Proximal methods with penalization effects in Banach spaces

Rolando Gárciga Otero* Alfredo Iusem†

January 25, 2002

Abstract

In this paper we introduce a boundary coercive condition on the regularizing function of the proximal point method in Banach spaces, which has a penalization effect and is useful for solving certain variational inequality problems. As its finite dimensional counterpart (proximal point method with Bregman distances) it avoids explicit consideration of the constraints, because the feasibility is taken care of by the regularizing function, whose derivative diverges on the boundary of the feasible set. The algorithm is therefore an interior point one. We show that this effect is guaranteed even for an inexact version of the method, namely the hybrid extragradient proximal point method proposed by Solodov and Svaiter for finite dimensional spaces. We give a full convergence analysis, and examples of regularizing functions satisfying the required conditions, for the cases of the feasible set being a closed ball or a polyhedron.

1 Introduction

In a reflexive Banach space B, an appropriate extension of the proximal point method for finding zeroes of a maximal monotone operator $T : B \to \mathcal{P}(B^*)$, where B^* is the dual of B, is achieved by taking x^{k+1} as the unique $x \in B$ such that

$$\lambda_k[f'(x^k) - f'(x)] \in T(x), \tag{1}$$

where $f : B \to \mathbb{R} \cup \{\infty\}$ is a strictly convex and Gâteaux differentiable function on the interior of its domain satisfying some technical assumptions (H1–H4 in Sention 2), and f' is its Gâteaux derivative (see [20, 1, 6, 5, 16]). Under such assumptions, it was proved in [5] that if dom $T \subset$ int(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. These results were extended in [16] where schemes allowing for inexact solutions of (1) are considered, and further

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

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

in [25]. But in all these works the function f is used only as a regularizing term, and no attempt is made of introducing in f information on a feasible set C. On the other hand, in finite dimensional or Hilbert spaces, where the square of the norm leads always to simpler computations, the scheme given by (1) 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 we consider several alternative hypotheses on f, called H3(a), H3(b) and H3(c) (see Section 2).

These methods, usually known as Proximal Point methods with Bregman functions (PPB for short), can be traced back to [14], [15] and [13]. Later on they have been studied e.g. in [8], [10], [17], [21] which consider the case of convex optimization under progressively weaker assumptions on the problem data or the Bregman function, and in [4], [12], [18] for the more general problem of finding zeroes of monotone operators. In [4] the notion of coercive Bregman function is appropriately extended to Hilbert spaces. Convergence results in [4] can be summarized as follows. If VIP(T,C) has solutions and T is pseudo- and paramonotone, then the sequence generated by PPB is bounded and all weak accumulation points are optimal solutions. Moreover, there exists a unique accumulation point when f' is weak-to-weak* continuous. In the finite dimensional case the pseudomonotonicity assumption is redundant (see [26]).

The main objective of this work is to extend the penalty approach in proximal-like methods for solving variational inequality problems to Banach spaces, and specifically to the inexact variant of the proximal method presented as Algorithm II in [16] (i.e. the hybrid extragradient-proximal point method). We emphasize that the main difficulty lies in identifying properties of the regularizing function which allow us to establish the convergence results but at the same time are such that examples of functions satisfying them can be found for relevant and nontrivial instances of feasible sets. Also, such functions should be such that the algorithm becomes implementable, at least in principle. It is possible to present properties which look less convoluted than our H3(a)–H3(c), but then the task of providing significant and manageable examples seems unsurmountable.

The material is organized as follows: Section 2 contains some preliminary material and lists all relevant properies of the regularizing function f, including the boundary coerciveness property H6, essential for penalization purposes. In Section 3.1 we prove some consequences of H6. In Section 3.2 we present the algorithm and establish that the generated sequence is well defined. The convergence properties of the algorithm are analyzed in Section 3.3. The main difference with respect to the convergence analysis in [16] 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 cannot be expected to hold. Our convergence analysis uses alternative weaker conditions instead of H3. In Section 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 4.1 satisfies H3(a). For the case in which C is a closed ball, the function discussed in Section 4.3 does not satisfy H3(a) but rather H3(b), and the penalization effect can be achieved, but with additional conditions imposed upon the regularizing coefficients: they must converge to zero in a predetermined manner, which is discussed in Section 4.2. In Section 4.3 we establish the convergence properties of the algorithm for the case of the ball with the additional condition on the regularizing parameters. Section 5 contains our conclusions and some open issues.

2 Preliminaries

Le B be a real reflexive Banach space with norm denoted by $\|\cdot\|$, B^* its topological dual (with the operator norm denoted by $\|\cdot\|_*$), and $\langle\cdot,\cdot\rangle$ the dual product in $B^* \times B$ (i.e $\langle\phi, x\rangle = \phi(x)$ for all $\phi \in B^*$ and all $x \in B$). Given $\{x^k\} \subset B$, we use the notation $x^k \xrightarrow[k\to\infty]{s} x$ (respectively $x^k \xrightarrow[k\to\infty]{w} x$) meaning that the sequence $\{x^k\}$ converges in the strong (respectively weak) topology of B to $x \in B$.

We recall that a point-to-set operator $T : B \to \mathcal{P}(B^*)$ is said to be *monotone* if $\langle w - w', x - x'\rangle \geq 0$ for all $x,\ x' \in B$ and all $w \in T(x),\ w' \in T(x')$. If this relation holds with strict inequality when $x \neq x'$ then T is called *strictly* monotone. There is a class of operators which is in between monotonicity and strict monotonicity: paramonotone operators.

Definition 2.1. 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:

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

Definition 2.2. An operator $T : B \to \mathcal{P}(B^*)$ is said to be *pseudomonotone* if its domain, $\operatorname{dom} T = \{x \in B \mid T(x) \neq \emptyset\}$, is closed and convex and the following condition is satisfied:

$$\begin{aligned} &\{x^k\} \subset \operatorname{dom} T,\ w^k \in T(x^k),\ x^k \xrightarrow[k\to\infty]{w} x^o \text{ and } \operatorname{limsup}_{k\to\infty}\langle w^k, x^k - x^o\rangle \leq 0 \\ &\Longrightarrow \forall y \in \operatorname{dom} T,\ \exists w^o \in T(x^o) \text{ such that } \langle w^o, x^o - y\rangle \leq \operatorname{liminf}_{k\to\infty}\langle w^k, x^k - y\rangle\ . \end{aligned}$$

A monotone operator is called *m*aximal 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. A well known example of a pseudo-, para- and maximal monotone operator is the subdifferential operator, $\partial f(\cdot) : B \to \mathcal{P}(B^*)$, of any proper lower semicontinuous convex function f (see e.g. Theorem 3.25 of [23] for maximality and [5] for pseudo- and paramonotonicity). In particular the *normalizing operator* of a nonempty closed and convex set C, i.e. $N_C = \partial I_C$, has such properties. Here I_C is the *indicator function* of C and the operator $N_C : B \to \mathcal{P}(B^*)$ can then be defined as follows.

$$N_C(x) = \begin{cases} \{z \in B^* |\ \langle z, y - x\rangle \leq 0,\ \forall y \in C\} & \text{if } x \in C \\ \emptyset & \text{otherwise.} \end{cases} \tag{2}$$

A wide variety of problems, such as convex optimization problems, are particular cases of the variational inequality problem, which we define next.

Definition 2.3. 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 $y \in C$.

Assume that $f : B \to \mathbb{R} \cup \{\infty\}$ is lower semicontinuous, with int(dom f) $\neq \emptyset$ and that f is G-differentiable on int(dom f). We denote such class of functions as $\mathcal{F}$. The Bregman distance and the modulus of total convexity associated to f are defined as follows.

Definition 2.4 (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 2.5. *Let $f \in \mathcal{F}$. Then*

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

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

$$D_f(y,x) - D_f(y,z) - D_f(z,x) = \langle f'(x) - f'(z), z - y\rangle. \tag{3}$$

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

$$D_f(w,z) = D_f(w,x) + \langle f'(x) - f'(z), w - y\rangle + D_f(y,z) - D_f(y,x). \tag{4}$$

Proof. The results can be found e.g. in [7]: item (i) is proved in 1.1.9 and (ii) in 1.3.9 of [7]. The equality in (4) follows from (ii) and its denomination *four-point equality* is taken from [26]. $\square$

Definition 2.6 (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.

When $\nu_f(x,t) > 0$ for all $x \in$ int(dom f) and $t > 0$ then f is called *totally convex.*

Definition 2.7. 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 $t > 0$, it holds that $\inf_{x\in\tilde{E}} \nu_f(x,t) > 0$.

The several assumptions on f to be considered in the sequel are the following:

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

H2: f is uniformly totally convex on $\mathrm{int}(\mathrm{dom}\, f)$, or equivalently (see [7], 2.1.2),

sequential consistency[1]: For all $\{x^k\}$, $\{y^k\} \subset \mathrm{int}(\mathrm{dom}\, f)$ such that $\{x^k\}$ is bounded and $\lim_{k\to\infty} D(y^k, x^k) = 0$, it holds that $x^k - y^k \xrightarrow[k\to\infty]{\mathrm{s}} 0$.

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

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

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

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

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

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

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 $\lim_{k\to\infty} D_f(w, z^k) = +\infty$ (here C is the feasible set of the variational inequality problem VIP(T,C) for which the method is intended)

We make a few comments on H1–H6. As mentioned before, in the unconstrained case (i.e. for the method without penalization purposes) H1-H4 are used for establishing that the generated sequence is well defined and bounded, and that its weak accumulation points are solutions, while H5 implies uniqueness of the accumulation point; excepting for H3, they play the same roles in the current situation. H6, related to condition B6 (boundary coerciveness) of [17], [4], introduces the feasible set C and entails the penalization effect: it implies that $\|f'(x)\|$ diverges as x approaches ∂C, but at the same time it is inconsistent with H3, because under H6 f' cannot be uniformly continuous on balls whose boundary intersects ∂C. Thus, we are forced to replace H3 by weaker conditions. H3(a), introduced in [5], is weaker than H3: when H2 and H3 hold, H3(a) follows from the three point equality (3). H3(b) is even weaker: it follows from H3(a), by taking $\rho_k = 1$ for all k, but it is less desirable, because under it convergence of the method may require choices of the regularization parameters λ_k, giving raise to numerically less stable subproblems. We include it

[1] H2 also implies sequential consistency with boundedness of the sequence $\{x^k\}$ replaced by boundedness of $\{y^k\}$. See e.g. [16], Proposition 5.

because there are relevant examples of sets C (e.g. balls) for which functions satisfying H3(b) are available, but we know no examples satisfying H3(a). H3(c) is also weaker than H3(a), and it is enough when there exist solutions in the interior of C. It is preferably to H3(b), because under it no additional demands on the λ_k's are needed, but the most interesting instances of VIP(T,C) are those in which all solutions belong to the boundary of C (otherwise the constraints imbedded in C are to some extent redundant).

3 Proximal Point methods with in-built penalization in Banach spaces

In this section $T : B \to \mathcal{P}(B^*)$ denotes a maximal monotone operator 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. Let S denote the solution set of this problem, i.e. the set of zeroes of the maximal monotone operator $T + N_C$.

3.1 The boundary coerciveness condition

The following lemma illustrates the way in which H6 will be used in our analysis.

Lemma 3.1. *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 2.4,

$$\begin{aligned} D_f(x, z^k) &\leq D_f(x, z^k) + D_f(z^k, x) = \langle f'(z^k) - f'(x), z^k - x\rangle \\ &= \langle f'(z^k), z^k - x\rangle - \langle f'(x), z^k - x\rangle \leq \left\|f'(z^k) - f'(x)\right\|_* \left\|z^k - x\right\|. \end{aligned}$$

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 $\{z^k\} \subset$ int(C) converging strongly to z. Hence, $z \notin$ dom f', because otherwise the norm-to-weak* continuity of f' :dom $f' \to B^*$, i.e. demicontinuity (see Corollary 1.1 of [24]), together with (i), imply that for any $x \in$ 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. □

3.2 Inexact Proximal Point-Extragradient method with in-built penalization in Banach spaces.

In this subsection we present the method under consideration. It accepts inexact solutions of the subproblems, with the criterion introduced in the inexact hybrid proximal point-extragradient method, proposed by Solodov and Svaiter [26], for solving the variational inequality problem VIP(T,C) in finite dimensional spaces. We follow here the presentation of [16], where the method was first studied in nonhilbertian Banach spaces, but without the penalization effect, i.e. without demending that f satisfies H6. The algorithm requires an exogenous constant $\sigma \in [0,1)$, an exogenous bounded sequence $\{\lambda_k\} \subset \mathbb{R}_{++}$ and an auxiliary function $f \in \mathcal{F}$. It is defined as follows:

Algorithm: 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
$$e^k + \lambda_k[f'(x^k) - f'(\tilde{x}^k)] \in T(\tilde{x}^k), \tag{5}$$
where e^k is any vector in B^* which satisfies
$$D_f\big(\tilde{x}^k, (f')^{-1}\big[f'(\tilde{x}^k) - \lambda_k^{-1}e^k\big]\big) \le \sigma D_f(\tilde{x}^k, x^k), \tag{6}$$
with D_f as in Definition 2.4.
3. If $\tilde{x}^k = x^k$, then stop. Otherwise,
$$x^{k+1} = (f')^{-1}\big[f'(\tilde{x}^k) - \lambda_k^{-1}e^k\big]. \tag{7}$$

The next proposition establishes that the generated sequence is well defined.

Proposition 3.2. *Take $f \in \mathcal{F}$ satisfying H4 and H6 and such that dom $f = C$. Then*

i) The algorithm above is well defined: for all k there exists $\tilde{x}^k$ satisfying (5) *and* (6). *Moreover, any such $\tilde{x}^k$ belongs to int(C). Also, given $\tilde{x}^k$ the next iterate x^{k+1} is uniquely determined by* (7) *(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) $D_f\big(\tilde{x}^k, x^{k+1}\big) \le \sigma D_f(\tilde{x}^k, x^k)$.

Proof. We proceed by induction, noting that $x^0 \in \operatorname{int}(C)$ by assumption. Indeed, assume that $x^k \in \operatorname{int}(C)$. Take $e^k = 0 \in B^*$, in which case equations (5) and (6) take the form:

$$0 \in T(\tilde{x}^k) + \lambda_k[f'(\tilde{x}^k) - f'(x^k)] \tag{8}$$
$$0 = D_f(\tilde{x}^k, \tilde{x}^k) \leq \sigma D_f(\tilde{x}^k, x^k). \tag{9}$$

Lemma 2.10 and Corollary 3.1 of [5] establish that inclusion (8) has a (unique) solution $\tilde{x}^k$, while (9) trivially holds for any $\sigma \geq 0$. In view of Lemma 3.1(ii) and H6, we get dom f' =int(C), and consequently $\tilde{x}^k \in \operatorname{int}(C)$ and f' is strictly monotone on $\operatorname{int}(C)$, because of strict convexity of f. Thus, we get from H4 that the next iterate x^{k+1} is uniquely determined by (7) 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 (5). Since $\tilde{x}^k \in \operatorname{int}(C)$ by item (i), $N_C(\tilde{x}^k) = 0$ (see (2)), and hence (a) holds. (7) together with (6) imply (c). By (7) 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. $\square$

3.3 Convergence analysis

The following proposition establishes the fundamental global properties of the algorithm with the penalization effect.

Proposition 3.3. *Let $\{x^k\}$, $\{\tilde{x}^k\}$ be the sequences generated by the algorithm (5)–(7). Assume that $f \in \mathcal{F}$ satisfies H1, H4, H6 and that dom $f = C$. 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) $\tilde{x}^k - x^k \xrightarrow[k\to\infty]{\mathrm{s}} 0$, *and consequently* $\{\tilde{x}^k\}$ *is bounded,*
b) $x^{k+1} - x^k \xrightarrow[k\to\infty]{\mathrm{s}} 0$.

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 (4), we get

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

$$\leq D_f(\bar{x},x^k) - \lambda_k^{-1}\langle v^k, \tilde{x}^k - \bar{x}\rangle + (\sigma-1)D_f(\tilde{x}^k,x^k) \leq D_f(\bar{x},x^k) - (1-\sigma)D_f(\tilde{x}^k,x^k), \quad (10)$$

where the second equality follows from Proposition 3.2(ii)-(b), the first inequality from Proposition 3.2(ii)-(c) and the last one from Proposition 3.2(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 (10) that $\{D_f(\bar{x},x^k)\}$ is a nonnegative, nonincreasing sequence, henceforth convergent, and that $\{x^k\}$ is contained in a level set of $D_f(\bar{x},\cdot)$, which is bounded by H1. Also, using again (10),

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

from which (iii) and (iv) follow easily. Item (v) follows from (iv) and Proposition 3.2(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 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.

Proposition 3.4. *Suppose that the algorithm* (5)–(7) *stops after k steps. Then the point $x^k \in B$ generated by the algorithm is a solution of VIP(T,C).*

Proof. The algorithm 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), $e^k = 0$, which in turn implies, by (5), $0 \in T(\tilde{x}^k)$. Consequently Proposition 3.2(ii)-(a) ensures that $0 \in T(\tilde{x}^k) + N_C(\tilde{x}^k)$, i.e. $\tilde{x}^k$ is a zero of $T+N_C$. We conclude that $\tilde{x}^k$ is a solution of VIP(T,C). □

Next we prove the convergence properties in the case of infinite sequences, under H3(a) for VIP(T,C), and H3 in the optimization case.

Theorem 3.5. *Assume that $f \in \mathcal{F}$ satisfies H1, H2, H3, H4 and H6 and that dom $f = C$. Let $\{x^k\}$ be the sequence generated by the algorithm* (5)–(7), *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 some $x^* \in S$.

Proof. i) 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 3.2(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

$$\langle v^k, \tilde{x}^k - \bar{x} \rangle \geq g(\tilde{x}^k) - g(\bar{x}) \geq 0. \tag{11}$$

In view of Proposition 3.3(iii) together with the fact that $\lambda_k \leq \bar{\lambda}$, we obtain, taking limits in (11), 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 3.3 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 the 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.

Consider now assumption (b) and apply Proposition 3.2(ii)-(b) and the four-point property (Proposition 2.5(iii)) to get

$$\langle v^k, \tilde{x}^k - x^\infty \rangle = \langle v^k, \tilde{x}^k - \bar{x} \rangle + \langle v^k, \bar{x} - x^\infty \rangle = \langle v^k, \tilde{x}^k - \bar{x} \rangle + \lambda_k \langle f'(x^k) - f'(x^{k+1}), \tilde{x}^k - \bar{x} \rangle$$

$$= \langle v^k, \tilde{x}^k - \bar{x} \rangle + \lambda_k \left[D_f(x^\infty, x^k) - D_f(x^\infty, x^{k+1}) + D_f(\bar{x}, x^k) - D_f(\bar{x}, x^{k+1}) \right]. \tag{12}$$

In view of Proposition 3.3, taking $x^{j_k} \xrightarrow[k\to\infty]{\text{w}} x^\infty$ and applying H3(a), we get

$$\limsup_k \langle v^{j_k}, \tilde{x}^{j_k} - x^\infty \rangle \leq 0.$$

Observe that $\bar{x} \in S \subset \mathrm{dom}(T + N_C)$. Then, by pseudomonotonicity of $T + N_C$ (see [22], p. 97) and Definition 2.2), there exists $v \in [T + N_C](x^\infty)$ satisfying

$$\langle v, x^\infty - \bar{x} \rangle \leq \liminf_j \langle v^{k_j}, \tilde{x}^{k_j} - \bar{x} \rangle = 0, \tag{13}$$

where the equality in (13) follows from Proposition 3.3. 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 $T + N_C$ (see [19]). $\square$

The next proposition shows that our convergence result holds under weaker assumptions on both f and T when there exist solutions of VIP(T,C) in the interior of C.

Proposition 3.6. *If $S \cap int(C) \neq \emptyset$, the results of Theorem 3.5 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 $S \cap \operatorname{int}(C) \neq \emptyset$, we can take $\bar{x} \in S \cap \operatorname{int}(C)$. From Proposition 3.3(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 3.3(ii), $\{x^k\}$ is bounded. Since f satisfies H3(c), f' is uniformly continuous on the bounded set $\{x^k\}$, so that

$$v^k = \lambda_k [f'(x^k) - f'(x^{k+1})] \xrightarrow[k\to\infty]{\mathrm{s}} 0, \tag{14}$$

using Proposition 3.3(vi)-(b) and Proposition 3.2(ii)-(b). Since $\tilde{x}^k \in \operatorname{int}(C)$, we get that $v^k \in [T + N_C](\tilde{x}^k)$, in view of Proposition 3.2(ii)-(a). Thus, taking a weak cluster point x^∞ of $\{x^k\}$ and a subsequence $\{x^{j_k}\}$ of $\{x^k\}$ such that $x^{j_k} \xrightarrow[k\to\infty]{\mathrm{w}} x^\infty$, we get that $\tilde{x}^{j_k} \xrightarrow[k\to\infty]{\mathrm{w}} x^\infty$, and the result follows then from (14) and demiclosedness of the graph of $T + N_C$ (see , e.g. [22], p. 105). □

4 Particular examples

In Section 3.3 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 this section satisfy them. We present in this section examples of sets C and functions f which satisfy H3(a) or H3(b), i.e. covering situations in which T is not the subdifferential of a convex functions and where all solutions of VIP(T,C) belong to ∂C.

4.1 Penalization in polyhedra

This subsection is devoted to the particular case when C is a polyhedron. We present an example of an $f \in \mathcal{F}$ satisfying all required properties, including H3(a). 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

$$C = \{x \in B \mid \langle v^i, x \rangle \geq \alpha_i,\ i = 1, ..., p\}, \tag{15}$$

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

$$h(x) = \begin{cases} f_2(x) + \sum_{i=1}^{p} \left(\langle v^i, x \rangle - \alpha_i\right) \log\left(\langle v^i, x \rangle - \alpha_i\right) & \text{if } x \in C \\ \infty & \text{otherwise,} \end{cases} \tag{16}$$

where $f_2(x) = \frac{1}{2} \|x\|_B^2$. We state next some properties of f_2.

Proposition 4.1. *If B is uniformly smooth and uniformly convex, then f_2 satisfies H1, H2, H3 and H4.*

Proof. See Proposition 2 of [16]. $\square$

Proposition 4.2. *Let B be uniformly smooth and uniformly convex and take h as in* (16)*. Then*

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

$$h'(x) = f_2'(x) + \sum_{i=1}^{p}[1 + log\left(\langle v^i, x\rangle - \alpha_i\right)]v^i. \tag{17}$$

ii) h satisfies H1, H2, H3(a), H3(c), H4 and H6.

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 $\eta : \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$, by $\eta(z) = \sum_{\ell=1}^{p} z_\ell \log z_\ell$, continuously extended to the boundary of $\mathbb{R}^p_+$ with the convention that $0 \log 0 = 0$. Then we have $h = \eta \circ \zeta + f_2$, where $\zeta : B \to \mathbb{R}^p$ is the application with ζ_i as its i-th component, ($1 \le i \le p$). Thus, the results presented here are consequences of [16], Proposition 2 together with the properties of the Bregman function η. In fact, strict convexity of h 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 of derivatives in view of the locally uniform convexity of B. Thus,

$$h'(x) = f_2' + \sum_{i=1}^{p}[1 + \log(\zeta_i(x))]v^i, \tag{18}$$

and (i) holds. Observe also that

$$D_h(y, x) = D_\eta(\zeta(y), \zeta(x)) + D_{f_2}(y, x), \tag{19}$$

so that H1 and H2 are direct consequences of properties H1 and H2 of f_2 (Proposition 4.1). Concerning H3(a), take $\{x^k\}$, $\{y^k\} \subset \text{int}(C)$ such that $x^k \xrightarrow[k\to\infty]{\text{w}} x^\infty$, $y^k \xrightarrow[k\to\infty]{\text{w}} x^\infty$ and $\lim_{k\to\infty} D_h(y^k, x^k) = 0$. Since h satisfies H2, $x^k - y^k \xrightarrow[k\to\infty]{\text{s}} 0$. In view of the three-point equality, Proposition 2.5(ii), 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,

$$\begin{aligned}\lim_{k\to\infty} |D_{f_2}(x^\infty, x^k) - D_{f_2}(x^\infty, y^k)| &\le \lim_{k\to\infty} D_{f_2}(y^k, x^k) + \lim_{k\to\infty} |\langle f_2'(x^k) - f_2'(y^k), y^k - x^\infty\rangle| \\ &\le \lim_{k\to\infty} \left\| f_2'(x^k) - f_2'(y^k)\right\|_* \left\| y^k - x^\infty\right\| = 0,\end{aligned} \tag{20}$$

using boundedness of the weakly convergent sequence $\{y^k\}$ and the fact that f_2 satisfies H3 (Proposition 4.1). Note also that $\zeta : B \to \mathbb{R}^p$ is weakly continuous, so that $\{\zeta(x^k)\}$ and $\{\zeta(y^k)\}$ are sequences in $\mathbb{R}^p$ both of which converge to $\zeta(x^\infty)$. Thus, in view of Lemma 2.1.3 of [9], 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

$$\lim_{k\to\infty} \Big[D_\eta(\zeta(x^\infty), \zeta(x^k)) - D_\eta(\zeta(x^\infty), \zeta(y^k))\Big] = 0. \tag{21}$$

It follows from equations (20) and (21) that h satisfies H3(a).

H4 is a consequence of Lemma 2.10 and Corollary 3.1 of [5] applied to the operator $T = \partial[\eta \circ \zeta] : B \to \mathcal{P}(B^*)$, which is maximal monotone, and to the regularizing function f_2, which satisfies H4 by Proposition 4.1, after observing that int(dom f_2)$\cap$ dom $T \neq \emptyset$.

For proving H6, choose $\{z^k\} \subset \text{int}(C)$ such that $\lim_{k\to\infty} d(z^k, \partial C) = 0$, with $\{z^k\}$ bounded. Observe first that in view of equation (19), for all $i \in \{1, \ldots, p\}$ it holds that

$$\begin{aligned} D_h(w, z^k) &= D_{f_2}(w, z^k) + \sum_{i=1}^{p} D_{\eta_i}(\zeta_i(w), \zeta_i(z^k)) \geq D_{\eta_i}(\zeta_i(w), \zeta_i(z^k)) \\ &= \zeta_i(w) \log\left(\frac{\zeta_i(w)}{\zeta_i(z^k)}\right) + \zeta_i(z^k) - \zeta_i(w). \end{aligned} \tag{22}$$

Assume that for some $w \in \text{int}(C)$ the sequence $\{D_h(w, z^k)\}$ does not diverge to infinity. Then, there exists a subsequence $\{z^{k_j}\}$ of $\{z^k\}$ and a real number M, such that $D_h(w, z^{k_j}) \leq M$ for all j. Since for any $x \in \text{int}(C)$, $\zeta_i(x) > 0$, in view of (22), it necessarily holds that

$$\liminf_j \zeta_i(z^{k_j}) > 0, \tag{23}$$

for all $i \in \{1, ..., p\}$. Note also that there exists a sequence $\{y^k\} \in \partial C$ satisfying $z^k - y^k \xrightarrow[k\to\infty]{s} 0$. 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,

$$\zeta(z^{k'}) = \zeta(z^{k'}) - \zeta_i(y^{k'}) = \langle v^i, z^{k'} - y^{k'} \rangle. \tag{24}$$

We conclude from (24) that $\lim_{k'} \zeta_i(z^{k'}) = 0$, in contradiction with (23). Hence, $\lim_{k\to\infty} D_h(w, z^k) = +\infty$ for all $w \in \text{int}(C)$.

Concerning H3(c), let $A \subset \text{int}(C)$ be a 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 $\{x^k\}, \{y^k\} \subset A$, 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 by Propostion 4.1, we also have $f_2'(x^k) - f_2'(y^k) \xrightarrow[k\to\infty]{s} 0$. From (18), we get

$$\begin{aligned} \left\| c'(x^k) - c'(y^k) \right\|_* &= \left\| f_2'(x^k) - f_2'(y^k) + \sum_{i=1}^{p} [\log(\zeta_i(x^k)) - \log(\zeta_i(y^k))] v^i \right\|_* \\ &\leq \left\| f_2'(x^k) - f_2'(y^k) \right\|_* + \sum_{i=1}^{p} \left| \log\left(\frac{\zeta_i(x^k))}{\zeta_i(y^k)} \right) \right| \left\| v^i \right\|_* \xrightarrow{k} 0. \end{aligned} \tag{25}$$

It follows from (25) that h' is uniformly continuous on A. □

Remark 4.3. If we replace in the definition (16) of h the function f_2 by f_r, defined as $f_r(x) = \|x\|^r$ with $r > 1$, the results of Proposition 4.2 still hold, because f_r also satisfies H1–H4 (see [16], Proposition 2).

Remark 4.4. Concerning weak-to-weak* continuity, we know from [3] that when B is a Hilbert space then h', as defined in (16), has this property. The same holds when $B = \ell_p$ and f_2 is replaced by $f_p = \|\cdot\|_p^p$ in (16). In fact, in this case f_p' is weak-to-weak* continuous (see [2], Proposition 8.2), so that, in view of (17), h' is weak-to-weak* continuous too. For other infinite dimensional Banach spaces h should fail to satisfy H5, in view of the counterexamples for f_r presented in [7].

Corollary 4.5. *Let B be a uniformly convex and uniformly smooth Banach space, and take C as defined in* (15) *and $f = h$, with h as in* (16)*. Assume that VIP(T,C) has solutions. If T is pseudo- and paramonotone then the sequence $\{x^k\}$ generated by the algorithm* (5)–(7) *has weak accumulation points and all of them are solutions of VIP(T,C). Moreover, if $B = H$ is a Hilbert space, or $B = \ell_p$ and f_2 is replaced by f_p in* (16)*, then the sequence $\{x^k\}$ is weakly convergent to a solution of VIP(T,C).*

Proof. By Proposition 4.2 and Remarks 4.3 and 4.4, f satisfies the hypotheses required in Theorem 3.5 and Proposition 3.6 for the announced results to hold. □

4.2 An approach depending on the regularizing parameters

Let us go back to the convergence analysis of the algorithm (5)–(7), i.e. to the proof of Theorem 3.5. Equation (12) 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 the results in Proposition 3.3 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, (12) ensures that

$$\limsup_k \langle v^k, \tilde{x}^k - x^\infty\rangle = \limsup_k \lambda_k \left[D_f(x^\infty, x^k) - D_f(x^\infty, x^{k+1}) \right], \tag{26}$$

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) seems appropriate when a pseudomonotonicity argument is being used, but in its absence, we can enforce the same effect by playing with the sequence of regularizing parameters $\{\lambda_k\}$, and using assumptions H3(b) and H3(c).

Theorem 4.6. *Let $\{x^k\}$ be the sequence generated by the algorithm* (5)–(7)*, with $f \in \mathcal{F}$ satisfying H1, H2, H3(b), H3(c), H4 and H6, and such that dom $f = C$. 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 $\{x^k\}$ is bounded and all its weak accumulation points are solutions of VIP(T,C).*

Proof. Let $\{x^k\}$ be the sequence generated by the algorithm, which is bounded by Proposition 3.3(ii). If $S \cap \mathrm{int}(C) \neq \emptyset$, the result follows from Proposition 3.6, wich uses H3(c), but neither pseudo- nor paramonotonicity, nor H3(b). Thus, we can assume that

$$S \cap \mathrm{int}(C) = \emptyset. \tag{27}$$

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 (27). 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 as in H3(b), it holds that

$$\limsup_k \lambda_k \left[D_f(x^\infty, x^k) - D_f(x^\infty, x^{k+1})\right] \leq 0. \tag{28}$$

In view of Proposition 3.3 and (28), taking a subsequence $\{x^{j_k}\}$ of $\{x^k\}$ such that $x^{j_k} \xrightarrow[k\to\infty]{\mathrm{w}} x^\infty$, we get from (12) that $\limsup_k \langle v^{j_k}, \tilde{x}^{j_k} - x^\infty\rangle \leq 0$. We complete the proof as in Theorem 3.5 after (12), using pseudo- and paramonotonicity of T. $\square$

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, causing numerical unstability if the problem VIP(T,C) is ill conditioned.

In the case in which C is a closed ball, we know of no function satisfying both conditions H3(a) and H6, but we do have functions satisfying H3(b) instead. We look at this case in the following subsection.

4.3 Penalization in balls

We present next a particular case for which a function satisfying H3(b) is available (but not one satisfying H3(a)). Let $C := \{x \in B : \|x\| \leq 1\}$ and consider the function $g : B \to \mathbb{R} \cup \{+\infty\}$, defined as

$$g(x) = \begin{cases} 1 - \sqrt{1 - \|x\|^2} & \text{if } \|x\| \leq 1 \\ \infty & \text{otherwise.} \end{cases} \tag{29}$$

The following proposition establishes some properties of this function.

Proposition 4.7. *Let B be a smooth real reflexive Banach space. Take g as in (29) and $f_2(x) = (1/2)\|x\|^2$. Then*

i) g satisfies H1.

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

$$g'(x) = \frac{1}{\sqrt{1-\|x\|^2}} f_2'(x) \tag{30}$$

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).

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

$$\nu_g(x,t) \geq \nu_{f_2}(x,t). \tag{31}$$

v) g satisfies H6.

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

$$\varphi(t) = \frac{t}{\sqrt{1-t^2}}. \tag{32}$$

Then $g = \psi \circ \|\cdot\|_B$, with $\psi(t) = \int_0^t \varphi(s)ds$ if $t \in [0,1]$ and $\psi(t) = \infty$ otherwise. Consider also the (extended) duality mapping $J_\varphi : B \to B^*$, defined by

$$J_\varphi(x) = \begin{cases} \{x^* \in B^* | \langle x^*, x\rangle = \|x^*\|_* \|x\|, \|x^*\|_* = \varphi(\|x\|)\} & \text{if } \|x\| < 1 \\ \emptyset & \text{otherwise.} \end{cases}$$

Therefore, $J_\varphi(x) = \partial[\psi \circ \|\cdot\|](x)$ for all $x \in B$, and we obtain (ii) and (iii) in view of known properties of duality mappings (see e.g. [11]). Item (i) is trivial, since C is bounded.

For (iv), choose f as the restriction of f_2 to the interior of C and observe that $g = \Phi \circ f_2$, where $\Phi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is defined as $\Phi(t) = 1 - \sqrt{1-2t}$, if $t \in (-\infty, 1/2]$, and $\Phi(t) = \infty$ otherwise. Since Φ is strictly convex and $\Phi'(t) \geq \Phi'(0) = 1$ for all $t \in [0, 1/2)$, we can use total convexity of f_2 to get $\nu_g(x,t) \geq \Phi'(f_2(x))\nu_{f_2}(x,t) \geq \nu_{f_2}(x,t)$ (see [7], 1.2.7).

In order to prove (v), choose $\{x^k\} \subset \mathrm{int}(C)$, $x \in \mathrm{int}(C) = \{x \in B|\ \|x\| < 1\}$ with $\lim_{k\to\infty} \left\|x^k\right\| = 1$. Note that

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

Observe that $\|x\| < 1$, and therefore $\left\|x^k\right\| - \|x\| \geq \alpha > 0$ for large enough k. The result then follows from the fact that $\left\|g'(x^k)\right\|_* = \varphi(\left\|x^k\right\|)$. □

We present next an example showing that, in general, g does not satisfy H3(a).

Example 4.8. 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 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

$$||x^k||_p = \mu_k < 1 \quad \text{and} \quad ||y^k||_p = \delta_k < 1 \tag{33}$$

and that

$$x^k \xrightarrow[k\to\infty]{\mathrm{w}} 2^{-\frac{1}{p}}e^1 \quad \text{and} \quad y^k \xrightarrow[k\to\infty]{\mathrm{w}} 2^{-\frac{1}{p}}e^1. \tag{34}$$

Thus, we have found two sequences, namely $\{x^k\}$ and $\{y^k\}$, contained in $\mathrm{int}(C)$, 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

$$\lim_{k\to\infty} D_g(x^k, y^k) = 0 \quad \text{and} \quad \lim_{k\to\infty} \left[D_g(x, x^k) - D_g(x, y^k)\right] = -\infty, \tag{35}$$

which implies that g does not satisfy H3(a). Consequently, g does not satisfy H3 either.

We prove next that g satisfies H3(b).

Proposition 4.9. *Let B be a strictly convex and smooth reflexive Banach space and take g as given by* (29)*. 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. $\lim_k \left\|x^k\right\| = 1$), the sequence of parameters $\{\rho_k\} \subset \mathbb{R}_{++}$, can be taken as*

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

Proof. Take $\{x^k\} \subset \mathrm{int}(C) = B(0,1)$, with $\lim_k d(x^k, \partial C) = 0$. Let $\hat{x}$ be a any weak accumulation point of $\{x^k\}$. Take also $\{y^k\} \subset \mathrm{int}(C)$. Then, with φ as in (32), we get from Proposition 4.7(ii) and Definition 2.4 that

$$D_g(\hat{x}, x^k) - D_f(\hat{x}, y^k) \leq D_g(\hat{x}, x^k) = \sqrt{1 - \|x^k\|^2} - \sqrt{1 - \|\hat{x}\|^2} - \langle g'(x^k), \hat{x} - x^k\rangle$$

$$\leq \sqrt{1 - \|x^k\|^2} + \langle g'(x^k), x^k\rangle - \langle g'(x^k), \hat{x}\rangle \leq \sqrt{1 - \|x^k\|^2} + \varphi(\left\|x^k\right\|)\left\|x^k\right\| + \varphi(\left\|x^k\right\|)\|\hat{x}\|$$

$$\leq \sqrt{1 - \|x^k\|^2} + 2\varphi(\left\|x^k\right\|).$$

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

$$\limsup_k \rho_k \left[D_g(\hat{x}, x^k) - D_g(\hat{x}, y^k)\right] \leq \limsup_k \rho_k \left[\sqrt{1 - \|x^k\|^2} + 2\varphi(\left\|x^k\right\|)\right]$$

$$\leq \limsup_k 2(1 - \left\|x^k\right\|)\frac{\|x^k\|}{\sqrt{1 - \|x^k\|^2}} \leq 2\limsup_k \sqrt{\frac{(1 - \|x^k\|)^2}{1 - \|x^k\|^2}}$$

$$= 2\limsup_k \sqrt{\frac{1 - \|x^k\|}{1 + \|x^k\|}} \leq 2\limsup_k \sqrt{1 - \|x^k\|} = 0.$$

□

Remark 4.10. Note that all results in this section hold trivially for any ball $B[0,\beta]$ (i.e. with β 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 functions of the form $g(x) = b - (b^r - \|x\|^r)^{1/r}$ with $r > 1$. 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 (29). 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,$B[x_0,\beta]$) can be reduced to VIP($\tilde{T}$,$B[0,\beta]$) with $\tilde{T}(x) = T(x+x_0)$. Thus, it is not worthwhile to go deeper in the analysis of this case.

Corollary 4.11. *Let B be uniformly convex and uniformly smooth. Consider $C = B[0,1]$ and assume that VIP(T,B$[0,1]$) has solutions, that all of them lie on the boudary of C and that T is pseudo- and paramonotone. Take the regularizing function f as defined by* (29) *and assume that, at step k, 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, the whole sequence $\{x^k\}$ converges strongly to a point in $S \cap \partial C$.*

Proof. Take $C = B[0,1]$ and f as defined in (29). From Proposition 4.7 and Proposition 4.9 we know that dom $f = C$ and that f satisfies H1, H2, H3(c), H4 and H6, and also H3(b) with any $\rho_k \leq 1 - \left\|x^k\right\|$. Then, we can apply Theorem 4.6 with $\lambda_k = \rho_k \leq 1 - \left\|x^k\right\|$, to conclude that $\{x^k\}$ has weak accumulation points and all of them are solutions of VIP(T,C).

Recall that we are assuming 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 because $S = [T + N_C]^{-1}(0)$ (see, e.g. [22], p. 105), strict convexity of B implies that S is a singleton, say $\{\bar{x}\}$. 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. $\lim_{k\to\infty} \left\|x^k\right\|_B = \|\bar{x}\|_B$ and $x^k \xrightarrow[k\to\infty]{\text{w}} \bar{x}$. Since B is locally uniformly convex, it satisfies property (h), as defined in p. 49 of [11], and it follows that the convergence is strong, i.e. $x^k \xrightarrow[k\to\infty]{\text{s}} \bar{x}$. $\square$

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 $\lim_k \lambda_k = 0$, and the regularization effect becomes indeed negligeable for large enough k.

5 Concluding remarks

We summarize next some relevant features of the results presented in this work and comment on 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 in reflexive Banach spaces, extending its applicability domain, originally restricted to finite dimensional spaces ([26]).

Open issues:

- It is relevant to find examples of regularizing funcions whose domains are closed and convex sets other than balls and polyhedra.
- For the situation where no regularizing function satisfying H3(a) is known, H3(b) is not totally satisfactory. Is there any way to prove the convergence properties of the method allowing for regularizing parameters bounded away from zero?

References

[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] 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.

[3] Burachik, R. S. *Generalized proximal point methods for the variational inequality problem.* PhD thesis, IMPA, Rio de Janeiro, 1995.

[4] 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.

[5] Burachik, R. S., and Scheimberg, S. A proximal point algorithm for the variational inequality problem in Banach spaces. *SIAM J. Control Optim. 39*, 5 (2001), 1633–1649.

[6] 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.

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

[8] Censor, Y., and Zenios, S. A. Proximal minimization algorithm with D-functions. *J. Optim. Theory Appl. 73*, 3 (1992), 451–464.

[9] Censor, Y., and Zenios, S. A. *Parallel optimization. Theory, algorithms, and applications.* Oxford University Press, New York, 1997.

[10] Chen, G., and Teboulle, M. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM J. Optim. 3*, 3 (1993), 538–543.

[11] Cioranescu, I. *Geometry of Banach spaces, duality mappings and nonlinear problems.* Kluwer Academic Publishers Group, Dordrecht, 1990.

[12] Eckstein, J. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Math. Oper. Res. 18*, 1 (1993), 202–226.

[13] Eggermont, P. P. B. Multiplicative iterative algorithms for convex programming. *Linear Algebra Appl. 130* (1990), 25–42.

[14] Eriksson, J. An iterative primal-dual algorithm for linear programming. Tech. Rep. 85-10, Department of Mathematics, Linköping University, 1985.

[15] Erlander, S. Entropy in linear programs. *Math. Programming 21*, 2 (1981), 137–151.

[16] Gárciga Otero, R., and Iusem, A. N. Inexact versions of proximal point and augmented Lagrangian algorithms in Banach spaces. *Numer. Funct. Anal. Optim. 22*, 5-6 (2001), 609–640.

[17] Iusem, A. N. Some properties of generalized proximal point methods for quadratic and linear programming. *J. Optim. Theory Appl. 85*, 3 (1995), 593–612.

[18] Iusem, A. N. On some properties of generalized proximal point methods for variational inequalities. *J. Optim. Theory Appl. 96*, 2 (1998), 337–362.

[19] Iusem, A. N. On some properties of paramonotone operators. *J. Convex Anal. 5*, 2 (1998), 269–278.

[20] Kassay, G. The proximal points algorithm for reflexive Banach spaces. *Studia Univ. Babeş-Bolyai Math. 30* (1985), 9–17.

[21] Kiwiel, K. C. Proximal minimization methods with generalized Bregman functions. *SIAM J. Control Optim. 35*, 4 (1997), 1142–1168.

[22] Pascali, D., and Sburlan, S. *Nonlinear mappings of monotone type.* Martinus Nijhoff Publishers, The Hague, 1978.

[23] Phelps, R. R. *Convex functions, monotone operators and differentiability.* Springer-Verlag, Berlin, 1989.

[24] Rockafellar, R. T. Local boundedness of nonlinear, monotone operators. *Michigan Math. J. 16* (1969), 397–407.

[25] Silva, G. J. P., and Svaiter, B. F. An inexact generalized proximal point method in Banach spaces. To appear.

[26] Solodov, M. V., and Svaiter, B. F. An inexact hybrid generalized proximal point algorithms and some new results on the theory of Bregman functions. *Math. Oper. Res. 51* (2000), 479–494.