathcal{P}(B^*)$ *is maximal monotone, then its graph is demiclosed.*

Proof. See page 105 of [52], for instance. ■

1.4 Convex Analysis

We will analyze our proposed algorithms using some tools of convex analysis which we present next. We follow [11], [14], [18] and [24] in this section.

Definition 1.4.1. *Let* $g : B \to \mathbb{R} \cup \{+\infty\}$.

i) *g is called convex if* $g(\alpha x+(1-\alpha)y) \leq \alpha g(x)+(1-\alpha)g(y)$ *for all* $\alpha \in (0,1)$ *and all* $x, y \in B$.

ii) *g is called concave if* $-g$ *is convex.*

iii) *g is called strictly convex if* $g(\alpha x+(1-\alpha)y) < \alpha g(x)+(1-\alpha)g(y)$ *for all* $\alpha \in (0,1)$ *and all* $x, y \in B$ *with* $x \neq y$.

iv) *The set* $\partial g(x) = \{v \in B^* : g(x) + \langle v, y-x\rangle \leq g(y),\ \forall y \in B\}$ *is called the subdifferential of g at x.*

Proposition 1.4.2. *Assume that B is a reflexive Banach space. Let* $g : B \to \mathbb{R} \cup \{+\infty\}$ *be a convex function whose domain has nonempty interior. Then the following statements hold.*

i) *For any* $x \in dom(g)$, *the subdifferential* $\partial g(x)$ *is convex and weak* closed.*

ii) *If g is continuous on* $int(dom(g))$, *then for each* $x \in int(dom(g))$, *the set* $\partial g(x)$ *is nonempty and weak* compact.*

iii) *If g is lower semicontinuous, then it is locally Lipschitz on* $int(dom(g))$.

Proof. See pages 6 and 8 of [14]. ∎

Proposition 1.4.3. *Assume that $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a convex function whose domain has nonempty interior. Then g is locally Lipschitz on $int(dom(g))$, and consequently $\partial g(x)$ is a nonempty subset of $\mathbb{R}^n$ for each $x \in int(dom(g))$.*

Proof. See Proposition 1.4.2 and page 174 of [24]. ∎

We will need the following property of the subdifferential of convex functions.

Proposition 1.4.4. *Let $g_i : B \to \mathbb{R} \cup \{+\infty\}$ $(i = 1, 2)$ be two proper, convex, and lower semicontinuous functions. Assume that there exists $x \in dom(g_1) \cap dom(g_2)$ at which one of the functions is continuous. For each $y \in dom(g_1) \cap dom(g_2)$, it holds that*

$$\partial(g_1 + g_2)(y) = \partial g_1(y) + \partial g_2(y).$$

Proof. See Theorem 3.5.7 of [11]. ∎

Definition 1.4.5. *Let $C \subset B$. The indicator function of C at x is defined as*

$$I_C(x) = \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{otherwise.} \end{cases}$$

Definition 1.4.6. *Let $C \subset B$. The normal cone of set C is the operator $N_C : B \to \mathcal{P}(B^*)$ defined at x as*

$$N_C(x) = \begin{cases} \{z \in B^* \mid \langle z, y - x \rangle \leq 0, \ \forall y \in C\} & \text{if } x \in C \\ \emptyset & \text{otherwise.} \end{cases} \tag{1.1}$$

It is known that $\partial I_C(x) = N_C(x)$ for each $x \in B$.

Proposition 1.4.7. *Assume that $g : B \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous convex function. In this case, the subdifferential operator, $\partial g : B \to \mathcal{P}(B^*)$, is a maximal monotone operator.*

Proof. See page 168 of [18], for instance. ∎

Definition 1.4.8. *Let $g : B \to \mathbb{R} \cup \{+\infty\}$ be a convex and Gâteaux differentiable function on $int(dom(g)) \neq \emptyset$.*

i) The Bregman distance with respect to g is the function $D_g : dom(g) \times int(dom(g)) \to \mathbb{R}_+$ *defined as* $D_g(x,y) = g(x) - g(y) - \langle g'(y), x-y \rangle$.

ii) The modulus of total convexity of g is the function $\nu_g : int(dom(g)) \times [0, +\infty) \to \mathbb{R}_+$ *defined as* $\nu_g(x,t) = \inf\{D_g(y,x) : y \in B, \|y-x\| = t\}$.

iii) g *is said to be a totally convex function if* $\nu_g(x,t) > 0$ *for all* $t > 0$ *and all* $x \in int(dom(g))$.

iv) g *is said to be a uniformly totally convex function on* $E \subset int(dom(g))$ *if for all* $t > 0$ *and all bounded subsets* $C \subset E$, *it holds that* $\inf_{x \in C} \nu_g(x,t) > 0$.

Example 1.4.9. *Assume that B is a uniformly smooth and uniformly convex Banach space. It has been shown in* [15] *that* $h(x) = \|x\|^r$ *is uniformly totally convex for each* $r > 1$.

It is obvious that positivity of $D_g(x,y)$ for all $x, y \in B$ $(x \neq y)$ is equivalent to strict convexity of g.

The following properties of Bregman distance are used in the convergence analysis of our methods.

Proposition 1.4.10. *Assume that* $g : B \to \mathbb{R} \cup \{+\infty\}$ *is convex, lower semicontinuous and that* $int(dom(g)) \neq \emptyset$. *The following statements are true.*

i) The function g is Gâteaux differentiable on int(dom(g)) if and only if the operator $\partial g(\cdot) : int(dom(g)) \to B^*$ *is a point-to-point operator, i.e., single valued. In the affirmative case,* $\partial g(x) = \{g'(x)\}$ *for each* $x \in int(dom(g))$, *and*

a) D_g *is well defined and* $D_g(\cdot, x)$ *is convex,*

b) for any $x, z \in int(dom(g))$ *and* $y \in dom(g)$,

$$D_g(y,x) - D_g(y,z) - D_g(z,x) = \langle g'(x) - g'(z), z - y \rangle,$$

c) if g is strictly convex, then $D_g(y,x) > 0$ *for all* $y \in dom(g)$ *with* $x \neq y$.

ii) If g is Fréchet differentiable, then $g' : int(dom(g)) \to B^*$ *is norm-to-norm continuous and* D_g *is continuous on* $int(dom(g)) \times int(dom(g))$.

Proof. The first statement in (i) was proved in 1.1.10 of [14], (i)(a) in 1.1.3 of [14], (i)(b) in 1.3.9 of [14], and (i)(c) in 1.1.9 of [14]. (ii) follows from the corollary in page 20 of [53] ■

Proposition 1.4.11. *Assume that* $g : B \to \mathbb{R} \cup \{+\infty\}$ *is convex, lower semicontinuous and that* $int(dom(g)) \neq \emptyset$. *Additionally assume that* $x \in int(dom(g))$, $s \geq 0$, *and that* $t \geq 0$. *In this situation,* $\nu_g(x, st) \geq s\nu_g(x,t)$.

Proof. See page 18 of [14]. ∎

Definition 1.4.12. *Assume that $T : B \to \mathcal{P}(B^*)$ is a point-to-set operator and the function $g : B \to \mathbb{R}$ is Fréchet differentiable and totally convex. T is θ-undermonotone if there exists $\theta \geq 0$ such that $\langle u - v, x - y\rangle \leq \theta\langle g'(x) - g'(y), x - y\rangle$ for all $x, y \in dom(T)$ and for all $u \in T(x), v \in T(y)$.*

1.5 Regularization Functions

Our convergence analysis demands an auxiliary function $g : B \to \mathbb{R} \cup \{+\infty\}$, which is strictly convex, lower semicontinuous, and Gâteaux differentiable in the interior of its domain. We will denote the family of such functions by $\mathcal{F}$. We denote as g' the Gâteaux derivative of g.

We will present next some additional conditions on g, called regularization conditions, needed in convergence analysis (see Theorem 2.3.5 and Theorem 2.3.8) and for regularizing a given problem (see Proposition 1.8.1 and Proposition 2.1.1). We will call such a g a regularization function.

H1: The level sets of $D_g(x, \cdot)$ are bounded for all $x \in dom(g)$.

H2: $\inf_{x\in C} \nu_g(x, t) > 0$ for all bounded set $C \subset int(dom(g))$ and all $t > 0$.

H3: g' is uniformly continuous on bounded subsets of $int(dom(g))$.

H4: g' is onto, i.e., for all $y \in B^*$, there exists $x \in int(dom(g))$ such that $g'(x) = y$.

H5: $\lim_{\|x\|\to+\infty}[g(x) - \rho\|x - z\|] = +\infty$ for all fixed $z \in C$ and $\rho \geq 0$ in which $C \subset dom(g)$.

H6: Take $C \subset int(dom(g))$, if $\{y^j\}_{j=0}^{\infty}$ and $\{z^j\}_{j=0}^{\infty}$ are sequences in C which converge weakly to y and z, respectively, and if $y \neq z$, then

$$\liminf_{j\to\infty} \left|\langle g'(y^j) - g'(z^j), y - z\rangle\right| > 0.$$

These properties, with the exception of H5, were identified in [26]. We make a few remarks on them. H2 is known to hold when g is lower semicontinuous and uniformly convex on bounded sets (see [16]). H5, introduced here for the first time in connection with Bregman functions and distances, is a form of coercivity. It has been proved in page 75 of [14] that sequential weak-to-weak* continuity of g' ensures H6.

It is important to check that functions satisfying these properties are available in a wide class of Banach spaces. The prototypical example is $g(x) = \frac{1}{2}\|x\|^2$, in which case g' is the duality operator, and the identity operator in the case of Hilbert space. It is convenient to deal with a general g rather than just the square of the norm because in Banach spaces this function lacks the privileged status it enjoys in Hilbert spaces. In the spaces $\mathcal{L}^p$ and ℓ_p, for instance, the function $g(x) = (1/p)\|x\|^p$ leads to simpler calculations than the square of the norm. It has been shown in [26] that the function $g(x) = r\|x\|^s$, works satisfactorily in any reflexive, uniformly smooth and uniformly convex Banach space, for any $r > 0$, $s > 1$, as established in the following proposition.

Proposition 1.5.1.

i) If B is a uniformly smooth and uniformly convex Banach space, then $g(x) = r\|x\|^s$ satisfies H1–H5 *for all $r > 0$ and all $s > 1$.*

ii) If B is a Hilbert space, then $g(x) = \frac{1}{2}\|x\|^2$ satisfies H6*. The same holds for $g(x) = \frac{1}{p}\|x\|^p$ when $B = \ell_p$ $(1 < p < \infty)$.*

Proof. The result of item (i) for properties H1 through H4, as well as item (ii), were proved in Proposition 2 of [26]. For H5, note that for the function of interest we have

$$g(x)-\rho\|x-z\| = r\|x\|^s-\rho\|x-z\| \geq r\|x\|^s-\rho\|x\|-\rho\|z\| = \|x\|\left[r\|x\|^{s-1}-\rho\right]-\rho\|z\|, \tag{1.2}$$

and the rightmost expression of (1.2) goes to ∞ as $\|x\| \to \infty$ because $s - 1 > 0$. ■

We remark that the only problematic property is H6, in the sense that the only example we have of a nonhilbertian Banach space for which we know functions satisfying it is ℓ_p with $1 < p < \infty$. As we will see in Section 2.3 and Section 2.4, most of our convergence results demand only H1–H5.

We will utilize the following facts in our convergence analysis.

Proposition 1.5.2. *Assume that g satisfies* H2*. Let $\{x^j\}_{j=0}^{\infty} \subset int(dom(g))$ and $\{y^j\}_{j=0}^{\infty} \subset dom(g)$ be two sequences such that at least one of them is bounded. If $\lim_{j\to\infty} D_g(y^j, x^j) = 0$, then $\lim_{j\to\infty}\|x^j - y^j\| = 0$.*

Proof. See Proposition 5 in [26]. ■

Proposition 1.5.3. *If g satisfies* H3*, then both g and g' are bounded on bounded subsets of B.*

Proof. See Proposition 4 in [26]. ■

1.6 Bregman Projection

Here, we present some properties of the Bregman projection on Banach spaces which is a generalization of the orthogonal projection in Hilbert spaces. A full discussion about this issue can be found in [14]

Definition 1.6.1. *Assume that B is a reflexive Banach space. Let $g \in \mathcal{F}$ be a totally convex function on $int(dom(g))$ satisfying* H1*, defined in Section 1.5. The Bregman projection of $x \in int(dom(g))$ onto $C \subset dom(g)$, denoted by $\Pi_C^g(x)$, is defined as unique point in C satisfying*

$$\Pi_C^g(x) = \operatorname*{argmin}_{y \in C} D_g(y, x). \tag{1.3}$$

We mention that the uniqueness of the Bregman projection is a consequence of strictly convexity of g, which ensures that $D_g(\cdot, x)$ is strictly convex on $dom(g)$, containing C. Note that total convexity of g implies strict convexity of g. The next proposition characterizes the Bregman projection.

Proposition 1.6.2. *Assume that B is a reflexive Banach space. Let $g \in \mathcal{F}$ be a totally convex function on $int(dom(g))$ satisfying* H1*. Take $C \subset dom(g)$. In this situation, the following two statements holds.*

i) The operator $\Pi_C^g : int(dom(g)) \to C$ is well defined.

ii) Assume that $C \subset int(dom(g))$ and $x \in int(dom(g))$. We have that $\bar{x} = \Pi_C^g(x)$ if and only if $g'(x) - g'(\bar{x}) \in N_C(\bar{x})$, or equivalently, $\bar{x} \in C$ and

$$\langle g'(x) - g'(\bar{x}), z - \bar{x} \rangle \leq 0 \quad \forall z \in C.$$

Proof. See page 70 of [14]. ∎

The following result, dealing with the Bregman projection onto hyperplanes, is a consequence of Proposition 1.6.2 and the definition of Bregman distance. For this purpose, we define $H = \{y \in B : \langle v, y - \tilde{y} \rangle = 0\}$, $H^+ = \{y \in B : \langle v, y - \tilde{y} \rangle \geq 0\}$, and $H^- = \{y \in B : \langle v, y - \tilde{y} \rangle \leq 0\}$, for each fix $\tilde{y} \in B$ and each fix $v \in B^*$.

Lemma 1.6.3. *Take $g \in \mathcal{F}$ totally convex. Assume that g satisfies* H1 *and $dom(g) = B$. Then for all $v \in B^* \setminus \{0\}$, $\tilde{y} \in B$, $x \in H^+$ and $\bar{x} \in H^-$, it holds that $D_g(\bar{x}, x) \geq D_g(\bar{x}, z) + D_g(z, x)$ where z is the unique minimizer of $D_g(\cdot, x)$ on H.*

Proof. See Lemma 1 of [26]. ∎

It is worthwhile mentioning that $D_g(x, y) = \frac{1}{2}\|x - y\|^2$ if $g(x) = \frac{1}{2}\|x\|^2$ and B is a Hilbert space. As a result of this fact, on Hilbert spaces, the Bregman projection coincides with the orthogonal projection on closed convex subsets.

1.7 Optimality Condition for Convex Minimization Problems

The first order optimality condition for convex minimization problem defined below, is an important tool in the study of equilibrium problem, both in reformulating equilibrium problems in terms of variational inequality problems (see Section 2.4), and for proving convergence of augmented Lagrangian methods (see Section 3.2). All the concepts introduced in this section can be found in [24] for the case of finite dimensional spaces, and in [33] for the case of infinite dimensional spaces.

Definition 1.7.1. *Let $h_0 : B \to \mathbb{R} \cup \{+\infty\}$ be a convex function. The convex minimization problem consists of finding an $\bar{x} \in C$ such that it solves*

$$\begin{array}{rl} \min & h_0(x) \\ s.t. & x \in C, \end{array} \tag{1.4}$$

where $C \subset B$ is a convex subset of B. Such an $\bar{x}$ is called a solution or a global minimizer of this problem.

Theorem 1.7.2. *$\bar{x} \in C$ is a solution of the problem* (1.4) *if and only if*

$$\langle \bar{u}, x - \bar{x} \rangle \geq 0 \qquad \forall x \in C,$$

for some $\bar{u} \in \partial h_0(\bar{x})$, or, equivalently, $\bar{x} \in C$ satisfies

$$0 \in \partial h_0(\bar{x}) + N_C(\bar{x}).$$

Proof. See Theorem 3.27 of [33]. ■

Assume that $h_0 : \mathbb{R}^n \to \mathbb{R}$ and that $C \subset \mathbb{R}^n$ is represented by finitely many convex inequalities

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

where $h_i : \mathbb{R}^n \to \mathbb{R}$ is convex $(1 \leq i \leq m)$. In this case, there are several conditions on the constraint functions h_i $(1 \leq i \leq m)$ under which $N_C(\bar{x})$ becomes a polyhedral cone. These conditions on the constraints are known as *constraint qualifications* (CQs), which guarantee existence of Lagrangian multipliers, reduce Theorem 1.7.2 to a computable result.

One of the CQs used in this thesis is Slater's condition, defined as follows.

Definition 1.7.3. *We say that the set C defined in* (1.5) *satisfies Slater's CQ if there exists $w \in \mathbb{R}^n$ such that $h_i(w) \leq 0$ for $i \in I$, and $h_i(w) < 0$ for $i \notin I$, where I is the* (*possibly empty*) *set of indices i such that the function h_i is affine.*

Using Definition 1.7.3, we can reformulate Theorem 1.7.2 as the classical Karush-Kuhn-Tucker Theorem, which deals with the nonsmooth case.

Theorem 1.7.4. *Assume that $\bar{x} \in C$ is a solution of the convex minimization problem* (1.4)*, with C defined as in* (1.5)*. If C satisfies Slater's CQ, then there exists a vector $\bar{\lambda} \in \mathbb{R}^m_+$ such that*

$$\begin{array}{c} 0 \in \partial h_0(\bar{x}) + \sum_{i=1}^m \bar{\lambda}_i \partial h_i(\bar{x}), \\ h(\bar{x}) \leq 0 \ \ (1 \leq i \leq m), \\ \bar{\lambda}_i h_i(\bar{x}) = 0 \ \ (1 \leq i \leq m). \end{array} \tag{1.6}$$

Conversely, if there exists a point $(\bar{x}, \bar{\lambda}) \in C \times \mathbb{R}^m_+$ satisfying statement (1.6)*, then $\bar{x}$ solves the convex minimization problem* (1.4)*.*

Proof. See Theorem 2.2.5 in Chapter VII of [24]. ∎

1.8 Proximal Point Methods for Finding Zeroes of Operators

This section is devoted to proximal point method for solving the problem of finding zeroes of point-to-set operators $T : B \to \mathcal{P}(B^*)$, defined as

$$\text{find} \ \ x \in B \ \ \text{such that} \ \ 0 \in T(x). \tag{1.7}$$

Let $T : H \to \mathcal{P}(H)$ be a point-to-set operator, where H is a Hilbert space. Rockafellar proposed the following iterative algorithm for finding zeroes of T in [59]. The algorithm generates a sequence $\{x^j\}_{j=0}^{\infty} \subset H$, starting from some $x^0 \in H$, where x^{j+1} is the unique zero of the operator T^j, defined as

$$T^j(x) = T(x) + \gamma_j(x - x^j), \tag{1.8}$$

with $\{\gamma_j\}_{j=0}^{\infty}$ being a bounded sequence of positive real numbers, called regularization sequence. It has been proved in [59] that for a maximal monotone T, the sequence $\{x^j\}_{j=0}^{\infty}$ is weakly convergent to a zero of T when T has zeroes, and is unbounded otherwise. Such weak convergence is global, i.e. the result just announced holds in fact for any $x^0 \in H$.

Now consider the case of proximal point methods for problem (1.7) in Banach spaces. In this setup, the formula for T^j given by (1.8) does not work any more, because $T(x)$ is a subset of B^*, while $\gamma_j(x - x^j)$ belongs to B. Thus, instead of (1.8), one works with

$$T^j(x) = T(x) + \gamma_j[g'(x) - g'(x^j)], \tag{1.9}$$

where $g \in \mathcal{F}$.

The following proposition shows that (1.9) is a regularization for the problem (1.7) in the sense that T^j attains a unique zero.

Proposition 1.8.1. *Let $g \in \mathcal{F}$ such that $dom(T) \subset int(dom(g))$. If g satisfies* H4, *then for each $\lambda > 0$ and each $z \in int(dom(g))$ there exists a unique solution of the problem $\lambda g'(z) \in T(x) + \lambda g'(x)$.*

Proof. See Corollary 3.1 of [12]. ■

Since the computation of each iterate requires solution of a regularized problem, it is important to establish convergence results assuming inexact solutions of the subproblems. This issue was already dealt with in [59], where it was assumed that the j-th subproblem was solved with an error bounded by a certain $\varepsilon_j > 0$, and the convergence results were preserved assuming that $\sum_{j=1}^{\infty} \varepsilon_j < \infty$. More precisely, in this inexact proximal point method in a Hilbert space H, having a current iterate $x^j \in H$, instead of finding a zero of operator (1.8), which is considered as the next iterate in the case of the exact proximal point method, one finds $x^{j+1} \in H$ and $v^{j+1} \in T(x^{j+1})$ such that $v^{j+1} + \gamma_j(x^{j+1} - x^j) - e^j = 0$, where e^j is the error at iteration j, satisfying

$$\left\|e^j\right\| \leq \varepsilon_j, \quad \sum_{j=1}^{\infty} \varepsilon_j < \infty.$$

This summability condition is undesirable. One drawback of this error criterion is that quite often there is no constructive way to enforce it. Another drawback of assumption $\sum_{j=1}^{\infty} \varepsilon_j < \infty$ is that it requires increasing accuracy along the iterative process. From the algorithmic standpoint, one would prefer to have computable error tolerance condition which is related to the progress of the algorithm at every given step when applied to the given problem. A better error criterion, introduced in [62], allows for constant relative errors and can be described, in its Hilbert space version, as follows. Instead of solving (1.8), which gives

$$\gamma_j(x^j - x^{j+1}) \in T(x^{j+1}),$$

one finds first an approximate zero of T_j, say $\hat{x}^j$, which can be taken as any point in the space satisfying

$$e^j + \gamma_j(x^j - \hat{x}^j) \in T(\hat{x}^j), \tag{1.10}$$

where the error term e^j satisfies

$$\left\|e^j\right\| \leq \sigma\gamma_j \max\left\{\left\|x^j - \hat{x}^j + \gamma_j^{-1}e^j\right\|, \left\|x^j - \hat{x}^j\right\|\right\}, \tag{1.11}$$

for some $\sigma \in [0, 1)$ (this σ can be seen as a constant relative error tolerance). Then the next iterate is obtained as the orthogonal projection of x^j onto the hyperplane $H_j = \{x \in H : \langle v^j, x - \hat{x}^j \rangle = 0\}$ with $v^j = \gamma_j (x^j - \hat{x}^j) + e^j$, more precisely

$$x^{j+1} = x^j - \frac{\langle v^j, x^j - \hat{x}^j \rangle}{\|v^j\|^2} v^j.$$

This inexact procedure, as well as a related one introduced in [63], were extended to Banach spaces in [26] for the same problem. Next we explain these two algorithms, to be called *Inexact Proximal Point+Bregman Projection Method* and *Inexact Proximal Point-Extragradient Method* in [26], which play very important roles in construction of our proximal point algorithms for equilibrium problems. Before doing this, we explain why σ in (1.11) is called *constant relative error tolerance*. For estimating the relative error in (1.11), we look at the ratios between e^j and $e^j + \gamma_j (x^j - \hat{x}^j)$, e^j and $\gamma_j (x^j - \hat{x}^j)$, i.e., the following two quantities, which can be seen as measures of the relative error:

$$\frac{\|e^j\|}{\|\gamma_j (x^j - \hat{x}^j) + e^j\|}, \quad \frac{\|e^j\|}{\gamma_j \|x^j - \hat{x}^j\|}.$$

In this sense, (1.11) is equivalent to saying that the bound for the relative error in solving the subproblem (1.10) can be fixed at σ, and need not tend to zero.

We describe next the algorithms presented in [26].

Algorithm I: Inexact Proximal Point+Bregman Projection Method for (1.7)

1. Choose an exogenous bounded sequence $\{\gamma_j\}_{j=0}^{\infty} \subset \mathbb{R}_{++}$ and a relative error tolerance $\sigma \in [0, 1]$. Initialize the algorithm with $x^0 \in B$.

2. Given x^j, find $\tilde{x}^j \in B$ such that

$$\gamma_j [g'(x^j) - g'(\tilde{x}^j)] - e^j \in T(\tilde{x}^j),$$

where e^j is any vector in B^* which satisfies

$$\left\|e^j\right\|_* \le \sigma \gamma_j \begin{cases} D_g(\tilde{x}^j, x^j) & \text{if } \|x^j - \tilde{x}^j\| < 1 \\ \nu_g(x^j, 1) & \text{if } \|x^j - \tilde{x}^j\| \ge 1, \end{cases}$$

with D_g, ν_g as in Definition 1.4.8(i)–(ii).

3. Let

$$v^j = \gamma_j [g'(x^j) - g'(\tilde{x}^j)] - e^j.$$

If $v^j = 0$ or $\tilde{x}^j = x^j$, then stop. Otherwise, take $H_j = \{x \in B : \langle v^j, x - \tilde{x}^j \rangle = 0\}$ and define

$$x^{j+1} = \operatorname*{argmin}_{x \in H_j} D_g(x, x^j).$$

Algorithm II: Inexact Proximal Point-Extragradient Method for (1.7)

1. Choose an exogenous bounded sequence $\{\gamma_j\}_{j=0}^{\infty} \subset \mathbb{R}_{++}$ and a relative error tolerance $\sigma \in [0, 1)$. Initialize the algorithm with $x^0 \in B$.

2. Given x^j, find $\tilde{x}^j \in B$ such that

$$\lambda_j[g'(x^j) - g'(\tilde{x}^j)] + e^j \in T(\tilde{x}^j),$$

where e^j is any vector in B^* which satisfies

$$D_g(\tilde{x}^j, (g')^{-1}[g'(\tilde{x}^j) - \gamma_j^{-1} e^j]) \leq \sigma D_g(\tilde{x}^j, x^j).$$

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

$$x^{j+1} = (g')^{-1}[g'(\tilde{x}^j) - \gamma_j^{-1} e^j].$$

The convergence results for Algorithm I and Algorithm II are given in the next two theorems and were proved in [26].

Theorem 1.8.2. *Assume that $T : B \to \mathcal{P}(B^*)$ is a maximal monotone operator. Take $g \in \mathcal{F}$ with dom$(g) = B$, satisfying* H1–H4, *$\{\gamma_j\}_{j=0}^{\infty} \subset (0, \bar{\gamma}]$, and $\sigma \in [0, 1]$. Let $\{x^j\}_{j=0}^{\infty}$ be the sequence generated by Algorithm* I. *If problem* (1.7) *has solutions, then*

i) $\{x^j\}_{j=0}^{\infty}$ has weak accumulation points and all of them are solutions of problem (1.7)*.*

ii) If g satisfies H6*, then the whole sequence $\{x^j\}_{j=0}^{\infty}$ is weakly convergent to a solution of problem* (1.7)*.*

Proof. See Theorem 2 of [26] ∎

Theorem 1.8.3. *Assume that $T : B \to \mathcal{P}(B^*)$ is a maximal monotone operator. Take $g \in \mathcal{F}$ with dom$(T) \subset$ int(dom(g)), satisfying* H1–H4, *$\{\gamma_j\}_{j=0}^{\infty} \subset (0, \bar{\gamma}]$, and $\sigma \in [0, 1)$. If problem* (1.7) *has solutions, then*

i) the sequence $\{x^j\}_{j=0}^{\infty}$ generated by Algorithm II *has weak accumulation points, all of which are solutions of problem* (1.7)*.*

ii) If g satisfies H6*, then the whole sequence $\{x^j\}_{j=0}^{\infty}$ is weakly convergent to a solution of problem* (1.7)*.*

Proof. See Theorem 3 of [26] ∎

1.9 Equilibrium Problems

In this section we formally define the equilibrium problem.

Definition 1.9.1. *Let B be a real Hausdorff topological vector space, and $K \subset B$ a nonempty closed and convex set. 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, denoted by EP(f, K)*, consists of finding $x^* \in K$ such that $f(x^*, y) \geq 0$ for all $y \in K$.*

Such an x^* satisfying the assumptions of Definition 1.9.1 is called a solution of EP(f, K). The set of solutions of EP(f, K) will be denoted by $S_E(f, K)$.

Through this thesis we will assume that B is a reflexive Banach space.

The study of the equilibrium problem demands some additional properties on EP(f, K). We present them next.

P4: There exists $\theta \geq 0$, and a totally convex and Fréchet differentiable function $g : B \to \mathbb{R}$ such that $f(x, y) + f(y, x) \leq \theta \langle g'(x) - g'(y), x - y \rangle$ for all $x, y \in K$.

P4$^\bullet$: $f(x, y) + f(y, x) \leq 0$ for all $x, y \in K$.

P4*: Whenever $f(x, y) \geq 0$ with $x, y \in K$, it holds that $f(y, x) \leq 0$.

P4': For all $x^1, \ldots, x^m \in K$ which are pairwise different, and all $\lambda_1, \ldots, \lambda_m$ which are strictly positive and such that $\sum_{i=1}^m \lambda_i = 1$, it holds that

$$\min_{1 \leq i \leq m} f\left(x^i, \sum_{j=1}^m \lambda_j x^j\right) < 0.$$

P4" : For all $x^1, \ldots, x^m \in K$ and all $\lambda_1, \ldots, \lambda_m \geq 0$ such that $\sum_{i=1}^m \lambda_i = 1$, it holds that

$$\sum_{i=1}^m \lambda_i f\left(x^i, \sum_{k=1}^m \lambda_k x^k\right) \leq 0.$$

P5 : For any sequence $\{x^j\}_{j=0}^\infty \subseteq K$ satisfying $\lim_{j \to \infty} \|x^j\| = +\infty$, there exists $u \in K$ and $j_0 \in \mathbb{N}$ such that $f(x^j, u) \leq 0$ for all $j \geq j_0$.

Definition 1.9.2. *Consider* EP(f, K)*. The function f is called θ-undermonotone (monotone, pseudomonotone, respectively) if it satisfies* P4 (P4$^\bullet$, P4**, respectively).*

In Section 1.10 we justify why the three properties P4, P4$^\bullet$ and P4*, given in Definition 1.9.2, are called θ-undermonotoncity, monotoncity, and pseudomonotonicity, respectively.

We comment now on some relations among P4, P4*, P4', and P4". For this purpose we consider the following illustrative example and the following proposition, taken from [32], which will be relevant in this work.

Example 1.9.3. *Let $K = [1/2, 1] \subset \mathbb{R}$ and define $f : K \times K \to \mathbb{R}$ as*

$$f(x, y) = x(x - y). \tag{1.12}$$

Proposition 1.9.4. *Under* P1–P3*,*

i) P4$^\bullet$ *implies any one among* P4*,* P4**, and* P4"*,*

ii) P4**,* P4' *and* P4" *are mutually independent,*

iii) none of the converse implications in (i) holds,

iv) P4' *does not imply* P4$^\bullet$*.*

Proof.

i) Elementary.

ii) See Section 2 of [27].

iii) Using Example 1.9.3, we first show that it satisfies P4, P4*, and it does not satisfies P4$^\bullet$. Note that $f(x, y) + f(y, x) = (x - y)^2$ so that f is not monotone, but it is immediate that it is 1-undermonotone. The fact that it satisfies P1–P3 is also immediate. For P4*, note that $f(x, y) \geq 0$ with $x, y \in K$ implies, since $x \geq 1/2$, that $x - y \geq 0$, in which case, using now that $y \geq 1/2$, one has $f(y, x) = y(y - x) \leq 0$, and so f is pseudomonotone. In order to verify that P4" does not imply P4$^\bullet$ we consider EP(f, K) where $K = \mathbb{R}^n$ and $f : K \times K \to \mathbb{R}$ defined as

$$f(x, y) = -\alpha \|x\|^2 + \beta \|y\|^2 + (\alpha - \beta)\langle x, y\rangle,$$

with $\beta > \alpha > 0$. We can easily observe 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 $f(x, y) + f(y, x) = (\beta - \alpha)\|x - y\|^2$. So f satisfies P4" and it is θ-undermonotone with $\theta = \beta - \alpha$, but it is not monotone, since $\beta - \alpha > 0$.

iv) It is an easy consequence of Example 1.9.3.

■

Now we make a few remarks on above properties. Under P1, concavity of $f(\cdot, y)$ for all y is sufficient for P4" to hold. We will prove that θ-undermonotonicity ensures regularity of our proposed proximal point method (See Proposition 2.1.1). In general, assumptions P4*, P4', P4", and P5 are introduced to guarantee existence of solutions for EP(f, K). In fact, property P5 has been identified in [27] as being necessary and sufficient for the existence of solutions of EP(f, K) under any monotonicity properties of f among P4*, P4' and P4". Indeed, we have the following result.

Theorem 1.9.5. *Assume that f satisfies* P1–P3*. Assume also that any one among* P4'*,* P4" *and* P4* *holds. Then* EP(f, K) *has solutions if and only if* P5 *holds.*

Proof. See Theorem 4.3 of [27]. ■

One can easily verify that the convex minimization problem (1.4) can be reformulated in terms of equilibrium problem EP(f, K) by defining $f(x, y) = h_0(y) - h_0(x)$ and $K = C$. More precisely, an x solves problem (1.4) if and only if x solves EP(f, K).

1.10 Variational Inequality Problems

We have already mentioned that the equilibrium problem includes as particular cases convex minimization problems, fixed point problems, complementarity problems, Nash equilibrium problems, variational inequality problems and vector minimization problems (see, e.g., [8], [30] and [49]). Nevertheless, the prototypical example of an equilibrium problem is the variational inequality problem. Since it plays an important role in the rest of this work (see Section 2.4), we describe it now in some detail. For this purpose we present the formal definition of a variational inequality problem which can be found in [11], for instance.

Definition 1.10.1. *Let B be a reflexive Banach space and B^* its topological dual space. Assume that $C \subset B$ is a nonempty closed and convex set. Given a point-to-set operator $T : C \to \mathcal{P}(B^*)$, the variational inequality problem denoted by* VIP(T, C) *consists of finding $x^* \in C$ such that for at least some $v^* \in T(x^*)$ it holds that*

$$\langle v^*, x - x^* \rangle \geq 0 \quad \forall x \in C.$$

Such an x^* satisfying the assumptions of Definition 1.10.1 is called a solution of VIP(T, C). The set of solutions of VIP(T, C) will be denoted by $S_V(T, C)$.

Remark 1.10.2. *We point out that* $x^* \in S_V(T,C)$ *if and only if* $0 \in T(x^*) + N_C(x^*)$, *where* N_C *is the normal cone of set* C *defined in* (1.1). *In other words, solving* $\mathrm{VIP}(T,C)$ *is equivalent to the problem of finding zeroes of point-to-set operator* $T+N_C$.

Now we investigate the reformulation of $\mathrm{VIP}(T,C)$ in terms of an equilibrium problem. First assume that T in $\mathrm{VIP}(T,C)$ is a point-to-point and continuous operator. We consider $\mathrm{EP}(f,C)$ with $f(x,y) = \langle T(x), y-x\rangle$. Then f satisfies P1–P3, and $\mathrm{EP}(f,C)$ is equivalent to the variational inequality problem $\mathrm{VIP}(T,C)$ in the sense that $S_V(T,C) = S_E(f,C)$. Now we assume that T in $\mathrm{VIP}(T,C)$ is a point-to-set operator such that $T(x)$ is compact in weak* topology for each $x \in C$. In this situation, we define

$$f(x,y) = \sup_{u \in T(x)} \langle u, y-x\rangle. \tag{1.13}$$

The fact that this f is well defined, it satisfies P1–P3 and that $S_E(f,C) = S_V(T,C)$ have been verified in Proposition 1.15 of [64].

It is easy to check that monotonicity of f defined in (1.13) (see Definition 1.9.2) is equivalent to monotonicity of T (see Definition 1.3.1(vi)). In addition, one can easily verify that if T is pseudomonotone (see Definition 1.3.1(viii)) and single-valued then f, as defined in 1.13, is pseudomonotone (see Definition 1.3.1(iv)), and the converse statement holds also when T is point-to-set. For this reason, a function f satisfying P4$^\bullet$ (respectively P4*) will be said to be monotone (respectively pseudomonotone). Along the same line, a function f satisfying P4 will be called θ-undermonotone.

1.11 Proximal Point Methods for Equilibrium Problems in Hilbert Spaces

Several approaches have already been considered for extending the proximal point method to the realm of equilibrium problems. In most cases, at each iteration some regularized problem is solved instead of $\mathrm{EP}(f,K)$. For instance we can mention the one proposed in [48] (see also [47]) for finite dimensional spaces, where the regularized problem can be rewritten as

$$\bar{f}(x,y) = f(\bar{x},y) + \gamma\langle x-\bar{x}, y-x\rangle.$$

Another proximal point method proposed in [20] for finite dimensional spaces, which its regularized problem consists of minimizing $\bar{f}(\bar{x},y)$ over $y \in K$ with

$$\bar{f}(x,y) = f(x,y) + \gamma D_g(y,x).$$

The basic idea of our proximal point methods, proposed in Chapter 2 for solving the equilibrium problem in Banach space, comes from the one presented in [32] for Hilbert space, where the regularized equilibrium problem is $\mathrm{EP}(\tilde{f}, K)$ with $\tilde{f} : K \times K \to \mathbb{R}$ defined as

$$\tilde{f}(x, y) = f(x, y) + \gamma\langle x - \bar{x}, y - x\rangle, \tag{1.14}$$

where γ is a positive real number. At iteration j, given $x^j \in K$, one solves the problem $\mathrm{EP}(\bar{f}_j, K)$, where the regularized function $\bar{f}_j$ is defined as

$$\bar{f}_j(x, y) = f(x, y) + \gamma_j\langle x - x^j, y - x\rangle, \tag{1.15}$$

for some $\gamma_i > 0$. It is established in [32] that $\bar{f}_j$ satisfies the required properties, and that $\mathrm{EP}(\bar{f}_j, K)$ has a unique solution, which is the next iterate x^{j+1}. The convergence properties of sequence $\{x^j\}_{j=0}^{\infty}$ generated by the method are given in the next theorem. Before this theorem we need the following two definitions taken from [32] and [36], respectively.

Definition 1.11.1. $\{z^j\}_{j=0}^{\infty} \subset K$ *is an asymptotically solving sequence for* $\mathrm{EP}(f, K)$ *if* $\liminf_{j\to\infty} f(z^j, y) \geq 0$ *for all* $y \in K$.

Definition 1.11.2. *Consider* $\mathrm{EP}(f, K)$. *The dual problem of* $\mathrm{EP}(f, K)$ *consists of finding* $y^* \in K$ *such that* $f(x, y^*) \leq 0$ *for all* $x \in K$. *The solution set of this dual problem will be denoted as* $S^d(f, K)$.

The next proposition guarantees that $\mathrm{EP}(\bar{f}_j, K)$, with $\bar{f}_j$ defined in (1.15), is a regularization of $\mathrm{EP}(f, K)$.

Proposition 1.11.3. *Take* f *satisfying* P1–P4. *Assume that* $\gamma > \theta$. *Then* $\mathrm{EP}(\tilde{f}, K)$ *has a unique solution, with* $\tilde{f}$ *defined in* (1.14).

Proof. See Proposition 3 of [32]. ■

Theorem 1.11.4. *Consider* $\mathrm{EP}(f, K)$, *where* f *satisfies* P1–P3. *Assume also that* $\{\gamma_j\}_{j=0}^{\infty} \subset (\theta, \bar{\gamma}]$ *where* $\bar{\gamma}$ *is a positive real number. Given* $x^j \in K$, *define* x^{j+1} *as solution of* $\mathrm{EP}(\bar{f}_j, K)$ *with* $\bar{f}_j$ *as in* (1.15). *In this situation, the following statements hold.*

i) If f *satisfies* P4 *then the sequence* $\{x^j\}_{j=0}^{\infty}$ *is well defined.*

ii) If $S^d(f, K) \neq \emptyset$ *then the sequence* $\{x^j\}_{j=0}^{\infty}$ *is bounded and*

$$\lim_{j\to\infty} \left\|x^{j+1} - x^j\right\| = 0.$$

iii) Under the assumptions of items (i) and (ii) the sequence $\{x^j\}_{j=0}^{\infty}$ *is an asymptotically solving sequence for* $\mathrm{EP}(f, K)$.

iv) If additionally $f(\cdot, y)$ is weakly upper semicontinuous for all $y \in K$, then all weak cluster points of $\{x^j\}_{j=0}^{\infty}$ solve $\mathrm{EP}(f, K)$.

v) If additionally $S_E(f, K) = S^d(f, K)$ then the sequence $\{x^j\}_{j=0}^{\infty}$ is weakly convergent to some solution $\hat{x}$ of $\mathrm{EP}(f, K)$.

Proof. See Theorem 1 of [32]. ■

We mention that [36], [43], [44], and [45] consider the regularized bifunction $\tilde{f}$ given by (1.14) and develop basically the same iterative procedure as (1.15).

We will see in Chapter 2 that when we move to Banach spaces, instead of $\bar{f}_j$ as defined in (1.15), we will solve at iteration j the equilibrium problem $\mathrm{EP}(f_j, K)$, with f_j given by

$$f_j(x, y) = f(x, y) + \gamma_j \langle g'(x) - g'(x^j), y - x \rangle,$$

where $g : B \to \mathbb{R}$ is an auxiliary function satisfying appropriate assumptions (see H1–H6). When B is a Hilbert space and $g(x) = \frac{1}{2}\|x\|^2$, we get $f_j = \bar{f}_j$. In order to generate an inexact algorithm, we will consider perturbations of f_j. Indeed, in our algorithm f_j will be replaced by a perturbed f_j^e,

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

where $e^j \in B^*$, the error vector at the j-th iteration, will be subject to bounds similar to (1.11), to be presented in Section 2.2.

1.12 Augmented Lagrangian Methods for Convex Minimization Problems

We describe the augmented Lagrangian methods for convex minimization problems, which is the departure point for the study of augmented Lagrangian methods for the equilibrium problem.

Consider the minimization problem

$$\begin{aligned} &\min h_0(x) \\ &\text{s.t. } h_i(x) \le 0 \ (1 \le i \le m), \end{aligned} \tag{1.16}$$

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

The Lagrangian for (1.16) is the function $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ given by

$$L(x, \lambda) = h_0(x) + \sum_{i=1}^{m} \lambda_i h_i(x), \tag{1.17}$$

and the dual problem associated to (1.16) is the convex minimization problem given by

$$\begin{aligned} &\min -\psi(y) \\ &\text{s.t. } y \in \mathbb{R}^m_+, \end{aligned} \tag{1.18}$$

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

$$\psi(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda). \tag{1.19}$$

The augmented Lagrangian associated to the problem given by (1.16) is the function $\bar{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_{++} \to \mathbb{R}$ defined as

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

where $\mathbb{R}_{++}$ is the set of positive real numbers. The augmented Lagrangian method requires an exogenous sequence of regularization parameters $\{\gamma_j\}_{j=0}^{\infty} \subset \mathbb{R}_{++}$. The method starts with some $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^m_+$, and given $(x^j, \lambda^j) \in \mathbb{R}^n \times \mathbb{R}^m_+$, 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 (1.16) and the dual problem (1.18) have solutions, and that the sequence $\{x^j\}_{j=0}^{\infty}$ is well defined, in the sense that all the unconstrained minimization subproblems are solvable, it has been proved that the sequence $\{\lambda^j\}_{j=0}^{\infty}$ converges to a solution of the dual problem (1.18) and that the cluster points of the sequence $\{x^j\}_{j=0}^{\infty}$ (if any) solve the primal problem (1.16) (see, e.g., [25] or [58]).

Another augmented Lagrangian method for the same problem, with better convergence properties, is the proximal augmented Lagrangian method (see [58]; this method is called "doubly augmented Lagrangian" in [25]). 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

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

The method uses an exogenous sequence $\{\gamma_j\}_{j=0}^{\infty} \subset \mathbb{R}_{++}$ as before, and it starts with $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^m_+$. Given $(x^j, \lambda^j) \in \mathbb{R}^n \times \mathbb{R}^m_+$, 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 $\{x^j\}_{j=0}^{\infty}$, $\{\lambda^j\}_{j=0}^{\infty}$ converge to a primal and a dual solution respectively (see, e.g., [25] or [58]).

The connection between the augmented Lagrangian methods for convex optimization and the proximal point method for finding zeroes of operators, introduced in Section 1.8, can be described as follows. Let $\{x^j\}_{j=0}^{\infty}$, $\{\lambda^j\}_{j=0}^{\infty}$ 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 (1.19). The sequence $\{z^j\}_{j=0}^{\infty}$ generated by the proximal point for finding zeroes of T coincides with $\{\lambda^j\}_{j=0}^{\infty}$, assuming that $\lambda^0 = z^0$, and that the same sequence $\{\gamma_j\}_{j=0}^{\infty}$ is used for both methods (see, e.g.,[25] or [58]). Hence, the convergence of $\{\lambda^j\}_{j=0}^{\infty}$ to some solution of the dual problem (1.18) follows from the convergence of the sequence $\{z^j\}_{j=0}^{\infty}$, generated by the proximal point method, to a zero of T. Thus, we observe that the main tool used in [58] for establishing the above mentioned convergence results is the proximal point algorithm.

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 (1.17) and $N_{\mathbb{R}^m_+}$ is the normal cone of the nonnegative orthant of $\mathbb{R}^m$ defined as $\mathbb{R}^m_+ = \{(x_1, \ldots, x_m) \in \mathbb{R}^m \mid x_i \geq 0 \ \ (1 \leq i \leq m)\}$. In this situation, the sequence $\{z^j\}_{j=0}^{\infty}$ generated by the proximal point method coincides with the sequence $\{(x^j, \lambda^j)\}_{j=0}^{\infty}$ generated by the proximal augmented Lagrangian method, assuming again that $z^0 = (x^0, \lambda^0)$, and that the same regularization sequence $\{\gamma_j\}_{j=0}^{\infty}$ is used in both algorithms (see, e.g., [25] or [58]).

We mention that the convergence analysis of the augmented Lagrangian methods for equilibrium problems, to be discussed in Chapter 3, invokes the convergence theorems of proximal point methods for equilibrium problems to be presented in Chapter 2.

Chapter 2

Inexact Proximal Point Methods for Equilibrium Problems in Banach Spaces

This chapter is devoted to an exact and two inexact proximal point methods for equilibrium problems. The results of this chapter can be found in [29].

We start this chapter by showing that a regularization for equilibrium problems exists when proper conditions are met. We present Algorithm IPPBP standing for *Inexact Proximal Point+Bregman Projection Method*, and Algorithm IPPE standing for *Inexact Proximal Point-Extragradient Method*, for solving $\text{EP}(f, K)$, and we establish their convergence properties. Finally, we introduce a reformulation of $\text{EP}(f, K)$ in terms of a certain variational inequality problem, or, equivalently, a problem of finding zeroes of point-to-set operators, which enables us to get rid of some assumptions in our convergence analysis.

In the sequel we will frequently use the properties H1–H6, introduced in Section 1.5, as well as the properties P1–P5 and variants of P4, introduced in Section 1.9, which we list here for easier reference.

H1: The level sets of $D_g(x, \cdot)$ are bounded for all $x \in dom(g)$.

H2: $\inf_{x \in C} \nu_g(x, t) > 0$ for all bounded set $C \subset int(dom(g))$ and all $t > 0$.

H3: g' is uniformly continuous on bounded subsets of $int(dom(g))$.

H4: g' is onto, i.e., for all $y \in B^*$, there exists $x \in B$ such that $g'(x) = y$.

H5: $\lim_{\|x\| \to \infty}[g(x) - \rho \|x - z\|] = +\infty$ for all fixed $z \in K$ and $\rho \geq 0$.

H6: If $\{y^j\}_{j=0}^{\infty}$ and $\{z^j\}_{j=0}^{\infty}$ are sequences in K which converge weakly to y and z, respectively and $y \neq z$, then

$$\liminf_{j\to\infty} \left|\langle g'(y^j) - g'(z^j), y - z\rangle\right| > 0.$$

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 $y \in K$.

P4: There exists $\theta \geq 0$ such that $f(x,y) + f(y,x) \leq \theta\langle g'(x) - g'(y), x - y\rangle$ for all $x, y \in K$, where $g : B \to \mathbb{R}$ is a Fréchet differentiable function satisfying some regularity properties among H1–H6.

P4$^\bullet$: $f(x,y) + f(y,x) \leq 0$ for all $x, y \in K$.

P4*: Whenever $f(x,y) \geq 0$ with $x, y \in K$, it holds that $f(y,x) \leq 0$.

P4': For all $x^1, \dots, x^m \in K$ which are pairwise different, and all $\lambda_1, \dots, \lambda_m$ which are strictly positive and such that $\sum_{i=1}^m \lambda_i = 1$, it holds that

$$\min_{1\leq i\leq m} f\left(x^i, \sum_{j=1}^m \lambda_j x^j\right) < 0. \tag{2.1}$$

P4" : For all $x^1, \dots, x^m \in K$ and all $\lambda_1, \dots, \lambda_m \geq 0$ such that $\sum_{i=1}^m \lambda_i = 1$, it holds that

$$\sum_{i=1}^m \lambda_i f\left(x^i, \sum_{k=1}^m \lambda_k x^k\right) \leq 0. \tag{2.2}$$

P5 : For any sequence $\{x^j\}_{j=0}^\infty \subseteq K$ satisfying $\lim_{j\to\infty} \|x^j\| = +\infty$, there exists $u \in K$ and $j_0 \in \mathbb{N}$ such that $f(x^j, u) \leq 0$ for all $j \geq j_0$.

In this chapter, unless otherwise stated, we assume that B is a reflexive Banach space, that K is a closed convex subset of B with nonempty relative interior, and that g belongs to $\mathcal{F}$, as defined in Section 1.5, with $dom(g) = B$.

2.1 Regularization of Equilibrium Problems

In this section we associate to a given equilibrium problem EP(f,K), another equilibrium problem which under some reasonable assumptions can be considered as a regularization of EP(f,K) in the sense that this new equilibrium problem attains a unique solution. For constructing such a regularization for EP(f,K), we take $\bar{x} \in B, e \in B^*, \lambda \in \mathbb{R}_{++}$, and we define EP$(\tilde{f},K)$ with

$$\tilde{f}(x,y) = f(x,y) + \gamma\langle g'(x) - g'(\bar{x}), y - x\rangle - \langle e, y - x\rangle, \tag{2.3}$$

where $g : B \to \mathbb{R}$ is a totally convex and Fréchet differentiable function. If $\mathrm{EP}(\tilde{f}, K)$ is the regularized equilibrium problem associated to $\mathrm{EP}(f, K)$, we say that γ is a regularization parameter and $\tilde{f}$ is a regularized bifunction for $\mathrm{EP}(f, K)$. We also say that $\mathrm{EP}(\tilde{f}, K)$ is an exact regularized equilibrium problem for $\mathrm{EP}(f, K)$ if $e = 0$, and otherwise we say that it is an inexact regularized equilibrium problem for $\mathrm{EP}(f, K)$. In this situation, $\tilde{f}$ is called an exact regularized bifunction (respectively inexact regularized bifunction) whenever $e = 0$ (respectively $e \neq 0$).

The following proposition guarantees that, under adequate monotonicity assumptions, the function $\tilde{f}$ introduced in (2.3) is a regularization of f.

Proposition 2.1.1. *Consider* $\mathrm{EP}(f, K)$ *satisfying* P1–P4. *Assume that* K *has nonempty relative interior. Fix* $(\bar{x}, e) \in B \times B^*$ *and* $\gamma > \theta$, *where* θ *is the undermonotonicity constant in* P4. *Assume that* $g : B \to \mathbb{R}$ *satisfies* H1–H2 *and* H5. *If* $\tilde{f} : K \times K \to \mathbb{R}$ *is defined as* (2.3), *then* $\mathrm{EP}(\tilde{f}, K)$ *has a unique solution.*

Proof. We first prove existence of solutions. In view of Theorem 1.9.5, it suffices to show that $\tilde{f}$ satisfies P1–P3, P4* and P5. It follows from (2.3) that $\tilde{f}$ inherits P1–P3 from f. Now we claim that $\tilde{f}$ satisfies P4*. Note that

$$\begin{aligned}\tilde{f}(x,y) + \tilde{f}(y,x) &= f(x,y) + f(y,x) - \gamma\langle g'(x) - g'(y), x - y\rangle \\ &\leq (\theta - \gamma)\langle g'(x) - g'(y), x - y\rangle \leq 0,\end{aligned} \tag{2.4}$$

using the definition of $\tilde{f}$ in the equality and the fact that f satisfies P4 in the first inequality. The second inequality follows from the facts that g is strictly convex, as a consequence of H2, and $\gamma > \theta$. It follows from (2.4) that P4* holds for $\tilde{f}$. In order to apply Theorem 1.9.5, it suffices to establish that $\tilde{f}$ satisfies P5. So, we take a sequence $\{x^j\}_{j=0}^{\infty} \subset K$ such that $\lim_{j\to\infty} \|x^j\| = +\infty$. We claim that P5 holds with u equal to the Bregman projection of $\bar{x}$ onto K, as defined in (1.3). Note that

$$\begin{aligned}\tilde{f}(x^j, u) &= f(x^j, u) - \gamma\langle g'(x^j) - g'(\bar{x}), x^j - u\rangle - \langle e, u - x^j\rangle \\ &= f(x^j, u) - \gamma\langle g'(u) - g'(\bar{x}), x^j - u\rangle - \gamma\langle g'(x^j) - g'(u), x^j - u\rangle - \langle e, u - x^j\rangle \\ &\leq -f(u, x^j) - \gamma\langle g'(u) - g'(\bar{x}), x^j - u\rangle + (\theta - \gamma)\langle g'(x^j) - g'(u), x^j - u\rangle \\ &+ \|e\|_* \left\|u - x^j\right\| \\ &\leq -f(u, x^j) + (\theta - \gamma)\langle g'(x^j) - g'(u), x^j - u\rangle + \|e\|_* \left\|u - x^j\right\|,\end{aligned} \tag{2.5}$$

using P4 and Cauchy-Schwartz inequality in the first inequality and Proposition 1.6.2(ii) in the second one, taking into account that $\gamma > \theta$ and $x^j \in K$. We introduce now some notation for the marginals of f. For $x \in K$, define $F_x : K \to \mathbb{R}$ as

$$F_x(y) = f(x, y). \tag{2.6}$$

Take $\hat{x} \in ri(K)$ (see Definition 1.1.16), which is nonempty by our assumption. Since F_u is convex by P2, its subdifferential at $\hat{x}$, denoted as $\partial F_u(\hat{x})$, is nonempty (see Proposition 1.4.2(ii)–(iii)). Take $\hat{v} \in \partial F_u(\hat{x})$. By the definition of subdifferential, we have

$$\langle \hat{v}, x^j - \hat{x} \rangle \leq F_u(x^j) - F_u(\hat{x}) = f(u, x^j) - f(u, \hat{x}). \tag{2.7}$$

In view of (2.7),

$$\begin{aligned} -f(u, x^j) &\leq \langle \hat{v}, \hat{x} - x^j \rangle - f(u, \hat{x}) \\ &\leq \|\hat{v}\|_* \left\|\hat{x} - x^j\right\| - f(u, \hat{x}) \\ &\leq \|\hat{v}\|_* \left\|u - x^j\right\| + \|\hat{v}\|_* \|\hat{x} - u\| - f(u, \hat{x}), \end{aligned} \tag{2.8}$$

using Cauchy-Schwartz inequality and the triangle inequality in the second and third inequalities, respectively.

We now find an upper bound for the second term in the last expression of (2.5).

$$\langle g'(x^j) - g'(u), x^j - u \rangle = D_g(x^j, u) + D_g(u, x^j) \geq D_g(x^j, u) =$$
$$g(x^j) - g(u) - \langle g'(u), x^j - u \rangle \geq g(x^j) - g(u) - \|g'(u)\|_* \left\|x^j - u\right\|, \tag{2.9}$$

using the definition of Bregman distance in the first equality, nonnegativity of D_g, which follows from H2, in the first inequality, and Cauchy-Schwartz inequality in the second inequality.

Now we utilize the inequalities obtained in (2.8) and (2.9) to get an upper bound for the rightmost expression in (2.5).

$$\tilde{f}(x^j, u) \leq (\theta - \gamma) \left[g(x^j) - \frac{\|\hat{v}\|_* + \|e\|_* + (\gamma - \theta) \|g'(u)\|_*}{\gamma - \theta} \left\|x^j - u\right\| \right]$$
$$+ (\gamma - \theta) g(u) + \|\hat{v}\|_* \|\hat{x} - u\| - f(u, \hat{x}). \tag{2.10}$$

Let $z = u$ and $\rho = (\gamma - \theta)^{-1}(\|\hat{v}\|_* + \|e\|_*) + \|g'(u)\|_*$. Since $\theta - \gamma < 0$, $\lim_{j\to\infty} \|x^j\| = +\infty$ and g satisfies H5, it follows from (2.10) that $\lim_{j\to\infty} \tilde{f}(x^j, u) = -\infty$. Therefore, $\tilde{f}(x^j, u) \leq 0$ for large enough j. We have shown that $\tilde{f}$ satisfies P5. Hence, $\tilde{f}$ satisfies all the assumptions of Theorem 1.9.5, which implies that EP($\tilde{f}, K$) has solutions.

Now we establish uniqueness of the solution. Assume that $\tilde{x}$ and $\tilde{x}'$ solve EP($\tilde{f}, K$). Using the definition of $\tilde{f}$, we have that

$$0 \leq \tilde{f}(\tilde{x}, \tilde{x}') = f(\tilde{x}, \tilde{x}') + \gamma \langle g'(\tilde{x}) - g'(\bar{x}), \tilde{x}' - \tilde{x} \rangle - \langle e, \tilde{x}' - \tilde{x} \rangle. \tag{2.11}$$

$$0 \leq \tilde{f}(\tilde{x}', \tilde{x}) = f(\tilde{x}', \tilde{x}) + \gamma \langle g'(\tilde{x}') - g'(\bar{x}), \tilde{x} - \tilde{x}' \rangle - \langle e, \tilde{x} - \tilde{x}' \rangle. \tag{2.12}$$

Adding (2.11) and (2.12), and using P4, we get

$$0 \leq f(\tilde{x}, \tilde{x}') + f(\tilde{x}', \tilde{x}) - \gamma \langle g'(\tilde{x}) - g'(\tilde{x}'), \tilde{x} - \tilde{x}' \rangle \leq (\theta - \gamma) \langle g'(\tilde{x}) - g'(\tilde{x}'), \tilde{x} - \tilde{x}' \rangle \leq 0.$$

Since $\theta - \gamma < 0$, we obtain $\langle g'(\tilde{x}) - g'(\tilde{x}'), \tilde{x} - \tilde{x}' \rangle = 0$, implying that $\tilde{x} = \tilde{x}'$, because g is strictly convex as a consequence of H2. ■

2.2 Inexact Versions of the Proximal Point Methods for EP(f, K)

We start by presenting an exact proximal point method for equilibrium problem in Banach spaces. Consider EP(f, K), where $K \subset B$ is closed and convex and $f : K \times K \to \mathbb{R}$ satisfies P1–P4 and any one among P4', P4" and P4*. The algorithm requires two exogenous data: an auxiliary function $g : B \to \mathbb{R}$ satisfying the regularity properties H1–H6 and a sequence of regularization parameters $\{\gamma_j\}_{j=0}^{\infty} \subset (\theta, \bar{\gamma}]$, where θ is the constant of undermonotonicity appeared in P4 and $\bar{\gamma} > \theta$. In fact, assumptions H1, H2, and H5 are essential to provide a regularization for equilibrium problems (See Proposition 2.1.1). Assumptions H1–H5 are used to show that all cluster points of the sequence generated by our methods solve the equilibrium problem, and assumption H6 implies the uniqueness of such cluster points (See Theorem 2.3.5 and Theorem 2.3.8)

The algorithm generates a sequence $\{x^j\}_{j=0}^{\infty} \subset K$ as follows: x^0 is an arbitrary point in K, and, given x^j, x^{j+1} is the solution of EP(f_j, K) with

$$f_j(x, y) = f(x, y) + \gamma_j \langle g'(x) - g'(x^j), y - x \rangle.$$

Existence and uniqueness of x^{j+1} are consequences of Proposition 2.1.1, with $\gamma = \gamma_j$, $\bar{x} = x^j$ and $e = 0$. When B is a Hilbert space and $g(x) = \frac{1}{2} \|x\|^2$ this method reduces to the one analyzed in [32]. The convergence properties of the method can be summarized as follow.

Proposition 2.2.1. *Consider* EP(f, K) *such that* K *has nonempty relative interior. Assume that* f *satisfies* P1–P4 *and additionally any one among* P4', P4" *and* P4*. *Suppose that* $g : B \to \mathbb{R}$ *satisfies* H1–H5 *defined in the beginning of this chapter and an exogenous sequence* $\{\gamma_j\}_{j=0}^{\infty} \subset (\theta, \bar{\gamma}]$, *where* θ *is the undermonotonicity constant in* P4. *Let* $\{x^j\}_{j=0}^{\infty}$ *be the sequence generated by the above exact proximal point method. In this situation, the following statements hold.*

i) If EP(f, K) *has solutions, then the sequence* $\{x^j\}_{j=0}^{\infty}$ *is bounded and asymptotically solves the problem* EP(f, K). *In other words,* $\liminf_{j \to \infty} f(x^j, y) \geq 0$ *for all* $y \in K$.

ii) If additionally $f(\cdot, y)$ *is weakly upper semicontinuous for all* $y \in K$, *then all weak cluster points of* $\{x^j\}_{j=0}^{\infty}$ *are solutions of* EP(f, K).

iii) If additionally either g satisfies H6 *or* EP(f, K) *has a unique solution, then* $\{x^j\}_{j=0}^{\infty}$ *is weakly convergent to a solution of* EP(f, K).

We will not prove this proposition, because the exact method is indeed a particular case of the two inexact methods which we present next, together with their convergence analysis.

Both inexact algorithms fit in the following scheme: given x^j, an auxiliary point $\tilde{x}^j$ is computed as the exact solution of a perturbed problem EP(f_j^e, K), where

$$f_j^e(x, y) = f_j(x, y) - \langle e^j, y - x\rangle,$$

and $e^j \in B^*$ is an arbitrary error vector whose norm is "small", i.e. bounded by an appropriate function of the data available at iteration j, namely γ_j, x^j and $\tilde{x}^j$. The error vector e^j is then used for building a hyperplane H_j which separates x^j from set $S_E(f, K)$. Then x^{j+1} is obtained by either projecting x^j onto H_j with respect to the Bregman distance D_g (in the case of Algorithm IPPBP), or by taking a step from x^j in the direction of H_j with respect to the metric induced by D_g (in the case of Algorithm IPPE). We mention parenthetically that seeing inexact solutions of a given problem as exact solutions of a perturbed problem is a frequent tool in Numerical Analysis, like in the error analysis of methods for solving systems of linear equations (see, e.g., [51]).

The function g and the sequence $\{\gamma_j\}_{j=0}^{\infty}$ appearing in the following two algorithms satisfy the same assumptions as those used in the exact algorithm given above. Additionally, we will need a relative error bound $\sigma \in [0, 1]$ for Algorithm IPPBP. In the case of Algorithm IPPE, we assume that $\sigma \in [0, 1)$.

Algorithm IPPBP: Inexact Proximal Point+Bregman Projection Method for EP(f, K)

1. Take an exogenous bounded sequence $\{\gamma_j\}_{j=0}^{\infty} \subset \mathbb{R}_{++}$ and a relative error tolerance $\sigma \in [0, 1]$. Initialize the algorithm with $x^0 \in B$.

2. Given $x^j \in int(dom(g)) = B$, find a pair $(\tilde{x}^j, e^j) \in B \times B^*$ such that $\tilde{x}^j$ solves EP$(\tilde{f}_j^e, K)$ with

$$f_j^e(x, y) = f(x, y) + \gamma_j\langle g'(x) - g'(x^j), y - x\rangle - \langle e^j, y - x\rangle, \tag{2.13}$$

i.e.

$$f_j^e(\tilde{x}^j, y) \geq 0 \quad \forall y \in K, \tag{2.14}$$

and e^j satisfies

$$\left\|e^j\right\|_* \leq \sigma\gamma_j \begin{cases} D_g(\tilde{x}^j, x^j) & \text{if } \|x^j - \tilde{x}^j\| < 1 \\ \nu_g(x^j, 1) & \text{if } \|x^j - \tilde{x}^j\| \geq 1, \end{cases} \tag{2.15}$$

with D_g, ν_g as in Definition 1.4.8(i) and Definition 1.4.8(ii), respectively.

3. Let

$$v^j = \gamma_j[g'(x^j) - g'(\tilde{x}^j)] + e^j. \tag{2.16}$$

If $v^j = 0$ or $\tilde{x}^j = x^j$, then stop. Otherwise, take $H_j = \{x \in B : \langle v^j, x - \tilde{x}^j\rangle = 0\}$ and define

$$x^{j+1} = \operatorname*{argmin}_{x \in H_j} D_g(x, x^j). \tag{2.17}$$

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

1. Take an exogenous bounded sequence $\{\gamma_j\}_{j=0}^{\infty} \subset \mathbb{R}_{++}$ and a relative error tolerance $\sigma \in [0, 1)$. Initialize the algorithm with $x^0 \in B$.

2. Given x^j, find a pair $(\tilde{x}^j, e^j) \in B \times B^*$ such that $\tilde{x}^j$ solves EP(f_j^e, K) with

$$f_j^e(x, y) = f(x, y) + \gamma_j\langle g'(x) - g'(x^j), y - x\rangle - \langle e^j, y - x\rangle, \tag{2.18}$$

i.e.

$$f_j^e(\tilde{x}^j, y) \geq 0 \quad \forall y \in K, \tag{2.19}$$

and e^j satisfies

$$D_g(\tilde{x}^j, (g')^{-1}[g'(\tilde{x}^j) - \gamma_j^{-1}e^j]) \leq \sigma D_g(\tilde{x}^j, x^j). \tag{2.20}$$

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

$$x^{j+1} = (g')^{-1}[g'(\tilde{x}^j) - \gamma_j^{-1}e^j]. \tag{2.21}$$

It is worthwhile to observe that the inexact subproblems of both Algorithm IPPBP (i.e., (2.14)–(2.15)) and Algorithm IPPE (i.e., (2.19)–(2.20)) are solvable. Indeed, given any $e^j \in B^*$, Proposition 2.1.1 ensures the existence of a unique solution for EP(f_j^e, K), say $\tilde{x}^j$. If we take, in particular, $e^j = 0$, then the left hand sides of both (2.15) and (2.20) vanish, and so the inequalities in (2.15) and (2.20) are satisfied. We also note that for $e^j = 0$ both (2.17) and (2.21) give $x^{j+1} = \tilde{x}^j$, so that with this choice of e^j both algorithms reduce to the exact algorithm introduced at the beginning of this section.

2.3 Convergence Analysis

First we settle the issue of finite termination of these algorithms.

Proposition 2.3.1. *Suppose that Algorithm* IPPBP (*respectively Algorithm* IPPE) *stops after j steps. Then $\tilde{x}^j$ generated by Algorithm* IPPBP (*respectively Algorithm* IPPE) *is a solution of* EP(f, K).

Proof. Algorithm IPPBP stops at j-th iteration in two cases: if $v^j = 0$, in which case, by (2.13) and (2.16), $\tilde{x}^j \in S_E(f, K)$, or if $\tilde{x}^j = x^j$, in which case, by (2.15), $e^j = 0$, which in turn implies, by (2.16), $v^j = 0$ and we are back to the first case. Consequently, $\tilde{x}^j$ is a solution of EP(f, K). Finite termination in Algorithm IPPE occurs only if $\tilde{x}^j = x^j$, in which case $D_g(\tilde{x}^j, x^j) = 0$, and therefore, by (2.20), $e^j = 0$, which in turn implies, by (2.18)–(2.19), $f(\tilde{x}^j, y) \geq 0$ for all $y \in K$, so that $\tilde{x}^j \in S_E(f, K)$. ■

As we have already mentioned in Section 1.8, the convergence analysis of the proximal point method for finding zeroes of maximal monotone operators requires monotonicity of the operator, or some variant thereof. For equilibrium problems, we will work under assumptions weaker than monotonicity (i.e., P4$^\bullet$). We will assume θ-undermonotonicity (property P4) and additionally any one among the three variants of pseudomonotonicity introduced in Section 1.9, namely P4*, P4' and P4". The fact that our convergence analysis works under any of these three assumptions is a consequence of the following result.

Proposition 2.3.2. *Assume that any one among* P4', P4" *and* P4* *holds. If* P2 *is satisfied, then* $f(y, x^*) \leq 0$ *for all* $y \in K$ *and all* $x^* \in S_E(f, K)$.

Proof. First consider the case P4'. Fixing $y \in K$, we have that $f(x^*, \lambda x^* + (1-\lambda)y) \geq 0$ for all $0 \leq \lambda \leq 1$, since $x^* \in S_E(f, K)$. Now take $x^1 = y$, $x^2 = x^*$ and $m = 2$ in (2.1), obtaining $f(y, \lambda y + (1-\lambda)x^*) < 0$. Using lower semicontinuity of $f(y, \cdot)$ which holds by property P2 when $\lambda \to 0^+$, we obtain the desired result. When P4" holds, we take $y \in K$ and then put $x^1 = y$, $x^2 = x^*$, and $m = 2$ in (2.2), getting

$$\lambda f(y, \lambda y + (1-\lambda)x^*) + (1-\lambda) f(x^*, \lambda y + (1-\lambda)x^*) \leq 0. \tag{2.22}$$

Since the second term in the left hand side of (2.22) is nonnegative because $x^* \in S_E(f, K)$, we conclude that $f(y, \lambda y + (1-\lambda)x^*) \leq 0$ for all $\lambda \in (0, 1)$. We use again P2 with $\lambda \to 0^+$ for obtaining the result. The proof for the case P4* follows from the fact that $f(x^*, y) \geq 0$ for all $y \in K$ and all $x^* \in S_E(f, K)$. ■

From now on we treat separately Algorithm I and Algorithm II in the next two subsections.

2.3.1 Convergence Analysis of Algorithm IPPBP

We start with a result similar to Lemma 2 in [26].

Lemma 2.3.3. *Let* $\{x^j\}_{j=0}^{\infty}$, $\{\tilde{x}^j\}_{j=0}^{\infty}$, $\{\gamma_j\}_{j=0}^{\infty}$, *and* σ *be as in Algorithm* IPPBP *and assume that* g *satisfies* H2. *For all* j, *it holds that*

$$\left\|e^j\right\|_* \left\|x^j - \tilde{x}^j\right\| \leq \sigma\gamma_j D_g(\tilde{x}^j, x^j) \leq \gamma_j D_g(x^j, \tilde{x}^j). \tag{2.23}$$

Proof. We consider two cases. First, if $\|x^j - \tilde{x}^j\| < 1$ then we have that

$$\left\|e^j\right\|_* \left\|x^j - \tilde{x}^j\right\| \leq \left\|e^j\right\|_*,$$

so that the leftmost inequality in (2.23) follows trivially from (2.15). For the second case, namely $\|x^j - \tilde{x}^j\| \geq 1$, we will use the fact that $\nu_g(x, st) \geq s\nu_g(x,t)$ for all $s \geq 1, t \geq 0, x \in B$ (see Proposition 1.4.11). Then,

$$\sigma\gamma_j D_g(\tilde{x}^j, x^j) \geq \sigma\gamma_j \nu_g(x^j, \left\|x^j - \tilde{x}^j\right\|) \geq \sigma\gamma_j \left\|x^j - \tilde{x}^j\right\| \nu_g(x^j, 1) \geq \left\|x^j - \tilde{x}^j\right\| \left\|e^j\right\|_*,$$

using Definition 1.4.8(ii) in the first inequality, the above stated property of ν_g in the second one and (2.15) in the third. Thus, the first inequality of (2.23) is proved, and the second one holds because $\sigma \in [0,1]$. ■

We continue by establishing some basic properties of the sequences $\{x^j\}_{j=0}^{\infty}$, $\{\tilde{x}^j\}_{j=0}^{\infty}$ and $\{v^j\}_{j=0}^{\infty}$ generated by Algorithm IPPBP.

Proposition 2.3.4. *Consider* EP(f,K) *such that* K *has nonempty relative interior. Assume that* f *satisfies* P1–P4 *and also any one among* P4', P4" *and* P4*. *Take* $g : B \to \mathbb{R}$ *satisfying* H1–H5, $\sigma \in [0,1]$, *and an exogenous sequence* $\{\gamma_j\}_{j=0}^{\infty} \subset (\theta, \bar{\gamma}]$, *where* θ *is the undermonotonicity constant in* P4. *Let* $\{x^j\}_{j=0}^{\infty}$ *be the sequence generated by Algorithm* IPPBP. *If* EP(f,K) *has solutions, then the following statements are true.*

i) The sequence $\{D_g(x^*, x^j)\}_{j=0}^{\infty}$ *is nonincreasing and convergent for all* $x^* \in S_E(f,K)$.

ii) $\{x^j\}_{j=0}^{\infty}$ *is bounded.*

iii) $\{x^{j+1} - x^j\}_{j=0}^{\infty}$ *converges strongly to* 0.

iv) $\{\gamma_j^{-1} e^j\}_{j=0}^{\infty}$ *is bounded.*

v) $\{x^j - \tilde{x}^j\}_{j=0}^{\infty}$ *converges strongly to* 0.

vi) $\{v^j\}_{j=0}^{\infty}$ *converges strongly to* 0.

Proof. Take $x^* \in S_E(f,K)$. Let $H_j^- = \{x \in B : \langle v^j, x - \tilde{x}^j \rangle \leq 0\}$. Since $\tilde{x}^j \in S_E(f_j^e, K)$ with f_j^e given by (2.13), we have $f_j^e(\tilde{x}^j, y) \geq 0$ or equivalently $f(\tilde{x}^j, y) \geq \langle v^j, y - \tilde{x}^j \rangle$ for all $y \in K$, with v^j as in (2.16). In particular, $f(\tilde{x}^j, x^*) \geq \langle v^j, x^* - \tilde{x}^j \rangle$.

Using Proposition 2.3.2, we have that $0 \geq \langle v^j, x^* - \tilde{x}^j \rangle$, so that $x^* \in H_j^-$. By definition of v^j and D_g we have that

$$\begin{aligned} \langle v^j, x^j - \tilde{x}^j \rangle &= \gamma_j \langle g'(x^j) - g'(\tilde{x}^j), x^j - \tilde{x}^j \rangle + \langle e^j, x^j - \tilde{x}^j \rangle \\ &= \gamma_j [D_g(\tilde{x}^j, x^j) + D_g(x^j, \tilde{x}^j)] + \langle e^j, x^j - \tilde{x}^j \rangle \\ &\geq \gamma_j D_g(\tilde{x}^j, x^j) + \left[\gamma_j D_g(x^j, \tilde{x}^j) - \left\|e^j\right\|_* \left\|x^j - \tilde{x}^j\right\|\right], \end{aligned} \tag{2.24}$$

where the last inequality follows from the definition of the norm in B^*. Applying now Lemma 2.3.3, we get from (2.24) $\langle v^j, x^j - \tilde{x}^j \rangle \geq \gamma_j D_g(\tilde{x}^j, x^j) \geq 0$, so that $x^j \in H_j^+$, and in view of (2.17), we are able to apply Lemma 1.6.3 with $\tilde{y} = \tilde{x}^j \in B$, $x = x^j$, $v = v^j$, and $z = x^{j+1}$, obtaining

$$D_g(x^*, x^j) \geq D_g(x^*, x^{j+1}) + D_g(x^{j+1}, x^j) \geq D_g(x^*, x^{j+1}). \tag{2.25}$$

The result of (i) follows from (2.25), taking into account the nonnegativity of D_g. We also obtain from (2.25) that $\{x^j\}_{j=0}^{\infty} \subseteq \{y \in B : D_g(x^*, y) \leq D_g(x^*, x^0)\}$, and hence the sequence is bounded because of H1, establishing (ii).

We prove now (iii). Taking limits in (2.25) as $j \to \infty$ and using (i), we get

$$\lim_{j\to\infty} D_g(x^{j+1}, x^j) = 0. \tag{2.26}$$

Since g satisfies H2, we conclude from Proposition 1.5.2 that (iii) holds.

To prove (iv), we consider the set $E \subset B$ defined as

$$E = \{y \in B : \left\|x^j - y\right\| \leq 1 \text{ for some } j\}.$$

In view of (ii), E is bounded. By Definition 1.4.8(i) and Proposition 1.5.3, D_g is bounded on $E \times E$, say by $\zeta > 0$. From (2.15), taking into account that $\sigma \in [0, 1]$, we get

$$\left\|\gamma_j^{-1} e^j\right\|_* \leq \begin{cases} D_g(\tilde{x}^j, x^j) & \text{if } \|x^j - \tilde{x}^j\| < 1 \\ \nu_g(x^j, 1) & \text{if } \|x^j - \tilde{x}^j\| \geq 1. \end{cases} \tag{2.27}$$

If $\|x^j - \tilde{x}^j\| \leq 1$ then both x^j and $\tilde{x}^j$ belong to E and we get from (2.27) that $\left\|\gamma_j^{-1} e^j\right\|_* \leq \zeta$; otherwise, we take any $y \in B$ such that $\|y - x^j\| = 1$, so that $y \in E$, and get from Definition 1.4.8(ii) that $\nu_g(x^j, 1) \leq D_g(y, x^j) \leq \zeta$, so that, in view of (2.27), $\left\|\gamma_j^{-1} e^j\right\|_* \leq \zeta$ also in this case. We conclude that $\{\gamma_j^{-1} e^j\}_{j=0}^{\infty}$ is bounded.

We proceed to prove (v). Note that $0 \leq \left|\langle \gamma_j^{-1} e^j, x^{j+1} - x^j \rangle\right| \leq \left\|\gamma_j^{-1} e^j\right\|_* \|x^{j+1} - x^j\|$. Since $\{\gamma_j^{-1} e^j\}_{j=0}^{\infty}$ is bounded by (iv) and $\{x^{j+1} - x^j\}_{j=0}^{\infty}$ converges strongly to 0 by (iii), we get

$$\lim_{j\to\infty} \langle \gamma_j^{-1} e^j, x^{j+1} - x^j \rangle = 0. \tag{2.28}$$

Observe that

$$D_g(x^{j+1},x^j) - D_g(x^{j+1},\tilde{x}^j) - D_g(\tilde{x}^j,x^j) = \langle g'(x^j) - g'(\tilde{x}^j), \tilde{x}^j - x^{j+1}\rangle =$$

$$\gamma_j^{-1}\langle v^j, \tilde{x}^j - x^{j+1}\rangle - \langle \gamma_j^{-1}e^j, \tilde{x}^j - x^{j+1}\rangle = \langle \gamma_j^{-1}e^j, x^{j+1} - \tilde{x}^j\rangle, \quad (2.29)$$

where the first equality follows from Proposition 1.4.10(b), the second one from (2.16) and the last one from the fact that $x^{j+1} \in H_j = \{x \in B : \langle v^j, x - \tilde{x}^j\rangle = 0\}$. Thus, using (2.29),

$$D_g(x^{j+1},x^j) - D_g(x^{j+1},\tilde{x}^j) = \langle \gamma_j^{-1}e^j, x^{j+1} - x^j\rangle + D_g(\tilde{x}^j,x^j) + \langle \gamma_j^{-1}e^j, x^j - \tilde{x}^j\rangle$$

$$\geq \langle \gamma_j^{-1}e^j, x^{j+1} - x^j\rangle + \frac{1}{\gamma_j}\left[\gamma_j D_g(\tilde{x}^j,x^j) - \left\|e^j\right\|_* \left\|x^j - \tilde{x}^j\right\|\right] \geq \langle \gamma_j^{-1}e^j, x^{j+1} - x^j\rangle, \quad (2.30)$$

where the last inequality follows from the leftmost inequality in (2.23), since $\sigma \in [0,1]$. Taking limits as $j \to \infty$ in the leftmost and rightmost expressions of (2.30), we obtain, using (2.26) and (2.28),

$$\lim_{j\to\infty} D_g(x^{j+1},\tilde{x}^j) = 0, \quad (2.31)$$

and therefore, we conclude, using H2 and Proposition 1.5.2, that $\{x^{j+1} - \tilde{x}^j\}_{j=0}^{\infty}$ converges strongly to 0, so that in view of (iii), $\{x^j - \tilde{x}^j\}_{j=0}^{\infty}$ converges strongly to 0, establishing (v).

Now we prove (vi). By (v), there exists $j_0 \in \mathbb{N}$ such that, $\|x^j - \tilde{x}^j\| < 1$, for all $j \geq j_0$, and consequently for $j \geq j_0$ our error criterium (2.15) implies that

$$\left\|e^j\right\|_* \leq \sigma\gamma_j D_g(\tilde{x}^j,x^j). \quad (2.32)$$

Next, we take limit as j goes to ∞ in the leftmost and rightmost expressions of (2.29). The rightmost one converges to 0 by (iii)–(v), the first term in the leftmost expression converges to 0 by (2.26) and the second term converges to 0 by (2.31). It follows that $\lim_{j\to\infty} D_g(\tilde{x}^j,x^j) = 0$, and then, since $\gamma_j \leq \bar{\gamma}$, we get from (2.32) that e^j is strongly convergent to 0. From (v) and H3, we get that $\{g'(x^j) - g'(\tilde{x}^j)\}_{j=0}^{\infty}$ also converges strongly to 0. It follows then from (2.16) that v^j is strongly convergent to 0 as well. ■

Now we proceed to state and prove our convergence result for Algorithm IPPBP.

Theorem 2.3.5. *Consider* EP(f,K) *such that K has nonempty relative interior. Assume that f satisfies* P1–P4 *and additionally any one among* P4', P4" *and* P4*. *Take $g : B \to \mathbb{R}$ satisfying* H1–H5, $\sigma \in [0,1]$, *and an exogenous sequence* $\{\gamma_j\}_{j=0}^{\infty} \subset (\theta,\bar{\gamma}]$, *where θ is the undermonotonicity constant in* P4. *Let* $\{x^j\}_{j=0}^{\infty}$ *be the sequence generated by Algorithm* IPPBP. *If* EP(f,K) *has solutions, then the following statements hold.*

i) $\{\tilde{x}^j\}_{j=0}^{\infty}$ *is an asymptotically solving sequence for* EP(f,K).

ii) If $f(\cdot, y)$ is weakly upper semicontinuous for all $y \in K$, then all cluster points of $\{x^j\}_{j=0}^\infty$ solve EP(f, K)*.*

iii) If in addition either g satisfies H6 *or* EP(f, K) *has a unique solution, then the whole sequence $\{x^j\}_{j=0}^\infty$ is weakly convergent to some solution x^* of* EP(f, K)*.*

Proof. i) Fix $y \in K$. Since $\tilde{x}^j$ solves EP(f_j^e, K), by the definition of f_j^e and Cauchy-Schwartz inequality, we have that

$$\begin{aligned} 0 \leq f_j^e(\tilde{x}^j, y) &= f(\tilde{x}^j, y) + \langle \gamma_j[g'(\tilde{x}^j) - g'(x^j)] - e^j, y - \tilde{x}^j \rangle \\ &= f(\tilde{x}^j, y) + \langle -v^j, y - \tilde{x}^j \rangle \leq f(\tilde{x}^j, y) + \left\| v^j \right\|_* \left\| y - \tilde{x}^j \right\|. \end{aligned} \tag{2.33}$$

By Proposition 2.3.4(ii) and Proposition 2.3.4(v), we know that the sequence $\{\tilde{x}^j\}_{j=0}^\infty$ and therefore, the sequence $\{y - \tilde{x}^j\}_{j=0}^\infty$, are bounded for each fixed y. Consequently, taking limits in (2.33) as $j \to \infty$ and using Proposition 2.3.4(vi) we get

$$0 \leq \liminf_{j \to \infty} f(\tilde{x}^j, y), \tag{2.34}$$

for all $y \in K$.

ii) By Proposition 2.3.4(ii) and Proposition 2.3.4(v), $\{x^j\}_{j=0}^\infty$ has weak cluster points, all of which are also weak cluster points of $\{\tilde{x}^j\}_{j=0}^\infty$. These weak cluster points belong to K, which, being closed and convex, is weakly closed as stated in Proposition 1.1.14(iii). Let $\tilde{x}$ be a weak cluster points of $\{\tilde{x}^j\}_{j=0}^\infty$, say the weak limit of the subsequence $\{\tilde{x}^{\ell_j}\}_{\ell=0}^\infty$ of $\{\tilde{x}^j\}_{j=0}^\infty$. Since $f(\cdot, y)$ is weakly upper semicontinuous, we obtain from (2.34) that $0 \leq \limsup_{\ell \to \infty} f(\tilde{x}^{\ell_j}, y) \leq f(\tilde{x}, y)$ for all $y \in K$. As a result, $\tilde{x}$ belongs to $S_E(f, K)$.

iii) If EP(f, K) has a unique solution, then the result follows from (ii). Otherwise, assume that $\hat{x}$ is another weak cluster point of $\{x^j\}_{j=0}^\infty$, say the weak limit of the subsequence $\{x^{i_j}\}_{i=0}^\infty$ of $\{x^j\}_{j=0}^\infty$. By (ii), both $\tilde{x}$ and $\hat{x}$ solve EP(f, K). By Proposition 2.3.4(i), both $D_g(\hat{x}, x^j)$ and $D_g(\tilde{x}, x^j)$ converge, say to $\eta \geq 0$ and $\mu \geq 0$, respectively. Using the definition of D_g, we have that

$$\langle g'(x^{\ell_j}) - g'(x^{i_j}), \hat{x} - \tilde{x} \rangle = D_g(\hat{x}, x^{i_j}) - D_g(\hat{x}, x^{\ell_j}) + D_g(\tilde{x}, x^{\ell_j}) - D_g(\tilde{x}, x^{i_j}).$$

Therefore

$$\left| \langle g'(x^{\ell_j}) - g'(x^{i_j}), \hat{x} - \tilde{x} \rangle \right| \leq \left| D_g(\hat{x}, x^{i_j}) - D_g(\hat{x}, x^{\ell_j}) \right| + \left| D_g(\tilde{x}, x^{\ell_j}) - D_g(\tilde{x}, x^{i_j}) \right|. \tag{2.35}$$

Taking limits in (2.35) as $j \to \infty$, we get

$$\liminf_{j \to \infty} \left| \langle g'(x^{\ell_j}) - g'(x^{i_j}), \hat{x} - \tilde{x} \rangle \right| \leq |\eta - \eta| + |\mu - \mu| = 0,$$

which contradicts H6. As a result, $\tilde{x} = \hat{x}$. ∎

2.3.2 Convergence Analysis of Algorithm IPPE

The next proposition establishes the basic property of the generated sequence, in terms of the Bregman distance from its iterates to any point of the solution set.

Proposition 2.3.6. *Consider* EP(f, K) *where* f *satisfies* P1–P4 *and any one among* P4', P4" *and* P4*. *Take* $g : B \to \mathbb{R}$ *satisfying* H1–H5. *Let* $\{x^j\}_{j=0}^{\infty}$, $\{\tilde{x}^j\}_{j=0}^{\infty}$, $\{\gamma_j\}_{j=0}^{\infty}$ *and* σ *be as in Algorithm* IPPE. *Then*

$$D_g(x^*, x^{j+1}) \leq D_g(x^*, x^j) - \gamma_j^{-1}\langle v^j, \tilde{x}^j - x^*\rangle - (1-\sigma)D_g(\tilde{x}^j, x^j) \leq D_g(x^*, x^j), \quad (2.36)$$

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

Proof. As in the case of Algorithm IPPBP, we define the auxiliary vector v^j as

$$v^j = \gamma_j[g'(x^j) - g'(\tilde{x}^j)] + e^j, \quad (2.37)$$

so that (2.21) can be rewritten as

$$0 = \gamma_j^{-1}v^j + g'(x^{j+1}) - g'(x^j). \quad (2.38)$$

Replacing (2.21) in (2.20), we obtain

$$D_g(\tilde{x}^j, x^{j+1}) \leq \sigma D_g(\tilde{x}^j, x^j). \quad (2.39)$$

Note that

$$\begin{aligned} D_g(x^*, x^{j+1}) &= D_g(x^*, x^j) + \langle g'(x^j) - g'(x^{j+1}), x^* - \tilde{x}^j\rangle + D_g(\tilde{x}^j, x^{j+1}) - D_g(\tilde{x}^j, x^j) \\ &= D_g(x^*, x^j) + \langle \gamma_j^{-1}v^j, x^* - \tilde{x}^j\rangle + D_g(\tilde{x}^j, x^{j+1}) - D_g(\tilde{x}^j, x^j) \\ &\leq D_g(x^*, x^j) - \gamma_j^{-1}\langle v^j, \tilde{x}^j - x^*\rangle - (1-\sigma)D_g(\tilde{x}^j, x^j), \end{aligned} \quad (2.40)$$

using the definition of D_g in the first equality, (2.38) in the second equality and (2.39) in the inequality. We have proved the leftmost inequality in (2.36).

Since D_g is nonnegative and $\sigma \in [0, 1)$, we obtain from (2.40)

$$D_g(x^*, x^{j+1}) \leq D_g(x^*, x^j) - \gamma_j^{-1}\langle v^j, \tilde{x}^j - x^*\rangle. \quad (2.41)$$

On the other hand, $\tilde{x}^j \in S_E(f_j^e, K)$. Consequently

$$f(\tilde{x}^j, y) + \gamma_j^{-1}\langle -v^j, y - \tilde{x}^j\rangle \geq 0, \quad (2.42)$$

for all $y \in K$. In particular, (2.42) holds for $y = x^*$, so that $f(\tilde{x}^j, x^*) \geq \gamma_j^{-1}\langle v^j, x^* - \tilde{x}^j\rangle$ which is equivalent to $f(\tilde{x}^j, x^*) \geq -\gamma_j^{-1}\langle v^j, \tilde{x}^j - x^*\rangle$. In view of Proposition 2.3.2, any one among P4', P4" and P4* implies that $f(\tilde{x}^j, x^*) \leq 0$. Henceforth, we have that $-\gamma_j^{-1}\langle v^j, \tilde{x}^j - x^*\rangle \leq 0$. Replacing this inequality in (2.41), we obtain the rightmost inequality in (2.36). ∎

The remaining convergence properties of the sequences $\{x^j\}_{j=0}^{\infty}, \{\tilde{x}^j\}_{j=0}^{\infty}, \{v^j\}_{j=0}^{\infty}$ are established in the following proposition.

Proposition 2.3.7. *Consider* EP(f, K) *such that* K *has nonempty relative interior. Assume that* f *satisfies* P1–P4 *and also any one among* P4', P4" *and* P4*. *Take* $g : B \to \mathbb{R}$ *satisfying* H1–H5, $\sigma \in [0, 1)$, *and an exogenous sequence* $\{\gamma_j\}_{j=0}^{\infty} \subset (\theta, \bar{\gamma}]$, *where* θ *is the undermonotonicity constant in* P4. *Let* $\{x^j\}_{j=0}^{\infty}$ *be the sequence generated by Algorithm* IPPE. *If* EP(f, K) *has solutions, then the following statements are true.*

i) $\{D_g(x^*, x^j)\}_{j=0}^{\infty}$ *is nonincreasing and convergent for all* $x^* \in S_E(f, K)$.

ii) $\{x^j\}_{j=0}^{\infty}$ *is bounded.*

iii) $\sum_{j=0}^{\infty} \gamma_j^{-1}\langle v^j, \tilde{x}^j - x^*\rangle < \infty$ *where* v^j *given by* (2.37).

iv) $\sum_{j=0}^{\infty} D_g(\tilde{x}^j, x^j) < \infty$.

v) $\sum_{j=0}^{\infty} D_g(\tilde{x}^j, x^{j+1}) < \infty$.

vi) $\{\tilde{x}^j - x^j\}_{j=0}^{\infty}$ *converges strongly to* 0, *and consequently* $\{\tilde{x}^j\}_{j=0}^{\infty}$ *is bounded.*

vii) $\{x^{j+1} - x^j\}_{j=0}^{\infty}$ *converges strongly to* 0.

viii) $\{v^j\}_{j=0}^{\infty}$ *converges strongly to* 0.

Proof. Take $x^* \in S_E(f, K)$. By Proposition 2.3.6, $\{D_g(x^*, x^j)\}_{j=0}^{\infty}$ is a nonnegative and nonincreasing sequence, henceforth convergent, and $\{x^j\}_{j=0}^{\infty}$ is contained in a level set of $D_g(x^*, \cdot)$, which is bounded by H1, establishing (i)–(ii). Invoking again Proposition 2.3.6,

$$0 \leq \gamma_j^{-1}\langle v^j, \tilde{x}^j - x^*\rangle + (1 - \sigma)D_g(\tilde{x}^j, x^j) \leq D_g(x^*, x^j) - D_g(x^*, x^{j+1}),$$

from which (iii) and (iv) follow easily. Item (v) follows from (iv) and (2.39). For (vi) and (vii), observe that $\lim_{j\to\infty} D_g(\tilde{x}^j, x^j) = \lim_{j\to\infty} D_g(\tilde{x}^j, x^{j+1}) = 0$ as a consequence of (iv), (v). Since $\{x^j\}_{j=0}^{\infty}$ is bounded by (i), we can apply Proposition 1.5.2 to obtain the strong convergence of $\{\tilde{x}^j - x^j\}_{j=0}^{\infty}$ and $\{\tilde{x}^j - x^{j+1}\}_{j=0}^{\infty}$ to 0, which entails the strong convergence of $\{x^j - x^{j+1}\}_{j=0}^{\infty}$ to 0. Finally, (viii) is obtained from (vii) and (2.38) taking limits with $j \to \infty$, using the facts that H3 holds and $\{\gamma_j\}_{j=0}^{\infty} \subset (\theta, \bar{\gamma}]$. ∎

The following theorem completes the convergence analysis of Algorithm IPPE.

Theorem 2.3.8. *Consider* EP(f, K) *such that* K *has nonempty relative interior. Assume that* f *satisfies* P1–P4 *and also any one among* P4', P4" *and* P4*. *Take* $g : B \to \mathbb{R}$ *satisfying* H1–H5, $\sigma \in [0, 1)$, *and an exogenous sequence* $\{\gamma_j\}_{j=0}^{\infty} \subset (\theta, \bar{\gamma}]$, *where* θ *is the undermonotonicity constant in* P4. *Let* $\{x^j\}_{j=0}^{\infty}$ *be the sequence generated by Algorithm* IPPE. *If* EP(f, K) *has solutions, then the following statements hold.*

i) $\{\tilde{x}^j\}_{j=0}^{\infty}$ *is an asymptotically solving sequence for* EP(f, K).

ii) If $f(\cdot, y)$ *is weakly upper semicontinuous for all* $y \in K$, *then all cluster points of* $\{x^j\}_{j=0}^{\infty}$ *solve* EP(f, K).

iii) If in addition either g *satisfies* H6 *or* EP(f, K) *has a unique solution, then the whole sequence* $\{x^j\}_{j=0}^{\infty}$ *is weakly convergent to some solution* x^* *of* EP(f, K).

Proof. The proof is similar to the one of Theorem 2.3.5, using now Proposition 2.3.7 instead of Proposition 2.3.4. ■

We comment that when B is a strictly convex and smooth Banach space and we take $g(x) = \|x\|^p$, then we have an explicit formula for $(g')^{-1}$, in term of ϕ', where ϕ is defined as $\phi(x) = \frac{1}{q}\|x\|_*^q$ with $\frac{1}{p} + \frac{1}{q} = 1$. Indeed, $(g')^{-1} = p^{1-q}\phi'$.

Now we look at the two additional conditions imposed in items (ii) and (iii) of Theorem 2.3.5 and Theorem 2.3.8. Not much can be done about assumption H6 on g, required for uniqueness of the weak limit points of $\{x^j\}_{j=0}^{\infty}$: it is rather demanding (as mentioned before, functions satisfying it are known only in very special Banach spaces), but it seems to be inherent to proximal point methods in Banach spaces, even in the exact version of the method for finding zeroes of monotone operators (e.g., in [12]). The situation is different in connection with weak upper semicontinuity of $f(\cdot, y)$, required for establishing that the weak limit points of $\{x^j\}_{j=0}^{\infty}$ are solutions of the equilibrium problem. It is also quite demanding, but we mention that it holds at least in two significant cases: when B is finite dimensional (where it follows from P3, since in this case weak and strong continuity coincide) and when $f(\cdot, y)$ is concave for all $y \in K$. Moreover, it is possible to replace it by a much weaker hypothesis, through a reformulation of the equilibrium problem, to be developed in the following section, but only for the case of a monotone f, i.e. satisfying P4$^\bullet$ instead of P4.

2.4 A Reformulation of the Equilibrium Problem

We will establish here that $S_E(f, K)$ coincides with the set of zeroes of a certain point-to-set operator (see also Remark 1.10.2). We will prove that when f is monotone the graph of this operator enjoys the demiclosedness property, in the sense of Definition 1.3.1(v), which enables us to get rid of the weak upper semicontinuity of $f(\cdot, y)$ in the convergence analysis of our algorithms.

Throughout this section we also assume that $\partial F_x(y) \neq \emptyset$ for all $x, y \in K$, where the function F_x is given by (2.6) for every $x \in K$. This is the case, for instance, if $f(x, \cdot)$ can be extended, preserving its convexity and its continuity, to some open subset W of B, containing K, for all $x \in K$ as described in Proposition 1.4.2(ii)–(iii). We associate to EP(f, K) the operator $T^f : B \to \mathcal{P}(B^*)$ defined as

$$T^f(x) = \partial F_x(x) + N_K(x), \tag{2.43}$$

where F_x is as in (2.6), i.e. $F_x(y) = f(x, y)$, and $\partial F_x(y)$ denotes its subdifferential at the point y.

Proposition 2.4.1. *The set of zeroes of T^f is equal to $S_E(f, K)$.*

Proof. By definition, we have that

$$x^* \in S_E(f, K) \Leftrightarrow F_{x^*}(x^*) = f(x^*, x^*) = 0 \leq f(x^*, y) = F_{x^*}(y) \quad \forall y \in K.$$

So, we have that $x^* \in S_E(f, K)$ if and only if x^* minimizes the function F_{x^*} over K. Since F_{x^*} and K are convex, taking into account Theorem 1.7.2, the necessary and sufficient condition for x^* to be a minimizer of F_{x^*} over K is the existence of $v^* \in \partial F_{x^*}(x^*)$ such that $\langle v^*, y - x^* \rangle \geq 0$ for all $y \in K$. In view of (1.1), that is precisely equivalent to saying that $0 \in \partial F_{x^*}(x^*) + N_K(x^*) = T^f(x^*)$ which completes the demonstration. ■

Corollary 2.4.2. *Consider the sequence $\{\tilde{x}^j\}_{j=0}^{\infty}$ generated by either Algorithm* IPPBP *or Algorithm* IPPE. *Then $\tilde{x}^j$ is the zero of $T^{f_j^e}$, as defined in* (2.43), *with f_j^e as in* (2.13), *i.e.*

$$e^j + \gamma_j [g'(x^j) - g'(\tilde{x}^j)] \in \partial F_{\tilde{x}^j}(\tilde{x}^j) + N_K(\tilde{x}^j). \tag{2.44}$$

Proof. Since $\tilde{x}^j$ is the solution of EP(f_j^e, K), by Proposition 2.4.1, it is a zero of $T^{f_j^e}$. The rule of subdifferential calculus, expressed in Proposition 1.4.4, applied to the convex function

$$\varphi(y) = f_j^e(x, y) = f(x, y) + \gamma_j \langle g'(x) - g'(x^j), y - x \rangle - \langle e^j, y - x \rangle$$

lead to (2.44), after taking $x = \tilde{x}^j$ and evaluating both N_K and the subdifferential of φ at the same point, namely $\tilde{x}^j$. ■

Corollary 2.4.2 allows us to rewrite the iterative step of our algorithms. For the case of Algorithm IPPBP, Step 2 is equivalent to finding a pair $(\tilde{x}^j, e^j) \in B \times B^*$ such that

$$e^j + \gamma_j [g'(x^j) - g'(\tilde{x}^j)] \in \partial F_{\tilde{x}^j}(\tilde{x}^j) + N_K(\tilde{x}^j), \tag{2.45}$$

$$\left\| e^j \right\|_* \leq \sigma \gamma_j \begin{cases} D_g(\tilde{x}^j, x^j) & \text{if } \|x^j - \tilde{x}^j\| < 1 \\ \nu_g(x^j, 1) & \text{if } \|x^j - \tilde{x}^j\| \geq 1. \end{cases} \tag{2.46}$$

Now we must explore the connection between the monotonicity properties of function f and operator T^f. Consider $\theta = 0$, $g : B \to \mathbb{R}$ satisfying H1–H5 and define $U^f : B \to \mathcal{P}(B^*)$ as

$$U^f(x) = \begin{cases} \partial F_x(x) & \text{if } x \in K \\ \emptyset & \text{otherwise.} \end{cases} \tag{2.47}$$

It is essential to note that U^f *is not* the subdifferential of a convex function; rather, at each point x it is the subdifferential of a certain convex function, namely F_x, but this function changes with the argument of the operator. Thus, the monotonicity of U is not guaranteed "a priori", but we have the following elementary result.

Proposition 2.4.3. *If f satisfies* P1–P2 *and* P4$^\bullet$*, then U^f is monotone.*

Proof. In view of (2.47), we only need to worry with points $x, y \in K$. Take $x, y \in K, u \in U^f(x)$ and $v \in U^f(y)$. Using, P1, P2 and the definition of ∂F_x, we obtain

$$\langle u, y - x \rangle \leq F_x(y) - F_x(x) = f(x, y) - f(x, x) = f(x, y), \tag{2.48}$$

$$\langle v, x - y \rangle \leq F_y(x) - F_y(y) = f(y, x) - f(y, y) = f(y, x). \tag{2.49}$$

Adding (2.48) and (2.49) we get

$$-\langle u - v, x - y \rangle = \langle u, y - x \rangle + \langle v, x - y \rangle \leq f(x, y) + f(y, x) \leq 0, \tag{2.50}$$

using P4$^\bullet$ in the last inequality. We conclude from (2.50) that $0 \leq \langle u - v, x - y \rangle$, establishing monotonicity of U^f. ∎

In order to establish the desired demiclosedness property, we need maximal monotonicity of U^f. This is less elementary, and we will invoke again the existence result in Proposition 2.1.1.

The following result has been established in Theorem 4.5.7 of [11] for the case of $g(x) = \frac{1}{2} \|x\|^2$, $\lambda = 1$. See also Remark 10.8 in [61].

Proposition 2.4.4. *If $T : B \to \mathcal{P}(B^*)$ is monotone, $g : B \to \mathbb{R}$ satisfies* H2 *and the operator $T + \lambda g'$ is onto for some $\lambda > 0$, then T is maximal monotone.*

Proof. Take a monotone operator $\bar{T}$ such that $T \subset \bar{T}$, and a pair (v, z) such that $v \in \bar{T}(z)$. We must show that $v \in T(z)$. Define $b = v + \lambda g'(z)$. Since $T + \lambda g'$ is onto, there exists $x \in B$ such that

$$b = v + \lambda g'(z) \in T(x) + \lambda g'(x) \subseteq \bar{T}(x) + \lambda g'(x). \tag{2.51}$$

On the other hand, since $v \in \bar{T}(z)$, we have that

$$b = v + \lambda g'(z) \in \bar{T}(z) + \lambda g'(z). \tag{2.52}$$

Since g' is strictly monotone by H2, the same holds for $\bar{T} + \lambda g'$, and it follows therefore from (2.51) and (2.52) that $x = z$. Thus, making $x = z$ in the first inclusion of (2.51), we get $v + \lambda g'(z) \in T(z) + \lambda g'(z)$, which implies that $v \in T(z)$. It follows that $\bar{T} \subset T$, and hence T is maximal. ■

Now we use Proposition 2.1.1 and Proposition 2.4.4 to establish maximal monotonicity of T^f.

Proposition 2.4.5. *If f satisfies* P1–P3, P4$^\bullet$ *and g satisfies* H1–H2, H4–H5, *then T^f, as defined in* (2.43), *is maximal monotone.*

Proof. Consider $T^f = U^f + N_K$ as in (2.43). We want to use Proposition 2.4.4, for which we need to show that T^f is monotone and that $T^f + \lambda g'$ is onto for some $\lambda > 0$. Note that U^f is monotone by Proposition 2.4.3. Since N_K is certainly monotone, it follows that T^f is monotone as well. Now we address the surjectivity issue. Take any $\lambda > 0$ and $b \in B^*$. We want to prove the existence of some $x \in K$ such that $b \in (T^f + \lambda g')(x)$. Consider $\tilde{f}$ as in (2.3) with $e = 0$, $\bar{x} \in B$ and $\gamma = \lambda$ satisfying $g'(\bar{x}) = \lambda^{-1} b$, i.e.

$$\tilde{f}(x, y) = f(x, y) + \lambda \langle g'(x) - \lambda^{-1} b, y - x \rangle. \tag{2.53}$$

Note that such an $\bar{x}$ exists by H4, and that $\mathrm{EP}(\tilde{f}, K)$ has a unique solution by Proposition 2.1.1, say $\hat{x}$. Since $\tilde{f}(\hat{x}, y) \geq 0$ for all $y \in K$ and $\tilde{f}(\hat{x}, \hat{x}) = 0$, $\hat{x}$ solves the following convex minimization problem:

$$\min \tilde{f}(\hat{x}, y) \;\; s.t. \;\; y \in K.$$

Thus, $\hat{x}$ satisfies the first order optimality condition for this problem (see Theorem 1.7.2), which is, in view of (2.53) and the differentiability of g,

$$0 \in \partial F_{\hat{x}}(\hat{x}) + \lambda [g'(\hat{x}) - \lambda^{-1} b] + N_K(\hat{x}),$$

or equivalently

$$b \in \partial F_{\hat{x}}(\hat{x}) + \lambda g'(\hat{x}) + N_K(\hat{x}) = T^f(\hat{x}) + \lambda g'(\hat{x}).$$

We have established surjectivity of $T^f + \lambda g'$. We now apply Proposition 2.4.4 to conclude that T^f is maximal monotone. ∎

Now we can get rid of the weak upper semicontinuity assumption in Theorem 2.3.5 and Theorem 2.3.8, using the demiclosedness property of maximal monotone operators announced in Proposition 1.3.2.

Theorem 2.4.6. *Assume that*

i) f *satisfies* P1–P3, *and* P4$^\bullet$,

ii) $g : B \to \mathbb{R}$ *satisfies* H1–H5,

iii) $\{\gamma_j\}_{j=0}^\infty$ *is contained in* $(0, \bar{\gamma}]$ *for some* $\bar{\gamma} > 0$,

iv) $\sigma \in [0, 1)$,

v) $f(x, \cdot)$ *can be extended, for all* $x \in K$, *to an open set* $W \supset K$, *while preserving its convexity,*

vi) K *has nonempty relative interior,*

vii) $\mathrm{EP}(f, K)$ *has solutions,*

then, for all $x^0 \in K$, *the sequence* $\{x^j\}_{j=0}^\infty$ *generated by either Algorithm* IPPBP *or Algorithm* IPPE *is bounded and all its weak cluster points are solutions of* $\mathrm{EP}(f, K)$. *If moreover* g *satisfies* H6 *or* $\mathrm{EP}(f, K)$ *has a unique solution, then* $\{x^j\}_{j=0}^\infty$ *is weakly convergent to a solution of* $\mathrm{EP}(f, K)$.

Proof. We are within the assumptions of Proposition 2.3.4 and Proposition 2.3.7 with $\theta = 0$. Clearly, P4$^\bullet$ is the same as P4 with $\theta = 0$, and, as discussed in Proposition 1.9.4(i), P4$^\bullet$ implies P4*, for instance. Consider first Algorithm IPPBP. The sequence $\{x^j\}_{j=0}^\infty$ is bounded by Proposition 2.3.4(ii). Let $\bar{x}$ be a cluster point of $\{x^j\}_{j=0}^\infty$. Let $\{x^{\ell_j}\}_{\ell=0}^\infty$ be a subsequence of $\{x^j\}_{j=0}^\infty$ weakly convergent to $\bar{x}$. By Proposition 2.3.4(v), the subsequence $\{\tilde{x}^{\ell_j}\}_{\ell=0}^\infty$ is also weakly convergent to $\bar{x}$. In view of Corollary 2.4.2, (2.16) and (2.43), we have $v^{\ell_j} \in T^f(\tilde{x}^{\ell_j})$. By Proposition 2.3.4(vi), $\{v^{\ell_j}\}_{\ell=0}^\infty$ is strongly convergent to 0. Since T^f is maximal monotone by Proposition 2.4.5, its graph is demiclosed by Proposition 1.3.2. It follows from Definition 1.3.1(v) that $0 \in T^f(\bar{x})$, and therefore $\bar{x} \in S_E(f, K)$, in view of Proposition 2.4.1. Uniqueness of the weak cluster point of $\{x^j\}_{j=0}^\infty$ when g satisfies H6 follows exactly as in the proof of Theorem 2.3.5(iii). The case of Algorithm IPPE is dealt with in a similar way, invoking now Proposition 2.3.7 instead of Proposition 2.3.4. ∎

We remark that the difference between Theorem 2.3.5 and Theorem 2.3.8 on one side, and Theorem 2.4.6, besides the fact that the proof of the latter requires the reformulation of the equilibrium problem as a variational inequality one, lies in the technical assumption on the extension of $f(x, \cdot)$ to an open set containing K, which replaces weak upper semicontinuity of $f(\cdot, y)$, as the tool for establishing optimality of the weak cluster points of the generated sequence.

At this point we mention that, under the reformulation, Algorithm IPPBP and Algorithm IPPE coincide with Algorithm I and Algorithm II described in Section 1.8, (i.e., Algorithm I and Algorithm II proposed in [26]) designed for finding zeroes of maximal monotone operators in Banach spaces. Thus, we could have omitted the proof of Theorem 2.4.6, which has been included just for making Chapter 2 more self-contained. On the other hand, the results in [26] demand monotonicity of the operator, akin to monotonicity of f in the case of equilibrium problem, while in Section 2.3 we worked under the weaker assumptions of pseudomonotonicity and θ-undermonotonicity (see Example 1.9.3). An inexact proximal point method for finding zeroes of nonmonotone operators in Banach spaces has appeared in [22], but in this reference the operator is assumed to be θ-*hypomonotone*. In the context of equilibrium problem, this is the same as stating that $[(T^f)^{-1} + \theta(g')^{-1}]^{-1}$ is monotone (see Lemma 1 of [22]), while our θ-undermonotonicity assumption entails that $T^f + \theta g'$ is monotone. Also, the relations between θ and the regularization parameters γ_j in [22] and in this thesis are different. A thorough discussion on the connection between our exact proximal point method and the use of proximal point methods for finding zeroes of operators applied to solving equilibrium problems via the reformulation can be found in [32].

The reformulation is also useful for visualizing the way in which our error criteria work. In practice, one assumes that one has some algorithm for solving the subproblem $\mathrm{EP}(f_j, K)$, which generates a sequence, say $\{x^{j,k}\}_{k=0}^{\infty}$. At each iteration, one should check whether $x^{j,k}$ satisfies the error criteria in order to be accepted as an approximate solution of $\mathrm{EP}(f_j, K)$. In view of (2.45), (2.46), for the case of Algorithm IPPBP one should verify whether there exists some e^j such that the pair $(x^{j,k}, e^j) \in B \times B^*$ satisfies

$$e^j + \gamma_j[g'(x^j) - g'(x^{j,k})] \in \partial F_{x^{j,k}}(x^{j,k}) + N_K(x^{j,k}),$$

$$\left\|e^j\right\|_* \leq \sigma\gamma_j \begin{cases} D_g(x^{j,k}, x^j) & \text{if } \left\|x^j - x^{j,k}\right\| < 1 \\ \nu_g(x^j, 1) & \text{if } \left\|x^j - x^{j,k}\right\| \geq 1. \end{cases}$$

If so, we take $\tilde{x}^j = x^{j,k}$ and continue with the computation of x^{j+1}; otherwise we take another step of the inner loop and repeat the check with $x^{j,k+1}$.

In order to support the given notion of approximate solution, it is important to verify that, at least in some "nice" cases, any feasible point close enough to the exact solution of the subproblem $\mathrm{EP}(f_j, K)$ will satisfy our criterion for an approximate solution, i.e. will solve $\mathrm{EP}(f_j^e, K)$ for some appropriate e^j. In such a case, if we use in the inner loop an algorithm known to converge to the exact solution of the subproblem, after a finite number of steps of the auxiliary algorithm we will end up with a vector satisfying our criteria for being an approximate solution.

We show next that this situation occurs in the smooth case, by which we mean that the function $F_x(y) := f(x,y)$ is Fréchet differentiable and that the boundary ∂K of K is smooth, in the following sense:

Definition 2.4.7. *Assume that K is a closed convex set with nonempty relative interior. We say that the boundary of K, denoted by ∂K and defined as $\partial K = K \setminus int(K)$, is smooth if there exists a Fréchet differentiable convex function $h : B \to \mathbb{R}$ such that $K = \{x \in B : h(x) \leq 0\}$ and $h'(x) \neq 0$ for all $x \in \partial K$.*

Proposition 2.4.8. *If the boundary of K is smooth then $N_K(x) = \{th'(x) : t \geq 0\}$ for all $x \in \partial K$.*

Proof. Without loss of generality we can assume that $int(K) \neq \emptyset$ (otherwise, we consider the relative interior of the set K). Take $x \in \partial K$, so that $h(x) = 0$. The fact that the halfline through $h'(x)$ is contained in $N_K(x)$ follows easily from the definition of $N_K(x)$ given by (1.1), the definition of subdifferential, and Proposition 1.4.10(i). In other words, from Definition 1.4.1(iv) and the fact that $h(x) = 0$, we have that

$$\langle h'(x), y - x\rangle \leq h(y) - h(x) = h(y) \leq 0 \quad \forall\, y \in K, \tag{2.54}$$

where the rightmost inequality is a consequence of the definition of K. For the reverse inclusion, we first verify that $-h'(x) \notin N_K(x)$. By contradiction, we assume that $-h'(x) \in N_K(x)$, which implies

$$\langle -h'(x), z - x\rangle \leq 0 \quad \forall\, z \in K. \tag{2.55}$$

Using (2.54) and (2.55), we get

$$\langle h'(x), y - z\rangle \leq 0 \quad \forall\, y, z \in K,$$

which is equivalent to saying that

$$\langle h'(x), y - z\rangle = 0 \quad \forall\, y, z \in K,$$

but this is a contradiction with hypothesis $h'(x) \neq 0$. Now we assume that there exists $\Lambda \in N_K(x)$ such that Λ and $h'(x)$ are linearly independent. In this situation, we claim that there exists $z \in B$ such that

$$\langle h'(x), z\rangle < 0, \tag{2.56}$$

$$\langle \Lambda, z\rangle > 0. \tag{2.57}$$

To prove the above assertion, we define subsets $A_1^<$, $A_1^>$, $A_2^<$ and $A_2^>$ of Banach space B as

$$A_1^< = \{y \in B : \langle h'(x), y\rangle < 0\}, \;\; A_2^< = \{y \in B : \langle \Lambda, y\rangle < 0\},$$

$$A_1^> = \{y \in B : \langle h'(x), y\rangle > 0\}, \;\; A_2^> = \{y \in B : \langle \Lambda, y\rangle > 0\},$$

and we show that $A_1^< \cap A_2^> \neq \emptyset$, or, equivalently $A_1^> \cap A_2^< \neq \emptyset$. By contradiction, we assume that $A_1^< \cap A_2^> = \emptyset$, or, equivalently $A_1^> \cap A_2^< = \emptyset$. It follows from the structure of $A_1^<$, $A_1^>$, $A_2^<$ and $A_2^>$ that $A_1^< \subset cl(A_2^<) = A_2^< \cup \ker(\Lambda)$ and $A_1^> \subset cl(A_2^>) = A_2^> \cup \ker(\Lambda)$, because $B = A_2^> \cup cl(A_2^<) = A_2^< \cup cl(A_2^>)$. Hence,

$$cl(A_1^<) = A_1^< \cup \ker(h'(x)) \subset A_2^< \cup \ker(\Lambda) = cl(A_2^<),$$

$$cl(A_1^>) = A_1^> \cup \ker(h'(x)) \subset A_2^> \cup \ker(\Lambda) = cl(A_2^>),$$

which implies

$$cl(A_1^<) \cap cl(A_1^>) = \ker(h'(x)) \subset \ker(\Lambda) = cl(A_2^<) \cap cl(A_2^>).$$

As a result of this fact, $\ker(h'(x)) \subset \ker(\Lambda)$. In the same line, one can show that $\ker(\Lambda) \subset \ker(h'(x))$, consequently, $\ker(h'(x)) = \ker(\Lambda)$, contradicting the linear independence of Λ and $h'(x)$ announced in Lemma 1.1.17 (take $n = 1$ and $\Lambda_1 = h'(x)$ in Lemma 1.1.17). Thus there exists some $z \in B$ satisfying (2.56) and (2.57). Let $y = x + tz$ with respect to such a z, and then consider the statement

$$h(y) = h(x + tz) = h(x) + t\langle h'(x), z\rangle + o(t) = t\langle h'(x), z\rangle + o(t),$$

which follows from differentiability of h and the fact that $h(x) = 0$. Consequently, we conclude from (2.56) that $y \in K$ for small enough $t > 0$, so that, in view of (1.1), $0 \geq \langle \Lambda, y - x\rangle = t\langle \Lambda, z\rangle$, which contradicts (2.57). ■

Theorem 2.4.9. *Under the hypotheses of Theorem 2.3.5 and Theorem 2.3.8, assume additionally that K has smooth boundary, $F_x : K \to \mathbb{R}$ defined as*

$$F_x(y) = f(x, y),$$

is Fréchet differentiable for each $x \in K$ and that $(g')^{-1}$ is continuous (for Algorithm IPPE*). Let $\{x^j\}_{j=0}^{\infty}$ be the sequence generated either by Algorithm* IPPBP *or Algorithm* IPPE*. Assume that x^j is not a solution of* EP(f, K) *and let $\hat{x}^j$ be the unique solution of* EP(f_j, K) *with f_j defined as*

$$f_j(x, y) = f(x, y) + \gamma_j\langle g'(x) - g'(x^j), y - x\rangle.$$

Then there exists $\delta_j > 0$ such that *i) if $\hat{x}^j$ belongs to $int(K)$, then any $x \in B(\hat{x}^j, \delta_j) \cap K$ solves the subproblem* (2.14)–(2.15) *in the case of Algorithm* IPPBP *and* (2.19)–(2.20) *in the case of Algorithm* IPPE*,*

ii) if $\hat{x}^j$ belongs to ∂K, then any $x \in B(\hat{x}^j, \delta_j) \cap \partial K$ solves the subproblem (2.14)–(2.15) *in the case of Algorithm* IPPBP *and* (2.19)–(2.20) *in the case of Algorithm* IPPE*.*

Proof. Note that if we take $e^j = 0$, then $\mathrm{EP}(f_j^e, K)$ reduces to the exact regularized equilibrium problem $\mathrm{EP}(f_j, K)$. If $\{x^j\}_{j=0}^{\infty}$ is a sequence generated by either Algorithm IPPBP or Algorithm IPPE, define $\Lambda_j : B \to B^*$ as

$$\Lambda_j(x) = -F'_x(x) - \gamma_j[g'(x) - g'(x^j)], \tag{2.58}$$

where F'_x denotes the Fréchet derivative of F_x.
i) We first consider the case of Algorithm IPPBP. Note that $\hat{x}^j$ is the unique solution of $\mathrm{EP}(f_j, K)$ by Proposition 2.1.1. Define $\Phi_j : B \to \mathbb{R}$ as

$$\Phi_j(x) = \sigma\gamma_j \min\{D_g(x, x^j), v_g(x^j, 1)\}.$$

Since $x^j \neq \hat{x}^j$, because otherwise x^j solves $\mathrm{EP}(f, K)$, we have that $D_g(\hat{x}^j, x^j) > 0$ and, by H2,

$$\Phi_j(\hat{x}^j) > 0. \tag{2.59}$$

Assume that $\hat{x}^j \in int(K)$. In this circumstance, there exists $\delta_j > 0$ such that $B(\hat{x}^j, \delta_j) \subset K$ and $N_K(x) = \{0\}$ for all $x \in B(\hat{x}^j, \delta_j)$. Since $\hat{x}^j$ is the exact solution of the j-th subproblem, it satisfies (2.44) with $e^j = 0$, $N_K(\hat{x}^j) = 0$, which implies, in view of (2.58), that

$$\Lambda(\hat{x}^j) = 0. \tag{2.60}$$

Define $\Psi_j : K \to \mathbb{R}$ as

$$\Psi_j(x) = \|\Lambda_j(x)\|_* - \Phi_j(x).$$

By (2.59)–(2.60),

$$\Psi_j(\hat{x}^j) = -\Phi_j(\hat{x}^j) < 0.$$

Continuity of F'_x and g' ensure continuity of both Λ_j and Φ_j, and therefore of Ψ_j as well. By (2.60) and continuity of Ψ_j we can choose the above δ_j small enough such that $\Psi_j(x) \leq 0$ for all $x \in B(\hat{x}^j, \delta_j) \cap K$ or equivalently $\|\Lambda_j(x)\|_* \leq \Phi_j(x)$ for all $x \in B(\hat{x}^j, \delta_j) \cap K$. In such a case, any pair (x, e) with $x \in B(\hat{x}^j, \delta_j) \cap K$, $e = \Lambda_j(x)$, will satisfy (2.45)–(2.46), with (x, e) substituting for $(\tilde{x}^j, e^j)$. Thus any such x can be taken as the $\tilde{x}^j$ required by Algorithm IPPBP.

For the case of Algorithm IPPE, we replace Ψ_j by $\bar{\Psi}_j : K \to \mathbb{R}$ defined as

$$\bar{\Psi}_j(x) = D_g(x, (g')^{-1}[g'(x^j) - \gamma_j^{-1} F'_x(x)]) - \sigma D_g(x, x^j),$$

and proceed with the same argument.

ii) We start with Algorithm IPPBP. Assume that $\hat{x}^j \in \partial K$. Since $\hat{x}^j$ is the exact solution of the j-th subproblem, it satisfies (2.44) with $e^j = 0$, so that, in view of (2.58), $\Lambda(\hat{x}^j) \in N_K(\hat{x}^j)$. Since ∂K is smooth, we get from Proposition 2.4.8 that $N_K(\hat{x}^j) = \{th'(\hat{x}^j) : t \geq 0\}$, so that there exists $t^* \geq 0$ such that $\Lambda_j(\hat{x}^j) = t^* h'(\hat{x}^j)$. Since $h'(\hat{x}^j) \neq 0$ by Definition 2.4.7, we get $t^* = \|\Lambda_j(\hat{x}^j)\|_* / \|h'(\hat{x}^j)\|_*$ and therefore

$$\frac{\|\Lambda_j(\hat{x}^j)\|_*}{\|h'(\hat{x}^j)\|_*} h'(\hat{x}^j) - \Lambda_j(\hat{x}^j) = 0. \tag{2.61}$$

Define $\hat{\Psi}_j : K \to \mathbb{R}$ as

$$\hat{\Psi}_j(x) = \left\| \frac{\|\Lambda_j(x)\|_*}{\|h'(x)\|_*} h'(x) - \Lambda_j(x) \right\|_* - \Phi_j(x).$$

$\hat{\Psi}_j$ is continuous by our smoothness assumption, and $\hat{\Psi}_j(\hat{x}^j) = -\Phi_j(\hat{x}^j) < 0$ by (2.61). Therefore, we can choose $\delta_j > 0$ such that $\hat{\Psi}_j(x) \leq 0$ for all $x \in B(\hat{x}^j, \delta_j) \cap \partial K$ or equivalently

$$\left\| \frac{\|\Lambda_j(x)\|_*}{\|h'(x)\|_*} h'(x) - \Lambda_j(x) \right\|_* \leq \Phi_j(x), \tag{2.62}$$

for all $x \in B(\hat{x}^j, \delta_j) \cap \partial K$. Let now $\xi(x) = \|\Lambda_j(x)\|_* / \|h'(x)\|_*$ and consider now a pair (x, e) with $x \in B(\hat{x}^j, \delta_j) \cap \partial K$ and $e = \xi(x)h'(x) - \Lambda_j(x)$. It follows from (2.62) and Proposition 2.4.8 that this pair satisfies (2.15) and (2.44) (with x instead of $\tilde{x}^j$). We also note that according to Proposition 2.4.1 an x satisfies (2.14) if and only if it satisfies (2.44), henceforth we establish the result.

For the case of Algorithm IPPE, we define, instead of $\hat{\Psi}_j$, the function $\tilde{\Psi}_j : K \to \mathbb{R}$ as

$$\tilde{\Psi}_j(x) = D_g\left(x, (g')^{-1}\left[g'(x) + \gamma_j^{-1}\Lambda_j(x) - \gamma_j^{-1}\frac{\|\Lambda_j(x)\|_*}{\|h'(x)\|_*} h'(x)\right]\right) - \sigma D_g(x, x^j),$$

and then argue as above. ■

The following example illustrates the fact that if ∂K is not smooth then the result of Theorem 2.4.9 fails to hold. We consider only Algorithm IPPBP (a similar example can be constructed for Algorithm IPPE).

Example 2.4.10. *Consider* $K = [\frac{1}{2}, 1] \times [\frac{1}{2}, 1] \subseteq \mathbb{R}^2$. *Take* $x = (x_1, x_2) \in K$ *and* $y = (y_1, y_2) \in K$. *Define* $f : K \times K \to \mathbb{R}$ *as* $f(x, y) = \langle x, x - y \rangle = x_1(x_1 - y_1) + x_2(x_2 - y_2)$.

Note that f satisfies P1–P4 with $g(x) = \frac{1}{2}\|x\|^2$. In fact, one can easily show that f is 1-undermonotone. It is clear that ∂K is not smooth. We also have that $S_E(f, K) = \{(1, 1)\}$. Put $x^0 = (\frac{1}{2}, \frac{1}{2})$, $\gamma_0 = \frac{10}{9} > 1 = \theta$ and $\sigma = \frac{2}{5}$. Then we have $f_0(x, y) = \langle \frac{10}{9}x^0 - \frac{1}{9}x, x - y\rangle$ and $S_E(f_0, K) = \{(1, 1)\}$. Take $e^0 = (e^0_1, e^0_2) \in \mathbb{R}^2$, so that $f^e_0(x, y) = \langle \frac{10}{9}x^0 - \frac{1}{9}x + e^0, x - y\rangle$. We claim that the unique solution of $\text{EP}(f^e_0, K)$ is $(1, 1)$, independently of the error vector e^j, thus falsifying the result of Theorem 2.4.9. Indeed, we need to verify that $(\frac{4}{9}, \frac{4}{9}) + e^0 \in N_K((1, 1))$ for all $e^0 \in \mathbb{R}^2$ such that

$$\|e^0\| \leq \frac{4}{9} D_g\left((1, 1), (\frac{1}{2}, \frac{1}{2})\right) = \frac{1}{9},$$

using the reformulation technique, i.e., two statements (2.45) and (2.46). Since

$$N_K((1, 1)) = \mathbb{R}^2_+,$$

as can be easily checked, the condition above becomes $(\frac{4}{9}, \frac{4}{9}) + e^0 \geq 0$, which certainly holds whenever $\|e^0\| \leq \frac{1}{9}$.

Chapter 3

Inexact Augmented Lagrangian Methods for Equilibrium Problems

This chapter is devoted to exact and inexact augmented Lagrangian methods for equilibrium problem. The material can be found in [28].

We introduce Algorithm IALE, standing for *Inexact Augmented Lagrangian-Extragradient Method*, for solving $\text{EP}(f, K)$. We establish the convergence properties of Algorithm IALE through the construction of an appropriate proximal point method for a certain equilibrium problem. We also construct and analyze a variant of Algorithm IALE, called Algorithm LIAL, standing for *Linearized Inexact Augmented Lagrangian-Extragradient Method*. We finish this chapter with some remarks.

For the sake of easier reference, the same as Chapter 2, we list here properties P1–P5 and the variants of P4, introduced in Section 1.9, which we will frequently use throughout this chapter. Before recalling these properties, we remark that we are going to work on $\mathbb{R}^n$ from now on. In other words, we take $B = \mathbb{R}^n$. In this setting, the most convenient auxiliary function g to be used in Algorithm IPPBP and Algorithm IPPE, as presented in Chapter 2, is $g(x) = \frac{1}{2}\|x\|^2$, which satisfies properties H1–H6 introduced in Section 1.5 and provides simpler computations, because $g'(x) = x$. So, we will restrict ourselves to this auxiliary function.

We point out that, since at each iteration of augmented Lagrangian methods one solves an unconstrained problem, it is reasonable to assume that the objective function f can be extended to the whole space. Thus, we assume that f can be extended to $\mathbb{R}^n \times \mathbb{R}^n$, while preserving the following properties.

P1: $f(x, x) = 0$ for all $x \in \mathbb{R}^n$.

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

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

P4: There exists $\theta \geq 0$ such that

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

P4" : For all $x^1, \ldots, x^m \in \mathbb{R}^n$ and all $\lambda_1, \ldots, \lambda_m \geq 0$ such that $\sum_{i=1}^m \lambda_i = 1$, it holds that

$$\sum_{i=1}^m \lambda_i f\left(x^i, \sum_{k=1}^m \lambda_k x^k\right) \leq 0.$$

We will not consider properties P4' and P4* in this chapter because Proposition 3.2.2 below is no longer true if P4' or P4* substitutes for P4". We also remind that both P4 and P4" are weaker than monotonicity of f (see Example 1.9.3). We will assume that the closed convex set K in EP(f,K) is defined by a finite set of convex inequality constraints. This is the most convenient formulation for developing computationally implementable augmented Lagrangian methods. More precisely, K is defined as

$$K = \{x \in \mathbb{R}^n : h_i(x) \leq 0 \;\; (1 \leq i \leq m)\}, \tag{3.1}$$

where $h_i : \mathbb{R}^n \to \mathbb{R}$ is convex $(1 \leq i \leq m)$. We will also assume that this set of constraints satisfies any standard constraint qualification, for instance the Slater's condition given in Definition 1.7.3.

3.1 Algorithm IALE for Equilibrium Problems

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

$$\mathcal{L}((x,\lambda),(y,\mu)) = f(x,y) + \sum_{i=1}^m \lambda_i h_i(y) - \sum_{i=1}^m \mu_i h_i(x). \tag{3.2}$$

It is worthwhile to mention that when we consider the convex minimization problem (1.4) with feasible set $C = K$ (which is as a particular case of EP(f,K) with $f(x,y) = h_0(y) - h_0(x)$), (3.2) 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 convex minimization problem defined in (1.17).

We introduce now our proximal augmented Lagrangian bifunction for EP(f,K). For this purpose, we first define the real valued function $s_i : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_{++} \to \mathbb{R}$ $(1 \leq i \leq m)$ as

$$s_i(x,y,\lambda,\gamma) = \frac{\gamma}{2}\left[\left(\max\left\{0, \lambda_i + \frac{h_i(y)}{\gamma}\right\}\right)^2 - \left(\max\left\{0, \lambda_i + \frac{h_i(x)}{\gamma}\right\}\right)^2\right], \qquad (3.3)$$

and then we propose $\widetilde{\mathcal{L}} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}$ as the proximal augmented Lagrangian bifunction for EP(f, K), defined as,

$$\widetilde{\mathcal{L}}(x,y,\lambda,z,\gamma) = f(x,y) + \gamma\langle x - z, y - x\rangle + \gamma \sum_{i=1}^{m} s_i(x,y,\lambda,\gamma). \qquad (3.4)$$

Now we present an exact augmented Lagrangian method for EP(f, K). Take an exogenous bounded sequence $\{\gamma_j\}_{j=0}^{\infty} \subset \mathbb{R}_{++}$. 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 equilibrium problem EP$(\tilde{\mathcal{L}}_j, \mathbb{R}^n)$ with $\tilde{\mathcal{L}}_j$ given by

$$\widetilde{\mathcal{L}}_j(x,y) = \widetilde{\mathcal{L}}(x,y,\lambda^j,x^j,\gamma_j) = f(x,y) + \gamma_j\langle x - x^j, y - x\rangle + \sum_{i=1}^{m} s_i(x,y,\lambda^j,\gamma_j). \qquad (3.5)$$

Then, the dual variables are updated as

$$\lambda_i^{j+1} = \max\left\{0, \lambda_i^j + \frac{h_i(x^{j+1})}{\gamma_j}\right\} \quad (1 \le i \le m). \qquad (3.6)$$

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

Algorithm IALE: Inexact Augmented Lagrangian-Extragradient Method for EP(f, K)

1. Take an exogenous bounded sequence $\{\gamma_j\}_{j=0}^{\infty} \subset \mathbb{R}_{++}$ 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 $(x^j, \lambda^j) \in \mathbb{R}^n \times \mathbb{R}^m_+$, find a pair $(\tilde{x}^j, e^j) \in \mathbb{R}^n \times \mathbb{R}^n$ such that $\tilde{x}^j$ solves EP$(\widetilde{\mathcal{L}}^e_j, \mathbb{R}^n)$, where $\widetilde{\mathcal{L}}^e_j$ is defined as

$$\widetilde{\mathcal{L}}^e_j(x,y) = f(x,y) + \gamma_j\langle x - x^j, y - x\rangle + \sum_{i=1}^{m} s_i(x,y,\lambda^j,\gamma_j) - \langle e^j, y - x\rangle, \qquad (3.7)$$

with s_i as given by (3.3), and e^j satisfies

$$\left\|e^j\right\| \le \sigma\gamma_j \left\|\tilde{x}^j - x^j\right\|. \qquad (3.8)$$

3. Define λ^{j+1} as

$$\lambda_i^{j+1} = \max\left\{0, \lambda_i^j + \frac{h_i(\tilde{x}^j)}{\gamma_j}\right\} \quad (1 \le i \le m). \tag{3.9}$$

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

$$x^{j+1} = \tilde{x}^j - \frac{1}{\gamma_j} e^j. \tag{3.10}$$

We mention that the exact augmented Lagrangian method defined by (3.5)–(3.6), can be realized as a particular instance of Algorithm IALE by taking $e^j = 0$ for $j = 0, 1, \ldots$. Thus, the convergence analysis for the inexact methods, to be presented next, holds also for the exact method.

3.2 Convergence Analysis of Algorithm IALE

We will use Algorithm IPPE, introduced in Section 2.2 for solving EP(f, K), as an auxiliary tool in the convergence analysis of Algorithm IALE. So, we start this section by reformulating Algorithm IPPE in finite dimensional spaces, with $g(x) = \frac{1}{2}\|x\|^2$ as regularization function.

Algorithm A: Inexact Proximal Point-Extragradient Method for EP(f, K) in $\mathbb{R}^n$

1. Consider an exogenuous bounded sequence of regularization parameters $\{\gamma_j\}_{j=0}^{\infty} \subset \mathbb{R}_{++}$ and a relative error tolerance $\sigma \in [0, 1)$. Initialize the algorithm with $x^0 \in \mathbb{R}^n$.

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

$$f_j^e(x, y) = f(x, y) + \gamma_j \langle x - x^j, y - x\rangle - \langle e^j, y - x\rangle, \tag{3.11}$$

and

$$\left\|e^j\right\| \le \sigma\gamma_j \left\|\hat{x}^j - x^j\right\|. \tag{3.12}$$

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

$$x^{j+1} = \hat{x}^j - \gamma_j^{-1} e^j. \tag{3.13}$$

The convergence result for Algorithm A, proved in Theorem 2.3.8, reduces to the following one.

Theorem 3.2.1. *Consider* EP(f,K) *satisfying* P1–P4 *and* P4". *Take an exogenous sequence* $\{\gamma_j\}_{j=0}^{\infty} \subset (\theta, \bar{\gamma}]$ *for some* $\bar{\gamma} > \theta$*, where* θ *is the undermonotonicity constant in* P4*, and a relative error tolerance* $\sigma \in [0,1)$*. Let* $\{x^j\}_{j=0}^{\infty}$ *be the sequence generated by Algorithm* A*. If* EP(f,K) *has solutions, then* $\{x^j\}_{j=0}^{\infty}$ *converges to some solution* x^* *of* EP(f,K)*.*

Proof. The validity of this theorem certainly follows from Theorem 2.3.8, because the technical hypotheses H1–H6 required for auxiliary function $g(x) = \frac{1}{2}\|x\|^2$ and weakly upper semicontinuity of function $f(\cdot, y)$ hold automatically in the finite dimensional case, which is the one of interest here. ∎

We will apply Algorithm A for solving equilibrium problem EP$(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$, with $\mathcal{L}$ as in (3.2), for which we must check that this equilibrium problem satisfies P1–P4 and P4".

Proposition 3.2.2. *Assume that* f *satisfies* P1–P4 *on* $\mathbb{R}^n \times \mathbb{R}^n$*, and that* K *is given by* (3.1)*. Then* $\mathcal{L}$*, as defined in* (3.2)*, satisfies* P1–P4 *on* $(\mathbb{R}^n \times \mathbb{R}^m_+) \times (\mathbb{R}^n \times \mathbb{R}^m_+)$*. Additionally, if* f *satisfies* P4" *on* $\mathbb{R}^n \times \mathbb{R}^n$*, so does* (3.2) *on* $(\mathbb{R}^n \times \mathbb{R}^m_+) \times (\mathbb{R}^n \times \mathbb{R}^m_+)$*.*

Proof. It follows easily from (3.2) that EP$(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$ inherits P1–P3 from EP(f,K). Furthermore, (3.2) 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 $x^1, \dots, x^q \in \mathbb{R}^n$, $\lambda^1, \dots, \lambda^q \in \mathbb{R}^m_+$ and $t_1, \dots, t_q \ge 0$ such that $\sum_{\ell=1}^q t_\ell = 1$. Then

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

$$f\left(x^\ell, \sum_{k=1}^q t_k x^k\right) + \sum_{i=1}^m \lambda_i^\ell h_i\left(\sum_{k=1}^q t_k x^k\right) - \sum_{i=1}^m \sum_{k=1}^q t_k \lambda_i^k h_i(x^\ell) \le \tag{3.14}$$

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

using the convexity of the h_i's in the inequality. Multiplying the leftmost and rightmost expressions of (3.14) by t_ℓ and then summing with $1 \le \ell \le q$, we get

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

$$\sum_{\ell=1}^{q} t_\ell f\left(x^\ell, \sum_{k=1}^{q} t_k x^k\right) + \sum_{\ell=1}^{q}\sum_{i=1}^{m}\sum_{k=1}^{q} t_\ell t_k \lambda_i^\ell h_i(x^k) - \sum_{\ell=1}^{q}\sum_{i=1}^{m}\sum_{k=1}^{q} t_\ell t_k \lambda_i^k h_i(x^\ell). \quad (3.15)$$

The first term in the right hand site of (3.15) is nonpositive 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 A for solving EP($\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+$). In view of (3.11), the regularized function at iteration j is given by

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

$$f(x,y) + \sum_{i=1}^{m}\lambda_i h_i(y) - \sum_{i=1}^{m}\mu_i h_i(x) + \gamma_j\langle x-x^j, y-x\rangle + \gamma_j\langle \lambda-\lambda^j, \mu-\lambda\rangle - \langle e^j, y-x\rangle, \quad (3.16)$$

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 equilibrium problem EP($\widehat{\mathcal{L}}_j^e, \mathbb{R}^n \times \mathbb{R}^m_+$) with $\widehat{\mathcal{L}}_j^e$ as defined in (3.16), and the iterative formulae (3.12)–(3.13) take the form:

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

$$x^{j+1} = \hat{x}^j - \gamma_j^{-1} e^j, \quad (3.17)$$

$$\lambda^{j+1} = \hat{\lambda}^j. \quad (3.18)$$

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 IALE the λ_i^j's are updated through the closed formula (3.9), 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 Algorithm A (i.e., (3.17)–(3.18)), as well as the fact that (3.16) can be considered as a regularization for EP($\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+$).

Corollary 3.2.3. *Consider f satisfying* P1–P4. *Fix* $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m$, $(e, 0) \in \mathbb{R}^n \times \mathbb{R}^m$ *and that* $\gamma > \theta$, *where* θ *is the undermonotonicity constant in* P4. *In this situation,* $\mathrm{EP}(\widehat{\mathcal{L}}^e, \mathbb{R}^n \times \mathbb{R}^m_+)$ *has a unique solution, with* $\widehat{\mathcal{L}}^e$ *defined as*

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

$$f(x,y) + \sum_{i=1}^{m} \lambda_i h_i(y) - \sum_{i=1}^{m} \mu_i h_i(x) + \gamma\langle x-\bar{x}, y-x\rangle + \gamma\langle \lambda-\bar{\lambda}, \mu-\lambda\rangle - \langle e, y-x\rangle.$$

Proof. The result follows from Proposition 2.1.1 and Proposition 3.2.2. ■

Corollary 3.2.4. *Consider* $\mathrm{EP}(f, K)$ *with K given by* (3.1) *and f satisfying* P1–P4 *and* P4" *on* $\mathbb{R}^n \times \mathbb{R}^n$. *Take* $\{\gamma_j\}_{j=0}^{\infty} \subset (\theta, \bar{\gamma}]$ *for some* $\bar{\gamma} > \theta$, *where θ is the undermonotonicity constant of f, and that* $\sigma \in [0, 1)$. *Let* $\{(x^j, \lambda^j)\}_{j=0}^{\infty}$ *be the sequence generated by Algorithm* A *applied to* $\mathrm{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$ *(i.e., the sequence generated by* (3.17)–(3.18)*). If the problem* $\mathrm{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$ *has solutions, then* $\{(x^j, \lambda^j)\}_{j=0}^{\infty}$ *converges to some pair* $(x^*, \lambda^*) \in S_E(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$.

Proof. It follows from Theorem 3.2.1 and Proposition 3.2.2. ■

For the sake of simplicity, we recall statement (2.6) here, for $x \in \mathbb{R}^n$, we consider $F_x : \mathbb{R}^n \to \mathbb{R}$ defined as

$$F_x(y) = f(x, y). \tag{3.19}$$

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

Proposition 3.2.5. *Consider* $\mathrm{EP}(f, K)$. *The following two statements are equivalent.*

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

ii) x^* *minimizes* F_{x^*} *over* K, *with* F_{x^*} *as in* (3.19).

Proof. Assume that $x^* \in S_E(f, K)$. By (3.19) 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 (3.19), 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* $\mathrm{EP}(f, K)$.

Definition 3.2.6. *Assume that the set K is defined as (3.1). We say $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m$ is an optimal pair for $\mathrm{EP}(f, K)$ if*

$$0 \in \partial F_{x^*}(x^*) + \sum_{i=1}^{m} \lambda_i^* \partial h_i(x^*), \tag{3.20}$$

$$\lambda_i^* \geq 0 \ (1 \leq i \leq m), \tag{3.21}$$

$$h_i(x^*) \leq 0 \ (1 \leq i \leq m), \tag{3.22}$$

$$\lambda_i^* h_i(x^*) = 0 \ (1 \leq i \leq m). \tag{3.23}$$

The sets $\partial F_{x^*}(x^*)$ and $\partial h_i(x^*)$ denote the subdifferentials of the convex functions F_{x^*} and h_i, respectively, at the point x^*. We mention that, since we are assuming that both f and the h_i's are finite on the whole $\mathbb{R}^n$, there is no difficulty with the nonsmooth Lagrangian condition (3.20). In other words, Proposition 1.4.3 guarantees that both sets $\partial F_{x^*}(x^*)$ and $\partial h_i(x^*)$ are nonempty.

Note that (3.20)–(3.23) 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 $\mathrm{EP}(f, K)$; 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 $\mathrm{EP}(f, K)$.

The next two propositions and corollary establish the relations between solutions of $\mathrm{EP}(f, K)$, solutions of $\mathrm{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$ and optimal pairs for $\mathrm{EP}(f, K)$. We mention that the next proposition does not require a constraint qualification for the feasible set K, while Proposition 3.2.8 does.

Proposition 3.2.7. *Consider* $\mathrm{EP}(f, K)$ *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* $\mathrm{EP}(f, K)$.

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

Proof.

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

$$\min \mathcal{F}_{(x^*,\lambda^*)}(x,\lambda) \tag{3.24}$$

$$\text{s.t. } (x,\lambda) \in \mathbb{R}^n \times \mathbb{R}^m_+. \tag{3.25}$$

Since EP($\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+$) satisfies P1–P3 by Proposition 3.2.2, we conclude from Proposition 3.2.5 that the pair (x^*, λ^*) solves (3.24)–(3.25). Since the constraints of this problem are affine, the Slater's CQ of Definition 1.7.3 holds for this problem and, invoking Theorem 1.7.4, there exists a vector of KKT multipliers $\eta^* \in \mathbb{R}^m$ such that

$$0 \in \partial F_{x^*}(x^*) + \sum_{i=1}^{m} \lambda_i^* \partial h_i(x^*), \tag{3.26}$$

$$h_i(x^*) + \eta_i^* = 0 \ \ (1 \leq i \leq m), \tag{3.27}$$

$$\lambda^* \geq 0, \tag{3.28}$$

$$\eta^* \geq 0, \tag{3.29}$$

$$\lambda_i^* \eta_i^* = 0 \ \ (1 \leq i \leq m). \tag{3.30}$$

Note that (3.26) and (3.28) coincide with (3.20) and (3.21) respectively. Since $\eta_i = -h_i(x^*)$ by (3.27), we get (3.22) and (3.23) from (3.29) and (3.30) respectively.

(i)⇒ (ii) Now we assume that the pair (x^*, λ^*) satisfies (3.20)–(3.23). Taking $\eta_i^* = -h_i(x^*)$, we get (3.26)–(3.30). Since problem (3.24)–(3.25) is convex, according to Theorem 1.7.4, the KKT conditions are sufficient for optimality, so that the pair (x^*, λ^*) solves this problem. In view of Proposition 3.2.5, this pair solves EP($\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+$).

■

Proposition 3.2.8. *Consider* 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* (3.19)*, and the Slater's CQ of Definition 1.7.3 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* EP(f, K). *Conversely, if (x^*, λ^*) is an optimal pair for* EP(f, K) *then x^* minimizes F_{x^*} over K, with F_{x^*} as in* (3.19).

Proof. For the first statement, since CQ holds, we invoke again Theorem 1.7.4 to conclude that there exists a vector $\lambda^* \in \mathbb{R}^m$ such that (3.20)–(3.23) hold. It follows from Definition 3.2.6 that (x^*, λ^*) is an optimal pair for EP(f, K). Reciprocally, if (x^*, λ^*) is an optimal pair for EP(f, K), then (3.20)–(3.23) 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 by Theorem 1.7.4, and hence x^* solves this problem.
■

Corollary 3.2.9. *Consider* EP(f, K) *and assume that* f *satisfies* P1–P3 *on* $\mathbb{R}^n \times \mathbb{R}^n$. *If* $(x^*, \lambda^*) \in S_E(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$, *then* $x^* \in S_E(f, K)$. *Conversely, if* $x^* \in S_E(f, K)$ *and the Slater's* CQ *of Definition 1.7.3 holds, then there exists* $\lambda^* \in \mathbb{R}^m_+$ *such that* $(x^*, \lambda^*) \in S_E(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$.

Proof. It follows from Proposition 3.2.5, Proposition 3.2.7 and Proposition 3.2.8. ■

Corollary 3.2.9 shows that solving EP$(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$ is enough for solving EP(f, K). Next we prove the equivalence between Algorithm IALE and Algorithm A. In fact, we will prove that the sequence generated by Algorithm IALE for solving the latter problem coincides with the sequence generated by Algorithm A for solving the former, using the technical result proved in Proposition 2.1.1.

Theorem 3.2.10. *Assume that* EP(f, K) *satisfies* P1–P4 *on* $\mathbb{R}^n \times \mathbb{R}^n$. *Fix a sequence* $\{\gamma_j\}_{j=0}^{\infty} \subset \mathbb{R}_{++}$ *and a relative error tolerance* $\sigma \in [0, 1)$. *Let* $\{(x^j, \lambda^j)\}_{j=0}^{\infty}$ *be the sequence generated by Algorithm* IALE *applied to* EP(f, K), *with associated error vector* $e^j \in \mathbb{R}^n$, *and* $\{(\bar{x}^j, \bar{\lambda}^j)\}_{j=0}^{\infty}$ *the sequence generated by Algorithm* A *applied to* 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* $\{\gamma_j\}_{j=0}^{\infty}$ *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 A, we must solve EP$(\widehat{\mathcal{L}}^e_j, \mathbb{R}^n \times \mathbb{R}^m_+)$, with $\widehat{\mathcal{L}}^e_j$ as in (3.16), which has a unique solution by Corollary 3.2.3. Let $(\hat{x}^j, \hat{\lambda}^j)$ be the solution of this problem. By Proposition 3.2.5, $(\hat{x}^j, \hat{\lambda}^j)$ solves the convex minimization problem defined as

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

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

with $\widehat{\mathcal{F}}_{(\hat{x}^j, \hat{\lambda}^j)}(x, \lambda) = \widehat{\mathcal{L}}^e_j((\hat{x}^j, \hat{\lambda}^j), (x, \lambda))$. The constraints of this problem are affine, so that the Slater's CQ holds and therefore, taking into account Theorem 1.7.4, there exists a KKT vector $\eta^j \in \mathbb{R}^m$ such that

$$\gamma_j[\bar{x}^j - \hat{x}^j] + e^j \in \partial F_{\hat{x}^j}(\hat{x}^j) + \sum_{i=1}^{m} \hat{\lambda}^j_i \partial h_i(\hat{x}^j), \tag{3.31}$$

$$-h_i(\hat{x}^j) + \gamma_j[\hat{\lambda}^j_i - \bar{\lambda}^j_i] = \eta^j_i \ \ (1 \le i \le m), \tag{3.32}$$

$$\hat{\lambda}^j \ge 0, \tag{3.33}$$

$$\eta^j \geq 0, \tag{3.34}$$

$$\hat{\lambda}_i^j \eta_i^j = 0 \ \ (1 \leq i \leq m). \tag{3.35}$$

Using (3.32) to eliminate η^j, (3.31)–(3.35) can be rewritten, after some elementary calculations, as

$$\gamma_j[\bar{x}^j - \hat{x}^j] + e^j \in \partial F_{\hat{x}^j}(\hat{x}^j) + \sum_{i=1}^{m} \hat{\lambda}_i^j \partial h_i(\hat{x}^j), \tag{3.36}$$

$$\hat{\lambda}_i^j = \max\left\{0, \bar{\lambda}_i^j + \frac{h_i(\hat{x}^j)}{\gamma_j}\right\} \ \ (1 \leq i \leq m). \tag{3.37}$$

Replacing (3.37) in (3.36) we get

$$\gamma_j[\bar{x}^j - \hat{x}^j] + e^j \in \partial F_{\hat{x}^j}(\hat{x}^j) + \sum_{i=1}^{m} \max\left\{0, \bar{\lambda}_i^j + \frac{h_i(\hat{x}^j)}{\gamma_j}\right\} \partial h_i(\hat{x}^j) \ \ (1 \leq i \leq m). \tag{3.38}$$

Now we look at Step 2 of Algorithm IALE, which demands the solution $\tilde{x}^j$ of $\mathrm{EP}(\widetilde{\mathcal{L}}_j^e, \mathbb{R}^n)$. Applying now Proposition 3.2.5 to this problem and taking into account Theorem 1.7.2, we obtain that $\tilde{x}^k$ belongs to $S_E(\widetilde{\mathcal{L}}_j^e, \mathbb{R}^n)$ if and only if

$$\gamma_j[x^j - \tilde{x}^j] + e^j \in \partial F_{\tilde{x}^j}(\tilde{x}^j) + \sum_{i=1}^{m} \max\left\{0, \lambda_i^j + \frac{h_i(\tilde{x}^j)}{\gamma_j}\right\} \partial h_i(\tilde{x}^j). \tag{3.39}$$

Since $x^j = \bar{x}^j$, $\lambda^j = \bar{\lambda}^j$ by inductive hypothesis, we get from (3.38) that (3.39) holds with $\hat{x}^j$ substituting for $\tilde{x}^j$, and hence $\hat{x}^j$ also solves $\mathrm{EP}(\widetilde{\mathcal{L}}_j^e, \mathbb{R}^n)$. Since this problem has a unique solution by Corollary 3.2.3, we conclude that

$$\hat{x}^j = \tilde{x}^j. \tag{3.40}$$

Taking now into account on the one hand (3.10) in Step 4 of Algorithm IALE, and on the other hand (3.17) in Step 3 of Algorithm A we conclude, using again the inductive hypothesis and (3.40), that $x^{j+1} = \bar{x}^{j+1}$. Now we look at the updating of the dual variables. In view of (3.18) and (3.37), for Algorithm A we have

$$\bar{\lambda}_i^{j+1} = \hat{\lambda}_i^j = \max\left\{0, \bar{\lambda}_i^j + \frac{h_i(\hat{x}^j)}{\gamma_j}\right\}. \tag{3.41}$$

Comparing now (3.41) with (3.9) and taking into account (3.40) 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 IALE.

Proposition 3.2.11. *Suppose that Algorithm* IALE *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 IALE stops at the j-th iteration, then, in view of Step 4, $(x^j, \lambda^j) = (\tilde{x}^j, \lambda^{j+1})$. Using (3.8) 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

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

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

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

By Proposition 3.2.5, $\tilde{x}^j$ is an unconstrained minimizer of $\breve{F}_{\tilde{x}^j}$. Thus, in view of (3.9) and Theorem 1.7.2,

$$0 \in \partial\breve{F}_{\tilde{x}^j}(\tilde{x}^j) = \partial F_{\tilde{x}^j}(\tilde{x}^j) + \sum_{i=1}^{m} \max\left\{0, \lambda_i^j + \frac{h_i(\tilde{x}^j)}{\gamma_j}\right\} \partial h_i(\tilde{x}^j) = \partial F_{\tilde{x}^j}(\tilde{x}^j) + \sum_{i=1}^{m} \lambda_i^j \partial h_i(\tilde{x}^j), \tag{3.42}$$

with $F_{\tilde{x}^j}$ as in (3.19), using (3.9) and the fact that $\lambda^j = \lambda^{j+1}$, which also gives

$$\lambda_i^{j+1} = \lambda_i^j = \max\left\{0, \lambda_i^j + \frac{h_i(\tilde{x}^j)}{\gamma_j}\right\} \quad (1 \le i \le m). \tag{3.43}$$

It follows easily from (3.43) that

$$\lambda_i^j \ge 0, \quad \lambda_i^j h_i(\tilde{x}^j) = 0, \qquad h_i(\tilde{x}^j) \le 0 \qquad (1 \le i \le m). \tag{3.44}$$

In view of (3.42) and (3.44), $(\tilde{x}^j, \lambda^j)$ is an optimal pair for $\mathrm{EP}(f, K)$ and we conclude from Proposition 3.2.7 that $(\tilde{x}^j, \lambda^j) \in S_E(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$, which in turns implies, by Corollary 3.2.9, $\tilde{x}^j \in S_E(f, K)$. ∎

Now we use Theorem 3.2.10 for completing the convergence analysis of Algorithm IALE.

Theorem 3.2.12. *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* (3.1),

iii) the Slater's CQ *stated in Definition 1.7.3 holds for* K,

iv) $\{\gamma_j\}_{j=0}^{\infty} \subset (\theta, \bar{\gamma}]$ *for some* $\bar{\gamma} > \theta$*, where* θ *is the undermonotonicity constant of* f *in* P4*,*

v) $\sigma \in [0, 1)$*,*

vi) $\mathrm{EP}(f, K)$ *has solutions.*

Let $\{(x^j, \lambda^j)\}_{j=0}^{\infty}$ *be the sequence generated by Algorithm* IALE *for solving* $\mathrm{EP}(f, K)$*. In this situation, the sequence* $\{(x^j, \lambda^j)\}_{j=0}^{\infty}$ *converges to some optimal pair* (x^*, λ^*) *for* $\mathrm{EP}(f, K)$*, and consequently* $x^* \in S_E(f, K)$*.*

Proof. By Theorem 3.2.10 the sequence $\{(x^j, \lambda^j)\}_{j=0}^{\infty}$ coincides with the sequence generated by Algorithm IPPE applied to $\mathrm{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$. Since $\mathrm{EP}(f, K)$ has solutions and the Slater's CQ holds, Corollary 3.2.9 implies that $\mathrm{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$ has solutions. By Corollary 3.2.4, the sequence $\{(x^j, \lambda^j)\}_{j=0}^{\infty}$ converges to a solution (x^*, λ^*) of $\mathrm{EP}(\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+)$. By Proposition 3.2.7, (x^*, λ^*) is an optimal pair for $\mathrm{EP}(f, K)$. By Corollary 3.2.9 again, x^* belongs to $S_E(f, K)$. ∎

We comment now on the real meaning of the error vector e^j appearing in Algorithm IALE and Algorithm A. 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 (3.8). 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 (3.39) we have

$$e^j = F'_{\tilde{x}^{j,k}}(\tilde{x}^{j,k}) + \sum_{i=1}^{m} \max\left\{0, \lambda_i^j + \frac{h_i(\tilde{x}^{j,k})}{\gamma_j}\right\} h_i'(\tilde{x}^{j,k}) + \gamma_j[\tilde{x}^{j,k} - x^j]. \tag{3.45}$$

If $\tilde{x}^{j,k}$ were the exact solution of the j-th subproblem, then the right hand side of (3.45) would vanish. If $\tilde{x}^{j,k}$ is just an approximation of this solution, then the right hand side of (3.45) is nonzero, and we call it e^j. Then we perform the test in Step 2 of the algorithm. If e^j satisfies the inequality in (3.8), 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 (3.7)–(3.8), 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 EP(f, K) and assume that f is continuously differentiable. We look at Algorithm A as described in (3.11)–(3.13). 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 (3.11) and $e^j = 0$. We have proved in Theorem 2.4.9 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 (3.11) and (3.12) are satisfied with x, e substituting for $\tilde{x}^j, e^j$ respectively.

Observe now that the j-th subproblem of Algorithm IALE, namely EP$(\widetilde{\mathcal{L}}_j^e, \mathbb{R}^n)$, is unconstrained, i.e. $K = \mathbb{R}^n$, so that the condition $\breve{x}^j \in int(K)$ is automatically satisfied. Regarding the continuous differentiability of $\widetilde{\mathcal{L}}_j^e$, it follows from (3.3) and (3.7) 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 given by (3.3)). Thus Theorem 2.4.9 can be rephrased for the case of Algorithm IALE as follows.

Corollary 3.2.13. *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* $\{(x^j, \lambda^j)\}_{j=0}^{\infty}$ *be the sequence generated by Algorithm* IALE. *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* (3.7), *with $e^j = 0$. Then there exists $\delta_j > 0$ such that any $x \in B(\breve{x}^j, \delta_j)$ solves the subproblem* (3.7)–(3.8).

In view of Corollary 3.2.13, if the subproblems of Algorithm IALE 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 Algorithm IALE.

3.3 Linearized Augmented Lagrangian

An interesting feature of Algorithm IALE is that its convergence properties are not altered if the Lagrangian function

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

which has been already defined in (3.2), is replaced by its first order approximation as a function of the second argument. This linearization gives rise to a variant of Algorithm IALE 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 again use extensively the notation $F_x(y) = f(x,y)$, and in particular the gradient of F_x, denoted as $F'_x : \mathbb{R}^n \to \mathbb{R}^n$.

If we linearize the Lagrangian given by (3.2) as a function of y around $y = x$, we obtain the function $\bar{\mathcal{L}} : (\mathbb{R}^n \times \mathbb{R}^m) \times (\mathbb{R}^n \times \mathbb{R}^m) \to \mathbb{R}$ defined as

$$\bar{\mathcal{L}}((x,\lambda),(y,\mu)) = \langle F'_x(x), y - x\rangle + \sum_{i=1}^m \lambda_i \langle h'_i(x), y - x\rangle + \sum_{i=1}^m (\lambda_i - \mu_i) h_i(x). \quad (3.46)$$

We will denote $\bar{\mathcal{L}}$ as the *Linearized Lagrangian* function 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 function, given by (3.7), i.e.,

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

where

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

which has been already defined in (3.3), we obtain a kind of Algorithm IALE, to be called *Linearized Inexact Augmented Lagrangian-Extragradient Method* (it will be named Algorithm LIALE from now on), which we describe next.

Algorithm LIALE: Linearized Inexact Augmented Lagrangian-Extragradient Method for EP(f, K)

1. Take an exogenous bounded sequence $\{\gamma_j\}_{j=0}^{\infty} \subset \mathbb{R}_{++}$ 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 $(x^j, \lambda^j) \in \mathbb{R}^n \times \mathbb{R}^m_+$, define $\bar{s}_i : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}_{++} \to \mathbb{R}$ as

$$\bar{s}_i(x,y,\lambda,\gamma) = \max\left\{0, \lambda_i + \frac{h_i(x)}{\gamma}\right\} \langle h'_i(x), y - x\rangle \ \ (1 \le i \le m), \quad (3.47)$$

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

$$\bar{\mathcal{L}}^e_j(x,y) = \langle F'_x(x), y - x\rangle + \gamma_j \langle x - x^j, y - x\rangle + \gamma_j \sum_{i=1}^m \bar{s}_i(x, y, \lambda^j, \gamma_j) - \langle e^j, y - x\rangle, \ (3.48)$$

with $\bar{s}_i$ as in (3.47), and e^j satisfies

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

3. Define λ^{j+1} as

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

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

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

Observe that the only difference between Algorithm IALE and Algorithm LIALE appears in the bifunction defining the unconstrained equilibrium subproblems $\mathrm{EP}(\widetilde{\mathcal{L}}_j^e, \mathbb{R}^n)$ and $\mathrm{EP}(\bar{\mathcal{L}}_j^e, \mathbb{R}^n)$. In fact, in iteration j of Algorithm LIALE one solves $\mathrm{EP}(\bar{\mathcal{L}}_j^e, \mathbb{R}^n)$ with $\bar{\mathcal{L}}_j^e$ as in (3.48), while in the j-th iteration of Algorithm IALE one solves $\mathrm{EP}(\widetilde{\mathcal{L}}_j^e, \mathbb{R}^n)$ with $\widetilde{\mathcal{L}}_j^e$ as in (3.7).

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 3.2.1, the sequence generated by Algorithm A 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 3.3.1. *Consider* $\mathrm{EP}(f, K)$*. Assume that f satisfies* P1–P4 *and* P4" *on* $\mathbb{R}^n \times \mathbb{R}^n$*, and that both f and all the h_i's are continuously differentiable 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* (3.46).

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. To demonstrate that $\mathrm{EP}(\bar{\mathcal{L}}, \mathbb{R}^n \times \mathbb{R}_+^m)$ satisfies P4 and P4", we will invoke several times the fact that for a differentiable convex function $g : \mathbb{R}^n \to \mathbb{R}$, it holds that

$$g(x) + \langle g'(x), y - x\rangle \leq g(y) \quad \forall x, y \in \mathbb{R}^n,$$

which is a consequence of Definition 1.4.1(iv) and Theorem 1.4.10(i).

Now we start proving that P4 holds. Using (3.46), we get

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

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

$$\leq f(x,y) - f(x,x) + f(y,x) - f(y,y) = f(x,y) + f(y,x) \leq \theta \left\|x - y\right\|^2,$$

using (3.46) 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 $t_1, \dots, t_q \geq 0$ such that $\sum_{\ell=1}^q t_\ell = 1$. Then

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

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

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

$$f\left(x^\ell, \sum_{k=1}^q t_k x^k\right) + \sum_{i=1}^m \sum_{k=1}^q t_k \lambda_i^\ell h_i(x^k) - \sum_{i=1}^m \sum_{k=1}^q t_k \lambda_i^k h_i(x^\ell), \tag{3.49}$$

using convexity of F_{x^ℓ} 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 (3.49) using the same argument as in the proof of Proposition 3.2.2 after (3.14). ■

It is easy to check that Proposition 3.2.7, Proposition 3.2.8 and Corollary 3.2.9 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 (3.20) takes the form

$$0 = F'_{x^*}(x^*) + \sum_{i=1}^m \lambda_i^* h'_i(x^*).$$

It is a matter of routine to check that the proofs of Theorem 3.2.10, Proposition 3.2.11, Theorem 3.2.12 and Corollary 3.2.13 also remain valid for LIALE, resulting in the following convergence theorem.

Theorem 3.3.2. *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* (3.1),

iii) the Slater's CQ *of Definition 1.7.3 holds for the feasible set* K,

iv) $\{\gamma_j\}_{j=0}^{\infty} \subset (\theta, \bar{\gamma}]$ *for some* $\bar{\gamma} > \theta$*, where* θ *is the undermonotonicity constant of* f *in* P4*,*

v) $\sigma \in [0, 1)$*,*

vi) $\mathrm{EP}(f, K)$ *has solutions,*

vii) f *is continuously differentiable,*

viii) h_i *is differentiable* $(1 \leq i \leq m)$*.*

Let $\{(x^j, \lambda^j)\}_{j=0}^{\infty}$ *be the sequence generated by Algorithm* LIALE *applied to* $\mathrm{EP}(f, K)$*. In this situation, the sequence* $\{(x^j, \lambda^j)\}_{j=0}^{\infty}$ *converges to an optimal pair* (x^*, λ^*) *for* $\mathrm{EP}(f, K)$*, so that* x^* *belongs to* $S_E(f, K)$*. 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* LIALE*.*

3.4 Variants of Augmented Lagrangian Methods for Equilibrium Problems with Projection approach

In this section, we develop another two augmented Lagrangian methods from Algorithm IPPBP presented in Section 2.2, to be called *Inexact Augmented Lagrangian Projection Method* (Algorithm IALP from now on) and *Linearized Inexact Augmented Lagrangian Projection Method* (Algorithm LIALP from now on), for solving $\mathrm{EP}(f, K)$. The convergence analysis of Algorithm IALP and Algorithm LIALP invokes the convergence results for Algorithm IPPBP. In this case, instead of Step 4 of Algorithm IALE (or Algorithm LIALE), the solution $\hat{x}^j$ of the subproblem is used for constructing a hyperplane H_j which separates x^j from set $S_E(f, K)$, 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, because we use $g(x) = \frac{1}{2}\|x\|^2$ as the auxiliary function, as we explained in Section 1.6.

Next we propose Algorithm IALP and Algorithm LIALP for solving $\mathrm{EP}(f, K)$.

Algorithm IALP: Inexact Augmented Lagrangian+Projection Method for EP(f, K)

1. Take an exogenous bounded sequence $\{\gamma_j\}_{j=0}^{\infty} \subset \mathbb{R}_{++}$ 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 $(x^j, \lambda^j) \in \mathbb{R}^n \times \mathbb{R}_+^m$, find a pair $(\tilde{x}^j, e^j) \in \mathbb{R}^n \times \mathbb{R}^n$ such that $\tilde{x}^j$ solves $\mathrm{EP}(\widetilde{\mathcal{L}}_j^e, \mathbb{R}^n)$, where $\widetilde{\mathcal{L}}_j^e$ is defined as

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

with s_i as given by

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

and e^j satisfies

$$\left\|e^j\right\| \le \frac{1}{2}\sigma\gamma_j \begin{cases} \left\|(\tilde{x}^j - x^j, \tilde{\lambda}^j - \lambda^j)\right\|^2 & \text{if } \left\|(\tilde{x}^j - x^j, \tilde{\lambda}^j - \lambda^j)\right\| < 1 \\ 1 & \text{if } \left\|(\tilde{x}^j - x^j, \tilde{\lambda}^j - \lambda^j)\right\| \ge 1, \end{cases}$$

where

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

3. Define v^j as

$$v^j = \gamma_j(x^j - \tilde{x}^j, \lambda^j - \tilde{\lambda}^j) + (e^j, 0).$$

If $v^j = 0$ or $(x^j,\lambda^j) = (\tilde{x}^j, \tilde{\lambda}^j)$, then stop. Otherwise, take

$$(x^{j+1},\lambda^{j+1}) = \underset{(x,\lambda)\in H_j}{\operatorname{argmin}} \left\|(x - x^j, \lambda - \lambda^j)\right\|^2 = (x^j,\lambda^j) - \frac{\langle v^j, (x^j - \tilde{x}^j, \lambda^j - \tilde{\lambda}^j)\rangle}{\left\|v^j\right\|^2} v^j,$$

where

$$H_j = \{(x,\lambda) \in \mathbb{R}^n \times \mathbb{R}^m : \langle v^j, (x - \tilde{x}^j, \lambda - \tilde{\lambda}^j)\rangle = 0\}.$$

Algorithm LIALP: Linearized Inexact Augmented Lagrangian+Projection Method for EP(f, K)

1. Take an exogenous bounded sequence $\{\gamma_j\}_{j=0}^{\infty} \subset \mathbb{R}_{++}$ 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 $(x^j,\lambda^j) \in \mathbb{R}^n \times \mathbb{R}^m_+$, define $\bar{s}_i : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}_{++} \to \mathbb{R}$ as

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

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 $\bar{\mathcal{L}}_j^e : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is defined as

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

and e^j satisfies

$$\left\|e^j\right\| \le \frac{1}{2}\sigma\gamma_j \begin{cases} \left\|(\tilde{x}^j - x^j, \tilde{\lambda}^j - \lambda^j)\right\|^2 & \text{if } \left\|(\tilde{x}^j - x^j, \tilde{\lambda}^j - \lambda^j)\right\| < 1 \\ 1 & \text{if } \left\|(\tilde{x}^j - x^j, \tilde{\lambda}^j - \lambda^j)\right\| \ge 1, \end{cases}$$

where

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

3. Define v^j as

$$v^j = \gamma_j(x^j - \tilde{x}^j, \lambda^j - \tilde{\lambda}^j) + (e^j, 0).$$

If $v^j = 0$ or $(x^j, \lambda^j) = (\tilde{x}^j, \tilde{\lambda}^j)$, then stop. Otherwise, take

$$(x^{j+1}, \lambda^{j+1}) = \operatorname*{argmin}_{(x,\lambda)\in H_j} \left\|(x - x^j, \lambda - \lambda^j)\right\|^2 = (x^j, \lambda^j) - \frac{\langle v^j, (x^j - \tilde{x}^j, \lambda^j - \tilde{\lambda}^j)\rangle}{\left\|v^j\right\|^2} v^j,$$

where

$$H_j = \{(x,\lambda) \in \mathbb{R}^n \times \mathbb{R}^m : \langle v^j, (x - \tilde{x}^j, \lambda - \tilde{\lambda}^j)\rangle = 0\}.$$

In the finite dimensional setting, Algorithm IPPBP, introduced in Section 2.2, reduces to the following one, to be called Algorithm B, with auxiliary function $g(x) = \frac{1}{2}\left\|x\right\|^2$, which is a tool for the analysis of Algorithm IALP and Algorithm LIALP.

Algorithm B: Inexact Proximal Point+Orthogonal Projection Method in $\mathbb{R}^n$ for EP(f, K)

1. Consider an exogenuous bounded sequence of regularization parameters $\{\gamma_j\}_{j=0}^{\infty} \subset \mathbb{R}_{++}$ 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

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

i.e.

$$f_j^e(\tilde{x}^j, y) \geq 0 \quad \forall y \in K,$$

and e^j satisfies

$$\left\|e^j\right\| \leq \frac{1}{2}\sigma\gamma_j \begin{cases} \|\tilde{x}^j - x^j\|^2 & \text{if } \|x^j - \tilde{x}^j\| < 1 \\ 1 & \text{if } \|x^j - \tilde{x}^j\| \geq 1. \end{cases}$$

3. Let

$$v^j = \gamma_j(x^j - \tilde{x}^j) + e^j \in \mathbb{R}^n.$$

If $v^j = 0$ or $\tilde{x}^j = x^j$, then stop. Otherwise, take $H_j = \{x \in \mathbb{R}^n : \langle v^j, x - \tilde{x}^j \rangle = 0\}$ and define

$$x^{j+1} = \operatorname*{argmin}_{x \in H_j} \frac{1}{2}\left\|x - x^j\right\|^2 = x^j - \frac{\langle v^j, x^j - \tilde{x}^j \rangle}{\left\|v^j\right\|^2} v^j.$$

Now consider Algorithm IALP. For analyzing Algorithm IALP, we apply Algorithm B to the equilibrium problem EP($\mathcal{L}, \mathbb{R}^n \times \mathbb{R}^m_+$) with $\mathcal{L}$ given by (3.2). In view of Proposition 3.2.2, we know that this problem satisfies P1–P4 and P4”. Thus, we use

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

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

as the regularized function at iteration j. Then we 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 EP($\widehat{\mathcal{L}}_j^e, \mathbb{R}^n \times \mathbb{R}^m_+$) and

$$\left\|e^j\right\| = \left\|(e^j, 0)\right\| \leq \frac{1}{2}\sigma\gamma_j \begin{cases} \left\|(\hat{x}^j - x^j, \hat{\lambda}^j - \lambda^j)\right\|^2 & \text{if } \left\|(\hat{x}^j - x^j, \hat{\lambda}^j - \lambda^j)\right\| < 1 \\ \\ 1 & \text{if } \left\|(\hat{x}^j - x^j, \hat{\lambda}^j - \lambda^j)\right\| \geq 1. \end{cases}$$

Then, in Step 3 of Algorithm B, one puts

$$\hat{v}^j = \gamma_j(x^j - \hat{x}^j, \lambda^j - \hat{\lambda}^j) + (e^j, 0) \in \mathbb{R}^n \times \mathbb{R}^m,$$

and checks if $\hat{v}^j = 0$ or $(\hat{x}^j, \hat{\lambda}^j) = (x^j, \lambda^j)$. If so, the algorithm stops. Otherwise, one takes $\widehat{H}_j = \{x \in \mathbb{R}^n \times \mathbb{R}^m : \langle \hat{v}^j, (x - \hat{x}^j, \lambda - \hat{\lambda}^j)\rangle = 0\}$ and defines the next iterate as

$$(x^{j+1},\lambda^{j+1}) = \operatorname*{argmin}_{(x,\lambda)\in\widehat{H}_j} \frac{1}{2}\left\|(x-x^j,\lambda-\lambda^j)\right\|^2 = (x^j,\lambda^j) - \frac{\langle \hat{v}^j,(x^j-\hat{x}^j,\lambda^j-\hat{\lambda}^j)\rangle}{\|\hat{v}^j\|^2}\hat{v}^j$$

which is the orthogonal projection of (x^j,λ^j) onto the hyperplane $\widehat{H}_j$.

If we follow the argument used to prove the convergence properties of Algorithm IALE in Section 3.2, we obtain the following convergence theorem. We omit the proof for the sake of conciseness.

Theorem 3.4.1. *Consider* 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* (3.1)*,*

iii) the Slater's CQ *stated in Definition 1.7.3 holds for* K*,*

iv) $\{\gamma_j\}_{j=0}^{\infty}\subset(\theta,\bar{\gamma}]$ *for some* $\bar{\gamma}>\theta$*, where* θ *is the undermonotonicity constant of* f *in* P4*,*

v) $\sigma\in[0,1]$*,*

vi) EP(f,K) *has solutions.*

Let $\{(x^j,\lambda^j)\}_{j=0}^{\infty}$ *be the sequence generated by Algorithm* IALP *for solving* EP(f,K)*. In this situation, the sequence* $\{(x^j,\lambda^j)\}_{j=0}^{\infty}$ *converges to some optimal pair* (x^*,λ^*) *for* EP(f,K)*, and consequently* $x^*\in S_E(f,K)$*. Additionally, assume that*

vii) f *is continuously differentiable,*

viii) h_i *is differentiable* $(1\le i\le m)$*.*

In this case, if x^j *is not a solution of* EP(f,K) *and* $\breve{x}^j$ *is the unique solution of* EP$(\widetilde{\mathcal{L}}_j^e,\mathbb{R}^n)$ *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* IALP*.*

Similar to Theorem 3.4.1, one can prove the following theorem for Algorithm LIALP.

Theorem 3.4.2. *Consider* 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* (3.1)*,*

iii) the Slater's CQ *stated in Definition 1.7.3 holds for* K*,*

iv) $\{\gamma_j\}_{j=0}^{\infty} \subset (\theta, \bar{\gamma}]$ *for some* $\bar{\gamma} > \theta$*, where* θ *is the undermonotonicity constant of* f *in* P4*,*

v) $\sigma \in [0, 1]$*,*

vi) $\mathrm{EP}(f, K)$ *has solutions,*

vii) f *is continuously differentiable,*

viii) h_i *is differentiable* $(1 \leq i \leq m)$*.*

Let $\{(x^j, \lambda^j)\}_{j=0}^{\infty}$ *be the sequence generated by Algorithm* LIALP *applied to* $\mathrm{EP}(f, K)$*. In this situation, the sequence* $\{(x^j, \lambda^j)\}_{j=0}^{\infty}$ *converges to an optimal pair* (x^*, λ^*) *for* $\mathrm{EP}(f, K)$*, so that* x^* *belongs to* $S_E(f, K)$*. 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* LIALE*.*

We remind that we did not consider an error vector associated with the λ and μ arguments of $\mathcal{L}$ in the case of Algorithm IALE and Algorithm LIALE, because the λ_i^j's are updated via a closed formula in these two algorithms. For the same reason, we act in the same way for both Algorithm IALP and Algorithm LIALP.

3.5 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 thesis (Algorithm IALE, Algorithm LIALE, Algorithm IALP and Algorithm LIALP), 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 [30]. 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 [30] will be enough for satisfying our error criteria, as discussed in this chapter (see Corollary 3.2.13, Theorem 3.3.2, Theorem 3.4.1, and Theorem 3.4.2).

Another option consists of solving the system of equations resulting from (3.39) in the case that functions f and h_i's are continuously differentiable, namely

$$0 = \gamma_j(x - x^j) + F'_x(x) + \sum_{i=1}^{m} \max\left\{0, \lambda_i^j + \frac{h_i(x)}{\gamma_j}\right\} h'_i(x). \tag{3.50}$$

We observe that the right hand side of (3.50) is continuous but not differentiable, due to the presence of the maximum. However, there is a substantial choice of efficient methods for nonsmooth equations which can be used in this case like the ones have been appeared in [19].

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.

Bibliography

[1] Antipin, A.S. Equilibrium programming: proximal methods. *Computational Mathematics and Mathematical Physics* **37** (1997) 1285–1296.

[2] Antipin, A.S., Vasilev, F.P., Stukalov, A.S. A regularized Newton method for solving equilibrium programming problems with an inexactly specified set. *Computational Mathematics and Mathematical Physics* **47** (2007) 19–31.

[3] Bachman, G., Narici, L. *Functional Analysis.* Dover, New York (2000).

[4] Bertsekas, D.P. *Constrained optimization and Lagrangian multiplier methods.* Academic Press, New York (1982).

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

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

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

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

[9] Brezis, H. *Analyse fonctionnelle Théorie et application.* Masson, Paris (1983).

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

[11] Burachik, R.S., Iusem, A.N. *Set-valued analysis and enlargements of monotone operators.* Springer, New York (2007).

[12] Burachik, R.S., Scheimberg, S. A proximal point algorithm for the variational inequality problem in Banach spaces. *SIAM Journal on Control and Optimization* **39** (2001) 1633–1649.

[13] Butnariu, D., Iusem, A.N. On a proximal point method for convex optimization in Banach spaces. *Numerical Functional Analysis and Optimization* **18** (1997) 723–744.

[14] Butnariu, D., Iusem, A.N. *Totally convex functions for fixed points computation and infinite dimensional optimization.* Kluwer, Dordrecht (2000).

[15] Butnariu, D., Iusem, A.N., Resmerita, E. Total convexity for powers of the norm in uniformly convex Banach spaces. *Journal of Convex Analysis* **7** (2000) 315–334.

[16] Butnariu, D., Resmerita, E. Bregman distances, totally convex functions and a method for solving operator equations in Banach spaces. *Abstract and Applied Analysis* (2006) Art. ID 84919.

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

[18] Cioranescu, I. *Geometry of Banach spaces, duality mappings, and nonlinear problems.* Kluwe Academic Publishers Group, Dordrecht (1990).

[19] Facchinei, F., Pang, J.S. *Finite-dimensional Variational Inequalities and Complementarity Problems II*. Springer, Berlin (2003).

[20] Flåm, S.D., Antipin, A.S. Equilibrium programming using proximal-like algorithms. *Mathematical Programming* **78** (1997) 29–41.

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

[22] Gárciga Otero, R., Iusem, A.N. Proximal methods in Banach spaces without monotonicity. *Journal of Mathematical Analysis and Applications* **330** (2007) 433–450.

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

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

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

[26] 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.

[27] Iusem, A.N., Kassay, G., Sosa, W. On certain conditions for the existence of solutions of equilibrium problems. *Mathematical Programming* **116** (2009) 259–273.

[28] Iusem, A.N., Nasri, M. Augmented Lagrangian methods for equilibrium problems (submitted).

[29] Iusem, A.N., Nasri, M. Inexact proximal point methods for equilibrium problems in Banach spaces. *Numerical Functional Analysis and Optimization* **28** (2007) 1279–1308.

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

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

[32] Iusem, A.N., Sosa, W. On the proximal point method for equilibrium problems in Hilbert spaces (submitted).

[33] Jahn, J. *Introduction to the theory of nonlinear optimization.* Springer-Verlag (1994).

[34] Kassay, G. The proximal point algorithm for reflexive Banach spaces. *Studia Mathematica* **30** (1985) 9–17.

[35] Konnov, I. V. Application of the proximal point method to a system of extended primal-dual equilibrium problems. *Lecture Notes in Economics and Mathematical Systems* **563** (2006) 87–102.

[36] Konnov, I. V. Application of the proximal point method to nonmonotone equilibrium problems. *Journal of Optimization Theory and Applications* **119** (2003) 317–333.

[37] Konnov, I. V. Partial proximal point method for nonmonotone equilibrium problems. *Optimization Methods and Software* **21** (2006) 373–384.

[38] 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.

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

[40] Mastroeni, G. Gap functions for equilibrium problems. *Journal of Global Optimization* **27** (2003) 411–426.

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

[42] Moudafi, A. On finite and strong convergence of a proximal method for equilibrium problems. *Numerical Functional Analysis and Optimization* **28** (2007) 1279–1308.

[43] Moudafi, A. Proximal point methods extended to equilibrium problems. *Journal of Natural Geometry* **15** (1999) 91–100.

[44] Moudafi, A. Second-order differential proximal methods for equilibrium problems. *Journal of Inequalities in Pure and Applied Mathematics* **4** (2003) Article no. 18.

[45] Moudafi, A., Théra, M. Proximal and dynamical approaches to equilibrium problems. In *Ill-posed Variational Problems and Regularization Techniques*. Lecture Notes in Economics and Mathematical Systems, Springer, Berlin **477** (1999) 187–201.

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

[47] Noor, M.A. Auxiliary principle technique for equilibrium problems. *Journal of Optimization Theory and Aplications* **122** (2004) 371–386.

[48] Noor, M.A., Rassias, T.M. On nonconvex equilibrium problems. *Journal of Mathematical Analysis and Applications* **312** (2005) 289–299.

[49] Nasri, M., Sosa, W. Generalized Nash games and equilibrium problems (submitted).

[50] Oettli, W. A remark on vector-valued equilibria and generalized monotonicity. *Acta Mathematica Vietnamica* **22** (1997) 213–221.

[51] Ortega, J.M. *Numerical Analysis. A Second Course*. Academic Press, New York (1972).

[52] Pascali, D., Sburlan, S. *Nonlinear Mappings of Monotone Type*. Editura Academiei, Bucarest (1978).

[53] Phelps, R.R. *Convex Functions, Monotone Operators and Differentiability*. Springer-Verlag, Berlin (1989).

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

[55] Reed, M., Simon, B. *Methods of modern mathematical physics*. Academic Press, New York (1980).

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

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

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

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

[60] Rudin, W. *Functional Analysis*. McGraw Hill, New York (1973).

[61] Simons, S. *Minimax and Monotonicity*. Lecture Notes in Mathematics **1693**. Springer, Berlin (1998).

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

[63] 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.

[64] Sosa, W. *Iterative algorithms for the abstract equilibrium problem.* Ph.D. Thesis. IMPA, Rio de Janeiro, (1999).

[65] Zeidler, E. *Nonlinear Functional Analysis and its Applications I*. Springer-Verlag, New York (1985).