# Augmented Lagrangian methods for cone-constrained convex optimization in Banach spaces

Alfredo Iusem* Rolando Gárciga Otero†

May 21, 2001

**Abstract**

We consider a general optimization problem of the form $\min g(x)$ subject to $-G(x) \in K$, with $g : B_1 \to \mathbb{R}$, $G : B_1 \to \mathrm{B}_2$, where $B_1$ and $B_2$ are real reflexive Banach spaces and $K$ is a nonempty closed convex cone in $B_2$. We introduce an inexact augmented Lagrangian method for this problem, extending to cone constrained problems recently introduced error criteria for the subproblems of the proximal point method. We present a full convergence analysis.

**Keywords:** Augmented Lagrangian, cone-constrained optimization, inexact solutions.
**1991 Mathematics Subject Classification:** 90C25, 90C30, 49J40, 46M37.

## 1 Introduction

We are concerned in this paper with he following problem:

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

$$\text{s.t. } G(x) \precsim 0, \tag{2}$$

where $g : B_1 \to \mathbb{R}$, $G : B_1 \to \mathrm{B}_2$, $B_1$ and $B_2$ are real reflexive Banach spaces and "$\precsim$" denotes the partial order relation induced by a nonempty convex cone $K$ in $B_2$(i.e. $z \precsim z'$ if and only if $z' - z \in K$). We will introduce an augmented Lagrangian method for this problem. Augmented Lagrangian methods for the finite dimensional case (i.e. when $B_1 = \mathbb{R}^n$, $B_2 = \mathbb{R}^m$ and $K$ is the nonnegative orthant of $\mathbb{R}^m$), started with [9] (previously, augmented Lagrangian methods had

*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

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

been proposed for equality constrained problems, e.g. [13] and [21]). Such algorithms were further studied in [23], [24], [2], [17]. Its connection with the Proximal Point method was established in [25]. The augmented Lagrange functional $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_{++} \to \mathbb{R}$, used in [24] and [25], is defined as

$$L(x, y, \rho) = g(x) + \rho \sum_{i=1}^{m} \left[ \left( \max\{0, y_i + (2\rho)^{-1} g_i(x)\} \right)^2 - y_i^2 \right], \tag{3}$$

where $g_i : \mathbb{R}^n \to \mathbb{R}$ $(1 \leq i \leq m)$ denote the components of $G$, i.e. for the problem

$$\min\ g(x) \tag{4}$$

$$\text{s.t. } g_i(x) \leq 0 \qquad (1 \leq i \leq m). \tag{5}$$

The augmented Lagrangian method in [25] uses an exogenous bounded sequence $\{\lambda_k\} \subset \mathbb{R}_{++}$ and it generates a sequence $\{(x^k, y^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m$ through the iterative formulae

$$x^{k+1} \in \operatorname{argmin}_{x \in \mathbb{R}^n} L(x, y^k, \lambda_k), \tag{6}$$

$$y_i^{k+1} = \max\{0, y_i^k + (2\lambda_k)^{-1} g_i(x^{k+1})\} \quad (1 \leq i \leq m). \tag{7}$$

It is proved in [25] that, if $g$ and the $g_i$'s $(1 \leq i \leq m)$ are convex and problem (4)–(5) has both primal and dual solutions, then the sequence $\{y^k\}$ converges to a dual solution (i.e. to a vector of KKT multipliers for the constraints in (5)). Regarding the sequence $\{x^k\}$, it may fail to exist, even when the problem has both primal and dual solutions, and it may be unbounded, but if it exists, it is a minimizing sequence, i.e. $\lim_{k\to\infty} g(x^k) = \inf_{x \in F} g(x)$, where $F = \{x \in \mathbb{R}^n : g_i(x) \leq 0\ (1 \leq i \leq m)\}$, and its cluster points, if any, are solutions of problem (4)–(5). The results on the primal sequence can be improved by considering a doubly augmented Lagrangian, i.e. replacing (6) by

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \{L(x, y^k, \lambda_k) + \lambda_k \|x - x^k\|^2\}, \tag{8}$$

in which case it is proved in [25] that $\{y^k\}$ still enjoys the same convergence properties as before, while $\{x^k\}$ is now well defined and converges to a solution of problem (4)–(5), provided it has both primal and dual solutions.

The results in [24] were extended in [29] to the case in which $B_1$ is a real Banach space and $B_2$ a Hilbert space $H$, in the cone constrained format of (2). Denoting by $P_{K^*}$ the orthogonal projection onto $K^*$, where $K^*$ is the positive dual cone of $K$, the augmented Lagrangian $\Lambda : B_1 \times H \times \mathbb{R}_{++} \to \mathbb{R}$ of [29] takes the form

$$\Lambda(x, y, \rho) = g(x) + \frac{1}{2}\rho \|P_{K^*}[y + \rho^{-1} G(x)]\|^2 - \frac{1}{2}\rho \|y\|^2. \tag{9}$$

The methods in [29] update the primal variables through

$$x^{k+1} \in \operatorname{argmin}_{x \in B_1} \Lambda(x, y^k, \lambda_k). \tag{10}$$

Several alternatives are given for the updating of the dual variables, one of which, namely

$$y^{k+1} = P_{K^*}\big(y^k + \lambda_k^{-1} G(x^{k+1})\big), \tag{11}$$

is such that the method reduces to the algorithm in [25], i.e. (6)–(7), when $B_1 = \mathbb{R}^n$, $H = \mathbb{R}^m$ and $K = \mathbb{R}^m_+$. Several alternatives are presented for the regularization parameters $\lambda_k$, in some of which they are endogenously updated. The results in [29] strongly depend on the fact that each element $y \in H$ can be uniquely decomposed into two orthogonal components, $y = P_K(y) + P_{-K^*}(y)$. This is not true in our case, i.e. when $B_2$ is not hilbertian. Convergence results for all algorithms in this reference require either that $\lim_{k\to\infty} \lambda_k = 0$ or at least that $\lambda_k$ be small enough for large $k$. Additionally, since the Augmented Lagrangian $\Lambda$ in (9) lacks primal regularization, not much can be said about the primal sequence $\{x^k\}$, besides the fact that, if it exists, it is a minimizing one. Convergence of $\{y^k\}$ to a dual solution, under some regularity conditions on the problem, is established for all variants under consideration.

Recently, a case with nonhilbertian $B_2$ was studied in [8]. It considers an arbitrary reflexive Banach space $B_1$, $B_2 = \mathcal{L}^p(\Omega)$ and $K = \mathcal{L}^p_+(\Omega) = \{z \in \mathcal{L}^p(\Omega) |\ z(\omega) \geq 0\ \mu \text{ a.e.}\}$, where $(\Omega, \mathcal{A}, \mu)$ is a measure space and $p \in (1, \infty)$, so that the problem becomes

$$\min\ g(x)$$

$$\text{s.t. } G(x, \omega) \leq 0 \quad \mu \text{ a.e.},$$

with $G(\cdot, \omega)$ convex for all $\omega \in \Omega$. The regularizing function $h : \mathcal{L}^q(\Omega) \to \mathbb{R}$, with $q = p/(p-1)$, defined as $h(y) = 1/r \|y\|^r$ $(r > 1)$, is used to introduce the augmented Lagrangian $\bar{\Lambda} : B \times \mathcal{L}^q(\Omega) \times \mathbb{R}_{++} \to \mathbb{R}$ defined as

$$\bar{\Lambda}(x, y, \rho) = g(x) + \frac{\rho}{p} \left\| P_{\mathcal{L}^p_+(\Omega)}[h'(y) + \rho^{-1} G(x)] \right\|_p^p, \tag{12}$$

The Lagrangian of (12) is used in [8] in order to generate an augmented Lagrangian method. As before, $\{\lambda_k\} \subset \mathbb{R}_{++}$ is a bounded sequence. For the sake of simplicity, we present next the updating formulae only for the case $r = q$. They are

$$x^{k+1} \in \operatorname{argmin}_{x \in B} \bar{\Lambda}(x, y^k, \lambda_k), \tag{13}$$

$$y^{k+1}(\omega) = \big[\max\{0, [h'(y^k) + \lambda_k^{-1} G(x^{k+1})](\omega)\}\big]^{p-1}. \tag{14}$$

We observe that, up to constant terms in $x$, (12) reduces to (9) when $p = r = 2$, so that $B_2$ is hilbertian, and henceforth to (3), when $B_1$ is finite dimensional and $\Omega$ is finite. Thus, in such cases (13)–(14) reduce to (10)–(11) and to (6)–(7) respectively.

An interesting point is that, though $\mathcal{L}^p(\Omega)$ is not hilbertian, the cone $\mathcal{L}^p_+(\Omega)$ still induces a decomposition in $\mathcal{L}^p(\Omega)$, in the sense that each $y \in \mathcal{L}^p(\Omega)$ admits a unique decomposition $y = P_{\mathcal{L}^p_+(\Omega)}(y) + P_{-\mathcal{L}^p_+(\Omega)}(y)$.

The convergence analysis in [8], based upon the Proximal Point methods in Banach spaces studied in [1] and [7], establishes that, assuming that the problem has both primal and dual solutions, the sequence $\{y^k\}$ is bounded and all its weak cluster points are dual solutions, with uniqueness of the weak cluster point when $\Omega$ is countable, so that $\mathcal{L}^q(\Omega) = \ell_q$. There is no primal regularization in (12), so that the results on $\{x^k\}$ only prove that if it exists, then it is a minimizing sequence, and if it is bounded, then all its weak cluster point are primal solutions. These limitations were avoided in [14] by adding a regularizing term to the augmented Lagrangian, which play a role similar to the term $\|x - x^k\|^2$ in the doubly augmented Lagrangian (8) of [25].

In this paper, we introduce two main improvements over the results discussed above. The first one consists of considering a general real reflexive Banach space $B_2$, so that the cone $K$ possibly does not induce a decomposition in $B_2$. With this purpose, we consider the regularizing function $h_r : B_2^* \to \mathbb{R}$ defined as $h_r(y) = \frac{1}{r}||y||^r_{B_2^*}$ for some $r \in (1, \infty)$ and the auxiliary mapping $M_r : B_1 \times B_2^* \times \mathbb{R}_{++} \to B_2$ given by $M_r(x, y, \rho) = h_r'(y) + \rho^{-1}G(x)$, with which we define the augmented Lagrangian $\bar{L} : B_1 \times B_2^* \times \mathbb{R}_{++} \to \mathbb{R}$ as $\bar{L}(x, y, \rho) = g(x) + \rho\frac{1}{s}d(M_r(x, y, \rho), -K)^s$, where $s = r/(r-1)$, which allows us to conveniently extend the augmented Lagrangian method to this context.

Our second main goal consists of extending to the case of cone constrained convex optimization in Banach spaces inexact versions of the augmented Lagrangian methods, incorporating recently developed error criteria for the proximal point method, which allow for constant relative errors. Up to know, we have been discussing the exact versions of the methods (e.g. $x^{k+1}$ is an exact minimizer in (6), (10) and (13)). At least since [25], convergence of the Proximal Point method under inexact solutions of the subproblems has been considered. In [25], convergence of the proximal sequence $\{x^k\}$ to a solution of the problem is ensured under the assumption that $\sum_{k=0}^{\infty} ||\tilde{x}^k - x^k|| < \infty$, where $x^k$ is the inexact solution of the $k$-th proximal subproblem and $\tilde{x}^k$ the exact one. Other related conditions for the case of quadratic $h$, but including always summability conditions on the error, can be found in [25], and [15] for finite dimensional Hilbert and Banach spaces respectively, and, for the case of nonquadratic $h$, in [12], [16] and [3]. 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 [26] for the case of quadratic $h$ in Hilbert spaces and in [27], [5] for the case of nonquadratic $h$ in finite dimensional spaces. The error criteria for the Proximal Point method introduced in [26] and [27] were adapted for Banach spaces in [14], and consequently transferred to two augmented Lagrangian methods for the particular case of the problem studied in [8], i.e. with $B_2 = \mathcal{L}^p(\Omega)$ and $K = \mathcal{L}^p_+(\Omega) = \{z \in \mathcal{L}^p(\Omega) |\ z(\omega) \geq 0\ \mu \text{ a.e.}\}$. In this paper, we will consider such error criteria for our augmented Lagrangian method applied to

the general problem (1)–(2), i.e. with a general Banach space $B_2$ and a general cone $K \subset B_2$.

## 2 Preliminary results

Given a nonempty, closed and convex subset $C$ of a reflexive real Banach space $B$, the *metric projection* $P_C : B \to \mathcal{P}(C)$ of $B$ onto $C$ is defined as

$$P_C(x) = \text{argmin}_{y \in C} \frac{1}{2} ||x - y||_B^2, \tag{15}$$

i.e. $P_C(x) = \{z \in C \ : ||x - z||_B = d(x, C)\}$ where $d(\cdot, C) : B \to \mathbb{R}$ is the distance to $C$, given by

$$d(x, C) = \inf_{y \in C} ||x - y||_B. \tag{16}$$

The metric projection is single valued when $B$ is strictly convex. Existence of points attaining the distance in (16) (and so a nonempty valued $P_C$) is ensured by reflexivity of $B$. We recall next some properties of duality mappings in Banach spaces.

A *weight function* $\varphi$ 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* $\varphi$ 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 $\varphi(t) = t$, $J_\varphi$ is called the *normalized duality mapping* and denoted by $J$. We will use in the sequel the following property of a Banach space $B$, which is known to hold, e.g. when $B$ is locally uniformly convex.

(H) $\quad x^k \xrightarrow[k \to \infty]{s} x$ whenever $||x^k||_B \overset{k \to \infty}{\longrightarrow} ||x||_B$ and $x^k \xrightarrow[k \to \infty]{w} x$.

When $B$ is reflexive, 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 $\varphi^{-1}$, the inverse function of the weight $\varphi$, then $\varphi^{-1}$ is a weight function too and $J^*_{\varphi^{-1}}$, the duality mapping of weight $\varphi^{-1}$ 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 $\varphi_1$ and $\varphi_2$ respectively, then, for all $x \in B$, $\varphi_2(||x||) J_1 x = \varphi_1(||x||) J_2 x$.

J4: If $B$ is smooth (i.e., if $B^*$ is strictly convex), then $J_\varphi$ is single valued and norm-to-weak* continuous. Moreover, if $B^*$ satisfies property (H) then $J_\varphi$ is norm-to-norm continuous.

Properties J1–J4 have been proved in [11].

**Proposition 1.** *Let $B$ be a a real reflexive Banach space. Take $x \in B$ and a nonempty, closed and covex $C \subset B$. Then*

*i)* $z \in P_C(x)$ *if and only if* $J(x-z) \cap N_C(z) \neq \emptyset$, *where* $N_C$ *is the normalizing operator of* $C$, *i.e. the subdiferential of the indicator function* $I_C$ *of* $C$, *defined as* $I_C(x) = 0$ *if* $x \in C$, $I_C(x) = \infty$ *otherwise.*

*ii)* $d(\cdot, C)$ *is a convex function and its subdifferential at* $x$ *is given, for any* $z \in P_C(x)$, *by*

$$\partial d(\cdot, C) = \partial ||\cdot - z||_B \cap N_C(z).$$

**Proof.** For (i) see the equivalence between (a) and (c) in Theorem 3.1 of [19]. Item (ii) is proved in Theorem 1 of [6]. □

**Corollary 1.** *Let* $C$ *a nonempty, closed and convex subset of* $B$ *and consider the convex function* $h = \frac{1}{s} d(\cdot, C)^s : B \to \mathbb{R}$ *with* $s \in (1, +\infty)$. *If* $B$ *is a strictly convex and smooth reflexive Banach space then*

*i)* $h$ *is Gâteaux differentiable and its Gâteaux derivative* $h'$ *is given by*

$$h'(x) = ||x - P_C(x)||_B^{s-2} J(x - P_C(x)) = J_s(x - P_C(x))$$

*for all* $x \in B$, *where* $J_s$ *is the duality map of weight* $\varphi(t) = t^{s-1}$.

*ii) If* $B$ *and* $B^*$ *satisfy property (H), then* $h'$ *is norm-to-norm continuous, and hence* $h$ *is Fréchet differentiable.*

**Proof.** Strict convexity of $B$ ensures that the projection operator is single valued, so that from Proposition 1(ii) we get, for any $x \in B$,

$$\partial d(x, C) = \partial ||x - P_C(x)||_B \cap N_C(P_C(x)). \tag{17}$$

By the chain rule of differentiation, we have

$$\partial h(x) = d(x, C)^{s-1} \partial d(x, C). \tag{18}$$

Thus, in view of (17) and (18), we have $\partial h(x) = \{0\}$ when $x = P_C(x)$ (i.e $x \in C$), and it suffices to study the case of $x \neq P_C(x)$. Remembering that $J = \partial(1/2||\cdot||^2)$, we get

$$\partial h(x) = d(x, C)^{s-1} \left[ ||x - P_C(x)||^{-1} J(x - P_C(x)) \cap N_C(P_C(x)) \right] =$$

$$d(x, C)^{s-1} ||x - P_C(x)||^{-1} \left[ J(x - P_C(x)) \cap N_C(P_C(x)) \right] =$$

$$||x - P_C(x)||^{s-2} \left[ J(x - P_C(x)) \cap N_C(P_C(x)) \right] = ||x - P_C(x)||^{s-2} J(x - P_C(x)), \tag{19}$$

where we use the definition of $N_C$ in order to get $||x - P_C(x)|| N_C(P_C(x)) = N_C(P_C(x))$ in the second equality above, and (15) in the third equality. The last equality follows from Proposition

1(i). Note that we have, by definition of $J$, $||x - P_C(x)||^{s-2}\|J(x - P_C(x))\|_* = ||x - P_C(x)||^{s-1}$, so that (19) also holds for the case $x = P_C(x)$. Observe now that $J$ is single valued, by smoothness of $B$ and (J4). It follows from (19) that $h$ is Gâteaux differentiable, and the first equality in (i) is proved. The second equality in (i) follows from property J3.

For (ii), see Lemma 1 of [20] for continuity of the metric projection $P_C$ which, together with J4, ensures the norm-to-norm continuity of $h'$. □

Take now real Banach spaces $B_1$ and $B_2$ and let $K \subset B_2$ be a a nonempty, closed and convex cone. A mapping $M : B_1 \to B_2$ is said to be *$K$-convex* if and only if $\alpha M(x) + (1-\alpha)M(x') - M(\alpha x + (1-\alpha)x') \in K$, for all $x,\ x' \in X$ and all $\alpha \in [0,1]$. We remind that, given a linear operator $A : B_1 \to B_2$, the adjoint $A^* : B_2^* \to B_1^*$ is defined as $A^*(y) = y \circ A$.

**Proposition 2.** *Take $B_1$, $B_2$ and $K$ as above, assume that $B_2$ is reflexive, let $M : B_1 \to B_2$ be $K$-convex and define $V : B_1 \to \mathbb{R}_+$ as $V = (1/s)d(\cdot, -K)^s \circ M$, $(s > 1)$. Then*

*i) $V$ is convex.*

*ii) If $B_2$ is strictly convex and smooth and also $B_2$ and $B_2^*$ have property (H) (e.g. if $B_2$ is locally uniformly convex and locally uniformly smooth), then*

*a) If $M$ is Gâteaux differentiable, with Gâteaux derivative $M'$, then $V$ is Gâteaux differentiable too and its Gâteaux derivative is given by*

$$V'(x) = [M'(x)]^* J_s \left(M(x) - P_{-K}(M(x))\right).$$

*b) If $M$ is Fréchet differentiable, then $V$ is Fréchet differentiable, so that $V'$ is norm-to-norm continuous.*

**Proof.** i) Clearly, it suffices to prove convexity of, $\tilde{V} = d(\cdot, -K) \circ M$. Take $x,\ x' \in B_1$ and $\alpha \in [0,1]$. By reflexivity of $B$, there exist $y,\ y' \in -K$ such that $||M(x) - y||_{B_2} = d(M(x), -K)$ and $||M(x') - y'||_{B_2} = d(M(x'), -K)$. By convexity of $-K$, $\alpha y + (1-\alpha)y' \in -K$, and therefore

$$\bar{y} + \alpha y + (1-\alpha)y' \in -K \ \ for\ all\ \bar{y} \in -K. \tag{20}$$

Define $y_\alpha = M(\alpha x + (1-\alpha)x') - [\alpha M(x) + (1-\alpha)M(x')]$. By $K$-convexity of $M$, $y_\alpha \in -K$, and

$$\tilde{V}(\alpha x + (1-\alpha)x') \stackrel{def}{=} d(M(\alpha x + (1-\alpha)x'), -K) =$$

$$d(y_\alpha + [\alpha M(x) + (1-\alpha)M(x')], -K) \stackrel{def}{=} \inf_{\tilde{y} \in -K} ||y_\alpha + [\alpha M(x) + (1-\alpha)M(x')] - \tilde{y}||_{B_2}. \tag{21}$$

Choosing $\bar{y} = y_\alpha$ we get, from (20)-(21),

$$\tilde{V}(\alpha x + (1-\alpha)x') \leq ||y_\alpha + [\alpha M(x) + (1-\alpha)M(x')] - [y_\alpha + \alpha y + (1-\alpha)y']||_{B_2} =$$

$$||\alpha[M(x)-y]+(1-\alpha)[M(x')-y']||_{B_2} \leq \alpha||M(x)-y||_{B_2}+(1-\alpha)||M(x')-y'||_{B_2} =$$

$$\alpha d(M(x),-K)+(1-\alpha)d(M(x'),-K) \stackrel{def}{=} \alpha\tilde{V}(x)+(1-\alpha)\tilde{V}(x').$$

For (ii), take $C=-K$ in Corollary 1 and use the properties of $B_2$ and the chain rule of differentiation to get Gâteaux differentiability (respectively Fréchet differentiability) of $V$ and the formula of $V'$. In fact

$$V'(x)=J_s\left(M(x)-P_C(M(x))\right)\circ M'(x)=[M'(x)]^*J_s\left(M(x)-P_{-K}(M(x))\right).$$

□

We point out that locally uniform convexity is not a demanding condition in reflexive Banach spaces. A result by Troyanski ensures that if $B$ is reflexive then there exist equivalent norms in $B$ and $B^*$, such that both spaces, so renormalized, are still mutually dual and also locally uniformly convex (see [28]).

## 3 Cone-constrained convex optimization in Banach spaces

In this section $B_j$ $(0\leq j\leq m)$ will denote $m+1$ real reflexive Banach spaces, $B_j^*$ its respective topological duals, $K_j\subset B_j$ $(1\leq j\leq m)$ nonempty, closed an convex cones and $K_j^*$ the positive dual of $K_j$ (i.e. $K_j^*=\{y\in B_j^* \ : \langle y,z\rangle_j\geq 0 \text{ for all } z\in K_j\}$, where $\langle\cdot,\cdot\rangle_j$ denotes the respective dual product). We consider the partial order relation $\precsim_j$ $(1\leq j\leq m)$, defined in $B_j$ as $z\precsim_j z'$ if and only if $z'-z\in K_j$. We consider functions $g:B_0\to\mathbb{R}$ and $G_j:B_0\to \mathrm{B}_j$ $(1\leq j\leq m)$ such that:

(A1) $g$ is convex, and $G_j$ is $K_j$-convex $(1\leq j\leq m)$.

(A2) $g$ and $G_j$ are Fréchet differentiable functions with Gâteaux derivatives denoted by $g'$ and $G_j'$, $(1\leq j\leq m)$ respectively.

The general convex optimization problem (or primal problem) will be defined as

$$(P)\ \begin{cases}\min\ g(x)\\ \text{s.t. } G_j(x)\precsim_j 0,\ 1\leq j\leq m\end{cases}$$

and its dual $(D)$, as

$$(D)\ \begin{cases}\max\ \Phi(y)\\ \text{s.t. } y\succnsim_* 0.\end{cases}$$

The dual objective $\Phi : \Pi_{j=1}^m B_j^* \to \mathbb{R} \cup \{-\infty\}$ is defined as $\Phi(y) = \inf_{x \in B_0} L(x,y)$ with the *Lagrangian* $L : B_0 \times \Pi_{j=1}^m B_j^* \to \mathbb{R}$, given by

$$L(x,y) = g(x) + \langle y, G(x) \rangle = g(x) + \sum_{j=1}^m \langle y_j, G_j(x) \rangle_j, \tag{22}$$

where $G : B_0 \to \Pi_{j=1}^m B_j$ is the application with components $G_j$, i.e.

$$G(x) = \big(G_1(x), ..., G_m(x)\big) \tag{23}$$

for all $x \in B_0$ and "$\succsim_*$" denotes the partial order induced in the topological dual $B^*$ of the product space $B = \Pi_{j=1}^m B_j$ by the closed convex cone $K^* = K_1^* \times \cdots \times K_m^*$ (i.e $y \succsim_* y'$ if and only if $y - y' \in K^*$, or equivalently, if and only if $y_j - y_j' \in K_j^*$ for all $j$).

If $B_j = \mathbb{R}$ $(1 \leq j \leq m)$ and $K_j = \mathbb{R}_+$ then the primal and dual problem are just the usual ones in real-valued convex optimization.

### 3.1 Some duality results

We start with some basic properties of the Lagrangian defined by (22).

**Proposition 3.** *Let $L$ and $G$ be defined as in* (22) *and* (23) *respectively. Then*

*i) a) for all $y \in B^*$, $L(\cdot, y)$ is Fréchet differentiable and its Gâteaux derivative is given, for all $x \in B_0$, by*

$$L_x'(x,y) = g'(x) + [G'(x)]^*(y) = g'(x) + \sum_{j=1}^m [G_j'(x)]^*(y_j),$$

*b) for all $y \in K^*$, $L(\cdot, y)$ is convex with norm-to-norm continuous Gâteaux derivative,*

*ii) for all $x \in B_0$ the function $L(x,\cdot) : B^* \to \mathbb{R}$, given by $L(x,y) = g(x) + \langle y, G(x) \rangle$ is Fréchet differentiable and its Gâteaux derivative is given by $L_y'(x,y) = G(x)$ for all $y \in B^*$.*

**Proof.** Item (i-a) follows from (A2) and elementary properties of Fréchet derivatives. For item (i-b), note that $K_j$-convexity of $G_j$ ensures convexity of the function $\langle y_j, G_j(\cdot) \rangle : B_0 \to \mathbb{R}$ whenever $y_j \in K_j^*$, so that the result follows from (A1). Item (ii) is immediate, since $L(x,\cdot) : B^* \to \mathbb{R}$ is affine. □

A pair $(x,y) \in B_0 \times B^*$ is *feasible* if $x$ is *primal feasible*, i.e. $G_j(x) \precsim_j 0$ $(1 \leq j \leq m)$ and $y$ is *dual feasible*, i.e. $y \succsim_* 0$. A pair $(x,y) \in B_0 \times B^*$ is a *KKT-pair* if it is feasible and additionally

$$0 = L_x'(x,y) = g'(x) + y \circ G'(x) \qquad \text{(Lagrangian condition)}, \tag{24}$$

$$\langle y, G(x)\rangle = 0 \qquad \text{(complementarity)}. \tag{25}$$

A pair $(x, y) \in B_0 \times B^*$ is *optimal* if $x$ is an optimal solution of problem $(P)$ and $y$ an optimal solution of problem $(D)$. We present next some Lagrangian duality results.

**Proposition 4.** *i) If $y \in B^*$ is dual feasible and $(x, y) \in B_0 \times B^*$ is a saddle point of the Lagrangian, i.e.*

$$L(x, \tilde{y}) \leq L(x, y) \leq L(\tilde{x}, y) \tag{26}$$

*for all $\tilde{x} \in B_0$ and all $\tilde{y} \in K^*$, then $(x, y)$ is an optimal pair.*

*ii) If int(K)$\neq \emptyset$, there exists $\tilde{x} \in B_0$ such that $-G(\tilde{x}) \in$ int(K) and $x$ is an optimal solution of problem (P), then there exist an element $y \in K^*$ such that the Lagrangian has a saddle point at $(x, y)$.*

**Proof.** See Theorem 2 in section 8.4 and Theorem 1 in section 8.3 of [18]. □

Next we define the *saddle-point operator* $T_L : B_0 \times B^* \to \mathcal{P}(B_0^* \times B)$ as

$$T_L(x, y) = \big(L'_x(x, y), -L'_y(x, y) + N_{K^*}(y)\big) = \big(g'(x) + [G'(x)]^*(y), -G(x) + N_{K^*}(y)\big)\,, \tag{27}$$

where $N_{K^*} : B^* \to \mathcal{P}(B)$ denotes the normalizing operator of the cone $K^*$, given by

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

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

**Proposition 5.** *Under assumptions (A1) and (A2), the operator $T_L$, defined in (27), 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) conversely, if $x$ is an optimal solution of problem (P), int $K \neq \emptyset$, and there exists $\tilde{x} \in B_0$ such that $-G(\tilde{x}) \in$ int(K), then there exists an element $y \in K^*$ such that $0 \in T_L(x, y)$,*

*iv) $T_L$ is maximal monotone.*

**Proof.** For (i), it has been proved in [22] that $0 \in T_L(x, y)$ if and only if $(x, y)$ is a saddle point of $U(x, y) = L(x, y) - I_{K^*}(y)$. Thus, it suffices to prove that $(x, y)$ is a KKT-pair if and only if it is a saddle point of $U$. Suppose that $(x, y)$ is a saddle point of $U$. Then, $y$ is feasible because of the

term $I_{K^*}$ in $U$ and $(x, y)$ satisfies (26). The right inequality in (26) implies that the Lagrangian condition (24) holds. The left one implies that

$$0 \geq \langle G(x), z - y \rangle \tag{28}$$

for all $z \in K^*$, which entails feasibility of $x$ and also (25), so that $(x, y)$ is a KKT pair. Conversely, if $(x, y)$ is a KKT pair then the Lagrangian condition (24) implies, by convexity of the Fréchet differentiable function $L(\cdot, y) : B_0 \to \mathbb{R}$, the right inequality in (26), while (25) and primal feasibility of $x$ imply (28) for all $z \in K^*$, so that the left inequality in (26) also holds. Then $(x, y)$ is a saddle point of $L$ and henceforth of $U$, because of the dual feasibility of $y$, completing the proof of (i). Items (ii) and (iii) are direct consequences of Proposition 4. Item (iv) follows also from [22]. □

### 3.2 An augmented Lagrange functional

In order to define our *augmented Lagrangian* for problems $(P)$ and $(D)$, we consider regularizing functions $h_{r_j} : B_j^* \to \mathbb{R}$ defined as $h_{r_j}(y_j) = \frac{1}{r_j}||y_j||_{B_j^*}^{r_j}$ for some $r_j \in (1, \infty)$, $(1 \leq j \leq m)$. Next we introduce the auxiliary mappings $M_{r_j} : B_0 \times B_j^* \times \mathbb{R}_{++} \to B_j$ $(1 \leq j \leq m)$ given by

$$M_{r_j}(x, y_j, \rho) = h'_{r_j}(y_j) + \rho^{-1} G_j(x). \tag{29}$$

With this notation, the augmented Lagrangian $\bar{L} : B_0 \times B^* \times \mathbb{R}_{++} \to \mathbb{R}$ is defined as

$$\bar{L}(x, y, \rho) = g(x) + \rho \sum_{j=1}^{m} \frac{1}{s_j} d(M_{r_j}(x, y_j, \rho), -K_j)^{s_j}, \tag{30}$$

with $s_j = r_j/(r_j - 1)$. We will also use the mapping $M_r : B_0 \times B^* \times \mathbb{R}_{++} \to B$, with components given by $M_{r_j}$, $(1 \leq j \leq m)$, i.e.

$$M_r(x, y, \rho) = \big(M_{r_1}(x, y_1, \rho), ..., M_{r_m}(x, y_m, \rho)\big). \tag{31}$$

When $B_0$ is finite dimensional, $B_j = \mathbb{R}$, $K_j = \mathbb{R}_+$ and $r_j = 2$ $(1 \leq j \leq m)$, $L$ reduces to the standard augmented Lagrangian, up to constant terms in $x$ (e.g. [25]). When the $B_j$'s are Hilbertian and $r_j = 2$ $(1 \leq j \leq m)$, $L$ reduces to the augmented Lagrangian in [29]. When $m = 1$, $B_1 = \mathcal{L}^p(\Omega)$ for some $p > 1$ and $K_1 = \mathcal{L}^p_+(\Omega)$, $L$ reduces to the augmented Lagrangian in [14]. We establish next some elementary properties of $L$.

**Proposition 6.** *Take $\bar{L}$ as in* (30). *Then for any $(y, \rho) \in B^* \times \mathbb{R}_{++}$ the function $\bar{L}(\cdot, y, \rho) : B_0 \to \mathbb{R}$ satisfies*

*i) $\bar{L}(\cdot, y, \rho)$ is convex.*

*ii) If $B_j$ and $B_j^*$ $(1 \leq j \leq m)$ are strictly convex reflexive Banach spaces satisfying property (H), then*

*a)* $\bar{L}(\cdot, y, \rho)$ *is Fréchet differentiable and its Gâteaux derivative at any* $x \in B_0$ *is given by*

$$\bar{L}'_x(x, y, \rho) = g'(x) + \sum_{j=1}^{m} [G'_j(x)]^* J_{s_j}\left(M_{r_j}(x, y_j, \rho) - P_{-K_j}(M_{r_j}(x, y_j, \rho))\right).$$

*Moreover,* $\bar{L}'_x(\cdot, y, \rho)$ *is norn-to-norm continuous.*

*b) For any* $x \in B_0$, $\bar{L}'_x(x, y, \rho) = L'_x(x, Q(x, y, \rho))$, *where* $Q(x, y, \rho) \in B^*$ *is defined as*

$$Q_j(x, y, \rho) = J_{s_j}\big(M_{r_j}(x, y_j, \rho) - P_{-K_j}(M_{r_j}(x, y_j, \rho)\big) \quad (1 \leq j \leq m). \tag{32}$$

**Proof.** Fix $(y, \rho) \in B^* \times \mathbb{R}_{++}$, choose $C = -K_j$ and let $M = M_{r_j}(\cdot, y_j, \rho) : B_0 \to B_j$ with $M_{r_j}$ as in (29). Then (A1) ensures $K_j$ convexity of $M$ and, in view of Proposition 2(i), we get convexity of the function $V = \frac{1}{s_j} d(M_{r_j}(\cdot, y, \rho), -K_j)^{s_j} : B_0 \to \mathbb{R}$ for any $j \in \{1, ..., m\}$. (A1) implies also convexity of $g$, so that (i) is proved. For (ii-a) apply Proposition 2(ii) and the sum rule for Fréchet derivatives. Item (ii-b) follows from (ii-a) and Proposition 3(i-a). □

Proposition 6(ii-b) allows us to construct a primal-dual method for our general convex optimization problem with a closed formula for updating the dual variables. In fact, given an initial iterate $(x^0, y^0) \in B_0 \times K^*$ and a sequence $\{\lambda_k\} \subset \mathbb{R}_{++}$, we define an augmented Lagrangian method through the formulae:

$$x^{k+1} = \arg\min_{x \in B_0} \bar{L}(x, y^k, \lambda_k),$$

$$y_j^{k+1} = J_{s_j}\left(M_{r_j}(x^{k+1}, y_j^k, \lambda_k) - P_{-K_j}(M_{r_j}(x^{k+1}, y_j^k, \lambda_k))\right) \quad (1 \leq j \leq m).$$

This is an extension of (13)-(14), but as discused in [8], existence of primal iterates is not guaranteed. Then, following the approach in [25], [12], [14], we introduce a regularizing term for primal variables, through a strictly convex and Fréchet differentiable function $f : B_0 \to \mathbb{R}$, with Gâteaux derivative denoted by $f'$. In order to state the required properties of $f$, we introduce the *Bregman distance* related to $f$, $D_f : B_0 \times B_0 \to \mathbb{R}$ given by

$$D_f(x, y) = f(x) - f(y) - \langle f'(y), x - y \rangle, \tag{33}$$

and the *modulus of total convexity* $\nu_f : B_0 \times \mathbb{R}_+ \to \mathbb{R}$, defined as

$$\nu_f(x, t) = \inf_{y \in \{y \in B_0 : ||y - x|| = t\}} D_f(y, x). \tag{34}$$

$f$ is said to be *totally convex* if $\nu_f(x, t) > 0$ for all $x \in B_0$ and all $t > 0$. Each of our convergence results requires some of the following assumptions on $f$:

H1: The level sets of $D_f(x,\cdot)$ are bounded for all $x \in B_0$.

H2: $\inf_{x\in C}\nu_f(x,t) > 0$, for all bounded set $C \subset B_0$ and all $t \in \mathbb{R}_{++}$.

H3: $f'$ is uniformly continuous on bounded subsets of $B_0$.

H4: $f'$ is onto.

H5: $f' : B_0 \to B_0^*$ is weak-to-weak$^*$ continuous.

It has been proved in Proposition 2 of [14] that $f(x) = \|x\|^r$ satisfies H1–H4 in any uniformly convex and uniformly smooth Banach space for all $r > 1$. It satisfies also H5 when either $r = 2$ and the space is hilbertian, or the space is $\ell_p$ $(p > 1)$ and $r = p$.

With the help of $f$, and of $D_f$ as given by (33), we define the *doubly augmented Lagrangian* $\hat{L} : B_0 \times B_0 \times B^* \times \mathbb{R}_{++} \to \mathbb{R}$ as

$$\hat{L}(x,z,y,\rho) = \bar{L}(x,y,\rho) + \rho D_f(x,z), \tag{35}$$

which allows us to define the method below.

**Doubly augmented Lagrangian method**

1. Choose $(x^0, y^0) \in B_0 \times K^*$.
2. Given $(x^k, y^k)$, choose $\lambda_k > 0$ and define $x^{k+1}$ as

$$x^{k+1} = \operatorname{argmin}_{x\in B_0} \hat{L}(x, x^k, y^k, \lambda_k) = \operatorname{argmin}_{x\in B_0} \left[\bar{L}(x, y^k, \lambda_k) + \lambda_k D_f(x, x^k)\right]. \tag{36}$$

3. Define $y^{k+1}$ as

$$y_j^{k+1} = J_{s_j}\left(M_{r_j}(x^{k+1}, y_j^k, \lambda_k) - P_{-K_j}(M_{r_j}(x^{k+1}, y_j^k, \lambda_k))\right), \quad (1 \le j \le m). \tag{37}$$

We mention that in the three cases discussed after (31), algorithm (36)–(37) reduces to the exact versions of the augmented Lagrangian methods analyzed in [25], [29] and [14].

In the sequel, we will introduce an inexact version of the doubly augmented Lagrangian above, but we will need the following existence result for the exact version.

**Proposition 7.** *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* (36).

**Proof.** If $f$ satisfies H4, take $T = \bar{L}'_x(\cdot, y^k, \lambda_k)$ and use [4], Corollary 3.1. Otherwise, the result follows from Proposition 3.15 in [8]. □

### 3.3 Inexact versions of the doubly augmented Lagrangian method

We consider the regularizing function $F : B_0 \times B^* \to \mathbb{R}$ defined as $F(x,y) = f(x) + \sum_{j=1}^m h_{r_j}(y_j)$, where $h_{r_j}(y_j) = \frac{1}{r_j}||y_j||_{B_j^*}^{r_j}$, $(1 < r_j < \infty)$, $(1 \leq j \leq m)$, and $f : B_0 \to \mathbb{R}$ is a strictly convex and Fréchet differentiable function. We consider also $Q : B_0 \times B^* \times \mathbb{R}_{++} \to B^*$ as defined in (32) and $M_{r_j}$ as in (29). Next we extend the inexact augmented Lagrangian methods proposed in [14] to the context of cone constrained optimization.

**Algorithm I.**

1. Choose $z^0 = (x^0, y^0) \in B_0 \times B^*$ and $\sigma \in [0,1]$.

2. Given $z^k = (x^k, y^k)$, choose $\lambda_k > 0$ and find $\tilde{x}^k \in B_0$ such that

$$\|\hat{L}'_x(\tilde{x}^k, x^k, y^k, \lambda_k)\|_{B_0^*} \leq \sigma\lambda_k \begin{cases} D_F(\tilde{z}^k, z^k) & \text{if } ||z^k - \tilde{z}^k||_{B_0 \times B^*} < 1, \\ \nu_F(z^k, 1) & \text{if } ||z^k - \tilde{z}^k||_{B \times B^*} \geq 1, \end{cases} \tag{38}$$

   with $D_F$ and $\nu_F$ as in (33), (34) respectively, where

$$\tilde{z}^k = \left(\tilde{x}^k, Q(\tilde{x}^k, y^k, \lambda_k)\right). \tag{39}$$

3. Set $v^k = \left(\bar{L}'_x(\tilde{x}^k, y^k, \lambda_k), -G(\tilde{x}^k) + \lambda_k P_{-K}(M_r(\tilde{x}^k, y^k, \lambda_k))\right) \in B_0^* \times B$. If $v^k = 0$ or $\tilde{z}^k = z^k$, then stop. Otherwise, let $H_k = \{z \in B_0 \times B^* : \langle v^k, z - \tilde{z}^k \rangle = 0\}$, and

$$z^{k+1} = (x^{k+1}, y^{k+1}) = \text{argmin}_{z \in H_k} D_F(z, z^k). \tag{40}$$

Observe that the $j$-th component of $P_{-K}(M_r(\tilde{x}^k, y^k, \lambda_k))$ is given by $P_{-K_j}(M_{r_j}(x, y_j, \rho))$, which is already known at step 3 for each $j$. So we need no extra effort to compute $v^k$. Concerning the projection step, we mention that the existence of $z^{k+1}$ is ensured by total convexity of $f$, which follows from H2, and also that even though this projection step cannot be performed through a closed formula, in the cases of interest, namely $f(x) = ||x||^r$ $(r > 1)$ with uniformly convex $B_0$, 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 [10]).

As a result of the Bregman projection in (40), we can get a pair $(x^k, y^k)$ with unfeasible $y^k$ (i.e. $y^k \notin K^*$) We present next a dual feasible alternative.

**Algorithm II.**

1. Choose $(x^0, y^0) \in B_0 \times K^*$ and $\sigma \in [0,1)$.

2. Given $z^k = (x^k, y^k)$, choose $\lambda_k > 0$ and find $\tilde{x}^k \in B_0$ such that

$$D_f\big(\tilde{x}^k, (f')^{-1}\big[f'(x^k) - \lambda_k^{-1}\bar{L}'_x(\tilde{x}^k, y^k, \lambda_k)\big]\big) \leq \sigma \left[ D_f(\tilde{x}^k, x^k) + \sum_{j=1}^{m} D_{h_{r_j}}\big(Q_j(\tilde{x}^k, y^k, \lambda_k), y_j^k\big) \right] \tag{41}$$

3. If $(\tilde{x}^k, Q(\tilde{x}^k, y^k, \lambda_k)) = (x^k, y^k)$, then stop. Otherwise, define

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

We will need in the sequel the following result.

**Lemma 1.** *Assume that the Banach spaces $B_j$'s ($1 \leq j \leq m$) are strictly convex, smooth and reflexive. Then, for all $(x, y, \rho) \in B_0 \times B^* \times \mathbb{R}_{++}$ and all $j \in \{1, ...m\}$ , it holds that*

*a)* $h'_{r_j}\left(Q_j(x, y, \rho)\right) = h'_{r_j}(y_j) - \frac{1}{\rho}\left[-G(x) + \rho P_{-K_j}\big(M_{r_j}(x, y_j, \rho)\right]$,

*b)* $Q(x, y, \rho) \in K^*$,

*c)* $P_{-K}\big(M_r(x, y, \rho)\big) \in N_{K^*}\left(Q(x, y, \rho)\right)$.

**Proof.** Let $J^*_{r_j}$ denote the duality map of weight $\varphi(t) = t^{r_j - 1}$ on $B_j^*$. The assumptions on the $B_j$'s ensure that $J^*_{r_j} = h'_{r_j}$ and that $J^*_{r_j}$ is single valued, onto and invertible, with inverse $[J^*_{r_j}]^{-1}$ given by the duality map associated to the weight function $\varphi(t)^{-1} = t^{s_j - 1}$ with $s_j = r_j/(r_j - 1)$, i.e. $[J^*_{r_j}]^{-1} = J_{s_j}$ (see properties J4, J2 and J1 in section 2). Hence

$$h'_{r_j}\left(J_{s_j}\left(M_{r_j}(x, y_j, \rho) - P_{-K_j}\big(M_{r_j}(x, y_j, \rho)\big)\right)\right) = M_{r_j}(x, y_j, \rho) - P_{-K_j}\big(M_{r_j}(x, y_j, \rho)\big) =$$

$$h'_{r_j}(y_j) + \rho^{-1}G_j(x) - P_{-K_j}\big(M_{r_j}(x, y_j, \rho)\big) = h'_{r_j}(y_j) - \frac{1}{\rho}\left[-G_j(x) + \rho P_{-K_j}\big(M_{r_j}(x, y_j, \rho)\big)\right]$$

and (a) is proved.

For proving (b), take $\bar{z} \in -K$ such that $\bar{z} = P_{-K}\big(M_r(x, y, \rho)\big)$ and $\bar{y} = J_s\left(M_r(x, y, \rho) - \bar{z}\right)$. First, observe that, by Proposition 1(i), $J\left(M_r(x, y, \rho) - \bar{z}\right)$ belongs to the cone $N_{-K}(\bar{z})$. Hence $\bar{y} = ||M_r(x, y, \rho) - \bar{z}||_B^{s-2} J\left(M_r(x, y, \rho) - \bar{z}\right) \in N_{-K}(\bar{z})$. Therefore,

$$\langle \bar{y}, z - \bar{z} \rangle \leq 0, \tag{42}$$

for all $z \in -K$. Applying (42) with $z = 0$ and $z = 2\bar{z}$, which belong to $-K$ because $K$ is a closed cone, we get $\langle \bar{y}, \bar{z} \rangle = 0$, which in turn implies, in view of (42),

$$\langle \bar{y}, z \rangle \leq \langle \bar{y}, \bar{z} \rangle = 0,$$

for all $z \in -K$, or equivalently $\langle \bar{y}, z' \rangle \geq 0$ for all $z' \in K$, establishing (b).

Fix now an arbitrary $y \in K^*$. Then $\langle \bar{z}, y - \bar{y} \rangle = \langle \bar{z}, y \rangle - \langle \bar{z}, \bar{y} \rangle = \langle \bar{z}, y \rangle \leq 0$, where the inequality follows from the definition of $K^*$, together with the fact that $-\bar{z} \in K$, completing the proof. □

Item (b) of Lemma 1 corroborates dual feasibility of Algorithm II, wich also has, as an advantage over Algorithm I, a closed formula to update primal and dual variables. As a disadvantage, we observe that the error criterion in Algorithm II implicitly requires repeated extragradient steps in order to find an aproximate zero of $\hat{L}(\cdot, x^k, y^k, \lambda_k)$, while in Algorithm I this task is performed just once, at the projection step. We remark that for for $\sigma = 0$ both Algorithm I and II reduce the doubly augmented Lagrangian method of (36)-(37). This fact has two consequence: first, the results in our following convergence analysis for the inexact versions also hold for the exact method, and secondly, existence of (exact) solutions of the primal subproblems for the method given by (36)–(37), established in Proposition 7, implies existence of all iterates both for Algorithm I and II.

# 4 Convergence analysis of Algorithms I and II

The convergence properties of Algorithms I and II will be a consequence of their relation to two hybrid inexact versions of the proximal point algorithm, applied to the problem of finding zeroes of the saddle point operator. Such methods were studied by the authors in [14] and we state them here in order to facilitate the analysis.

The methods are intended for solving problem:

$$\text{Find } x \in B \text{ such that } 0 \in T(x), \tag{43}$$

where the point-to-set operator $T : B \to \mathcal{P}(B^*)$ is maximal monotone and $B$ is a real reflexive Banach space. Both methods require an exogenous constant $\sigma \in [0, 1]$ ($\sigma < 1$ for the second one), an exogenous bounded sequence of regularization parameters $\{\lambda_k\} \subset \mathbb{R}_{++}$ and an auxiliary strictly convex and Fréchet differentiable function $f : B \to \mathbb{R}$. They are defined as follows:

**Algorithm PI: Inexact Proximal Point + Bregman Projection Method**

1. Choose $x^0 \in B$.
2. Given $x^k$, find $\tilde{x}^k \in B$ such that

$$\lambda_k[f'(x^k) - f'(\tilde{x}^k)] - e^k \in T(\tilde{x}^k), \tag{44}$$

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

$$\|e^k\|_{B^*} \leq \sigma \lambda_k \begin{cases} D_f(\tilde{x}^k, x^k) & \text{if } ||x^k - \tilde{x}^k|| < 1, \\ \nu_f(x^k, 1) & \text{if } ||x^k - \tilde{x}^k|| \geq 1. \end{cases} \tag{45}$$

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

$$x^{k+1} = \text{argmin}_{x \in H_k} D_f(x, x^k). \tag{46}$$

**Algorithm PII: Inexact Proximal Point-Extragradient Method**

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

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

$$e^k + \lambda_k[f'(x^k) - f'(\tilde{x}^k)] \in T(\tilde{x}^k), \tag{47}$$

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

$$D_f\big(\tilde{x}^k, (f')^{-1}\big[f'(\tilde{x}^k) - \lambda_k^{-1} e^k\big]\big) \leq \sigma D_f(\tilde{x}^k, x^k). \tag{48}$$

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

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

We present next the main results on algorithms PI and PII, proved in Corollary 1 and Theorems 1, 2 and 3 of [14] respectively.

**Theorem 1.** *If $f$ satisfies H4, then, for all $k$ and all $\sigma \in [0, 1]$, the inexact proximal subproblems of both Algorithm PI (i.e.* (44)-(45)*) and Algorithm PII (i.e.* (47)-(48)*) have solutions.*

**Theorem 2.** *Suppose that Algorithm PI (respectively Algorithm PII) stops after $k$ steps. Then $\tilde{x}^k$ generated by Algorithm PI (respectively Algorithm PII) is a solution of* (43).

**Theorem 3.** *Take $f$ satisfying H1–H4, and $\{\lambda_k\} \subset (0, \bar{\lambda}]$. Let $\{x^k\}$ be the sequence generated by Algorithm PI (resp. Algorithm PII). If problem* (43) *has solutions, then*

*i) $\{x^k\}$ has weak accumulation points and all of them are solutions of* (43).

*ii) If $f$ also satisfies H5, then the whole sequence $\{x^k\}$ is weakly convergent to a solution of* (43).

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

**Proposition 8.** *Take $F$, $\{v^k\}$, $\{z^k\}$, $\{\tilde{z}^k\}$, $\{\lambda_k\}$ and $\sigma$ as in Algorithm I. Then*

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

*ii) Let* $e^k = \lambda_k[F'(z^k) - F'(\tilde{z}^k)] - v^k \in B_0^* \times B$. *Then*

$$\|e^k\|_{(B_0\times B^*)^*} \leq \sigma\lambda_k \begin{cases} D_F(\tilde{z}^k, z^k) & \text{if } ||z^k - \tilde{z}^k||_{B_0\times B^*} < 1, \\ \nu_F(z^k, 1) & \text{if } ||z^k - \tilde{z}^k||_{B_0\times B^*} \geq 1. \end{cases}$$

**Proof.** Let $v^k = (u^k, w^k) \in B^* \times B$. Take $Q$ as in (32). Then we have, by definition of $v^k$,

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

where the second equality follows from Proposition 6(ii-b) and the last one from (39), and also

$$w^k = -G(\tilde{x}^k) + \lambda_k P_{-K}(M_r(\tilde{x}^k, y^k, \lambda_k)) \in -G(\tilde{x}^k) + N_{K^*}(Q(\tilde{x}^k, y^k, \lambda_k)),$$

where the inclusion was proved in Lemma 1(c). So $v^k$ is an element of $T_L(\tilde{z}^k)$ and (i) holds.

In order to prove (ii), let $e^k = (\epsilon^k, \eta^k) \in B_0^* \times B$. Then we have that

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

if and only if

$$\begin{aligned} 0 &= u^k + \lambda_k[f'(\tilde{x}^k) - f'(x^k)] + \epsilon^k, \\ 0 &= w_j^k + \lambda_k[h'_{r_j}(Q_j(\tilde{x}^k, y^k, \lambda_k)) - h'_{r_j}(y_j^k)] + \eta_j^k, \quad (1 \leq j \leq m) \end{aligned}$$

if and only if

$$\begin{aligned} 0 &= \bar{L}'_x(\tilde{x}^k, y^k, \lambda_k) + \lambda_k[f'(\tilde{x}^k) - f'(x^k)] + \epsilon^k, \\ 0 &= -G_j(\tilde{x}^k) + \lambda_k P_{-K_j}(M_{r_j}(\tilde{x}^k, y_j^k, \lambda_k)) + \lambda_k[h'_{r_j}(Q_j(\tilde{x}^k, y^k, \lambda_k)) - h'_{r_j}(y_j^k)] + \eta_j^k, \quad (1 \leq j \leq m) \end{aligned}$$

if and only if, using (35) and Lemma 1(a),

$$\begin{aligned} 0 &= \hat{L}'_x(\tilde{x}^k, x^k, y^k, \lambda_k) + \epsilon^k, \\ 0 &= \eta_j^k, \quad (1 \leq j \leq m) \end{aligned}$$

if and only if

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

Thus, $\|e^k\|_{(B_0\times B^*)^*} = \|(\epsilon^k, 0)\|_{B_0^*\times B} = \|\epsilon^k\|_{B_0^*} = \|\hat{L}'_x(\tilde{x}^k, x^k, y^k, \lambda_k)\|_{B_0^*}$, and the result follows from (38). □

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

**Proposition 9.** *Take* $\{z^k\}$, $\{\tilde{x}^k\}$, $\{\lambda_k\}$, $\sigma$ *and* $F$ *as in Algorithm II. Define* $\tilde{z}^k = (\tilde{x}^k, Q(\tilde{x}^k, y^k, \lambda_k))$ *and* $e^k = \Big(\hat{L}'_x(\tilde{x}^k, x^k, y^k, \lambda_k), 0\Big)$ *, where* $Q$ *is defined as in* (32)*. Then*

*i)* $e^k + \lambda_k[F'(z^k) - F'(\tilde{z}^k)] \in T_L(\tilde{z}^k)$,

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

*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 (35) and Lemma 1(a)–(c) we have

$$
\begin{aligned}
& e^k + \lambda_k[F'(z^k) - F'(\tilde{z}^k)] \\
= \;& \Big(\hat{L}'_x(\tilde{x}^k, x^k, y^k, \lambda_k) + \lambda_k[f'(x^k) - f'(\tilde{x}^k)], \times_{j=1}^m \lambda_k[h'_{r_j}(y_j^k) - h'_{r_j}(Q_j(\tilde{x}^k, y^k, \lambda_k))]\Big) \\
= \;& \Big(\bar{L}'_x(\tilde{x}^k, y^k, \lambda_k), \times_{j=1}^m - G_j(\tilde{x}^k) + \lambda_k P_{-K_j}(M_{r_j}(\tilde{x}^k, y_j^k, \lambda_k))\Big) \\
= \;& \Big(L'_x(\tilde{x}^k, Q(\tilde{x}^k, y^k, \lambda_k)), -G(\tilde{x}^k) + \lambda_k P_{-K}(M_r(\tilde{x}^k, y^k, \lambda_k))\Big) \in T_L(\tilde{z}^k),
\end{aligned}
$$

establishing (i). Also,

$$
\begin{gathered}
D_F\big(\tilde{z}^k, (F')^{-1}\big[F'(\tilde{z}^k) - \lambda_k^{-1}e^k\big]\big) = \\
D_F\big(\big(\tilde{x}^k, Q(\tilde{x}^k, y^k, \lambda_k)\big), \big((f')^{-1}\big[f'(\tilde{x}^k) - \lambda_k^{-1}\hat{L}'_x(\tilde{x}^k, x^k, y^k, \lambda_k)\big], Q(\tilde{x}^k, y^k, \lambda_k)\big)\big) = \\
D_f\big(\tilde{x}^k, (f')^{-1}\big[f'(x^k) - \lambda_k^{-1}\bar{L}'_x(\tilde{x}^k, y^k, \lambda_k)\big]\big) + \sum_{j=1}^m D_{h_{r_j}}\big(Q_j(\tilde{x}^k, y^k, \lambda_k), Q_j(\tilde{x}^k, y^k, \lambda_k)\big) = \\
D_f\big(\tilde{x}^k, (f')^{-1}\big[f'(x^k) - \lambda_k^{-1}\bar{L}'_x(\tilde{x}^k, y^k, \lambda_k)\big]\big) \leq \sigma\big[D_f(\tilde{x}^k, x^k) + \sum_{j=1}^m D_{h_{r_j}}\big(Q_j(\tilde{x}^k, y^k, \lambda_k), y_j^k\big)\big] = \\
\sigma D_F(\tilde{z}^k, z^k),
\end{gathered}
$$

using the error criterion given by (41) 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

$$
\begin{aligned}
x^{k+1} &= (f')^{-1}\big[f'(\tilde{x}^k) - \lambda_k^{-1}\hat{L}'_x(\tilde{x}^k, x^k, y^k, \lambda_k)\big] \\
y_j^{k+1} &= (h'_{r_j})^{-1}\big[h'_{r_j}(Q_j(\tilde{x}^k, y^k, \lambda_k))\big], \quad j = \overline{1, m}
\end{aligned}
$$

if and only if

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

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

Propositions 8–9 show that Algorithms I, II inherit the properties of Algorithms PI, PII. We get therefore the following results for Algorithms I and II. The first one deals with the case of finite termination.

**Theorem 4.** *Suppose that Algorithm I (respectively Algorithm II) 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* (32).

**Proof.** Immediate from Proposition 8 (respectively 9), Theorem 2 and Proposition 5(ii). □

The next result establishes that our error criteria are robust, in the sense that any point sufficiently close to the exact solution of the primal subproblem satisfies the error criteria. As a consequence, if the subproblems are solved with any algorithm guaranteed to converge to the (unique) solution of the subproblem, then a finite number of iterations of such inner loop will be enough to generate a point satisfying the error criteria.

**Theorem 5.** *Assume that both $B$ and $B^*$ are strictly convex and reflexive Banach spaces, satisfying property (H). Let $\{z^k\}$ be the sequence generated by Algorithm I for conclusion (a) and by Algorithm II 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* (38)-(39).

*b) If $(f')^{-1}$ is continuous then there exists an open subset $U_k \subset B$ such that any $x \in U_k$ solves* (41).

**Proof.** Let $\bar{x}^k$ denote the exact solution of (36) whose existence is ensured by Proposition 7, and $\bar{z}^k = (\bar{x}^k, Q(\bar{x}^k, y^k, \lambda_k))$, where $Q$ is as in (32). Then $\bar{z}^k \neq z^k$, because otherwise, by Proposition 8, $0 \in T_L(z^k)$, in contradiction with the assumption that $z^k$ is not a KKT-pair. Hence, $D_F(\bar{z}^k, z^k) > 0$, with $F(x,y) = f(x) + \sum_{j=1}^m h_{r_j}(y_j)$, and by total convexity of $F$, resulting from Proposition 3(i) in [14], $\alpha_k := \sigma\lambda_k \min\{D_F(\bar{z}^k, z^k), \nu_F(z^k, 1)\} > 0$. Corollary 1 establishes continuity of $Q(\cdot, y^k, \lambda_k)$ and then 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

$$\psi_k(x) = \|L'_x(x, Q(x, y^k, \lambda_k)) + \lambda_k[f'(x) - f'(x^k)]\|_{B^*} -$$

$$\sigma\lambda_k \min\{D_f(x, x^k) + \sum_{j=1}^m D_{h_{r_j}}(Q_j(x, y^k, \lambda_k), y_j^k), \nu_F(x^k, 1)\}.$$

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 9 and, instead of $\psi_k$, the auxiliary function $\bar{\psi}_k : B \to \mathbb{R}$ defined as

$$\bar{\psi}_k(x) = D_f\big(x, (f')^{-1}\big[f'(x^k) - \lambda_k^{-1} L'_x(x, Q(x, y^k, \lambda_k))\big]\big) - \sigma\big[D_f(x, x^k) + \sum_{j=1}^{m} D_{h_{r_j}}\big(Q_j(x, y^k, \lambda_k), y_j^k\big)\big].$$

□

Next we present the main convergence results for Algorithms I and II.

**Theorem 6.** *Suppose that $B_j$ $(1 \leq j \leq m)$ are uniformly convex and uniformly smooth Banach spaces. Take $f : B_0 \to \mathbb{R}$ satisfying H1-H4 and $\lambda_k \leq \bar{\lambda}$. Let $\{z^k\}$ be the sequence generated by Algorithm I (respectively Algorithm II). 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) $B_j = \mathcal{L}^{p_j}(\Omega)$, $\Omega$ is countable and $r_j = q_j = p_j/(p_j - 1)$ $(1 \leq j \leq m)$, i.e. the regularization function for the dual variables, which belong to $\Pi_{j=1}^m \ell_{q_j}$, is $h_q(y) = \sum_{j=1}^m \frac{1}{q_j} ||y_j||_{q_j}^{q_j}$, or*

*b) $B_j$ is a Hilbert space and $r_j = 2$ $(1 \leq j \leq m)$, i.e. the the dual regularization function is $h_2(y) = \frac{1}{2} \sum_{j=1}^m ||y_j||_{B_j}^2$,*

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

**Proof.** By Proposition 8 (respectively 9) 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 $F : B_0 \times B^* \to \mathbb{R}$ given by $F(x, y) = f(x) + \sum_{j=1}^m ||y_j||_{B_j^*}^{r_j}$. By Propositions 2 and 3 of [14], $F$ satisfies H1–H4 and also satisfies H5 under conditions (a) or (b) of (ii). By our assumption on existence of KKT-pairs, Proposition 5(i) and 5(iii), $T_L$ is a maximal monotone operator with zeroes. The result follows then from Theorem 3 and Proposition 5(ii). □

**Remark**: We only request reflexivity of $B_0$ (and not strict convexity or smoothness) because we require good properties of the primal regularizing function $f$ (e.g. H1–H4). In the case of the image spaces $B_j$ $(1 \leq j \leq m)$ we have fixed the dual regularizing function $h_{r_j}$, and we can ensure that they will satisfy H1-H4 only when the $B_j$'s are uniformly convex and uniformly smooth Banach spaces. Uniform convexity is sufficient for H1-H2, reflexivity for H4 and uniform smoothness is necessary and sufficient for H3 in the case of any power of the norm greater than one. Using specific $h_{r_j}$'s allowed us to get explicit formulae for the dual updatings.

We comment next on our convergence results as compared to previous works. In the case of $m = 1$, $B_1 = \mathcal{L}^p(\Omega)$, $K_1 = \mathcal{L}^p_+(\Omega)$, our results coincide with those in [14]. For the case of hilbertian $B_j$'s ($1 \leq j \leq m$), our convergence results are much stronger than those in [29]. Convergence results for all algorithms in this reference require either that $\lim_{k\to\infty} \lambda_k = 0$ or at least that $\lambda_k$ be small enough, while our results hold for any choice of the sequence $\{\lambda_k\}$, either endogenous or exogenous, as long as it is bounded. Additionally, as we have already observed, since the methods in [29] lack primal regularization, not much can be said about the primal sequence $\{x^k\}$. Finally, methods connected with the augmented Lagrangian approach are analyzed in this reference under exact solution of the subproblems, i.e. without admitting errors.

### 4.1 Some results on the convergence rate of algorithm PI

In this subsection we consider problem (43) for the case in which $B$ is a Hilbert space $H$, and we look at Algorithm PI with $f = \frac{1}{2}||\cdot||^2_H$, which satisfies all properties H1-H5 (see [14]). Then the Bregman projection is exactly the metric projection and the method takes the form:

**Algorithm PI in Hilbert spaces**

1. Choose $x^0 \in H$ and $\sigma \in [0, 1)$.

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

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

   where $e^k$ is any vector in $H$ which satisfies

$$\|e^k\|_{B^*} \leq \frac{1}{2}\sigma\lambda_k \min\{||x^k - \tilde{x}^k||^2, 1\}. \tag{49}$$

3. Let $v^k = \lambda_k(x^k - \tilde{x}^k) - e^k$. If $v^k = 0$ or $\tilde{x}^k = x^k$, then stop. Otherwise,

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

In [14] we observed that, working in Banach spaces, we were forced to adopt the error criterion (49), which is somewhat more restrictive that the one originally proposed by Solodov and Svaiter for Hilbert spaces in [26], namely

$$\|e^k\|_{B^*} \leq \sigma\lambda_k \max\{||x^k - \tilde{x}^k||, \lambda_k^{-1}||v^k||\}, \tag{50}$$

with $\sigma \in (0, 1)$. Observe the square in the right hand side of (49), which is absent in (50), and which makes the right hand side of (49) much smaller than the one in (50) for large $k$. In compensation, we get a strictly better asymptotic constant for the linear convergence rate and

we can establish superlinear convergence when $\lim_{k\to\infty}\lambda_k = 0$, even with inexact solutions of the proximal subproblems, which has not been proved for the error criterion in [26]. We recall that $T^{-1}$ is *Lipschitz continuous at zero with modulus* $\theta > 0$ if $T^{-1}(0) = \{\bar{x}\}$ and there exists a positive constant $\delta$ such that $||y - \bar{x}|| \leq \theta||v||$ for all $y \in T^{-1}(v)$ and $||v|| \leq \delta$. Under this hypotesis it was proved in [26] that any sequence $\{x^k\}$ generated by Algorithm PI with $\|e^k\|_{B^*} \leq \sigma\lambda_k \max\{||x^k - \tilde{x}^k||, \lambda_k^{-1}||v^k||\}$ in place of (49) and $\lambda_k \leq \bar{\lambda}$, converges strongly to $\bar{x}$, and that the rate of convergence is Q-linear with asymptotic error constant bounded by

$$\eta = \left[1 - \left(\frac{1-\sigma}{1+\sigma}\right)^4 \frac{1}{(\theta\bar{\lambda}(1-\sigma)+1)^2}\right]^{1/2}, \tag{51}$$

or, more precisely,

$$||x^{k+1} - \bar{x}|| \leq \left[1 - \left(\frac{1-\sigma}{1+\sigma}\right)^4 \frac{1}{(\theta\lambda_k(1-\sigma)+1)^2}\right]^{1/2} ||x^k - \bar{x}|| \leq \eta||x^k - \bar{x}||, \tag{52}$$

where $\bar{x} = \lim_{k\to\infty} x^k$. With our more restrictive error criterion, we have the following more accurate result.

**Theorem 7.** *Suppose that $T^{-1}$ is Lipschitz continuous at zero with modulus $\theta$, then any sequence $\{x^k\}$ generated by Algorithm PI, with $\lambda_k \leq \bar{\lambda}$, converges strongly to $\bar{x}$, the solution of (43), the convergence is Q-linear with asymptotic error constant bounded by*

$$\eta_k = \left[1 - \left(\frac{1-\sigma/2}{1+\sigma/2}\right)^4 \frac{1}{(\theta\lambda_k(1-\sigma/2)+1)^2}\right]^{1/2}. \tag{53}$$

*Moreover, if $\lim_{k\to\infty}\lambda_k = 0$ then the convergence is superlinear.*

**Proof.** From the proof of Theorem 2 in [14] we know that $\lim_{k\to\infty} v^k = 0$ and

$$\lim_{k\to\infty} \|x^k - \tilde{x}^k\| = 0. \tag{54}$$

Then, in view of (49), there exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$ it holds that

$$\|e^k\|_{B^*} \leq \frac{1}{2}\sigma\lambda_k\|x^k - \tilde{x}^k\|^2. \tag{55}$$

In view of (54) and (55), we have

$$\|e^k\|_{B^*} \leq \frac{1}{2}\sigma\lambda_k\|x^k - \tilde{x}^k\| \tag{56}$$

for all $k \geq k_0$. By (56), (50) holds and therefore we can use the results in [26], namely the inequality in (52), first with $\tilde{\sigma} = \sigma/2$, and then with $\tilde{\sigma} = (\sigma/2)\|x^k - \tilde{x}^k\|$ getting, for all $k \geq k_0$, $\|x^{k+1} - \bar{x}\| \leq \eta_k \|x^k - \bar{x}\|$, which proves the first part of the theorem, and also

$$\|x^{k+1} - \bar{x}\| \leq \left[1 - \left(\frac{1 - (\sigma/2)\|x^k - \tilde{x}^k\|}{1 + (\sigma/2)\|x^k - \tilde{x}^k\|}\right)^4 \frac{1}{(L\lambda_k(1 - (\sigma/2)\|x^k - \tilde{x}^k\|) + 1)^2}\right]^{1/2} \|x^k - \bar{x}\|, \quad (57)$$

for all $k$. Since

$$\lim_{k\to\infty} \frac{1 - (\sigma/2)\|x^k - \tilde{x}^k\|}{1 + (\sigma/2)\|x^k - \tilde{x}^k\|} = 1,$$

and $\lim_{k\to\infty} \lambda_k = 0$ implies

$$\lim_{k\to\infty} \frac{1}{(L\lambda_k(1 - (\sigma/2)\|x^k - \tilde{x}^k\|) + 1)^2} = 1,$$

we get

$$\lim_{k\to\infty} \left[1 - \left(\frac{1 - (\sigma/2)\|x^k - \tilde{x}^k\|}{1 + (\sigma/2)\|x^k - \tilde{x}^k\|}\right)^4 \frac{1}{(L\lambda_k(1 - (\sigma/2)\|x^k - \tilde{x}^k\|) + 1)^2}\right]^{1/2} = 0.$$

In view of (57), the proof is complete. □

We point out that our $\eta_k$ of (53) is strictly smaller than $\eta$ in (51). In fact, if we define $\bar{\eta}$ with $\bar{\lambda}$ substituting for $\lambda_k$ in (53), then we still have $\bar{\eta} < \eta$. An easy computation shows that this statement is true if and only if the polynomial $p(\sigma) = \sigma^2 - 3\sigma + 2(1 + 2/(L\bar{\lambda}))$ is strictly positive in the interval $[0, 1)$, which is always true, because $p(\sigma) \geq 4/(L\bar{\lambda}) > 0$ for all $\sigma \in [0, 1]$. We emphasize that the results in Theorem 7 are strongly dependent on the square in the right hand side of (49).

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

[3] 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.

[4] Burachik, R.S., Scheimberg, S. A proximal point algorithm for the variational inequality problem in Banach spaces. Technical Report ES-470/98, COPPE-UFRJ, Rio de Janeiro (1998) (to be published in *SIAM Journal on Control and Optimization*).

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

[6] Burke, J. V., Ferris, M. C., Qian, M. On the Clarke subdifferential of the distance function of a closed set. *Journal of Mathematical Analysis and Applications* **166** (1992) 119-213.

[7] 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.

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

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

[10] Butnariu, D., Iusem, A.N. and Resmerita, E. Total convexity of the powers of the norm in uniformly convex Banach spaces (to be published in *Journal of Convex Analysis*).

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

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

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

[14] Iusem, A.N., Gárciga O., R. Inexact versions of proximal point and augmented Lagrangian algorithms in Banach spaces (to be published).

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

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

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

[18] Luenberger, D.G. *Optimization by vector space methods.* John Wiley & Sons, Inc. New York (1969).

[19] Penot, J. P., Ratsimahalo, R. Characterizations of metric projections in Banach spaces and applications. *Abstract and Applied Analysis* **3**, 1-2, (1998) 85-103.

[20] Phelps, R. R. Metric projections and the Gradient Projection method in Banach spaces. *SIAM Journal on Control and Optimization* **23** (1985) 973-977.

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

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

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

[24] Rockafellar, R.T. Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control and Optimization* **12** (1974) 268-285.

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

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

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

[28] Troyanski, S.L., On locally uniformly convex and differentiable norms in certain non-separable Banach spaces. *Studia Mathematica* **37** (1971) 173-180.

[29] Wierzbicki, A.P., Kurcyusz, S. Projection on a cone, penalty functionals and duality theory for problems with inequality constraints in Hilbert space. *SIAM Journal on Control and Optimization* **15** (1977) 25-56.