Instituto Nacional de Matemática Pura e Aplicada

Inexact versions of proximal point and cone-constrained augmented Lagrangians in Banach spaces

Rolando Gárciga Otero

Tese apresentada para a obtenção do grau de Doutor em Matemática

Rio de Janeiro

2001

To my family.

ACKNOWLEDGMENTS

I am grateful to the Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Instituto Nacional de Matemática Pura e Aplicada for supporting this research and providing excellent working conditions. I am particularly grateful to the members of the Optimization group at IMPA for their enlightening and suggestions, but specially to my adviser Professor Alfredo Noel Iusem, to whom I am eternally indebted: thank you for believing in me, thank you for teaching me.

I would like to express also my gratefulness to my mom, sisters and friends for their understanding and unconditional support through these long years of separations.

Abstract

We extend to Banach spaces the hybrid proximal-extragradient and proximal-projection methods for finding zeroes of maximal monotone operators, recently proposed by Solodov and Svaiter in finite dimension and Hilbert spaces respectively. The generalization of the hybrid-projection method makes it possible the use of regularizations other than the quadratic by using an appropriate error criterion, which allows for bounded relative error, and a Bregman projection instead of the metric projection. Boundedness of the sequence generated by both methods 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. A variant of the error for the hybrid-projection method let us establish superlinear convergence even with inexact solutions of the proximal subproblem in Hilbert spaces. We show that the hybrid steps of the Proximal Point methods, allowing for constant relative errors, are necessary in order to ensure boundedness of the generated sequence, even in the optimization case. Moreover, we show that such conditions do not imply that the sequence of errors results summable "a posteriori". We then transpose such methods to generate augmented Lagrangian methods for the following cone-constrained convex optimization problem in Banach spaces: $\min g(x)$ 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. Two alternative procedures are developed 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. Finally we add to the hybrid-extragradient method a penalization effect for solving variational inequality problems in Banach spaces, by introducing a boundary coercive condition on the regularizing function. We give examples of regularizing functions for the cases of the feasible set being closed balls and polyhedra. We get convergence results similar to those of the methods without penalization under the assumptions of pseudo- and paramonotonicity of the involved operator and existence of solutions.

Keywords: proximal point method, augmented Lagrangian, inexact solutions, maximal monotone operator, cone-constrained optimization.

Contents

Chapter 1

Introduction

Many problems reduce to finding zeroes of maximal monotone operators, originated e.g. in optimization or in variational inequalities. Illustrative examples are elliptic boundary value problems (see e.g. [49]), which have the Sobolev spaces, $W^{m,p}(\Omega)$, as its natural domain of definition. Thus, methods for finding zeros of maximal monotone operators in nonhilbertian spaces seem to be quite important, or at least as important as they are in hilbertian or finite dimensional spaces. Proximal point methods, discussed in detail in section 2.5 of Chapter 2, are among the main tools for finding zeroes of maximal monotone operators in Hilbert spaces (see [61]). We devote this thesis to the study of theoretical properties of variants of the proximal point method for finding zeroes of maximal monotone operators in Banach spaces. Such methods have been extensively studied in Hilbert spaces, but in nonhilbertian spaces this is not the case. In fact, up to our knowledge and to present days, there exists a small number of references on this issue [1, 9, 15, 40], and only one of then, [40], considers inexact variants of the algorithm. In [40] only partial convergence results are given and the error criteria require convergence to zero of a sequence $\{\delta_k\}$, where δ_k is some absolute measure of the error committed at iteration k. These criteria are undesirable, because they impose increasing accuracy of the procedure along the iterative process. Our first main objective consists of studying variants of the proximal point method in nonhilbertian spaces, emphasizing inexact variants allowing for constant relative errors, in the sense, e.g. of [64] and [62]. In order to attain this goal, we extend to Banach spaces two hybrid methods, namely those which combine a step of an inexact proximal point method with either an extragradient or a projection step, recently proposed by Solodov and Svaiter in finite dimension and Hilbert spaces respectively ([64, 62]).

For this purpose, we need an appropriate family of regularizing functions, which substitute for the square of the norm in Hilbert spaces. Such a family has been proposed for proximal methods which work in Banach spaces but less general than those studied in this thesis, because they do not include the option of inexact iterates. Among them, we mention exact proximal methods for optimization (e.g. [1], [15]), and also for variational inequalities (see [9]). We keep the technical assumptions on such family of regularization functions required in [9]. We present a whole family of explicit functions which satisfy such technical assumptions, for instance $f(x) = \|\cdot\|^r$, with $r > 1$, in any uniformly convex and uniformly smooth Banach space.

We comment that the extension to Banach spaces of the inexact methods with constant relative errors involves several nontrivial technicalities, related to the inexistence of inner products in nonhilbertian Banach spaces. In our modest opinion, we were able to overcome the obstacles presented by such technical difficulties.

Augmented Lagrangian, or more generally multiplier methods (see e.g. [6]) have been recognized as an efficient option for dealing with inequality constrained optimization problems, particularly when both the objective function and the feasible set are convex. There is a strong connection between augmented Lagrangian methods and proximal point methods [60, 3, 25, 38]. This connection led us to our second main objective: to translate these inexact proximal methods in Banach spaces into the setting of augmented Lagrangian methods. As a result, we get two inexact augmented Lagrangian methods, allowing for constant relative errors, which are applicable to smooth cone-constrained convex optimization in Banach spaces. We remark that previously existing results on augmented Lagrangian methods for cone constrained convex optimization appear only in [66], which deals only with the case of Hilbert spaces and offers only partial convergence results, e.g. not much is said about the primal sequence, besides the fact that if it exists it is a minimizing one, strong regularity conditions are needed on the problem data and the regularizing parameters are not freely chosen. In view of this fact, it happens that our results are original even for finite dimensional problems. Moreover, the results in [66] depend strongly on properties of the Hilbert space, while we consider a general reflexive Banach space for the operator constraints. As particular instances we get applications involving e.g. an infinite number of constrains, or the cone of the positive semidefinite matrices in an $n-$dimensional Euclidean space.

In finite dimensional or Hilbert spaces, proximal point methods with a nonquadratic regularizing function f have been proposed mainly with penalization purposes. It is assumed that the problem includes a closed and convex set C, i.e. if we are looking for the zeroes of $T + N_C$, where T is a maximal monotone operator and N_C is the normalizing cone of C, or equivalently we are dealing with the variational inequality problem with operator T and feasible set C. In such a situation, the use of an f whose domain is C and whose gradient diverges at the boundary of C makes the proximal subproblems unconstrained ([19, 21, 35, 41, 12, 25, 36, 62]). Extension of this approach to nonhilbertian spaces is a highly nontrivial question, and we make some progress in this direction, offering alternative sufficient conditions on the regularizing functions. We give explicit examples of functions satisfying some of these conditions for the cases in which the feasible region is either a ball or a polyhedron. In the case of the polyhedron, our explicit functions satisfy all desired properties. In the case of balls, our functions miss one of the sufficient conditions, but then we give an appropriate way to choose the regularizing parameters, which allows us to establish the desired convergence properties in the absence of the missing condition.

To get started we need some preliminaries, including a detailed analysis of previous works on the subject, which is achieved in Chapter 2.

Unless otherwise stated, all results in Chapters 3 through 6 are, to our knowledge, new.

In Chapter 3, we deal with the required properties of the regularizing functions, and establish that positive powers of the norm satisfy them, thus completing the results of [17, 13]. We introduce the notion of generalized duality mapping, which seems to be new. Such mappings recover some basic properties of duality mappings and we use them, also, for obtaining regularizing functions with penalty purposes for closed balls. A family of functions for the case in which the feasible set is a polyhedron is also given. All desired properties (labeled as H1-H6) are discussed, and we give counterexamples for the properties which do not hold, as well as alternative properties which substitute for the missing ones.

In Chapter 4 we extend to Banach spaces the methods proposed for Hilbert or finite dimensional spaces in [64, 62]: 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 criterion 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. A variant of the error for the hybrid-projection method, when compared with that in [64] in Hilbert spaces, gives better asymptotic constant for the linear convergence rate and let us establish superlinear convergence (under the standard assumptions of Lipschitz continuity of the inverse operator at zero and regularizing parameters converging to zero), even with inexact solutions of the proximal subproblems. We also establish in Section 4.4 of this chapter the necessity of the hybrid steps in order to guarantee boundedness of the iterates, even for optimization problems in Hilbert spaces. Moreover, we prove that the relative error measures of these hybrid methods do not result summable "a posteriori", showing its intrinsic advantage over criteria which require summability of the errors.

In Chapter 5, we transpose such methods to the setting of cone-constrained convex optimization, in order to generate augmented Lagrangian methods for 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 obtain 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.

Finally, in Chapter 6, we add to the hybrid Proximal-Extragradient method presented on Charter 4 a penalization effect in Banach spaces. This effect requires a boundary coerciveness condition (H6), in order to guarantee good definition and penalization in the method. A complete convergence analysis is performed. The analysis is separated in two cases: when there exist solutions in the interior of the feasible set and when all solutions lie in the boundary. In the first case, there is no need of additional assumptions on the maximal monotone operator, other than existence of solutions. In the second case we keep the standard assumptions of pseudo- and paramonotonicity of the operator, and condition H3.a on the regularizing function, as was done in [9] for the exact method. We prove that the generated sequence is bounded and that all its weak accumulation points are solutions. Condition H3.a is satisfied for the given family in the case of a polyhedron, but not for the case of a ball, a fact discussed in Sections 3.4 and 3.3 respectively. Thus, the case of a ball needs special attention, which is given in Section 6.5. The idea is to enforce condition H3.a by choosing appropriately the regularizing parameters at each step of the algorithm. As a consequence, the case of a ball is completely studied, getting strong convergence of the whole sequence to a solution, when there is no solution in the interior of the ball. We mention that for the optimization case we achieve the same results with a streamlined analysis, where condition H3.a is not needed. Moreover, in this case the sequence of inexact proximal iterates is a minimizing one.

Chapter 2

Preliminaries

2.1 Maximal Monotone Operators

From now on, B is a real 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 $x \in B$).

Definition 2.1.1. A point-to-set operator $T : B \to \mathcal{P}(B^*)$ is said to be *monotone* if

$$\langle w - w', x - x'\rangle \geq 0, \text{ for all } x,\ x' \in B \text{ and all } w \in T(x),\ w' \in T(x').$$

If the relation above holds with strict inequality when $x \neq x'$ then T is called *strictly* monotone.

Definition 2.1.2. 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:

$$\text{Find } x \in B \text{ such that } 0 \in T(x). \tag{2.1}$$

Definition 2.1.3. A *zero* of the operator $T : B \to \mathcal{P}(B^*)$ is any solution $x \in B$ of problem (2.1) above.

We recall that the strong topology on B is the one induced by the norm and the weak topology on B is that induced by the linear and bounded operators from B to $\mathbb{R}$ (i.e. the elements of B^*), in the sense that $\{x^n\} \subset B$ is weakly convergent to $\bar{x} \in B$ if and only if $\lim_{n\to\infty}\langle v, x^n - \bar{x}\rangle = 0$ for all $v \in B^*$. Thus, given $\{x^k\} \subset B$, we use the notation $x^k \xrightarrow[k\to\infty]{\text{s}} x$ (respectively $x^k \xrightharpoonup[k\to\infty]{\text{w}} x$) meaning that the sequence $\{x^k\}$ converges in the strong (respectively weak) topology of B to the element $x \in B$. The same for B^*, recalling that now the weak topology is induced by the elements of B^{**}, hence reflexivity of the real Banach space B ensures that this is also the so called *weak** topology of B^*.

Definition 2.1.4. An operator $T : B \to \mathcal{P}(B^*)$ is said to be *pseudomonotone* if its domain,

$$\text{dom } T = \{x \in B \mid T(x) \neq \emptyset\},$$

is closed and convex and the following condition is satisfied:

$$\{x^k\} \subset \text{dom } T,\ w^k \in T(x^k),\ x^k \xrightharpoonup[k\to\infty]{\text{w}} x^o \text{ and } \limsup_{k\to\infty}\langle w^k, x^k - x^o\rangle \leq 0$$
$$\Longrightarrow \forall y \in \text{dom } T,\ \exists w^o \in T(x^o) \text{ such that } \langle w^o, x^o - y\rangle \leq \liminf_{k\to\infty}\langle w^k, x^k - y\rangle\ .$$

Definition 2.1.5. Given a closed and convex subset C of B, a monotone operator $T : B \to \mathcal{P}(B^*)$ is *paramonotone* in C if it satisfies:

$$\langle w - w', z - z'\rangle = 0, \text{ with } z,\ z' \in C,\ w \in T(z),\ w' \in T(z')$$
$$\Longrightarrow w \in T(z') \text{ and } w' \in T(z).$$

Proposition 2.1.6. *If T is a maximal monotone mapping then, for any $x \in$ dom T, the image $T(x)$ is a closed and convex subset of B^*. Moreover, the graph of T, $G(T)$, is demi-closed, i.e. if $x^k \xrightarrow[k\to\infty]{\text{s}} x$ in B, $v^k \xrightharpoonup[k\to\infty]{\text{w}} v$ in B^* (or $x^k \xrightharpoonup[k\to\infty]{\text{w}} x$ in B, $v^k \xrightarrow[k\to\infty]{\text{s}} v$ in B^*) and $(x^k, v^k) \in G(T)$ for all k, then $(x, v) \in G(T)$.*

Proof. See, e.g. [50], page 105.

The following proposition contains some basic results concerning sums of point-to-set operators.

Proposition 2.1.7. *Let T_1 and T_2 be two point-to-set operators. Then, the sum $T_1 + T_2$ is:*

i) paramonotone in C when both T_1 and T_2 are paramonotone in C,

ii) pseudomonotone, when both operators are pseudomonotone and also the images $T_1(x)$ and $T_2(x)$ are convex and closed for all $x \in$ dom $T_1 \cap$ dom T_2,

iii) maximal monotone, when both operators are maximal monotone and int(dom T_1)$\cap$ dom $T_2 \neq \emptyset$.

Proof. For (i) see [37], for (ii) see [50], page 97, and for (iii) see [57].

2.2 The derivatives of Fréchet and Gâteaux

Let B_1, B_2 be two real Banach spaces. Take $x \in B_1$, an open neighborhood $U(x)$ of x and $f : U(x) \subset B_1 \to B_2$.

Definition 2.2.1. (1) The *directional derivative* of f at x in the direction $y \in B_1$ is the limit

$$\lim_{t \to 0^+} \frac{f(x+ty) - f(x)}{t} = f'_+(x, y),$$

when it exists.

(2) If there exists an operator in $L(B_1, B_2)$ (i.e. a continuous linear operator), to be denoted by $f'(x)$, such that

$$\lim_{t \to 0} \frac{f(x+ty) - f(x)}{t} = \langle f'(x), y \rangle$$

for all $y \in B_1$, then f is *Gâteaux differentiable* (or G-differentiable for short) at x.

(3) Moreover, if the limit above is uniform on the unit sphere, i.e.

$$\lim_{t \to 0} \sup_{\|y\|_1 = 1} \left\| \frac{f(x+ty) - f(x)}{t} - \langle f'(x), y \rangle \right\|_2 = 0,$$

then f is called *F-differentiable* at x (F for *Fréchet*).

Definition 2.2.1 has been taken from [22].

Proposition 2.2.2. *(Relations between F- and G-derivatives)*

i) Every F-derivative at x is also a G-derivative at x.

ii) If f' exists as a G-derivative in some neighborhood of x, and is continuous at x, then $f'(x)$ is also an F-derivative at x.

iii) If f' exists as an F-derivative at x, then f also is continuous at x.

Proof. See Proposition 4.8 of [67].

Proposition 2.2.3. *(Sum Rule) Suppose that $f,\ g : U(x) \subset B_1 \to B_2$ are G-differentiable (respectively F-differentiable) at x, where B_1 and B_2 are Banach spaces over $\mathbb{R}$. Then $\alpha f + \beta g$ is G-differentiable (respectively F-differentiable) for all $\alpha,\ \beta \in \mathbb{R}$ and*

$$(\alpha f + \beta g)'(x) = \alpha f'(x) + \beta g'(x).$$

Proof. See Proposition 4.9 in [67].

Proposition 2.2.4. *(Chain Rule) Let B_1, B_2 and B_3 be Banach spaces. Consider $x \in B_1$, $y = f(x) \in B_2$ and neighborhoods $U(x)$ of x and $V(y)$ of y. Suppose that we are given maps $f : U(x) \subset B_1 \to B_2$ and $g : V(y) \subset B_2 \to B_3$ with $f(U(x)) \subset V(y)$. This defines the composition map $g \circ f : U(x) \subset B_1 \to B_3$. Suppose that $g'(y)$ exist as an F-derivative. Then*

i) If $f'(x)$ exists as an F-derivative then the composite map $H = g \circ f$ is F-differentiable at x.

ii) If $f'(x)$ exists as a G-derivative, then H is G-differentiable at x and

$$H'(x) = g'(f(x)) \circ f'(x).$$

Proof. See Proposition 4.10 in [67].

We define next some classical notions of smooth, locally uniformly smooth and uniformly smooth Banach spaces, which depend on differentiability properties of the norm.

Definition 2.2.5. Let f denote the norm in the real Banach space B, i.e. $f = \|\cdot\|_B : B \to \mathbb{R}_+$.

(1) B is called *smooth* if f is G-differentiable on $B \backslash \{0\}$.

(2) If f is F-differentiable on $B \backslash \{0\}$, then B is called *locally uniformly smooth.*

(3) If the norm is uniformly F-differentiable on the unit sphere, i.e. f is F-differentiable and

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

then B is called *uniformly smooth.*

Definition 2.2.6. A real Banach space B is

(1) *strictly convex* if for all $x,\ y\ \in B,\ x \neq y,\ \|x\| = \|y\| = 1$, it holds that $\|\lambda x + (1-\lambda)y\| < 1$ for all $\lambda \in (0,1)$,

(2) *uniformly convex* if for each $\epsilon > 0$, there exists $\delta > 0$, so that $\|x\| = \|y\| = 1$ and $\|x - y\| \geq \epsilon$ imply $\|x + y\| \leq 2(1-\delta)$,

(3) *locally uniformly convex* when for any x in the unit sphere it holds that for all $\epsilon > 0$, there exists $\delta = \delta(x) > 0$, so that $\|y\| = 1$ and $\|x - y\| \geq \epsilon$ imply $\|x + y\| \leq 2(1 - \delta)$.

Remark 2.2.7. It is clear that uniform convexity implies both local and strict convexity. Moreover, the following geometric interpretation of uniform convexity can be given: if the sequence of the middle points of the line segments $[x^n, y^n]$ with $\|x^n\| = \|y^n\| = 1$, converges to a point in the boundary of the unit ball, then the length of the segments converges to zero. In particular, strict convexity means that the boundary of the unit ball contains no line segments. Well known examples of uniformly convex spaces are Hilbert spaces, $\mathcal{L}^p(\Omega)$ and Sobolev spaces $W^{m,p}(\Omega)$ $(1 < p < \infty)$.

There is a strong relation between the different definitions of smoothness and convexity. In fact,

Proposition 2.2.8. *Let B be a real reflexive Banach space. Then*

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

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

Proof. See Corollary 1.4 and Theorems 2.13 and 2.14 in pages 43, 52 and 53 respectively of [22].

Definition 2.2.9. We say that the Banach space B has *property* (h) if

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

Remark 2.2.10. Every locally uniformly convex Banach space has property (h). See e.g. [22], page 49.

Remark 2.2.11. 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 [65]).

2.3 Convex functions

A function $f : B \to \mathbb{R} \cup \{+\infty\}$ is *convex* if and only if

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y),$$

for any two points $x,\ y \in B$ and for any real $\alpha \in (0,1)$. If the relation above holds with strict inequality when $x \neq y$ then f is called *strictly convex*. When dom $f \neq \emptyset$, f is called *proper*. Moreover, f is *lower semicontinuous* if for each real number r, the set $\{x \in B |\ f(x) \leq r\}$ is closed. Recall also that a *subgradient* of f at the point $x \in$ dom f is an element $v \in B^*$ such that

$$\langle v, y - x \rangle \leq f(y) - f(x), \tag{2.2}$$

for all $y \in B$. The set of all subgradients of f at x is called the *subdifferential* of f at x and denoted $\partial f(x)$.

Proposition 2.3.1. *The subdifferential operator, $\partial f(\cdot) : B \to \mathcal{P}(B^*)$ of any proper lower semicontinuous convex function is maximal monotone.*

Proof. See e.g. Theorem 3.25 of [53].

Proposition 2.3.2. *Let f be a lower semicontinuous convex function with int(dom f) $\neq \emptyset$ and D_o the subset of dom (∂f), where ∂f is single valued. Then ∂f is* demicontinuous *on D_o (i.e. continuous as a single valued mapping from D_o in the strong topology to B^* in the weak* topology) and $D_o \subset$ int[dom (∂f)].*

Proof. By Proposition 2.3.1, the subdifferential operator, $\partial f(\cdot) : B \to \mathcal{P}(B^*)$ is maximal monotone with a nonempty interior of its domain. Then the results follow from Corollary 1.1 of [55].

Example 2.3.3 (The indicator function). Let $C \subset B$ be a nonempty, closed and convex set and $I_C : B \to \mathbb{R} \cup \{+\infty\}$ the *indicator function* of C, i.e.

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

I_C is a proper lower semicontinuous convex function with subdifferential given by

$$\partial I_C(x) = \begin{cases} \{z \in B |\ \langle z, y - x \rangle \leq 0,\ \forall y \in C\} & \text{if } x \in C \\ \emptyset & \text{otherwise.} \end{cases} \tag{2.3}$$

Definition 2.3.4. The operator $N_C : B \to \mathcal{P}(B^*)$ defined by equation (2.3) above, i.e. $N_C = \partial I_C$, is called the *normalizing operator* of the set C.

The normalizing operator of a nonempty, closed and convex set C is, then, a maximal monotone operator. It follows that the operator $\partial f + N_C$, where f is any proper, lower semicontinuous and convex function, is maximal monotone too, whenever int $(C)\cap$ dom $f \neq \emptyset$. More generally, the operator $T + N_C$ obtained by replacing ∂f by any maximal monotone operator T satisfying int $(C)\cap$ dom $T \neq \emptyset$, is also maximal monotone.

Definition 2.3.5. Given a proper lower semicontinuous convex function and a closed convex set C in B the *constrained convex optimization problem* is defined as $\min_{x\in C} f(x)$.

The zeroes of $T + N_C$ are related to the variational inequality problem, which we define next.

Definition 2.3.6. The *variational inequality problem* for T and C, denoted VIP(T,C), where $T : B \to \mathcal{P}(B^*)$ is a maximal monotone operator and C is any nonempty closed and convex set in B, is defined as:

Find $x \in C$ such that there exists $v \in T(x)$ with $\langle v, y - x\rangle \geq 0$ for all $y \in C$.

Unless otherwise stated, S will denote the set of solutions of VIP(T,C).

The following result is rather immediate.

Proposition 2.3.7. *The solutions of VIP(T,C) are precisely the zeroes of* $T + N_C$.

Example 2.3.8 (Distance function). Given a nonempty, closed and convex subset C of a reflexive real Banach space B the *distance* to C, $d(\cdot, C) : B \to \mathbb{R}$ is given by

$$d(x, C) = \inf_{y\in C} \|x - y\|_B \,. \tag{2.4}$$

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

$$\partial d(\cdot, C) = \partial \|\cdot - z\|_B \cap N_C(z), \tag{2.5}$$

a result that can be found in Theorem 1 of [14]. Here P_C denotes the *metric projection* onto C defined at any $x \in B$ as

$$P_C(x) = \arg\min_{y\in C} \frac{1}{2} \|x - y\|_B^2 \,, \tag{2.6}$$

i.e. $P_C(x) = \{z \in C \ : \|x - z\|_B = d(x, C)\}$.

The point-to-set operator $P_C : B \to \mathcal{P}(C)$ is single-valued when B is strictly convex. Existence of points attaining the distance in (2.4), and so a nonempty valued metric projection P_C, is ensured by reflexivity of B.

2.3.1 Differentiable convex functions

In this section, B is a Banach space and $f : B \to \mathbb{R} \cup \{\infty\}$ is lower semicontinuous, with int(dom f) $\neq \emptyset$. We assume that f is G-differentiable on int(dom f). We define the Bregman distance and the modulus of total convexity associated to f as follows.

Definition 2.3.9 (Bregman distance). The function D_f :dom $f\times$ int(dom f) $\to \mathbb{R}_+$, given by

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

is called the *Bregman distance* associated to f.

Proposition 2.3.10. *Let f be a lower semicontinuous convex function with int(dom f) $\neq \emptyset$.*

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

a) D_f is well defined and $D_f(\cdot, x)$ is convex.

b) (Three-point equality) For any $x,\ z \in$ int(dom f) and $y \in$ dom f, it holds that

$$D_f(y, x) - D_f(y, z) - D_f(z, x) = \langle f'(x) - f'(z), z - y\rangle. \tag{2.7}$$

c) (Four-point equality) For any $x, z \in$ int(dom f) and $y, w \in$ dom f, it holds that

$$D_f(w, z) = D_f(w, x) + \langle f'(x) - f'(z), w - y\rangle + D_f(y, z) - D_f(y, x). \tag{2.8}$$

d) If f is strictly convex, then, for any $x,\ y \in$ int(dom f) such that $x \neq y$ it holds $D_f(y, x) > 0$.

ii) If f is F-differentiable, then f' :int(dom f) $\to B^*$ is norm-to-norm continuous and D_f is continuous on int(dom f)$\times$ int(dom f)

Proof. The results in (i) are proved in [16]: item (i) appears in 1.1.10, (a) in 1.1.3, (b) in 1.3.9 and (d) in 1.1.9 of [16]. The relation in (2.8) is a consequence of (b) and its denomination *Four-point equality* is taken from [62]. The result in (ii) follows from the corollary in page 20 of [53].

Example 2.3.11 (Powers of the norm). Let $f_r(x) = \frac{1}{r}\|x\|_B^r$ with $r > 1$. f_r is a continuous convex function which, in view of Proposition 2.2.4 and Definition 2.2.5, is G-differentiable when B is smooth, and F-differentiable when B is locally uniformly smooth. See a complete proof in [17].

Definition 2.3.12 (Modulus of total convexity). ν_f :int(dom f) $\times \mathbb{R}_+ \to \mathbb{R}$, defined as

$$\nu_f(x,t) = \inf_{y \in S(x,t)} D_f(y,x),$$

with $S(x,t) = \{y \in B : \|y - x\| = t\}$, is called the *modulus of total convexity* of f.

Remark 2.3.13. The modulus of total convexity can be defined in a bigger domain and without the additional assumption of differentiability of f on the interior of its domain (see [16]).

We collect next a few results on the modulus of total convexity, needed in the sequel.

Proposition 2.3.14. *If $x \in$ int(dom f) then*

i) The domain of $\nu_f(x,\cdot)$ is an interval $[0,\tau(x))$ or $[0,\tau(x)]$ with $0 < \tau(x) \leq \infty$. Moreover $\tau(x)$ is finite if and only if dom f is bounded.

ii) For all $s \geq 1$ and all $t \geq 0$,

$$\nu_f(x, st) \geq s\nu_f(x,t). \tag{2.9}$$

iii) (Superadditivity) $\nu_f(x, t + t') \geq \nu_f(x,t) + \nu_f(x,t')$ for all $t,\ t' \in [0,\infty)$.

iv) $\nu_f(x,\cdot)$ is nondecreasing, and it is strictly increasing if and only if $\nu_f(x,t) > 0$ for all $t > 0$.

Proof. see [16], 1.2.2.

2.3.2 Totally convex functions

In this section, B is a Banach space and $f : B \to \mathbb{R} \cup \{\infty\}$ is lower semicontinuous and G-differentiable in the interior of its domain. We introduce the notion of *totally convex* function, closely related to the modulus of total convexity. We follow the presentation in [16].

Definition 2.3.15 (Total convexity). A function $f : B \to \mathbb{R} \cup \{+\infty\}$ is said to be *totally convex* if $\nu_f(x,t) > 0$ for all $t > 0$ and all $x \in$ int(dom f).

Remark 2.3.16. Total convexity of f ensures strict convexity of f in the interior of its domain. The converse result holds when f is continuous on a closed domain and B is finite dimensional. In particular both notions are equivalent when f is finite (i.e. dom $f = B$) and B is a finite dimensional space (see [16], 1.2.6).

Proposition 2.3.17. *If f is totally convex and lower semicontinuous with open domain D, and φ is a real, convex, differentiable and strictly increasing function defined on an open interval which contains $f(D)$, then the function $g = \varphi \circ f : B \to \mathbb{R} \cup \{+\infty\}$ (with the convention that $g(x) = +\infty$ if $x \notin D$) is totally convex and $\nu_g(x,t) \geq \varphi'(f(x))\nu_f(x,t)$, for all $x \in D$ and for all $t \geq 0$.*

Proof. See [16], 1.2.7.

In finite dimensional spaces we have a wide variety of totally convex functions, e.g. any strictly convex function, as discussed in Remark 2.3.16. We establish next total convexity of certain functions in infinite dimensional spaces.

Proposition 2.3.18. *Let B be a locally uniformly convex Banach space and consider f_r as in Example 2.3.11, with $r \geq 2$. Then f_r is totally convex and*

$$\nu_{f_r}(x,t) = \begin{cases} \|x\|^{r-2} \nu_{f_2}(x,t) & \text{if } x \neq 0, \\ \frac{1}{r}t^r & \text{if } x = 0. \end{cases} \tag{2.10}$$

Proof. This result follows from [16], 1.4.2, after noting that $\nu_{f_r} = \frac{1}{r}\nu_f$, with $f = rf_r = \|\cdot\|^r$, which is the case considered in [16].

2.3.3 Bregman projection

Definition 2.3.19 (Bregman projection). Let B be a Banach space and $f : B \to \mathbb{R} \cup \{\infty\}$ a strictly and totally convex function. Given a closed and convex subset $C \subset \text{dom } f$, the *Bregman projection* of a point $x \in \text{int}(\text{dom} f)$ onto C is the (necessarily unique) point $\Pi^f_C(x)$ satisfying

$$\Pi^f_C(x) = \arg\min_{y \in C} D_f(y,x).$$

Note that the uniqueness of the Bregman projection $\Pi^f_C(x)$ follows from strict convexity of the function f, which ensures that the function $D_f(\cdot,x) = f(\cdot) - f(x) - \langle f'(x), \cdot - x\rangle$ is strictly convex on the convex set dom f, which contains C. In view of Remark 2.3.16 this is, then, a direct consequence of the total convexity assumption on f when $C \subset \text{int}(\text{dom } f)$[1], e.g. dom $f = B$. In general it seems that strict convexity of f on the boundary of C is indeed needed for uniqueness of the Bregman projection. Another case where total convexity

[1]It also holds when C is a subset of the algebraic interior of the domain of f and f is totally convex in this set. See [16].

suffices for uniqueness occurs when the norm of the derivative of f goes to infinity in the boundary of C (this case will be considered in Chapter 6). Since these are the two cases to be considered in this thesis, we can assume in the sequel that uniqueness of the Bregman projection is ensured assuming just total convexity of f in the interior of its domain. On the other hand, total convexity is indeed sufficient for the existence of $\Pi_C^f(x)$ as the following results shows.

Proposition 2.3.20. *Suppose that the strictly convex function $f : B \to \mathbb{R} \cup \{+\infty\}$ is totally convex and that the nonempty set C is convex and closed. Then, the following statements hold:*

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

ii) If $C \subset$ int(dom f) for each $x \in$ int(dom f), we have $\bar{x} = \Pi_C^f(x)$ if and only if $f'(x) - f'(\bar{x}) \in N_C(\bar{x})$ or, equivalently, $\bar{x} \in C$ and

$$\langle f'(x) - f'(\bar{x}), z - \bar{x} \rangle \leq 0 \tag{2.11}$$

for all $z \in C$.

Proof. See [16], 2.1.5.

When the closed and convex set C is a *hyperplane* (i.e. $C = \{y \in B \mid \langle v, y - x\rangle = 0\}$ for some $v \in B^* \backslash \{0\}$ and $x \in B$), the normal cone at any $\bar{x} \in C$ is easily computable, in fact, it is the one dimensional subspace spanned by v. Hence, if C is the negative half-space of such hyperplane (i.e. $C = \{y \in B \mid \langle v, y - x\rangle \leq 0\}$), we get $N_C(\bar{x}) = \{sv \mid s \in \mathbb{R}_+\}$. Thus, if f is finite, F-differentiable and totally convex, then the Bregman projection $\bar{x} = \Pi_C^f(x)$ of any $x \in B$ is a zero of the maximal monotone operator $f'(\cdot) - f'(x) + \{sv \mid s \in \mathbb{R}_+\}$. When f' is easily computable, finding such a zero is reasonably easy, and thus Bregman projections onto hyperplanes become essentially computable.

2.3.4 Uniformly totally convex functions

Let B be a Banach space and $f : B \to \mathbb{R} \cup \{+\infty\}$ be a totally convex function on a subset E of int(dom f) (i.e. the restriction of f to E is totally convex as defined in Definition 2.3.15). It is interesting to study whether or not the modulus of total convexity of f is bounded away from zero on certain sets.

Definition 2.3.21. A function $f : B \to \mathbb{R} \cup \{+\infty\}$ is *uniformly totally convex* on $E \subset$ int(dom f), if for each bounded subset $\tilde{E} \subset E$ and for each real number $t > 0$, we have

$$\inf_{x \in \tilde{E}} \nu_f(x, t) > 0.$$

This property has been called by *sequential consistency* in [16] and *norm compatibility* in [17]. They were introduced with different definitions in those references, but such definitions are equivalent, as we state next.

Proposition 2.3.22. *Let $f : B \to \mathbb{R} \cup \{+\infty\}$ be a convex function and $E \subset int(dom\ f)$. Then the following affirmations are equivalent:*

a) *f is uniformly totally convex on E,*

b) Sequential consistency on E: *For all $\{x^k\}$, $\{y^k\} \subset E$ such that $\{x^k\}$ is bounded and* $\lim_{k\to\infty} D(y^k, x^k) = 0$, *it holds that $x^k - y^k \xrightarrow[k\to\infty]{s} 0$.*

Proof. See [16], 2.1.2.

Remark 2.3.23. Uniform total convexity also implies a sequential consistency similar to the one in Proposition 2.3.22(b), but with boundedness of the sequence $\{x^k\}$ replaced by boundedness of $\{y^k\}$ in the assumption. In fact, when $\{y^k\}$ is bounded the argument in the proof of Theorem 2.4 in [62] establishes that in such a case $\{x^k\}$ is bounded too.

Example 2.3.24 (Powers of the norm). Let B be an uniformly convex smooth Banach space and consider f_r defined as in Example 2.3.11. Then f_r is uniformly totally convex (see [17]). Moreover, when $B = \mathcal{L}^p(\Omega)$ $(1 < p < \infty)$, the following closed formulae, depending on the strictly convex real function $\varphi_s(t) = \frac{1}{s}t^s : \mathbb{R}_+ \to \mathbb{R}_+$ with $s > 1$, give lower bounds of ν_{f_r} (see [34]):

1. If $1 < p < r < 2$ then $\nu_{f_r}(0, t) = \frac{1}{r}t^r$ and for any $x \neq 0$
$$\nu_{f_r}(x, t) \geq \|x\|_p^{r-p} D_{\varphi_p}(t + \|x\|_p, \|x\|_p) \geq \|x\|_p^{r-p} \nu_{\varphi_p}(\|x\|_p, t). \tag{2.12}$$

2. If $1 < r \leq p \leq 2$ then
$$\nu_{f_r}(x, t) = D_{\varphi_r}(t + \|x\|_p, \|x\|_p) \geq \nu_{\varphi_r}(\|x\|_p, t). \tag{2.13}$$

3. If $1 < r \leq p$ and $p \geq 2$ then $\nu_{f_r}(0, t) = \frac{1}{r}t^r$ and for any $x \neq 0$
$$\nu_{f_r}(x, t) \geq \left(1 + \frac{t}{\|x\|_p}\right)^{r-p} D_{\varphi_r}\left([2^{1-p}t^p + \|x\|_p^p]^{\frac{1}{p}}, \|x\|_p\right). \tag{2.14}$$

4. If $p = r \geq 2$ then
$$\nu_{f_p}(x, t) \geq 2^{1-p}\varphi_p(t). \tag{2.15}$$

In particular, $\nu_{1/2\|\cdot\|^2}(x, t) \geq t^2/4$.

5. If $r \geq 2$ then $\nu_{f_r}(0,t) = \frac{1}{r}t^r$ and for any $x \neq 0$

$$\nu_{f_r}(x,t) \geq \|x\|_p^{r-2} \nu_{1/2\|\cdot\|^2}(x,t) \geq \frac{t^2}{4} \|x\|_p^{r-2}. \tag{2.16}$$

2.4 Duality mappings

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

$$J_\varphi(x) = \{x^* \in B^* | \langle x^*, x\rangle = \|x^*\|_* \|x\|, \|x^*\|_* = \varphi(\|x\|)\} \tag{2.17}$$

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:

J0: $J = I$, i.e. the identity map, if and only if B is a Hilbert space.

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

J2: The inverse function of the weight φ, 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 smooth (i.e., if B^* is strictly convex), then J_φ is single-valued and norm-to-weak* continuous. Moreover, if B^* satisfies property (h) then J_φ is norm-to-norm continuous.

J5: If B is strictly convex, then J is strictly monotone.

J6: If B is uniformly smooth, then J is uniformly continuous on bounded subsets of B.

Proofs of properties J0–J5 can be found in [22], and a proof of property J6 appears in [68].

Remark 2.4.1. There is a strong relation between the theory of duality mappings and the geometry of Banach spaces. In fact the converse statements of J5, J6 and of the first statement in J4 also hold (see [22] and [50] respectively).

Example 2.4.2 (Squared norm). Let B be an uniformly smooth Banach space and consider $f_2 = \frac{1}{2}\|\cdot\|_B^2$. Using properties J1, J2 and J6 for the weight function $\varphi(t) = t$, we get $f_2'(x) = J(x)$ for all $x \in B$. Hence, f_2' is uniformly continuous on bounded subsets of its domain B.

Metric projections can be characterized with the help of duality mappings. The next proposition contains one such characterization.

Proposition 2.4.3. *Let B be a a real reflexive Banach space. Take $x \in B$ and a nonempty, closed and convex $C \subset B$. Then, the following statements are equivalent:*

i) $z \in P_C(x)$,

ii) $J(x - z) \cap N_C(z) \neq \emptyset$, *where N_C is the normalizing operator of C, as in Definition 2.3.4.*

Proof. See the equivalence between (a) and (c) in Theorem 3.1 of [51]. □

Example 2.4.4. Let $H_{v,\alpha} = \{y \in B \mid \langle v, y\rangle = \alpha\}$ be the hyperplane given by $v \in B^*\backslash\{0\}$ and $\alpha \in \mathbb{R}$. Choose $r > 1$ and the weight function $\varphi(t) = rt^{r-1}$. If B is smooth then the function $f = \|\cdot\|^r$ is G-differentiable and $J_\varphi(x) = \{f'(x)\}$. Moreover, if B is uniformly convex then f is totally convex on B and the Bregman projection, $\Pi^f_{H_{v,\alpha}}$ is well defined. More interestingly, this is an example of reasonably effective computability of the Bregman projection. In fact,

$$\Pi^f_{H_{v,\alpha}} = J^*_{\varphi^{-1}}(sv + J_\varphi), \tag{2.18}$$

where $s \in \mathbb{R}$ is a solution of the equation in one real unknown

$$\langle v, J^*_{\varphi^{-1}}(sv + J_\varphi)\rangle = \alpha. \tag{2.19}$$

Thus the Bregman projection of a point onto a hyperplane with respect to a power of the norm bigger than 1 is effectively computable, up to the solution of one (generally nonlinear) equation in one real unknown (see [17]).

In Chapter 3 we will show that we can use more general weight functions and still get (generalized) duality mappings with similar properties, also strongly related to the geometry of the space, which represent the derivatives of certain convex functions incorporating information of particular proper subsets of the Banach space B, in such a way that they can be used for penalization purposes.

2.5 The exact Proximal Point Method

The Proximal Point method, whose origins can be traced back e.g. to [43], [48], [46], was presented in [61] as a method for finding zeroes of monotone operators in Hilbert spaces. The *exact* Proximal Point Method is defined as follows: 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 $z^0 \in H$, and then, given z^k, define z^{k+1} as the only $z \in H$ such that

$$0 \in T(z) + \lambda_k(z - z^k). \tag{2.20}$$

It was proved in [61] that when T has zeroes the sequence $\{x^k\}$ defined by (2.20) is weakly convergent to a zero of T. This convergence is not strong in general, a fact that was proved by Güler in [31], through an example in the space ℓ^2 for the optimization case (i.e. for $T = \partial f$ for some convex f).

Proposition 2.5.1. *(Güler) There exists a proper, closed and convex function in ℓ^2 which has minimizers and such that, given any bounded positive sequence $\{\lambda_k\}$, there exists a point $x \in dom\ f$ for which the exact Proximal Point algorithm* (2.20)*, starting at x, converges weakly but not strongly to a minimizer of f.*

Proof. See [31], Corollary 5.1.

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 (2.20) is achieved by taking x^{k+1} as the unique $x \in B$ such that

$$\lambda_k[f'(x^k) - f'(x)] \in T(x), \tag{2.21}$$

where $f : B \to \mathbb{R} \cup \{\infty\}$ is a strictly convex and Gâteaux differentiable function on the interior of its domain satisfying conditions H1–H3 below, and f' is its Gâteaux derivative. When $f = 1/2 \|x\|_B^2$ and B is a Hilbert space, (2.21) reduces to (2.20).

The assumptions on f are, essentially, the following:

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

H2: Uniform total convexity of f on int(dom f).

H3: The G-derivative of f, f', is uniformly continuous on bounded subsets of int(dom f).

Under these assumptions, it was proved in [9] that if dom $T \subset \text{int}(\text{dom } f)$ and 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. Actually, it is also requested in [9] that f' be onto, in order to ensure existence of a solution x of (2.21).

Proposition 2.5.2. *Let $f : B \to \mathbb{R} \cup \{\infty\}$ be a strictly convex and Gâteaux differentiable function on the interior of its domain satisfying dom $T \cap$ int(dom $f) \neq \emptyset$ and $T : B \to \mathcal{P}(B^*)$ a maximal monotone operator. If f' is surjective, then $T + f'$ is surjective. In particular, 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 [9], Lemma 2.10 and Corollary 3.1. □

This surjectivity requirement can be avoided[2], but we are going to use it for other purposes too, e.g. as a sufficient condition for good definition of extragradient-like steps. We add thus this condition under the label of H4.

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

As we will discuss in Section 2.6, we are also interested in applying the proximal point method to constrained problems with a function f whose derivative diverges on the boundary of the feasible set. In such a case, H3 will fail on any set intersecting such boundary. We consider thus some weaker variants of H3. One of them, introduced in [9], is the following:

H3.a: $\{x^k\}_k$, $\{y^k\}_k \subset \text{int}(\text{dom } f)$, $x^k \xrightarrow[k\to\infty]{w} x^\infty$, $y^k \xrightarrow[k\to\infty]{w} x^\infty$ and $\lim_k D_f(x^k, y^k) = 0$ $\Longrightarrow \lim_k \left[D_f(x^\infty, x^k) - D_f(x^\infty, y^k)\right] = 0$.

H3.a is weaker than H3 in the sense that when H2 holds, it can be proved, using the three point equality (2.7) and Proposition 2.3.22, that H3 implies H3.a. We observe that in [9] condition H3.a was used for guaranteeing convergence of the exact proximal point method when the operator T is pseudo- and paramonotone.

We will need an even weaker variant, namely

H3.b: For all bounded sequence $\{x^k\} \subset \text{int}(\text{ dom } f)$ such that $\lim_k d(x^k, \partial \text{ dom } f) = 0$ there exists a bounded sequence $\{\rho_k\} \subset \mathbb{R}_{++}$ such that, if $\hat{x}$ is a weak limit of $\{x^k\}$, $\{y^k\}_k \subset \text{int}(\text{ dom } f)$ is weakly convergent to $\hat{x}$ and $\lim_k D_f(y^k, x^k) = 0$ then

$$\text{limsup}_k \rho_k \left[D_f(\hat{x}, x^k) - D_f(\hat{x}, y^k)\right] \leq 0.$$

[2]It is redundant when T has zeroes. This fact was proved for the first time, as far as we know, by B.F. Svaiter.

Note that H3.a implies H3.b, taking $\rho_k = 1$ for all k. We will also use a variant which considers only certain subsets of dom f. We remind that for M, $M' \subset B$, $d(M, M') = \inf\{\|x - x'\| \ : x \in M, x' \in M'\}$.

H3.c: f' is uniformly continuous on any bounded set $M \subset$ dom f such that $d(M, \partial$ dom $f) > 0$.

Some additional condition on f must be imposed in order to establish uniqueness of weak cluster point of sequences generated by proximal point and related methods. One such condition is the weak-to-weak* continuity of f'. Another one, apparently weaker than weak-to-weak* continuity, has been used in [16] and we introduce it next, with the label of H5.

Definition 2.5.3 (Separability requirement). We say that the G-differentiable function $f : B \to \mathbb{R} \cup \{\infty\}$ satisfies the *separability requirement* if given sequences $\{x^k\}$, $\{y^k\} \subset$ int(dom f) converging weakly to x and y respectively, with $x \neq y$, it holds that

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

H5 is then defined as:

H5: f satisfies the separability requirement.

We mention that we found no advantage of H5 over the weak-to-weak* continuity of f'.

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 have discussed above (we will complete the proof that such f satisfies H1-H4 in Chapter 3).

On the other hand, H5 is much more infrequent. It is known that in a Hilbert space $f_2(x) = 1/2\,\|x\|^2$ is such that f_2' is the identity function (see Example 2.4.2 together with property J0 in section 1.4), which is, obviously, weak-to-weak* continuous. For $B = \ell_p$, $1 < p < +\infty$, it can be proved that for $f = \|\cdot\|_p^p$, f' is sequentially weak-to-weak* continuous (see, for example, Proposition 8.2 in [8]). Unfortunately, we have found counterexamples showing that in $B = \ell_p$ or $B = \mathcal{L}^p$ ($1 < p < +\infty$) the function $f = \|\cdot\|_p^r$ ($r > 1$) does not satisfy the *separability requirement*, excepting in the two cases just mentioned. Such examples are presented in Chapter 3. We emphasize that H5 is needed only to prove uniqueness of the weak accumulation points of the sequences generated by the methods under consideration in this thesis, and only as a sufficient condition.

We mention next some previous references on the Proximal Point method in Banach spaces. A scheme similar to (2.21) was considered in [23], [24], 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 $f(x) = 1/2 \|x\|_B^2$, the scheme given by (2.21) was analyzed in [40], 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 (2.21) was analyzed in [1] for the case of $f(x) = 1/2 \|x\|_B^2$ and in [15] 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 $\mathcal{L}^p$).

2.6 Generalized Proximal Point with penalization purposes

In finite dimensional or Hilbert spaces, where the square of the norm leads always to simpler computations, the scheme given by (2.21) 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 [19], [21], [35], [41] for the optimization case and in [12], [25], [36] for finding zeroes of monotone operators.

Definition 2.6.1 (Bregman function). Given $C \subset \mathbb{R}^p$ closed, convex and with nonempty interior, the function $c : C \to \mathbb{R}$ is said to be a *Bregman function* with *zone* C if the following conditions hold:

B1. c is strictly convex and continuous on C.

B2. c is continuously differentiable on $\mathrm{int}(C)$.

B3. For all $\gamma \in \mathbb{R}$ and all $z \in C$ the partial level sets $\Gamma(z, \gamma) = \{w \in \mathrm{int}(C) : D_c(z, w) \le \gamma\}$ are bounded.

B4. If $\{z^k\} \subset \mathrm{int}(C)$ and $\lim_{k\to\infty} z^k = \widetilde{z}$ then $\lim_{k\to\infty} D_c(\widetilde{z}, z^k) = 0$.

B5. If $\{w^k\} \subset C$ and $\{z^k\} \subset \text{int}(C)$ are sequences such that $\{w^k\}$ is bounded, $\lim_{k\to\infty} z^k = \widetilde{z}$ and $\lim_{k\to\infty} D_c(w^k, z^k) = 0$ then $\lim_{k\to\infty} w^k = \widetilde{z}$.

Bregman functions were introduced in [7]. The original definition also requires the left partial level sets

$$\Gamma(\gamma, z) = \{w \in C : D_c(w, z) \leq \gamma\}$$

to be bounded for any $z \in \text{int}(C)$. This condition is not needed to prove convergence of proximal methods (e.g [26]) and it is also a consequence of B1–B3 as observed in [62][3]. Condition B5 was imposed in all studies of Bregman function and related algorithms previous to [62] (e.g. the references cited above in this section). It was proved in [62] that B5 holds automatically as a consequence of B1–B2.

Definition 2.6.2 (Boundary coercive Bregman function). A Bregman function c is said to be boundary coercive if it additionally satisfies the following condition B6.

B6. *Boundary coercive*: if $\{z^k\} \subset \text{int}(C)$ is such that $\lim_{k\to\infty} z^k = \widetilde{z}$ and $\widetilde{z}$ belongs to the boundary of C, then $\lim_{k\to\infty} D_c(w, z^k) = \infty$ for all $w \in \text{int}(C)$.

B6 was introduced in [35]. It is equivalent, by Definition 2.3.9, to the following condition:

$$\lim_{k\to\infty} \langle c'(z^k), w - z^k\rangle = -\infty$$

and also to

$$\lim_{k\to\infty} \left\|c'(z^k)\right\| = +\infty.$$

Conditions B4–B6 hold automatically, as a consequence of B1–B3, when $C = \mathbb{R}^p$. We present next some relevant examples of Bregman functions in finite dimensional spaces.

Example 2.6.3. $C = \mathbb{R}^p$, $c(z) = \|z\|^2$, in which case $D_c(z, z') = \|z - z'\|^2$. More generally $c(z) = z^t M z$ with $M \in \mathbb{R}^{p\times p}$ symmetric and positive definite, in which case $D_c(z, z') = (z - z')^t M(z - z')$.

Example 2.6.4. $C = \mathbb{R}^p_+$, $c(z) = \sum_{\ell=1}^p z_\ell \log z_\ell$, continuously extended to the boundary of $\mathbb{R}^p_+$ with the convention that $0 \log 0 = 0$. In this case

$$D_c(z, z') = \sum_{\ell=1}^{p} [z_\ell \log(z_\ell/z'_\ell) + z'_\ell - z_\ell],$$

which is called the Kullback-Leibler distance, widely used in statistics (see [44]).

[3] The same holds in Banach spaces when c is totally convex. See [16], 2.1.1.

Example 2.6.5. $C = \mathbb{R}^p_+$, $c(z) = \sum_{\ell=1}^p (z_\ell^\alpha - z_\ell^\beta)$ with $\alpha \geq 1$, $\beta \in (0,1)$. For $\alpha = 2$, $\beta = 1/2$ we get

$$D_c(z, z') = \|z - z'\|^2 + (1/2) \sum_{\ell=1}^p \frac{\left(\sqrt{z_\ell} - \sqrt{z'_\ell}\right)^2}{\sqrt{z'_\ell}};$$

for $\alpha = 1$, $\beta = 1/2$ we have

$$D_c(z, z') = (1/2) \sum_{\ell=1}^p \frac{\left(\sqrt{z_\ell} - \sqrt{z'_\ell}\right)^2}{\sqrt{z'_\ell}}.$$

Let us consider now a variational inequality problem, VIP(T,C), as in Definition 2.3.6. Consider a coercive Bregman function with zone C and assume that int(C)$\cap$ dom T is nonempty. The *generalized proximal point method with Bregman functions* (GPPB from now on) for solving VIP(T,C) starts from any $z^0 \in$ int(C) and generates a sequence $\{z^k\}$ through the iterative formula

$$0 \in T_k(z^{k+1}),$$

where $T_k : \mathbb{R}^p \to \mathcal{P}(\mathbb{R}^p)$ is defined as

$$T_k(z) = T(z) + \lambda_k[\nabla c(z) - \nabla c(z^k)],$$

with λ_k as in the previous section (i.e. as required in the Proximal Point method, PP for short). In this case, there is no need to modify T by adding the normalizing operator N_C, because divergence of ∇c at the boundary of C guarantees that the whole sequence $\{z^k\}$ is contained in the interior of C. This is more evident if we consider the case of $T = \partial\varphi$ with a convex $\varphi : \mathbb{R}^m \to \mathbb{R}$. If the problem of interest is $\min \varphi(x)$ subject to $x \in E$, and h is a Bregman function with zone $E \subset \mathbb{R}^m$, then the iteration of GPPB becomes

$$y^{k+1} = \arg\min_{y \in \mathbb{R}^m} \left\{\varphi(y) + \lambda_k D_h(y, y^k)\right\}, \tag{2.22}$$

while the subproblems of PP for the same problem, after adding the normalizing operator N_C to the operator $\partial\varphi$, become

$$y^{k+1} = \arg\min_{y \in E} \left\{\varphi(y) + (\lambda_k/2) \left\|y - y^k\right\|^2 \right\}. \tag{2.23}$$

Note that the subproblems given by (2.22) are unconstrained, while the subproblems given by (2.23) are subject to the constraints $y \in E$. In the case of GPPB, these constraints are taken care of by D_h which, besides its regularization role, as in PP, has also a penalization effect.

GPPB can be traced back to [29] and [30], which considered methods related to GPPB with the Bregman function of Example 2.6.4 applied to linear programming. The next step was [28], which considered GPPB with the same Bregman function applied to general convex optimization problems. Relevant works on GPPB in its current formulation include [25], which considers GPPB for finding zeroes of monotone operators (or VIP(T,C) with solutions in the interior of C) and [19], [21], [35], [41], which study GPPB for the convex optimization problem under progressively weaker assumptions on the problem data or the Bregman function. GPPB for variational inequality problems has been analyzed in [12]. In this reference the notion of coercive Bregman function is appropriately extended to Hilbert spaces. Basically strong convergence is replaced by weak convergence in conditions B5 and B6. B4 takes the form of H3.a as defined in section 2.5 and a new condition is added, called *zone coerciveness*. Zone coerciveness is equivalent to the surjectivity of the G-derivative (see property H4 in the previous section 2.5). Convergence results in [12] can be summarized as follows. If VIP(T,C) has solutions and T is pseudo- and paramonotone, then the sequence generated by the GPPB is bounded and any weak accumulation point is a solution. Moreover, uniqueness follows from a separability requirement condition as H5 in section 2.5. In finite dimension, where zone coerciveness is stronger than boundary coerciveness, the first one is not needed and uniqueness of the accumulation point follows from B4. We give next an example of Bregman function presented in [11] for Hilbert spaces.

Example 2.6.6. Let H be a Hilbert space. Take $v^1,\ v^2, ...,\ v^p \in H\backslash\{0\}$ and $\alpha_1,\ \alpha_2, ...,\ \alpha_p \in \mathbb{R}$. Define $C = \{x \in H \mid \langle v^i, x\rangle \geq \alpha_i,\ i = 1, ..., p\}$ and assume that $\operatorname{int}(C)$ is nonempty. Consider the function $g : C \to \mathbb{R}$ defined by

$$g(x) = \frac{1}{2}\|x\|^2 + \sum_{i=1}^{p} \left(\langle v^i, x\rangle - \alpha_i\right) \log\left(\langle v^i, x\rangle - \alpha_i\right) \tag{2.24}$$

for $x \in C$ and infinity otherwise. The following properties hold:

B1. g is strictly convex and continuous on C.

B2. g is F-differentiable on $\operatorname{int}(C)$, with G-derivative given by

$$g'(x) = x + \sum_{i=1}^{p}[1 + log\left(\langle v^i, x\rangle - \alpha_i\right)]v^i. \tag{2.25}$$

B3. g satisfies H1.

B4. g satisfies H3.a.

B5. If $\{w^k\} \subset C$ and $\{z^k\} \subset \operatorname{int}(C)$ are sequences such that $\{w^k\}$ is bounded, $z^k \xrightarrow[k\to\infty]{\text{w}} \widetilde{z}$ and $\lim_{k\to\infty} D_g(w^k, z^k) = 0$ then $w^k \xrightarrow[k\to\infty]{\text{w}} \widetilde{z}$.

B6. (*boundary coerciveness*) If $\{z^k\} \subset \operatorname{int}(C)$ is such that $z^k \xrightarrow[k\to\infty]{\text{w}} \widetilde{z}$ and $\widetilde{z}$ belongs to the boundary of C, then $\lim_{k\to\infty}\langle g'(z^k), w - z^k\rangle = -\infty$, for all $w \in \operatorname{int}(C)$.

B7. (*zone coerciveness*) g satisfies H4.

B8. (*separability requirement*) g satisfies H5.

H1–H5 are as in the previous section. For a complete proof see 5.6 in [11].

Concerning the finite dimensional case, we point out that it has been proved in [62] not only that $B5$ is redundant, but also that the pseudomonotonicity assumption is redundant too. Moreover an inexact version of the GPPB, which we analyze in the next section (i.e. the proximal-extragradient method), is considered in this reference.

Other variants of the Proximal Point method in finite dimensional spaces, also with nonquadratic regularization terms but with iteration formulae different from (2.21), can be found in [4], [3], [10], [27], [38] and [39].

Other definitions of Bregman function, now in Banach spaces, can be found in [16] and [9], but without attempting penalization in the related methods, thus without caring about boundary coerciveness. In this work we are not interested in offering additional definitions of Bregman functions. The required properties for the regularizing function involved in the methods that we present are basically H1–H5 of section 2.7. A deeper analysis of the existence of such functions in Banach spaces, including also some form of boundary coerciveness, appears in Chapter 3. Related GPPB methods are presented in Chapter 6.

2.7 The inexact Proximal Point Method

Let H be a Hilbert space and $T : H \to \mathcal{P}(H)$ a maximal monotone operator. Given $\lambda_k > 0$, we define the *resolvent* operator P_k of T as

$$P_k = (T + \lambda_k I)^{-1}.$$

P_k is single valued, nonexpansive and its domain is the whole Hilbert space H (see Minty's Theorem in [47]). The exact Proximal Point method for finding a zero of T defines a sequence $\{z^k\} \subset H$, starting from an arbitrary $z^0 \in H$, through the iteration

$$z^{k+1} = P_k(z^k).$$

Inexact variants of the Proximal Point method in Hilbert spaces were considered as early as in [61], where the iteration

$$z^{k+1} \approx P_k(z^k) \tag{2.26}$$

is considered. Here $\|z^{k+1} - P_k(z^k)\| = \epsilon_k$ is the error in the k-th iteration, and convergence of the sequence $\{z^k\}$ to a zero of T is ensured under the assumption that $\sum_{k=0}^{\infty} \epsilon_k < \infty$. Such summability condition demands that the precision in the computations increase with the iteration index k. Other related conditions for the case of quadratic f, but including always summability conditions on the error, can be found in [61] and [40] for Hilbert and Banach spaces respectively, and, for the case of nonquadratic f, in [25], [41] for optimization and [13] for variational inequalities. The extension presented by Kassay in [40] was based on the following result, more general than Minty's Theorem: If B is a reflexive Banach space, $T : B \to \mathcal{P}(B^*)$ is maximal monotone, and $c > 0$ is an arbitrary constant, then $J + cT$ is surjective. Moreover, the operator $P_c = (J + cT)^{-1}$ is a single-valued maximal monotone operator (not necessarily nonexpansive) (see 2.11 in [50]).

The iteration considered in [40] is

$$z^{k+1} \approx \left[\left(\frac{1}{\alpha_k} J + c_k T \right)^{-1} \circ \frac{1}{\alpha_k} J \right] (z^k), \tag{2.27}$$

where $\{\alpha_k\}$, $\{c_k\}$ are sequences of positive numbers, with $\lim_{k\to\infty} \alpha_k = +\infty$ and $\{c_k\}$ bounded away from zero (i.e. $c_k \geq \underline{c} > 0$, for all k). We mention that the composition $(\frac{1}{\alpha_k} J(z^k) + c_k T)^{-1}$ with $\frac{1}{\alpha_k} J(z^k)$ in (2.27) is necessary because the first factor is defined on B^*. Translating this scheme into the format of (2.21), it reduces to finding an approximate zero of (2.21) with $f = 1/2 \|\cdot\|^2$ and $\lambda_k = 1/(c_k \alpha_k)$. The error criterion implicit in the $\approx$ symbol of (2.27) is less demanding than that in [61], but it is a measure of error of absolute type in each iteration that needs, a priori, to converge to zero; for instance, $\lim_k \epsilon_k = 0$. Boundedness of the sequence $\{x^k\}$ is not guaranteed under this condition.

Also, divergence of $\{\alpha_k\}$, and the relation between α_k and λ_k just discussed, mean that the algorithm in [40] demands that the exogenous regularization parameters λ_k be chosen so that $\lim_{k\to\infty} \lambda_k = 0$. This is a serious drawback, because the proximal point method basically replaces the inversion of T by a sequence of subproblems in each one of which a regularized operator must be inverted (e.g. $\lambda_k I + T$ in Hilbert spaces). Thus, the method is really useful when inversion of T is hard in some sense, i.e. when T is somehow ill-conditioned. When T is maximal monotone, $\lambda_k I + T$ is theoretically well behaved for any $\lambda_k > 0$, but numerically $\lambda_k I + T$ is almost as ill-conditioned as T for very small λ_k, because in such a case the term $\lambda_k I$ becomes negligeable as compared to T (consider e.g. the case of a linear and singular T). One of the main advantages of the Proximal Point method over other regularization schemes (e.g. Thikonov's regularization) is that it works well without requesting that the regularization parameters converge to zero (e.g. taking them constant), so that each subproblem requires inversion of an operator which will be as well behaved as desired if the regularization parameter λ_k is large enough. In this thesis, we also improve over Kassay's scheme in connection with this issue.

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 [64, 63] for the case of quadratic f in Hilbert spaces and in [62], [10] for the case of nonquadratic f in finite dimensional spaces (in [10], with a regularization different from the scheme given by (2.21) and in [63] as an extension of [62] to Hilbert spaces, but without attempting penalization).

We discuss next the error schemes in [64], [62], which we will extend to Banach spaces in Chapter 4. The basic feature of these methods is to combine, in each iteration, an approximate solution of the proximal problem with a projection [64] or an extragradient step [62, 63]. For this reason, these are hybrid methods (proximal-projection or proximal-extragradient). In both of these hybrid methods, the proximal problem (2.20) is decomposed in the following equivalent system:

$$v \in T(z), \tag{2.28}$$

$$v + \lambda_k(z - z^k) = 0. \tag{2.29}$$

An approximate solution of this system is used to compute the next iterate by means of a projection or an extragradient step. The error tolerance is given by a *fixed* $\sigma \in [0, 1)$. In the Hybrid Projection-Proximal method, [64], the approximate solution of (2.29) 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

$$e^k + \lambda_k(x^k - \tilde{x}^k) \in T(\tilde{x}^k), \tag{2.30}$$

where the error term e^k satisfies

$$\|e^k\| \leq \sigma\lambda_k \max\{\|x^k - \tilde{x}^k + \lambda_k^{-1}e^k\|, \|x^k - \tilde{x}^k\|\}. \tag{2.31}$$

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.

$$x^{k+1} = x^k - \frac{\langle v^k, x^k - \tilde{x}^k\rangle}{\|v^k\|^2} v^k. \tag{2.32}$$

The Hybrid Extragradient-Proximal method in [62] requires a regularizing Bregman function f with surjective derivative f'. The auxiliary point $\tilde{x}^k$ must satisfy

$$e^k + \lambda_k[f'(x^k) - f'(\tilde{x}^k)] \in T(\tilde{x}^k), \tag{2.33}$$

and the error criterion is given by

$$D_f\big(\tilde{x}^k, (f')^{-1}\big[f'(\tilde{x}^k) - \lambda_k^{-1}e^k\big]\big) \leq \sigma D_f(\tilde{x}^k, x^k), \tag{2.34}$$

where $(f')^{-1}$ denotes the inverse function of f'. Then the next iterate x^{k+1} is given by

$$x^{k+1} = (f')^{-1}\big[f'(\tilde{x}^k) - \lambda_k^{-1}e^k\big]. \tag{2.35}$$

It has been proved in [64] and [62] that the sequences $\{x^k\}$ defined by (2.30)–(2.32) and (2.33)–(2.35) 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 (2.20) or (2.26), both (2.31) and (2.34) hold with strict inequality for any $\tilde{x}^k \neq x^k$. 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.

2.8 Smooth cone-constrained convex optimization

Let B_j $(0 \leq j \leq m)$ denote $m+1$ real reflexive Banach spaces, B_j^* their 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. $K_j^* = \{y \in B_j^* \; : \langle y, z\rangle_j \geq 0$ 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 define next K-convexity of a mapping between Banach spaces.

Definition 2.8.1 (K-convexity). Take real Banach spaces B_0 and B_1 and let $K \subset B_1$ be a a nonempty, closed and convex cone. A mapping $M : B_0 \to B_1$ 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 $\alpha \in [0, 1]$. Equivalently,

$$M(\alpha x + (1-\alpha)x') \precsim_1 \alpha M(x) + (1-\alpha)M(x'),$$

where $\precsim_1$ is the partial order relation induced in B_1 by K.

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

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

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

$$L(x, y) = g(x) + \langle y, G(x) \rangle = g(x) + \sum_{j=1}^{m} \langle y_j, G_j(x) \rangle_j, \tag{2.36}$$

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

$$G(x) = \big(G_1(x), ..., G_m(x)\big) \tag{2.37}$$

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.

Definition 2.8.2 (Feasible pair). 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. $y \succsim_* 0$.

Definition 2.8.3 (Optimal pair). 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 summarize next some basic properties of the Lagrangian defined by (2.36) in the context already defined, i.e. under assumptions A1-A2. 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 2.8.4. *Let L and G be defined as in* (2.36) *and* (2.37) *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 (i.e. Propositions 2.2.3 and 2.2.4). 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) and Proposition 2.3.10(ii). Item (ii) is immediate, since $L(x, \cdot) : B^* \to \mathbb{R}$ is affine. □

Definition 2.8.5 (KKT-pair). A pair $(x, y) \in B_0 \times B^*$ is a *KKT-pair* if it is feasible and additionally

$$0 = L'_x(x,y) = g'(x) + y \circ G'(x) \qquad \text{(Lagrangian condition)}, \tag{2.38}$$

$$\langle y, G(x)\rangle = 0 \qquad \text{(complementarity)}. \tag{2.39}$$

We present next some Lagrangian duality results.

Proposition 2.8.6. *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.*

$$L(x, \tilde{y}) \le L(x,y) \le L(\tilde{x}, y) \tag{2.40}$$

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 [45]. □

Next we define the *saddle-point operator*

Definition 2.8.7 (Saddle-point operator). $T_L : B_0 \times B^* \to \mathcal{P}(B_0^* \times B)$, defined as

$$T_L(x,y) = \left(L'_x(x,y), -L'_y(x,y) + N_{K^*}(y)\right) = \left(g'(x) + [G'(x)]^*(y), -G(x) + N_{K^*}(y)\right), \tag{2.41}$$

is the *saddle-point operator* associated with the Lagrangian L, where $N_{K^*} : B^* \to \mathcal{P}(B)$ denotes the normalizing operator of the cone K^*.

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

Proposition 2.8.8. *Under assumptions (A1) and (A2), the operator* T_L*, defined in* (2.41)*, 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 [56] 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 (2.40). The right inequality in (2.40) implies that the Lagrangian condition (2.38) holds. The left one implies that

$$0 \geq \langle G(x), z - y \rangle \tag{2.42}$$

for all $z \in K^*$, which entails feasibility of x and also (2.39), so that (x,y) is a KKT pair. Conversely, if (x,y) is a KKT pair then the Lagrangian condition (2.38) implies, by convexity of the Fréchet differentiable function $L(\cdot,y) : B_0 \to \mathbb{R}$, the right inequality in (2.40), while (2.39) and primal feasibility of x imply (2.42) for all $z \in K^*$, so that the left inequality in (2.40) 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 2.8.6. Item (iv) has been proved in [56]. □

2.9 Augmented Lagrangian Methods

We are concerned now with the general optimization problem:

$$\min g(x) \tag{2.43}$$

$$\text{s.t. } G(x) \precsim 0, \tag{2.44}$$

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 in Chapter 4 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 [18] (previously, augmented Lagrangian methods had been proposed for equality constrained problems, e.g. [32] and [54]). Such algorithms were further studied in [58], [59], [5], [42]. Its connection with the Proximal Point method was established in [60] and treated, also, in [3], [6], [25], [38] and [33]. The augmented Lagrange functional $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_{++} \to \mathbb{R}$, used in [59] and [60], is defined as

$$L(x, y, \rho) = g(x) + \rho \sum_{i=1}^{m} \left[\left(\max\{0, y_i + (2\rho)^{-1} g_i(x)\} \right)^2 - y_i^2 \right], \tag{2.45}$$

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

$$\min g(x) \tag{2.46}$$

$$\text{s.t. } g_i(x) \le 0 \qquad (1 \le i \le m). \tag{2.47}$$

The augmented Lagrangian method in [60] 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

$$x^{k+1} \in \mathrm{argmin}_{x \in \mathbb{R}^n} L(x, y^k, \lambda_k), \tag{2.48}$$

$$y_i^{k+1} = \max\{0, y_i^k + (2\lambda_k)^{-1} g_i(x^{k+1})\} \quad (1 \le i \le m). \tag{2.49}$$

It is proved in [60] that, if g and the g_i's $(1 \le i \le m)$ are convex and problem (2.46)–(2.47) has both primal and dual solutions, then the sequence $\{y^k\}$ (when well defined) converges to a dual solution (i.e. to a vector of KKT multipliers for the constraints in

(2.47)). 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 $F = \{x \in \mathbb{R}^n : g_i(x) \le 0\ (1 \le i \le m)\}$, and its cluster points, if any, are solutions of problem (2.46)–(2.47). The results on the primal sequence can be improved by considering a doubly augmented Lagrangian, i.e. replacing (2.48) by

$$x^{k+1} = \operatorname{argmin}_{x\in\mathbb{R}^n}\{L(x, y^k, \lambda_k) + \lambda_k\|x - x^k\|^2\}, \tag{2.50}$$

in which case it is proved in [60] 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 (2.46)–(2.47), provided it has both primal and dual solutions.

The results in [59] were extended in [66] 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 (2.44). 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 [66] takes the form

$$\Lambda(x, y, \rho) = g(x) + \frac{1}{2}\rho\|P_{K^*}[y + \rho^{-1}G(x)]\|^2 - \frac{1}{2}\rho\|y\|^2. \tag{2.51}$$

The methods in [66] update the primal variables through

$$x^{k+1} \in \operatorname{argmin}_{x\in B_1}\Lambda(x, y^k, \lambda_k). \tag{2.52}$$

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

$$y^{k+1} = P_{K^*}\big(y^k + \lambda_k^{-1}G(x^{k+1})\big), \tag{2.53}$$

is such that the method reduces to the algorithm in [60], i.e. (2.48)–(2.49), when $B_1 = \mathbb{R}^n$, $H = \mathbb{R}^m$ and $K = \mathbb{R}^m_+$. Several alternatives are also presented for the regularization parameters λ_k, in some of which they are endogenously updated. The results in [66] 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 λ_k be small enough for large k. Additionally, since the Augmented Lagrangian Λ in (2.51) 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 [16]. 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

$$\bar{\Lambda}(x, y, \rho) = g(x) + \frac{\rho}{p} \left\| P_{\mathcal{L}^p_+(\Omega)}[h'(y) + \rho^{-1}G(x)] \right\|_p^p, \tag{2.54}$$

The Lagrangian of (2.54) is used in [16] 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

$$x^{k+1} \in \operatorname{argmin}_{x \in B} \bar{\Lambda}(x, y^k, \lambda_k), \tag{2.55}$$

$$y^{k+1}(\omega) = \left[\max\{0, [h'(y^k) + \lambda_k^{-1} G(x^{k+1})](\omega)\}\right]^{p-1}. \tag{2.56}$$

We observe that, up to constant terms in x, (2.54) reduces to (2.51) when $p = r = 2$, so that B_2 is hilbertian, and henceforth to (2.45), when B_1 is finite dimensional and Ω is finite. Thus, in such cases (2.55)–(2.56) reduce to (2.52)–(2.53) and to (2.48)–(2.49) 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 [16], based upon the Proximal Point methods in Banach spaces studied in [1] and [15], 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 Ω is countable, so that $\mathcal{L}^q(\Omega) = \ell_q$. There is no primal regularization in (2.54), 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 are avoided in Chapter 5 of this thesis, 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 (2.50) of [60].

In this thesis, 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 $h_r(y) = \frac{1}{r} \|y\|_{B_2^*}^r$ 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 error criteria developed on Chapter 3 for the proximal point method, which allow for constant relative errors, in the spirit of [64] and [62].

To our knowledge, all results presented in the following chapters are new, unless otherwise stated.

Chapter 3

On the existence of appropriate regularizing functions in general Banach spaces

We start with the definition of a basic class of regularizing functions.

Definition 3.0.1. We denote as $\mathcal{F}$ the set of functions $f : B \to \mathbb{R} \cup \{\infty\}$ which are strictly convex, lower semicontinuous and Gâteaux differentiable in the interior of their domain, which we assume to be nonempty.

As usual, f' will denote from now on the Gâteaux derivative of a function $f \in \mathcal{F}$. The methods we analyze in this thesis use, as an auxiliary device, functions $f \in \mathcal{F}$ which satisfy some assumptions, among which some or all of H1-H5, discussed in Chapter 2. Thus, it is important to exhibit functions which satisfy these properties in as large a class of Banach spaces as possible, and we focus our attention on functions of the form $f = \psi \circ \|\cdot\|_B : B \to \mathbb{R} \cup \{+\infty\}$, where $\psi : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{+\infty\}$ is defined by

$$\psi(t) = \begin{cases} \int_0^t \varphi(s)ds, & \text{if } t \in [0, b] \\ \infty & \text{otherwise,} \end{cases} \tag{3.1}$$

with $b \in (0, \infty]$, and φ is an *extended weight function* as defined next.

Definition 3.0.2. An *extended weight function* φ is a continuous and strictly increasing function $\varphi : [0, b) \to \mathbb{R}_+$ such that $\varphi(0) = 0$, where $b \in \mathbb{R}_+ \cup \{+\infty\}$.

A motivation for this choice are the functions $f_r(x) = \frac{1}{r} \|x\|_B^r$ with $r > 1$, which are potential candidates for regularizing functions. Observe that these functions are as in (3.1), with $\varphi(t) = t^{r-1}$, which is obviously a weight function according to the classical definition, as in section 2.4, and also in the sense of Definition 3.0.2. Note that, since extended weight functions are strictly increasing and nonnegative, ψ, as given by (3.1) is convex and strictly increasing, and hence $f = \psi \circ \|\cdot\|_B$ is convex and continuous.

Lemma 3.0.3. *The function $\psi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$, defined in 3.1, has the following differentiability properties:*

i) $\psi'_+(0) = \varphi(0) = 0$.

ii) $\psi'(t) = \varphi(t)$ *for all* $t \in (0, b)$.

iii) $\psi'_-(b) = \lim_{t\to b^-} \varphi(t) \overset{def}{=} \varphi(b^-) \in \mathbb{R} \cup \{+\infty\}$.

iv) $\partial\psi(t) = \emptyset$ *for all* $t \notin [0, b)$.

Proof. Elementary. □

When $b = \infty$ and $\varphi(b^-) = \infty$, we recover the definition of weight function as treated in section 2.4, but the case where the domain of the regularizing function f is not the whole space is important for us. Indeed, by allowing b to be finite, we add to the analysis the case where the domain of the candidate function f is a ball. It is not an essential modification and all discussed properties of duality mappings (see section 2.4) can be stated in an analogous way. For this purpose, we only need to extend, appropriately, the definition of duality map.

Definition 3.0.4. The extended duality mapping of weight φ is the mapping $J_\varphi : B \to B^*$, defined by

$$J_\varphi(x) = \begin{cases} \{x^* \in B^* | \langle x^*, x\rangle = \|x^*\|_* \|x\|, \|x^*\|_* = \varphi(\|x\|)\} & \text{if } \|x\| < b \\ \{x^* \in B^* | \langle x^*, x\rangle = \|x^*\|_* b, \|x^*\|_* \geq \varphi(b^-)\} & \text{if } \|x\| = b \\ \emptyset & \text{otherwise.} \end{cases}$$

The following lemma establishes the good definition and the respective version of the Asplud's theorem (i.e. J1 of section 2.4) for extended duality mappings.

Lemma 3.0.5. *The extended duality mapping, J_φ, of the extended weight function φ given by Definition 3.0.4 is well defined. Moreover,*

EJ1: $J_\varphi(x) = \partial[\psi \circ \|\cdot\|](x)$ *for all* $x \in B$.

Proof. If $b = \infty$ then there is nothing to add to the discussion in Remark 4.4 and Theorem 4.4 in pages 25 and 26 respectively of [22]. These results also apply to the case when $\|x\| < b$. Since $J_\varphi(x) = \emptyset$ when $\|x\| > b$ by Definition 3.0.4, we assume from now on, without loss of generality, that $\|x\| = b < \infty$. Let $\varphi(b^-) = \lim_{t\to b^-} \varphi(t)$. We consider two cases: $\varphi(b^-) = \infty$ and $\varphi(b^-) < \infty$. In the first one, we get $\partial\psi(b) = \emptyset$, by Lemma 3.0.3(iii) and $J_\varphi(x) = \emptyset$ for any $x \in B$ such that $\|x\| = b$, by Definition 3.0.4, because there exists no $x^* \in B^*$ with $\|x^*\| \geq +\infty$. Thus EJ1 holds for all $x \in B$ such that $\|x\| \geq b$. We study next the remaining case, i.e. both b and $\varphi(b^-)$ finite, and $\|x\| = b$. We prove first that $J_\varphi(x) \neq \emptyset$. Consider $x_0 = \varphi(b^-)x \in B$. By the Hahn-Banach Theorem (see e.g. page 776 of [67]), there exists $v \in B^*$ with $\|v\|_* = 1$ and $\langle v, x_0\rangle = \|x_0\|$. Hence, taking $x^* = \varphi(b^-)v$, we get an element x^* of B^* satisfying

$$\|x^*\| = \left\|\varphi(b^-)v\right\| = \varphi(b^-),$$

and

$$\langle x^*, x\rangle = \|x_0\| = \left\|\varphi(b^-)x\right\| = \|x^*\|\, b.$$

Thus, in view of Definition 3.0.4, $x^* \in J_\varphi(x)$. We proceed to prove the inclusion " $\subset$ " in EJ1. Take any $x^* \in J_\varphi(x)$. Then

$$\|x^*\| \geq \varphi(b^-) = \psi'_-(b) \geq \frac{\psi(\|y\|) - \psi(b)}{\|y\| - b},$$

for all $y \in B$ such that $\|y\| < b$, using Lemma 3.0.3(iii) in the equality. Thus,

$$\psi(\|y\|) - \psi(b) \geq \|x^*\| \left(\|y\| - b\right) = \|x^*\| \, \|y\| - \|x^*\|\, b$$

$$= \|x^*\| \, \|y\| - \langle x^*, x\rangle \geq \langle x^*, y\rangle - \langle x^*, x\rangle = \langle x^*, y - x\rangle$$

for all $y \in B$ such that $\|y\| < b$. It follows that $x^* \in \partial[\psi \circ \|\cdot\|](x)$. For proving the opposite inclusion, take $x^* \in \partial[\psi \circ \|\cdot\|](x)$, so that

$$\psi(\|y\|) - \psi(\|x\|) \geq \langle x^*, y - x\rangle \tag{3.2}$$

for all $y \in B$. Taking now any $y \in B$, such that $\|y\| = \|x\| = b$, we get

$$0 \geq \langle x^*, y - x\rangle = \langle x^*, y\rangle - \langle x^*, x\rangle,$$

so that $\langle x^*, y\rangle \leq \langle x^*, x\rangle$ for all $y \in B$ such that $\|y\| = b$. We conclude that

$$\|x^*\|\, b = \left(\sup_{\|z\|=1} \langle x^*, z\rangle\right) b = \sup_{\|y\|=b} \langle x^*, y\rangle \leq \langle x^*, x\rangle \leq \|x^*\|\, b,$$

and therefore,

$$\langle x^*, x\rangle = \|x^*\| \, b. \tag{3.3}$$

From (3.2) with $y = \frac{t}{b}x$ and $0 < t < b$, we get

$$\psi(t) - \psi(b) \geq \langle x^*, y - x\rangle = \frac{(t-b)}{b}\langle x^*, x\rangle = (t-b)\,\|x^*\|\,,$$

where the last equality follows from (3.3). Hence,

$$\|x^*\|\, b \geq \frac{\psi(t) - \psi(b)}{t-b} \tag{3.4}$$

for all $t \in (0, b)$. Taking limits in (3.4) with $t \to b$, we get, in view of Lemma 3.0.3(iii),

$$\|x^*\|\, b \geq \psi'_-(b) = \varphi(b^-). \tag{3.5}$$

Equations (3.3) and (3.5) ensure that $x^* \in J_\varphi(x)$, which completes the proof. □

We present next some properties of these extended duality mappings.

Proposition 3.0.6. *Let B be a reflexive Banach space and J_φ the extended duality mapping associated with the extended weight φ, with domain $[0, b)$. Then the following properties hold:*

EJ2: The inverse function of the extended weight φ, φ^{-1} is an extended weight function too and $J^*_{\varphi^{-1}}$, the extended duality mapping of weight φ^{-1} on B^*, is such that $x^* \in J_\varphi(x)$ with $x \in int(dom\ J_\varphi)$ if and only if $x \in J^*_{\varphi^{-1}}(x^*)$ and $x^* \in int(dom\ J^*_{\varphi^{-1}})$.

EJ3: For all $x \in B$ such that $\|x\| < b$, it holds that $\|x\|\, J_\varphi(x) = \varphi(\|x\|)J(x)$.

EJ4: If B is smooth (i.e., if B^* is strictly convex), then J_φ is single valued and norm-to-weak continuous on int(dom J_φ). Moreover, if B^* satisfies property (h), as in Definition 2.2.9, then J_φ :int(dom J_φ) $\to B^*$ is norm-to-norm continuous.*

EJ5: If B is strictly convex, then J_φ is strictly monotone on int(dom J_φ).

EJ6: If B is uniformly smooth, then J_φ is uniformly continuous on any bounded subset A of int(dom J_φ), (when b is finite and $\varphi(b^-) = \infty$ the result requires the additional assumption that $d(A, \partial\ domJ_\varphi) > 0$).

Proof. Properties EJ2–EJ5 are just properties J2-J5 of duality mappings (see section 2.4) adapted to the new context. They can be proved exactly as done in [22] for J2-J5 (Proposition 4.7(e), (f) and Corollary 4.5 in page 27 for EJ2, EJ3 and the first part of EJ4 respectively; Theorem 4.12 in page 30 and Proposition 5.2 in page 27 for the second and third statements in EJ4, and Theorem 1.8 in page 45 for EJ5). EJ6 is a direct consequence of J6, which can be easily proved, but we found no reference for it, even for the case of duality mappings with classical weight functions, and therefore we present next a proof.

If the result does not hold, then we can find a bounded subset A of dom J_φ 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 $x^k - y^k \xrightarrow[k\to\infty]{s} 0$ and $\|J_\varphi(x^k) - J_\varphi(y^k)\|_{B^*} \geq \epsilon$ 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 Definition 3.0.4 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, $M \in \mathbb{R}$ (with $M < b$ when b is finite and $\varphi(b^-) = \infty$, in view of the additional assumption in EJ6 for this case). Let $\bar{\varphi}(t) = \varphi(t)/t$, and $\bar{M} = \max_{t\in[m,M]} \bar{\varphi}(t)$. Using EJ3 and the definition of the normalized duality map J (see section 2.4), we have

$$\begin{aligned}
\epsilon &\leq \|J_\varphi(x^k) - J_\varphi(y^k)\|_* = \left\| \frac{\varphi(\|x^k\|)}{\|x^k\|} J(x^k) - \frac{\varphi(\|y^k\|)}{\|y^k\|} J(y^k) \right\|_* \\
&\leq \bar{\varphi}(\|x^k\|)\|J(x^k) - J(y^k)\|_* + \left|\bar{\varphi}(\|x^k\|) - \bar{\varphi}(\|y^k\|)\right| \|J(y^k)\|_* \\
&\leq \bar{M}\|J(x^k) - J(y^k)\|_* + \left|\bar{\varphi}(\|x^k\|) - \bar{\varphi}(\|y^k\|)\right| \|y^k\| \\
&\leq \bar{M}\|J(x^k) - J(y^k)\|_* + M \left|\bar{\varphi}(\|x^k\|) - \bar{\varphi}(\|y^k\|)\right|.
\end{aligned}$$

Since $\lim_{k\to\infty} |\bar{\varphi}(\|x^k\|) - \bar{\varphi}(\|y^k\|)| = 0$ by uniform continuity of $\bar{\varphi}$ in $[m, M] \subset (0, b)$, it follows that $\|J(x^k) - J(y^k)\|_* \geq \epsilon/(2\bar{M})$ for large enough k, contradicting J4 of section 2.4. □

Next we state some properties of functions of the form $f = \psi \circ \|\cdot\|_B$ with ψ as in (3.1) and φ as in Definition 3.0.2.

Corollary 3.0.7. *If the extended weight function φ is such that $\varphi(b^-) = \infty$ then J_φ is surjective. Particularly, if B is smooth then, $f = \psi \circ \|\cdot\|_B$, with ψ as in* (3.1)*, is G-differentiable*[1] *on int(dom f) and it satisfies H4.*

Proof. In such a case dom $\varphi^{-1} = [0, +\infty)$, so that, in view of EJ3, any $x^* \in B^*$ belongs to int(dom$J^*_{\varphi^{-1}}$), and therefore there exist $x \in$ int(dom$J^*_{\varphi^{-1}})x^*$. Hence $x \in$ int(domJ_φ) and $x^* \in J_\varphi(x)$. When B is smooth, J_φ is single valued on int(domJ_φ) (see EJ4). Thus EJ1 in Lemma 3.0.5 ensures that f is G-differentiable on int(dom f) = int(domJ_φ). □

[1] In fact, F-differentiable, provided that B^* satisfies property (h); see EJ4.

Corollary 3.0.8. *Let B be strictly convex. If the extended weight function φ is such that either*

i) $b = +\infty$ (i.e. dom $\varphi = [0, +\infty)$), or

ii) b is finite and there exists $r > 1$ such that

$$t \to \int_0^{t^{1/r}} \varphi(s)ds \ : [0, b] \to \mathbb{R}_+ \cup \{+\infty\} \text{ is convex,} \tag{3.6}$$

then $f = \psi \circ \|\cdot\|_B$, with ψ as in (3.1), is strictly convex on dom f.

Proof. i) If $b = +\infty$, then dom $J_\varphi = B$ and, in view of EJ1 in Lemma 3.0.5, together with EJ5, $\partial f : B \to \mathcal{P}(B^*)$ is strictly monotone. Hence f is strictly convex on $B =$ dom$f$.

ii) The function $\psi \circ (\cdot)^{1/r}$ is convex and strictly increasing. Since $r > 1$ strict convexity of B ensures strict convexity of $\|\cdot\|_B^r$. Then the composition $f = \left[\psi \circ (\cdot)^{1/r}\right] \circ \|\cdot\|_B^r$ is strictly convex on its domain.

□

Corollary 3.0.9. *Suppose that the extended weight function φ is such that one of the following conditions holds:*

a) b is finite,

(b) For all $t_0 \in \mathbb{R}_+$, $\lim_{t\to\infty} \left[\varphi(t)(t - t_o) - \int_{t_o}^t \varphi(s)ds\right] = \infty$,[2]

(c) φ is differentiable and $\varphi'(t) \geq \alpha > 0$ for all $t \in \mathbb{R}_+$,

(d) φ is differentiable, convex and $\varphi'(0) > 0$.

Then $f = \psi \circ \|\cdot\|_B$, with ψ as in (3.1), satisfies H1.

Proof. If b is finite then int(dom f) is bounded, so that $\{x \in \text{int(dom } f) \mid D_f(y, x) \leq \beta\}$ is bounded (when nonempty) for all $y \in$ dom f. Suppose now that $b = +\infty$ and that there exists $\beta \in [0, \infty)$ and $y \in$ dom f such that the right level set above is unbounded. Then we can find a sequence $\{x^k\} \subset$ int(dom f) satisfying $\beta \geq D_f(y, x^k)$ for all k. At this point we remark that the definition of the Bregman distance D_f can be stated without the assumption

[2]This condition is essentially the one used in Corollary 1(i) of [17], where a proof of this result is offered for the particular wheight function $\varphi(t) = rt^{r-1}$, $r > 1$.

of G-differentiability of f in int(dom f), as presented in Definition 2.3.9. In fact, following [17], we can assume that

$$D_f(y, x^k) = f(y) - f(x^k) - \inf_{v \in \partial f(x^k)} \langle v, y - x^k \rangle. \tag{3.7}$$

Taking any $v^k \in \partial f(x^k) = J_\varphi(x^k)$ (see EJ1), it follows that

$$\begin{aligned} \beta &\geq f(y) - f(x^k) - \langle v^k, y - x^k \rangle = \psi(\|y\|) - \psi(\|x^k\|) - \langle v^k, y \rangle + \langle v^k, x^k \rangle \\ &= \psi(\|y\|) - \psi(\|x^k\|) - \langle v^k, y \rangle + \|v^k\|_* \|x^k\| \\ &\geq \psi(\|y\|) - \psi(\|x^k\|) - \|v^k\|_* \|y\| + \|v^k\|_* \|x^k\| \\ &= \varphi(\|x^k\|)(\|x^k\| - \|y\|) - \int_{\|y\|}^{\|x^k\|} \varphi(s)ds. \end{aligned} \tag{3.8}$$

Thus, taking limits when k goes to infinity in the rightmost expression of (3.8), we get from condition (b) that $\beta = +\infty$, which is a contradiction. Observe, now, that condition (d) implies condition (c), which in turn implies condition (b), so that the proof is complete. □

Corollary 3.0.10. *Let B be uniformly smooth. Take $f = \psi \circ \|\cdot\|_B$, with ψ as in* (3.1).

i) If $b = \infty$ then f satisfies H3.

ii) if b is finite and $\varphi(b^-) < \infty$ then f' is uniformly continuous on int(dom f).

iii) if b is finite and $\varphi(b^-) = \infty$ then f' is uniformly continuous on any set $C \subset$ int(dom f) such that $C \subset B[0, M] = \{x \in B \ : \ \|x\| \leq M\}$ for some $M < b$.

Proof. The result follows from EJ1 and EJ6. □

We need in the sequel the following result on Gâteaux differentiable functions.

Proposition 3.0.11. *Let $f : B \to \mathbb{R}$ be Gâteaux differentiable. If f' is uniformly continuous on bounded sets, then*

i) both f and f' are bounded on bounded subsets of B.

ii) If $\{x^k\}$, $\{y^k\}$ are bounded sequences satisfying $x^k - y^k \xrightarrow[k\to\infty]{\text{s}} 0$, then

$$\lim_k \frac{D_f(y^k, x^k)}{\|y^k - x^k\|} = 0.$$

Proof. For (i), let U be a bounded subset of B. Take $\bar{x} \in B$ and $U \subseteq B(\bar{x}, \rho) = \{x \in B : \|x - \bar{x}\| \leq \rho\}$. Take δ such that $\|f'(x) - f'(x')\|_* \leq 1$ for all $x, x' \in B(\bar{x}, \rho)$ satisfying $\|x' - x\| \leq \delta$. 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 $|f(x)| \leq \hat{\gamma}$. To prove (ii) just observe that

$$D_f(y^k, x^k) \leq D_f(y^k, x^k) + D_f(x^k, y^k) \leq \left\| f'(x^k) - f'(y^k) \right\|_* \left\| x^k - y^k \right\|.$$

□

3.1 Powers of the norm

Next we summarize our results on the validity of H1–H5 for the function $f_r(x) = \frac{1}{r} \|x\|_B^r$ with $r > 1$. Remember that such conditions are:

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

H2: f is uniformly totally convex on int(dom f).

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

H4: f' is onto.

H5: f satisfies the separability requirement.

Proposition 3.1.1. *Take $f_r(x) = \frac{1}{r} \|\cdot\|_B^r : B \to \mathbb{R}$ with $r > 1$.*

i) If B is a uniformly smooth and uniformly convex Banach space, then f_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.

iii) If $B = \ell_p$ $(1 < p < \infty)$ then $f_p(x) = \frac{1}{p} \|x\|_p^p$ satisfies H5.

Proof. Consider the weight function $\varphi(t) = t^{r-1}$, which gives $\psi(t) = \int_0^t \varphi(s) ds = \frac{1}{r} t^r$ for all $t \in \mathbb{R}_+$. It follows from Corollaries 3.0.10 and 3.0.7 that f satisfies H3 and H4 respectively. H1 is a consequence of Corollary 3.0.9, but it is not a new result in this particular case. In fact it has been proved in [17] that, for all $r > 1$, f_r satisfies H1 and H2 when B is uniformly convex (see Example 2.3.24 for more information on H2).

For (ii), note that in the case of a Hilbert space, f_2' is the identity, which is certainly weak-to-weak continuous. Concerning f_p in ℓ_p, it has been proved in [8], Proposition 8.2, that f_p' is weak-to-weak* continuous too. Hence, in these two cases, i.e. (ii) and (iii), f satisfies the separability condition, which is implied by the weak-to-weak* continuity of f'. □

Unfortunately, examples in the following section show that for $B = \ell_p$ or $B = \mathcal{L}^p[\alpha, \beta]$ the function $f_r(x) = \frac{1}{r}\|x\|_p^r$, $1 < r < \infty$, does not satisfy H5, excepting in the two cases considered in Proposition 3.1.1(ii) and (iii).

3.2 Counterexamples for the separability requirement

In connection with the existence of functions defined on Banach spaces whose derivatives are weak-to-weak* continuous or which satisfy the separability requirement, i.e. H5, Proposition 3.1.1(ii)-(iii) gives us a partial answer. We discuss next the fact that for $B = \ell_p$ or $B = \mathcal{L}^p[a, b]$ $(1 < p < \infty)$, the function $f(x) = \|x\|_p^r$ $(r > 1)$ does not satisfy the separability condition (and so, a fortiori, f' is not weak-to-weak* continuous), excepting for the two cases considered in Proposition 3.1.1(ii)-(iii).

Example 3.2.1. $B = \ell_p$, $f(x) = \|x\|_p^r$, $r \neq p$. Let $\{e^k\}$ be the canonical basis of ℓ_p, i.e. $e_j^k = \delta_{kj}$ (Kronecker's delta), and consider the sequences $\{x^k\}$, $\{y^k\} \subset \ell_p$ defined as $x^k := 2e^1 + \alpha e^k$ and $y^k := e^1 + \beta e^k$, where

$$\alpha = 2^{\frac{p-1}{p-r}} \qquad \beta = 2^{\frac{r-1}{r-p}}. \tag{3.9}$$

Both $\{x^k\}$ and $\{y^k\}$ are contained in ℓ_p and converge weakly to $x = 2e^1$ and $y = e^1$ respectively. Since $f'(x)_j = r\,\|x\|_p^{r-p}\,|x_j|^{p-2}x_j$, it is easy to check that

$$f'(x^k) = r(2^p + \alpha^p)^{\frac{r}{p}-1}(2^{p-1}e^1 + \alpha^{p-1}e^k)$$

and

$$f'(y^k) = r(1 + \beta^p)^{\frac{r}{p}-1}(e^1 + \beta^{p-1}e^k)$$

for all k. Therefore, if $k > 1$, we have

$$\langle f'(x^k) - f'(y^k), x - y\rangle = \langle f'(x^k) - f'(y^k), e^1\rangle =$$
$$r\left[(2^p + \alpha^p)^{\frac{r}{p}-1}2^{p-1} - (1 + \beta^p)^{\frac{r}{p}-1}\right] = 0,$$

because of (3.9). Thus,

$$\liminf_{k\to\infty} |\langle f'(x^k) - f'(y^k), x - y\rangle| = 0,$$

which implies, since $x \neq y$, that f does not satisfy the separability condition H5.

For $B = \mathcal{L}^p[a,b]$, in which case $f'(x)(t) = r\,\|x\|_p^{r-p}\,|x(t)|^{p-2}x(t)$, we consider two cases: $p = 2$ and $p \neq 2$.

Example 3.2.2. $B = \mathcal{L}^2[a,b]$, $r \neq 2$. Since $p = 2$, B is a Hilbert space, and the natural extension of the counterexample above for ℓ_p works: let $\{e^k\}$ be a complete orthonormal system and take $x^k = 2e^1 + \alpha e^k$ and $y^k = e^1 + \beta e^k$, where α and β are as in (3.9), with $p = 2$. The same computation as in the case of ℓ_p holds, and f does not satisfy the separability requirement H5.

Example 3.2.3. $B = \mathcal{L}^p[a,b]$, $p \neq 2$. [3] If $p \neq 2$ we want to establish the result also for the case of $r = p$, and so the counterexample above is not enough. Take $h : \mathbb{R} \to \mathbb{R}$ such that h is periodic with period $b - a$, and let

$$\beta = \int_a^b h(t)dt, \quad \gamma = \int_a^b |h(t)|^p dt \quad \text{and} \quad \delta = \int_a^b |h(t)|^{p-2}h(t)dt.$$

We show next that, for $p \neq 2$, h can be chosen so that $\beta = 0$ and $\gamma, \delta \in (0, \infty)$. If $p > 2$, take

$$h(t) = \begin{cases} 3 & \text{if } \ t \in [a, a + (b-a)/4] \\ -1 & \text{if } \ t \in [a + (b-a)/4, b], \end{cases} \tag{3.10}$$

extended to the whole real line with period $b - a$. Then,

$$\beta = 0, \quad \gamma = \frac{3}{4}(b-a)(3^{p-1} + 1) > 0 \quad \text{and} \quad \delta = \frac{3}{4}(b-a)(3^{p-2} - 1) > 0.$$

If $p \in (1, 2)$, then multiply h as given by (3.10) by -1, so that we have the same values as before for β and γ, while $\delta = \frac{3}{4}(b-a)(1 - 3^{p-2}) > 0$.

Define $x^k(t) = h(a + k(t-a))$. Using the fact that $\beta = 0$, it is easy to establish that $\{x^k\}$ is weakly convergent to $x = 0$. Let

$$\theta = \left[\frac{\delta}{\gamma}\left(\frac{\gamma}{b-a}\right)^{r/p}\right]^{\frac{1}{r-1}}.$$

Define y^k, $y \in \mathcal{L}^q[a,b]$ as $y^k(t) = y(t) = \theta$ for all $t \in [a,b]$. It is easy to check that

$$\langle f'(x^k) - f'(y^k), x - y\rangle = r\theta\big[(b-a)^{r/p}\theta^{r-1} - \gamma^{r/p-1}\delta\big]. \tag{3.11}$$

[3] This counterexample is based upon [22], Remark 4.1.5.

In view of the definition of θ, we get from (3.11) that $\langle f'(x^k) - f'(y^k), x - y\rangle = 0$ for all k, so that

$$\liminf_{k\to\infty} |\langle f'(x^k) - f'(y^k), x - y\rangle| = 0$$

Since $x = 0 \neq y$, it follows that f does not satisfy the separability requirement H5.

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 in Chapter 4, 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 3.1.1, whose results trivially hold for all positive multiples of $\|\cdot\|_B^r$.

3.3 A candidate penalty function in a ball

In this section, we study the function $g : B \to \mathbb{R} \cup \{+\infty\}$, defined as

$$g(x) = \begin{cases} 1 - \sqrt{1 - \|x\|^2} & \text{if } \|x\| \leq 1 \\ \infty & \text{otherwise,} \end{cases} \tag{3.12}$$

in connection with properties H1, H2, H4 and the variants of H3 introduced in Section 2.5, in order to use it in extensions to Banach spaces of methods like GPPB of section 2.6. We will study such extensions in Chapter 6.

Let us consider the real function $\varphi : [0, 1) \to \mathbb{R}_+$ defined as

$$\varphi(t) = \frac{t}{\sqrt{1 - t^2}} \tag{3.13}$$

We show next that g is a function of the form $\psi \circ \|\cdot\|_B$, with ψ as in (3.1) associated to the weight function given by (3.13).

Lemma 3.3.1. *Let B be a real reflexive Banach space. Take φ as in* (3.13) *and g as in* (3.12). *Then φ is an extended weight function with $\varphi(1^-) = \lim_{t\to 1^-} \varphi(t) = \infty$ and $g = \psi \circ \|\cdot\|_B$ with ψ as in* (3.1)

Proof. Elementary. □

Let $C := \{x \in B : \|x\| \leq 1\}$. Observe that C is the domain of g.

Corollary 3.3.2. *Let B be a real reflexive Banach space. Take g as in* (3.12) *and $f_2 = (1/2)\|\cdot\|^2$. Then* *i) g satisfies H1.*

ii) If B is strictly convex and smooth then g is strictly convex, G-differentiable on int(dom g) and its derivative is given by

$$g'(x) = \frac{1}{\sqrt{1 - \|x\|^2}} f_2'(x), \tag{3.14}$$

for all $x \in int(C)$. Moreover, g satisfies H4 and the inverse function of g', namely $(g')^{-1} : B^* \to int(C)$, exists and it is the Gâteaux derivative of $[\sqrt{1 + (\cdot)^2} - 1] \circ \|\cdot\|_{B^*}$.

iii) If B^* is uniformly convex, then g satisfies H3.c.

Proof. Choose φ as in (3.13) and apply Lemma 3.3.1 to get $g = \psi \circ \|\cdot\|_B$ with ψ given by (3.1). Then results follows directly from Corollaries 3.0.9, 3.0.8, 3.0.7 and 3.0.10 respectively. The formula for g', (3.14), is a consequence of EJ3. The expression for the inverse of g' follows from the fact that $t \longmapsto \frac{t}{\sqrt{1+t^2}} : \mathbb{R}_+ \to [0, 1)$ is the inverse function of the weight φ, namely φ^{-1}, hence the inverse operator of $g' = J_\varphi : \text{int}(C) \to B^*$ is given by $J^*_{\varphi^{-1}}$ (see EJ2). □

Observe, now, that $g = \Phi \circ f_2$, where $f_2 : B \to \mathbb{R}$ is given by $f_2(x) = \frac{1}{2} \|\cdot\|_B^2$ and $\Phi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is defined as

$$\Phi(t) = \begin{cases} 1 - \sqrt{1 - 2t} & \text{if } t \in (-\infty, 1/2] \\ \infty & \text{otherwise.} \end{cases} \tag{3.15}$$

The function f_2 have been already studied on section 3.1 and it is certainly well behaved. We discuss next some relevant properties of Φ.

Lemma 3.3.3. *Take Φ as defined by (3.15). Then*

i) $\Phi'(t) = \frac{1}{\sqrt{1-2t}}$ for all $t \in (-\infty, 1/2)$, so Φ is differentiable on the interior of its domain,

ii) $\Phi''(t) = (1 - 2t)^{-\frac{3}{2}} > 0$ for all $t \in (-\infty, 1/2)$, and consequently

(a) Φ is strictly convex,

(b) $\Phi'(t) \geq \Phi'(0) = 1$ for all $t \in [0, 1/2)$.

Proof. Elementary. □

Corollary 3.3.4. *If B is uniformly convex and smooth, then g satisfies H2. In fact for all $x \in int(C)$ and all $t \geq 0$ it holds that*

$$\nu_g(x,t) \geq \Phi'(f_2(x))\nu_{f_2}(x,t) \geq \nu_{f_2}(x,t). \tag{3.16}$$

Proof. Choose f as the restriction of f_2 to the interior of C and use Propositions 2.3.17 and 2.3.18 to get the first inequality in (3.16). The second inequality follows immediately from Lemma 3.3.3(ii)-(b). Since f_2 satisfies H2 when B is uniformly convex and smooth (see Example 2.3.24), the result follows from (3.16). □

Concerning H3.a, we present now an example showing that, in general, g does not enjoy this property.

Example 3.3.5. Let $B = \ell_p$ with $p \in (1, \infty)$. Take $a \in (0,1)$ and define $\theta = [(\frac{a+1}{2})^{1/p} + 1]/2$. Take any sequence $\{\mu_k\} \subset [\theta, 1)$ such that $\lim_{k\to\infty} \mu_k = 1$, and define

$$\lambda_k = (2\mu_k^p - 1)^{\frac{1}{p}}, \quad \delta_k = 2\mu_k - 1, \quad \gamma_k = (2\delta_k^p - 1)^{\frac{1}{p}},$$

Let $\{e^k\}$ be the canonical basis of ℓ_p, as presented in Example 3.2.1, and consider the sequences $\{x^k\}$, $\{y^k\}$ given by

$$x^k = 2^{-\frac{1}{p}}(e^1 + \gamma_k e^k) \quad \text{and} \quad y^k = 2^{-\frac{1}{p}}(e^1 + \lambda_k e^k).$$

It is easy to check that

$$||x^k||_p = \mu_k < 1 \quad \text{and} \quad ||y^k||_p = \delta_k < 1 \tag{3.17}$$

and that

$$x^k \xrightarrow[k\to\infty]{\text{w}} 2^{-\frac{1}{p}}e^1 \quad \text{and} \quad y^k \xrightarrow[k\to\infty]{\text{w}} 2^{-\frac{1}{p}}e^1. \tag{3.18}$$

Thus, we have found two sequences, namely $\{x^k\}$ and $\{y^k\}$, contained in int(C), both of which are weakly convergent to the same point, namely $x = 2^{-\frac{1}{p}}e^1$. Since $g'(x)_j = \frac{1}{\sqrt{1-||x||^2}}\,||x||_p^{2-p}\,|x_j|^{p-2}x_j$, it is also easy to check that

$$\langle g'(y^k), x^k - y^k\rangle = \frac{1}{2}(\mu_k)^{2-p}\lambda_k^{p-1}\frac{(\lambda_k - \gamma_k)}{\sqrt{1-\mu_k^2}}. \tag{3.19}$$

Let us consider, now, the real function $\eta : [\xi, 1] \to \mathbb{R}$ defined as $\eta(t) = (2t^p - 1)^{1/p}$, with $\xi = (\frac{a+1}{2})^{1/p}$. Then

$$\eta'(t) = \frac{2t^{p-1}}{(2t^p - 1)^{\frac{p-1}{p}}} \quad \text{and} \quad \sup_{t\in[\xi,1]} |\eta'(t)| \leq \frac{2}{a^{\frac{p-1}{p}}} \stackrel{def}{=} M.$$

Hence, η is Lipschitz continuous with constant M and

$$|\lambda_k - \gamma_k| = |\eta(\mu_k) - \eta(\delta_k)| \leq M|\mu_k - \delta_k|. \tag{3.20}$$

From equations (3.19)-(3.20) we get

$$\begin{aligned}
|\langle g'(y^k), x^k - y^k\rangle| &= \frac{1}{2}\mu_k^{2-p}\lambda_k^{p-1}\sqrt{\frac{|\lambda_k - \gamma_k|^2}{|(1-\mu_k)(1+\mu_k)|}} \\
&= \frac{1}{2}(\mu_k)^{2-p}\lambda_k^{p-1}\sqrt{\frac{|\lambda_k - \gamma_k|}{|\mu_k - \delta_k|}}\sqrt{\frac{|\lambda_k - \gamma_k|}{(1+\mu_k)}} \\
&\leq \frac{1}{2}\sqrt{M}(\mu_k)^{2-p}\lambda_k^{p-1}\sqrt{|\lambda_k - \gamma_k|}.
\end{aligned} \tag{3.21}$$

Taking limits when k goes to infinity in the extreme expressions of (3.21), it follows that

$$\lim_{k\to\infty}\langle g'(y^k), x^k - y^k\rangle = 0. \tag{3.22}$$

Since

$$D_g(x^k, y^k) = g(x^k) - g(y^k) - \langle g'(y^k), x^k - y^k\rangle = \sqrt{1-\mu_k^2} - \sqrt{1-\delta_k^2} - \langle g'(y^k), x^k - y^k\rangle,$$

we get from (3.22) that

$$\lim_{k\to\infty} D_g(x^k, y^k) = 0. \tag{3.23}$$

It follows from (3.17), (3.18) and (3.23) that $\{x^k\}$ and $\{y^k\}$ satisfy the hypotheses of H3. We show next that $\lim_{k\to\infty}\left[D_g(x, x^k) - D_g(x, y^k)\right] \neq 0$. Easy computations show that

$$\langle g'(y^k) - g'(x^k), x - x^k\rangle =$$

$$2^{-\frac{p+4}{2p}}\lambda_k^{p-1}\gamma_k\mu_k^{2-p}\left[\left(\frac{\delta_k}{\mu_k}\right)^{2-p}\left(\frac{\gamma_k}{\lambda_k}\right)^{p-1}\sqrt{\frac{1+\mu_k}{1+\delta_k}} - \sqrt{2}\right]\frac{1}{\sqrt{1-\mu_k^2}}. \tag{3.24}$$

It follows from (3.24) that

$$\lim_{k\to\infty}\langle g'(y^k) - g'(x^k), x - x^k\rangle = -\infty. \tag{3.25}$$

By Proposition 2.3.10(i)-(c),

$$D_g(x, x^k) - D_g(x, y^k) = -D_g(x^k, y^k) + \langle g'(y^k) - g'(x^k), x - x^k\rangle. \tag{3.26}$$

We conclude from (3.26) and (3.25) that

$$\lim_{k\to\infty} \left[D_g(x, x^k) - D_g(x, y^k)\right] = -\infty. \tag{3.27}$$

Equation (3.27) implies that g does not satisfy H3.a. It follows easily from Proposition 2.3.10(i)-(c), that functions which are G-differentiable in the interior of its domain, and which satisfy H2 and H3, also satisfy H3.a. Thus, this example also shows that the function g under consideration does not satisfy H3 either.

The fact that g satisfies neither H3 nor H3.a (and, in fact, that we know no other function whose domain is the interior of a ball, and which satisfies all the required properties) will have as a consequence that our results in Chapter 6, for proximal methods with penalization, will be somewhat weaker that the result for the case in which the domain of the regularizing functions is the whole space B. In such situation condition H3.b will be used, which is satisfied by g as we show bellow.

Proposition 3.3.6. *Let B be a strictly convex and smooth reflexive Banach space and g be given by* (3.12). *Then g satisfies H3.b. Moreover, given any sequence $\{x^k\}_k \subset int(C) = B(0,1)$, with $\lim_k d(x^k, \partial C) = 0$ (i.e. $\lim_k \|x^k\| = 1$), the sequence of parameters $\{\rho_k\} \subset \mathbb{R}_{++}$, can be taken for all k, as*

$$\rho_k \le d(x^k, \partial C) = 1 - \|x^k\|.$$

Proof. Let $\{x^k\}_k \subset \mathrm{int}(C) = B(0,1)$, with $\lim_k d(x^k, \partial C) = 0$. Take any weak limit $\hat{x}$ of such sequence and $\{y^k\}_k \subset \mathrm{int}(C)$. Then, taking φ as in (3.13), we get from Corollary 3.3.2 (ii) and Definition 2.3.9

$$\begin{aligned}
D_g(\hat{x}, x^k) - D_f(\hat{x}, y^k) &\le D_g(\hat{x}, x^k) = \sqrt{1 - \|x^k\|^2} - \sqrt{1 - \|\hat{x}\|^2} - \langle g'(x^k), \hat{x} - x^k\rangle \\
&\le \sqrt{1 - \|x^k\|^2} + \langle g'(x^k), x^k\rangle - \langle g'(x^k), \hat{x}\rangle \\
&\le \sqrt{1 - \|x^k\|^2} + \varphi(\|x^k\|)\, \|x^k\| + \varphi(\|x^k\|)\, \|\hat{x}\| \\
&\le \sqrt{1 - \|x^k\|^2} + 2\varphi(\|x^k\|).
\end{aligned}$$

Thus, taking $\rho_k \leq 1 - \|x^k\|$, we get

$$\begin{aligned}
\limsup_k \rho_k \big[D_g(\hat{x}, x^k) - D_g(\hat{x}, y^k)\big] &\leq \limsup_k \rho_k \big[\sqrt{1 - \|x^k\|^2} + 2\varphi(\|x^k\|)\big] \\
&\leq \limsup_k 2(1 - \|x^k\|) \frac{\|x^k\|}{\sqrt{1 - \|x^k\|^2}} \\
&\leq 2 \limsup_k \sqrt{\frac{(1 - \|x^k\|)^2}{1 - \|x^k\|^2}} \\
&= 2 \limsup_k \sqrt{\frac{1 - \|x^k\|}{1 + \|x^k\|}} \leq 2 \limsup_k \sqrt{1 - \|x^k\|} = 0.
\end{aligned}$$

Next we present another property of g which will be useful in the analysis of Chapter 6.

Proposition 3.3.7. *Let B be a smooth Banach space and consider g as in* (3.12)*. If $\{x^k\}$ is any sequence in $int(C) = \{x \in B| \ \|x\| < 1\}$ such that $\lim_{k\to\infty} \|x^k\| = 1$ then $\lim_k \|g'(x^k)\| = +\infty$. Moreover, for any $x \in int(dom\ g)$ it holds that*

$$\lim_{k\to\infty} D_g(x, x^k) = \infty. \tag{3.28}$$

Proof. Since $\{x^k\}$ is bounded (and so is $\{J(x^k)\}$) equation (3.14) in Corollary 3.3.2 ensures the first statement. Observe also, that the result also follows from the fact that $g'(x^k) = \varphi(\|x^k\|)$ and the definition of φ, namely (3.13). For the second statement, note that

$$\begin{aligned}
D_g(x, x^k) &= g(x) - g(x^k) - \langle g'(x^k), x - x^k \rangle \\
&= \sqrt{1 - \|x^k\|^2} - \sqrt{1 - \|x\|^2} - \langle g'(x^k), x \rangle + \langle g'(x^k), x^k \rangle \\
&\geq \sqrt{1 - \|x^k\|^2} - \sqrt{1 - \|x\|^2} - \|g'(x^k)\|_* \|x\| + \|g'(x^k)\|_* \|x^k\| \\
&= \sqrt{1 - \|x^k\|^2} - \sqrt{1 - \|x\|^2} + \|g'(x^k)\|_* (\|x^k\| - \|x\|).
\end{aligned}$$

Observe that $\|x\| < 1$, and therefore $\|x^k\| - \|x\| \geq \alpha > 0$ for large enough k. The result then follows from the first statement, already established. ☐

Remark 3.3.8. It should be observed that all results in this section hold trivially for any ball $B[0, b]$ (i.e. with b not necessarily equal to 1). Moreover, there is no need to choose precisely the squared norm (i.e. f_2) for defining g. In other words, we can consider on functions of the form $g(x) = b - (b^r - \|x\|^r)^{1/r}$. Choosing, e.g.

$$\varphi(t) = \frac{t^{r-1}}{(b^r - t^r)^{\frac{r-1}{r}}} \quad \text{for } t \in [0, b) \quad \text{and} \quad \Phi(t) = b - (b^r - rt)^{1/r}$$

for all $t \in (-\infty, b^r/r]$, it can be seen that this more general g enjoys the same properties as the one given by (3.12). The same holds, more generally, with balls centered at an arbitrary point x_0, by considering $\hat{g}(x) = g(x - x_0)$, with g as above. In any case, the variational inequality problem VIP(T,$B[x_0, b]$) can be reduced to VIP($\tilde{T}$,$B[0, b]$) with $\tilde{T}(x) = T(x+x_0)$. Thus, it is not worthwhile to go deeper in the analysis of this case.

3.4 A candidate penalty function in a polyhedron

Let B be a real reflexive Banach space. Take $v^1, \ldots, v^p \in B^* \backslash \{0\}$ and $\alpha_1, \ldots, \alpha_p \in \mathbb{R}$. Define the closed and convex set C by

$$C = \{x \in B \mid \langle v^i, x\rangle \geq \alpha_i, \ i = 1, ..., p\}, \tag{3.29}$$

and assume that int(C) is nonempty. Consider the function $c : C \to \mathbb{R}$ defined by

$$c(x) = \begin{cases} \frac{1}{2}\|x\|_B^2 + \sum_{i=1}^p (\langle v^i, x\rangle - \alpha_i) \log(\langle v^i, x\rangle - \alpha_i) & \text{if } x \in C \\ \infty & \text{otherwise.} \end{cases} \tag{3.30}$$

It is known that this function has nice properties in Hilbert spaces (see Example 2.6.6). We state next some related properties on Banach spaces.

Proposition 3.4.1. *Let B be uniformly smooth and uniformly convex and take c as in* (3.30)*. Then*

i) c is strictly convex and continuous on C. Moreover, c is F-differentiable on int(C), with G-derivative given by

$$c'(x) = J(x) + \sum_{i=1}^{p}[1 + log\left(\langle v^i, x\rangle - \alpha_i\right)]v^i. \tag{3.31}$$

ii) c satisfies H1, H2, H3.a, H3.c and H4.

*iii) (*boundary coerciveness*) If $\{z^k\} \subset int(C)$ is such that*

(a) $z^k \xrightarrow[k\to\infty]{\mathrm{w}} \widetilde{z}$ and $\widetilde{z}$ belongs to the boundary of C, or

(b) $d(z^k, \partial C) \overset{k\to\infty}{\longrightarrow} 0$ (i.e. the distance from the terms of the sequence $\{z^k\}$ to the boundary of C goes to zero), with $\{z^k\}$ bounded,

then $\lim_{k\to\infty} \left\| c'(z^k) \right\|_* = +\infty$, $\lim_{k\to\infty}\langle c'(z^k), w - z^k\rangle = -\infty$ and $\lim_{k\to\infty} D_c(w, z^k) = \infty$ for all $w \in int(C)$.

Proof. Let $\zeta_i : B \to \mathbb{R}$ be given by $\zeta_i(x) = \langle v^i, x\rangle - \alpha_i$ for all $i = 1, ..., p$ and $\eta : \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$ the Bregman function in Example 2.6.4. Taking f_2 defined as, $f_2 = (1/2) \left\|\cdot\right\|^2$, we have $c = \eta \circ \zeta + f_2$, where $\zeta : B \to \mathbb{R}^p$ is the application with ζ_i as its i-th. component, $i = 1, ..., p$. Thus, the results presented here are consequences of Proposition 3.1.1(i) together with properties B1-B6 of the Bregman function η. In fact, strict convexity of c follows from the strict convexity of f_2 when B is strictly convex. Continuity is obvious and F-differentiability, as well as the expression for the G-derivative, are consequences of the sum and chain rules (Propositions 2.2.3 and 2.2.4 respectively) in view of the locally uniform convexity of B. Thus,

$$c'(x) = f_2' + \sum_{i=1}^{p} [1 + \log(\zeta_i(x))] v^i, \tag{3.32}$$

and (i) holds. Observe also that

$$D_c(y, x) = D_\eta(\zeta(y), \zeta(x)) + D_{f_2}(y, x), \tag{3.33}$$

so that H1 and H2 are direct consequences of properties H1 and H2 of f_2. Concerning H3.a, take $\{x^k\}$, $\{y^k\} \subset \mathrm{int}(C)$ such that $x^k \xrightarrow[k\to\infty]{\mathrm{w}} x^\infty$, $y^k \xrightarrow[k\to\infty]{\mathrm{w}} x^\infty$ and $\lim_{k\to\infty} D_c(y^k, x^k) = 0$.

Since c satisfies H2, $x^k - y^k \xrightarrow[k\to\infty]{\mathrm{s}} 0$, by Proposition 2.3.22. In view of the three-point equality, Proposition 2.3.10 (i)-(b), applied to f_2, we get

$$D_{f_2}(x^\infty, x^k) - D_{f_2}(x^\infty, y^k) = D_{f_2}(y^k, x^k) + \langle f_2'(x^k) - f_2'(y^k), y^k - x^\infty\rangle.$$

Hence,

$$\begin{aligned} \lim_{k\to\infty} |D_{f_2}(x^\infty, x^k) - D_{f_2}(x^\infty, y^k)| &\le \lim_{k\to\infty} D_{f_2}(y^k, x^k) + \lim_{k\to\infty} |\langle f_2'(x^k) - f_2'(y^k), y^k - x^\infty\rangle| \\ &\le \lim_{k\to\infty} \left\| f_2'(x^k) - f_2'(y^k) \right\|_* \left\| y^k - x^\infty \right\| = 0, \end{aligned} \tag{3.34}$$

using boundedness of the weakly convergent sequence $\{y^k\}$ and the fact that f_2 satisfies H3. Note also that $\zeta : B \to \mathbb{R}^p$ is weakly continuous, hence $\{\zeta(x^k)\}$ and $\{\zeta(y^k)\}$ are sequences in $\mathbb{R}^p$ both of which converge to $\zeta(x^\infty)$. Thus, in view of the fact that η satisfies B4, which has been proved in Lemma 2.1.3 of [20], it holds that

$$\lim_{k\to\infty} D_\eta(\zeta(x^\infty),\zeta(x^k)) = 0 \quad \text{and} \quad \lim_{k\to\infty} D_\eta(\zeta(x^\infty),\zeta(y^k)) = 0.$$

Consequently

$$\lim_{k\to\infty} \left[D_\eta(\zeta(x^\infty),\zeta(x^k)) - D_\eta(\zeta(x^\infty),\zeta(y^k))\right] = 0. \tag{3.35}$$

It follows from equations (3.34) and (3.35) that c satisfies H3.a.

H4 is a consequence of Proposition 2.5.2 applied to the operator $T = \partial[\eta\circ\zeta] : B \to \mathcal{P}(B^*)$, which is maximal monotone (see Proposition 2.3.1), and to the regularizing function f_2, which satisfies H4, after observing that int(dom f_2)$\cap$ dom $T \neq \emptyset$.

For (iii), observe first that in view of equation (3.33),

$$\begin{aligned} D_c(w,z^k) &= D_{f_2}(w,z^k) + \sum_{i=1}^p D_{\eta_i}(\zeta_i(w),\zeta_i(z^k)) \geq D_{\eta_i}(\zeta_i(w),\zeta_i(z^k)) \\ &= \zeta_i(w)\log\left(\frac{\zeta_i(w)}{\zeta_i(z^k)}\right) + \zeta_i(z^k) - \zeta_i(w) \qquad \forall i \in \{1,...,p\}. \end{aligned} \tag{3.36}$$

Take $\{z^k\} \subset \text{int}(C)$ satisfying (a) or (b), and assume that for some $w \in \text{int}(C)$ the sequence $\{D_c(w,z^k)\}$ does not diverge to infinity. Then, there exists a subsequence $\{z^{k_j}\}$ of $\{z^k\}$ and a real number M, such that $D_c(w,z^{k_j}) \leq M$ for all j. Since for any $x \in \text{int}(C)$, $\zeta_i(x) > 0$, in view of (3.36), it necessarily holds that

$$\liminf_j \zeta_i(z^{k_j}) > 0, \tag{3.37}$$

for all i in the finite set $\{1,...,p\}$.

Next we consider two cases: If (a) holds, take $z^k \xrightarrow[k\to\infty]{\text{w}} y \in \partial C$. Then $\zeta_i(y) = 0$ for some $i \in \{1,...,p\}$. Since ζ_i is weakly continuous, we get $\lim_k \zeta_i(z^k) = \zeta_i(y) = 0$, contradicting (3.37). If (b) holds, there exists a sequence $\{y^k\} \in \partial C$ satisfying $z^k - y^k \xrightarrow[k\to\infty]{\text{s}} 0$. Since for all k there exists $i = i(k) \in \{1,...,p\}$ such that $\zeta_{i(k)}(y^k) = 0$, we can find a subsequence $\{y^{k'}\}_{k'}$ of $\{y^{k_j}\}$ and a fixed index $i \in \{1,...,p\}$ satisfying $\zeta_i(y^{k'}) = 0$ for all k'. Thus,

$$\zeta(z^{k'}) = \zeta(z^{k'}) - \zeta_i(y^{k'}) = \langle v^i, z^{k'} - y^{k'}\rangle. \tag{3.38}$$

We conclude from (3.38) that $\lim_{k'} \zeta_i(z^{k'}) = 0$, in contradiction with (3.37), too. Hence,

$$\lim_{k\to\infty} D_c(w,z^k) = +\infty. \tag{3.39}$$

for all $w \in \text{int}(C)$

Observe now that boundedness of $\{z^k\}$ ensures boundedness of $\{c(x) - c(z^k)\}_k$. Thus, using the Definition 2.3.9 of D_c and (3.39), we get $\lim_{k\to\infty}\langle c'(z^k), x - z^k)\rangle = -\infty$.

Finally, the fact that $\left\|c'(z^k)\right\|_* \left\|x - z^k\right\| \geq -\langle c'(z^k), x-z^k)\rangle$ implies that $\lim_{k\to\infty} \left\|c'(z^k)\right\|_* = \infty$.

Concerning H3.c, let $A \subset \text{int}(C)$ be any bounded set such that $d(A, \partial C) > 0$. Then, there exists $\alpha > 0$ such that $\zeta_i(x) \geq \alpha$ for all $x \in A$ and all $i = 1, ..., p$. Choose $\{x^k\}, \{y^k\}_k \subset A$, satisfying $x^k - y^k \xrightarrow[k\to\infty]{s} 0$. Then, for all $i \in \{1, \ldots, p\}$,

$$\lim_{k\to\infty} [\zeta_i(x^k) - \zeta_i(y^k)] = \lim_{k\to\infty} \langle v^i, x^k - y^k\rangle = 0 \quad \text{and} \quad \lim_{k\to\infty} \frac{\zeta_i(x^k)}{\zeta_i(y^k)} = 1.$$

Since f_2 satisfies H3, we also have $f_2'(x^k) - f_2'(y^k) \xrightarrow[k\to\infty]{s} 0$. From (3.32), we get

$$\begin{aligned} \left\|c'(x^k) - c'(y^k)\right\|_* &= \left\| f_2'(x^k) - f_2'(y^k) + \sum_{i=1}^{p} [\log(\zeta_i(x^k)) - \log(\zeta_i(y^k))] v^i \right\|_* \\ &\leq \left\|f_2'(x^k) - f_2'(y^k)\right\|_* + \sum_{i=1}^{p} \left| \log\left(\frac{\zeta_i(x^k))}{\zeta_i(y^k)}\right)\right| \left\|v^i\right\|_* \xrightarrow{k} 0. \quad (3.40) \end{aligned}$$

It follows from (3.40) that c' is uniformly continuous on A. □

Remark 3.4.2. It is not necessary to use f_2 in the definition of c, i.e. in (3.30). In fact, we can use any power bigger than 1 of the norm, i.e. $\|\cdot\|^r$ $(r > 1)$, and still get the same results, since such powers of the norm share the required properties of f_2 (see Section 3.1).

Remark 3.4.3. Concerning H5, we already know from Example 2.6.6 that when B is a Hilbert space then c, as defined in (3.30), satisfies H5. The same holds when $B = \ell_p$ and f_2 is replaced by $f_p = \|\cdot\|_p^p$ in (3.30). In fact, in this case f_p' is weak-to-weak continuous (see Proposition 3.1.1(iii)), so that, choosing sequences $\{x^k\}$, $\{y^k\} \subset \text{int}(C)$ such that $x^k \xrightarrow[k\to\infty]{w} x$, $y^k \xrightarrow[k\to\infty]{w} x$ and $0 = \liminf_k |\langle c'(x^k) - c'(y^k), x - y\rangle|$, we get

$$\begin{aligned} 0 &= \liminf_k \left| \langle f_p'(x^k) - f_p'(y^k), x - y\rangle + \sum_{i=1}^{p} [\log(\zeta_i(x^k)) - \log(\zeta_i(y^k))][\zeta_i(x) - \zeta_i(y)] \right| \\ &= \langle f_p'(x) - f_p'(y), x - y\rangle + \sum_{i=1}^{p} [\log(\zeta_i(x)) - \log(\zeta_i(y))][\zeta_i(x) - \zeta_i(y)] \\ &\geq \langle f_p'(x) - f_p'(y), x - y\rangle. \end{aligned}$$

Thus $x = y$, because f_p is strictly convex.

In cases other than those in Remark 3.4.3, H5 should fail, in view of the counterexamples in subsection 3.2.

3.5 The separable case

We discuss next some properties of separable regularizing functions defined on the Cartesian product of two Banach spaces, which are also needed in the following chapters.

Proposition 3.5.1. [4] *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 $\tau_f(x) = \sup\{t \mid t \in \text{dom } \nu_f(x, \cdot)\}$ and $\tau_h(y) = \sup\{t \mid t \in \text{dom } \nu_h(y, \cdot)\}$. 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 $i = 1, \ldots, 4$, if both f and h satisfy Hi then F also satisfies Hi. Moreover, if both f' and h' are weak-to-weak continuous then F also satisfies $H5$.*

Proof. Take $\bar{z} = (\bar{x}, \bar{y}) \in \text{int}(\text{dom } F)$ and $t \in \mathbb{R}_+$. By definition of ν_F and D_F ((2.3.12) and (2.3.9)),

$$\begin{aligned} \nu_F(\bar{z}, t) &= \inf\{D_F(z, \bar{z}) \mid \|z - \bar{z}\|_{B_1 \times B_2} = t\} \\ &= \inf\{D_f(x, \bar{x}) + D_h(y, \bar{y}) \mid \|x - \bar{x}\|_{B_1} + \|y - \bar{y}\|_{B_2} = t\}. \end{aligned} \tag{3.41}$$

From Proposition 2.3.14(i) 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 (3.41) that

$$\begin{aligned} \nu_F(\bar{z}, t) &\geq \inf\{\nu_f(\bar{x}, \|x - \bar{x}\|_{B_1}) + \nu_h(\bar{y}, \|y - \bar{y}\|_{B_2}) \mid \|x - \bar{x}\|_{B_1} + \|y - \bar{y}\|_{B_2} = t\} \\ &= \inf\{\nu_f(\bar{x}, s) + \nu_h(\bar{y}, t - s) \mid s \in [0, t]\}. \end{aligned} \tag{3.42}$$

[4] We thank Professor B.F. Svaiter for the correct statements in (i)-(ii) of this proposition.

By (3.41), and for any $s \in \mathrm{dom}\, \nu_f(\bar{x}, \cdot)$ with $t - s \in \mathrm{dom}\, \nu_h(\bar{y}, \cdot)$, 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,

$$\nu_F(\bar{z}, t) \leq \inf\{\nu_f\left(\bar{x}, s\right) + \nu_h\left(\bar{y}, t - s\right) \mid s \in [0, t]\}. \tag{3.43}$$

Item (i) follows from (3.42) and (3.43).

For (ii) choose any $t_1 \in \mathrm{dom}\, \nu_f(\bar{x}, \cdot)$ and $t_2 \in \mathrm{dom}\, \nu_h(\bar{y}, \cdot)$ with $t_1 + t_2 = t$. Then $t_1 \geq t/2$ or $t_2 \geq t/2$. From Proposition 2.3.14(iv), 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,

$$\inf\{\nu_f\left(\bar{x}, t_1\right) + \nu_h\left(\bar{y}, t_2\right) \mid t_1 + t_2 = t\} \geq \min\{\nu_f(\bar{x}, t/2), \nu_h(\bar{y}, t/2)\}. \tag{3.44}$$

In view of (i) and (3.44) (ii) is true.

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

Chapter 4

Inexact versions of the Proximal Point method in Banach spaces

In this chapter $T : B \to \mathcal{P}(B^*)$ denotes a maximal monotone operator (see Definitions 2.1.1 and 2.1.2) and we are concerned with the problem of finding zeroes of T (see Definition 2.1.3). We start from the exact proximal point method proposed in [9] (see Section 2.5) and the hybrid methods proposed in [64, 62] (see Section 2.7). The methods in this chapter are not intended for penalty purposes, as those discussed in Section 2.6, whose extension to Banach spaces will be considered in Chapter 6. The novelty of this chapter lies in the extension to Banach spaces of the methods proposed for Hilbert or finite dimensional spaces in [64, 62], i.e. inexact proximal-like methods allowing for constant relative errors.

Our first approach will explore the interpretation, proposed in [64] for Hilbert spaces, of the proximal point method as a projection-type scheme, which we describe next.

It is not difficult to check that in the *exact* proximal point algorithm defined by (2.21), it holds that $x^{k+1} = \Pi_H^f(x)$, 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 Definition 2.3.19), of x^k onto H, which *separates* x^k from $T^{-1}(0)$. In fact x^{k+1} obviously belongs to H and $\langle f'(x^k) - f'(x^{k+1}), x - x^{k+1}\rangle \leq 0$ for all $x \in H$. Thus in view of proposition 2.3.20 (ii) $x^{k+1} = \Pi_H^f(x)$, provided that f is totally convex with dom $f = B$. Since $v^k = \lambda_k[f'(x^k) - f'(x^{k+1})] \in T(x^{k+1})$, monotonicity of T ensures that

$$\langle v^k, x^{k+1} - \bar{x}\rangle \geq 0, \quad \forall \bar{x} \in T^{-1}(0).$$

On the other hand, if x^k is not a zero of T (and thus $x^{k+1} \neq x^k$), strict convexity of f guarantees that

$$\langle f'(x^k) - f'(x^{k+1}), x^{k+1} - x^k\rangle < 0,$$

which means that the hyperplane H strictly separates the previous iterate x^k from the solution set of problem (2.1).

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 (2.1) 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 $f = B$, with $\mathcal{F}$ as in Definition 3.0.1 (see Chapter 2 for examples of such functions). 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

$$\lambda_k[f'(x^k) - f'(\tilde{x}^k)] - e^k \in T(\tilde{x}^k), \tag{4.1}$$

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

$$\|e^k\|_{B^*} \leq \sigma\lambda_k \begin{cases} D_f(\tilde{x}^k, x^k) & \text{if } \left\|x^k - \tilde{x}^k\right\| < 1 \\ \nu_f(x^k, 1) & \text{if } \left\|x^k - \tilde{x}^k\right\| \geq 1, \end{cases} \tag{4.2}$$

with D_f as in Definition 2.3.9 and ν_f as in Definition 2.3.12.

3. Let

$$v^k = \lambda_k[f'(x^k) - f'(\tilde{x}^k)] - e^k. \tag{4.3}$$

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

$$x^{k+1} = \Pi^f_{H_k}(x^k) = \arg\min_{x \in H_k} D_f(x, x^k), \tag{4.4}$$

with $\Pi^f_{H_k}$ as in Definition 2.3.19.

In order to compare our error criterion with the one in [64], i.e. (2.30)–(2.31), 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 criterion (4.2) takes the form $\|e^k\| \leq \sigma\lambda_k\frac{1}{2}\min\{\left\|x^k - \tilde{x}^k\right\|^2, 1\}$, a somewhat more restrictive condition than (2.31), because the right hand side of (2.31) involves the maximum between $\lambda_k^{-1}\|v^k\|$ and $\left\|x^k - \tilde{x}^k\right\|$. 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. As another improvement we get better convergence results for the case of Hilbert spaces, a fact discussed in section 4.3. We mention that we preserve the most important feature of (2.30)–(2.31), 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. [61]). Also, conditions on the distance from the iterates to the unknown exact solutions of the subproblems, as in [40], are avoided (see Section 2.7 for a deeper discussion of the algorithms in [61, 40]).

Concerning the projection step (4.4), 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 $f = B$ (see Proposition 2.3.20), 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 (2.18) and (2.19) in Example 2.4.4). 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 whether a vector e^k satisfies (4.2) or not, appear in Example 2.3.24.

In our second approach we extend to Banach spaces the error criterion introduced in [62] as an hybrid extragradient-inexact generalized proximal point method in $\mathbb{R}^n$ (see (2.34)). This algorithm requires an exogenous constant $\sigma \in [0, 1)$, an exogenous bounded sequence $\{\lambda_k\} \subset \mathbb{R}_{++}$ and an auxiliary function $f \in \mathcal{F}$, with $\mathcal{F}$ as in Definition 3.0.1, 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

$$e^k + \lambda_k[f'(x^k) - f'(\tilde{x}^k)] \in T(\tilde{x}^k), \tag{4.5}$$

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

$$D_f\big(\tilde{x}^k, (f')^{-1}\big[f'(\tilde{x}^k) - \lambda_k^{-1} e^k\big]\big) \leq \sigma D_f(\tilde{x}^k, x^k), \tag{4.6}$$

   with D_f as in Definition 2.3.9.

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

$$x^{k+1} = (f')^{-1}\big[f'(\tilde{x}^k) - \lambda_k^{-1} e^k\big]. \tag{4.7}$$

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 [9], means basically that we are not attempting any penalization effect 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 (4.6), (4.7), 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 4.0.2. *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.* (4.1)-(4.2)*) and Algorithm II (i.e.* (4.5)-(4.6)*) have solutions.*

Proof. Consider first the case of Algorithm I. Choose $\sigma = 0$, so that the right-hand side of (4.2) 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 2.5.2. Of course, the solution provided by this result satisfies (4.2) for any $\sigma \geq 0$, and so the result holds. The proof is analogous for Algorithm II. $\square$

Next we settle the issue of finite termination.

Theorem 4.0.3. *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* (2.1).

Proof. Algorithm I stops at the k-th iteration in two cases: if $v^k = 0$, in which case, by (4.1), $\tilde{x}^k$ is a solution of (2.1), or if $\tilde{x}^k = x^k$, in which case, by (4.2), $e^k = 0$, which in turn implies, by (4.3), $v^k = 0$ and we are back to the first case. Consequently, $\tilde{x}^k$ is a solution of (2.1). 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 (4.6), $e^k = 0$, which in turn implies, by (4.5), $0 \in T(\tilde{x}^k)$, so that $\tilde{x}^k$ is a solution of (2.1). $\square$

Proposition 4.0.2 ensures that exact solutions of (2.21) 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

$$\lambda_k f'(x^k) \in T(x) + \lambda_k f'(x) \tag{4.8}$$

with any feasible algorithm which generates a sequence strongly convergent to the unique solution of (4.8), a finite number of iterations of such algorithm will provide an appropriate next iterate both for Algorithm I and II.

Proposition 4.0.4. *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 any sequence generated either by Algorithm I or II. If x^k is not a solution of* (2.1), *then there exists a ball U_k around the exact solution of* (4.8) *such that any $x \in U_k \cap$ dom T solves* (4.1)–(4.2) *in the case of Algorithm I and* (4.5)–(4.6) *in the case of Algorithm II.*

Proof. Let $\bar{x}^k$ be the unique solution x of the inclusion (4.8). 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 $\psi_k : \text{dom } T \to \mathbb{R}$ defined as

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

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 $y \in \text{dom } T$ with $\left\|y - \bar{x}^k\right\| < \delta_k$. Let $U_k := \{y \in B \ : \left\|y - \bar{x}^k\right\| < \delta_k\}$. Then

$$\left\|T(y) + \lambda_k[f'(y) - f'(x^k)]\right\|_* \leq \sigma\lambda_k \min\{D_f(y, x^k), \nu_f(x^k, 1)\} \tag{4.9}$$

for all $y \in U_k \cap$ dom T. Finally, note that the right hand side of (4.9) is trivially less than or equal to the right hand side of (4.2) with y replacing $\tilde{x}^k$.

In the case of Algorithm II, proceed in the same way using Corollary 4.0.2 and, instead of ψ_k, the auxiliary function $\bar{\psi}_k : \text{dom } T \to \mathbb{R}$ defined as

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

□

4.1 Convergence analysis of Algorithm I

In this section 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 (2.1) is nonincreasing, which essentially entails global convergence of Algorithm I.

Lemma 4.1.1. *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 $z = \text{argmin}_{y \in H} D_f(y, x)$ 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 Proposition 2.3.20, exists and is unique when $H \subset$ dom $f = B$, 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 $z = \text{argmin}_{y \in H^-} D_f(y, x)$. 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 (see Example 2.3.3 and Definition 2.3.4). Consequently, since $\bar{x} \in H^-$, we have $\langle f'(x) - f'(z), z - \bar{x} \rangle \geq 0$. Therefore,

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

where the equality is an elementary identity, known as the *three-point property* (see (2.7) in Proposition 2.3.10). □

Lemma 4.1.2. *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*

$$\|e^k\|_* \left\|x^k - \tilde{x}^k\right\| \leq \sigma \lambda_k D_f(\tilde{x}^k, x^k) \leq \lambda_k D_f(\tilde{x}^k, x^k). \tag{4.10}$$

Proof. Let us consider two cases. First, if $\left\|x^k - \tilde{x}^k\right\| < 1$ then we have $\|e^k\|_* \left\|x^k - \tilde{x}^k\right\| \leq \|e^k\|_*$, so that the leftmost inequality in (4.10) follows trivially from (4.2). Second, if $\left\|x^k - \tilde{x}^k\right\| \geq 1$, then we use (2.9) and Definition 2.3.12 of ν_f, to get

$$\sigma \lambda_k D_f(\tilde{x}^k, x^k) \geq \sigma \lambda_k \nu_f(x^k, \left\|x^k - \tilde{x}^k\right\|) \geq \sigma \lambda_k \left\|x^k - \tilde{x}^k\right\| \nu_f(x^k, 1) \geq \left\|x^k - \tilde{x}^k\right\| \|e^k\|_*, \tag{4.11}$$

where the last inequality follows from (4.2). Thus, the first inequality of (4.10) is proved, and the second one holds because $\sigma \in [0, 1]$. □

Proposition 4.1.3. *Take $f \in \mathcal{F}$, satisfying H1 and H2 and such that dom $f = B$. Let $\{x^k\}$ be any sequence generated by Algorithm I. If problem (2.1) 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)* $\{x^k\}$ *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, (4.3) and (4.1), we have $\langle v^k, \bar{x} - \tilde{x}^k \rangle \leq 0$, so that $\bar{x} \in H_k^-$. Applying (4.3) and Definition 2.3.9 of D_f, we get

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

where the last inequality follows from the definition of the norm in B^*. Now, applying Lemma 4.1.2, 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 (4.4), we are able to apply Lemma 4.1.1 with $\tilde{y} = \tilde{x}^k \in B$, $x = x^k$, $v = v^k$ and $z = x^{k+1}$, obtaining

$$D_f(\bar{x}, x^k) \geq D_f(\bar{x}, x^{k+1}) + D_f(x^{k+1}, x^k) \geq D_f(\bar{x}, x^{k+1}). \tag{4.12}$$

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 (4.12),

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

establishing (iii). □

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

Theorem 4.1.4. *Take* $f \in \mathcal{F}$ *with* dom $f = B$ *satisfying H1–H4, and* $\{\lambda_k\} \subset (0, \bar{\lambda}]$. *Let* $\{x^k\}$ *be any sequence generated by Algorithm I. If problem* (2.1) *has solutions, then*

i) $\{x^k\}$ *has weak accumulation points and all of them are solutions of* (2.1).

ii) If f *also satisfies H5, then the whole sequence* $\{x^k\}$ *is weakly convergent to a solution of* (2.1).

Proof. By Proposition 4.1.3(ii), $\{x^k\}$ is bounded. Since B is reflexive, there exists at least one weak accumulation point. It follows from Proposition 4.1.3(iii) that $\lim_{k\to\infty} D_f(x^{k+1}, x^k) = 0$, and then, since f satisfies H2, we conclude from Proposition 2.3.22 that

$$x^{k+1} - x^k \xrightarrow[k\to\infty]{\text{s}} 0. \tag{4.13}$$

From (4.2) and Proposition 3.0.11(i) we get boundedness of the sequence $\{\lambda_k^{-1} e^k\}$, and therefore, using (4.13) and the fact that $0 \le |\langle \lambda_k^{-1} e^k, x^{k+1} - x^k\rangle| \le \|\lambda_k^{-1} e^k\|_* \left\|x^{k+1} - x^k\right\|$, we get

$$\langle \lambda_k^{-1} e^k, x^{k+1} - x^k\rangle \xrightarrow[k\to\infty]{\text{s}} 0. \tag{4.14}$$

Using the three-point property (i.e., the equality in (2.7)) and (4.3), we get

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

$$= \lambda_k^{-1}\langle v^k, \tilde{x}^k - x^{k+1}\rangle + \langle \lambda_k^{-1} e^k, \tilde{x}^k - x^{k+1}\rangle = \langle \lambda_k^{-1} e^k, \tilde{x}^k - x^{k+1}\rangle, \tag{4.15}$$

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

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

$$\ge \langle \lambda_k^{-1} e^k, x^k - x^{k+1}\rangle + \frac{1}{\lambda_k}\left[\lambda_k D_f(\tilde{x}^k, x^k) - \|e^k\|_* \left\|x^k - \tilde{x}^k\right\|\right] \ge \langle \lambda_k^{-1} e^k, x^k - x^{k+1}\rangle, \tag{4.16}$$

where the last inequality follows from Lemma 4.1.2. Taking limits as k goes to ∞ in the leftmost and rightmost expressions of (4.16), we obtain, using (4.14),

$$\lim_{k\to\infty} D_f(x^{k+1}, \tilde{x}^k) = 0, \tag{4.17}$$

and therefore, using H2, Proposition 2.3.22 and Remark 2.3.23,

$$x^{k+1} - \tilde{x}^k \xrightarrow[k\to\infty]{\text{s}} 0, \tag{4.18}$$

which, in view of (4.13), implies

$$x^k - \tilde{x}^k \xrightarrow[k\to\infty]{\text{s}} 0. \tag{4.19}$$

From (4.19) and H3, we get

$$f'(x^k) - f'(\tilde{x}^k) \xrightarrow[k\to\infty]{s} 0. \tag{4.20}$$

By (4.19), 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 criterion (4.2) implies that

$$\|e^k\|_* \leq \sigma \lambda_k D_f(\tilde{x}^k, x^k). \tag{4.21}$$

Next, we take limits as k goes to ∞ in the leftmost and rightmost expressions of (4.15). The rightmost one converges to 0 by (4.18) 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 4.1.3(iii), and the second term converges to 0 by (4.17). 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 (4.21) that $e^k \xrightarrow[k\to\infty]{s} 0$. It follows then from (4.3) and (4.20) that

$$v^k \xrightarrow[k\to\infty]{s} 0. \tag{4.22}$$

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 (4.19), $\tilde{x}^{j_k} \xrightharpoonup[k\to\infty]{w} \bar{x}$. It follows from (4.1) and (4.22) that $v^{j_k} \in T(\tilde{x}^{j_k})$ and $v^{j_k} \xrightarrow[k\to\infty]{s} 0$. Thus, $0 \in T(\bar{x})$, because the graph of a maximal monotone operator is demiclosed (see Proposition 2.1.6).

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 $x^{j_k} \xrightharpoonup[k\to\infty]{w} \bar{x}_1$ and $x^{i_k} \xrightharpoonup[k\to\infty]{w} \bar{x}_2$. By (i), $\bar{x}_1$ and $\bar{x}_2$ are solutions of (2.1). Thus, Proposition 4.2.2(i) guarantees the existence of $\xi_1, \xi_2 \in \mathbb{R}_+$ such that

$$\lim_{k\to\infty} D_f(\bar{x}_1, x^k) = \xi_1, \qquad \lim_{k\to\infty} D_f(\bar{x}_2, x^k) = \xi_2. \tag{4.23}$$

Now, using the four-point equality (2.8), we get

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

$$\leq \left|D_f(\bar{x}_1, x^{i_k}) - D_f(\bar{x}_1, x^{j_k})\right| + \left|D_f(\bar{x}_2, x^{i_k}) - D_f(\bar{x}_2, x^{j_k})\right|. \tag{4.24}$$

Taking limits as k goes to ∞ in the extreme expressions of (4.24) and using (4.23), we get that

$$\lim_{k\to\infty} |\langle f'(x^{i_k}) - f'(x^{j_k}), \bar{x}_1 - \bar{x}_2\rangle| = 0.$$

It follows from H5 that $\bar{x}_1 = \bar{x}_2$, establishing the uniqueness of the weak accumulation point of $\{x^k\}$.

□

Remark 4.1.5. In the case in which B is a Hilbert space, the error criterion in (4.2) becomes

$$\left\|e^k\right\| \leq \frac{1}{2}\sigma\lambda_k \min\{\left\|x^k - \tilde{x}^k\right\|^2, 1\}. \tag{4.25}$$

It could be argued that σ does not represent a truly relative error, because the norm of the error e^k in the left hand side of (4.25) is not squared, while the norm of the difference between the previous iterate x^k and the candidate $\tilde{x}^k$ in the right hand side is squared, so that σ is not "adimensional". This problem can be avoided by replacing the error criterion in (4.2) by:

$$\|e^k\|_* \left\|x^k - \tilde{x}^k\right\| \leq \sigma\lambda_k D_f(\tilde{x}^k, x^k), \tag{4.26}$$

$$\|e^k\|_* \leq \lambda_k \beta, \tag{4.27}$$

where β is an arbitrary positive constant, so that (4.27) is equivalent to requiring that $\{\lambda_k^{-1}e^k\}$ be bounded. Using Lemma 4.1.2 and Proposition 3.0.11 (ii), it can be easily checked that the proof of Theorem 4.1.4(i) remains valid when (4.2) is replaced by (4.26), (4.27), which, for the case of Hilbert spaces, give an error criterion closer to the one in [64], but we chose to use instead the more demanding error criterion in (4.2) for two reasons: the "dimensionality" objection becomes somewhat devoid of meaning in the case of nonhilbertian Banach spaces, which are our main object of interest in this thesis, and additionally (4.2) allows us to get superlinear convergence in some cases (see Section 4.3), which is not the case for (4.26), (4.27).

Remark 4.1.6 (Changing scales). Convergence results of this section can be easily extended to the case in which the error criteria in (4.2) is slightly modified as follows:

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

for some $\theta \geq 1$, thus rescaling the regions where the candidate point $\tilde{x}^k$ is being looking for.

Remark 4.1.7. It is worthwhile to mention that, in principle, the presence of $\sigma \in [0, 1]$ in Algorithm I could be considered superfluous, because if an error vector e^k satisfies (4.2) with $\sigma = 1$, then it satisfies it for all $\sigma \in [0, 1]$, so that it suffices to study the algorithm with $\sigma = 1$, or, equivalently, without σ in (4.2). From a computational viewpoint, on the other hand, the additional flexibility provided by σ might be useful. The idea is that the k-th subproblem will be solved with an algorithm guaranteed to converge to te exact solution of the subproblem, and that such inner loop will stop the first time that (4.2) is satisfied. In such a case, a value of σ smaller that 1 means that more inner steps will be taken, but also that the resulting $\tilde{x}^k$ will be closer to the exact solution, which might have an acceleration effect on the outer loop, meaning that resulting x^{k+1} could be closer to the solution set than the one obtained with $\sigma = 1$. We acknowledge that such an inner loop is not discussed in this work, but by explicitly including the relative error σ in our algorithm, we prepare the road for future consideration of such an inner loop, and for the correlated issue of good strategies for the choice of σ.

4.2 Convergence analysis of Algorithm II

From (4.5)–(4.7) with $v^k = e^k + \lambda_k[f'(x^k) - f'(\tilde{x}^k)]$, we obtain that sequences generated by Algorithm II satisfy

$$0 = \lambda_k^{-1} v^k + f'(x^{k+1}) - f'(x^k), \tag{4.28}$$

$$D_f\big(\tilde{x}^k, x^{k+1}\big) \leq \sigma D_f(\tilde{x}^k, x^k) \tag{4.29}$$

for all k. The following lemma essentially resumes the behavior of the error criterion in Algorithm II and entails its global convergence. This result was presented in Lemma 4.1 of [62] for the finite dimensional case, and its extension to Banach spaces is immediate.

Lemma 4.2.1. *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 the four-point equality (2.8), 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 (4.28) and the last inequality from (4.29). Using nonnegativity of D_f and the fact that $\sigma \in [0, 1)$, we get

$$D_f(\bar{x}, x^{k+1}) \leq D_f(\bar{x}, x^k) - \lambda_k^{-1}\langle v^k, \tilde{x}^k - \bar{x}\rangle, \tag{4.30}$$

and the result follows from (4.30) and monotonicity of T, since $v^k \in T(\tilde{x}^k)$. □

Proposition 4.2.2. *Let $f \in \mathcal{F}$, satisfying H1 and H4 and such that dom $T \subset$ int(dom f). If problem* (2.1) *has solutions, then*

i) $D_f(\bar{x}, x^k)$ *converges decreasingly, 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) $\tilde{x}^k - x^k \xrightarrow[k\to\infty]{s} 0$, *and consequently* $\{\tilde{x}^k\}$ *is bounded,*

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

Proof. Take $\bar{x} \in T^{-1}(0)$ (which is nonempty because problem (2.1) has solutions). Then, by Lemma 4.2.1, $\{D_f(\bar{x}, x^k)\}$ 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 4.2.1,

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

from which (iii) and (iv) follow easily. Item (v) follows from (iv) and (4.29). 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 2.3.22 to obtain $\tilde{x}^k - x^k \xrightarrow[k\to\infty]{s} 0$. In the same way $\tilde{x}^k - x^{k+1} \xrightarrow[k\to\infty]{s} 0$, implying that $x^k - x^{k+1} \xrightarrow[k\to\infty]{s} 0$. □

Proposition 4.2.2 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 4.2.3. *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* (2.1) *has solutions, then*

i) any sequence $\{x^k\}$ *generated by Algorithm II has weak accumulation points, all of which are solutions of* (2.1),

ii) if f also satisfies H5, then the whole sequence is weakly convergent to a solution of (2.1)[1].

[1]This is obviously true when T has only one zero, even if f does not satisfy H5.

Proof. By H3 and Proposition 4.2.2(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 (4.5), $v^k \in T(\tilde{x}^k)$ and then, using $\lambda_k \leq \bar{\lambda}$, we get from (4.28) that $v^k \xrightarrow[k\to\infty]{\text{s}} 0$. Thus, (i) follows from demiclosedness of the graph of maximal monotone operators (see Proposition 2.1.6). Concerning (ii), its proof follows exactly the argument in the proof of Theorem 4.1.4(ii).

□

4.3 Some results on the convergence rate of algorithm I

In this section we consider problem (2.1) for the case in which B is a Hilbert space H, and we look at Algorithm I with $f = \frac{1}{2}\|\cdot\|_H^2$, which satisfies all properties H1-H5 (see Proposition 3.1.1). Then the Bregman projection is exactly the metric projection (see (2.6) in Example 2.3.8 and Definition 2.3.19) and the method takes the form:

Algorithm I 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

$$\|e^k\| \leq \frac{1}{2}\sigma\lambda_k \min\{\|x^k - \tilde{x}^k\|^2, 1\}. \tag{4.31}$$

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 Algorithm I, working in Banach spaces, we adopted the error criterion (4.31), which is somewhat more restrictive that the one originally proposed by Solodov and Svaiter for Hilbert spaces in [64] (see section 2.7), namely

$$\|e^k\|_{B^*} \leq \sigma\lambda_k \max\{\|x^k - \tilde{x}^k\|, \lambda_k^{-1}\|v^k\|\}, \tag{4.32}$$

with $\sigma \in (0,1)$. Observe the square in the right hand side of (4.31), which is absent in (4.32), and which makes the right hand side of (4.31) much smaller than the one in (4.32) 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. 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 δ such that $\|y - \bar{x}\| \leq \theta \|v\|$ for all $y \in T^{-1}(v)$ and $\|v\| \leq \delta$. Under this hypothesis it was proved in [64] that any sequence $\{x^k\}$ generated by Algorithm I with $\|e^k\|_{B^*} \leq \sigma\lambda_k \max\{\|x^k - \tilde{x}^k\|, \lambda_k^{-1}\|v^k\|\}$ in place of (4.31) 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

$$\eta = \left[1 - \left(\frac{1-\sigma}{1+\sigma}\right)^4 \frac{1}{(\theta\bar{\lambda}(1-\sigma)+1)^2}\right]^{1/2}, \tag{4.33}$$

or, more precisely,

$$\|x^{k+1} - \bar{x}\| \leq \left[1 - \left(\frac{1-\sigma}{1+\sigma}\right)^4 \frac{1}{(\theta\lambda_k(1-\sigma)+1)^2}\right]^{1/2} \|x^k - \bar{x}\| \leq \eta \|x^k - \bar{x}\|, \tag{4.34}$$

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

Theorem 4.3.1. *Suppose that T^{-1} is Lipschitz continuous at zero with modulus θ. Then any sequence $\{x^k\}$ generated by Algorithm I, with $\lambda_k \leq \bar{\lambda}$, converges strongly to $\bar{x}$, the solution of* (2.1) *and the convergence is Q-linear with asymptotic error constant bounded by*

$$\eta_k = \left[1 - \left(\frac{1-\sigma/2}{1+\sigma/2}\right)^4 \frac{1}{(\theta\lambda_k(1-\sigma/2)+1)^2}\right]^{1/2}. \tag{4.35}$$

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

Proof. From the proof of Theorem 4.1.4(i) in section 4.1 we know that $\lim_{k\to\infty} v^k = 0$ and

$$\lim_{k\to\infty} \|x^k - \tilde{x}^k\| = 0. \tag{4.36}$$

Then, in view of (4.31), there exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$ it holds that

$$\|e^k\|_{B^*} \leq \frac{1}{2}\sigma\lambda_k\|x^k - \tilde{x}^k\|^2. \tag{4.37}$$

In view of (4.36) and (4.37), we have

$$\|e^k\|_{B^*} \leq \frac{1}{2}\sigma\lambda_k\|x^k - \tilde{x}^k\| \tag{4.38}$$

for all $k \geq k_0$. By (4.38), (4.32) holds and therefore we can use the results in [64], namely the inequality in (4.34), 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 statement of the theorem, and also

$$\|x^{k+1} - \bar{x}\| \leq \left[1 - \left(\frac{1-(\sigma/2)\|x^k - \tilde{x}^k\|}{1+(\sigma/2)\|x^k - \tilde{x}^k\|}\right)^4 \frac{1}{(L\lambda_k(1-(\sigma/2)\|x^k - \tilde{x}^k\|)+1)^2}\right]^{1/2} \|x^k - \bar{x}\|, \tag{4.39}$$

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 (4.39), the proof is complete. □

We point out that our η_k of (4.35) is strictly smaller than η in (4.33). In fact, if we define $\bar{\eta}$ with $\bar{\lambda}$ substituting for λ_k in (4.35), 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 4.3.1 are strongly dependent on the square in the right hand side of (4.31).

4.4 On the need for hybrid steps in Hybrid Proximal Point methods

The main objective of this section is to answer the following questions, concerning the hybrid methods I and II, and consequently the methods in [64, 62, 63]:

1) Is boundedness of the iterates guaranteed without the hybrid steps?

   Or, equivalently:

   Does the classical PPA converge under the "relative error criteria" of the hybrid methods?

2) Do the error measures, mainly the distances from the inexact solutions to the resolvent, i.e. $\|\tilde{z}^k - P_k(z^k)\|$, become summable "a posteriori"?

It should be observed that Solodov and Svaiter gave in [64] a negative answer to question (1), taking $T(x, y) = (x, -y)$ in $\mathbb{R}^2$. This operator T is not a subdifferential, so that question (1) remained open for the optimization case. We prove here that, even for the optimization case and in Hilbert spaces, the answer for both questions is negative in general. Our examples require a function g such that the sequence $\{x^k\}$ generated by the exact method (2.20) with $T = \partial g$ satisfies

$$\sum_{k=0}^{\infty} \left\|x^{k+1} - x^k\right\| = \infty. \tag{4.40}$$

A sufficient condition for (4.40) to hold is that the convergence of $\{x^k\}$ to its limit be weak but not strong, which can occur only in infinite dimensional spaces. Thus, we take $H = \ell_2$ in the examples. Of course, (4.40) could hold also for strongly convergent sequences generated by the exact method, but we have found no example of such a situation. So the issue remains open for the optimization case in finite dimensional spaces.

Next we present a full statement of both methods (I and II) for the optimization case (i.e. $T = \partial g$ for some proper, closed and convex $g : H \to \mathbb{R} \cup \{\infty\}$). Choose the regularizing function $f = (1/2) \|\cdot\|^2$ and error criterion in Remark 4.1.5 for Algorithm I. Thus instances of Algorithm I and II can be essentially given as follows:

1. Choose $z^0 \in H$.

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

$$\tilde{v}^k \in \partial g(\tilde{z}^k), \tag{4.41}$$
$$e^k = \tilde{v}^k + \lambda_k(\tilde{z}^k - z^k), \tag{4.42}$$
$$\|e^k\| \leq \sigma\lambda_k\|\tilde{z}^k - z^k\|. \tag{4.43}$$

3. Hybrid step:

$$z^{k+1} = z^k - \frac{\langle \tilde{v}^k, z^k - \tilde{z}^k \rangle}{\|\tilde{v}^k\|^2} \tilde{v}^k. \quad (Projection) \tag{4.44}$$

$$z^{k+1} = z^k - \lambda_k^{-1} \tilde{v}^k. \quad (Extragradient) \tag{4.45}$$

We state next our main result.

Theorem 4.4.1. *There exist a Hilbert space H and a proper, closed and convex function g in H which has minimizers and such that, given any bounded positive sequence $\{\lambda_k\}$, there exists a point $z \in$ dom g such that*

i) The inexact Proximal Point algorithm allowing for constant relative errors as in (4.41)-(4.43)*, without the hybrid steps, starting at z, may generate unbounded sequences of iterates.*

ii) The Hybrid Proximal Point methods allowing for constant relative error in (4.41)-(4.43) *and* (4.44) *or* (4.41)-(4.43) *and* (4.45)*, starting at z, accept inexact iterates with nonsummable distances to the resolvent.*

Proof. Let f be the function provided by Proposition 2.5.1 and $\{x^k\}$ the sequence generated by the exact Proximal Point Algorithm (2.20) starting at point x, also given by Proposition 2.5.1. Then the sequence $\{x^k\}$ converges weakly to some x^*, but does not converge strongly. Therefore,

$$\sum_{k=1}^{\infty} \|x^{k+1} - x^k\| = \infty, \tag{4.46}$$

because otherwise the sequence $\{x^k\}$ would converge strongly to x^*. Since $\{x^k\}$ is generated by Algorithm (2.20), there exists a sequence $\{u^k\}_{k\geq 1} \in \ell^2$ such that for all k:

$$0 = u^{k+1} + \lambda_k (x^{k+1} - x^k), \tag{4.47}$$

$$u^{k+1} \in \partial f(x^{k+1}). \tag{4.48}$$

Let us consider the product space $H = \ell^2 \times \mathbb{R}$ provided with the 2-norm $\|\cdot\|_2$ (i.e. $\|(x,t)\|_2 = (\|x\|_{\ell^2}^2 + |t|^2)^{1/2}$). Trivially, H is a Hilbert space. Define the function $g : H \to \mathbb{R} \cup \{\infty\}$ as

$$g(x,t) = f(x),$$

for all $(x,t) \in H$. Consequently g is a proper, closed and convex function with subdifferential at any (x,t) given by $\partial g(x,t) = (\partial f(x), 0)$. In particular, if S is the set of minimizers of f, then $S \times \mathbb{R}$ is the set of minimizers of g, hence nonempty.

Now we are ready to prove item (i). Choose any $t^0 \in \mathbb{R}$ and define the sequence $\{t^k\} \in \mathbb{R}$ as

$$t^{k+1} = t^k + \frac{\sigma}{\sqrt{1-\sigma^2}} \|x^{k+1} - x^k\| = t^0 + \frac{\sigma}{\sqrt{1-\sigma^2}} \sum_{n=0}^{k} \|x^{n+1} - x^n\|. \tag{4.49}$$

Define also, for $k = 0, 1, \cdots$

$$z^k = (x^k, t^k), \quad e^k = \lambda_k(0, t^{k+1} - t^k), \quad v^{k+1} = (u^{k+1}, 0). \tag{4.50}$$

Then

$$e^k = v^{k+1} + \lambda_k(z^{k+1} - z^k), \tag{4.51}$$

$$v^{k+1} \in \partial g(z^{k+1}), \tag{4.52}$$

where (4.51) follows from (4.47), (4.52) from (4.48) and the residual e^k satisfies

$$\begin{aligned} \|e^k\|_2 &= \lambda_k |t^{k+1} - t^k| = \lambda_k \frac{\sigma}{\sqrt{1-\sigma^2}} \|x^{k+1} - x^k\| \\ &= \lambda_k \sigma \left(\|x^{k+1} - x^k\|^2 + |t^{k+1} - t^k|^2 \right)^{1/2} = \lambda_k \sigma \|z^{k+1} - z^k\|, \end{aligned}$$

using (4.49). It follows that the pair (z^{k+1}, v^{k+1}) satisfies (4.41)-(4.43), and can be accepted as an inexact solution of the k-th proximal subproblem. Hence, the sequence $\{z^k\}$ is generated by the inexact Proximal Point Algorithm, applied to an optimization problem and using the relative error tolerance of the hybrid methods. Trivially, from (4.46), (4.49) and (4.50),

$$\lim_{k \to \infty} \|z^k\| = \infty,$$

which proves item (i).

For (ii) let us consider first the extragradient case, i.e. the method in (4.41)-(4.43) and (4.45). Through an induction argument, it can be proved that starting at $z^0 = (x, t^0)$ the inexact solutions can be taken as

$$\tilde{z}^k = (x^{k+1}, t^0 + \frac{\sigma}{\sqrt{1-\sigma^2}} \|x^{k+1} - x^k\|), \ \tilde{v}^k = (u^{k+1}, 0),$$

and the iterates as

$$z^k = (x^k, t^0).$$

Then $P_k(z^k) = z^{k+1}$ for all k and

$$\|\tilde{z}^k - P_k(z^k)\| = \|\tilde{z}^k - z^{k+1}\| = |\tilde{t}^k - t^0| = \frac{\sigma}{\sqrt{1-\sigma^2}}\|x^{k+1} - x^k\|,$$

which implies $\sum_{k=0}^{\infty} \|\tilde{z}^k - P_k(z^k)\| = \infty$. Note that z^{k+1} is also the projection of z^k onto $H_k = \{z \in H : \langle \tilde{v}^k, z - \tilde{z}^k \rangle = 0\}$, hence the same example works for the Hybrid Projection Proximal Algorithm. □

Remark 4.4.2. Observe that in the proof of item (ii), $\|e^k\| = \lambda_k |\tilde{t}^k - t^0|$ for all k. Consequently

$$\sum_{k=0}^{\infty} \|e^k\| \geq \underline{\lambda} \sum_{k=0}^{\infty} \|\tilde{z}^k - P_k(z^k)\| = \infty,$$

provided that $\lambda_k \geq \underline{\lambda} > 0$ for all k, so that the residuals are not summable either.

Remark 4.4.3. Though the example given by Theorem (4.4.1) does not prove nonsummability of the distances of the hybrid iterates to the resolvent, because $z^{k+1} = P_k(z^k)$ for all k, a slight modification of it (e.g. with $g(x,t) = f(x) + t^2/2$) provides an example in which the condition $\sum_{k=0}^{\infty} \|z^{k+1} - P_k(z^k)\| < \infty$ does not hold either.

4.5 Concluding remarks

In this section we summarize the main results of this chapter and present some issues which are left for future research.

Features:

- We get inexact versions of the Proximal Point Method for finding zeroes of maximal monotone operators in a context of Banach spaces.
- The extension is achieved with the use of regularizing functions with no more demanding assumptions that those used in the exact method in [9]. Moreover, we extended the family of known functions with the required properties.
- The inexact versions extend the Hybrid methods of Solodov and Svaiter ([64, 62]), keeping the advantages of the relative error criteria.
- The Hybrid Proximal-Bregman projection allows the use of regularizing functions other than the square of the norm.

- We exhibit a variant of the error criterion for the Hybrid Proximal-Bregman projection method in Hilbert spaces with the quadratic regularization, which has superlinear convergence, even with inexact solutions of the subproblems.
- We establish the necessity of the hybrid steps, in these hybrid methods, to get boundedness of the iterates for the optimization problem in Hilbert spaces.
- We prove that the relative error measures of these hybrid methods do not become summable "a posteriori".

Open research:

- Explore the enlargement of maximal monotone operators (see [63]) for Algorithm I. This should offer robustness of the method with less demanding assumptions (see Proposition 4.0.4).

Chapter 5

Augmented Lagrangians methods for cone-constrained convex optimization in Banach spaces

In this chapter 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. We consider functions $g : B_0 \to \mathbb{R}$ and $G_j : B_0 \to \mathrm{B}_j$ $(1 \leq j \leq m)$ satisfying assumptions (A1) and (A2) (see Section 2.8):

(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.

As discussed in Section 2.8, the general convex optimization problem (or primal problem) is

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

and it is studied with the help of its dual (D)

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

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 equation (2.36), i.e.

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

where $G : B_0 \to \Pi_{j=1}^{m} B_j$ is the application with components G_j, i.e. given by (2.37).

5.1 Fundamentals

Proposition 5.1.1. *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 Example 2.3.8 we get, for any $x \in B$,

$$\partial d(x, C) = \partial \|x - P_C(x)\|_B \cap N_C(P_C(x)). \tag{5.1}$$

By the chain rule of differentiation, we have

$$\partial h(x) = d(x, C)^{s-1} \partial d(x, C). \tag{5.2}$$

Thus, in view of (5.1) and (5.2), 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 the normalized duality map J satisfies $J = \partial(1/2 \|\cdot\|^2)$, we get

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

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

$$\|x - P_C(x)\|^{s-2} [J(x - P_C(x)) \cap N_C(P_C(x))] = \|x - P_C(x)\|^{s-2} J(x - P_C(x)), \qquad (5.3)$$

where we use Definition 2.3.4 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 (2.6) in the third equality. The last equality follows from Proposition 2.4.3. 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 (5.3) 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 (5.3) 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 [52] 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. Let $M : B_1 \to B_2$ be a K-convex mapping as defined in Definition 2.8.1.

Proposition 5.1.2. *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

$$\bar{y} + \alpha y + (1 - \alpha)y' \in -K \;\; for\ all\ \bar{y} \in -K. \qquad (5.4)$$

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

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

$$d(y_\alpha + [\alpha M(x) + (1-\alpha)M(x')], -K) \stackrel{def}{=} \inf_{\tilde{y} \in -K} \|y_\alpha + [\alpha M(x) + (1-\alpha)M(x')] - \tilde{y}\|_{B_2}. \tag{5.5}$$

Choosing $\bar{y} = y_\alpha$ we get, from (5.4)-(5.5),

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

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

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

For (ii), take $C = -K$ in Proposition 5.1.1 and use the properties of B_2 and the chain rule of differentiation (Proposition 2.2.4) to get Gâteaux differentiability (respectively Fréchet differentiability) of V and the formula of V'. In fact

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

□

5.2 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

$$M_{r_j}(x, y_j, \rho) = h'_{r_j}(y_j) + \rho^{-1}G_j(x). \tag{5.6}$$

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

$$\bar{L}(x, y, \rho) = g(x) + \rho \sum_{j=1}^{m} \frac{1}{s_j} d(M_{r_j}(x, y_j, \rho), -K_j)^{s_j}, \tag{5.7}$$

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.

$$M_r(x, y, \rho) = \big(M_{r_1}(x, y_1, \rho), ..., M_{r_m}(x, y_m, \rho)\big). \tag{5.8}$$

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. [60]). When the B_j's are Hilbertian and $r_j = 2$ $(1 \leq j \leq m)$, L reduces to the augmented Lagrangian in [66] (see Section 2.9 for a deeper discussion on such cases). We establish next some elementary properties of $\bar{L}$.

Proposition 5.2.1. *Take* $\bar{L}$ *as in* (5.7). *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 norm-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*

$$Q_j(x, y, \rho) = J_{s_j}\left(M_{r_j}(x, y_j, \rho) - P_{-K_j}(M_{r_j}(x, y_j, \rho)\right) \quad (1 \leq j \leq m). \qquad (5.9)$$

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 (5.6). Then (A1) ensures K_j convexity of M and, in view of Proposition 5.1.2(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 5.1.2(ii) and the sum rule for Fréchet derivatives (Proposition 2.2.3). Norm-to-norm continuity of $\bar{L}'_x(\cdot, y, \rho)$ is, then, a consequence of (i) and Proposition 2.3.10(ii). Item (ii)-(b) follows from (ii)-(a) and Proposition 2.8.4(i)-(a). □

Proposition 5.2.1(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:

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

$$y^{k+1} = Q(x^{k+1}, y^k, \lambda_k).$$

This is an extension of (2.55)-(2.56), but as discussed in Section 2.9, existence of primal iterates is not guaranteed. Therefore, following the approach in [60], [25] (see Section 2.9), 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'. Each of our convergence results requires some of the assumptions H1–H5 on f (see Section 2.5 for definition of such assumptions and Section 3.1 for examples of functions satisfying such properties).

With the help of f, and of D_f as given by Definition 2.3.9, we define the *doubly augmented Lagrangian* $\hat{L} : B_0 \times B_0 \times B^* \times \mathbb{R}_{++} \to \mathbb{R}$ as

$$\hat{L}(x, z, y, \rho) = \bar{L}(x, y, \rho) + \rho D_f(x, z), \tag{5.10}$$

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

$$x^{k+1} = \arg\min_{x \in B_0} \hat{L}(x, x^k, y^k, \lambda_k) = \arg\min_{x \in B_0} \left[\bar{L}(x, y^k, \lambda_k) + \lambda_k D_f(x, x^k)\right]. \tag{5.11}$$

3. Define y^{k+1} as

$$y_j^{k+1} = J_{s_j}\left(M_{r_j}(x^{k+1}, y_j^k, \lambda_k) - P_{-K_j}(M_{r_j}(x^{k+1}, y_j^k, \lambda_k))\right) \quad (1 \le j \le m). \tag{5.12}$$

We mention that in the two cases discussed after (5.8), algorithm (5.11)–(5.12) reduces to the exact versions of the augmented Lagrangian methods analyzed in [60] and [66] (see Section 2.9).

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 5.2.2. *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* (5.11).

Proof. If f satisfies H4, take $T = \bar{L}'_x(\cdot, y^k, \lambda_k)$ and use [9], Corollary 3.1. Otherwise, the result follows from Proposition 3.15 in [16]. □

5.3 Inexact versions of the doubly augmented Lagrangian method

We consider the regularizing function $F : B_0 \times B^* \to \mathbb{R}$ defined as

$$F(x, y) = f(x) + \sum_{j=1}^{m} h_{r_j}(y_j), \tag{5.13}$$

where $h_{r_j}(y_j) = \frac{1}{r_j} \|y_j\|_{B_j^*}^{r_j}$, $(1 < r_j < \infty)$, $(1 \le j \le m)$, 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 (5.9) and M_{r_j} as in (5.6).

Algorithm DAL-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

$$\|\hat{L}'_x(\tilde{x}^k, x^k, y^k, \lambda_k)\|_{B_0^*} \le \sigma\lambda_k \begin{cases} D_F(\tilde{z}^k, z^k) & \text{if } \left\|z^k - \tilde{z}^k\right\|_{B_0 \times B^*} < 1, \\ \nu_F(z^k, 1) & \text{if } \left\|z^k - \tilde{z}^k\right\|_{B_0 \times B^*} \ge 1, \end{cases} \tag{5.14}$$

   with D_F and ν_F as in Definitions 2.3.9, 2.3.12 respectively, where

$$\tilde{z}^k = \left(\tilde{x}^k, Q(\tilde{x}^k, y^k, \lambda_k)\right). \tag{5.15}$$

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

$$z^{k+1} = (x^{k+1}, y^{k+1}) = \operatorname{argmin}_{z \in H_k} D_F(z, z^k). \tag{5.16}$$

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 discussion of Algorithm I in Chapter 4 for more details).

As a result of the Bregman projection in (5.16), we can get a pair (x^k, y^k) with infeasible y^k (i.e. $y^k \notin K^*$) We present next a dual feasible alternative.

Algorithm DAL-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

$$D_f\big(\tilde{x}^k, (f')^{-1}\big[f'(x^k) - \lambda_k^{-1}\bar{L}'_x(\tilde{x}^k, y^k, \lambda_k)\big]\big) \leq \sigma \left[D_f(\tilde{x}^k, x^k) + \sum_{j=1}^{m} D_{h_{r_j}}\big(Q_j(\tilde{x}^k, y^k, \lambda_k), y_j^k\big)\right] \tag{5.17}$$

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

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

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

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 Chapter 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 (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 2.4.3, $J\left(M_r(x, y, \rho) - \bar{z}\right)$ 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,

$$\langle \bar{y}, z - \bar{z} \rangle \leq 0, \tag{5.18}$$

for all $z \in -K$. Applying (5.18) 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 (5.18),

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

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 5.3.1 corroborates dual feasibility of Algorithm DAL-II, which also has, as an advantage over Algorithm DAL-I, a closed formula to update primal and dual variables. As a disadvantage, we observe that the error criterion in Algorithm DAL-II implicitly requires repeated extragradient steps in order to find an approximate zero of $\hat{L}(\cdot, x^k, y^k, \lambda_k)$, while in Algorithm DAL-I this task is performed just once, at the projection step. We remark that for $\sigma = 0$ both Algorithm DAL-I and DAL-II reduce to the doubly augmented Lagrangian method of (5.11)-(5.12). This fact has two consequences: 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 (5.11)–(5.12), established in Proposition 5.2.2, implies existence of all iterates both for Algorithm DAL-I and DAL-II.

5.4 Convergence analysis of Algorithms DAL-I and DAL-II

The convergence properties of Algorithms DAL-I and DAL-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 (see Definition 2.41). See Chapter 4 for a complete study of such methods, called Algorithms I and II.

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

Proposition 5.4.1. *Take F, $\{v^k\}$, $\{z^k\}$, $\{\tilde{z}^k\}$, $\{\lambda_k\}$ and σ as in Algorithm DAL-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*

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

Proof. Let $v^k = (u^k, w^k) \in B^* \times B$. Take Q as in (5.9). 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 5.2.1(ii)-(b) and the last one from (5.15), 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 5.3.1(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

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

if and only if

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

if and only if, using (5.10) and Lemma 5.3.1(a),

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

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 (5.14). □

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

Proposition 5.4.2. *Take* $\{z^k\}$, $\{\tilde{x}^k\}$, $\{\lambda_k\}$, σ *and* F *as in Algorithm DAL-II. Define* $\tilde{z}^k = (\tilde{x}^k, Q(\tilde{x}^k, y^k, \lambda_k))$ *and* $e^k = \left(\hat{L}'_x(\tilde{x}^k, x^k, y^k, \lambda_k), 0\right)$ *, where* Q *is defined as in* (5.9)*. 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 (5.10) and Lemma 5.3.1(a)–(c) we have

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

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

establishing (i). Also,

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

$$D_F\big((\tilde{x}^k, Q(\tilde{x}^k, y^k, \lambda_k)), ((f')^{-1}\big[f'(\tilde{x}^k) - \lambda_k^{-1}\hat{L}'_x(\tilde{x}^k, x^k, y^k, \lambda_k)\big], Q(\tilde{x}^k, y^k, \lambda_k))\big) =$$

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

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

$$\sigma D_F(\tilde{z}^k, z^k),$$

using the error criterion given by (5.17) in the inequality, and the separability of F, so that (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

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

if and only if

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

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

Propositions 5.4.1–5.4.2 show that Algorithms DAL-I, DAL-II inherit the properties of Algorithms I, II. We get therefore the following results for Algorithms DAL-I and DAL-II. The first one deals with the case of finite termination.

Theorem 5.4.3. *Suppose that Algorithm DAL-I (respectively Algorithm DAL-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* (5.9).

Proof. Immediate from Proposition 5.4.1 (respectively 5.4.2), Theorem 4.0.3 and Proposition 2.8.8(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 5.4.4. *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 DAL-I for conclusion (a) and by Algorithm DAL-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 (5.14)-(5.15).

b) If $(f')^{-1}$ is continuous then there exists an open subset $U_k \subset B$ such that any $x \in U_k$ solves (5.17).

Proof. Let $\bar{x}^k$ denote the exact solution of (5.11) whose existence is ensured by Proposition 5.2.2, and $\bar{z}^k = (\bar{x}^k, Q(\bar{x}^k, y^k, \lambda_k))$, where Q is as in (5.9). Then $\bar{z}^k \neq z^k$, because otherwise, by Proposition 5.4.1, $0 \in T_L(z^k)$, 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 Propositions 3.5.1(ii) and 3.1.1(i), $\alpha_k := \sigma\lambda_k \min\{D_F(\bar{z}^k, z^k), \nu_F(z^k, 1)\} > 0$. Proposition 5.1.1 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 $x \in U_k := \{x \in B \; : \left\|x - \bar{x}^k\right\| < \delta_k\}$. The result in (a) follows then from the inequality above. Item (b) is proved with a similar argument, using now Proposition 5.4.2 and, instead of ψ_k, the auxiliary function $\bar{\psi}_k : B \to \mathbb{R}$ defined as

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

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

□

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

Theorem 5.4.5. *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 DAL-I (respectively Algorithm DAL-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' is weak-to-weak continuous and either

a) $B_j = \mathcal{L}^{p_j}(\Omega)$, Ω 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 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 5.4.1 (respectively 5.4.2) 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 3.5.1(iii) and 3.1.1, F satisfies H1–H4 and also satisfies H5 under conditions (a) or (b) of (ii). By our assumption on existence of KKT-pairs, Proposition 2.8.8(i) and (iii), T_L is a maximal monotone operator with zeroes. The result follows then from Theorem 4.1.4 (respectively Theorem 4.2.3 for Algorithm DAL-II) and Proposition 2.8.8(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 ($1 \le j \le m$) 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.

5.5 Concluding remarks

We summarize in this section what we identify as the main features of our results in this chapter, and present some suggestions for further research.

Features:

- We present two inexact augmented Lagrangian methods, allowing for constant relative error, for cone-constrained convex optimization.
- The results are entirely new when the constrains act on nonhilbertian Banach spaces.

We comment next on our convergence results as compared to previous works. For the case of hilbertian B_j's ($1 \le j \le m$), our convergence results are much stronger than those in [66].

- Convergence results for all algorithms in [66] require either that $\lim_{k\to\infty} \lambda_k = 0$ or at least that λ_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 [66] 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.

- For a general cone, our methods are original and valuable even in finite dimension.

- In finite dimension, when the cone is the nonnegative orthant and the primal regularizing function is the square of the norm, we get back essentially the methods in [33]. Thus, we have extended the work of this reference to more general cones and regularizing functions.

We also improve over the results presented in [16], which are restricted to the particular cone $K = \mathcal{L}^p_+(\Omega) = \{z \in \mathcal{L}^p(\Omega)|\ z(\omega) \geq 0\ \mu \text{ a.e.}\}$, where $(\Omega, \mathcal{A}, \mu)$ is a measure space and $p \in (1, \infty)$. In this reference the constraints act on $\mathcal{L}^p(\Omega)$ and not much is said about the primal sequence, because the method lacks primal regularization.

Open research:

- Study the use of regularizing functions for the dual variables other than the powers of the norm.

- Explore specific applications, e.g. related to the cone of the positive semidefinite matrices in an n-dimensional Euclidean space.

Chapter 6

Proximal Point methods with in-built penalization in Banach spaces

In this chapter $T : B \to \mathcal{P}(B^*)$ denotes a maximal monotone operator (see Definitions 2.1.1 and 2.1.2) and C is a convex and closed subset of the reflexive real Banach space B. Suppose that int$(C)\cap$ dom$(T) \neq \emptyset$ and consider the variational inequality problem VIP(T,C), as in Definition 2.3.6. Let S denote the solution set of this problem, i.e. the set of zeroes of the maximal monotone operator $T + N_C$. In Section 2.6 we have discussed proximal point methods for finite dimensional spaces in which the domain of the derivative of the regularizing function f is the interior of C, and a boundary coerciveness condition is satisfied, in which case the sequence generated by the method remains in the interior of C, making the proximal subproblems unconstrained, in the sense that the normalizing cone N_C (and consequently the feasible set C) need not be explicitly considered at each iteration, because the information on C is embedded in f. The main objective of this chapter is to extend this approach to Banach spaces, and specifically to the inexact variant of the proximal method presented as Algorithm II in Chapter 4. The material of this chapter is organized as follows: in Section 6.1 we introduce a boundary coerciveness condition, which we call H6, and must be satisfied by the regularizing function f in order to achieve the penalization effect. In Section 6.2 we discuss good definition of Algorithm II of Chapter 4, when the regularizing function f is not defined in the whole space and it satisfies the new assumption H6. Convergence properties of this algorithm are analyzed in Section 6.3. The main difference with respect to the convergence analysis in Chapter 4 is the following: when the domain of f' is the interior of C, $\|f'(x)\|$ diverges when x approaches the boundary of C, in which case condition H3 in Section 2.5 cannot be expected to hold. We mentioned in that section that the alternative weaker condition H3.a was introduced in [9], and our convergence analysis in this chapter uses this condition instead of H3. In Section 6.4 we consider examples of convex sets C in

Banach spaces for which there exist functions satisfying the assumptions in our convergence results. When C is a polyhedron with nonempty interior, the function f proposed in Section 3.4 does satisfy such conditions. On the other hand, the function discussed in Section 3.3 for the case in which C is a closed ball, does not satisfy H3.a. However, the effect for which H3.a is intended can be achieved using this f, but with additional conditions imposed upon the regularizing coefficients: they must converge to zero in a predetermined manner. In Section 6.5 we establish the convergence properties of the algorithm for the case of the ball with the function introduced in Section 3.3 and the additional condition on the regularizing parameters.

6.1 A boundary coerciveness condition

Let f denote a function in $\mathcal{F}$, with dom $f = C$ and satisfying condition H6 below, which can be seen as a boundary coerciveness condition in the sense of Section 2.6.

H6: If $\{z^k\} \subset \mathrm{int}(C)$ is bounded and $\lim_{k\to\infty} d(z^k, \partial C) = 0$ (i.e. the distance from the terms of the sequence $\{z^k\}$ to the boundary of C goes to zero) then $\lim_{k\to\infty} D_f(w, z^k) = +\infty$ for all $w \in \mathrm{int}(C)$.

The following lemma illustrates the way in which H6 will be used in our analysis.

Lemma 6.1.1. *Take $f \in \mathcal{F}$, with dom $f = C$.*

i) Assume that $x \in int(C)$ and that $\{z^k\} \subset int(C)$ is a bounded sequence. If $\lim_k D_f(x, z^k) = \infty$ then $\lim_{k\to\infty} \left\|f'(z^k)\right\|_* = \infty$ and $\lim_{k\to\infty}\langle f'(z^k), x - z^k\rangle = -\infty$.

ii) If f satisfies H6 then dom f' =int(C).

Proof. In view of Definition 2.3.9,

$$\begin{aligned} D_f(x, z^k) &\leq D_f(x, z^k) + D_f(z^k, x) = \langle f'(z^k) - f'(x), z^k - x\rangle \\ &= \langle f'(z^k), z^k - x\rangle - \langle f'(x), z^k - x\rangle \leq \left\|f'(z^k) - f'(x)\right\|_* \left\|z^k - x\right\|. \end{aligned}$$

Since both $\{z^k - x\}$ and $\{\langle f'(x), z^k - x\rangle\}$ are bounded, it holds that $\lim_{k\to\infty}\langle f'(z^k), z^k - x\rangle = \infty$ and $\lim_{k\to\infty} \left\|f'(z^k) - f'(x)\right\|_* = \infty$, when the hypothesis in (i) is satisfied.

Assume now that f satisfies H6. We claim that dom f' =int(C). Indeed, for any $z \in \partial C$ there exists a sequence $\{z^k\} \subset \mathrm{int}(C)$ converging strongly to z Hence, $z \notin$ dom f', because otherwise the norm-to-weak continuity of f' :dom $f' \to B^*$, which results from Proposition 2.3.2, together with (i), imply that for any $x \in \mathrm{int}(C)$ it holds that $-\infty = \lim_k\langle f'(z^k), x - z^k\rangle = \langle f'(z), x - z\rangle$, which is a contradiction, establishing the claim.

Remark 6.1.2. By Lemma 6.1.1, the G-derivative of any function f satisfying H6 "diverges" at the boundary of C: for any sequence on the interior of C, which approaches the boundary of C in the sense of the norm, it holds that the norm of the G-derivative of f goes to infinity. It is this behavior of f' near the boundary of C, when f satisfies H6, which makes the proximal subproblems unconstrained.

6.2 Inexact Proximal Point-Extragradient method with in-built penalization in Banach spaces.

In this section we rewrite Algorithm II of Chapter 4 for the variational inequality problem VIP(T,C). This algorithm requires an exogenous constant $\sigma \in [0, 1)$, an exogenous bounded sequence $\{\lambda_k\} \subset \mathbb{R}_{++}$ and an auxiliary function $f \in \mathcal{F}$, with $\mathcal{F}$ as in Definition 3.0.1, such that dom $f = C$ and f satisfies H4 and H6. It is defined as follows:

Algorithm II: Inexact Proximal Point-Extragradient Method

1. Choose $x^0 \in \mathrm{int}(C)$.

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

$$e^k + \lambda_k[f'(x^k) - f'(\tilde{x}^k)] \in T(\tilde{x}^k), \tag{6.1}$$

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

$$D_f\big(\tilde{x}^k, (f')^{-1}\big[f'(\tilde{x}^k) - \lambda_k^{-1}e^k\big]\big) \leq \sigma D_f(\tilde{x}^k, x^k), \tag{6.2}$$

   with D_f as in Definition 2.3.9.

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

$$x^{k+1} = (f')^{-1}\big[f'(\tilde{x}^k) - \lambda_k^{-1}e^k\big]. \tag{6.3}$$

Proposition 6.2.1. *Let $f \in \mathcal{F}$ be a function with dom $f = C$ and such that f satisfies H6 and H4. Then*

i) Algorithm II above is well defined: for all k inclusion (6.1) *has at least one solution $\tilde{x}^k$ satisfying* (6.2)*. Moreover, any such solution $\tilde{x}^k$ belongs to int(C), the next iterate x^{k+1} is uniquely determined by* (6.3) *(unless the algorithm stops at iteration k), and it belongs to int(C).*

ii) For each k, $v^k = e^k + \lambda_k[f'(x^k) - f'(\tilde{x}^k)]$ satisfies

(a) $v^k \in T(\tilde{x}^k) = [T + N_C](\tilde{x}^k)$,

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

(c) $D_f\big(\tilde{x}^k, x^{k+1}\big) \leq \sigma D_f(\tilde{x}^k, x^k)$.

Proof. We proceed by induction, noting that $x^0 \in \text{int}(C)$ by assumption. Indeed, assume that $x^k \in \text{int}(C)$. Take $e^k = 0 \in B^*$, in which case equations (6.1) and (6.2) take the form:

$$0 \in T(\tilde{x}^k) + \lambda_k[f'(\tilde{x}^k) - f'(x^k)] \tag{6.4}$$

$$0 = D_f\big(\tilde{x}^k, \tilde{x}^k\big) \leq \sigma D_f(\tilde{x}^k, x^k). \tag{6.5}$$

Proposition 2.5.2 establishes that inclusion (6.4) has a (unique) solution $\tilde{x}^k$, while (6.5) trivially holds for any $\sigma \geq 0$. In view of Lemma 6.1.1(ii) and H6, we get dom f' =int(C), and consequently $\tilde{x}^k \in \text{int}(C)$ and f' is strictly monotone on int(C), because of strict convexity of f. Thus, we get from H4 that the next iterate x^{k+1} is uniquely determined by (6.3) and it belongs to the interior of C.

For (ii), take $v^k = e^k + \lambda_k[f'(x^k) - f'(\tilde{x}^k)]$. Then $v^k \in T(\tilde{x}^k)$ by (6.1). Since $\tilde{x}^k \in \text{int}(C)$ by item (i), $N_C(\tilde{x}^k) = 0$ (see Definition 2.3.4), and hence (a) holds. (6.3) together with (6.2) imply (c). By (6.3) we get $f'(x^{k+1}) = f'(\tilde{x}^k) - \lambda_k^{-1} e^k$, which ensures (b), in view of the definition of v^k. □

6.3 Convergence analysis

Definition 6.3.1. Let $\mathcal{F}_C$ be the set of functions f in $\mathcal{F}$ satisfying H4 and H6 and such that dom $f = C$. We recall that $\mathcal{F}$ is defined in Definition 3.0.1.

We mention that the set $\mathcal{F}_C$ consists essentially of those functions which are needed for the good definition of Algorithm II with penalization effect on the feasible set C (see Proposition 6.2.1). We present next the fundamental global properties of Algorithm II with the penalization effect.

Proposition 6.3.2. *Let $\{x^k\}$, $\{\tilde{x}^k\}$ be the sequences generated by Algorithm II with $f \in \mathcal{F}_C$. Assume that f satisfies H1. If VIP(T,C) has solutions, then*

i) $D_f(\bar{x}, x^k)$ converges decreasingly, for all $\bar{x} \in S = [T + N_C]^{-1}(0)$,

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

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

Proof. Take any $\bar{x} \in S = (T + N_C)^{-1}(0)$. Let $v^k = e^k = \lambda_k[f'(x^k) - f'(\tilde{x}^k)]$. Then, by the four-point equality (2.8), we get

$$\begin{aligned} D_f(\bar{x}, x^{k+1}) &= D_f(\bar{x}, x^k) + \langle f'(x^k) - f'(x^{k+1}), \bar{x} - \tilde{x}^k\rangle + D_f(\tilde{x}^k, x^{k+1}) - D_f(\tilde{x}^k, x^k) \\ &= D_f(\bar{x}, x^k) + \langle \lambda_k^{-1} v^k, \bar{x} - \tilde{x}^k\rangle + D_f(\tilde{x}^k, x^{k+1}) - D_f(\tilde{x}^k, x^k) \\ &\leq D_f(\bar{x}, x^k) - \lambda_k^{-1}\langle v^k, \tilde{x}^k - \bar{x}\rangle + (\sigma - 1)D_f(\tilde{x}^k, x^k) \\ &\leq D_f(\bar{x}, x^k)x\rangle - (1-\sigma)D_f(\tilde{x}^k, x^k), \end{aligned} \tag{6.6}$$

where the second equality follows from Proposition 6.2.1(ii)-(b), the first inequality from Proposition 6.2.1(ii)-(c) and the last one from Proposition 6.2.1(ii)-(a), monotonicity of $T + N_C$, and the fact that $0 \in (T + N_C)(\bar{x})$.

Using nonnegativity of D_f and the fact that $\sigma \in [0, 1)$, we get from (6.6) that $\{D_f(\bar{x}, x^k)\}$ 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 (6.6),

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

from which (iii) and (iv) follow easily. Item (v) follows from (iv) and Proposition 6.2.1(ii)-(c). 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 2.3.22 to obtain $\tilde{x}^k - x^k \xrightarrow[k\to\infty]{s} 0$. In the same way $\tilde{x}^k - x^{k+1} \xrightarrow[k\to\infty]{s} 0$, implying that $x^k - x^{k+1} \xrightarrow[k\to\infty]{s} 0$. □

Next we settle the issue of finite termination.

Theorem 6.3.3. *Suppose that Algorithm II stops after k steps. Then the point* $x^k \in B$ *generated by Algorithm II is a solution of VIP(T,C).*

Proof. Algorithm I stops at the k-th iteration if $\tilde{x}^k = x^k$, in which case $D_f(\tilde{x}^k, x^k) = 0$, and therefore, by (6.2), $e^k = 0$, which in turn implies, by (6.1), $0 \in T(\tilde{x}^k)$. Consequently Proposition 6.2.1(ii)-(a) ensures that $0 \in T(\tilde{x}^k) + N_C(\tilde{x}^k)$, i.e. $\tilde{x}^k$ is a zero of $T + N_C$. By Proposition 2.3.7, we conclude that $\tilde{x}^k$ is a solution of VIP(T,C).

Theorem 6.3.4. *Take $f \in \mathcal{F}_C$, satisfying H1 and H2. Let $\{x^k\}$ be the sequence generated by Algorithm II and $\{\lambda_k\} \subset (0, \bar{\lambda}]$. If $S \neq \emptyset$ (i.e. VIP(T,C) has solutions), then*

i) If any of the following conditions holds

(a) $T = \partial g$, where g is a lower semicontinuous convex function (i.e. the optimization case), or

(b) T is pseudo- and paramonotone and f satisfies H3.a,

then $\{x^k\}$ has weak accumulation points and all of them are solutions of VIP(T,C). Moreover in case (a), $\{\tilde{x}^k\}$ is a minimizing sequence *(i.e. $\lim_k g(\tilde{x}^k) = \inf_{x\in C} g(x)$).*

ii) If f also satisfies H5, then the whole sequence $\{x^k\}$ is weakly convergent to an element of S.

Proof. Take any $\bar{x} \in S$ and let $v^k = e^k + \lambda_k[f'(x^k) - f'(\tilde{x}^k)]$. If $T = \partial g$, then Proposition 6.2.1(ii)-(a) implies that $v^k \in \partial g(\tilde{x}^k)$. Hence,

$$g(\bar{x}) \geq g(\tilde{x}^k) + \langle v^k, \bar{x} - \tilde{x}^k \rangle.$$

Since $g(\tilde{x}^k) \geq g(\bar{x}) = g^* = \min_{x\in C} g(x)$, we get

$$\langle v^k, \tilde{x}^k - \bar{x} \rangle \geq g(\tilde{x}^k) - g(\bar{x}) \geq 0. \tag{6.7}$$

In view of Proposition 6.3.2(iii) together with the fact that $\lambda_k \leq \bar{\lambda}$, we obtain, taking limits in (6.7), that

$$\lim_{k\to\infty} g(\tilde{x}^k) = g(\bar{x}) = g^*,$$

which proves that $\{\tilde{x}^k\}$ is a minimizing sequence. By Proposition 6.3.2 we have that $\{x^k\}$ is bounded and that $x^k - \tilde{x}^k \xrightarrow[k\to\infty]{\text{s}} 0$. Therefore, taking any weak limit x^∞ of $\{x^k\}$, there exists a subsequence $\{\tilde{x}^{j_k}\}$ such that $\tilde{x}^{j_k} \xrightarrow[k\to\infty]{\text{w}} x^\infty$, which in turn implies

$$g(x^\infty) \leq \liminf_k g(\tilde{x}^{j_k}) = g^*,$$

because of lower semicontinuity of g. Since $\{\tilde{x}^k\} \subset C$, which is weakly closed, we conclude that $x^\infty \in C$. Thus, x^∞ is a solution of the optimization problem.

Assume now that (b) is true and apply Proposition 6.2.1(ii)-(b) and the four-point property (Proposition 2.3.10(i)-(c)) to get

$$\begin{aligned}
\langle v^k, \tilde{x}^k - x^\infty \rangle &= \langle v^k, \tilde{x}^k - \bar{x} \rangle + \langle v^k, \bar{x} - x^\infty \rangle \\
&= \langle v^k, \tilde{x}^k - \bar{x} \rangle + \lambda_k \langle f'(x^k) - f'(x^{k+1}), \tilde{x}^k - \bar{x} \rangle \\
&= \langle v^k, \tilde{x}^k - \bar{x} \rangle \\
&+ \lambda_k \left[D_f(x^\infty, x^k) - D_f(x^\infty, x^{k+1}) + D_f(\bar{x}, x^k) - D_f(\bar{x}, x^{k+1}) \right].
\end{aligned} \tag{6.8}$$

In view of Proposition 6.3.2, taking $x^{j_k} \xrightharpoonup[k\to\infty]{\text{w}} x^\infty$ and applying H3.a, we get

$$\limsup_k \langle v^{j_k}, \tilde{x}^{j_k} - x^\infty \rangle \leq 0.$$

Observe that $\bar{x} \in S \subset \text{dom}(T+N_C)$. Then, by pseudomonotonicity of $T+N_C$ (see Proposition 2.1.7 (i) and Definition 2.1.4), there exists $v \in [T+N_C](x^\infty)$ satisfying

$$\langle v, x^\infty - \bar{x} \rangle \leq \liminf_j \langle v^{k_j}, \tilde{x}^{k_j} - \bar{x} \rangle = 0, \tag{6.9}$$

where the equality in (6.9) follows from Proposition 6.3.2. Since $0 \in [T+N_C](\bar{x})$, monotonicity of this operator ensures that $\langle v, x^\infty - \bar{x} \rangle = 0$. The result then follows from paramonotonicity of $T + N_C$. □

Proposition 6.3.5. *If $S \cap int(C) \neq \emptyset$, the results of Theorem 6.3.4 holds under assumption (b) without requesting either pseudo- or paramonotonicity of T and with the weaker assumption H3.c instead of H3.a.*

Proof. Since $S\cap$ int$(C) \neq \emptyset$, we can take $\bar{x} \in S\cap$ int(C). From Proposition 6.3.2(i), $D_f(\bar{x}, x^k)$ is a convergent sequence, bounded above by $D_f(\bar{x}, x^0)$. Hence, in view of H6, there exists $\epsilon > 0$ such that $d(x^k, \partial C) \geq \epsilon$ for all k. By Proposition 6.3.2(ii), $\{x^k\}$ is bounded. Since f satisfies H3.c, f' is uniformly continuous on the bounded set $\{x^k\}$, so that

$$v^k = \lambda_k[f'(x^k) - f'(x^{k+1})] \xrightarrow[k\to\infty]{\text{s}} 0, \tag{6.10}$$

using Proposition 6.3.2(vi)-(b) and Proposition 6.2.1(ii)-(b). Since $\tilde{x}^k \in \text{int}(C)$, we get that $v^k \in [T+N_C](\tilde{x}^k)$, in view of Proposition 6.2.1(ii)-(a). Thus, taking a weak cluster point x^∞ of $\{x^k\}$ and a subsequence $\{x^{j_k}\}$ of $\{x^k\}$ such that $x^{j_k} \xrightharpoonup[k\to\infty]{\text{w}} x^\infty$, we get that $\tilde{x}^{j_k} \xrightharpoonup[k\to\infty]{\text{w}} x^\infty$ and the result follows then from (6.10) and demiclosedness of the graph of $T + N_C$. □

6.4 Penalization in polyhedra

In Section 6.3, we have shown that for the particular cases in which the operator T is a subdifferential, or the variational inequality problem, VIP(T,C) has solutions in the interior of C, the properties required on the regularizing function are not too demanding. In fact all examples of regularizing functions given in Chapter 3 satisfy them. When the solutions of VIP(T,C) are on the boundary of C, we use the more demanding condition H3.a. This is the situation in this section, devoted to the particular case when C is a polyhedron as defined in (3.29), i.e.

$$C = \{x \in B \mid \langle v^i, x \rangle \geq \alpha_i,\ i = 1, ..., p\},$$

where $v^1, \ldots, v^p \in B^*\backslash\{0\}$ and $\alpha_1, \ldots, \alpha_p \in \mathbb{R}$. Assume that C has nonempty interior and let f be defined by (3.30), i.e.

$$f(x) = \begin{cases} \frac{1}{r} \|x\|_B^r + \sum_{i=1}^p (\langle v^i, x\rangle - \alpha_i) \log(\langle v^i, x\rangle - \alpha_i) & \text{if } x \in C \\ \infty & \text{otherwise.} \end{cases}$$

Then, as proved in Proposition 3.4.1(ii) together with Remark 3.4.2, f satisfies H3.a when the Banach space B is uniformly convex and uniformly smooth. Indeed, all required properties are satisfied and we can apply Algorithm II for this choice of f.

Corollary 6.4.1. *Let B be a uniformly convex and uniformly smooth Banach space, and take C and f as defined in* (3.29) *and* (3.30) *respectively. Assume that VIP(T,C) has solutions. If T is pseudo- and paramonotone then the sequence $\{x^k\}$ generated by Algorithm II has weak accumulation points and all of them are solutions of VIP(T,C). Moreover, if $S \cap int(C) \neq \emptyset$, then the result holds without the pseudo- and paramonotonicity assumptions on T.*

Proof. By Proposition 3.4.1 and Remark 3.4.2, f satisfies the hypotheses required in Theorem 6.3.4 and Proposition 6.3.5 for the announced results to hold. □

6.5 An approach depending on the regularizing parameters

In the case in which C is a closed ball, we know of no function satisfying both conditions H3.a and H6. The variational inequality problem with a closed ball as a feasible set is used in this section to illustrate how to use H3.b instead of H3.a.

Let us go back to the convergence analysis of Algorithm II, i.e. to the proof of Theorem 6.3.4. Equation (6.8) deserves special attention for two reasons. First, if x^∞ denotes any weak limit of $\{\tilde{x}^k\}$, the behavior of $\langle v^k, \tilde{x}^k - x^\infty\rangle$ is essential for the pseudomonotonicity argument. Second, since results in Proposition 6.3.2 do not depend on assumption H3.a, we know that for any solution $\bar{x}$ it holds that

$$\lim_{k\to\infty} \langle v^k, \tilde{x}^k - \bar{x}\rangle = 0 \quad \text{and} \quad \lim_{k\to\infty} D_f(\bar{x}, x^k) - D_f(\bar{x}, x^{k+1}) = 0.$$

Hence, (6.8) ensures that

$$\limsup_k \langle v^k, \tilde{x}^k - x^\infty\rangle = \limsup_k \lambda_k \left[D_f(x^\infty, x^k) - D_f(x^\infty, x^{k+1})\right], \tag{6.11}$$

which means that $\langle v^k, \tilde{x}^k - x^\infty\rangle$ behaves asymptotically as $D_f(x^\infty, x^k) - D_f(x^\infty, x^{k+1})$ multiplied by the factor λ_k. Thus, a condition like H3.a looks appropriate when a pseudomonotonicity argument is being used, but in its absence, we can enforce the same effect by playing with the the sequence of regularizing parameters, and using assumption H3.b defined in Section 2.5.

Theorem 6.5.1. *Let $\{x^k\}$ be the sequence generated by Algorithm II with $f \in \mathcal{F}_C$ satisfying H1, H2, H3.b and H3.c. Assume that VIP(T,C) has solutions and that T is pseudo- and paramonotone. If we take $\lambda_k = \rho_k$ for all k, with ρ_k as in H3.b, then all weak accumulation points of $\{x^k\}$ are solutions of VIP(T,C).*

Proof. Let $\{x^k\}$ be the sequence generated by Algorithm II, which is bounded by Proposition 6.3.2(ii). If $S\cap\text{int}(C) \neq \emptyset$, the result follows from Proposition 6.3.5 using neither pseudo- nor paramonotonicity, nor H3.b, but H3.c instead. Thus, we can assume that

$$S \cap \text{int}(C) = \emptyset. \tag{6.12}$$

Note first that if $\liminf_k d(x^k, \partial C) > 0$ then, since f' is uniformly continuous on the bounded set $\{x^k\}$ by H3.c, we get that there exist weak accumulation points of $\{x^k\}$ in the interior of C and that they are solutions, contradicting (6.12). Thus, we assume that $\lim_k d(x^k, \partial C) = 0$. Take any weak limit point x^∞ of $\{x^k\}$. Since $\lambda_k = \rho_k$ with ρ_k given by H3.b, it holds that

$$\limsup_k \lambda_k \left[D_f(x^\infty, x^k) - D_f(x^\infty, x^{k+1})\right] \leq 0. \tag{6.13}$$

In view of Proposition 6.3.2 and (6.13), taking a subsequence $\{x^{j_k}\}$ of $\{x^k\}$ such that $x^{j_k} \xrightarrow[k\to\infty]{w} x^\infty$, we get from (6.8) that $\limsup_k \langle v^{j_k}, \tilde{x}^{j_k} - x^\infty\rangle \leq 0$. We complete the proof as in Theorem 6.3.4 after (6.8), using pseudo- and paramonotonicity of T. □

We mention that this choice of $\{\lambda_k\}$ might force $\lim_k \lambda_k = 0$, in which case the regularizing effect might become negligeable for large k.

We present next a particular case for which a function satisfying H3.b is available. In this example C is the closed unit ball of center zero in B and f is the function given by (3.12), i.e.

$$f(x) = \begin{cases} 1 - \sqrt{1 - \|x\|^2} & \text{if } \|x\| \leq 1 \\ \infty & \text{otherwise,} \end{cases}$$

It was proved in Section 3.3 that f fails to satisfy H3.a (see Example 3.3.5) but it does satisfy H3.b (see Proposition 3.3.6).

Corollary 6.5.2. *Let B be uniformly convex and uniformly smooth. Consider $C = B[0,1]$ and assume that VIP(T,$B[0,1]$) has solutions and that T is pseudo- and paramonotone. Take the regularizing function f as above (i.e. as defined in* (3.12)*) and assume that at step k of Algorithm II λ_k is taken such that $\lambda_k \leq 1 - \left\|x^k\right\|$. Then $\{x^k\}$ has weak accumulation points and all of them are solutions of VIP(T,C). Moreover, if $S \cap int(C) = \emptyset$ then the whole sequence $\{x^k\}$ converges strongly to a point in $S \cap \partial C$.*

Proof. Take $C = B[0,1]$ and f defined in (3.12). From Corollary 3.3.2, Corollary 3.3.4, Proposition 3.3.7 and Proposition 3.3.6 we know that f belongs to $\mathcal{F}_C$ and satisfies H1, H2, H3.c and H3.b with any $\rho_k \leq 1 - \left\|x^k\right\|$. Then, we can apply Theorem 6.5.1, with $\lambda_k = \rho_k \leq 1 - \left\|x^k\right\|$, to get that $\{x^k\}$ has weak accumulation points and all of them are solutions of VIP(T,C).

Assume now that $S \cap \text{int}(C) = \emptyset$, so that $S \subset \partial C = S[0,1]$ (i.e. all solutions belong to the unit sphere). Since S is convex by Proposition 2.1.6, because $S = [T + N_C]^{-1}(0)$, strict convexity of B implies that S is a singleton, say $\{\bar{x}\}$, in view of Definition 2.2.6(i) and Remark 2.2.7. Thus, the whole sequence $\{x^k\}$ is weakly convergent to $\bar{x}$. Therefore,

$$1 = \|\bar{x}\| \leq \liminf_k \left\|x^k\right\| \leq 1,$$

i.e. $\left\|x^k\right\|_B \overset{k\to\infty}{\longrightarrow} \|\bar{x}\|_B$ and $x^k \xrightarrow[k\to\infty]{\text{w}} \bar{x}$. Since B is locally uniformly convex, it satisfies property (h) (see Definition 2.2.9 and Remark 2.2.10). Hence, the convergence is strong. i.e. $x^k \xrightarrow[k\to\infty]{\text{s}} \bar{x}$. □

Observe that in this case, if there exist solutions, but none of them belongs to the interior of C, then the choice $\lambda_k \leq 1 - \left\|x^k\right\|$ implies $\lim_k \lambda_k = 0$.

6.6 Concluding remarks

We summarize next some relevant features of the results presented in this chapter and present some aspects that remain open for further research.

Features:

- It is possible to get the penalization effect for proximal methods in Banach spaces with the help of appropriate regularizing functions.
- We give sufficient conditions on regularizing functions which achieve the penalization effect. We present explicit cases of such functions for closed balls and polyhedra.

- Our results are valid also for inexact solutions of the proximal subproblems under error criteria with constant relative error.
- We prove that the Hybrid Proximal-Extragradient method with in-built penalization works not only in Hilbert spaces but also in reflexive Banach spaces, extending its domain of application originally stated in finite dimension by Solodov and Svaiter ([62]).

Open research:

- It is relevant to find examples for feasible sets and associated regularizing functions, other than balls and polyhedra.
- For the situation where there exists no regularizing function satisfying assumptions H3.a, H3.b is not totally satisfactory. Is there any way to prove the convergence properties of the method without letting the regularizing parameters converge to zero?
- In this chapter we have included in-built penalization only for the Hybrid Proximal-Extragradient method. It is worthwhile to attempt a similar analysis for the Hybrid Proximal-Bregman Projection method (Algorithm I in Chapter 4).

Bibliography

[1] Alber, Y. I., Burachik, R. S., and Iusem, A. N. A proximal point method for nonsmooth convex optimization problems in Banach spaces. *Abstr. Appl. Anal. 2*, 1-2 (1997), 97–120.

[2] Araujo, A. The nonexistence of smooth demand in general Banach spaces. *J. Math. Econom. 17*, 4 (1988), 309–319.

[3] Auslender, A., Teboulle, M., and Ben-Tiba, S. Interior proximal and multiplier methods based on second order homogeneous kernels. *Math. Oper. Res. 24*, 3 (1999), 645–668.

[4] Auslender, A., Teboulle, M., and Ben-Tiba, S. A logarithmic-quadratic proximal method for variational inequalities. *Comput. Optim. Appl. 12*, 1-3 (1999), 31–40.

[5] Bertsekas, D. P. On penalty and multiplier methods for constrained minimization. *SIAM J. Control Optim. 14*, 2 (1976), 216–235.

[6] Bertsekas, D. P. *Constrained optimization and Lagrange multiplier methods.* Academic Press Inc., New York, 1982.

[7] Bregman, L. M. A relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming. *Ž. Vyčisl. Mat. i Mat. Fiz. 7* (1967), 620–631.

[8] Browder, F. E. Nonlinear operators and nonlinear equations of evolution in Banach spaces. In *Nonlinear functional analysis (Proc. Sympos. Pure Math., Vol. XVIII, Part 2, Chicago, Ill., 1968).* Amer. Math. Soc., Providence, R. I., 1976, pp. 1–308.

[9] Burachik, R., and Scheimberg, S. A proximal point algorithm for the variational inequality problem in Banach spaces. *SIAM J. Control Optim. 39*, 5 (2001), 1633–1649.

[10] Burachik, R., and Svaiter, B. A relative error tolerance for a family of generalized proximal point methods. To appear.

[11] Burachik, R. S. *Generalized proximal point methods for the variational inequality problem.* PhD thesis, IMPA, Rio de Janeiro, 1995.

[12] Burachik, R. S., and Iusem, A. N. A generalized proximal point algorithm for the variational inequality problem in a Hilbert space. *SIAM J. Optim. 8*, 1 (1998), 197–216.

[13] Burachik, R. S., Iusem, A. N., and Svaiter, B. F. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Anal. 5*, 2 (1997), 159–180.

[14] Burke, J. V., Ferris, M. C., and Qian, M. On the Clarke subdifferential of the distance function of a closed set. *J. Math. Anal. Appl. 166*, 1 (1992), 199–213.

[15] Butnariu, D., and Iusem, A. N. On a proximal point method for convex optimization in Banach spaces. *Numer. Funct. Anal. Optim. 18*, 7-8 (1997), 723–744.

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

[17] Butnariu, D., Iusem, A. N., and Resmerita, E. Total convexity for powers of the norm in uniformly convex Banach spaces. *J. Convex Anal. 7*, 2 (2000), 319–334.

[18] Buys, J. D. *Dual algorithms for constrained optimization problems.* Rijksuniversiteit te Leiden, Leiden, 1972. Doctoral dissertation, Faculty of Science, University of Leiden.

[19] Censor, Y., and Zenios, S. A. Proximal minimization algorithm with D-functions. *J. Optim. Theory Appl. 73*, 3 (1992), 451–464.

[20] Censor, Y., and Zenios, S. A. *Parallel optimization. Theory, algorithms, and applications.* Oxford University Press, New York, 1997.

[21] Chen, G., and Teboulle, M. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM J. Optim. 3*, 3 (1993), 538–543.

[22] Cioranescu, I. *Geometry of Banach spaces, duality mappings and nonlinear problems.* Kluwer Academic Publishers Group, Dordrecht, 1990.

[23] Cohen, G. Auxiliary problem principle and decomposition of optimization problems. *J. Optim. Theory Appl. 32*, 3 (1980), 277–305.

[24] Cohen, G. Auxiliary problem principle extended to variational inequalities. *J. Optim. Theory Appl. 59*, 2 (1988), 325–333.

[25] Eckstein, J. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Math. Oper. Res. 18*, 1 (1993), 202–226.

[26] Eckstein, J. Approximate iterations in Bregman-function-based proximal algorithms. *Math. Programming 83*, 1, Ser. A (1998), 113–123.

[27] Eckstein, J., Humes Jr, C., and Silva, P. Rescaling and stepsize selection in proximal methods using separable generalized distances. *SIAM J. Optim.* To appear.

[28] Eggermont, P. P. B. Multiplicative iterative algorithms for convex programming. *Linear Algebra Appl. 130* (1990), 25–42.

[29] Eriksson, J. An iterative primal-dual algorithm for linear programming. Tech. Rep. 85-10, Department of Mathematics, Linköping University, 1985.

[30] Erlander, S. Entropy in linear programs. *Math. Programming 21*, 2 (1981), 137–151.

[31] Güler, O. On the convergence of the proximal point algorithm for convex minimization. *SIAM J. Control Optim. 29*, 2 (1991), 403–419.

[32] Hestenes, M. R. Multiplier and gradient methods. *J. Optim. Theory Appl. 4* (1969), 303–320.

[33] Humes Jr, C., Silva, P., and Svaiter, B. Some inexact hybrid proximal augmented Lagrangian methods. To appear.

[34] Isnard, C. A., and Iusem, A. N. On mixed Hölder-Minkowski inequalities and total convexity of certain functions in $L^p(\Omega)$. *Math. Inequal. Appl. 3*, 4 (2000), 519–537.

[35] Iusem, A. N. Some properties of generalized proximal point methods for quadratic and linear programming. *J. Optim. Theory Appl. 85*, 3 (1995), 593–612.

[36] Iusem, A. N. On some properties of generalized proximal point methods for variational inequalities. *J. Optim. Theory Appl. 96*, 2 (1998), 337–362.

[37] Iusem, A. N. On some properties of paramonotone operators. *J. Convex Anal. 5*, 2 (1998), 269–278.

[38] Iusem, A. N., Svaiter, B. F., and Teboulle, M. Entropy-like proximal methods in convex programming. *Math. Oper. Res. 19*, 4 (1994), 790–814.

[39] Iusem, A. N., and Teboulle, M. Convergence rate analysis of nonquadratic proximal methods for convex and linear programming. *Math. Oper. Res. 20*, 3 (1995), 657–677.

[40] Kassay, G. The proximal points algorithm for reflexive Banach spaces. *Studia Univ. Babeş-Bolyai Math. 30* (1985), 9–17.

[41] Kiwiel, K. C. Proximal minimization methods with generalized Bregman functions. *SIAM J. Control Optim. 35*, 4 (1997), 1142–1168.

[42] Kort, B. W., and Bertsekas, D. P. Combined primal-dual and penalty methods for convex programming. *SIAM J. Control Optim. 14*, 2 (1976), 268–294.

[43] Krasnosel′skiĭ, M. A. Two remarks on the method of successive approximations. *Uspehi Mat. Nauk (N.S.) 10*, 1(63) (1955), 123–127.

[44] Liese, F., and Vajda, I. *Convex statistical distances.* BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1987.

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

[46] Martinet, B. Régularisation d'inéquations variationnelles par approximations successives. *Rev. Française Informat. Recherche Opérationnelle 4*, Ser. R-3 (1970), 154–158.

[47] Minty, G. J. Monotone (nonlinear) operators in Hilbert space. *Duke Math. J. 29* (1962), 341–346.

[48] Moreau, J.-J. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France 93* (1965), 273–299.

[49] Mosco, U. Perturbation of variational inequalities. In *Nonlinear Functional Analysis (Proc. Sympos. Pure Math., Vol. XVIII, Part 1, Chicago, Ill., 1968).* Amer. Math. Soc., Providence, R.I., 1970, pp. 182–194.

[50] Pascali, D., and Sburlan, S. *Nonlinear mappings of monotone type.* Martinus Nijhoff Publishers, The Hague, 1978.

[51] Penot, J.-P., and Ratsimahalo, R. Characterizations of metric projections in Banach spaces and applications. *Abstr. Appl. Anal. 3*, 1-2 (1998), 85–103.

[52] Phelps, R. R. Metric projections and the gradient projection method in Banach spaces. *SIAM J. Control Optim. 23*, 6 (1985), 973–977.

[53] Phelps, R. R. *Convex functions, monotone operators and differentiability.* Springer-Verlag, Berlin, 1989.

[54] Powell, M. J. D. A method for nonlinear constraints in minimization problems. In *Optimization (Sympos., Univ. Keele, Keele, 1968).* Academic Press, London, 1969, pp. 283–298.

[55] Rockafellar, R. T. Local boundedness of nonlinear, monotone operators. *Michigan Math. J. 16* (1969), 397–407.

[56] Rockafellar, R. T. Monotone operators associated with saddle-functions and minimax problems. In *Nonlinear Functional Analysis (Proc. Sympos. Pure Math., Vol. XVIII, Part 1, Chicago, Ill., 1968).* Amer. Math. Soc., Providence, R.I., 1970, pp. 241–250.

[57] Rockafellar, R. T. On the maximality of sums of nonlinear monotone operators. *Trans. Amer. Math. Soc. 149* (1970), 75–88.

[58] Rockafellar, R. T. The multiplier method of Hestenes and Powell applied to convex programming. *J. Optim. Theory Appl. 12* (1973), 555–562.

[59] Rockafellar, R. T. Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM J. Control 12* (1974), 268–285.

[60] Rockafellar, R. T. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res. 1*, 2 (1976), 97–116.

[61] Rockafellar, R. T. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim. 14*, 5 (1976), 877–898.

[62] Solodov, M., and Svaiter, B. An inexact hybrid generalized proximal point algorithms and some new results on the theory of Bregman functions. *Math. Oper. Res. 51* (2000), 479–494.

[63] Solodov, M. V., and Svaiter, B. F. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal. 7*, 4 (1999), 323–345.

[64] Solodov, M. V., and Svaiter, B. F. A hybrid projection-proximal point algorithm. *J. Convex Anal. 6*, 1 (1999), 59–70.

[65] Troyanski, S. L. On locally uniformly convex and differentiable norms in certain non-separable Banach spaces. *Studia Math. 37* (1970/71), 173–180.

[66] Wierzbicki, A. P., and Kurcyusz, S. Projection on a cone, penalty functionals and duality theory for problems with inequality constraints in Hilbert space. *SIAM J. Control Optim. 15*, 1 (1977), 25–56.

[67] Zeidler, E. *Nonlinear functional analysis and its applications. I.* Springer-Verlag, New York, 1986.

[68] Zeidler, E. *Nonlinear functional analysis and its applications. II/B.* Springer-Verlag, New York, 1990.