Inexact versions of proximal point and augmented Lagrangian algorithms in Banach spaces

Alfredo Iusem* Rolando Gárciga O.†

October 15, 2001

Abstract

We generalize a proximal-like method for finding zeroes of maximal monotone operators in Hilbert spaces with quadratic regularization due to Solodov and Svaiter, making it possible the use of other kind of regularizations and extending it to Banach spaces. In particular, we introduce an appropriate error criterium to obtain an inexact proximal iteration based on Bregman functions and construct a hyperplane which strictly separates the current iterate from the solution set. A Bregman projection onto this hyperplane is then used to obtain the next iterate. Boundedness of the sequence and optimality of the weak accumulation points are established under suitable assumptions on the regularizing function, which hold for any power greater than one of the norm of any uniformly smooth and uniformly convex Banach space, without any assumption on the operator other than existence of zeroes. These assumptions let us, also, obtain similar results in Banach spaces for the Hybrid Extragradient-Generalized Proximal Point method, proposed by Solodov and Svaiter for finite dimensional spaces. We then transpose such methods to generate augmented Lagrangian methods for $\mathcal{L}^p$-constrained convex optimization problems in Banach spaces, obtaining two alternative procedures which allow for inexact solutions of the primal subproblems. Boundedness of both the primal and the dual sequences, and optimality of primal and dual weak accumulation points, are then established, assuming only existence of Karush-Kuhn-Tucker pairs.

Keywords: proximal point method, augmented Lagrangian, inexact solutions, Bregman projection, maximal monotone operator.

*Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro, RJ, 22460-320, Brazil (**iusp@impa.br**). The work of this author was partially supported by CNPq grant no. 301280/86 and by Marcel Dassault chair, Centro de Modelamiento Matemático, Universidad de Chile.

†Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro, RJ, 22460-320, Brazil (**rgarciga@impa.br**).

1 Introduction

The Proximal Point method, whose origins can be traced back e.g. to [33], [35], was presented in [40] as a method for finding zeroes of monotone operators in Hilbert spaces in the following way: Given a Hilbert space H and a maximal monotone point-to-set operator $T : H \to \mathcal{P}(H)$, take a bounded sequence of regularization parameters $\{\lambda_k\} \subset \mathbb{R}_{++}$ and any initial $x^0 \in H$, and then, given x^k, define x^{k+1} as the only $x \in H$ such that

$$\lambda_k(x^k - x) \in T(x). \tag{1}$$

It was proved in [40] that when T has zeroes the sequence $\{x^k\}$ defined by (1) is weakly convergent to a zero of T. If we have a Banach space B, instead of H, in which case we must take $T : B \to \mathcal{P}(B^*)$, where B^* is the dual of B, then an appropriate extension of (1) is achieved by taking x^{k+1} as the unique $x \in B$ such that

$$\lambda_k[f'(x^k) - f'(x)] \in T(x), \tag{2}$$

where $f : B \to \mathbb{R} \cup \{\infty\}$ is a convex and Gâteaux differentiable function satisfying conditions H1–H4 in Section 2, and f' is its Gâteaux derivative. When $f = ||x||_B^2$ and B is a Hilbert space, (2) reduces to (1).

Under H1–H4, it was proved in [10] that if T has zeroes then the sequence $\{x^k\}$ is bounded and all its weak cluster points are zeroes of T, and that there exists a unique weak cluster point when f' is weak-to-weak* continuous. Functions satisfying H1-H4 exist in most Banach spaces. This is the case, for instance, of $f(x) = ||x||_B^r$ with $r > 1$ where B is any uniformly smooth and uniformly convex Banach space, as we will see in Section 2.

We mention next some previous references on the Proximal Point method in Banach spaces. A scheme similar to (2) was considered in [19], [20], but requesting a condition on f much stronger than H1–H4, namely strong convexity. It was proved in [2] that if there exists $f : B \to \mathbb{R}$ which is strongly convex and twice continuously differentiable at least at one point, then B is hilbertian, so that the analysis in these references holds only for Hilbert spaces. With $f(x) = ||x||_B^2$, the scheme given by (2) was analyzed in [30], where only partial convergence results are given (excluding, e.g., boundedness of $\{x^k\}$). For the case of optimization (i.e. when T is the subdifferential of a convex function $h : B \to \mathbb{R}$, in which case the zeroes of T are the minimizers of h), the scheme given by (2) was analyzed in [1] for the case of $f(x) = ||x||_B^2$ and in [12] for the case of a general f. Consideration of functions other than the square of the norm is convenient because they provide simpler computations in some cases (e.g. $f(x) = ||x||_p^p$ in ℓ_p or $\mathcal{L}^p$).

In finite dimensional or Hilbert spaces, where the square of the norm leads always to simpler computations, the scheme given by (2) with a nonquadratic f has been proposed mainly with penalization purposes: if the problem includes a feasible set C (e.g. constrained convex optimization or monotone variational inequalities) then the use of an f whose domain is C and whose gradient diverges at the boundary of C, forces the sequence $\{x^k\}$ to stay in the interior of C and makes the subproblems unconstrained. Divergence of the gradient precludes validity of H3, and so several additional hypothesis on f are required. These methods, usually known as Proximal Point methods with Bregman functions, have been studied e.g. in [16], [17], [25], [31] for the optimization case and in [8], [21], [26] for finding zeroes of monotone operators. Other variants of the Proximal Point method in finite dimensional spaces, also with nonquadratic regularization terms but with iteration formulae different from (2), can be found in [3], [4], [11], [22], [28] and [29].

The previous discussion refers to the case of exact solution of the subproblems, i.e. when the inclusion given in (2) is exactly solved in order to find x^{k+1}. Inexact variants in Hilbert spaces were considered as early as in [40], where the iteration

$$e^k + \lambda_k(x^k - x^{k+1}) \in T(x^{k+1}) \tag{3}$$

is considered. Here e^k is the error in the k-th iteration, and convergence of the sequence $\{x^k\}$ to a zero of T is ensured under the assumption that $\sum_{k=0}^{\infty} ||x^{k+1} - x|| < \infty$, where x^{k+1} is given by (3) and x is the unique solution of (1). Other related conditions for the case of quadratic f, but including always summability conditions on the error, can be found in [40] and [30] for Hilbert and Banach spaces respectively, and, for the case of nonquadratic f, in [21], [31] for optimization and [9] for variational inequalities. Such summability conditions cannot be verified in actual implementations, and also demand that the precision in the computations increase with the iteration index k. New error schemes, accepting constant relative errors, and introducing computationally checkable inequalities such that any vector which satisfies them can be taken as the next iterate, have been recently presented in [42] for the case of quadratic f in Hilbert spaces and in [43], [11] for the case of nonquadratic f in finite dimensional spaces (in [11], with a regularization different from the scheme given by (2)). We discuss next the error schemes in [42], [43], which we will extend to Banach spaces in Section 3.

In [42], the approximate solution of (3) is not taken as the next iterate, but as an auxiliary point $\tilde{x}^k$, i.e. $\tilde{x}^k$ is any vector such that

$$e^k + \lambda_k(x^k - \tilde{x}^k) \in T(\tilde{x}^k), \tag{4}$$

where the error term e^k satisfies

$$||e^k|| \leq \sigma\lambda_k \max\{||x^k - \tilde{x}^k + \lambda_k^{-1}e^k||, ||x^k - \tilde{x}^k||\}, \tag{5}$$

for some $\sigma \in [0, 1)$. Then the next iterate is obtained as the orthogonal projection of x^k onto the hyperplane $H_k = \{x \in H : \langle v^k, x - \tilde{x}^k\rangle = 0\}$ with $v^k = \lambda_k(x^k - \tilde{x}^k) + e^k$, i.e.

$$x^{k+1} = x^k - \frac{\langle v^k, x^k - \tilde{x}^k\rangle}{||v^k||^2} v^k. \tag{6}$$

The algorithm in [43] requires the definition of the *Bregman distance* D_f :dom $f\times$ int(dom f) $\to \mathbb{R}$, given by

$$D_f(x, y) = f(x) - f(y) - \langle f'(y), x - y\rangle. \tag{7}$$

The derivative f' is required to be onto, the auxiliary point $\tilde{x}^k$ must satisfy

$$e^k + \lambda_k[f'(x^k) - f'(\tilde{x}^k)] \in T(\tilde{x}^k), \tag{8}$$

and the error criterium is given by

$$D_f\big(\tilde{x}^k, (f')^{-1}\big[f'(\tilde{x}^k) - \lambda_k^{-1}e^k\big]\big) \le \sigma D_f(\tilde{x}^k, x^k), \tag{9}$$

for some $\sigma \in [0,1)$, where $(f')^{-1}$ denotes the inverse function of f'. Then the next iterate x^{k+1} is given by

$$x^{k+1} = (f')^{-1}\big[f'(\tilde{x}^k) - \lambda_k^{-1}e^k\big]. \tag{10}$$

It has been proved in [42] and [43] that the sequences $\{x^k\}$ defined by (4)–(6) and (8)–(10) respectively, converge to a zero of T whenever T has zeroes. When $e^k = 0$, which occurs only when $\tilde{x}^k$ is the exact solution of (1) or (2), both (5) and (9) hold with strict inequality for any $\tilde{x}^k \neq x^k$, so that these inequalities can be seen as Armijo-type rules which will be satisfied by any vector close enough to the exact solution of the subproblem. The constant σ can be interpreted as a relative error, measuring the ratio of a measure of proximity between the candidate point $\tilde{x}^k$ and the exact solution, and a measure of proximity between the same candidate point and the previous iterate.

The augmented Lagrangian method for equality constrained optimization problems (nonconvex, in general) was introduced in [23], [37]. Its extension to inequality constrained problems started with [15] and was continued in [5], [32], [39]. Its connections with the Proximal Point Method are treated in [41], [4], [6], [21] and [28]. A method for the finite dimensional case, with $f(x) = ||x||^2$ and inexact solution of the subproblems in the spirit of [42], [43] appears in [24], which is also a starting point for our work.

Augmented Lagrangian methods for constrained convex optimization in Banach spaces were introduced in [13] in the following way. Given a Banach space B, a measure space $(\Omega, \mathcal{A}, \mu)$ and functions $g : B \to \mathbb{R}$, $G : B \times \Omega \to \mathbb{R}$, such that $g(\cdot)$ and $G(\cdot, \omega)$ are convex and differentiable for all $\omega \in \Omega$, and both $G(x, \cdot)$ and its derivative with respect to the first argument $G'(x, \cdot)$ belong to $\mathcal{L}^p(\Omega)$ for all $x \in B$, we consider problem (P), consisting of minimizing $g(x)$ subject to $G(x, \omega) \le 0$ for almost all $\omega \in \Omega$. This formulation allows both for a finite and an infinite number of constrains, according to the cardinality of Ω.

The method introduced in [13] requires a bounded sequence $\{\lambda_k\} \subset \mathbb{R}_{++}$ and a function $f : \mathcal{L}^q(\Omega) \to \mathbb{R}$ satisfying assumptions H1–H4, with $1/p + 1/q = 1$ (for the sake of simplicity, we present in the following formulae only the case corresponding to $f(y) = \frac{1}{q}||y||_q^q$). The augmented Lagrangian $\bar{L} : B \times \mathcal{L}^q(\Omega) \times \mathbb{R}_{++} \to \mathbb{R}$ is defined as

$$\bar{L}(x, y, \rho) = g(x) + \frac{\rho}{p}\left\|\max\{0, M(x, y, \rho)\}\right\|_p^p, \tag{11}$$

where $M : B \times \mathcal{L}^q(\Omega) \times \mathbb{R}_{++} \to \mathcal{L}^p(\Omega)$ is defined as $M(x, y, \rho)(\omega) = f'(y)(\omega) + \rho^{-1}G(x, \omega)$, and it generates sequences $\{x^k\} \subset B$, $\{y^k\} \subset \mathcal{L}^q(\Omega)$, starting from any $y^0 \in \mathcal{L}^q(\Omega)$ such that $y^0(\omega) \geq 0$ a.e., through the formulae

$$x^{k+1} \in \operatorname{argmin}_{x \in B} \bar{L}(x, y^k, \lambda_k), \tag{12}$$

$$y^{k+1}(\omega) = \left[\max\{0, M(x^{k+1}, y^k, \lambda_k)(\omega)\}\right]^{p-1}. \tag{13}$$

Algorithm (12)–(13) is analyzed with the help of the dual problem (D), defined as max $\Phi(y)$ subject to $y(\omega) \geq 0$ a.e., where $\Phi : \mathcal{L}^q(\Omega) \to \mathbb{R} \cup \{-\infty\}$ is defined as $\Phi(y) = \inf_{x \in B} L(x, y)$, with the Lagrangian $L : B \times \mathcal{L}^q(\Omega) \to \mathbb{R}$ given by

$$L(x, y) = g(x) + \int_\Omega G(x, \omega) y(\omega) d\mu(\omega).$$

Again, as in the case of its finite dimensional counterpart, coercivity assumptions on the data functions are required to ensure existence of the iterates x^k, and boundedness of the sequence $\{x^k\}$. It is proved in [13] that the sequence $\{y^k\}$ generated by (12)–(13) coincides with the sequence generated by (2) with $T = \partial(-\Phi)$ and $f(y) = \frac{1}{q}||y||_q^q$, starting from the same y^0. It follows from the results in [10] or [12] on the convergence of the Proximal Point method in Banach spaces, and the duality results in [13], that $\{y^k\}$ is bounded and all its weak cluster points are solutions of (D). Based upon this result, it is proved in [13] that all weak cluster points of $\{x^k\}$, if any, solve (P).

In this paper we will extend the results in [10] and [13] in the following way. First we will present two Proximal Point methods for finding zeroes of maximal monotone operators in Banach spaces with inexact solutions of the subproblems which allow for constant relative errors, thus extending to Banach spaces the algorithms presented in [42] and [43]. Then, in Section 4, we will translate these methods into the setting of augmented Lagrangian methods in Banach spaces, i.e. for solving problem (P) above, allowing again for inexact solution of the subproblems in (12), and adding also a regularization term to $\bar{L}$ as given by (11), which will allow us to establish existence and boundedness of the primal sequence $\{x^k\}$, without coercivity requirements on the problem data. We will prove, in Section 5, that both the primal and the dual sequence are bounded and that all their weak cluster points are solutions of (P) and (D) respectively.

2 On the existence of appropriate regularizing functions in general Banach spaces

From now on, B is a reflexive Banach space with norm denoted by $||\cdot||$ or $||\cdot||_{\mathrm{B}}$, B^* its topological dual (with the operator norm denoted by $\|\cdot\|_*$ or $\|\cdot\|_{\mathrm{B}^*}$) and $\langle\cdot,\cdot\rangle$ denotes the dual product in $B^* \times B$ (i.e $\langle \phi, x\rangle = \phi(x)$ for all $\phi \in B^*$ and all $x \in B$). Given Banach spaces B_1, B_2, we consider the Banach space $B_1 \times B_2$ with the product norm $\|(x, y)\|_{B_1 \times B_2} = \|x\|_{B_1} + \|y\|_{B_2}$.

In this section, $f : B \to \mathbb{R} \cup \{\infty\}$ is strictly convex, lower semicontinuous and Gâteaux differentiable in the interior of its domain, and f' is its Gâteaux derivative. We will denote the family of such functions as $\mathcal{F}$. We will use in the sequel the *modulus of total convexity* ν_f :int(dom $f) \times \mathbb{R}_+ \to \mathbb{R}$ defined as

$$\nu_f(x,t) = \inf_{y \in \{y \in B : ||y-x||=t\}} D_f(y,x), \tag{14}$$

with D_f as in (7). The methods we analyze in this paper use, as an auxiliary device, functions $f \in \mathcal{F}$ which satisfy some or all of the following assumptions:

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

H2: $\inf_{x \in C} \nu_f(x,t) > 0$, for all bounded set $C \subset$ int(dom f) and all $t \in \mathbb{R}_{++}$.

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

H4: For all $y \in B^*$, there exists $x \in$ int(dom f) such that $f'(x) = y$.

H5: f' : int(dom $f) \to B^*$ is weak-to-weak* continuous.

It is important to exhibit functions which satisfy these properties in as large a class of Banach spaces as possible, and we focuse our attention on functions of the form $f_r(x) = \frac{1}{r}||x||_B^r$ with $r > 1$, for which we recall next some properties of duality mappings in Banach spaces.

A *weight function* φ is a continuous and strictly increasing function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\varphi(0) = 0$ and $\lim_{t \to \infty} \varphi(t) = \infty$. The *duality mapping of weight* φ is the mapping $J_\varphi : B \to B^*$, defined by $J_\varphi x = \{x^* \in B^* | \langle x^*, x\rangle = \|x^*\|_* ||x||, \|x^*\|_* = \varphi(||x||)\}$ for all $x \in B$. When $\varphi(t) = t$, J_φ is called the *normalized duality mapping* and denoted by J. If J_φ is the duality mapping of weight φ in a reflexive Banach space B, then the following properties hold:

J1: $J_\varphi x = \partial \psi(||x||)$ for all $x \in B$, where $\psi(t) = \int_0^t \varphi(s) ds$.

J2: If there exist φ^{-1}, the inverse function of the weight φ, then φ^{-1} is a weight function too and $J^*_{\varphi^{-1}}$, the duality mapping of weight φ^{-1} on B^*, is such that $x^* \in J_\varphi x$ if and only if $x \in J^*_{\varphi^{-1}} x^*$.

J3: If J_1 and J_2 are duality maps of weights φ_1 and φ_2 respectively, then, for all $x \in B$, $\varphi_2(||x||) J_1 x = \varphi_1(||x||) J_2 x$.

J4: If B is uniformly smooth, then J is uniformly continuous on bounded subsets of B.

Properties J1–J3 have been proved in [18] and property J4 in [44]. We need to show that J4 holds for duality mappings with an arbitrary weight φ, which can be easily proved, but we found no reference for it, and therefore we present next a proof.

Proposition 1. *Let B be a uniformly smooth Banach space and J_φ a duality mapping of weight φ on B. Then J_φ is uniformly continuous on bounded subsets of B (considering the strong topology both in B and B^*).*

Proof. If the result does not hold, then we can find a bounded subset A of B such that J_φ is not uniformly continuous on A, i.e. there exist $\epsilon > 0$, and sequences $\{x^k\}$, $\{y^k\} \subset A$ such that $x^k - y^k \xrightarrow[k\to\infty]{\mathrm{s}} 0$ and $\|J_\varphi x^k - J_\varphi y^k\|_{\mathrm{B}^*} \geq \epsilon$ for all k.

Note, first, that if $\lim_{k\to\infty} ||x^k|| = 0$, then $\lim_{k\to\infty} ||y^k|| = 0$ too, and from the definition of J_φ it follows that $\lim_{k\to\infty} \|J_\varphi x^k - J_\varphi y^k\|_* = 0$, which is a contradiction. Thus, we can assume, without loss of generality, that $0 < m \leq ||x^k||, ||y^k|| \leq M$, for some m, $M \in \mathbb{R}$. Let $\bar\varphi(t) = \varphi(t)/t$, and $\bar M = \max_{t\in[m,M]} \bar\varphi(t)$. Using J3 and the definition of J, we have

$$\epsilon \leq \|J_\varphi x^k - J_\varphi y^k\|_* = \left\| \frac{\varphi(\|x^k\|)}{\|x^k\|} J x^k - \frac{\varphi(\|y^k\|)}{\|y^k\|} J y^k \right\|_* \leq$$

$$\bar\varphi(\|x^k\|)\|Jx^k - Jy^k\|_* + \left|\bar\varphi(\|x^k\|) - \bar\varphi(\|y^k\|)\right| \|Jy^k\|_* \leq \bar M \|Jx^k - Jy^k\|_* + \left|\bar\varphi(\|x^k\|) - \bar\varphi(\|y^k\|)\right| \|y^k\|.$$

Since $\lim_{k\to\infty} |\bar\varphi(\|x^k\|) - \bar\varphi(\|y^k\|)| = 0$ by uniform continuity of $\bar\varphi$ in $[m, M]$, and $\{y^k\}$ is bounded, it follows that $\|Jx^k - Jy^k\|_* \geq \epsilon/(2\bar M)$ for large enough k, contradicting J4. $\square$

Our results on the validity of H1–H5 for $f_r(x) = \frac{1}{r}||x||_B^r$ are summarized in the following proposition.

Proposition 2. *i) If B is a uniformly smooth and uniformly convex Banach space, then $f_r(x) = \frac{1}{r}||x||_B^r$ satisfies H1, H2, H3 and H4 for all $r > 1$.*

ii) If B is a Hilbert space, then $f_2(x) = \frac{1}{2}||x||^2$ satisfies H5. If $B = \ell_p$ ($1 < p < \infty$) then $f_p(x) = \frac{1}{p}||x||_p^p$ satisfies H5.

Proof. It has been proved in [14] that, for all $r > 1$, f_r is Gâteaux differentiable when B is smooth (i.e. when B^* is strictly convex), and that it satisfies H1 and H2 when B is uniformly convex. Consider the weight function $\varphi(t) = t^{r-1}$. Since $\int_0^t \varphi(s)ds = \frac{1}{r}t^r$, we get from J1 that $J_\varphi = f_r'$ when B is smooth. Proposition 1 and J1 with this weight φ easily imply that f_r satisfies H3 for all $r > 1$. Note that φ is invertible for $r > 1$, so that J_φ is onto by J2. It follows that f_r satisfies H4, and thus (i) holds. For (ii), note that in the case of a Hilbert space, f_2' is the identity, which is certainly weak-to-weak continuous. The result for f_p in ℓ_p has been proved in [7], Proposition 8.2. $\square$

Unfortunately, it has been proved in [13] that for $B = \ell_p$ or $B = \mathcal{L}^p[\alpha, \beta]$ the function $f_r(x) = \frac{1}{r}||x||_p^r$ does not satisfy H5, excepting in the two cases considered in Proposition 2(ii). We remark that, as it will be seen, properties H1–H4 are required for establishing existence and uniqueness of the iterates of the algorithms under consideration, and also boundedness of the generated sequences, while H5 is required only for uniqueness of the weak cluster points of such sequences. We mention also that the factor $\frac{1}{r}$ in the definition of f_r is inessential for Proposition 2, whose results trivially hold for all positive multiples of $||\cdot||_B^r$.

We discuss next some properties of functions satisfying some of the assumptions above. A function $f : B \to \mathbb{R}$ is said to be *totally convex* if $\nu_f(x,t) > 0$ for all $t > 0$ and all $x \in B$. It has been proved in ([13], 1.2.2) that if f is totally convex then, for all $s \geq 1$, all $t \geq 0$ and all $x \in \operatorname{int}(\operatorname{dom} f)$,

$$\nu_f(x, st) \geq s\nu_f(x,t). \tag{15}$$

Proposition 3. *Let B_1, B_2 be real Banach spaces, $f : B_1 \to \mathbb{R}\cup\{\infty\}$ and $h : B_2 \to \mathbb{R}\cup\{\infty\}$ two proper convex functions with int(dom f), int(dom h) $\neq \emptyset$. Define $F : B_1 \times B_2 \to \mathbb{R}\cup\{\infty\}$ as $F(x,y) = f(x) + h(y)$. Then*

i) for any $z = (x,y) \in$ int(dom F) the domain of $\nu_F(z,\cdot)$ is an interval $[0,\tau(z))$ or $[0,\tau(z)]$ with $0 < \tau(z) = \tau_f(x)+\tau_h(y) \leq \infty$, where $\tau_f(x) = \sup\{t \mid t \in$ dom $\nu_f(x,\cdot)\}$ and $\tau_h(y) = \sup\{t \mid t \in$ dom $\nu_h(y,\cdot)\}$. Moreover $\nu_F(z,t) = \inf\{\nu_f(x,s) + \nu_h(y,t-s) \mid s \in [0,t]\}$,

ii) for any $z = (x,y) \in$ int(dom F) and $t \in$ dom $\nu_F(z,\cdot)$

$$\nu_F(z,t) \geq \min\{\nu_f(x,t/2), \nu_h(y,t/2)\},$$

therefore, if both f and h are totally convex then F is totally convex.

iii) For $i = 1,\ldots,5$, if both f and h satisfy Hi then F also satisfies Hi.

Proof. Take $\bar{z} = (\bar{x},\bar{y}) \in \operatorname{int}(\operatorname{dom} F)$ and $t \in \mathbb{R}_+$. By definition of ν_F and D_F ((14) and (7)),

$$\begin{aligned} \nu_F(\bar{z},t) &= \inf\{D_F(z,\bar{z}) \mid \|z-\bar{z}\|_{B_1\times B_2} = t\} \\ &= \inf\{D_f(x,\bar{x}) + D_h(y,\bar{y}) \mid \|x-\bar{x}\|_{B_1} + \|y-\bar{y}\|_{B_2} = t\}. \end{aligned} \tag{16}$$

From Proposition 1.2.2. of [13] we know that the domain of $\nu_f(x,\cdot)$ is an interval $[0,\tau(x))$ or $[0,\tau(x)]$ with $0 < \tau_f(x) \leq \infty$ and the same for the domain of $\nu_h(y,\cdot)$ for some $0 < \tau_h(y) \leq \infty$, which proves the first statement in (i). Since

$$\nu_f(\bar{x}, \|x-\bar{x}\|_{B_1}) \leq D_f(x,\bar{x}) \quad \text{and} \quad \nu_h(\bar{y}, \|y-\bar{y}\|_{B_2}) \leq D_h(y,\bar{y}),$$

we get from (16) that

$$\begin{aligned} \nu_F(\bar{z},t) &\geq \inf\{\nu_f(\bar{x}, \|x-\bar{x}\|_{B_1}) + \nu_h(\bar{y}, \|y-\bar{y}\|_{B_2}) \mid \|x-\bar{x}\|_{B_1} + \|y-\bar{y}\|_{B_2} = t\} \\ &= \inf\{\nu_f(\bar{x},s) + \nu_h(\bar{y},t-s) \mid s \in [0,t]\}. \end{aligned} \tag{17}$$

By (16), and for any $s \in \text{dom}\ \nu_f(\bar{x},\cdot)$ with $t-s \in \text{dom}\ \nu_h(\bar{y},\cdot)$, we have

$$\begin{aligned}
\nu_F(\bar{z},t) &\leq \inf\{D_f(x,\bar{x}) + D_h(y,\bar{y}) \mid \|x-\bar{x}\|_{B_1} = s, \|y-\bar{y}\|_{B_2} = t-s\} \\
&= \inf\left\{D_f(x,\bar{x}) \mid \|x-\bar{x}\|_{B_1} = s\right\} + \inf\left\{D_h(y,\bar{y}) \mid \|y-\bar{y}\|_{B_2} = t-s\right\} \\
&= \nu_f\left(\bar{x},s\right) + \nu_h\left(\bar{y},t-s\right).
\end{aligned}$$

Consequently,

$$\nu_F(\bar{z},t) \leq \inf\{\nu_f\left(\bar{x},s\right) + \nu_h\left(\bar{y},t-s\right) \mid s \in [0,t]\}. \tag{18}$$

Item (i) follows from (17) and (18).

For (ii) choose any $t_1 \in \text{dom}\ \nu_f(\bar{x},\cdot)$ and $t_2 \in \text{dom}\ \nu_h(\bar{y},\cdot)$ with $t_1 + t_2 = t$. Then $t_1 \geq t/2$ or $t_2 \geq t/2$. From Proposition 1.2.2. of [13], we get $\nu_f(\bar{x},t_1) \geq \nu_f(\bar{x},t/2)$ or $\nu_h(\bar{y},t_2) \geq \nu_h(\bar{y},t/2)$. By nonnegativity of ν_f and ν_h we get $\nu_f(\bar{x},t_1) + \nu_h(\bar{y},t_2) \geq \min\{\nu_f(\bar{x},t/2), \nu_h(\bar{y},t/2)\}$. Thus,

$$\inf\{\nu_f\left(\bar{x},t_1\right) + \nu_h\left(\bar{y},t_2\right) \mid t_1 + t_2 = t\} \geq \min\{\nu_f(\bar{x},t/2), \nu_h(\bar{y},t/2)\}. \tag{19}$$

In view of (i) and (19) (ii) is true.

The proof of (iii) is elementary, using (ii) for the case of H2. □

Proposition 4. *Take $f : B \to \mathbb{R}$ Gâteaux differentiable. If f' is uniformly continuous on bounded sets, then both f and f' are bounded on bounded subsets of B.*

Proof. Let U be a bounded subset of B. Take $\bar{x} \in B$ and $\rho > 0$ such that $U \subseteq B(\bar{x},\rho) = \{x \in B\ : ||x-\bar{x}|| \leq \rho\}$. Take δ such that , for all $x, x' \in B(\bar{x},\rho)$ satisfying $||x'-x|| \leq \delta$, it holds that $\|f'(x) - f'(x')\|_* \leq 1$. Define $\bar{\gamma}$, $\hat{\gamma}$ as $\bar{\gamma} = \|f'(\bar{x})\|_* + 1 + \rho/\delta$, $\hat{\gamma} = \rho\bar{\gamma} + |f(\bar{x})|$. Then, for any $x \in U$, an easy computation shows that $\|f'(x)\|_* \leq \bar{\gamma}$, and, invoking the Mean Value Theorem applied to the restriction of f to the segment between x and $\bar{x}$, it follows that $|f(x)| \leq \hat{\gamma}$. □

Proposition 5. *If $f \in \mathcal{F}$ satisfies H2 then, for all $\{x^k\}$, $\{y^k\} \subset$ int(dom f) such that $\{x^k\}$ (or $\{y^k\}$) is bounded and $\lim_{k\to\infty} D(y^k,x^k) = 0$, it holds that $x^k - y^k \xrightarrow[k\to\infty]{\text{s}} 0$.*

Proof. See [13], Lemma 2.1.2 for the case in which $\{x^k\}$ is bounded. When $\{y^k\}$ is bounded the argument in the proof of Theorem 2.4 in [43] establishes that in such a case $\{x^k\}$ is bounded too, so that the result also holds. □

3 Inexact versions of the Proximal Point method

A point-to-set operator $T : B \to \mathcal{P}(B^*)$ is said to be *monotone* if $\langle w - w', x - x'\rangle \geq 0$, for all x, $x' \in B$ and all $w \in T(x)$, $w' \in T(x')$. A monotone operator is called *maximal* if its *graph*, $G(T) = \{(v,x) \in B^* \times B | v \in T(x)\}$, is not properly contained in the graph of any other monotone operator. Such operators are of interest because a wide variety of problems, such as convex optimization problems and monotone variational inequalities, are particular cases of the following fundamental problem:

$$\text{Find } x \in B \text{ such that } 0 \in T(x). \tag{20}$$

As discussed in Section 1, the Proximal Point method for solving this problem in Hilbert spaces has been extensively studied. This is not the case for nonhilbertian Banach spaces. In fact, there exists a small number of references on this issue ([30], [12], [10]) and only the first one of them considers inexact solutions of the subproblems, which is an issue of high practical relevance, and the main subject of this section.

Our first approach will explore the interpretation, proposed in [42] for Hilbert spaces, of the proximal point method as a projection-type scheme. It is not difficult to check that in the algorithm defined by (2), it holds that $x^{k+1} = \text{argmin}_{x\in H} D_f(x, x^k)$, where $H = \{x \in B | \langle f'(x^k) - f'(x^{k+1}), x - x^{k+1}\rangle = 0\}$, i.e. x^{k+1} is the *Bregman projection* (see [13], 2.1.5), of x^k onto H, which separates x^k from $T^{-1}(0)$. We will generate an inexact generalized proximal iteration in order to construct an appropriate separating hyperplane. A Bregman projection onto this hyperplane is then used to obtain the next iterate. The method is intended for solving problem (20) and requires an exogenous constant $\sigma \in [0, 1]$, an exogenous bounded sequence $\{\lambda_k\} \subset \mathbb{R}_{++}$ and an auxiliary function $f \in \mathcal{F}$ such that dom $f = B$. It is defined as follows:

Algorithm I: Inexact Proximal Point + Bregman Projection Method

1. Choose $x^0 \in B$.

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

$$\lambda_k[f'(x^k) - f'(\tilde{x}^k)] - e^k \in T(\tilde{x}^k), \tag{21}$$

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

$$\|e^k\|_{B^*} \le \sigma\lambda_k \begin{cases} D_f(\tilde{x}^k, x^k) & \text{if } ||x^k - \tilde{x}^k|| < 1, \\ \nu_f(x^k, 1) & \text{if } ||x^k - \tilde{x}^k|| \ge 1. \end{cases} \tag{22}$$

3. Let

$$v^k = \lambda_k[f'(x^k) - f'(\tilde{x}^k)] - e^k. \tag{23}$$

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

$$x^{k+1} = \text{argmin}_{x\in H_k} D_f(x, x^k). \tag{24}$$

In order to compare our error criterium with the one in [42], i.e. (4)–(5), we must consider the specific case where B is a Hilbert space, and $f = \frac{1}{2}||\cdot||^2$, which is the one considered in this reference. In this situation, our error criterium (22) takes the form $\|e^k\| \leq \sigma\lambda_k\frac{1}{2}\min\{||x^k - \tilde{x}^k||^2, 1\}$, a somewhat more restrictive condition than (5), because the right hand side of (5) involves the maximum between $\lambda_k^{-1}\|v^k\|$ and $||x^k - \tilde{x}^k||$. This seems to be the price paid for our extension: Banach instead of Hilbert spaces and a whole class of regularizing functions instead of just the square of the norm. We mention that we preserve the most important feature of (4)–(5), namely a bound σ for the relative error which needs not go to zero, contrasting with tolerances which must converge to zero or even be summable (e.g. [40]). Also, conditions on the distance from the iterates to the unknown exact solutions of the subproblems, as in [30], are avoided.

Concerning the projection step (24), we mention that the existence of $x^{k+1} \in \text{int}(\text{dom } f)$ is ensured by total convexity of f, which follows from H2, and condition dom $f = B$, and also that even though this projection step cannot be performed through a closed formula, in the cases of interest, namely $f_r(x) = ||x||^r$ $(r > 1)$ with uniformly convex B, it reduces to solving a nonlinear equation of the form $\Phi(s) = 0$, where $\Phi : \mathbb{R}_+ \to \mathbb{R}$ is a continuous function given by a closed formula (see [14]). Also, closed formulae for lower bounds of $\nu_f(x,t)$ in terms of $||x||$ and t, for $f(x) = ||x||_p^r$ $(r > 1)$ and $B = \ell_p$ or $B = \mathcal{L}^p(\Omega)$, which allow to establish that an error vector e^k satisfies (22), appear in [27].

In our second approach we extend to Banach spaces the error criterium introduced in [43] as an hybrid extragradient-inexact generalized proximal point method in $\mathbb{R}^n$ (see (9)). This algorithm requires an exogenous constant $\sigma \in [0,1)$, an exogenous bounded sequence $\{\lambda_k\} \subset \mathbb{R}_{++}$ and an auxiliary function $f \in \mathcal{F}$ such that dom $T \subset \text{int}(\text{dom } f)$. It is defined as follows:

Algorithm II: Inexact Proximal Point-Extragradient Method

1. Choose $x^0 \in \text{int}(\text{dom } f)$.

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

$$e^k + \lambda_k[f'(x^k) - f'(\tilde{x}^k)] \in T(\tilde{x}^k), \tag{25}$$

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

$$D_f\big(\tilde{x}^k, (f')^{-1}\big[f'(\tilde{x}^k) - \lambda_k^{-1}e^k\big]\big) \leq \sigma D_f(\tilde{x}^k, x^k). \tag{26}$$

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

$$x^{k+1} = (f')^{-1}\big[f'(\tilde{x}^k) - \lambda_k^{-1}e^k\big]. \tag{27}$$

Before proceeding to the convergence analysis, we make some general remarks. First, the assumption that dom $T \subset \text{int}(\text{dom } f)$, used also in [10], means basically that we are not attempting any penalization efect of the proximal iteration. Second, we remark that assumptions H1–H5 on f will be used in the convergence analysis of our algorithms (we do not require them in the statement of the algorithms in order to isolate which of them are required for each specific result). In this sense, for instance, existence of $(f)^{-1}$, as required in (26), (27), will be a consequence of H4 for any $f \in \mathcal{F}$. We mention that for the case of strictly convex and smooth B and $f(x) = ||x||_B^r$, we have an explicit formula for $(f')^{-1}$, in terms of g', where $g(\cdot) = \frac{1}{s}\|\cdot\|_{B^*}^s$ with $\frac{1}{s} + \frac{1}{r} = 1$, namely $(f')^{-1} = r^{1-s}g'$. In fact, H4 is sufficient for existence of the iterates generated by both algorithms, as the following results show.

Proposition 6. *Let $f \in \mathcal{F}$ such that $domT \subset int(domf)$. If f satisfies H4, then for all $\lambda > 0$ and all $z \in int(dom\ f)$ there exists a unique solution x of the problem $\lambda f'(z) \in T(x) + \lambda f'(x)$.*

Proof. See [10], Corollary 3.1. □

Corollary 1. *Take $f \in \mathcal{F}$ such that $dom\ T \subset int(dom\ f)$. If f satisfies H4, then, for all k and all $\sigma \in [0, 1]$, the inexact proximal subproblems of both Algorithm I (i.e.* (21)-(22)*) and Algorithm II (i.e.* (25)-(26)*) have solutions.*

Proof. Consider first the case of Algorithm I. Choose $\sigma = 0$, so that the right-hand side of (22) vanishes. Hence e^k must be zero too and the proximal subproblem takes the form: find $\tilde{x}^k \in B$ such that $\lambda_k[f'(x^k) - f'(\tilde{x}^k)] \in T(x^k)$, which always has a unique solution by Proposition 6. Of course, the solution provided by this result satisfies (22) for any $\sigma \geq 0$, and so the result holds. The proof is analogous for Algorithm II. □

Next we settle the issue of finite termination.

Theorem 1. *Suppose that Algorithm I (respectively Algorithm II) stops after k steps. Then $\tilde{x}^k$ generated by Algorithm I (respectively Algorithm II) is a solution of* (20)*.*

Proof. Algorithm I stops at the k-th iteration in two cases: if $v^k = 0$, in which case, by (21), $\tilde{x}^k$ is a solution of (20), or if $\tilde{x}^k = x^k$, in which case, by (22), $e^k = 0$, which in turn implies, by (23), $v^k = 0$ and we are back to the first case. Consequently, $\tilde{x}^k$ is a solution of (20). Finite termination in Algorithm II occurs only if $\tilde{x}^k = x^k$, in which case $D_f(\tilde{x}^k, x^k) = 0$, and therefore, by (26), $e^k = 0$, which in turn implies, by (25), $0 \in T(\tilde{x}^k)$, so that $\tilde{x}^k$ is a solution of (20). □

Corollary 1 ensures that exact solutions of (2) satisfy the error criteria in Algorithms I and II, with $e^k = 0$, so our subproblems will certainly have solutions without any assumption on the operator T. Next we show that for both methods, when T is single valued and continuous, there exists a ball around the exact solution of the subproblem such that any point in its intersection with dom T can be taken as an inexact solution. As a consequence, if we solve the inclusion

$$\lambda_k f'(x^k) \in T(x) + \lambda_k f'(x) \tag{28}$$

with any feasible algorithm which generates a sequence strongly convergent to the unique solution of (28), a finite number of iterations of such algorithm will provide an appropriate next iterate both for Algorithm I and II.

Proposition 7. *Assume that $f \in \mathcal{F}$, H4 holds, dom $T \subset$ int(dom f) and T is single valued and continuous. Assume additionally that f is totally convex and Fréchet differentiable for the case of Algorithm I, and that $(f')^{-1}$ is continuous for the case of Algorithm II. Let $\{x^k\}$ be the sequence generated either by Algorithm I or II. If x^k is not a solution of* (20)*, then there exists a ball U_k around the exact solution of* (28) *such that any $x \in U_k \cap$ dom T solves* (21)–(22) *in the case of Algorithm I and* (25)–(26) *in the case of Algorithm II.*

Proof. Let $\bar{x}^k$ be the unique solution x of the inclusion (28). Consider first the case of Algorithm I. Since $x^k \neq \bar{x}^k$, because otherwise $0 \in T(x^k)$, we have $D_f(\bar{x}^k, x^k) > 0$ and, by total convexity of f, $\alpha_k := \sigma\lambda_k \min\{D_f(\bar{x}^k, x^k), \nu_f(x^k, 1)\} > 0$. Let us consider, now, the function $\psi_k : \mathrm{dom}\ T \to \mathbb{R}$ defined as

$$\psi_k(y) = ||T(y) + \lambda_k[f'(y) - f'(x^k)]||_* - \sigma\lambda_k \min\{D_f(y, x^k), \nu_f(x^k, 1)\}.$$

Observe that ψ_k is continuous by continuity of T, and that $\psi_k(\bar{x}^k) = -\alpha_k < 0$, so that there exist $\delta_k > 0$ such that $\psi_k(y) \leq 0$ for all $y \in \mathrm{dom}\ T$ with $||y - \bar{x}^k|| < \delta_k$. Let $U_k := \{y \in B : ||y - \bar{x}^k|| < \delta_k\}$. Then

$$||T(y) + \lambda_k[f'(y) - f'(x^k)]||_* \leq \sigma\lambda_k \min\{D_f(y, x^k), \nu_f(x^k, 1)\} \tag{29}$$

for all $y \in U_k \cap$ dom T. Finally, note that the right hand side of (29) is trivially less than or equal to the right hand side of (22) with y replacing $\tilde{x}^k$.

In the case of Algorithm II, proceed in the same way using Corollary 1 and, instead of ψ_k, the auxiliary function $\bar{\psi}_k : \mathrm{dom}\ T \to \mathbb{R}$ defined as

$$\bar{\psi}_k(y) = D_f\big(y, (f')^{-1}\big[f'(x^k) - \lambda_k^{-1} T(y)\big]\big) - \sigma D_f(y, x^k).$$

□

3.1 Convergence analysis of Algorithm I

In this subsection we establish the convergence properties of Algorithm I. The following lemma will be used to prove that the Bregman distance with respect to f from the iterates to the solution set of (20) is nonincreasing, which essentially entails global convergence of Algorithm I.

Lemma 1. *Take $f \in \mathcal{F}$ totally convex and such that dom $f = B$. Then for all $v \in B^* \backslash \{0\}$, $\tilde{y} \in B$, $x \in H^+$, $\bar{x} \in H^-$, it holds that $D_f(\bar{x}, x) \geq D_f(\bar{x}, z) + D_f(z, x)$, where $z = \mathrm{argmin}_{y \in H} D_f(y, x)$ and $H = \{y \in B : \langle v, y - \tilde{y}\rangle = 0\}$, $H^+ = \{y \in B : \langle v, y - \tilde{y}\rangle \geq 0\}$ and $H^- = \{y \in B : \langle v, y - \tilde{y}\rangle \leq 0\}$.*

Proof. First, we observe that z is well defined, because it is the Bregman projection of y onto H, which, according to 2.1.5 in [13], exists and is unique when $H \subset$ dom f, and f is totally convex. Note that $D_f(\cdot, x)$ is strictly convex and continuous because $f \in \mathcal{F}$. Consequently, since $x \in H^+$, it holds that $z = \mathrm{argmin}_{y \in H^-} D_f(y, x)$. It follows that z satisfies the first order optimality condition for min $D_f(y,x)$ subject to $y \in H^-$, namely $f'(x) - f'(z) \in N_{H^-}(z)$, where N_{H^-} is the normalizing operator of H^-, i.e. the subdifferential of the indicator function $I_{H^-} : B \to \mathbb{R} \cup \{\infty\}$ defined as $I_{H^-}(x) = 0$ if $x \in H^-$, $I_{H^-}(x) = \infty$ otherwise. Consequently, since $\bar{x} \in H^-$, we have $\langle f'(x) - f'(z), z - \bar{x}\rangle \geq 0$. Therefore,

$$D_f(\bar{x},x) - D_f(\bar{x},z) - D_f(z,x) = \langle f'(x) - f'(z), z - \bar{x}\rangle \geq 0, \tag{30}$$

where the equality in (30) is an elementary identity, known as the *three-point property* (see e.g. [17]). □

Lemma 2. *Let $\{x^k\}$, $\{\tilde{x}^k\}$, $\{\lambda_k\}$ and σ be as in Algorithm I and assume that $f \in \mathcal{F}$ is totally convex. For all k it holds that*

$$\|e^k\|_* ||x^k - \tilde{x}^k|| \leq \sigma \lambda_k D_f(\tilde{x}^k, x^k) \leq \lambda_k D_f(\tilde{x}^k, x^k). \tag{31}$$

Proof. Let us consider two cases. First, if $||x^k - \tilde{x}^k|| < 1$ then we have $\|e^k\|_* ||x^k - \tilde{x}^k|| \leq \|e^k\|_*$, so that the leftmost inequality in (31) follows trivially from (22). Second, if $||x^k - \tilde{x}^k|| \geq 1$, then we use (15) and the definition of ν_f (i.e. (14)), to get

$$\sigma\lambda_k D_f(\tilde{x}^k, x^k) \geq \sigma\lambda_k \nu_f(x^k, ||x^k - \tilde{x}^k||) \geq \sigma\lambda_k ||x^k - \tilde{x}^k|| \nu_f(x^k, 1) \geq ||x^k - \tilde{x}^k|| \|e^k\|_*, \tag{32}$$

where the last inequality follows from (22). Thus, the first inequality of (31) is proved, and the second one holds because $\sigma \in [0,1]$. □

Proposition 8. *Take $f \in \mathcal{F}$, satisfying H1 and H2 and such that dom $f = B$. Let $\{x^k\}$ be the sequence generated by Algorithm I. If problem (20) has solutions (i.e. $T^{-1}(0) \neq \emptyset$), then*

i) For all $\bar{x} \in T^{-1}(0)$, $D_f(\bar{x}, x^k)$ is nonincreasing and convergent,

ii) The sequence $\{x^k\}$ is bounded,

iii) $\sum_{k=0}^{\infty} D_f(x^{k+1}, x^k) < \infty$.

Proof. Take $\bar{x} \in T^{-1}(0)$. Let $H_k^- = \{x \in B |\ \langle v^k, x - \tilde{x}^k\rangle \leq 0\}$, with $\tilde{x}^k$ as in step 2 of the algorithm. By monotonicity of T, (23) and (21), we have $\langle v^k, \bar{x} - \tilde{x}^k\rangle \leq 0$, so that $\bar{x} \in H_k^-$. Applying (23) and the definition of D_f (i.e. (7)), we get

$$\begin{aligned}
\langle v^k, x^k - \tilde{x}^k\rangle &= \lambda_k \langle f'(x^k) - f'(\tilde{x}^k), x^k - \tilde{x}^k\rangle - \langle e^k, x^k - \tilde{x}^k\rangle \\
&= \lambda_k [D_f(\tilde{x}^k, x^k) + D_f(x^k, \tilde{x}^k)] - \langle e^k, x^k - \tilde{x}^k\rangle \\
&\geq \lambda_k D_f(x^k, \tilde{x}^k) + [\lambda_k D_f(\tilde{x}^k, x^k) - \|e^k\|_* ||x^k - \tilde{x}^k||].
\end{aligned}$$

where the last inequality follows from the definition of the norm in B^*. Now, applying Lemma 2, we have $\langle v^k, x^k - \tilde{x}^k\rangle \geq \lambda_k D_f(x^k, \tilde{x}^k) \geq 0$, so that $x^k \in H_k^+$, and in view of (24), we are able to apply Lemma 1 with $\tilde{y} = \tilde{x}^k \in B$, $x = x^k$, $v = v^k$ and $z = x^{k+1}$, obtaining

$$D_f(\bar{x}, x^k) \geq D_f(\bar{x}, x^{k+1}) + D_f(x^{k+1}, x^k) \geq D_f(\bar{x}, x^{k+1}). \tag{33}$$

Consequently $\{D_f(\bar{x}, x^k)\}$ is a nonnegative, nonincreasing sequence (in fact, strictly decreasing unless $\tilde{x}^k = x^k$), henceforth convergent, establishing (i), and also $\{x^k\} \subset \{y \in B | \; D_f(\bar{x}, y) \leq D_f(\bar{x}, x^0)\}$, which is a bounded set by H1, establishing (ii). Finally, by (33),

$$\sum_{k=0}^{n-1} D_f(x^{k+1}, x^k) \leq \sum_{k=0}^{n-1} D_f(\bar{x}, x^k) - D_f(\bar{x}, x^{k+1}) = D_f(\bar{x}, x^0) - D_f(\bar{x}, x^n) \leq D_f(\bar{x}, x^0),$$

establishing (iii). □

Now we state and prove the main result of this subsection.

Theorem 2. *Take $f \in \mathcal{F}$ with dom $f = B$ satisfying H1–H4, and $\{\lambda_k\} \subset (0, \bar{\lambda}]$. Let $\{x^k\}$ be the sequence generated by Algorithm I. If problem* (20) *has solutions, then*

i) $\{x^k\}$ has weak accumulation points and all of them are solutions of (20).

ii) If f also satisfies H5, then the whole sequence $\{x^k\}$ is weakly convergent to a solution of (20).

Proof. By Proposition 8(ii), $\{x^k\}$ is bounded. Since B is reflexive, there exists at least one weak accumulation point. It follows from Proposition 8(iii) that $\lim_{k\to\infty} D_f(x^{k+1}, x^k) = 0$, and then, since f satisfies H2, we conclude from Proposition 5 that

$$x^{k+1} - x^k \xrightarrow[k\to\infty]{\text{s}} 0. \tag{34}$$

From (22) and Proposition 4 we get boundedness of the sequence $\{\lambda_k^{-1} e^k\}$, and therefore, using (34) and the fact that $0 \leq |\langle \lambda_k^{-1} e^k, x^{k+1} - x^k \rangle| \leq \|\lambda_k^{-1} e^k\|_* ||x^{k+1} - x^k||$, we get

$$\langle \lambda_k^{-1} e^k, x^{k+1} - x^k \rangle \xrightarrow[k\to\infty]{\text{s}} 0. \tag{35}$$

Using the three-point property (i.e., the equality in (30)) and (23), we get

$$D_f(x^{k+1}, x^k) - D_f(x^{k+1}, \tilde{x}^k) - D_f(\tilde{x}^k, x^k) = \langle f'(x^k) - f'(\tilde{x}^k), \tilde{x}^k - x^{k+1} \rangle$$

$$= \lambda_k^{-1} \langle v^k, \tilde{x}^k - x^{k+1} \rangle + \langle \lambda_k^{-1} e^k, \tilde{x}^k - x^{k+1} \rangle = \langle \lambda_k^{-1} e^k, \tilde{x}^k - x^{k+1} \rangle, \tag{36}$$

where the last equality follows from the fact that $x^{k+1} \in H_k = \{x \in B : \langle v^k, x - \tilde{x}^k \rangle = 0\}$. Then

$$D_f(x^{k+1}, x^k) - D_f(x^{k+1}, \tilde{x}^k) = \langle \lambda_k^{-1} e^k, x^k - x^{k+1} \rangle + D_f(\tilde{x}^k, x^k) - \langle \lambda_k^{-1} e^k, x^k - \tilde{x}^k \rangle$$

$$\geq \langle \lambda_k^{-1} e^k, x^k - x^{k+1} \rangle + \frac{1}{\lambda_k} \left[\lambda_k D_f(\tilde{x}^k, x^k) - \|e^k\|_* ||x^k - \tilde{x}^k|| \right] \geq \langle \lambda_k^{-1} e^k, x^k - x^{k+1} \rangle, \tag{37}$$

where the last inequality follows from Lemma 2. Taking limits as k goes to ∞ in the leftmost and rightmost expressions of (37), we obtain, using (35),

$$\lim_{k\to\infty} D_f(x^{k+1}, \tilde{x}^k) = 0, \tag{38}$$

and therefore, using H2 and Proposition 5,

$$x^{k+1} - \tilde{x}^k \xrightarrow[k\to\infty]{\text{s}} 0, \tag{39}$$

which, in view of (34), implies

$$x^k - \tilde{x}^k \xrightarrow[k\to\infty]{\text{s}} 0. \tag{40}$$

From (40) and H3, we get

$$f'(x^k) - f'(\tilde{x}^k) \xrightarrow[k\to\infty]{\text{s}} 0. \tag{41}$$

By (40), there exists $k_0 \in \mathbb{N}$ such that, $||x^k - \tilde{x}^k|| < 1$, for all $k \geq k_0$, and consequently for $k \geq k_0$ our error criterium (22) implies that

$$||e^k||_* \leq \sigma \lambda_k D_f(\tilde{x}^k, x^k). \tag{42}$$

Next, we take limits as k goes to ∞ in the leftmost and rightmost expressions of (36). The rightmost one converges to 0 by (39) and boundedness of the sequence $\{\lambda_k^{-1} e^k\}$, the first term in the leftmost expression converges to 0 as a consequence of Proposition 8(iii), and the second term converges to 0 by (38). It follows that $\lim_{k\to\infty} D_f(\tilde{x}^k, x^k) = 0$, and then, since $\lambda_k \leq \bar{\lambda}$, we get from (42) that $e^k \xrightarrow[k\to\infty]{\text{s}} 0$. It follows then from (23) and (41) that

$$v^k \xrightarrow[k\to\infty]{\text{s}} 0. \tag{43}$$

Next, we prove (i). Let $\bar{x}$ be a weak accumulation point of $\{x^k\}$, so that there exist a subsequence $\{x^{j_k}\}$ of $\{x^k\}$ which is weakly convergent to $\bar{x}$. By (40), $\tilde{x}^{j_k} \xrightarrow[k\to\infty]{\text{w}} \bar{x}$. It follows from (21) and (43) that $v^{j_k} \in T(\tilde{x}^{j_k})$ and $v^{j_k} \xrightarrow[k\to\infty]{\text{s}} 0$. Thus, $0 \in T(\bar{x})$, because the graph of a maximal monotone operator is demiclosed (see e.g. [36], page 105).

Item (ii) is proved exactly as in Proposition 2.1.9 of [13]. For the sake of completeness, we will give a proof, but only in the next subsection, namely in the proof of Theorem 3(ii), because it requires a preliminary result which fits better there. □

3.2 Convergence analysis of Algorithm II

From (25)–(27) with $v^k = e^k + \lambda_k[f'(x^k) - f'(\tilde{x}^k)]$, we obtain

$$0 = \lambda_k^{-1} v^k + f'(x^{k+1}) - f'(x^k), \tag{44}$$

$$D_f(\tilde{x}^k, x^{k+1}) \leq \sigma D_f(\tilde{x}^k, x^k) \tag{45}$$

for all k. The following lemmas essentially resume the behavior of the error criterium in Algorithm II and entail its global convergence. These results were presented in Corollary 2.6 and Lemma 4.1 of [43] for the finite dimensional case, and their extension to Banach spaces is immediate.

Lemma 3. *(Four-point Lemma) Take $f \in \mathcal{F}$. For any $x, z \in int\ (dom\ f)$ and $y, w \in dom\ f$, it holds that*

$$D_f(w,z) = D_f(w,x) + \langle f'(x) - f'(z), w - y\rangle + D_f(y,z) - D_f(y,x).$$

Proof. Immediate from the definition of D_f (i.e. (7)). □

Lemma 4. *Take $f \in \mathcal{F}$, satisfying H4, such that $dom\ T \subset int(dom\ f)$, $\bar{x} \in T^{-1}(0)$ and $\{x^k\}$, $\{\tilde{x}^k\}$, $\{\lambda_k\}$ and σ as in Algorithm II. Let $v^k = e^k + \lambda_k[f'(x^k) - f'(\tilde{x}^k)]$. Then*

$$D_f(\bar{x}, x^{k+1}) \leq D_f(\bar{x}, x^k) - \lambda_k^{-1}\langle v^k, \tilde{x}^k - \bar{x}\rangle - (1-\sigma) D_f(\tilde{x}^k, x^k) \leq D_f(\bar{x}, x^k).$$

Proof. From Lemma 3, we get

$$\begin{aligned} D_f(\bar{x}, x^{k+1}) &= D_f(\bar{x}, x^k) + \langle f'(x^k) - f'(x^{k+1}), \bar{x} - \tilde{x}^k\rangle + D_f(\tilde{x}^k, x^{k+1}) - D_f(\tilde{x}^k, x^k) \\ &= D_f(\bar{x}, x^k) + \langle \lambda_k^{-1} v^k, \bar{x} - \tilde{x}^k\rangle + D_f(\tilde{x}^k, x^{k+1}) - D_f(\tilde{x}^k, x^k) \\ &\leq D_f(\bar{x}, x^k) - \lambda_k^{-1}\langle v^k, \tilde{x}^k - \bar{x}\rangle + (\sigma - 1) D_f(\tilde{x}^k, x^k), \end{aligned}$$

where the second equality follows from (44) and the last inequality from (45). Using nonnegativity of D_f and the fact that $\sigma \in [0,1)$, we get

$$D_f(\bar{x}, x^{k+1}) \leq D_f(\bar{x}, x^k) - \lambda_k^{-1}\langle v^k, \tilde{x}^k - \bar{x}\rangle, \tag{46}$$

and the result follows from (46) and monotonicity of T, since $v^k \in T(\tilde{x}^k)$. □

Proposition 9. *Let $f \in \mathcal{F}$, satisfying H1 and H4 and such that $dom\ T \subset int(dom\ f)$. If problem* (20) *has solutions, then*

i) *$D_f(\bar{x}, x^k)$ converges decrescently, for all $\bar{x} \in T^{-1}(0)$,*

ii) *the sequence $\{x^k\}$ is bounded,*

iii) $\sum_{k=0}^{\infty} \lambda_k^{-1}\langle v^k, \tilde{x}^k - \bar{x}\rangle < \infty$, *with* $v^k = e^k + \lambda_k[f'(x^k) - f'(\tilde{x}^k)]$,

iv) $\sum_{k=0}^{\infty} D_f(\tilde{x}^k, x^k) < \infty$,

v) $\sum_{k=0}^{\infty} D_f(\tilde{x}^k, x^{k+1}) < \infty$,

vi) *if f satisfies H2, then*

a) $\tilde{x}^k - x^k \xrightarrow[k\to\infty]{\text{s}} 0$, *and consequently* $\{\tilde{x}^k\}$ *is bounded,*

b) $x^{k+1} - x^k \xrightarrow[k\to\infty]{\text{s}} 0$.

Proof. Take $\bar{x} \in T^{-1}(0)$ (which is nonempty because problem (20) has solutions). Then, by Lemma 4, $\{D_f(\bar{x}, x^k)\}$ is a nonnegative, nonincreasing sequence, henceforth convergent, and $\{x^k\}$ is contained in a level set of $D_f(\bar{x}, \cdot)$, which is bounded by H1. Also, using again Lemma 4,

$$\lambda_k^{-1}\langle v^k, \tilde{x}^k - \bar{x}\rangle + (1-\sigma)D_f(\tilde{x}^k, x^k) \le D_f(\bar{x}, x^k) - D_f(\bar{x}, x^{k+1}),$$

from which (iii) and (iv) follow easily. Item (v) follows from (iv) and (45). For (vi), observe that $\lim_{k\to\infty} D_f(\tilde{x}^k, x^k) = 0$ and that $\{x^k\}$ is bounded, so that we can apply H2 and Proposition 5 to obtain $\tilde{x}^k - x^k \xrightarrow[k\to\infty]{\text{s}} 0$. In the same way $\tilde{x}^k - x^{k+1} \xrightarrow[k\to\infty]{\text{s}} 0$, implying that $x^k - x^{k+1} \xrightarrow[k\to\infty]{\text{s}} 0$. □

Proposition 9 ensures existence of weak accumulation points of the sequence $\{x^k\}$ and, also, that they coincide with those of the sequence $\{\tilde{x}^k\}$.

Theorem 3. *Take $f \in \mathcal{F}$, satisfying H1, H2, H3 and H4, and such that dom $T \subset$ int(dom f), and $\{\lambda_k\} \subset (0, \bar{\lambda}]$. If problem* (20) *has solutions, then*

i) The sequence $\{x^k\}$ generated by Algorithm II has weak accumulation points, all of which are solutions of (20)*,*

ii) if f also satisfies H5, then the whole sequence is weakly convergent to a solution of (20).

Proof. By H3 and Proposition 9(ii) and (vi-b), we have $f'(x^{k+1}) - f'(x^k) \xrightarrow[k\to\infty]{\text{s}} 0$. Let $v^k = e^k + \lambda_k[f'(x^k) - f'(\tilde{x}^k)]$. By (25), $v^k \in T(\tilde{x}^k)$ and then, using $\lambda_k \le \bar{\lambda}$, we get from (44) that $v^k \xrightarrow[k\to\infty]{\text{s}} 0$. Thus, (i) follows from demiclosedness of the graph of maximal monotone operators.

In order to prove (ii), let $\bar{x}_1$ and $\bar{x}_2$ be two weak accumulation points of $\{x^k\}$, so that there exist subsequences $\{x^{j_k}\}$, and $\{x^{i_k}\}$ of $\{x^k\}$ such that $x^{j_k} \xrightarrow[k\to\infty]{\text{w}} \bar{x}_1$ and $x^{i_k} \xrightarrow[k\to\infty]{\text{w}} \bar{x}_2$. By (i), $\bar{x}_1$ and $\bar{x}_2$ are solutions of (20). Thus, Proposition 9(i) guarantees the existence of $\xi_1, \xi_2 \in \mathbb{R}_+$ such that

$$\lim_{k\to\infty} D_f(\bar{x}_1, x^k) = \xi_1, \qquad \lim_{k\to\infty} D_f(\bar{x}_2, x^k) = \xi_2. \tag{47}$$

Now, using Lemma 3, we get

$$|\langle f'(x^{i_k}) - f'(x^{j_k}), \bar{x}_1 - \bar{x}_2\rangle| = \left|D_f(\bar{x}_1, x^{i_k}) - D_f(\bar{x}_1, x^{j_k}) - [D_f(\bar{x}_2, x^{i_k}) - D_f(\bar{x}_2, x^{j_k})]\right|$$

$$\le \left|D_f(\bar{x}_1, x^{i_k}) - D_f(\bar{x}_1, x^{j_k})\right| + \left|D_f(\bar{x}_2, x^{i_k}) - D_f(\bar{x}_2, x^{j_k})\right|. \tag{48}$$

Taking limits as k goes to ∞ in the extreme expressions of (48) and using (47), we get that $\lim_{k\to\infty}$ $|\langle f'(x^{i_k}) - f'(x^{j_k}), \bar{x}_1 - \bar{x}_2\rangle| = 0$, and then, since $\bar{x}_1$ and $\bar{x}_2$ belong to dom $T \subset$ int(dom f), H5 and nonnegativity of D_f imply that

$$0 = \langle f'(\bar{x}_1) - f'(\bar{x}_2), \bar{x}_1 - \bar{x}_2\rangle = D_f(\bar{x}_1, \bar{x}_2) + D_f(\bar{x}_2, \bar{x}_1) \ge D_f(\bar{x}_1, \bar{x}_2). \tag{49}$$

It follows from (49) and H2 that $\bar{x}_1 = \bar{x}_2$, establishing the uniqueness of the weak accumulation points of $\{x^k\}$. □

4 Inexact doubly augmented Lagrangians for $\mathcal{L}^p$-constrained convex optimization problems

In this section we present an application of Algorithms I and II in the context of the augmented Lagrangian method for convex optimization problems in infinite dimensional spaces with a possibly infinite number of constraints, introduced in ([13], 3.4) Let us consider a measure space $(\Omega, \mathcal{A}, \mu)$ and functions $g : B \to \mathbb{R}$, $G : \Omega \times B \to \mathbb{R}$, such that:

i) g and $G(\omega, \cdot)$ are convex and Fréchet differentiable for all $\omega \in \Omega$.

ii) For all $x \in B$, both $G(\cdot, x) : \Omega \to \mathbb{R}$ and $G'_x(\cdot, x) : \Omega \to \mathrm{B}^*$ are p-integrable for some $p \in (1, \infty)$, where $G'_x(\omega, \cdot)$ denotes the Fréchet derivative of $G(\omega, \cdot)$ (in the case of G'_x integrability is understood in the sense of the Bochner Integral, see, e.g., [34]).

The primal problem (P) is defined as

$$(P) \quad \begin{cases} \min g(x) \\ \ \text{s.t.} \quad G(\omega, x) \leq 0 \ \text{ a.e. in } \Omega, \end{cases}$$

and its dual (D), as

$$(D) \quad \begin{cases} \max \Phi(y) \\ \ \text{s.t.} \ \ y \in \mathcal{L}^q(\Omega), \qquad y(\omega) \geq 0 \ \text{ a.e. in } \Omega, \end{cases}$$

where $q = p/(p-1)$, the dual objective $\Phi : \mathcal{L}^q(\Omega) \to \mathbb{R} \cup \{-\infty\}$ is defined as $\Phi(y) = \inf_{x \in B} L(x, y)$, and the *Lagrangian* $L : B \times \mathcal{L}^q(\Omega) \to \mathbb{R}$, is given by

$$L(x, y) = g(x) + \int_\Omega G(\omega, x) y(\omega) d\mu(\omega).$$

A pair $(x, y) \in B \times \mathcal{L}^q(\Omega)$ is called optimal if x is an optimal solution of problem (P) and y an optimal solution of problem (D). A pair $(x, y) \in B \times \mathcal{L}^q(\Omega)$ is called a KKT-pair if:

$$G(\omega, x) \leq 0 \text{ a.e. in } \Omega \qquad \text{(primal feasibility)}, \tag{50}$$

$$y(\omega) \geq 0 \text{ a.e. in } \Omega \qquad \text{(dual feasibility)}, \tag{51}$$

$$0 = L'_x(x, y) = g'(x) + \int_\Omega G'_x(\omega, x) y(\omega) d\mu(\omega) \qquad \text{(Lagrangian condition)}, \tag{52}$$

$$\int_\Omega G(\omega, x) y(\omega) d\mu(\omega) = 0 \qquad \text{(complementarity)}. \tag{53}$$

See ([13], 3.3) for this formulation of the Karush-Kuhn-Tucker conditions. Note also, that $L(x,\cdot) : \mathcal{L}^q(\Omega) \to \mathbb{R}$ is Fréchet differentiable for all $x \in B$. In fact,

$$\frac{L(x, y + t\tilde{y}) - L(x,y)}{t} = \int_\Omega G(\omega, x)\tilde{y}(\omega) d\mu(\omega),$$

for all $\tilde{y} \in \mathcal{L}^q(\Omega)$, so that $L'_y(x,y) = G(\cdot, x)$ for all $y \in \mathcal{L}^q(\Omega)$. Next we define the *saddle-point operator* $T_L : B \times \mathcal{L}^q(\Omega) \to \mathcal{P}(B^* \times \mathcal{L}^p(\Omega))$ as

$$\begin{aligned} T_L(x,y) &= \big(L'_x(x,y), -L'_y(x,y) + N_{\mathcal{L}^q_+(\Omega)}(y)\big) \\ &= \left(g'(x) + \int_\Omega G'_x(\omega, x) y(\omega) d\mu(\omega), -G(\cdot, x) + N_{\mathcal{L}^q_+(\Omega)}(y)\right), \end{aligned} \tag{54}$$

where $\mathcal{L}^q_+(\Omega) = \{y \in \mathcal{L}^q(\Omega) |\ y(\omega) \geq 0 \text{ a.e.}\}$, and $N_{\mathcal{L}^q_+(\Omega)} : \mathcal{L}^q(\Omega) \to \mathcal{P}(\mathcal{L}^p(\Omega))$ denotes the normalizing operator of the cone $\mathcal{L}^q_+(\Omega)$, given by

$$N_{\mathcal{L}^q_+(\Omega)}(y) = \begin{cases} \{s \in \mathcal{L}^p(\Omega) |\ \int_\Omega s(\omega)[z(\omega) - y(\omega)] d\mu(\omega) \leq 0,\ \forall z \in \mathcal{L}^q_+(\Omega)\} & \text{if } y \in \mathcal{L}^q_+(\Omega) \\ \emptyset & \text{otherwise.} \end{cases}$$

The following proposition resumes some elementary properties of the saddle-point operator.

Proposition 10. *The operator T_L, defined in* (54)*, satisfies*

i) $0 \in T_L(x,y)$ *if and only if* (x,y) *is a KKT-pair.*

ii) If $0 \in T_L(x,y)$*, then* (x,y) *is an optimal pair.*

iii) T_L *is a maximal monotone operator.*

Proof. Note that $0 \in T_L(x,y)$ if and only if $(x,y) \in B \times \mathcal{L}^q(\Omega)$ is such that:

$$\begin{aligned} 0 &= g'(x) + \int_\Omega G'_x(\omega, x) y(\omega) d\mu(\omega) \\ 0 &\in -G(\cdot, x) + N_{\mathcal{L}^q_+(\Omega)}(y), \end{aligned}$$

or equivalently

$$0 = g'(x) + \int_\Omega G'_x(\omega, x) y(\omega) d\mu(\omega), \tag{55}$$

$$0 \geq \int_\Omega G(\omega, x)[z(\omega) - y(\omega)] d\mu(\omega) \qquad \text{for all } z \in \mathcal{L}^q_+(\Omega), \tag{56}$$

$$y \in \mathcal{L}_+^q(\Omega). \tag{57}$$

For the "if" part of (i), observe that these conditions are immediately implied by (50)–(53). For the "only if" part of (i), note first that (55) is just (52) and (57) is (51). If (50) does not hold, then we can find a subset $W \subset \Omega$ with finite, nonzero measure such that $G(\omega, x) > 0$, for all $\omega \in W$. Take $z \in \mathcal{L}^q(\Omega)$ defined as $z(\omega) = y(\omega) + \mathcal{X}_W(\omega)$, where $\mathcal{X}_W$ denotes the characteristic function of the set W, i.e. $\mathcal{X}_W(w) = 1$ if $w \in W$, $\mathcal{X}_W(w) = 0$ otherwise. Then $z \in \mathcal{L}_+^q(\Omega)$, and

$$0 < \int_{\mathrm{W}} G(\omega, x) d\mu(\omega) = \int_\Omega G(w, x)\mathcal{X}_W(w) d\mu(\omega) = \int_\Omega G(w, x)[z(\omega) - y(\omega)] d\mu(\omega),$$

in contradiction with (56), (57), so that (50) holds, which, in turn, implies

$$\int_\Omega G(\omega, x) y(\omega) d\mu(\omega) \leq 0. \tag{58}$$

Take $z \in \mathcal{L}_+^q(\Omega)$ defined as $z(\omega) = 0$ for all $\omega \in \Omega$, and get from (56)

$$-\int_\Omega G(\omega, x) y(\omega) d\mu(\omega) \leq 0. \tag{59}$$

In view of (58) and (59), (53) holds and the proof of (i) is completed. Item (ii) follows from (i) and Proposition 3.3.2 of [13], and (iii) follows from results in [38] applied to the saddle function $L(x, y)$. □

As we discussed in Section 1, the augmented Lagrangian method introduced in [13], of which (12)-(13) is a particular case, requires rather demanding conditions, like coercivity on the data functions, in order to ensure existence of the primal iterates x^k, and boundedness of the sequence $\{x^k\}$. Also, it is assumed in its convergence proof that the primal subproblems are exactly solved. We present next two variants which overcome these limitations. The first obstacle can be removed by introducing a well behaved regularizing term for the primal subproblems, i.e. by considering a doubly augmented Lagrangian. In the finite dimensional case, this idea was first proposed in [41]. We will use the regularizing function $h_r : \mathcal{L}^q(\Omega) \to \mathbb{R}$ defined as $h_r(y) = \frac{1}{r}||y||_q^r$ for some $r \in (1, \infty)$. The *augmented Lagrangian* $\bar{L} : B \times \mathcal{L}^q(\Omega) \times \mathbb{R}_{++} \to \mathbb{R}$ is defined as

$$\bar{L}(x, y, \rho) = g(x) + \frac{\rho}{s} \left\| M_r^+(x, y, \rho) \right\|_p^s,$$

where $s = r/(r-1)$, $M_r : B \times \mathcal{L}^q(\Omega) \times \mathbb{R}_{++} \to \mathcal{L}^p(\Omega)$ is given by

$$M_r(x, y, \rho)(\omega) = h_r'(y)(\omega) + \rho^{-1} G(\omega, x), \tag{60}$$

and $M_r^+(x, y, \rho)(\omega) = \max\{0, M_r(x, y, \rho)(\omega)\}$. The *doubly augmented Lagrangian* $\hat{L} : B \times B \times \mathcal{L}^q(\Omega) \times \mathbb{R}_{++} \to \mathbb{R}$ is defined as

$$\hat{L}(x, z, y, \rho) = \bar{L}(x, y, \rho) + \rho D_f(x, z), \tag{61}$$

where $f \in \mathcal{F}$ is such that dom $f = B$. Our exact method with primal regulatization is defined as:

Doubly augmented Lagrangian method

1. Choose $(x^0, y^0) \in B \times \mathcal{L}^q_+(\Omega)$.

2. Given (x^k, y^k), choose $\lambda_k > 0$ and define x^{k+1} as

$$x^{k+1} = \text{argmin}_{x \in B} \hat{L}(x, x^k, y^k, \lambda_k) = \text{argmin}_{x \in B} \left[\bar{L}(x, y^k, \lambda_k) + \lambda_k D_f(x, x^k) \right]. \tag{62}$$

3. Define y^{k+1} as

$$y^{k+1}(\omega) = \left\| M_r^+(x^{k+1}, y^k, \lambda_k) \right\|_p^{s-p} \left[M_r^+(x^{k+1}, y^k, \lambda_k)(\omega) \right]^{p-1}.$$

Though this version still requires exact solution of the subproblems, it solves the issue of the existence and uniqueness of the primal iterates. In fact, we have the following result.

Proposition 11. *Let $f \in \mathcal{F}$ such that dom $f = B$. If either f satisfies H4, or f is totally convex and g is bounded from below, then there exists a unique solution of each primal subproblem* (62).

Proof. If f satisfies H4, take $T = \bar{L}'_x(\cdot, y^k, \lambda_k)$ and use Proposition 6. In the other case the result follows from Proposition 3.15 in [13]. □

Next we translate the error criteria studied in Section 3 to this setting in order to admit inexact solution of the primal subproblems in the doubly augmented Lagrangian method just presented. We consider the regularizing function $F : B \times \mathcal{L}^q(\Omega) \to \mathbb{R}$ defined as $F(x, y) = f(x) + h_r(y)$, with $h_r(y) = \frac{1}{r} ||y||_q^r$ $(1 < r < \infty)$, and $f \in \mathcal{F}$ such that dom $f = B$. We get the following algorithms.

Algorithm III.

1. Choose $z^0 = (x^0, y^0) \in B \times \mathcal{L}^q(\Omega)$ and $\sigma \in [0, 1]$.

2. Given $z^k = (x^k, y^k)$, choose $\lambda_k > 0$ and find $\tilde{x}^k \in B$ such that

$$||\hat{L}'_x(\tilde{x}^k, x^k, y^k, \lambda_k)||_{B^*} \leq \sigma \lambda_k \begin{cases} D_F(\tilde{z}^k, z^k) & \text{if } ||z^k - \tilde{z}^k||_{B \times \mathcal{L}^q(\Omega)} < 1, \\ \nu_F(z^k, 1) & \text{if } ||z^k - \tilde{z}^k||_{B \times \mathcal{L}^q(\Omega)} \geq 1, \end{cases} \tag{63}$$

where

$$\tilde{z}^k = \left(\tilde{x}^k, \left\| M_r^+(\tilde{x}^k, y^k, \lambda_k) \right\|_p^{s-p} \left[M_r^+(\tilde{x}^k, y^k, \lambda_k) \right]^{p-1} \right). \tag{64}$$

3. Set $v^k = \big(\bar{L}'_x(\tilde{x}^k, y^k, \lambda_k), -G(\cdot, \tilde{x}^k) + \lambda_k M_r^-(\tilde{x}^k, y^k, \lambda_k)\big) \in B^* \times \mathcal{L}^p(\Omega)$. If $v^k = 0$ or $\tilde{z}^k = z^k$, then stop. Otherwise,
$$z^{k+1} = (x^{k+1}, y^{k+1}) = \operatorname{argmin}_{z \in H_k} D_F(z, z^k),$$
where $H_k = \{z \in B \times \mathcal{L}^q(\Omega) : \langle v^k, z - \tilde{z}^k \rangle = 0\}$.

We need some notation for our next algorithm, which uses the same regularizing function $F = f + h_r$ as Algorithm III. Define $Q : B \times \mathcal{L}^q(\Omega) \times \mathbb{R}_{++} \to \mathcal{L}^q(\Omega)$ as
$$Q(x, y, \rho) = \left\| M_r^+(x, y, \rho) \right\|_p^{s-p} \left[M_r^+(x, y, \rho) \right]^{p-1}, \tag{65}$$
with $s = r/(r-1)$ and M_r as in (60).

Algorithm IV.

1. Choose $(x^0, y^0) \in B \times \mathcal{L}_+^q(\Omega)$ and $\sigma \in [0, 1)$.
2. Given $z^k = (x^k, y^k)$, choose $\lambda_k > 0$ and find $\tilde{x}^k \in B$ such that
$$D_f\big(\tilde{x}^k, (f')^{-1}\big[f'(x^k) - \lambda_k^{-1} \bar{L}'_x(\tilde{x}^k, y^k, \lambda_k)\big]\big) \le \sigma\big[D_f(\tilde{x}^k, x^k) + D_{h_r}\big(u^k, y^k\big)\big], \tag{66}$$
where $u^k = Q(\tilde{x}^k, y^k, \lambda_k)$, with Q as in (65).
3. If $(\tilde{x}^k, u^k) = (x^k, y^k)$, then stop. Otherwise, define
$$\begin{aligned} y^{k+1} &= u^k = \left\| M_r^+(\tilde{x}^k, y^k, \lambda_k) \right\|_p^{s-p} \left[M_r^+(\tilde{x}^k, y^k, \lambda_k) \right]^{p-1}, \\ x^{k+1} &= (f')^{-1}\big[f'(x^k) - \lambda_k^{-1} L'_x(\tilde{x}^k, y^{k+1})\big]. \end{aligned}$$

The following lemma is essential in our analysis. Let $M_r^-(x, y, \rho)(\omega) = \min\{0, M_r(x, y, \rho)(\omega)\}$.

Lemma 5. *For all $(x, y, \rho) \in B \times \mathcal{L}^q(\Omega) \times \mathbb{R}_{++}$ it holds that*

a) $h'_r(Q(x, y, \rho)) = M_r^+(x, y, \rho)$,

b) $h'_r(Q(x, y, \rho)) - h'_r(y) = -\frac{1}{\rho}\big[-G(\cdot, x) + \rho M_r^-(x, y, \rho)\big]$,

c) $M_r^-(x, y, \rho) \in N_{\mathcal{L}_+^q(\Omega)}(Q(x, y, \rho))$,

d) $\bar{L}'_x(x, y, \rho) = L'_x(x, Q(x, y, \rho))$.

Proof. It is well known that $h_r'(y) = ||y||_q^{r-q}|y|^{q-1}\text{sign}(y)$ for all $y \in \mathcal{L}^q(\Omega)$ (see e.g. Proposition 4.9 of [18] together with Proposition 2.1(a),(c)). Substituting $Q(x, y, \rho)$ in this formula and using the facts that $q = p/(p-1)$, $s = r/(r-1)$, we get

$$h_r'(Q(x,y,\rho)) = \left\| \left\| M_r^+(x,y,\rho) \right\|_p^{s-p} \left[M_r^+(x,y,\rho) \right]^{p-1} \right\|_q^{r-q} \left[\left\| M_r^+(x,y,\rho) \right\|_p^{s-p} \left[M_r^+(x,y,\rho) \right]^{p-1} \right]^{q-1}$$

$$= \left\| M_r^+(x,y,\rho) \right\|_p^{(s-p)(r-1)} \left\| M_r^+(x,y,\rho) \right\|_p^{\frac{p(r-q)}{q}} M_r^+(x,y,\rho) = M_r^+(x,y,\rho),$$

establishing (a). Item (b) follows from (a) and (60). We proceed to prove (c). Take any $z \in \mathcal{L}_+^q(\Omega)$ and note that

$$\int_\Omega M_r^-(x,y,\rho)(\omega)[z(\omega) - Q(x,y,\rho)(\omega)]d\mu(\omega) = \int_\Omega M_r^-(x,y,\rho)(\omega)z(\omega)d\mu(\omega)$$

$$- \left\| M_r^+ \right\|_p^{s-p} \int_\Omega M_r^-(x,y,\rho)(\omega) \left[M_r^+(x,y,\rho)(\omega) \right]^{p-1} d\mu(\omega) = \int_\Omega M_r^-(x,y,\rho)(\omega)z(\omega)d\mu(\omega) \le 0.$$

Then the proof follows from the definition of $N_{\mathcal{L}_+^q(\Omega)}(Q(x,y,\rho))$. Item (d) has been proved in Proposition 3.8.1 of [13]. □

We point out that the exact solution of (62) obviously satisfies the error criteria both in Algorithms III and IV. We also mention that when $\sigma = 0$ both Algorithms III and IV reduce to the doubly augmented Lagrangian, because in such a case $\tilde{x}^k$ coincides with the exact solution of (62) and step 3 becomes redundant in both methods. Thus, our following convergence results for Algorithms III and IV also hold for the doubly augmented Lagrangian method.

5 Convergence analysis of Algorithms III and IV

The following proposition establishes the relation between Algorithm III and Algorithm I applied to the Lagrangian operator T_L.

Proposition 12. *Take F, $\{v^k\}$, $\{z^k\}$, $\{\tilde{z}^k\}$, $\{\lambda_k\}$ and σ as in Algorithm III. Then*

i) $v^k \in T_L(\tilde{z}^k)$,

ii) Let $e^k = \lambda_k[F'(z^k) - F'(\tilde{z}^k)] - v^k \in B^* \times \mathcal{L}^p(\Omega)$. *Then*

$$\|e^k\|_{(B\times\mathcal{L}^q(\Omega))^*} \le \sigma\lambda_k \begin{cases} D_F(\tilde{z}^k, z^k) & \textit{if } ||z^k - \tilde{z}^k||_{B\times\mathcal{L}^q(\Omega)} < 1, \\ \nu_F(z^k, 1) & \textit{if } ||z^k - \tilde{z}^k||_{B\times\mathcal{L}^q(\Omega)} \ge 1. \end{cases}$$

Proof. Let $v^k = (v_1^k, v_2^k) \in B^* \times \mathcal{L}^p(\Omega)$. Take Q as in (65). Then we have, by definition of v^k,

$$v_1^k = \bar{L}_x'(\tilde{x}^k, y^k, \lambda_k) = L_x'(\tilde{x}^k, Q(\tilde{x}^k, y^k, \lambda_k)) = L_x'(\tilde{z}^k),$$

where the second equality follows from Lemma 5(d) and the last one from (64), and also

$$v_2^k = -G(\cdot, \tilde{x}^k) + \lambda_k M_r^-(\tilde{x}^k, y^k, \lambda_k) \in -G(\cdot, \tilde{x}^k) + N_{\mathcal{L}_+^q}(Q(\tilde{x}^k, y^k, \lambda_k)),$$

where the inclusion was proved in Lemma 5(c). So v^k is an element of $T_L(\tilde{z}^k)$ and (i) holds.

In order to prove (ii), let $e^k = (e_1^k, e_2^k) \in B^* \times \mathcal{L}^p(\Omega)$. Then we have that

$$e^k = \lambda_k[F'(z^k) - F'(\tilde{z}^k)] - v^k$$

if and only if

$$\begin{aligned} 0 &= v_1^k + \lambda_k[f'(\tilde{x}^k) - f'(x^k)] + e_1^k, \\ 0 &= v_2^k + \lambda_k[h_r'(Q(\tilde{x}^k, y^k, \lambda_k)) - h_r'(y^k)] + e_2^k \end{aligned}$$

if and only if

$$\begin{aligned} 0 &= \bar{L}_x'(\tilde{x}^k, y^k, \lambda_k) + \lambda_k D_f'(\tilde{x}^k, x^k) + e_1^k, \\ 0 &= -G(\cdot, \tilde{x}^k) + \lambda_k M_r^-(\tilde{x}^k, y^k, \lambda_k) + \lambda_k[h_r'(Q(\tilde{x}^k, y^k, \lambda_k)) - h_r'(y^k)] + e_2^k \end{aligned}$$

if and only if, using (61) and Lemma 5(b),

$$\begin{aligned} 0 &= \hat{L}_x'(\tilde{x}^k, x^k, y^k, \lambda_k) + e_1^k, \\ 0 &= e_2^k \end{aligned}$$

if and only if

$$e^k = \left(-\hat{L}_x'(\tilde{x}^k, x^k, y^k, \lambda_k), 0\right).$$

Thus, $\|e^k\|_{(B\times\mathcal{L}^q(\Omega))^*} = \|(e_1^k, 0)\|_{B^*\times\mathcal{L}^p(\Omega)} = \|e_1^k\|_{B^*} = \|\hat{L}_x'(\tilde{x}^k, x^k, y^k, \lambda_k)\|_{B^*}$, and the result follows from (63). □

We prove next that Algorithm IV is a particular instance of Algorithm II applied to the saddle-point operator T_L.

Proposition 13. *Let $\{x^k\}$, $\{y^k\}$, $\{\tilde{x}^k\}$, $\{\lambda_k\}$, σ and f be as in Algorithm IV. Define $F : B \times \mathcal{L}^q(\Omega) \to \mathbb{R}$ as $F(x,y) = f(x) + h_r(y)$, $e^k = \left(\hat{L}_x'(\tilde{x}^k, x^k, y^k, \lambda_k), 0\right)$, $\tilde{z}^k = (\tilde{x}^k, Q(\tilde{x}^k, y^k, \lambda_k))$ and $z^k = (x^k, y^k)$, where Q is defined as in (65). Then*

i) $e^k + \lambda_k[F'(z^k) - F'(\tilde{z}^k)] \in T_L(\tilde{z}^k)$, *ii)* $D_F\big(\tilde z^k,(F')^{-1}\big[F'(\tilde z^k)-\lambda_k^{-1}e^k\big]\big)\le \sigma D_F(\tilde z^k,z^k)$,

iii) $z^{k+1}=(F')^{-1}\big[F'(\tilde z^k)-\lambda_k^{-1}e^k\big]$.

Proof. Using the definition of $\hat L$ in (61) and Lemma 5(b)–(d) we have

$$\begin{aligned}
& e^k+\lambda_k[F'(z^k)-F'(\tilde z^k)] \\
= \;& \Big(\hat L'_x(\tilde x^k,x^k,y^k,\lambda_k)+\lambda_k[f'(x^k)-f'(\tilde x^k)],\lambda_k[h'_r(y^k)-h'_r(Q(\tilde x^k,y^k,\lambda_k))]\Big) \\
= \;& \Big(\bar L'_x(\tilde x^k,y^k,\lambda_k),-G(\cdot,\tilde x^k)+\lambda_k M_r^-(\tilde x^k,y^k,\lambda_k)\Big) \\
= \;& \Big(L'_x(\tilde x^k,Q(\tilde x^k,y^k,\lambda_k)),-G(\cdot,\tilde x^k)+\lambda_k M_r^-(\tilde x^k,y^k,\lambda_k)\Big)\in T_L(\tilde z^k),
\end{aligned}$$

establishing (i). Also,

$$\begin{aligned}
& D_F\big(\tilde z^k,(F')^{-1}\big[F'(\tilde z^k)-\lambda_k^{-1}e^k\big]\big)= \\
& D_F\big(\big(\tilde x^k,Q(\tilde x^k,y^k,\lambda_k)\big),\big((f')^{-1}\big[f'(\tilde x^k)-\lambda_k^{-1}\hat L'_x(\tilde x^k,x^k,y^k,\lambda_k)\big],Q(\tilde x^k,y^k,\lambda_k)\big)\big)= \\
& D_f\big(\tilde x^k,(f')^{-1}\big[f'(x^k)-\lambda_k^{-1}\bar L'_x(\tilde x^k,y^k,\lambda_k)\big]\big)+D_{h_r}\big(Q(\tilde x^k,y^k,\lambda_k),Q(\tilde x^k,y^k,\lambda_k)\big)= \\
& D_f\big(\tilde x^k,(f')^{-1}\big[f'(x^k)-\lambda_k^{-1}\bar L'_x(\tilde x^k,y^k,\lambda_k)\big]\big)\le\sigma\big[D_f(\tilde x^k,x^k)+D_{h_r}\big(Q(\tilde x^k,y^k,\lambda_k),y^k\big)\big]=\sigma D_F(\tilde z^k,z^k),
\end{aligned}$$

using the error criterium given by (66) in the inequality, and the separability of F, and so (ii) holds. Finally, observe that

$$z^{k+1}=(F')^{-1}\big[F'(\tilde z^k)-\lambda_k^{-1}e^k\big]$$

if and only if

$$\begin{aligned}
y^{k+1} &= (h'_r)^{-1}\big[h'_r(Q(\tilde x^k,y^k,\lambda_k))\big], \\
x^{k+1} &= (f')^{-1}\big[f'(x^k)-\lambda_k^{-1}\bar L'_x(\tilde x^k,y^k,\lambda_k)\big]
\end{aligned}$$

if and only if

$$\begin{aligned}
y^{k+1} &= Q(\tilde x^k,y^k,\lambda_k), \\
x^{k+1} &= (f')^{-1}\big[f'(x^k)-\lambda_k^{-1}L'_x(\tilde x^k,Q(\tilde x^k,y^k,\lambda_k))\big],
\end{aligned}$$

which hold by the definition of Algorithm IV. We have proved (iii). □

Propositions 12–13 show that Algorithms III, IV inherit the properties of Algorithms I, II.

Theorem 4. *Suppose that Algorithm III (respectively Algorithm IV) stops after k steps. Then $\tilde x^k$ is an optimal solution of problem (P) and $Q(\tilde x^k,y^k,\lambda_k)$ is an optimal solution of problem (D), with Q as given by* (65).

Proof. Immediate from Proposition 12 (respectively 13) and Theorem 1. □

Theorem 5. *Let $\{z^k\}$ be the sequence generated by Algorithm III for conclusion (a) and by Algorithm IV for conclusion (b). Assume that f satisfies H4. If z^k is not a KKT-pair for (P)–(D), then*

a) If f is totally convex and Fréchet differentiable then there exists an open subset $U_k \subset B$ such that any $x \in U_k$ solves (63)-(64).

b) If $(f')^{-1}$ is continuous then there exists an open subset $U_k \subset B$ such that any $x \in U_k$ solves (66).

Proof. Let $\bar{x}^k$ denote the exact solution of (63) whose existence is ensured by Proposition 11, and $\bar{z}^k = (\bar{x}^k, Q(\bar{x}^k, y^k, \lambda_k))$, where Q is as in (65). Then $\bar{z}^k \neq z^k$, because otherwise, by Proposition 12, $0 \in T_L(z^k)$, in contradiction with the assumption that z^k is not a KKT-pair. Hence, $D_F(\bar{z}^k, z^k) > 0$, with $F(x,y) = f(x) + h_r(y)$, and by total convexity of F, resulting from Proposition 3(ii), $\alpha_k := \sigma\lambda_k \min\{D_F(\bar{z}^k, z^k), \nu_F(z^k, 1)\} > 0$. Observe that the assumptions on the data functions of problem (P) and Fréchet differentiability of f ensure continuity of the function $\psi_k : B \to \mathbb{R}$ defined as

$$\psi_k(x) = \|L'_x(x, Q(x, y^k, \lambda_k)) + \lambda_k[f'(x) - f'(x^k)]\|_{B^*} -$$
$$\sigma\lambda_k \min\{D_f(x, x^k) + D_{h_r}(Q(x, y^k, \lambda_k), y^k), \nu_F(x^k, 1)\}.$$

Also, $\psi_k(\bar{x}^k) = 0 - \alpha_k < 0$, and consequently there exists $\delta_k > 0$ such that $\psi_k(x) \leq 0$ for all $x \in U_k := \{x \in B : ||x - \bar{x}^k|| < \delta_k\}$. The result in (a) follows then from the inequality above. Item (b) is proved with a similar argument, using now Proposition 13 and, instead of ψ_k, the auxiliary function $\bar{\psi}_k : B \to \mathbb{R}$ defined as

$$\bar{\psi}_k(x) = D_f\big(x, (f')^{-1}\big[f'(x^k) - \lambda_k^{-1} L'_x(x, Q(x, y^k, \lambda_k))\big]\big)$$
$$-\sigma\big[D_f(x, x^k) + D_{h_r}\big(Q(x, y^k, \lambda_k), y^k\big)\big].$$

□

Theorem 6. *Take $f \in \mathcal{F}$ satisfying H1-H4 and such that dom $f = B$, and $\lambda_k \leq \bar{\lambda}$. Let $\{z^k\}$ be the sequence generated by Algorithm III (respectively Algorithm IV). If there exist KKT-pairs for problems (P) and (D), then*

i) The sequence $\{z^k\} = \{(x^k, y^k)\}$ is bounded and all its weak accumulation points are optimal pairs for problems (P) and (D).

ii) If f also satisfies H5 and either

a) Ω is countable and $r = q$ (i.e. the dual variables belong to ℓ_q and the dual regularization function is $h_q(y) = \frac{1}{q}||y||_q^q$), or *b) $p = q = r = 2$, (i.e. the dual variables belong to $\mathcal{L}^2(\Omega)$ and the dual regularization function is $h_2(y) = \frac{1}{2}||y||_2^2$),*

then the whole sequence $\{z^k\} = \{(x^k, y^k)\}$ converges weakly to an optimal pair.

Proof. By Proposition 12 (respectively 13) the sequence $\{z^k\}$ is a particular instance of the sequences generated by Algorithm I (respectively Algorithm II) for finding zeroes of the operator T_L, with regularizing function $F : B \times \mathcal{L}^q(\Omega) \to \mathbb{R}$ given by $F(x, y) = f(x) + ||y||_q^r$. By Propositions 2(i) and 3(iii), F satisfies H1–H4. By Proposition 2(ii), it also satisfies H5 under conditions (a) or (b) of (ii). By our assumption on existence of KKT-pairs, Proposition 10(i) and 10(iii), T_L is a maximal monotone operator with zeroes. The result follows then from Theorem 2 (respectively 3) and Proposition 10(ii). □

References

[1] Alber, Ya.I, Burachik, R.S., Iusem, A.N. A proximal point method for nonsmooth convex optimization problems in Banach spaces. *Abstract and Applied Analysis* **2** (1997) 97-120.

[2] Araujo, A. The non-existence of smooth demands in general Banach spaces. *Jornal of Mathematical Economics* **17** (1988) 309-319.

[3] Auslender, A., Teboulle, M., Ben-Tiba, S., A logarithmic-quadratic proximal method for variational inequalities. *Computational Optimization and Applications* **12** 1-3 (1999) 31-40.

[4] Auslender, A., Teboulle, M., Ben-Tiba, S. Interior proximal and multiplier methods based on second order homogeneous kernels. *Mathematics of Operations Research* **24** 3 (1999) 645-668.

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

[6] Bertsekas, D.P. *Constrained Optimization and Lagrange Multipliers.* Academic Press, New York (1982).

[7] Browder, F.E. Nonlinear operators and nonlinear equations of evolution in Banach spaces. *Proceedings of Symposia in Pure Mathematics, American Mathematical Society* **18**,2 (1976).

[8] Burachik, R.S., Iusem, A.N. A generalized proximal point algorithm for the variational inequality problem in a Hilbert space. *SIAM Journal on Optimization* **8** (1998) 197-216.

[9] Burachik, R.S., Iusem, A.N., Svaiter, B.F. Enlargement of monotone operators with applications to variational inequalities *Set-Valued Analysis* **5** (1997) 159-180.

[10] Burachik, R.S., Scheimberg, S. A proximal point algorithm for the variational inequality problem in Banach spaces. *SIAM Journal on Control and Optimization* **39** 5 (2001) 1633-1649.

[11] Burachik, R.S., Svaiter, B.F. A relative error tolerance for a family of generalized proximal point methods (to be published).

[12] 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.

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

[14] Butnariu, D., Iusem, A.N. and Resmerita, E. Total convexity of the powers of the norm in uniformly convex Banach spaces. *Journal of Convex Analysis* **7** 2 (2000) 319-334.

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

[16] Censor, Y., Zenios, S.A. The proximal minimization algorithm with D-functions. *Journal of Optimization Theory and Applications* **73** (1992) 451-464.

[17] Chen, G., Teboulle, M. Convergence analysis of a proximal-like optimization algorithm using Bregman functions. *SIAM Journal on Optimization* **3** (1993) 538-543.

[18] Ciorãnescu, I. *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*. Kluwer, Dordrecht, The Netherlands (1990).

[19] Cohen, G. Auxiliary problem principle and decomposition in optimization problems. *Journal of Optimization Theory and Applications* **32** (1980) 277-305.

[20] Cohen, G. Auxiliary problem principle extended to variational inequalities. *Journal of Optimization Theory and Applications* **59** (1988) 325-333.

[21] Eckstein, J. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Mathematics of Operations Research* **18** (1993) 202-226.

[22] Eckstein, J., Humes Jr, C., Silva, P.J.S. Rescaling and stepsize selection in proximal methods using separable generalized distances. Technical Report RRR-35-99, RUTCOR, Rutgers University (1999).

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

[24] Humes Jr, C., Silva, P.J.S., Svaiter, B.F. Some inexact hybrid proximal augmented Lagrangian methods (to be published).

[25] Iusem, A.N. On some properties of generalized proximal point methods for quadratic and linear programming. *Journal of Optimization Theory and Applications* **85** (1995) 593-612.

[26] Iusem, A.N. On some properties of generalized proximal point methods for variational inequalities. *Journal of Optimization Theory and Applications* **96** (1998) 337-362.

[27] Iusem, A.N., Isnard, C.A. On mixed Hölder-Minkowski inequalities and total convexity of certain functions in $\mathcal{L}^p(\Omega)$. *Mathematical Inequalities and Applications* **3** 4 (2000) 519-537.

[28] Iusem, A.N., Svaiter, B.F., Teboulle, M. Entropy-like proximal methods in convex programming. *Mathematics of Operations Research* **19** (1994) 790-814.

[29] Iusem, A.N., Teboulle, M. Convergence rate analysis of nonquadratic proximal and augmented Lagrangian methods for convex and linear programming. *Mathematics of Operations Research* **20** (1995) 657-677.

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

[31] Kiwiel, K. Proximal minimization methods with generalized Bregman functions. *SIAM Journal on Control and Optimization* **35** (1997) 1142-1168.

[32] 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.

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

[34] Mikusinski, J. *The Bochner Integral*. Academic Press, New York (1978).

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

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

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

[38] Rockafellar, R.T. Monotone operators associated with saddle functions and minimax problems. *Proceedings of Symposia in Pure Mathematics, American Mathematical Society* **18**, Part 1 (1970) 241-250.

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

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

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

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

[43] Solodov, M.V., Svaiter, B.F. An inexact hybrid generalized proximal point algorithms and some new results on the theory of Bregman functions. *Mathematics of Operations Research* **51** (2000) 479-494.

[44] Zeidler, E. *Nonlinear Functional Analysis and its Applications* II/B. Springer, New York (1990).