Inexact proximal point methods for equilibrium problems in Banach spaces

Alfredo N. Iusem *and Mostafa Nasri †

July 3, 2007

Abstract

We introduce two inexact proximal-like methods for solving equilibrium problems in reflexive Banach spaces and establish their convergence properties, proving that the sequence generated by each one of them converges to a solution of the equilibrium problem under reasonable assumptions.

Keywords: Bregman distance, Equilibrium problem, Inexact solutions, Maximal monotone operator, Proximal point method.

1 Introduction

Let B be a reflexive Banach space and $K \subset B$ a nonempty closed and convex set. Given $f : K \times K \longrightarrow \mathbb{R}$ such that

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

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

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

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

The equilibrium problem encompasses, among its particular cases, convex minimization problems, fixed point problems, complementarity problems, Nash equilibrium problems, variational inequality problems and vector minimization problems (see, e.g., [4], [19]).

The prototypical example of an equilibrium problem is a variational inequality problem. Since it plays an important role in the sequel, we describe it now in some detail. Consider a continuous

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

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

$T : B \to B^*$, where B^* is the dual of B, and define $f(x,y) = \langle T(x), y-x\rangle$. Then f satisfies P1-P3, and EP(f,K) is equivalent to the variational inequality problem VIP(T,K), consisting of finding a point $x^* \in K$ such that $\langle T(x^*), x-x^*\rangle \geq 0$ for all $x \in K$. We can consider also the case of a point-to-set operator $T : B \to \mathcal{P}(B^*)$, if it is maximal monotone. In this case VIP(T,K) consists of finding $x^* \in K$ such that $\langle v^*, x-x^*\rangle \geq 0$ for some $v^* \in T(x^*)$ and all $x \in K$. In this situation, we define $f(x,y) = \sup_{u\in T(x)}\langle u, y-x\rangle$. Though it is less immediate, this f is well defined and it still satisfies P1-P3. Finiteness of f follows from monotonicity of T, and upper semicontinuity of $f(\cdot,y)$ from maximality (via demiclosedness of the graph of maximal monotone operators).

The equilibrium problem has been rather widely studied, but most of the work on the issue deals with conditions for the existence of solutions (see, e.g., [1], [2], [3], [5], [11],[12], [13], [16] and [19]).

In terms of computational methods for the equilibrium problem, only a few references can be found in the literature. Among those of interest, we mention the algorithms introduced in [19] and [25].

In the current paper we introduce inexact versions of a proximal point method for solving EP(f,K) in reflexive Banach spaces. We comment next on this method.

The proximal point algorithm, whose origins can be traced back to [23] and [24], attained its basic formulation in the work of Rockafellar [29], where it is presented as an algorithm for finding zeroes of a maximal monotone point-to-set operator $T : H \to \mathcal{P}(H)$, where H is a Hilbert space. The algorithm generates a sequence $\{x^j\} \subset H$, starting from some $x^0 \in H$, where x^{j+1} is the unique zero of the operator T^j defined as

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

with $\{\gamma_j\}$ being a bounded sequence of positive real numbers, called regularization coefficients. It has been proved in [29] that for a maximal monotone T, the sequence $\{x^j\}$ 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 $x^0 \in H$.

Proximal point methods for the same problem but in Banach spaces, can be traced back to [22] and were furtherly developed in [7] and [8]. In this set-up, one works with maximal monotone point-to-set operators $T : B \to \mathcal{P}(\mathcal{B}^*)$, where B is Banach space and B^* its dual, so that the formula for T^j given by (1) does not work any more, because $T(x)$ is a subset of B^*, while $\gamma_k(x - x^j)$ belongs to B. Thus, instead of (1), one works with

$$T^j(x) = T(x) + \gamma_j[g'(x) - g'(x^j)], \tag{2}$$

where $g : B \to \mathbb{R}$ is Fréchet differentiable satisfying some regularity properties (see H1-H6 in Section 2). The prototypical example is $g(x) = (1/2)\|x\|^2$, in which case g' is the duality operator, and the identity operator in the case of Hilbert spaces, for which the definitions of T^j given by (1) and (2) coincide. 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 [15] that the function $g(x) = r\|x\|^s$, works satisfactorily in any reflexive, uniformly smooth and uniformly convex Banach space, for any $r > 0$, $s > 1$. The convergence analysis recovers most of the already mentioned results which hold in Hilbert spaces (see, e.g., [7]).

A proximal point method for equilibrium problems in a Hilbert space H has been recently proposed in [20]. At iteration j, given $x^j \in H$, one solves the problem $\mathrm{EP}(\bar{f}_j, K)$, where the regularized function $\bar{f}_j$ is defined as

$$\bar{f}_j(x,y) = f(x,y) + \gamma_j \langle x - x^j, y - x\rangle. \tag{3}$$

It is established in [20] 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 sequence $\{x^j\}$ generated by the method is shown to be weakly convergent to a solution of the problem, under appropriate assumptions on f, when the problem has solutions.

Since the computation of each iterate requires solution of a regularized equilibrium problem (which will be called a *subproblem* in the sequel), it is important to establish convergence results assuming inexact solutions of the subproblems. This issue was already dealt with in [29], 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. A better error criterion, introduced in [32], allows for constant relative errors and can be described, in its Hilbert space version, as follows. Instead of solving (1), 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

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

where the error term e^j satisfies

$$\left\|e^j\right\| \le \sigma\gamma_j \max\left\{\left\|x^j - \hat{x}^j + \gamma_j^{-1}e^j\right\|, \left\|x^j - \hat{x}^j\right\|\right\}, \tag{4}$$

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 $v^j = \gamma_j(x^j - \hat{x}^j) + e^j$, i.e.

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

This inexact procedure, as well as a related one introduced in [33], were extended to Banach spaces in [15].

In this paper we will build upon the results of [15] and [20], obtaining inexact proximal point methods for equilibrium problems in Banach spaces. Instead of $\bar{f}_j$ as defined in (3), we will solve at iteration j the regularized problem $\mathrm{EP}(f_j, K)$, with f_j given by

$$f_j(x,y) = f(x,y) + \gamma_j \langle g'(x) - g'(x^j), y - x \rangle, \tag{5}$$

where $g : B \to \mathbb{R}$ is an auxiliary function satisfying appropriate assumptions (see H1-H6 in Section 2). When B is a Hilbert space and $g(x) = (1/2)\|x\|^2$, 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,

$$f_j^e(x,y) = f_j(x,y) - \langle e^j, y - x \rangle = f(x,y) + \gamma_j \langle g'(x) - g'(x^j), y - x \rangle - \langle e^j, y - x \rangle, \tag{6}$$

where $e^j \in B^*$, the error vector at the j-th iteration, will be subject to bounds similar to (4), to be presented in Section 4.

The outline of this paper is the following. We present some preliminary material in Section 2. In Section 3 we show that (6) can be considered as a regularization of $\mathrm{EP}(f, K)$, i.e. $\mathrm{EP}(f_j^e, K)$ has a unique solution for each j whenever certain conditions are satisfied. In Section 4 we present two inexact proximal point algorithms for solving $\mathrm{EP}(f, K)$ and establish their convergence properties in Section 5. In Section 6 we present a reformulation of $\mathrm{EP}(f, K)$ which allows us to refine our convergence results.

2 Preliminary results

We state first some standard notation. B will denote a reflexive Banach space with norm $\|\cdot\|$ (or $\|\cdot\|_B$), B^* will be the dual of B, with norm $\|\cdot\|_*$, We denote the open ball with radius δ centered at $x \in B$ by $B(x, \delta) \subset B$ and $B^*(x^*, \delta) \subset B^*$ will denote the open ball with radius δ centered at $x^* \in B^*$, respectively. The symbol $\langle \cdot, \cdot \rangle$ will indicate the duality coupling in $B^* \times B$, defined by $\langle \phi, x \rangle = \phi(x)$ for all $x \in B$ and all $\phi \in B^*$. For a Gâteaux differentiable function $g : B \to \mathbb{R}$, $g' : B \to B^*$ will denote its Gâteaux derivative.

Our results require some conditions both on the data problem, namely the function f, and on the auxiliary function g. We will present next such conditions, starting with g. First, we introduce the Bregman distance associated to a convex and Gâteaux differentiable function $g : B \to \mathbb{R}$, together with some related notations, taken from [9].

Definition 2.1. *The Bregman distance with respect to g is the function $D_g : B \times B \to \mathbb{R}$ defined by $D_g(x,y) = g(x) - g(y) - \langle g'(y), x - y \rangle$.*

Definition 2.2. *The modulus of total convexity of g is the function $\nu_g : B \times [0, +\infty) \to [0, +\infty]$ defined by $\nu_g(x,t) = \inf\{D_g(y,x) : \|y - x\| = t\}$.*

Next we present some assumptions on g which will be needed in our convergence analysis.

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

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

H3: g' is uniformly continuous on bounded subsets of B.

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 $\{y^j\}$ and $\{z^j\}$ are sequences in K which converge weakly to y and z, respectively and $y \neq z$, then

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

These properties, with the exception of H5, were identified in [15]. We make a few remarks on them. Observe that positivity of $D_g(x,y)$ for all $x, y \in B$ is equivalent to strict convexity of g. Thus, condition H2 implies strict convexity of g (functions satisfying a weaker condition, namely positivity of $\nu_g(x,t)$ for all $x \in B$ and all $t > 0$, are called *totally convex* in [9]). H5, introduced here for the first time in connection with Bregman functions and distances, is a form of coercivity. It has been proved in [9], p.75, 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. We have the following result.

Proposition 2.3.

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 $g_2(x) = \frac{1}{2}\|x\|^2$ satisfies H6. The same holds for $g_p(x) = \frac{1}{p}\|x\|_p^p$ 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 [15]. For H5, note that for the function of interest we have

$$g(x) - \rho\|x - z\| = r\|x\|^s - \rho\|x - z\| \geq r\|x\|^s - \rho\|x\| - \rho\|z\| = \|x\|\,[r\|x\|^{s-1} - \rho] - \rho\|z\|, \quad (7)$$

and the rightmost expression of (7) 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 5, most of our convergence results demand only H1-H5.

The following results are consequences of some of the properties above. The first one deals with Bregman projections onto hyperplanes. For fixed $0 \neq v \in B^*, \tilde{y} \in B$, let $H = \{y \in B : \langle v, y - \tilde{y}\rangle = 0\}$, $H^+ = \{y \in B : \langle v, y - \tilde{y}\rangle \geq 0\}$, and $H^- = \{y \in B : \langle v, y - \tilde{y}\rangle \leq 0\}$.

Proposition 2.4. *Assume that g satisfies H2. Then for all $0 \neq v \in B^*$, $\tilde{y} \in B$, $x \in H^+$ 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 $D_g(\cdot, \tilde{y})$ on H.*

Proof. See Lemma 1 in [15], where the result is proved under an assumption weaker than H2, namely total convexity of g. □

The next result deals with Bregman projections onto closed and convex sets.

Proposition 2.5. *If g satisfies H1-H2 and $C \subset B$ is closed and convex, then for all $\bar{x} \in B$ there exists a unique solution $\hat{x}$ of the problem $\min D_g(x, \bar{x})$ s.t. $x \in C$, which satisfies $\langle g'(\bar{x}) - g'(\hat{x}), y - \hat{x}\rangle \leq 0$ for all $y \in C$.*

Proof. See [9], p. 70. □

The point $\hat{x}$ is said to be the *Bregman projection of $\bar{x}$ onto C.*

Proposition 2.6. *Assume that g satisfies H2. Let $\{x^k\}, \{y^k\} \subset B$ be two sequences such that at least one of them is bounded. If $\lim_{k\to\infty} D_g(y^k, x^k) = 0$, then $\lim_{k\to\infty} \left\|x^k - y^k\right\| = 0$.*

Proof. See Proposition 5 in [15]. □

Proposition 2.7. *If g satisfies H3, then both g and g' are bounded on bounded subsets of B.*

Proof. See Proposition 4 in [15]. □

We move now to conditions on the bifunction f, for which we start by recalling some definitions and facts related to maximal monotone operators. Let $T : B \to \mathcal{P}(B^*)$ be a point-to-set operator. The *graph* of T is the set $G(T) = \{(v, x) \in B^* \times B : v \in T(x)\}$. A point $x \in B$ is a zero of T if $0 \in T(x)$. T is said to be *monotone* if $\langle w - w', x - x'\rangle \geq 0$, for all x, $x' \in B$ and all $w \in T(x)$, $w' \in T(x')$. It is *maximal monotone* if its graph is not properly contained in the graph of any other monotone operator.

Convergence results for the proximal point method applied to the problem of finding zeros of operators require that they be monotone, or some variant thereof. We need to transpose such monotonicity concepts to the realm of equilibrium problems. We consider first the following alternatives:

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: 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 $x, y \in K$.

In the prototypical example of an equilibrium problem, i.e., $f(x,y) = \sup_{u\in T(x)}\langle u, y-x\rangle$ for some $T : B \to \mathcal{P}(B^*)$, it is easy to check that P4$^\bullet$ is equivalent to monotonicity of T. Thus, a function f satisfying P4$^\bullet$ will be said to be *monotone.* We remind that an operator $T : B \to \mathcal{P}(B^*)$ is said to be *pseudomonotone* when $\langle u, x-y\rangle \leq 0$ for some $x, y \in H$ and some $u \in T(x)$ implies that $\langle v, x-y\rangle \geq 0$ for all $v \in T(y)$. It is easy to check that if T is pseudomonotone and single-valued then f, as defined above, satisfies P4*, and the converse statement holds also when T is point-to-set. For this reason, a function f satisfying P4* will be said to be *pseudomonotone.* Along the same line, a function f satisfying P4 will be called *θ-undermonotone.*

We comment now on some relations among P4, P4* and P4$^\bullet$. It has been proved in Proposition 1 of [20] that under P1-P3, P4$^\bullet$ implies P4*. On the other hand, P4* does not imply P4$^\bullet$, as the following example, also taken from [20], shows.

Example 2.8. *Let $K = [1/2, 1] \subset \mathbb{R}$ and define $f : K \times K \to \mathbb{R}$ as $f(x,y) = x(x-y)$. This f satisfies P1, P2 and P3. It is pseudomonotone and 1-undermonotone, but not monotone.*

Two other monotonicity-like properties which have been considered in the literature are the following:

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

$$\min_{1\leq i\leq m} f(x^i, \Sigma_{j=1}^m \lambda_j x^j) < 0. \tag{8}$$

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

$$\Sigma_{i=1}^m \lambda_i f(x^i, \Sigma_{k=1}^m \lambda_k x^k) \leq 0. \tag{9}$$

It is immediate that both P4' and P4" are weaker than P4$^\bullet$, i.e., than monotonicity. Examples are given in [16] showing that P4', P4" and P4* are mutually independent.

The following property has been identified in [16] as being necessary and sufficient for the existence of solutions of EP(f, K) under some monotonicity properties of f.

P5: For any sequence $\{x^j\} \subseteq K$ 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 $j \geq j_0$.

Indeed, we have the following result:

Theorem 2.9. *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 [16]. ☐

3 Regularization of equilibrium problems

The following proposition guarantees that, under adequate monotonicity assumptions, the function f_j^e introduced in (6) is a regularization of f.

Proposition 3.1. *Consider f satisfying P1-P4. Fix $\bar{x} \in B, e \in B^*$ and $\gamma > \theta$, where θ is the undermonotonicity constant in P4. Take $g : B \to \mathbb{R}$ satisfying H1-H2 and H5. If $\tilde{f} : K \times K \to \mathbb{R}$ is defined as*

$$\tilde{f}(x, y) = f(x, y) + \gamma\langle g'(x) - g'(\bar{x}), y - x\rangle - \langle e, y - x\rangle, \tag{10}$$

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

Proof. We first prove existence of solutions. In view of Theorem 2.9, it suffices to show that $\tilde{f}$ satisfies P1-P3, P4* and P5. It follows from (10) that $\tilde{f}$ inherits P1-P3 from f. Now we claim that $\tilde{f}$ satisfies P4*. Note that

$$\begin{aligned} \tilde{f}(x, y) + \tilde{f}(y, x) &= f(x, y) + f(y, x) - \gamma\langle g'(x) - g'(y), x - y\rangle \\ &\le (\theta - \gamma)\langle g'(x) - g'(y), x - y\rangle \le 0, \end{aligned} \tag{11}$$

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 (11) that P4* holds for $\tilde{f}$. In order to apply Theorem 2.9, it suffices to establish that $\tilde{f}$ satisfies P5. So, we take a sequence $\{x^j\}$ 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 after Proposition 2.5. Note that

$$\begin{aligned} &\tilde{f}(x^j, u) = f(x^j, u) - \gamma\langle g'(x^j) - g'(\bar{x}), x^j - u\rangle - \langle e, u - x^j\rangle = f(x^j, u) \\ &- \gamma\langle g'(u) - g'(\bar{x}), x^j - u\rangle - \gamma\langle g'(x^j) - g'(u), x^j - u\rangle - \langle e, u - x^j\rangle \le \\ &- f(u, x^j) - \gamma\langle g'(u) - g'(\bar{x}), x^j - u\rangle + (\theta - \gamma)\langle g'(x^j) - g'(u), x^j - u\rangle \\ &+ \|e\|_* \left\|u - x^j\right\| \le -f(u, x^j) + (\theta - \gamma)\langle g'(x^j) - g'(u), x^j - u\rangle + \|e\|_* \left\|u - x^j\right\|, \end{aligned} \tag{12}$$

using P4 and Cauchy-Schwartz inequality in the first inequality and Proposition 2.5 in the second one, taking into account that $\gamma > 0$ 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

$$F_x(y) = f(x, y). \tag{13}$$

Take $\hat{x} \in \mathrm{ri}(K)$ (the relative interior of K), which is nonempty by convexity of K. Since F_u is convex by P2, its subdifferential at $\hat{x}$, denoted as $\partial F_u(\hat{x})$, is nonempty. Take $\hat{v} \in \partial F_u(\hat{x})$. By the definition of subdifferential, we have

$$\langle \hat{v}, x^j - \hat{x}\rangle \le F_u(x^j) - F_u(\hat{x}) = f(u, x^j) - f(u, \hat{x}). \tag{14}$$

In view of (14),

$$-f(u,x^j) \leq \langle \hat{v}, \hat{x}-x^j\rangle - f(u,\hat{x}) \leq \|\hat{v}\|_* \left\|\hat{x}-x^j\right\| - f(u,\hat{x}) \leq \|\hat{v}\|_* \left\|u-x^j\right\| + \|\hat{v}\|_* \|\hat{x}-u\| - f(u,\hat{x}), \tag{15}$$

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 rightmost expression of (12).

$$\begin{aligned} &\langle g'(x^j) - g'(u), x^j - u\rangle = D_g(x^j,u) + D_g(u,x^j) \geq D_g(x^j,u) = \\ &g(x^j) - g(u) - \langle g'(u), x^j - u\rangle \geq g(x^j) - g(u) - \left\|g'(u)\right\|_* \left\|x^j - u\right\|, \end{aligned} \tag{16}$$

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

Now we utilize the inequalities obtained in (15) and (16) to get an upper bound for the rightmost expression in (12).

$$\begin{aligned} &\tilde{f}(x^j,u) \leq (\theta-\gamma)\left[g(x^j) - \frac{\|\hat{v}\|_* + \|e\|_* + (\gamma-\theta)\|g'(u)\|_*}{\gamma-\theta}\left\|x^j-u\right\|\right] \\ &+ (\gamma-\theta)g(u) + \|\hat{v}\|_* \|\hat{x}-u\| - f(u,\hat{x}). \end{aligned} \tag{17}$$

Let $z = u$ and $\rho = (\gamma-\theta)^{-1}(\|\hat{v}\|_* + \|e\|_*) + \|g'(u)\|_*$. Since $\theta - \gamma < 0$, $\lim_{j\to\infty}\left\|x^j\right\| = +\infty$ and g satisfies H5, it follows from (17) 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 2.9, 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

$$0 \leq \tilde{f}(\tilde{x},\tilde{x}') = f(\tilde{x},\tilde{x}') + \gamma\langle g'(\tilde{x}) - g'(\bar{x}), \tilde{x}'-\tilde{x}\rangle - \langle e, \tilde{x}'-\tilde{x}\rangle. \tag{18}$$

$$0 \leq \tilde{f}(\tilde{x}',\tilde{x}) = f(\tilde{x}',\tilde{x}) + \gamma\langle g'(\tilde{x}') - g'(\bar{x}), \tilde{x}-\tilde{x}'\rangle - \langle e, \tilde{x}-\tilde{x}'\rangle. \tag{19}$$

Adding (18) and (19), 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. □

4 Inexact versions of the proximal point method

We start by presenting an exact proximal point method for equilibrium problems 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 P4*.

The algorithm requires two exogenous data: an auxiliary function $g : B \to \mathbb{R}$ satisfying H1-H5 and a sequence of regularization parameters $\{\gamma_j\} \subset (\theta, \bar{\gamma}]$, where θ is the constant of undermonotonicity appearing in P4 and $\bar{\gamma} > \theta$.

The algorithm generates a sequence $\{x^j\} \subset B$ as follows. x^0 is an arbitrary point in K, and, given x^j, x^{j+1} is the solution of $\mathrm{EP}(f_j, K)$ 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 3.1, with $\gamma = \gamma_j$, $\bar{x} = x^j$. When B is a Hilbert space and $g(x) = (1/2) \|x\|^2$ this method reduces to the one analyzed in [20]. The convergence properties of the method can be summarized as follow.

Proposition 4.1.

i) If $\mathrm{EP}(f, K)$ *has solutions, then the sequence* $\{x^j\}$ *is bounded and asymptotically solves the problem, in the sense that* $\liminf_{j \to \infty} f(x^j, y) \geq 0$ *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* $\{x^j\}$ *are solutions of* $\mathrm{EP}(f, K)$.

iii) If additionally g *satisfies H6, then* $\{x^j\}$ *is weakly convergent to a solution of* $\mathrm{EP}(f, K)$.

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 $\mathrm{EP}(f_j^e, K)$, where $f_j^e(x, y) = f_j(x, y) + \langle e^j, x - y \rangle$ and e^j 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 $\mathrm{S}(f, K)$. Then x^{j+1} is obtained by either projecting x^j onto H_j with respect to the Bregman distance D_g (Algorithm I), or by taking a step from x^j in the direction of H_j with respect to the metric induced by D_g (Algorithm II). 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., [26]).

The function g and the sequence $\{\gamma_j\}$ 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 $\sigma \in (0, 1)$.

Algorithm I: Inexact Proximal Point+Bregman Projection Method

1. Choose $x^0 \in K$.
2. Given x^j, find a pair $\tilde{x}^j \in B, e^j \in B^*$ such that $\tilde{x}^j$ solves $\mathrm{EP}(\tilde{f}_j^e, K)$ with

$$f_j^e(x, y) = f(x, y) + \gamma_j \langle g'(x) - g'(x^j), y - x \rangle - \langle e^j, y - x \rangle, \tag{20}$$

i.e.

$$f_j^e(\tilde{x}^j, y) \geq 0 \;\forall y \in K, \tag{21}$$

and e^j satisfies

$$\left\|e^j\right\|_* \leq \begin{cases} \sigma\gamma_j D_g(\tilde{x}^j, x^j) & \text{if } \left\|x^j - \tilde{x}^j\right\| < 1 \\ \sigma\gamma_j \nu_g(x^j, 1) & \text{if } \left\|x^j - \tilde{x}^j\right\| \geq 1, \end{cases} \tag{22}$$

with D_g, ν_g as in Definition 2.1 and 2.2, respectively.

3. Let

$$v^j = \gamma_j[g'(x^j) - g'(\tilde{x}^j)] + e^j. \tag{23}$$

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

$$x^{j+1} = \operatorname*{arg\,min}_{x \in H_j} D_g(x, x^j). \tag{24}$$

Algorithm II: Inexact Proximal Point-Extragradient Method

1. Choose $x^0 \in K$.
2. Given x^j, find a pair $\tilde{x}^j \in B, e^j \in B^*$ such that $\tilde{x}^j$ solves EP(f_j^e, K) with

$$f_j^e(x, y) = f(x, y) + \gamma_j\langle g'(x) - g'(x^j), y - x\rangle - \langle e^j, y - x\rangle, \tag{25}$$

i.e.

$$f_j^e(\tilde{x}^j, y) \geq 0 \;\forall y \in K, \tag{26}$$

and e^j satisfies

$$D_g(\tilde{x}^j, (g')^{-1}[g'(\tilde{x}^j) - \gamma_j^{-1}e^j]) \leq \sigma D_g(\tilde{x}^j, x^j). \tag{27}$$

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

$$x^{j+1} = (g')^{-1}[g'(\tilde{x}^j) - \gamma_j^{-1}e^j]. \tag{28}$$

It is worthwhile to observe that the inexact subproblems of both Algorithm I (i.e. (21)-(22)) and Algorithm II (i.e. (26)-(27)) are solvable. Indeed, given any $e^j \in B$, Proposition 3.1 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 (22) and (27) vanish, and so the inequalities in (22) and (27) are satisfied. We also note that for $e^j = 0$ both (24) and (28) 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.

5 Convergence Analysis

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

Proposition 5.1. *Suppose that Algorithm I (respectively Algorithm II) stops after j steps. Then $\tilde{x}^j$ generated by Algorithm I (respectively Algorithm II) is a solution of* $\mathrm{EP}(f,K)$.

Proof. Algorithm I stops at j-th iteration in two cases: if $v^j = 0$, in which case, by (20) and (23), $\tilde{x}^j \in \mathrm{S}(f,K)$, or if $\tilde{x}^j = x^j$, in which case, by (22), $e^j = 0$, which in turn implies, by (23), $v^j = 0$ and we are back to the first case. Consequently, $\tilde{x}^j$ is a solution of $\mathrm{EP}(f,K)$. Finite termination in Algorithm II occurs only if $\tilde{x}^j = x^j$, in which case $D_g(\tilde{x}^j, x^j) = 0$, and therefore, by (27), $e^j = 0$, which in turn implies, by (25)-(26), $f(\tilde{x}^j, y) \geq 0$ for all $y \in K$, so that $\tilde{x}^j \in \mathrm{S}(f,K)$. □

The convergence analysis of the proximal point method for finding zeroes of 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 2, 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 5.2. *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 $x^* \in \mathrm{S}(f,K)$.*

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 $x^* \in \mathrm{S}(f,K)$. Now take $x^1 = y$, $x^2 = x^*$ and $m = 2$ in (8), obtaining $f(y, \lambda y + (1-\lambda)x^*) < 0$. Using property P2 with $\lambda \to 0^+$, 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 (9), getting

$$\lambda f(y, \lambda y + (1-\lambda)x^*) + (1-\lambda) f(x^*, \lambda y + (1-\lambda)x^*) \leq 0. \tag{29}$$

Since the second term in the left hand side of (29) is nonnegative because $x^* \in \mathrm{S}(f,K)$, 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 $x^* \in \mathrm{S}(f,K)$. □

From now on we treat separately Algorithms I and II.

5.1 Convergence Analysis of Algorithm I

We start with a result similar to Lemma 2 in [15].

Lemma 5.3. *Let $\{x^j\}$, $\{\tilde{x}^j\}$, $\{\gamma_j\}$, and σ be as in Algorithm I and assume that g satisfies H2. For all j, it holds that*

$$\left\|e^j\right\|_* \left\|x^j - \tilde{x}^j\right\| \leq \sigma \gamma_j D_g(\tilde{x}^j, x^j) \leq \gamma_j D_g(x^j, \tilde{x}^j). \tag{30}$$

Proof. We consider two cases. First, if $\|x^j - \tilde{x}^j\| < 1$ then we have that $\|e^j\|_* \|x^j - \tilde{x}^j\| \le \|e^j\|_*$, so that the leftmost inequality in (30) follows trivially from (22). For the second case, namely $\|x^j - \tilde{x}^j\| \ge 1$, we will use the fact that $\nu_g(x, st) \ge s\nu_g(x,t)$ for all $s \ge 1, t \ge 0, x \in B$, which has been proved in [9], p. 18. Then,

$$\sigma\gamma_j D_g(\tilde{x}^j, x^j) \ge \sigma\gamma_j \nu_g(x^j, \|x^j - \tilde{x}^j\|) \ge \sigma\gamma_j \|x^j - \tilde{x}^j\| \nu_g(x^j, 1) \ge \|x^j - \tilde{x}^j\| \|e^j\|_*, \tag{31}$$

using Definition 2.2 in the first inequality, the above stated property of ν_g in the second one and (22) in the third. Thus, the first inequality of (30) is proved, and the second one holds because $\sigma \in [0,1]$. □

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

Proposition 5.4. *Consider* $\mathrm{EP}(f,K)$. *Assume that f satisfies P1-P4 and also any one among P4', P4" and P4*. Take $g: B \to \mathbb{R}$ satisfying H1-H5 and an exogenous sequence $\{\gamma_j\} \subset (\theta, \bar{\gamma}]$, where θ is the undermonotonicity constant in P4. Let $\{x^j\}$ be the sequence generated by Algorithm I. If* $\mathrm{EP}(f,K)$ *has solutions, then*

i) For all $x^* \in \mathrm{S}(f,K)$, $D_g(x^*, x^j)$ is nonincreasing and convergent.

ii) $\{x^j\}$ is bounded.

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

iv) $\{\gamma_j^{-1} e^j\}$ is bounded.

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

vi) $\{v^j\}$ converges strongly to 0.

Proof. Take $x^* \in \mathrm{S}(f,K)$. Let $H_j^- = \{x \in B : \langle v^j, x - \tilde{x}^j\rangle \le 0\}$. Since $\tilde{x}^j \in \mathrm{S}(f_j^e, K)$ with f_j^e given by (20), we have $f_j^e(\tilde{x}^j, y) \ge 0$ or equivalently $f(\tilde{x}^j, y) \ge \langle v^j, y - \tilde{x}^j\rangle$ for all $y \in K$, with v^j as in (23). In particular, $f(\tilde{x}^j, x^*) \ge \langle v^j, x^* - \tilde{x}^j\rangle$. Using Proposition 5.2, we have that $0 \ge \langle v^j, x^* - \tilde{x}^j\rangle$, so that $x^* \in H_j^-$. By definition of v^j and D_g we have that

$$\begin{aligned}\langle v^j, x^j - \tilde{x}^j\rangle = \gamma_j\langle g'(x^j) - g'(\tilde{x}^j), x^j - \tilde{x}^j\rangle + \langle e^j, x^j - \tilde{x}^j\rangle = \gamma_j[D_g(\tilde{x}^j, x^j) + D_g(x^j, \tilde{x}^j)] + \\ \langle e^j, x^j - \tilde{x}^j\rangle \ge \gamma_j D_g(\tilde{x}^j, x^j) + \left[\gamma_j D_g(x^j, \tilde{x}^j) - \|e^j\|_* \|x^j - \tilde{x}^j\|\right],\end{aligned} \tag{32}$$

where the last inequality follows from the definition of the norm in B^*. Applying now Lemma 5.3, we get from (32) $\langle v^j, x^j - \tilde{x}^j\rangle \ge \gamma_j D_g(\tilde{x}^j, x^j) \ge 0$, so that $x^j \in H_j^+$, and in view of (24), we are able to apply Lemma 2.4 with $\tilde{y} = \tilde{x}^j \in B$, $x = x^j$, $v = v^j$, and $z = x^{j+1}$, obtaining

$$D_g(x^*, x^j) \ge D_g(x^*, x^{j+1}) + D_g(x^{j+1}, x^j) \ge D_g(x^*, x^{j+1}). \tag{33}$$

The result of (i) follows from (33), taking into account the nonnegativity of D_g. We also obtain from (33) that $\{x^j\} \subseteq \{y \in B : D_g(x^*, y) \le D_g(x^*, x^0)\}$, and hence the sequence is bounded because of H1, establishing (ii).

We prove now (iii). Taking limits in (33) as $j \to \infty$ and using (i), we get

$$\lim_{j\to\infty} D_g(x^{j+1}, x^j) = 0. \tag{34}$$

Since g satisfies H2, we conclude from Proposition 2.6 that (iii) holds.

For proving (iv), we consider the set $A \subset B$ defined as $A = \{y : \|x^j - y\| \le 1 \text{ for some } j\}$. In view of (ii), A is bounded. By Definition 2.1 and Proposition 2.7, D_g is bounded on $A \times A$, say by $\zeta > 0$. From (22), taking into account that $\sigma \le 1$, we get

$$\left\|\gamma_j^{-1} e^j\right\|_* \le \begin{cases} D_g(\tilde{x}^j, x^j) & \text{if } \|x^j - \tilde{x}^j\| < 1 \\ \nu_g(x^j, 1) & \text{if } \|x^j - \tilde{x}^j\| \ge 1, \end{cases} \tag{35}$$

If $\|x^j - \tilde{x}^j\| \le 1$ then both x^j and $\tilde{x}^j$ belong to A and we get from (35) that $\left\|\gamma_j^{-1} e^j\right\|_* \le \zeta$; otherwise, we take any $y \in B$ such that $\|y - x^j\| = 1$, so that $y \in A$, and get from Definition 2.2 that $\nu_g(x^j, 1) \le D_g(y, x^j) \le \zeta$, so that, in view of (35), $\left\|\gamma_j^{-1} e^j\right\|_* \le \zeta$ also in this case. We conclude that $\{\gamma_j^{-1} e^j\}$ is bounded.

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

$$\lim_{j\to\infty} \langle \gamma_j^{-1} e^j, x^{j+1} - x^j \rangle = 0. \tag{36}$$

Observe that

$$\begin{aligned} &D_g(x^{j+1}, x^j) - D_g(x^{j+1}, \tilde{x}^j) - D_g(\tilde{x}^j, x^j) = \langle g'(x^j) - g'(\tilde{x}^j), \tilde{x}^j - x^{j+1} \rangle \\ &= \gamma_j^{-1} \langle v^j, \tilde{x}^j - x^{j+1} \rangle - \langle \gamma_j^{-1} e^j, \tilde{x}^j - x^{j+1} \rangle = \langle \gamma_j^{-1} e^j, x^{j+1} - \tilde{x}^j \rangle, \end{aligned} \tag{37}$$

where the first equality follows from the definition of D_g, the second one from (23) 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 (37),

$$\begin{aligned} &D_g(x^{j+1}, x^j) - D_g(x^{j+1}, \tilde{x}^j) = \langle \gamma_j^{-1} e^j, x^{j+1} - x^j \rangle + D_g(\tilde{x}^j, x^j) + \langle \gamma_j^{-1} e^j, x^j - \tilde{x}^j \rangle \\ &\ge \langle \gamma_j^{-1} e^j, x^{j+1} - x^j \rangle + \frac{1}{\gamma_j} \left[\gamma_j D_g(\tilde{x}^j, x^j) - \|e^j\|_* \|x^j - \tilde{x}^j\| \right] \ge \langle \gamma_j^{-1} e^j, x^{j+1} - x^j \rangle, \end{aligned} \tag{38}$$

where the last inequality follows from the leftmost inequality in (30), since $\sigma \in [0, 1]$. Taking limits as $j \to \infty$ in the leftmost and rightmost expressions of (38), we obtain, using (34) and (36),

$$\lim_{j\to\infty} D_g(x^{j+1}, \tilde{x}^j) = 0, \tag{39}$$

and therefore, we conclude, using H2 and Proposition 2.6, that $\{x^{j+1} - \tilde{x}^j\}$ converges strongly to 0, so that in view of (ii), $\{x^j - \tilde{x}^j\}$ converges strongly to 0, establishing (v).

Now we prove (vi). By (v), there exists $j_0 \in \mathbb{N}$ such that, $\left\|x^j - \tilde{x}^j\right\| < 1$, for all $j \geq j_0$, and consequently for $j \geq j_0$ our error criterium (22) implies that

$$\left\|e^j\right\|_* \leq \sigma\gamma_j D_g(\tilde{x}^j, x^j). \tag{40}$$

Next, we take limit as j goes to ∞ in the leftmost and rightmost expressions of (37). The rightmost one converges to 0 by (iii)-(v), the first term in the leftmost expression converges to 0 by (34) and the second term converges to 0 by (39). 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 (40) that e^j is strongly convergent to 0. From (v) and H3, we get that $\{g'(x^j) - g'(\tilde{x}^j)\}$ also converges strongly to 0. It follows then from (23) that v^j is strongly convergent to 0 as well. □

Now we proceed to state and prove our convergence result for Algorithm I. We remind that a sequence $\{z^j\} \subset K$ is an *asymptotically solving sequence* for $\mathrm{EP}(f, K)$ if $\liminf_{j\to\infty} f(z^j, y) \geq 0$ for all $y \in K$.

Theorem 5.5. *Consider* $\mathrm{EP}(f, K)$*. Assume that* f *satisfies P1-P4 and additionally any one among P4', P4" and P4*. Take* $g : B \to \mathbb{R}$ *satisfying H1-H5 and an exogenous sequence* $\{\gamma_j\} \subset (\theta, \bar{\gamma}]$*, where* θ *is the undermonotonicity constant in P4. Let* $\{x^j\}$ *be the sequence generated by Algorithm I. If* $\mathrm{EP}(f, K)$ *has solutions, then*

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

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

iii) If in addition g *satisfies H6, then the whole sequence* $\{x^j\}$ *is weakly convergent to some solution* x^* *of* $\mathrm{EP}(f, K)$*.*

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

$$\begin{aligned} 0 \leq f_j^e(\tilde{x}^j, y) &= f(\tilde{x}^j, y) + \langle \gamma_j[g'(\tilde{x}^j) - g'(x^j)] - e^k, y - \tilde{x}^j \rangle \\ &= f(\tilde{x}^j, y) + \langle -v^j, y - \tilde{x}^j \rangle \leq f(\tilde{x}^j, y) + \left\|v^j\right\|_* \left\|y - \tilde{x}^j\right\|. \end{aligned} \tag{41}$$

By Proposition 5.4(ii) and (v), we know that the sequence $\{\tilde{x}^j\}$ and therefore, the sequence $\{y - \tilde{x}^j\}$, are bounded for each fixed y. Consequently, taking limits in (41) as $j \to \infty$ and using Proposition 5.4(vi) we get

$$0 \leq \liminf_{j\to\infty} f(\tilde{x}^j, y), \tag{42}$$

for all $y \in K$.

ii) By Proposition 5.4(ii) and (v), $\{x^j\}$ has weak cluster points, all of which are also weak cluster points of $\{\tilde{x}^j\}$. These weak cluster points belong to K, which, being closed and convex, is weakly closed. Let $\tilde{x}$ be a weak cluster points of $\{\tilde{x}^j\}$, say the weak limit of the subsequence $\{\tilde{x}^{\ell_j}\}$ of $\{\tilde{x}^j\}$. Since $f(\cdot, y)$ is weakly upper semicontinuous, we obtain from (42) that $0 \leq \limsup_{j\to\infty} f(\tilde{x}^{\ell_j}, y) \leq f(\tilde{x}, y)$ for all $y \in K$. As a result, $\tilde{x}$ belongs to $\mathrm{S}(f, K)$.

iii) Assume that $\hat{x}$ is another weak cluster point of $\{x^j\}$, say the weak limit of the subsequence $\{x^{i_j}\}$ of $\{x^j\}$. By (ii), both $\tilde{x}$ and $\hat{x}$ solve $\mathrm{EP}(f, K)$. By Proposition 5.4(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

$$\left|\langle g'(x^{\ell_j}) - g'(x^{i_j}), \hat{x} - \tilde{x}\rangle\right| \leq \left|D_g(\hat{x}, x^{i_j}) - D_g(\hat{x}, x^{\ell_j})\right| + \left|D_g(\tilde{x}, x^{\ell_j}) - D_g(\tilde{x}, x^{i_j})\right|. \tag{43}$$

Taking limits in (43) 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, $\tilde{x} = \hat{x}$. □

5.2 Convergence Analysis of Algorithm II

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 5.6. *Consider* $\mathrm{EP}(f, K)$ *where* f *satisfies P1-P3 and any one among P4', P4" and P4*. Take* $g : B \to \mathbb{R}$ *satisfying H1-H5. Let* $\{x^j\}$, $\{\tilde{x}^j\}$, $\{\gamma_j\}$ *and* σ *be as in Algorithm II. Then*

$$D_g(x^*, x^{j+1}) \leq D_g(x^*, x^j) - \gamma_j^{-1}\langle v^j, \tilde{x}^j - x^*\rangle - (1 - \sigma) D_g(\tilde{x}^j, x^j) \leq D_g(x^*, x^j), \tag{44}$$

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

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

$$v^j = \gamma_j[g'(x^j) - g'(\tilde{x}^j)] + e^j, \tag{45}$$

so that (28) can be rewritten as

$$0 = \gamma_j^{-1} v^j + g'(x^{j+1}) - g'(x^j). \tag{46}$$

Replacing (28) in (27), we obtain

$$D_g(\tilde{x}^j, x^{j+1}) \leq \sigma D_g(\tilde{x}^j, x^j). \tag{47}$$

Note that

$$\begin{aligned} &D_g(x^*, x^{j+1}) = D_g(x^*, x^j) + \langle g'(x^j) - g'(x^{j+1}), x^* - \tilde{x}^j\rangle + D_g(\tilde{x}^j, x^{j+1}) - D_g(\tilde{x}^j, x^j) \\ &= D_g(x^*, x^j) + \langle \gamma_j^{-1} v^j, x^* - \tilde{x}^j\rangle + D_g(\tilde{x}^j, x^{j+1}) - D_g(\tilde{x}^j, x^j) \\ &\leq D_g(x^*, x^j) - \gamma_j^{-1}\langle v^j, \tilde{x}^j - x^*\rangle - (1-\sigma) D_g(\tilde{x}^j, x^j), \end{aligned} \tag{48}$$

using the definition of D_g in the first equality, (46) in the second equality and (47) in the inequality. We have proved the leftmost inequality in (44).

Since D_g is nonnegative and $\sigma \in [0,1)$, we obtain from (48)

$$D_g(x^*, x^{j+1}) \leq D_g(x^*, x^j) - \gamma_j^{-1}\langle v^j, \tilde{x}^j - x^*\rangle. \tag{49}$$

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

$$f(\tilde{x}^j, y) + \gamma_j^{-1}\langle -v^j, y - \tilde{x}^j\rangle \geq 0, \tag{50}$$

for all $y \in K$. In particular, (50) 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 5.2, 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 (49), we obtain the rightmost inequality in (44). □

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

Proposition 5.7. *Consider* $\mathrm{EP}(f,K)$. *Assume that f satisfies P1-P4 and also any one among P4', P4" and P4*. Take $g : B \to \mathbb{R}$ satisfying H1-H5, and an exogenous sequence $\{\gamma_j\} \subset (\theta, \bar{\gamma}]$, where θ is the undermonotonicity constant in P4. Let $\{x^j\}$ be the sequence generated by Algorithm II. If* $\mathrm{EP}(f,K)$ *has solutions, then*

i) For all $x^* \in \mathrm{S}(f,K)$, $D_g(x^*, x^j)$ is nonincreasing and convergent.

ii) $\{x^j\}$ is bounded.

iii) $\sum_{j=0}^{\infty} \gamma_j^{-1}\langle v^j, \tilde{x}^j - x^*\rangle < \infty$ where v^j given by (45).

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) $\{\tilde{x}^j - x^j\}$ converges strongly to 0, and consequently $\{\tilde{x}^j\}$ is bounded,

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

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

Proof. Take $x^* \in \mathrm{S}(f,K)$. By Proposition 5.6, $\{D_g(x^*,x^j)\}$ is a nonnegative and nonincreasing sequence, henceforth convergent, and $\{x^j\}$ is contained in a level set of $D_g(x^*,\cdot)$, which is bounded by H1, establishing (i)-(ii). Invoking again Proposition 5.6,

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

from which (iii) and (iv) follow easily. Item (v) follows from (iv) and (47). 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 $\{x^j\}$ is bounded by (i), we can apply Proposition 2.6 to obtain the strong convergence of $\{\tilde{x}^j - x^j\}$ and $\{\tilde{x}^j - x^{j+1}\}$ to 0, which entails the strong convergence of $\{x^j - x^{j+1}\}$ to 0. Finally, (viii) is obtained from (vii) and (46) taking limits with $j\to\infty$, using the facts that H3 holds and $\{\gamma_j\}\subset(\theta,\bar{\gamma}]$. □

The following theorem completes the convergence analysis of Algorithm II.

Theorem 5.8. *Consider* $\mathrm{EP}(f,K)$. *Assume that f satisfies P1-P4 and also any one among P4', P4" and P4*. Take $g: B\to\mathbb{R}$ satisfying H1-H5, and an exogenous sequence $\{\gamma_j\}\subset(\theta,\bar{\gamma}]$, where θ is the undermonotonicity constant in P4. Let $\{x^j\}$ be the sequence generated by Algorithm II. If* $\mathrm{EP}(f,K)$ *has solutions, then*

- *i)* $\{\tilde{x}^j\}$ *is an asymptotically solving sequence for* $\mathrm{EP}(f,K)$.
- *ii) If $f(\cdot,y)$ is weakly upper semicontinuous for all $y\in K$, then all cluster points of $\{x^j\}$ solve* $\mathrm{EP}(f,K)$.
- *iii) If in addition g satisfies H6, then the whole sequence $\{x^j\}$ is weakly convergent to some solution x^* of* $\mathrm{EP}(f,K)$.

Proof. The proof is similar to the one of Theorem 5.5, using now Proposition 5.7 instead of Proposition 5.4. □

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

Now we look at the two additional conditions imposed in items (ii) and (iii) of Theorems 5.5 and 5.8. Not much can be done about assumption H6 on g, required for uniqueness of the weak limit points of $\{x^j\}$: 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 [7]). The situation is different in connection with weak upper semicontinuity of $f(\cdot,y)$, required for establishing that the weak limit points of $\{x^j\}$ 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 but 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.

6 A Reformulation of the Equilibrium Problem

We will establish here that $\mathrm{S}(f, K)$ coincides with the set of zeroes of a certain point-to-set operator. We will prove then that when f is monotone the graph of this operator has a certain property, namely *demiclosedness*, which enables us to get rid of the weak upper semicontinuity of $f(\cdot, y)$ in the convergence analysis of our algorithms.

We denote as $N_K : B \to \mathcal{P}(B^*)$ the normal operator of K, i.e. the subdifferential of the indicator function I_K, which vanishes at points of K and takes the value $+\infty$ outside K. The explicit formula of N_K given by

$$N_K(x) = \begin{cases} \{z \in B^* | \ \langle z, y - x \rangle \leq 0, \ \forall y \in K\} & \text{if } x \in K \\ \emptyset & \text{otherwise.} \end{cases} \tag{51}$$

We also assume that $\partial F_x(y) \neq \emptyset$ for all $x, y \in K$. This is the case, for instance, if $f(x, \cdot)$ can be extended, preserving its convexity, to some open subset W of B, containing K, for all $x \in K$. We associate to $\mathrm{EP}(f, K)$ the operator $T^f : B \to \mathcal{P}(B^*)$ defined as

$$T^f(x) = \partial F_x(x) + N_K(x), \tag{52}$$

where F_x is as in (13), i.e. $F_x(y) = f(x, y)$, and $\partial F_x(y)$ denotes its subdifferential at the point y.

Proposition 6.1. *The set of zeroes of T^f is equal to $\mathrm{S}(f, K)$.*

Proof. By definition, we have that $x^* \in \mathrm{S}(f, K)$ if and only if $F_{x^*}(x^*) = f(x^*, x^*) = 0 \leq f(x^*, y) = F_{x^*}(y)$ for all $y \in K$. So, $x^* \in \mathrm{S}(f, K)$ if and only if x^* minimizes the function F_{x^*} over K. Since F_{x^*} and K are convex, 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 (51), that is precisely equivalent to saying that $0 \in \partial F_{x^*}(x^*) + N_K(x^*) = T^f(x^*)$. □

Corollary 6.2. *Consider the sequence $\{\tilde{x}^j\}$ generated by either Algorithm I or II. Then $\tilde{x}^j$ is the zero of $T^{f_j^e}$, as defined in* (52)*, with f_j^e as in* (20)*, i.e.*

$$e^j + \gamma_j[g'(x^j) - g'(\tilde{x}^j)] \in \partial F_{\tilde{x}^j}(\tilde{x}^j) + N_K(\tilde{x}^j). \tag{53}$$

Proof. Since $\tilde{x}^j$ is the solution of $\mathrm{EP}(f_j^e, K)$, by Proposition 6.1 it is a zero of $T^{f_j^e}$. The rules of subdifferential calculus 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 (53), 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 6.2 allows us to rewrite the iterative step of our algorithms. For the case of Algorithm I, Step 2 is equivalent to finding a pair $(\tilde{x}j, e^j)$ such that

$$e^j + \gamma_j[g'(x^j) - g'(\tilde{x}^j)] \in \partial F_{\tilde{x}^j}(\tilde{x}^j) + N_K(\tilde{x}^j), \tag{54}$$

$$\left\|e^j\right\|_* \leq \begin{cases} \sigma\gamma_j D_g(\tilde{x}^j, x^j) & \text{if } \left\|x^j - \tilde{x}^j\right\| < 1 \\ \sigma\gamma_j \nu_g(x^j, 1) & \text{if } \left\|x^j - \tilde{x}^j\right\| \geq 1. \end{cases} \tag{55}$$

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

$$U^f(x) = \begin{cases} \partial F_x(x) & \text{if } x \in K \\ \emptyset & \text{otherwise.} \end{cases} \tag{56}$$

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 granted "a priori", but we have the following elementary result.

Proposition 6.3. *If f satisfies P1-P3 and P4$^\bullet$, then U^f is monotone.*

Proof. In view of (56), 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

$$\langle u, y - x\rangle \leq F_x(y) - F_x(x) = f(x,y) - f(x,x) = f(x,y), \tag{57}$$

$$\langle v, x - y\rangle \leq F_y(x) - F_y(y) = f(y,x) - f(y,y) = f(y,x). \tag{58}$$

Adding (57) and (58) we get

$$-\langle u - v, x - y\rangle = \langle u, y - x\rangle + \langle v, x - y\rangle \leq f(x,y) + f(y,x) \leq 0, \tag{59}$$

using P4$^\bullet$ in the last inequality. We conclude from (59) 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 3.1.

The following result has been established in Theorem 4.5.7 of [6] for the case of $g(x) = (1/2)\|x\|^2$, $\lambda = 1$. See also Remark 10.8 in [31].

Proposition 6.4. *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

$$b = v + \lambda g'(z) \in T(x) + \lambda g'(x) \subseteq \bar{T}(x) + \lambda g'(x). \tag{60}$$

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

$$b = v + \lambda g'(z) \in \bar{T}(z) + \lambda g'(z). \tag{61}$$

Since g' is strictly monotone by H2, the same holds for $\bar{T} + \lambda g'$, and it follows therefore from (60) and (61) that $x = z$. Thus, making $x = z$ in the first inclusion of (60), 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 Propositions 3.1 and 6.4 to establish maximal monotonicity of T^f.

Proposition 6.5. *If f satisfies P1-P3, P4$^\bullet$ and g satisfies H1-H2, H4-H5, then T^f, as defined in* (52)*, is maximal monotone.*

Proof. Consider $T^f = U^f + N_K$ as in (52). We want to use Proposition 6.4, 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 6.3. 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 (10) with $e = 0$, $\bar{x} \in B$ and $\gamma = \lambda$ satisfying $g'(\bar{x}) = \lambda^{-1} b$, i.e.

$$\tilde{f}(x, y) = f(x, y) + \lambda\langle g'(x) - \lambda^{-1} b, y - x\rangle. \tag{62}$$

Note that such an $\bar{x}$ exists by H4, and that $\mathrm{EP}(\tilde{f}, K)$ has a unique solution by Proposition 3.1, say $\hat{x}$. Since $\tilde{f}(x, y) \geq 0$ for all $y \in K$ and $\tilde{f}(x, x) = 0$, $\hat{x}$ solves the following convex minimization problem: $\min \tilde{f}(x, \cdot)$ s.t. $y \in K$. Thus, $\hat{x}$ satisfies the first order optimality condition for this problem, which is, in view of (62) 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 6.4 to conclude that T^f is maximal monotone. □

We introduce now the demiclosedness property announced above.

Definition 6.6. *Given $T : B \to \mathcal{P}(B^*)$, the graph of T is said to be demiclosed, when the following property holds: if $\{x^j\} \subset B$ is weakly convergent to $x^* \in B$, $\{v^j\} \subset B^*$ is strongly convergent to $v^* \in B^*$ and $v^j \in T(x^j)$ for all j, then $v^* \in T(x^*)$.*

Proposition 6.7. *If $T : B \to \mathcal{P}(B^*)$ is maximal monotone, then its graph is demiclosed.*

Proof. See, e.g., [27], p. 105. □

Now we can get rid of the weak upper semicontinuity assumption in Theorems 5.5, 5.8.

Theorem 6.8. *Assume that*

i) f satisfies P1, P2, P3, and P4$^\bullet$,

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

iii) $\{\gamma_j\}$ is contained in $(0, \bar{\gamma}]$ for some $\bar{\gamma} > 0$,

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

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

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

Proof. We are within the assumptions of Propositions 5.4, 5.7 with $\theta = 0$. Clearly, P4$^\bullet$ is the same as P4 with $\theta = 0$, and, as discussed in Section 2, P4$^\bullet$ implies P4*, for instance. Consider first Algorithm I. The sequence $\{x^j\}$ is bounded by Proposition 5.4(ii). Let $\bar{x}$ be a cluster point of $\{x^j\}$. Let $\{x^{\ell_j}\}$ be a subsequence of $\{x^j\}$ weakly convergent to $\bar{x}$. By Proposition 5.4(v), the subsequence $\{\tilde{x}^{\ell_j}\}$ is also weakly convergent to $\bar{x}$. In view of Corollary 6.2, (23) and (52), we have $v^{\ell_j} \in T^f(\tilde{x}^{\ell_j})$. By Proposition 5.4(vi), $\{v^{\ell_j}\}$ is strongly convergent to 0. Since T^f is maximal monotone by Proposition 6.5, its graph is demiclosed by Proposition 6.7. It follows from Definition 6.6 that $0 \in T^f(\bar{x})$, and therefore $\bar{x} \in \mathrm{S}(f, K)$, in view of Proposition 6.1. Uniqueness of the weak cluster point of $\{x^j\}$ when g satisfies H6 follows exactly as in the proof of Theorem 5.5(iii). The case of Algorithm II is dealt with in a similar way, invoking now Proposition 5.7 instead of Proposition 5.4. □

We remark that the difference between Theorems 5.5, 5.8 on one side, and Theorem 6.8, 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, Algorithms I and II coincide with Algorithms I and II in [15], designed for finding zeroes of maximal monotone operators in Banach spaces. Thus, we could have omitted the proof of Theorem 6.8, which has been included just for making this paper more self-contained. On the other hand, the results in [15] demand monotonicity of the operator, akin to monotonicity of f in the case of equilibrium problems, while in Section 5 we worked under the weaker assumptions of pseudomonotonicity and θ-undermonotonicity (cf. Example 2.8). An inexact proximal point method for finding zeroes of non-monotone operators in Banach spaces has appeared in [14], but in this reference the operator is assumed to be θ-*hypomonotone*. In the context of equilibrium problems, this is the same as stating that $(T^f)^{-1} + \theta g'$ is monotone, 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 [14] and in this paper are different. A thorough discussion on the connection between our algorithms and the use of proximal point methods for finding zeroes of operators applied to solving equilibrium problems via the reformulation can be found in [20].

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 $\{x^{j,k}\}_{k\in\mathbb{N}}$. 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 (54), (55), for the case of Algorithm I one should verify whether there exists some e^j such that the pair $(x^{j,k}, e^j)$ satisfies

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

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

If so, we take $\tilde{x}^j = x^{j,k}$ and continue with the computation fo 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 continuously Fréchet differentiable and that the boundary ∂K of K is smooth, in the following sense:

Definition 6.9. *We say that the boundary of a closed and convex set K, denoted by ∂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 6.10. *If the boundary of K is smooth then $N_K(x) = \{th'(x) : t \geq 0\}$ for all $x \in \partial K$.*

Proof. Take $x \in \partial K$. The fact that the halfline through $h'(x)$ is contained in $N_K(x)$ follows easily from (51) and convexity of h. For the reverse inclusion, assume that there exists $w \in N_K(x)$ such that w and $h'(x)$ are linearly independent. It follows from the Convex Separation Theorem that there exists $z \in B$ such that

$$\langle h'(x), z\rangle < 0, \tag{63}$$

$$\langle w, z\rangle > 0, \tag{64}$$

because otherwise Ker $w =$ Ker $h'(x)$, contradicting the linear independence of w and $h'(x)$ (cf. Lemma 3.9 in [30]). Let $y = x + tz$. It follows from (63) and differentiability of h that $y \in K$ for small enough $t > 0$, so that, in view of (51), $0 \geq \langle w, y - x\rangle = t\langle w, z\rangle$, which contradicts (64). □

Theorem 6.11. *Under the hypotheses of Theorems 5.5, 5.8, assume additionally that K has smooth boundary, F_x is continuously Fréchet differentiable and that $(g')^{-1}$ is continuous (for Algorithm II). Let $\{x^j\}$ be the sequence generated either by Algorithm I or II. Assume that x^j is not a solution of* $\mathrm{EP}(f, K)$ *and let $\hat{x}^j$ be the unique solution of* $\mathrm{EP}(f_j, K)$ *with f_j as defined in (5). 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 (21)-(22) in the case of Algorithm I and (26)-(27) in the case of Algorithm II,*
- *ii) if $\hat{x}^j$ belongs to ∂K, then any $x \in B(\hat{x}^j, \delta_j) \cap \partial K$ solves the subproblem (21)-(22) in the case of Algorithm I and (26)-(27) in the case of Algorithm II.*

Proof. Note that if we take $e^j = 0$, then $\mathrm{EP}(f_j^e, K)$ reduces to $\mathrm{EP}(f_j, K)$. If $\{x^k\}$ is a sequence generated by either Algorithm I or II, define $\Lambda_j : B \to B^*$ as

$$\Lambda_j(x) = -F_x'(x) - \gamma_j[g'(x) - g'(x^j)], \tag{65}$$

where F_x' denotes the Fréchet derivative of F_x.
i) We first consider the case of Algorithm I. Note that $\hat{x}^j$ is the unique solution of $\mathrm{EP}(f_j, K)$ by Proposition 3.1. 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,

$$\Phi_j(\hat{x}^j) > 0. \tag{66}$$

Assume that $\hat{x}^j \in \mathrm{int}(K)$. In this case, 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 (53) with $e^j = 0$, $N_K(\hat{x}^j) = 0$, which implies, in view of (65), that

$$\Lambda(\hat{x}^j) = 0. \tag{67}$$

Define $\Psi_j : K \to \mathbb{R}$ as $\Psi_j(x) = \|\Lambda_j(x)\|_* - \Phi_j(x)$. By (66), (67),

$$\Psi_j(\hat{x}^j) = -\Phi_j(\hat{x}^j) < 0. \tag{68}$$

Continuity of F'_x and g' ensure continuity of both Λ_j and Φ_j, and therefore of Ψ_j as well. By (67) and continuity of Ψ_j we can choose the above δ_j small enough such that $\Psi_j(x) \le 0$ for all $x \in B(\hat{x}^j, \delta_j) \cap K$ or equivalently $\|\Lambda_j(x)\|_* \le \Phi_j(x)$ for all $x \in B(\hat{x}^j, \delta_j) \cap K$. In such a case, any pair (x, e) with $x \in B(\hat{x}^j, \delta_j)$, $e = \Lambda_j(x)$, will satisfy (54), (55), 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 I.

For the case of Algorithm II, 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 I. Assume that $\hat{x}^j \in \partial K$. Since $\hat{x}^j$ is the exact solution of the j-th subproblem, it satisfies (53) with $e^j = 0$, so that, in view of (65), $\Lambda(\hat{x}^j) \in N_K(\hat{x}^j)$. Since ∂K is smooth, we get from Proposition 6.10 that $N_K(\hat{x}^j) = \{th'(\hat{x}^j) : t \ge 0\}$, so that there exists $t^* \ge 0$ such that $\Lambda_j(\hat{x}^j) = t^* h'(\hat{x}^j)$. Since $h'(\hat{x}^j) \neq 0$ by Definition 6.9, we get $t^* = \left\|\Lambda_j(\hat{x}^j)\right\|_* / \left\|h'(\hat{x}^j)\right\|_*$ and therefore

$$\frac{\left\|\Lambda_j(\hat{x}^j)\right\|_*}{\|h'(\hat{x}^j)\|_*} h'(\hat{x}^j) - \Lambda_j(\hat{x}^j) = 0. \tag{69}$$

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 (69). Therefore, we can choose $\delta_j > 0$ such that $\hat{\Psi}_j(x) \le 0$ for all $x \in B(\hat{x}^j, \delta_j) \cap \partial K$ or equivalently

$$\left\| \frac{\|\Lambda_j(x)\|_*}{\|h'(x)\|_*} h'(x) - \Lambda_j(x) \right\|_* \le \Phi_j(x), \tag{70}$$

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 (70) that this pair satisfies (22) and (53), and henceforth (21) (with x instead of $\tilde{x}^j$), establishing the result.

For the case of Algorithm II, 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. $\square$

The following example illustrates the fact that if ∂K is not smooth then the result of Theorem 6.11 fails to hold. We consider only Algorithm I (a similar example can be constructed for Algorithm II).

Example 6.12. *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 $g(x) = \frac{1}{2} \|x\|_2^2$. In fact, one can easily show that f is 1-undermonotone. It is clear that ∂K is not smooth. We also have that $\mathrm{S}(f, K) = \{(1, 1)\}$. 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 $\mathrm{S}(f_0, K) = \{(1, 1)\}$. Take $e^0 = (e_1^0, e_2^0) \in \mathbb{R}^2$, so that $f_0^e(x, y) = \langle \frac{10}{9} x^0 - \frac{1}{9} x + e^0, x - y \rangle$. We claim that the unique solution of $\mathrm{EP}(f_0^e, K)$ is $(1, 1)$, independently of the error vector e^j, thus falsifying the result of Theorem 6.11. 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 $\|e^0\|_2 \leq \frac{4}{9} D_g((1, 1), (\frac{1}{2}, \frac{1}{2})) = \frac{1}{9}$. 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 $\|e^0\|_2 \leq \frac{1}{9}$.

Acknowledgment

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

References

[1] Bianchi, M., Pini, R. A note on equilibrium problems with properly quasimonotone bifunctions. *Journal of Global Optimization* **20** (2001) 67-76.

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

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

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

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

[6] Burachik, R.S., Iusem, A.N. *Set-Valued Mappings and Enlargements of Monotone Operators*. Springer, Berlin (2007).

[7] 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) 1615-1632.

[8] 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.

[9] Butnariu, D., Iusem, A.N. *Totally convex functions for fixed points computation and infinite dimensional optimization.* Kluwer, Dordrecht (2000).

[10] Fan, K. A minimax inequality and applications. In *Inequality III* (O. Shisha, editor) Academic Press, New York (1972) 103-113.

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

[12] Flores-Bazán, F. Ideal, weakly efficient solutions for convex vector optimization problems. *Mathematical Programming* **93** (2002) 453-475.

[13] Flores-Bazán, F. Existence theory for finite dimensional pseudomonotone equilibrium problems. *Acta Applicandae Mathematicae* **77** (2003) 249-297.

[14] Gárciga Otero, R., Iusem, A.N. Proximal methods in Banach spaces without monotonicity. *Journal of Mathematical Analysis and Applications* **330** (2007) 433-450.

[15] 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.

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

[17] Iusem, A.N., Pennanen, T., Svaiter, B.F. Inexact variants of the proximal point method without monotonicity. *SIAM Journal on Optimization* **13** (2003) 1080-1097.

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

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

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

[21] Kaplan, A., Tichatschke, R. Proximal point methods and nonconvex optimization. *Journal of Global Optimization* **13** (1998) 389-406.

[22] Kassay, G. The proximal point algorithm for reflexive Banach spaces. *Studia Mathematica* **30** (1985) 9-17.

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

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

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

[26] Ortega, J.M. *Numerical Analysis. A Second Course*. Academic Press, New York (1972).

[27] Pascali, D., Sburlan, S. *Nonlinear Mappings of Monotone Type*. Editura Academiei, Bucarest (1978).

[28] Pennanen, T. Local convergence of the proximal point method and multiplier methods without monotonicity. *Mathematics of Operations Research* **27** (2002) 170-191.

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

[30] Rudin, W. *Functional Analysis*. McGraw Hill, New York (1973).

[31] Simons, S. *Minimax and Monotonicity*. Lecture Notes in Mathematics **1693**. Springer, Berlin (1998).

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

[33] 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.