

Bregman Forward-Backward Operator Splitting*

Minh N. Bui and Patrick L. Combettes

North Carolina State University, Department of Mathematics, Raleigh, NC 27695-8205, USA
mnbui@ncsu.edu and plc@math.ncsu.edu

Dedicated to Terry Rockafellar on the occasion of his 85th birthday

Abstract. We establish the convergence of the forward-backward splitting algorithm based on Bregman distances for the sum of two monotone operators in reflexive Banach spaces. Even in Euclidean spaces, the convergence of this algorithm has so far been proved only in the case of minimization problems. The proposed framework features Bregman distances that vary over the iterations and a novel assumption on the single-valued operator that captures various properties scattered in the literature. In the minimization setting, we obtain rates that are sharper than existing ones.

Keywords. Banach space, Bregman distance, forward-backward splitting, Legendre function, monotone operator.

1 Introduction

Throughout, $\mathcal{X}$ is a reflexive real Banach space with topological dual $\mathcal{X}^*$. We are concerned with the following monotone inclusion problem (see Section 2.1 for notation and definitions).

Problem 1.1 Let $A: \mathcal{X} \rightarrow 2^{\mathcal{X}^*}$ and $B: \mathcal{X} \rightarrow 2^{\mathcal{X}^*}$ be maximally monotone, let $f \in \Gamma_0(\mathcal{X})$ be essentially smooth, and let D_f be the Bregman distance associated with f . Set $C = (\text{int dom } f) \cap \text{dom } A$ and $\mathcal{S} = (\text{int dom } f) \cap \text{zer}(A + B)$. Suppose that $C \subset \text{int dom } B$, $\mathcal{S} \neq \emptyset$, B is single-valued on $\text{int dom } B$, and there exist $\delta_1 \in [0, 1[$, $\delta_2 \in [0, 1]$, and $\kappa \in [0, +\infty[$ such that

$$\begin{aligned} (\forall x \in C)(\forall y \in C)(\forall z \in \mathcal{S})(\forall y^* \in Ay)(\forall z^* \in Az) \\ \langle y - x, By - Bz \rangle \leq \kappa D_f(x, y) + \langle y - z, \delta_1(y^* - z^*) + \delta_2(By - Bz) \rangle. \end{aligned} \quad (1.1)$$

The objective is to

$$\text{find } x \in \text{int dom } f \text{ such that } 0 \in Ax + Bx. \quad (1.2)$$

The central problem (1.2) has extensive connections with various areas of mathematics and its applications. In Hilbert spaces, if B is cocoercive, a standard method for solving (1.2) is the forward-backward algorithm, which operates with the update $x_{n+1} = (\text{Id} + \gamma A)^{-1}(x_n - \gamma Bx_n)$ [17]. This

*Contact author: P. L. Combettes, plc@math.ncsu.edu, phone: +1 (919) 515 2671. This work was supported by the National Science Foundation under grant DMS-1818946.

iteration is not applicable beyond Hilbert spaces since A maps to $\mathcal{X}^* \neq \mathcal{X}$. In addition, there has been a significant body of work (see, e.g., [3, 6, 8, 12, 13, 16, 18, 19, 23]) showing the benefits of replacing standard distances by Bregman distances, even in Euclidean spaces. Given a sequence $(\gamma_n)_{n \in \mathbb{N}}$ in $]0, +\infty[$ and a suitable sequence of differentiable convex functions $(f_n)_{n \in \mathbb{N}}$, we propose to solve (1.2) via the iterative scheme

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = (\nabla f_n + \gamma_n A)^{-1}(\nabla f_n(x_n) - \gamma_n Bx_n), \quad (1.3)$$

which consists of first applying a forward (explicit) step involving B and then a backward (implicit) step involving A . Let us note that the convergence of such an iterative process has not yet been established, even in finite-dimensional spaces with a single function $f_n = f$ and constant parameters $\gamma_n = \gamma$. Furthermore, the novel scheme (1.3) will be shown to unify and extend several iterative methods which have thus far not been brought together:

- The Bregman monotone proximal point algorithm

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = (\nabla f + \gamma_n A)^{-1}(\nabla f(x_n)) \quad (1.4)$$

of [6] for finding a zero of A in $\text{int dom } f$, where f is a Legendre function.

- The variable metric forward-backward splitting method

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = (U_n + \gamma_n A)^{-1}(U_n x_n - \gamma_n Bx_n) \quad (1.5)$$

of [15] for finding a zero of $A + B$ in a Hilbert space, where $(U_n)_{n \in \mathbb{N}}$ is a sequence of strongly positive self-adjoint bounded linear operators.

- The splitting method

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = (\nabla f_n + \gamma_n \partial \varphi)^{-1}(\nabla f_n(x_n) - \gamma_n \nabla \psi(x_n)) \quad (1.6)$$

of [18] for finding a minimizer of the sum of the convex functions φ and ψ in $\text{int dom } f$.

- The Renaud–Cohen algorithm

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = (\nabla f + \gamma A)^{-1}(\nabla f(x_n) - \gamma Bx_n) \quad (1.7)$$

of [20] for finding a zero of $A + B$ in a Hilbert space, where f is real-valued and strongly convex.

Problems which cannot be solved by algorithms (1.4)–(1.7) will be presented in Example 2.9 as well as in Sections 3.2 and 3.4. New results on the minimization setting will be presented in Section 3.3.

The goal of the present paper is to investigate the asymptotic behavior of (1.3) under mild conditions on A , B , and $(f_n)_{n \in \mathbb{N}}$. Let us note that the convergence proof techniques used in the above four frameworks do not extend to (1.3). For instance, the tools of [18] rely heavily on functional inequalities involving φ and ψ . On the other hand, the approach of [15] exploits specific properties of quadratic kernels in Hilbert spaces, while [6] relies on Bregman monotonicity properties of the iterates that will no longer hold in the presence of B . Finally, the proofs of [20] depend on the strong convexity of f , the underlying Hilbertian structure, and the fact that the updating equation is governed by a fixed operator. Our analysis will not only capture these frameworks but also provide new methods to solve problems beyond their reach. It hinges on the theory of Legendre functions and the new condition (1.1), which will be seen to cover in particular various properties such as the cocoercivity assumption used in the standard forward-backward method in Hilbert spaces [7, 17], as well as the seemingly unrelated assumptions used in [6, 15, 18, 20] to study (1.4)–(1.7).

The main result on the convergence of (1.3) is established in Section 2 for the general scenario described in Problem 1.1. Section 3 is dedicated to special cases and applications. In the context of minimization problems, convergence rates on the worst behavior of the method are obtained.

2 Main results

2.1 Notation and definitions

The norm of $\mathcal{X}$ is denoted by $\|\cdot\|$ and the canonical pairing between $\mathcal{X}$ and $\mathcal{X}^*$ by $\langle \cdot, \cdot \rangle$. If $\mathcal{X}$ is Hilbertian, its scalar product is denoted by $\langle \cdot | \cdot \rangle$. The symbols $\rightharpoonup$ and $\rightarrow$ denote respectively weak and strong convergence. The set of weak sequential cluster points of a sequence $(x_n)_{n \in \mathbb{N}}$ in $\mathcal{X}$ is denoted by $\mathfrak{W}(x_n)_{n \in \mathbb{N}}$.

Let $M: \mathcal{X} \rightarrow 2^{\mathcal{X}^*}$ be a set-valued operator. Then $\text{gra } M = \{(x, x^*) \in \mathcal{X} \times \mathcal{X}^* \mid x^* \in Mx\}$ is the graph of M , $\text{dom } M = \{x \in \mathcal{X} \mid Mx \neq \emptyset\}$ the domain of M , $\text{ran } M = \{x^* \in \mathcal{X}^* \mid (\exists x \in \mathcal{X}) x^* \in Mx\}$ the range of M , and $\text{zer } M = \{x \in \mathcal{X} \mid 0 \in Mx\}$ the set of zeros of M . Moreover, M is monotone if

$$(\forall (x_1, x_1^*) \in \text{gra } M) (\forall (x_2, x_2^*) \in \text{gra } M) \quad \langle x_1 - x_2, x_1^* - x_2^* \rangle \geq 0, \quad (2.1)$$

and maximally monotone if, furthermore, there exists no monotone operator from $\mathcal{X}$ to $2^{\mathcal{X}^*}$ the graph of which properly contains $\text{gra } M$.

A function $f: \mathcal{X} \rightarrow]-\infty, +\infty]$ is coercive if $\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty$ and supercoercive if $\lim_{\|x\| \rightarrow +\infty} f(x)/\|x\| = +\infty$. $\Gamma_0(\mathcal{X})$ is the class of lower semicontinuous convex functions $f: \mathcal{X} \rightarrow]-\infty, +\infty]$ such that $\text{dom } f = \{x \in \mathcal{X} \mid f(x) < +\infty\} \neq \emptyset$. Now let $f \in \Gamma_0(\mathcal{X})$. The conjugate of f is the function $f^* \in \Gamma_0(\mathcal{X}^*)$ defined by $f^*: \mathcal{X}^* \rightarrow]-\infty, +\infty]$: $x^* \mapsto \sup_{x \in \mathcal{X}} (\langle x, x^* \rangle - f(x))$, and the subdifferential of f is the maximally monotone operator

$$\partial f: \mathcal{X} \rightarrow 2^{\mathcal{X}^*}: x \mapsto \{x^* \in \mathcal{X}^* \mid (\forall y \in \mathcal{X}) \langle y - x, x^* \rangle + f(x) \leq f(y)\}. \quad (2.2)$$

In addition, f is a Legendre function if it is essentially smooth in the sense that ∂f is both locally bounded and single-valued on its domain, and essentially strictly convex in the sense that ∂f^* is locally bounded on its domain and f is strictly convex on every convex subset of $\text{dom } \partial f$ [5]. Suppose that f is Gâteaux differentiable on $\text{int dom } f \neq \emptyset$. The Bregman distance associated with f is

$$D_f: \mathcal{X} \times \mathcal{X} \rightarrow [0, +\infty] \\ (x, y) \mapsto \begin{cases} f(x) - f(y) - \langle x - y, \nabla f(y) \rangle, & \text{if } y \in \text{int dom } f; \\ +\infty, & \text{otherwise.} \end{cases} \quad (2.3)$$

Given $\alpha \in]0, +\infty[$, we define

$$\mathcal{C}_\alpha(f) = \{g \in \Gamma_0(\mathcal{X}) \mid \text{dom } g = \text{dom } f, g \text{ is Gâteaux differentiable on } \text{int dom } f, D_g \geq \alpha D_f\}. \quad (2.4)$$

2.2 On condition (1.1)

The following proposition provides several key illustrations of the pertinence of (1.1) in terms of capturing concrete scenarios.

Proposition 2.1 *Consider the setting of Problem 1.1. Then (1.1) holds in each of the following cases:*

- (i) $\delta_1 \in [0, 1[$, $\delta_2 = 1$, and $(\forall x \in C)(\forall y \in C)(\forall z \in \mathcal{S}) \langle z - x, By - Bz \rangle \leq \kappa D_f(x, y)$.
- (ii) $\delta_1 = 0$, $\delta_2 = 1$, and $B = \partial\psi$, where $\psi \in \Gamma_0(\mathcal{X})$ satisfies

$$(\forall x \in C)(\forall y \in C)(\forall z \in \mathcal{S}) \quad D_\psi(x, y) \leq \kappa D_f(x, y) + D_\psi(x, z) + D_\psi(z, y). \quad (2.5)$$

(iii) $\delta_1 = 0$, $\delta_2 = 1$, and there exists $\psi \in \Gamma_0(\mathcal{X})$ such that $B = \partial\psi$ and $(\forall x \in C)(\forall y \in C) D_\psi(x, y) \leq \kappa D_f(x, y)$.

(iv) $\text{dom } B = \mathcal{X}$, there exists $\beta \in]0, +\infty[$ such that

$$(\forall (x, x^*) \in \text{gra}(A + B))(\forall (y, y^*) \in \text{gra}(A + B)) \quad \langle x - y, x^* - y^* \rangle \geq \beta \|Bx - By\|^2, \quad (2.6)$$

f is Fréchet differentiable on $\mathcal{X}$, ∇f is α -strongly monotone on $\text{dom } A$ for some $\alpha \in]0, +\infty[$, $\varepsilon \in]0, 2\beta[$, $\kappa = 1/(\alpha(2\beta - \varepsilon))$, and $\delta_1 = \delta_2 = (2\beta - \varepsilon)/(2\beta)$.

(v) $A + B$ is strongly monotone with constant $\mu \in]0, +\infty[$, B is Lipschitzian on $\text{dom } B = \mathcal{X}$ with constant $\nu \in]0, +\infty[$, f is Fréchet differentiable on $\mathcal{X}$, ∇f is α -strongly monotone on $\text{dom } A$ for some $\alpha \in]0, +\infty[$, $\varepsilon \in]0, 2\mu/\nu^2[$, $\kappa = \nu^2/(\alpha(2\mu - \varepsilon\nu^2))$, and $\delta_1 = \delta_2 = (2\mu - \varepsilon\nu^2)/(2\mu)$.

(vi) $\text{dom } B = \mathcal{X}$, $\beta \in]0, +\infty[$, f is Fréchet differentiable on $\mathcal{X}$, ∇f is α -strongly monotone on $\text{dom } A$ for some $\alpha \in]0, +\infty[$, $\varepsilon \in]0, 2\beta[$, $\kappa = 1/(\alpha(2\beta - \varepsilon))$, $\delta_1 = 0$, $\delta_2 = (2\beta - \varepsilon)/(2\beta)$, and one of the following is satisfied:

[a] B is β -cocoercive, i.e.,

$$(\forall x \in \mathcal{X})(\forall y \in \mathcal{X}) \quad \langle x - y, Bx - By \rangle \geq \beta \|Bx - By\|^2. \quad (2.7)$$

[b] B is ν -Lipschitzian for some $\nu \in]0, +\infty[$, and angle bounded with constant $1/(4\beta\nu)$, i.e.,

$$(\forall x \in \mathcal{X})(\forall y \in \mathcal{X})(\forall z \in \mathcal{X}) \quad \langle y - z, Bz - Bx \rangle \leq \frac{1}{4\beta\nu} \langle x - y, Bx - By \rangle. \quad (2.8)$$

[c] B is $(1/\beta)$ -Lipschitzian and there exists $\psi \in \Gamma_0(\mathcal{X})$ such that $B = \nabla\psi$.

Proof. (i): Let $x \in C$, $y \in C$, and $z \in \mathcal{S}$. Then $\langle y - x, By - Bz \rangle = \langle z - x, By - Bz \rangle + \langle y - z, By - Bz \rangle \leq \kappa D_f(x, y) + \langle y - z, \delta_2(By - Bz) \rangle$. In view of the monotonicity of A , we obtain (1.1).

(ii) $\Rightarrow$ (i): In the light of [9, Proposition 4.1.5 and Corollary 4.2.5], ψ is Gâteaux differentiable on $\text{int dom } \psi$ and $B = \nabla\psi$ on $\text{int dom } \psi = \text{int dom } B \supset C$. Hence, we derive from (2.5), (2.3), and [6, Proposition 2.3(ii)] that

$$(\forall x \in C)(\forall y \in C)(\forall z \in \mathcal{S}) \quad \kappa D_f(x, y) \geq D_\psi(x, y) - D_\psi(x, z) - D_\psi(z, y) = \langle z - x, By - Bz \rangle. \quad (2.9)$$

(iii) $\Rightarrow$ (ii): Clear.

(iv): It results from [9, Theorem 4.2.10] that ∇f is continuous. Thus, using the strong monotonicity of ∇f on $\text{dom } A$, we obtain

$$(\forall x \in \overline{\text{dom } A})(\forall y \in \overline{\text{dom } A}) \quad \langle x - y, \nabla f(x) - \nabla f(y) \rangle \geq \alpha \|x - y\|^2. \quad (2.10)$$

Given x and y in $\overline{\text{dom } A}$, define $\phi: \mathbb{R} \rightarrow \mathbb{R}: t \mapsto f(y + t(x - y))$, and observe that, since $\overline{\text{dom } A}$ is convex [24, Theorem 3.11.12], $[x, y] \subset \overline{\text{dom } A}$ and therefore (2.10) yields

$$\begin{aligned} D_f(x, y) &= \int_0^1 \phi'(t) dt - \langle x - y, \nabla f(y) \rangle \\ &= \int_0^1 \langle x - y, \nabla f(y + t(x - y)) - \nabla f(y) \rangle dt \end{aligned}$$

$$\begin{aligned}
&\geq \int_0^1 t\alpha\|x-y\|^2 dt \\
&= \frac{\alpha}{2}\|x-y\|^2.
\end{aligned} \tag{2.11}$$

In turn, using (2.6) and (2.11), we deduce that

$$\begin{aligned}
&(\forall x \in C)(\forall(y, y^*) \in \text{gra } A)(\forall(z, z^*) \in \text{gra } A) \\
&\langle y-x, By-Bz \rangle \leq \left\| \frac{y-x}{\sqrt{2\beta-\varepsilon}} \right\| \left\| \sqrt{2\beta-\varepsilon}(By-Bz) \right\| \\
&\leq \frac{\|y-x\|^2}{2(2\beta-\varepsilon)} + \frac{2\beta-\varepsilon}{2}\|By-Bz\|^2
\end{aligned} \tag{2.12}$$

$$\leq \kappa D_f(x, y) + \langle y-z, \delta_1(y^*-z^*) + \delta_2(By-Bz) \rangle. \tag{2.13}$$

(v) $\Rightarrow$ (iv): Set $\beta = \mu/\nu^2$. Then

$$\begin{aligned}
&(\forall(x, x^*) \in \text{gra}(A+B))(\forall(y, y^*) \in \text{gra}(A+B)) \\
&\langle x-y, x^*-y^* \rangle \geq \mu\|x-y\|^2 \geq \beta\|Bx-By\|^2.
\end{aligned} \tag{2.14}$$

(vi): We consider each case separately.

[a]: By arguing as in (2.11), we obtain $(\forall x \in \text{dom } A)(\forall y \in \text{dom } A) D_f(x, y) \geq (\alpha/2)\|x-y\|^2$. It thus follows from (2.12) and (2.7) that

$$\begin{aligned}
&(\forall x \in C)(\forall(y, y^*) \in \text{gra } A)(\forall(z, z^*) \in \text{gra } A) \\
&\langle y-x, By-Bz \rangle \leq \frac{\|y-x\|^2}{2(2\beta-\varepsilon)} + \frac{2\beta-\varepsilon}{2}\|By-Bz\|^2 \\
&\leq \kappa D_f(x, y) + \langle y-z, \delta_2(By-Bz) \rangle.
\end{aligned} \tag{2.15}$$

[b] $\Rightarrow$ [a]: We derive from [1, Proposition 4] that B is cocoercive with constant β .

[c] $\Rightarrow$ [a]: This follows from [1, Corollaire 10]. $\square$

Remark 2.2 Condition (iv) in Proposition 2.1 first appeared in [20] and does not seem to have gotten much notice in the literature. The cocoercivity condition (vi)[a] was first used in [17] to prove the weak convergence of the classical forward-backward method in Hilbert spaces. Finally, in reflexive Banach space minimization problems, (iii) appears in [18]; see also [3] for the Euclidean case.

Remark 2.3 Condition (iii) is satisfied in particular when $\mathcal{X}$ is a Hilbert space, $f = \|\cdot\|^2/2$, $\text{dom } \psi = \mathcal{X}$, and $\nabla\psi$ is Lipschitzian [7, Theorem 18.15], in which case it is known as the “descent lemma.” Condition (ii) can be viewed as an extension of this standard descent lemma involving triples (x, y, z) and an arbitrary Bregman distance D_f in reflexive Banach spaces. Let us underline that (ii) is more general than (iii). Indeed, consider the setting of Problem 1.1 with the following additional assumptions: $\mathcal{X}$ is a Hilbert space, $0 \in \text{int dom } f$, A is the normal cone operator of some self-dual cone K , and there exists a Gâteaux differentiable convex function $\psi: \mathcal{X} \rightarrow \mathbb{R}$ such that

$$B = \nabla\psi, \quad \text{Argmin } \psi = \{0\}, \quad \text{and} \quad \nabla\psi(K) \subset K. \tag{2.16}$$

Then $C = (\text{int dom } f) \cap \text{dom } A \subset K$ and $\mathcal{S} = \{0\}$. Further, for every $x \in C$ and every $y \in C$, (2.16) yields $D_\psi(x, y) - D_\psi(x, 0) - D_\psi(0, y) = \langle -x | \nabla\psi(y) - \nabla\psi(0) \rangle = \langle -x | \nabla\psi(y) \rangle \leq 0 \leq D_f(x, y)$. Therefore, (2.5) is satisfied. On the other hand, (iii) does not hold in general. For instance, take $\mathcal{X} = \mathbb{R}$, $K = [0, +\infty[$, $f = |\cdot|^2/2$, and $\psi = |\cdot|^{3/2}$.

2.3 Forward-backward splitting for monotone inclusions

The formal setting of the proposed Bregman forward-backward splitting method is as follows.

Algorithm 2.4 Consider the setting of Problem 1.1. Let $\alpha \in]0, +\infty[$, let $(\gamma_n)_{n \in \mathbb{N}}$ be in $]0, +\infty[$, and let $(f_n)_{n \in \mathbb{N}}$ be in $\mathcal{C}_\alpha(f)$. Suppose that the following hold:

- [a] $\inf_{n \in \mathbb{N}} \gamma_n > 0$, $\sup_{n \in \mathbb{N}} (\kappa \gamma_n) \leq \alpha$, and $\sup_{n \in \mathbb{N}} (\delta_1 \gamma_{n+1} / \gamma_n) < 1$.
- [b] There exists a summable sequence $(\eta_n)_{n \in \mathbb{N}}$ in $[0, +\infty[$ such that $(\forall n \in \mathbb{N}) D_{f_{n+1}} \leq (1 + \eta_n) D_{f_n}$.
- [c] For every $n \in \mathbb{N}$, ∇f_n is strictly monotone on C and $(\nabla f_n - \gamma_n B)(C) \subset \text{ran}(\nabla f_n + \gamma_n A)$.

Take $x_0 \in C$ and set $(\forall n \in \mathbb{N}) x_{n+1} = (\nabla f_n + \gamma_n A)^{-1}(\nabla f_n(x_n) - \gamma_n Bx_n)$.

Let us establish basic asymptotic properties of Algorithm 2.4, starting with the fact that its viability domain is C .

Proposition 2.5 Let $(x_n)_{n \in \mathbb{N}}$ be a sequence generated by Algorithm 2.4 and let $z \in \mathcal{S}$. Then $(x_n)_{n \in \mathbb{N}}$ is a well-defined sequence in C and the following hold:

- (i) $(D_{f_n}(z, x_n))_{n \in \mathbb{N}}$ converges.
- (ii) $\sum_{n \in \mathbb{N}} (1 - \kappa \gamma_n / \alpha) D_{f_n}(x_{n+1}, x_n) < +\infty$ and $\sum_{n \in \mathbb{N}} (1 - \kappa \gamma_n / \alpha) D_f(x_{n+1}, x_n) < +\infty$.
- (iii) $\sum_{n \in \mathbb{N}} \langle x_{n+1} - z, \gamma_n^{-1}(\nabla f_n(x_n) - \nabla f_n(x_{n+1})) - Bx_n + Bz \rangle < +\infty$.
- (iv) $\sum_{n \in \mathbb{N}} (1 - \delta_2) \langle x_n - z, Bx_n - Bz \rangle < +\infty$.
- (v) Suppose that one of the following is satisfied:
 - [a] C is bounded.
 - [b] f is supercoercive.
 - [c] f is uniformly convex.
 - [d] f is essentially strictly convex with $\text{dom } f^*$ open and ∇f^* weakly sequentially continuous.
 - [e] $\mathcal{X}$ is finite-dimensional and $\text{dom } f^*$ is open.
 - [f] f is essentially strictly convex and $\rho = \inf_{\substack{x \in \text{int dom } f \\ y \in \text{int dom } f \\ x \neq y}} \frac{D_f(x, y)}{D_f(y, x)} \in]0, +\infty[$.

Then $(x_n)_{n \in \mathbb{N}}$ is bounded.

Proof. Take $n \in \mathbb{N}$, and suppose that (y^*, y_1) and (y^*, y_2) belong to $\text{gra}(\nabla f_n + \gamma_n A)^{-1}$. Then $y^* \in (\nabla f_n + \gamma_n A)y_1$ and $y^* \in (\nabla f_n + \gamma_n A)y_2$. However, by virtue of condition [c] in Algorithm 2.4, $\nabla f_n + \gamma_n A$ is strictly monotone. Therefore, since $\langle y_1 - y_2, y^* - y^* \rangle = 0$, we infer that $y_1 = y_2$. Hence

$$(\nabla f_n + \gamma_n A)^{-1} \text{ is single-valued on } \text{dom}(\nabla f_n + \gamma_n A)^{-1} = \text{ran}(\nabla f_n + \gamma_n A). \quad (2.17)$$

Moreover, it follows from [9, Proposition 4.2.2] and (2.4) that

$$\text{ran}(\nabla f_n + \gamma_n A)^{-1} = \text{dom } \nabla f_n \cap \text{dom } A = (\text{int dom } f_n) \cap \text{dom } A = C. \quad (2.18)$$

Next, we observe that, since $x_0 \in C \subset \text{int dom } B$, $\nabla f_0(x_0) - \gamma_0 Bx_0$ is a singleton. Furthermore, in view of condition [c] in Algorithm 2.4, $\nabla f_0(x_0) - \gamma_0 Bx_0 \in \text{ran}(\nabla f_0 + \gamma_0 A)$. We thus deduce from (2.17) that $x_1 = (\nabla f_0 + \gamma_0 A)^{-1}(\nabla f_0(x_0) - \gamma_0 Bx_0)$ is uniquely defined. In addition, (2.18) yields $x_1 \in \text{ran}(\nabla f_0 + \gamma_0 A)^{-1} = C$. The conclusion that $(x_n)_{n \in \mathbb{N}}$ is a well-defined sequence in C follows by invoking these facts inductively.

(i)–(iv): Condition [a] in Algorithm 2.4 entails that there exists $\varepsilon \in]0, 1[$ such that

$$\delta_1 \gamma_{n+1} \leq (1 - \varepsilon) \gamma_n. \quad (2.19)$$

Now take $x_0^* \in Ax_0$ and set

$$\begin{cases} x_{n+1}^* = \gamma_n^{-1}(\nabla f_n(x_n) - \nabla f_n(x_{n+1})) - Bx_n \\ \Delta_n = D_{f_n}(z, x_n) + \delta_1 \gamma_n \langle x_n - z, x_n^* + Bz \rangle \\ \theta_n = (1 - \kappa \gamma_n / \alpha) D_{f_n}(x_{n+1}, x_n) \\ \quad + \varepsilon \gamma_n \langle x_{n+1} - z, x_{n+1}^* + Bz \rangle + (1 - \delta_2) \gamma_n \langle x_n - z, Bx_n - Bz \rangle. \end{cases} \quad (2.20)$$

In view of (2.20),

$$(x_{n+1}, x_{n+1}^*) \in \text{gra } A. \quad (2.21)$$

In turn, since $(z, -Bz) \in \text{gra } A$ and A is monotone,

$$\langle x_{n+1} - z, x_{n+1}^* + Bz \rangle \geq 0. \quad (2.22)$$

Hence, invoking condition [a] in Algorithm 2.4 and the monotonicity of B , we obtain $\theta_n \geq 0$. Next, since $z \in \text{int dom } f = \text{int dom } f_n$ by (2.4), we derive from (2.20) and [6, Proposition 2.3(ii)] that

$$\begin{aligned} 0 &= \langle x_{n+1} - z, \nabla f_n(x_n) - \nabla f_n(x_{n+1}) - \gamma_n Bx_n - \gamma_n x_{n+1}^* \rangle \\ &= \langle x_{n+1} - z, \nabla f_n(x_n) - \nabla f_n(x_{n+1}) \rangle + \gamma_n \langle z - x_{n+1}, Bx_n - Bz \rangle - \gamma_n \langle x_{n+1} - z, x_{n+1}^* + Bz \rangle \\ &= D_{f_n}(z, x_n) - D_{f_n}(z, x_{n+1}) - D_{f_n}(x_{n+1}, x_n) + \gamma_n \langle z - x_{n+1}, Bx_n - Bz \rangle \\ &\quad - \gamma_n \langle x_{n+1} - z, x_{n+1}^* + Bz \rangle. \end{aligned} \quad (2.23)$$

Thus, since $(z, -Bz) \in \text{gra } A$ and $f_n \in \mathcal{C}_\alpha(f)$, we infer from (2.19), (2.22), (2.21), and (1.1) that

$$\begin{aligned} &D_{f_n}(z, x_{n+1}) + \delta_1 \gamma_{n+1} \langle x_{n+1} - z, x_{n+1}^* + Bz \rangle \\ &\leq D_{f_n}(z, x_{n+1}) + \gamma_n \langle x_{n+1} - z, x_{n+1}^* + Bz \rangle - \varepsilon \gamma_n \langle x_{n+1} - z, x_{n+1}^* + Bz \rangle \\ &= D_{f_n}(z, x_n) - D_{f_n}(x_{n+1}, x_n) + \gamma_n \langle z - x_{n+1}, Bx_n - Bz \rangle - \varepsilon \gamma_n \langle x_{n+1} - z, x_{n+1}^* + Bz \rangle \\ &= D_{f_n}(z, x_n) - D_{f_n}(x_{n+1}, x_n) + \gamma_n \langle x_n - x_{n+1}, Bx_n - Bz \rangle - \gamma_n \langle x_n - z, Bx_n - Bz \rangle \\ &\quad - \varepsilon \gamma_n \langle x_{n+1} - z, x_{n+1}^* + Bz \rangle \\ &\leq D_{f_n}(z, x_n) - D_{f_n}(x_{n+1}, x_n) + \kappa \gamma_n D_f(x_{n+1}, x_n) + \delta_1 \gamma_n \langle x_n - z, x_n^* + Bz \rangle \\ &\quad + \delta_2 \gamma_n \langle x_n - z, Bx_n - Bz \rangle - \gamma_n \langle x_n - z, Bx_n - Bz \rangle - \varepsilon \gamma_n \langle x_{n+1} - z, x_{n+1}^* + Bz \rangle \\ &\leq D_{f_n}(z, x_n) + \delta_1 \gamma_n \langle x_n - z, x_n^* + Bz \rangle - (1 - \kappa \gamma_n / \alpha) D_{f_n}(x_{n+1}, x_n) \\ &\quad - \varepsilon \gamma_n \langle x_{n+1} - z, x_{n+1}^* + Bz \rangle - (1 - \delta_2) \gamma_n \langle x_n - z, Bx_n - Bz \rangle \\ &= \Delta_n - \theta_n. \end{aligned} \quad (2.24)$$

Consequently, by condition [b] in Algorithm 2.4 and (2.22),

$$\Delta_{n+1} = D_{f_{n+1}}(z, x_{n+1}) + \delta_1 \gamma_{n+1} \langle x_{n+1} - z, x_{n+1}^* + Bz \rangle$$

$$\begin{aligned}
&\leq (1 + \eta_n)(D_{f_n}(z, x_{n+1}) + \delta_1 \gamma_{n+1} \langle x_{n+1} - z, x_{n+1}^* + Bz \rangle) \\
&\leq (1 + \eta_n)(\Delta_n - \theta_n) \\
&\leq (1 + \eta_n)\Delta_n - \theta_n.
\end{aligned} \tag{2.25}$$

Hence, [7, Lemma 5.31] asserts that

$$(\Delta_n)_{n \in \mathbb{N}} \text{ converges and } \sum_{n \in \mathbb{N}} \theta_n < +\infty. \tag{2.26}$$

In turn, we infer from (2.20) and condition [a] in Algorithm 2.4 that

$$\begin{cases} \sum_{n \in \mathbb{N}} (1 - \kappa \gamma_n / \alpha) D_{f_n}(x_{n+1}, x_n) < +\infty \\ \sum_{n \in \mathbb{N}} \langle x_{n+1} - z, x_{n+1}^* + Bz \rangle < +\infty \\ \sum_{n \in \mathbb{N}} (1 - \delta_2) \langle x_n - z, Bx_n - Bz \rangle < +\infty. \end{cases} \tag{2.27}$$

Thus, since $(f_n)_{n \in \mathbb{N}}$ lies in $\mathcal{C}_\alpha(f)$, we obtain $\sum_{n \in \mathbb{N}} (1 - \kappa \gamma_n / \alpha) D_f(x_{n+1}, x_n) < +\infty$. It results from (2.26) and (2.20) that $(D_{f_n}(z, x_n))_{n \in \mathbb{N}}$ converges.

(v): Recall that $(x_n)_{n \in \mathbb{N}}$ lies in C .

[a]: Clear.

[b]: We derive from (i) that $(D_f(z, x_n))_{n \in \mathbb{N}}$ is bounded. In turn, [5, Lemma 7.3(viii)] asserts that $(x_n)_{n \in \mathbb{N}}$ is bounded.

[c]: It results from [24, Theorem 3.5.10] that there exists a function $\phi: [0, +\infty[\rightarrow [0, +\infty]$ that vanishes only at 0 such that $\lim_{t \rightarrow +\infty} \phi(t)/t \rightarrow +\infty$ and

$$(\forall x \in \text{int dom } f)(\forall y \in \text{dom } f) \quad \langle y - x, \nabla f(x) \rangle + f(x) + \phi(\|x - y\|) \leq f(y). \tag{2.28}$$

Hence, in the light of (i), $\sup_{n \in \mathbb{N}} \phi(\|x_n - z\|) \leq \sup_{n \in \mathbb{N}} D_f(z, x_n) \leq (1/\alpha) \sup_{n \in \mathbb{N}} D_{f_n}(z, x_n) < +\infty$ and $(x_n)_{n \in \mathbb{N}}$ is therefore bounded.

[d]: Suppose that there exists a subsequence $(x_{k_n})_{n \in \mathbb{N}}$ of $(x_n)_{n \in \mathbb{N}}$ such that $\|x_{k_n}\| \rightarrow +\infty$. We deduce from [5, Lemma 7.3(vii)] and (i) that

$$\sup_{n \in \mathbb{N}} D_{f^*}(\nabla f(x_n), \nabla f(z)) = \sup_{n \in \mathbb{N}} D_f(z, x_n) \leq \frac{1}{\alpha} \sup_{n \in \mathbb{N}} D_{f_n}(z, x_n) < +\infty. \tag{2.29}$$

However, f^* is a Legendre function by virtue of [5, Corollary 5.5] and $\nabla f(z) \in \text{int dom } f^*$ by virtue of [5, Theorem 5.10]. Thus, [5, Lemma 7.3(v)] guarantees that $D_{f^*}(\cdot, \nabla f(z))$ is coercive. It therefore follows from (2.29) that $(\nabla f(x_{k_n}))_{n \in \mathbb{N}}$ is bounded, and then from the reflexivity of $\mathcal{X}^*$ that $\mathfrak{W}(\nabla f(x_{k_n}))_{n \in \mathbb{N}} \neq \emptyset$. In turn, there exist a subsequence $(x_{l_{k_n}})_{n \in \mathbb{N}}$ of $(x_{k_n})_{n \in \mathbb{N}}$ and $x^* \in \mathcal{X}^*$ such that $\nabla f(x_{l_{k_n}}) \rightharpoonup x^*$. The weak lower semicontinuity of f^* and (2.29) yield $D_{f^*}(x^*, \nabla f(z)) \leq \liminf D_{f^*}(\nabla f(x_{l_{k_n}}), \nabla f(z)) < +\infty$. Therefore

$$\nabla f(x_{l_{k_n}}) \rightharpoonup x^* \in \text{dom } f^* = \text{int dom } f^*. \tag{2.30}$$

Moreover, [5, Theorem 5.10] asserts that $\nabla f^*(x^*) \in \text{int dom } f$ and $(\forall n \in \mathbb{N}) \nabla f^*(\nabla f(x_n)) = x_n$. Hence, (2.30) and the weak sequential continuity of ∇f^* imply that $x_{l_{k_n}} = \nabla f^*(\nabla f(x_{l_{k_n}})) \rightharpoonup \nabla f^*(x^*)$. This yields $\sup_{n \in \mathbb{N}} \|x_{l_{k_n}}\| < +\infty$ and we reach a contradiction.

[e]: A consequence of [5, Lemma 7.3(ix)] and (i).

[f]: It results from [5, Lemma 7.3(v)] that $D_f(\cdot, z)$ is coercive. In turn, since $\sup_{n \in \mathbb{N}} D_f(x_n, z) \leq (1/\rho) \sup_{n \in \mathbb{N}} D_f(z, x_n) < +\infty$ by (i), $(x_n)_{n \in \mathbb{N}}$ is bounded. $\square$

As seen in Proposition 2.5, by construction, an orbit of Algorithm 2.4 lies in C and therefore in $\text{int dom } f$. Next, we proceed to identify sufficient conditions that guarantee that their weak sequential cluster points are also in $\text{int dom } f$.

Proposition 2.6 *Let $(x_n)_{n \in \mathbb{N}}$ be a sequence generated by Algorithm 2.4 and suppose that one of the following holds:*

- [a] $\overline{\text{dom } f} \cap \overline{\text{dom } A} \subset \text{int dom } f$.
- [b] f is essentially strictly convex with $\text{dom } f^*$ open and ∇f^* weakly sequentially continuous.
- [c] f is strictly convex on $\text{int dom } f$ and $\rho = \inf_{\substack{x \in \text{int dom } f \\ y \in \text{int dom } f \\ x \neq y}} \frac{D_f(x, y)}{D_f(y, x)} \in]0, +\infty[$.
- [d] $\mathcal{X}$ is finite-dimensional.

Then $\mathfrak{W}(x_n)_{n \in \mathbb{N}} \subset \text{int dom } f$.

Proof. Suppose that $x \in \mathfrak{W}(x_n)_{n \in \mathbb{N}}$, say $x_{k_n} \rightharpoonup x$, and fix $z \in \mathcal{S}$.

[a]: Since $\overline{\text{dom } f}$ is closed and convex, it is weakly closed [10, Corollary II.6.3.3(i)]. Hence, since Proposition 2.5 asserts that $(x_n)_{n \in \mathbb{N}}$ lies in $C \subset \text{dom } f$, we infer that $\mathfrak{W}(x_n)_{n \in \mathbb{N}} \subset \overline{\text{dom } f}$. Likewise, since $\overline{\text{dom } A}$ is a closed convex set [24, Theorem 3.11.12] and $(x_n)_{n \in \mathbb{N}}$ lies in $C \subset \text{dom } A$, we obtain $\mathfrak{W}(x_n)_{n \in \mathbb{N}} \subset \overline{\text{dom } A}$. Altogether, $\mathfrak{W}(x_n)_{n \in \mathbb{N}} \subset \overline{\text{dom } f} \cap \overline{\text{dom } A} \subset \text{int dom } f$.

[b]: Using an argument similar to that of the proof of Proposition 2.5(v)[d], we infer that there exist a strictly increasing sequence $(l_{k_n})_{n \in \mathbb{N}}$ in $\mathbb{N}$ and $x^* \in \text{int dom } f^*$ such that $x_{l_{k_n}} \rightharpoonup \nabla f^*(x^*)$. Thus, appealing to [5, Theorem 5.10], we conclude that $x = \nabla f^*(x^*) \in \text{int dom } f$.

[c]: Proposition 2.5(i) and the weak lower semicontinuity of $D_f(\cdot, z)$ yield

$$D_f(x, z) \leq \liminf D_f(x_{k_n}, z) \leq (1/\rho) \liminf D_f(z, x_{k_n}) \leq (\alpha\rho)^{-1} \lim D_{f_{k_n}}(z, x_{k_n}) < +\infty. \quad (2.31)$$

Thus $x \in \text{dom } f$. We show that $\text{dom } f$ is open. Suppose that there exists $y \in \text{dom } f \setminus \text{int dom } f$, let $(\alpha_n)_{n \in \mathbb{N}}$ be a sequence in $]0, 1[$ such that $\alpha_n \rightarrow 1$, and set $(\forall n \in \mathbb{N}) y_n = \alpha_n y + (1 - \alpha_n)z$. Then $\{y_n\}_{n \in \mathbb{N}} \subset]y, z[\subset (\text{int dom } f) \setminus \{z\}$ [10, Proposition II.2.6.16]. Moreover, $y_n \rightarrow y$ and, by convexity of f , $(\forall n \in \mathbb{N}) D_f(y_n, z) \leq \alpha_n(f(y) - f(z) - \langle y - z, \nabla f(z) \rangle)$. Hence

$$\lim D_f(y_n, z) \leq f(y) - f(z) - \langle y - z, \nabla f(z) \rangle = D_f(y, z). \quad (2.32)$$

However, it results from the lower semicontinuity of f that $\liminf D_f(y_n, z) = \liminf (f(y_n) - f(z)) - \lim \langle y_n - z, \nabla f(z) \rangle \geq f(y) - f(z) - \langle y - z, \nabla f(z) \rangle = D_f(y, z)$. Hence, (2.32) forces

$$\lim D_f(y_n, z) = D_f(y, z). \quad (2.33)$$

In addition, by convexity of f , $(\forall n \in \mathbb{N}) D_f(z, y_n) \geq \alpha_n(f(z) - f(y) - \langle z - y, \nabla f(y_n) \rangle)$. However, [5, Theorem 5.6] and the essential smoothness of f entail that

$$\langle z - y, \nabla f(y_n) \rangle = \langle z - y, \nabla f(y + (1 - \alpha_n)(z - y)) \rangle \rightarrow -\infty. \quad (2.34)$$

Thus,

$$+\infty = \lim \left(\alpha_n (f(z) - f(y) - \langle z - y, \nabla f(y_n) \rangle) \right) \leq \lim D_f(z, y_n). \quad (2.35)$$

It results from (2.33) and (2.35) that $0 < \rho \leq \lim D_f(y_n, z)/D_f(z, y_n) = 0$, so that we reach a contradiction. Consequently, $\text{dom } f$ is open and hence $x \in \text{dom } f = \text{int dom } f$.

[d]: Proposition 2.5(i) ensures that $(x_{k_n})_{n \in \mathbb{N}}$ is a sequence in $\text{int dom } f$ such that $(D_f(z, x_{k_n}))_{n \in \mathbb{N}}$ is bounded. Therefore, [4, Theorem 3.8(ii)] and the essential smoothness of f yield $x \in \text{int dom } f$. $\square$

Definition 2.7 Algorithm 2.4 is focusing if, for every $z \in \mathcal{S}$,

$$\left\{ \begin{array}{l} (D_{f_n}(z, x_n))_{n \in \mathbb{N}} \text{ converges} \\ \sum_{n \in \mathbb{N}} \langle x_{n+1} - z, \gamma_n^{-1}(\nabla f_n(x_n) - \nabla f_n(x_{n+1})) - Bx_n + Bz \rangle < +\infty \\ \sum_{n \in \mathbb{N}} (1 - \delta_2) \langle x_n - z, Bx_n - Bz \rangle < +\infty \\ \sum_{n \in \mathbb{N}} (1 - \kappa \gamma_n / \alpha) D_{f_n}(x_{n+1}, x_n) < +\infty \end{array} \right. \Rightarrow \mathfrak{W}(x_n)_{n \in \mathbb{N}} \subset \text{zer}(A+B). \quad (2.36)$$

Our main result establishes the weak convergence of the orbits of Algorithm 2.4.

Theorem 2.8 Let $(x_n)_{n \in \mathbb{N}}$ be a sequence generated by Algorithm 2.4 and suppose that the following hold:

- [a] $(x_n)_{n \in \mathbb{N}}$ is bounded.
- [b] $\mathfrak{W}(x_n)_{n \in \mathbb{N}} \subset \text{int dom } f$.
- [c] Algorithm 2.4 is focusing.
- [d] One of the following is satisfied:

1/ $\mathcal{S}$ is a singleton.

2/ There exists a function g in $\Gamma_0(\mathcal{X})$ which is Gâteaux differentiable on $\text{int dom } g \supset C$, with ∇g strictly monotone on C , and such that, for every sequence $(y_n)_{n \in \mathbb{N}}$ in C and every $y \in \mathfrak{W}(y_n)_{n \in \mathbb{N}} \cap C$, $y_{k_n} \rightharpoonup y \Rightarrow \nabla f_{k_n}(y_{k_n}) \rightharpoonup \nabla g(y)$.

Then $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in $\mathcal{S}$.

Proof. It results from [a] and the reflexivity of $\mathcal{X}$ that

$$(x_n)_{n \in \mathbb{N}} \text{ lies in a weakly sequentially compact set.} \quad (2.37)$$

On the other hand, [c] and items (i)–(iv) in Proposition 2.5 yield $\mathfrak{W}(x_n)_{n \in \mathbb{N}} \subset \text{zer}(A + B)$. In turn, it results from [b] that

$$\emptyset \neq \mathfrak{W}(x_n)_{n \in \mathbb{N}} \subset \mathcal{S} \subset C. \quad (2.38)$$

In view of [7, Lemma 1.35] applied in $\mathcal{X}^{\text{weak}}$, it remains to show that $\mathfrak{W}(x_n)_{n \in \mathbb{N}}$ is a singleton. If [d]1/ holds, this follows from (2.38). Now suppose that [d]2/ holds, and take y_1 and y_2 in $\mathfrak{W}(x_n)_{n \in \mathbb{N}}$, say $x_{k_n} \rightharpoonup y_1$ and $x_{l_n} \rightharpoonup y_2$. Then $y_1 \in \mathcal{S}$ and $y_2 \in \mathcal{S}$ by virtue of (2.38), and we therefore deduce from Proposition 2.5(i) that $(D_{f_n}(y_1, x_n))_{n \in \mathbb{N}}$ and $(D_{f_n}(y_2, x_n))_{n \in \mathbb{N}}$ converge. However, condition [b] in Algorithm 2.4 and [7, Lemma 5.31] assert that $(D_{f_n}(y_1, y_2))_{n \in \mathbb{N}}$ converges. Hence, appealing to [6, Proposition 2.3(ii)], it follows that $(\langle y_1 - y_2, \nabla f_n(x_n) - \nabla f_n(y_2) \rangle)_{n \in \mathbb{N}} = (D_{f_n}(y_2, x_n) + D_{f_n}(y_1, y_2) - D_{f_n}(y_1, x_n))_{n \in \mathbb{N}}$ converges. Set

$\ell = \lim \langle y_1 - y_2, \nabla f_n(x_n) - \nabla f_n(y_2) \rangle$. Since $(x_n)_{n \in \mathbb{N}}$ is a sequence in C , we infer from (2.38) and [d]2/ that $\ell \leftarrow \langle y_1 - y_2, \nabla f_{l_n}(x_{l_n}) - \nabla f_{l_n}(y_2) \rangle \rightarrow \langle y_1 - y_2, \nabla g(y_2) - \nabla g(y_2) \rangle = 0$, which yields $\ell = 0$. However, invoking [d]2/, we obtain $\ell \leftarrow \langle y_1 - y_2, \nabla f_{k_n}(x_{k_n}) - \nabla f_{k_n}(y_2) \rangle \rightarrow \langle y_1 - y_2, \nabla g(y_1) - \nabla g(y_2) \rangle$. It therefore follows that $\langle y_1 - y_2, \nabla g(y_1) - \nabla g(y_2) \rangle = 0$ and hence from the strict monotonicity of ∇g on C that $y_1 = y_2$. $\square$

Example 2.9 We provide an example with operating conditions that are not captured by any of the methods described in (1.4)–(1.7). Let $p \in]1, +\infty[$, let $(\chi_n)_{n \in \mathbb{N}}$ be a sequence in $[1, +\infty[$ such that $\chi_n \rightarrow 1$, and let $(\eta_n)_{n \in \mathbb{N}}$ be a summable sequence in $[0, +\infty[$ such that $(\forall n \in \mathbb{N}) \chi_{n+1} \leq (1 + \eta_n)\chi_n$. We denote by $z = (\zeta_k)_{k \in \mathbb{N}}$ a sequence in $\ell^p(\mathbb{N})$. Set $\mathcal{X} = \ell^p(\mathbb{N}) \times \mathbb{R}$, hence $\mathcal{X}^* = \ell^{p/(p-1)}(\mathbb{N}) \times \mathbb{R}$, and define the Legendre functions

$$(\forall n \in \mathbb{N}) \quad f_n: \mathcal{X} \rightarrow]-\infty, +\infty]: (z, \xi) \mapsto \begin{cases} \frac{\chi_n}{p} \|z\|^p + 1 - \xi + \xi \ln \xi, & \text{if } \xi > 0; \\ \frac{\chi_n}{p} \|z\|^p + 1, & \text{if } \xi = 0; \\ +\infty, & \text{if } \xi \leq 0 \end{cases} \quad (2.39)$$

and

$$f = g: \mathcal{X} \rightarrow]-\infty, +\infty]: (z, \xi) \mapsto \begin{cases} \frac{1}{p} \|z\|^p - \xi + \xi \ln \xi, & \text{if } \xi > 0; \\ \frac{1}{p} \|z\|^p, & \text{if } \xi = 0; \\ +\infty, & \text{if } \xi \leq 0. \end{cases} \quad (2.40)$$

Now let $\psi: \mathcal{X} \rightarrow [0, +\infty[: (z, \xi) \mapsto \|z\|^p/p$, set $B = \nabla \psi$, and let $A: \mathcal{X} \rightarrow 2^{\mathcal{X}^*}$ be any maximally monotone operator such that

$$\text{dom } A \subset \ell^p(\mathbb{N}) \times]0, +\infty[\quad \text{and} \quad \text{zer}(A + B) \neq \emptyset. \quad (2.41)$$

Let us check that this setting conforms to that of Theorem 2.8. First, Proposition 2.1(iii) implies that (1.1) is satisfied with $\delta_1 = 0$ and $\delta_2 = \kappa = 1$. Next, we note that $\text{int dom } f = \ell^p(\mathbb{N}) \times]0, +\infty[$, that $(f_n)_{n \in \mathbb{N}}$ lies in $\mathcal{C}_1(f)$, and that condition [b] in Algorithm 2.4 holds. Furthermore, we derive from (2.39) that

$$(\forall n \in \mathbb{N}) \quad \nabla f_n: \ell^p(\mathbb{N}) \times]0, +\infty[\rightarrow \mathcal{X}^*: (z, \xi) \mapsto \left(\chi_n (\text{sign}(\zeta_k) |\zeta_k|^{p-1})_{k \in \mathbb{N}}, \ln \xi \right) \quad (2.42)$$

and we observe that

$$(\forall n \in \mathbb{N}) \quad \text{ran } \nabla f_n = \mathcal{X}^* \quad \text{and} \quad \text{dom}(\gamma_n A) \subset \text{dom } \nabla f_n. \quad (2.43)$$

It therefore follows from the Brézis–Haraux theorem [11, Théorème 4] that

$$(\forall n \in \mathbb{N}) \quad \text{ran}(\nabla f_n + \gamma_n A) = \mathcal{X}^*, \quad (2.44)$$

and hence that condition [c] in Algorithm 2.4 holds. It remains to verify condition [d]2/ in Theorem 2.8. Set $\varphi: \ell^p(\mathbb{N}) \rightarrow [0, +\infty[: z \mapsto \|z\|^p/p$ and $(\forall n \in \mathbb{N}) \varphi_n: \ell^p(\mathbb{N}) \rightarrow [0, +\infty[: z \mapsto \chi_n \|z\|^p/p$. Take a sequence $(z_n, \xi_n)_{n \in \mathbb{N}}$ in $\text{dom } A$ and a point $(z, \xi) \in \text{dom } A$ such that $(z_n, \xi_n) \rightharpoonup (z, \xi)$. We have $\xi_n \rightarrow \xi$ and $(\forall k \in \mathbb{N}) \zeta_{n,k} \rightarrow \zeta_k$. Now let $(e_k)_{k \in \mathbb{N}}$ be the canonical Schauder basis of $\ell^p(\mathbb{N})$. Then

$$(\forall k \in \mathbb{N}) \quad \langle e_k, \nabla \varphi_n(z_n) \rangle = \chi_n \text{sign}(\zeta_{n,k}) |\zeta_{n,k}|^{p-1} \rightarrow \text{sign}(\zeta_k) |\zeta_k|^{p-1} = \langle e_k, \nabla \varphi(z) \rangle \quad (2.45)$$

and $(\nabla \varphi_n(z_n))_{n \in \mathbb{N}}$ is bounded. It therefore follows from [2, Théorème VIII-2] that $\nabla \varphi_n(z_n) \rightharpoonup \nabla \varphi(z)$ and, in turn, that $\nabla f_n(z_n, \xi_n) \rightharpoonup \nabla g(z, \xi)$ by (2.40) and (2.42). Note that the above setting is not covered by the assumptions underlying (1.4)–(1.7): the fact that $B \neq 0$ excludes [6], the fact that $\mathcal{X}$ is not a Hilbert space excludes [15] and [20], and [18] is excluded because A is not a subdifferential.

3 Special cases and applications

We illustrate the general scope of Theorem 2.8 by recovering apparently unrelated results and also by deriving new ones. Sufficient conditions for [a] and [b] in Theorem 2.8 to hold can be found in Propositions 2.5(v) and 2.6, respectively. As to checking the focusing condition [c], the following fact will be useful.

Lemma 3.1 [13, Proposition 2.1(iii)] *Let $M_1: \mathcal{X} \rightarrow 2^{\mathcal{X}^*}$ and $M_2: \mathcal{X} \rightarrow 2^{\mathcal{X}^*}$ be maximally monotone, let $(a_n, a_n^*)_{n \in \mathbb{N}}$ be a sequence in $\text{gra } M_1$, let $(b_n, b_n^*)_{n \in \mathbb{N}}$ be a sequence in $\text{gra } M_2$, let $x \in \mathcal{X}$, and let $y^* \in \mathcal{X}^*$. Suppose that $a_n \rightharpoonup x$, $b_n^* \rightharpoonup y^*$, $a_n^* + b_n^* \rightarrow 0$, and $a_n - b_n \rightarrow 0$. Then $x \in \text{zer}(M_1 + M_2)$.*

3.1 Recovering existing frameworks for monotone inclusions

In this section, we show that the existing results of [6, 15, 20] discussed in the Introduction can be recovered from Theorem 2.8. As will be clear from the proofs, more general versions of these results can also be derived at once from Theorem 2.8. First, we derive from Theorem 2.8 the convergence of the Bregman-based proximal point algorithm (1.4) studied in [6, Section 5.5].

Corollary 3.2 *Let $A: \mathcal{X} \rightarrow 2^{\mathcal{X}^*}$ be maximally monotone, let $f \in \Gamma_0(\mathcal{X})$ be a supercoercive Legendre function such that $\emptyset \neq \text{zer } A \subset \text{dom } A \subset \text{int dom } f$ and ∇f is weakly sequentially continuous, and let $(\gamma_n)_{n \in \mathbb{N}}$ be a sequence in $]0, +\infty[$ such that $\inf_{n \in \mathbb{N}} \gamma_n > 0$. Suppose that, for every bounded sequence $(y_n)_{n \in \mathbb{N}}$ in $\text{int dom } f$,*

$$D_f(y_{n+1}, y_n) \rightarrow 0 \quad \Rightarrow \quad \nabla f(y_{n+1}) - \nabla f(y_n) \rightarrow 0. \quad (3.1)$$

Take $x_0 \in C$ and set $(\forall n \in \mathbb{N}) x_{n+1} = (\nabla f + \gamma_n A)^{-1}(\nabla f(x_n))$. Then $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in $\text{zer } A$.

Proof. We apply Theorem 2.8 with $B = 0$, $\alpha = 1$, $\kappa = \delta_1 = \delta_2 = 0$, and $(\forall n \in \mathbb{N}) f_n = f$. First, (1.1) together with conditions [a] and [b] in Algorithm 2.4 are trivially fulfilled. On the other hand, since f is a Legendre function and $\text{dom } A \subset \text{int dom } f$, condition [c] in Algorithm 2.4 follows from [6, Theorem 3.13(iv)(d)]. Next, condition [a] in Theorem 2.8 follows from Proposition 2.5(v)[b]. Furthermore, in view of the weak sequential continuity of ∇f , condition [d]2/ in Theorem 2.8 is satisfied with $g = f$. Next, to show that the algorithm is focusing, suppose that $\sum_{n \in \mathbb{N}} D_f(x_{n+1}, x_n) < +\infty$ and take $x \in \mathfrak{W}(x_n)_{n \in \mathbb{N}}$, say $x_{k_n} \rightharpoonup x$. Since $(x_n)_{n \in \mathbb{N}}$ is a bounded sequence in $\text{int dom } f$, we derive from (3.1) that $\nabla f(x_{n+1}) - \nabla f(x_n) \rightarrow 0$. In turn, since $\inf_{n \in \mathbb{N}} \gamma_n > 0$, it follows that $\gamma_n^{-1}(\nabla f(x_{n+1}) - \nabla f(x_n)) \rightarrow 0$. However, by construction, $(\forall n \in \mathbb{N}) \gamma_{k_n-1}^{-1}(\nabla f(x_{k_n-1}) - \nabla f(x_{k_n})) \in Ax_{k_n}$. Therefore, upon invoking Lemma 3.1 (with $M_1 = A$ and $M_2 = 0$), we obtain $x \in \text{zer } A$ and the algorithm is therefore focusing. This also shows that $\mathfrak{W}(x_n)_{n \in \mathbb{N}} \subset \text{zer } A \subset \text{int dom } f$. Condition [b] in Theorem 2.8 is thus satisfied. $\square$

The next application of Theorem 2.8 is a variable metric version of the Hilbertian forward-backward method (1.5) established in [15, Theorem 4.1].

Corollary 3.3 *Let $\mathcal{X}$ be a real Hilbert space, let $A: \mathcal{X} \rightarrow 2^{\mathcal{X}}$ be maximally monotone, let α and β be in $]0, +\infty[$, and let $B: \mathcal{X} \rightarrow \mathcal{X}$ satisfy*

$$(\forall x \in \mathcal{X})(\forall y \in \mathcal{X}) \quad \langle x - y \mid Bx - By \rangle \geq \beta \|Bx - By\|^2. \quad (3.2)$$

Further, for every $n \in \mathbb{N}$, let $U_n: \mathcal{X} \rightarrow \mathcal{X}$ be a bounded linear operator which is α -strongly monotone and self-adjoint. Suppose that $\text{zer}(A + B) \neq \emptyset$ and that there exists a summable sequence $(\eta_n)_{n \in \mathbb{N}}$ in $[0, +\infty[$

such that

$$(\forall n \in \mathbb{N})(\forall x \in \mathcal{X}) \quad \langle x \mid U_{n+1}x \rangle \leq (1 + \eta_n)\langle x \mid U_nx \rangle. \quad (3.3)$$

Let $\varepsilon \in]0, 2\beta[$ and let $(\gamma_n)_{n \in \mathbb{N}}$ be a sequence in $]0, +\infty[$ such that $0 < \inf_{n \in \mathbb{N}} \gamma_n \leq \sup_{n \in \mathbb{N}} \gamma_n \leq (2\beta - \varepsilon)\alpha$. Define a sequence $(x_n)_{n \in \mathbb{N}}$ via the recursion

$$x_0 \in \text{dom } A \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = (U_n + \gamma_n A)^{-1}(U_n x_n - \gamma_n Bx_n). \quad (3.4)$$

Then $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in $\text{zer}(A + B)$.

Proof. Set $f = \|\cdot\|^2/2$, $C = \text{dom } A$, and $\mathcal{S} = \text{zer}(A + B)$. In addition, for every $n \in \mathbb{N}$, define $f_n: \mathcal{X} \rightarrow \mathbb{R}: x \mapsto \langle x \mid U_nx \rangle/2$. Let us apply Theorem 2.8 with $\kappa = 1/(2\beta - \varepsilon)$, $\delta_1 = 0$, and $\delta_2 = (2\beta - \varepsilon)/(2\beta) \in]0, 1[$. First, $f \in \Gamma_0(\mathcal{X})$ is a supercoercive Legendre function with $\text{dom } f = \mathcal{X}$ and, for every $n \in \mathbb{N}$, since $\nabla f_n = U_n$ is α -strongly monotone, $f_n \in \mathcal{C}_\alpha(f)$. Furthermore, it follows from Proposition 2.1(vi)[a] that (1.1) is fulfilled. We also observe that condition [a] in Algorithm 2.4 is satisfied. Next, by (3.3) and the assumption that the operators $(U_n)_{n \in \mathbb{N}}$ are self-adjoint,

$$\begin{aligned} (\forall n \in \mathbb{N})(\forall x \in \mathcal{X})(\forall y \in \mathcal{X}) \quad D_{f_{n+1}}(x, y) &= \frac{1}{2} \langle x - y \mid U_{n+1}(x - y) \rangle \\ &\leq \frac{1 + \eta_n}{2} \langle x - y \mid U_n(x - y) \rangle \\ &= D_{f_n}(x, y) \end{aligned} \quad (3.5)$$

and condition [b] in Algorithm 2.4 therefore holds. Now take $n \in \mathbb{N}$. Since $\nabla f_n = U_n$ is maximally monotone with $\text{dom } \nabla f_n = \mathcal{X}$ and A is maximally monotone, [7, Corollary 25.5(i)] entails that $\nabla f_n + \gamma_n A$ is maximally monotone. Thus, since $\nabla f_n + \gamma_n A$ is α -strongly monotone, [7, Proposition 22.11(ii)] implies that $\text{ran}(\nabla f_n + \gamma_n A) = \mathcal{X}$ and it follows that condition [c] in Algorithm 2.4 is satisfied. Next, in view of Proposition 2.5(v)[b], $(x_n)_{n \in \mathbb{N}}$ is bounded, while $\mathfrak{W}(x_n)_{n \in \mathbb{N}} \subset \mathcal{X} = \text{int dom } f$. Now set $\mu = \sup_{n \in \mathbb{N}} \|U_n\|$. For every $n \in \mathbb{N}$, since it results from (3.3) and [7, Fact 2.25(iii)] that

$$(\forall x \in \mathcal{X}) \quad \|x\| \leq 1 \quad \Rightarrow \quad \langle x \mid U_nx \rangle \leq \left(\prod_{k \in \mathbb{N}} (1 + \eta_k) \right) \langle x \mid U_0x \rangle \leq \left(\prod_{k \in \mathbb{N}} (1 + \eta_k) \right) \|U_0\|, \quad (3.6)$$

we derive from [7, Fact 2.25(iii)] that $\|U_n\| \leq \|U_0\| \prod_{k \in \mathbb{N}} (1 + \eta_k)$. Hence $\mu < +\infty$ and therefore, appealing to [14, Lemma 2.3(i)], there exists an α -strongly monotone self-adjoint bounded linear operator $U: \mathcal{X} \rightarrow \mathcal{X}$ such that $(\forall w \in \mathcal{X}) \quad U_n w \rightarrow U w$. Define $g: \mathcal{X} \rightarrow \mathbb{R}: x \mapsto \langle x \mid Ux \rangle/2$. Then $\nabla g = U$ is strongly monotone (and thus strictly monotone). Furthermore, given $(y_n)_{n \in \mathbb{N}}$ in C and $y \in \mathfrak{W}(y_n)_{n \in \mathbb{N}} \cap C$, say $y_{k_n} \rightharpoonup y$, we have

$$(\forall w \in \mathcal{X}) \quad \langle w \mid \nabla f_{k_n}(y_{k_n}) \rangle = \langle U_{k_n} w \mid y_{k_n} \rangle \rightarrow \langle U w \mid y \rangle = \langle w \mid U y \rangle = \langle w \mid \nabla g(y) \rangle \quad (3.7)$$

and thus $\nabla f_{k_n}(y_{k_n}) \rightharpoonup \nabla g(y)$. Therefore, condition [d]2/ in Theorem 2.8 is satisfied. Let us now verify that (3.4) is focusing. Towards this goal, take $z \in \mathcal{S}$ and suppose that $\sum_{n \in \mathbb{N}} (1 - \delta_2) \langle x_n - z \mid Bx_n - Bz \rangle < +\infty$ and $\sum_{n \in \mathbb{N}} (1 - \kappa \gamma_n / \alpha) D_{f_n}(x_{n+1}, x_n) < +\infty$. Since $\delta_2 < 1$ and $\sup_{n \in \mathbb{N}} (\kappa \gamma_n) < \alpha$, we infer from (3.2) that

$$\sum_{n \in \mathbb{N}} \|Bx_n - Bz\|^2 \leq \frac{1}{\beta} \sum_{n \in \mathbb{N}} \langle x_n - z \mid Bx_n - Bz \rangle < +\infty \quad (3.8)$$

and $\sum_{n \in \mathbb{N}} \|x_{n+1} - x_n\|^2 = 2 \sum_{n \in \mathbb{N}} D_f(x_{n+1}, x_n) \leq (2/\alpha) \sum_{n \in \mathbb{N}} D_{f_n}(x_{n+1}, x_n) < +\infty$. It follows that

$$\|U_n(x_{n+1} - x_n)\| \leq \mu \|x_{n+1} - x_n\| \rightarrow 0. \quad (3.9)$$

Now take $x \in \mathfrak{W}(x_n)_{n \in \mathbb{N}}$, say $x_{k_n} \rightharpoonup x$, and set $(\forall n \in \mathbb{N}) x_{n+1}^* = \gamma_n^{-1} U_n(x_n - x_{n+1}) - Bx_n$. It results from (3.4) that $(x_{k_n+1}^*, x_{k_n+1}^*)_{n \in \mathbb{N}}$ lies in $\text{gra } A$ and from (3.9) that $x_{k_n+1} \rightharpoonup x$. Moreover, (3.9) yields $x_{k_n+1}^* + Bx_{k_n} \rightarrow 0$. Altogether, Lemma 3.1 (applied to the sequences $(x_{k_n+1}^*, x_{k_n+1}^*)_{n \in \mathbb{N}}$ in $\text{gra } A$ and $(x_{k_n}, Bx_{k_n})_{n \in \mathbb{N}}$ in $\text{gra } B$) guarantees that $x \in \text{zer}(A + B)$. Consequently, Theorem 2.8 asserts that $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in $\mathcal{S}$. $\square$

Example 3.4 The classical forward-backward method is obtained by setting $U_n \equiv \text{Id}$ in Corollary 3.3, which yields

$$x_0 \in \text{dom } A \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = (\text{Id} + \gamma_n A)^{-1}(x_n - \gamma_n Bx_n). \quad (3.10)$$

The case when the proximal parameters $(\gamma_n)_{n \in \mathbb{N}}$ are constant was first addressed in [17].

We now turn to the Renaud–Cohen algorithm (1.7) and recover [20, Theorem 3.4].

Corollary 3.5 *Let $\mathcal{X}$ be a real Hilbert space, let $A: \mathcal{X} \rightarrow 2^{\mathcal{X}}$ and $B: \mathcal{X} \rightarrow \mathcal{X}$ be maximally monotone, and let $f: \mathcal{X} \rightarrow \mathbb{R}$ be convex and Fréchet differentiable. Suppose that $\text{zer}(A + B) \neq \emptyset$, that ∇f is 1-strongly monotone on $\text{dom } A$ and Lipschitzian on bounded sets, and that there exists $\beta \in]0, +\infty[$ such that*

$$(\forall (x, x^*) \in \text{gra}(A + B)) (\forall (y, y^*) \in \text{gra}(A + B)) \quad \langle x - y \mid x^* - y^* \rangle \geq \beta \|Bx - By\|^2. \quad (3.11)$$

Let $\gamma \in]0, 2\beta[$, take $x_0 \in \text{dom } A$, and set $(\forall n \in \mathbb{N}) x_{n+1} = (\nabla f + \gamma A)^{-1}(\nabla f(x_n) - \gamma Bx_n)$. Suppose, in addition, that ∇f is weakly sequentially continuous. Then $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in $\text{zer}(A + B)$.

Proof. Let $\varepsilon \in]0, 2\beta[$ be such that $\gamma < 2\beta - \varepsilon$. We apply Theorem 2.8 with $C = \text{dom } A$, $\alpha = 1$, $\kappa = 1/(2\beta - \varepsilon)$, $\delta_1 = \delta_2 = (2\beta - \varepsilon)/(2\beta) \in]0, 1[$, and $(\forall n \in \mathbb{N}) f_n = f$ and $\eta_n = 0$. Proposition 2.1(iv) asserts that (1.1) is satisfied. Furthermore, as shown in the proof of Proposition 2.1(iv),

$$(\forall x \in \overline{\text{dom } A}) (\forall y \in \overline{\text{dom } A}) \quad D_f(x, y) \geq \frac{1}{2} \|x - y\|^2. \quad (3.12)$$

Next, note that conditions [a] and [b] in Algorithm 2.4 are trivially satisfied. Since $\nabla f + \gamma A$ is strongly monotone and since, by [7, Corollary 25.5(i)], $\nabla f + \gamma A$ is maximally monotone, it follows from [7, Proposition 22.11(ii)] that $\text{ran}(\nabla f + \gamma A) = \mathcal{X}$ and therefore that condition [c] in Algorithm 2.4 holds. We observe that condition [b] in Theorem 2.8 is trivially satisfied and that condition [a] in Theorem 2.8 follows from (3.12) and Proposition 2.5(i). Furthermore, since ∇f is weakly sequentially continuous and 1-strongly monotone on C , condition [d]2/ in Theorem 2.8 is satisfied with $g = f$. Now take $z \in \text{zer}(A + B)$ and suppose that $\sum_{n \in \mathbb{N}} (1 - \kappa\gamma) D_f(x_{n+1}, x_n) < +\infty$, $\sum_{n \in \mathbb{N}} (1 - \delta_2) \langle x_n - z \mid Bx_n - Bz \rangle < +\infty$, and $\sum_{n \in \mathbb{N}} \langle x_{n+1} - z \mid \gamma^{-1}(\nabla f(x_n) - \nabla f(x_{n+1})) - Bx_n + Bz \rangle < +\infty$. Then, since $\kappa\gamma < 1$ and $\delta_2 < 1$, it follows that

$$\sum_{n \in \mathbb{N}} D_f(x_{n+1}, x_n) < +\infty \quad \text{and} \quad \sum_{n \in \mathbb{N}} \langle x_n - z \mid Bx_n - Bz \rangle < +\infty, \quad (3.13)$$

and therefore that

$$\sum_{n \in \mathbb{N}} \langle x_{n+1} - z \mid \gamma^{-1}(\nabla f(x_n) - \nabla f(x_{n+1})) - Bx_n + Bx_{n+1} \rangle < +\infty. \quad (3.14)$$

Since $(z, 0) \in \text{gra}(A + B)$ and since the sequence $(x_{n+1}, \gamma^{-1}(\nabla f(x_n) - \nabla f(x_{n+1})) - Bx_n + Bx_{n+1})_{n \in \mathbb{N}}$ lies in $\text{gra}(A + B)$ by construction, it follows from (3.11) and (3.14) that $\sum_{n \in \mathbb{N}} \|Bx_n - Bz\|^2 < +\infty$.

On the other hand, since $(x_n)_{n \in \mathbb{N}}$ lies in $\text{dom } A$ by Proposition 2.5, we deduce from (3.12) and (3.13) that $x_{n+1} - x_n \rightarrow 0$. In turn, it results from the Lipschitz continuity of ∇f on the bounded set $\{x_n\}_{n \in \mathbb{N}}$ that $\nabla f(x_n) - \nabla f(x_{n+1}) \rightarrow 0$. Now take $x \in \mathfrak{W}(x_n)_{n \in \mathbb{N}}$, say $x_{k_n} \rightharpoonup x$, and set $(\forall n \in \mathbb{N})$ $x_{n+1}^* = \gamma^{-1}(\nabla f(x_n) - \nabla f(x_{n+1})) - Bx_n$. Then $(x_{k_n+1}, x_{k_n+1}^*)_{n \in \mathbb{N}}$ lies in $\text{gra } A$. Furthermore, $x_{k_n+1}^* + Bx_{k_n} = \gamma^{-1}(\nabla f(x_{k_n}) - \nabla f(x_{k_n+1})) \rightarrow 0$ and, since $x_n - x_{n+1} \rightarrow 0$, $x_{k_n+1} \rightharpoonup x$. Thus, applying Lemma 3.1 with the sequences $(x_{k_n+1}, x_{k_n+1}^*)_{n \in \mathbb{N}}$ and $(x_{k_n}, Bx_{k_n})_{n \in \mathbb{N}}$ yields $x \in \text{zer}(A + B)$, and we conclude that condition [c] in Theorem 2.8 is satisfied as well. $\square$

3.2 The finite-dimensional case

We discuss the finite-dimensional case, a setting in which the assumptions can be greatly simplified and the results presented below are new.

Corollary 3.6 *Let $(x_n)_{n \in \mathbb{N}}$ be a sequence generated by Algorithm 2.4. In addition, suppose that the following hold:*

- [a] $\mathcal{X}$ is finite-dimensional.
- [b] f is essentially strictly convex and $\text{dom } f^*$ is open.
- [c] $(\text{int dom } f) \cap \overline{\text{dom } A} \subset \text{int dom } B$.
- [d] $\sup_{n \in \mathbb{N}} (\kappa \gamma_n) < \alpha$.
- [e] *There exists a function g in $\Gamma_0(\mathcal{X})$ which is differentiable on $\text{int dom } g \supset \text{int dom } f$, with ∇g strictly monotone on C , and such that, for every sequence $(y_n)_{n \in \mathbb{N}}$ in C and every sequential cluster point $y \in \text{int dom } f$ of $(y_n)_{n \in \mathbb{N}}$, $y_{k_n} \rightarrow y \Rightarrow \nabla f_{k_n}(y_{k_n}) \rightarrow \nabla g(y)$.*

Then $(x_n)_{n \in \mathbb{N}}$ converges to a point in $\mathcal{S}$.

Proof. It follows from Proposition 2.5(v)[e] that $(x_n)_{n \in \mathbb{N}}$ is bounded and from Proposition 2.6[d] that $\mathfrak{W}(x_n)_{n \in \mathbb{N}} \subset \text{int dom } f$. In view of Theorem 2.8, it remains to show that Algorithm 2.4 is focusing. Towards this goal, let $z \in \mathcal{S}$, and suppose that $(D_{f_n}(z, x_n))_{n \in \mathbb{N}}$ converges and $\sum_{n \in \mathbb{N}} (1 - \kappa \gamma_n / \alpha) D_{f_n}(x_{n+1}, x_n) < +\infty$, and let x be a sequential cluster point of $(x_n)_{n \in \mathbb{N}}$, say $x_{k_n} \rightarrow x$. Using [d] and the fact that $(f_n)_{n \in \mathbb{N}}$ lies in $\mathcal{C}_\alpha(f)$, we obtain

$$(D_f(z, x_n))_{n \in \mathbb{N}} \text{ is bounded and } \sum_{n \in \mathbb{N}} D_{f_n}(x_{n+1}, x_n) < +\infty. \quad (3.15)$$

Since $(x_{k_n})_{n \in \mathbb{N}}$ lies in $\text{int dom } f$, [4, Theorem 3.8(ii)] and (3.15) imply that

$$x \in \text{int dom } f \quad (3.16)$$

and [5, Theorem 5.10] thus yields

$$\nabla f(x_{k_n}) \rightarrow \nabla f(x) \in \text{int dom } f^*. \quad (3.17)$$

Next, it results from [b], [5, Lemma 7.3(vii)], and (3.15) that

$$(D_{f^*}(\nabla f(x_n), \nabla f(z)))_{n \in \mathbb{N}} = (D_f(z, x_n))_{n \in \mathbb{N}} \text{ is bounded.} \quad (3.18)$$

Therefore, since $\nabla f(z) \in \text{int dom } f^*$ [5, Theorem 5.10] and since f^* is a Legendre function [5, Corollary 5.5], it results from [5, Lemma 7.3(v)] that $(\nabla f(x_{k_n+1}))_{n \in \mathbb{N}}$ is bounded. In turn, there exists a strictly increasing sequence $(l_{k_n})_{n \in \mathbb{N}}$ in $\mathbb{N}$ and a point $x^* \in \mathcal{X}^*$ such that

$$\nabla f(x_{l_{k_n}+1}) \rightarrow x^*. \quad (3.19)$$

By lower semicontinuity of $D_{f^*}(\cdot, \nabla f(z))$ and (3.18), $x^* \in \text{dom } f^*$. On the other hand, appealing to [5, Lemma 7.3(vii)] and (3.15), we obtain

$$0 \leq D_{f^*}(\nabla f(x_{l_{k_n}}), \nabla f(x_{l_{k_n}+1})) = D_f(x_{l_{k_n}+1}, x_{l_{k_n}}) \leq \frac{1}{\alpha} D_{f_{l_{k_n}}}(x_{l_{k_n}+1}, x_{l_{k_n}}) \rightarrow 0. \quad (3.20)$$

Thus, since $(\nabla f(x_n))_{n \in \mathbb{N}}$ lies in $\text{int dom } f^*$ by virtue of Proposition 2.5 and [5, Theorem 5.10], we derive from [4, Theorem 3.9(iii)], (3.17), and (3.19) that $x^* = \nabla f(x)$ and, hence, from (3.19) that $\nabla f(x_{l_{k_n}+1}) \rightarrow \nabla f(x)$. It thus follows from [5, Theorem 5.10] that $x_{l_{k_n}+1} \rightarrow x$. In turn, by using respectively [e] with the sequences $(x_n)_{n \in \mathbb{N}}$ and $(x_{n+1})_{n \in \mathbb{N}}$, we get $\nabla f_{l_{k_n}}(x_{l_{k_n}}) \rightarrow \nabla g(x)$ and $\nabla f_{l_{k_n}}(x_{l_{k_n}+1}) \rightarrow \nabla g(x)$. Now set $(\forall n \in \mathbb{N})$ $x_{n+1}^* = \gamma_n^{-1}(\nabla f_n(x_n) - \nabla f_n(x_{n+1})) - Bx_n$. Then, by construction of $(x_n)_{n \in \mathbb{N}}$, $(\forall n \in \mathbb{N})$ $(x_{n+1}, x_{n+1}^*) \in \text{gra } A$. In addition, since $\inf_{n \in \mathbb{N}} \gamma_n > 0$ and $\nabla f_{l_{k_n}}(x_{l_{k_n}}) - \nabla f_{l_{k_n}}(x_{l_{k_n}+1}) \rightarrow \nabla g(x) - \nabla g(x) = 0$, we deduce that $x_{l_{k_n}+1}^* + Bx_{l_{k_n}} \rightarrow 0$. On the other hand, since $(x_n)_{n \in \mathbb{N}}$ lies in $\text{dom } A$ and $x_{k_n} \rightarrow x$, it follows that $x \in \overline{\text{dom } A}$ and therefore, by (3.16) and [c], that $x \in \text{int dom } B$. Hence, using [21, Corollary 1.1], we obtain $Bx_{l_{k_n}} \rightarrow Bx$. Altogether, Lemma 3.1 (applied to the sequence $(x_{l_{k_n}+1}, x_{l_{k_n}+1}^*)_{n \in \mathbb{N}}$ in $\text{gra } A$ and the sequence $(x_{l_{k_n}}, Bx_{l_{k_n}})_{n \in \mathbb{N}}$ in $\text{gra } B$) asserts that $x \in \text{zer}(A + B)$. In view of Theorem 2.8, we conclude that $(x_n)_{n \in \mathbb{N}}$ converges to a point in $\mathcal{S}$. $\square$

3.3 Forward-backward splitting for convex minimization

In this section, we study the convergence of (1.6). Our results improve on and complement those of [18].

Problem 3.7 Let $\varphi \in \Gamma_0(\mathcal{X})$, let $\psi \in \Gamma_0(\mathcal{X})$, and let $f \in \Gamma_0(\mathcal{X})$ be essentially smooth. Set $C = (\text{int dom } f) \cap \text{dom } \partial \varphi$ and $\mathcal{S} = (\text{int dom } f) \cap \text{Argmin}(\varphi + \psi)$. Suppose that $\varphi + \psi$ is coercive, $\emptyset \neq C \subset \text{int dom } \psi$, $\mathcal{S} \neq \emptyset$, ψ is Gâteaux differentiable on $\text{int dom } \psi$, and there exists $\kappa \in]0, +\infty[$ such that

$$(\forall x \in C)(\forall y \in C) \quad D_\psi(x, y) \leq \kappa D_f(x, y). \quad (3.21)$$

The objective is to find a point in $\mathcal{S}$.

In the context of Problem 3.7, given $\gamma \in]0, +\infty[$ and $g \in \mathcal{C}_\alpha(f)$, we define $\text{prox}_{\gamma\varphi}^g = (\nabla g + \gamma\partial\varphi)^{-1}$.

Algorithm 3.8 Consider the setting of Problem 3.7. Let $\alpha \in]0, +\infty[$, let $(\gamma_n)_{n \in \mathbb{N}}$ be in $]0, +\infty[$, and let $(f_n)_{n \in \mathbb{N}}$ be in $\mathcal{C}_\alpha(f)$. Suppose that the following hold:

- [a] There exists $\varepsilon \in]0, 1[$ such that $0 < \inf_{n \in \mathbb{N}} \gamma_n \leq \sup_{n \in \mathbb{N}} \gamma_n \leq \alpha(1 - \varepsilon)/\kappa$.
- [b] There exists a summable sequence $(\eta_n)_{n \in \mathbb{N}}$ in $[0, +\infty[$ such that $(\forall n \in \mathbb{N})$ $D_{f_{n+1}} \leq (1 + \eta_n)D_{f_n}$.
- [c] For every $n \in \mathbb{N}$, $\text{int dom } f_n = \text{dom } \partial f_n$ and ∇f_n is strictly monotone on C .

Take $x_0 \in C$ and set $(\forall n \in \mathbb{N})$ $x_{n+1} = \text{prox}_{\gamma_n\varphi}^{f_n}(\nabla f_n(x_n) - \gamma_n \nabla \psi(x_n))$.

Theorem 3.9 Let $(x_n)_{n \in \mathbb{N}}$ be a sequence generated by Algorithm 3.8 and suppose that the following hold:

[a] $\mathfrak{W}(x_n)_{n \in \mathbb{N}} \subset \text{int dom } f$.

[b] One of the following is satisfied:

1/ $\mathcal{S}$ is a singleton.

2/ There exists a function g in $\Gamma_0(\mathcal{X})$ which is Gâteaux differentiable on $\text{int dom } g \supset C$, with ∇g strictly monotone on C , and such that, for every sequence $(y_n)_{n \in \mathbb{N}}$ in C and every $y \in \mathfrak{W}(y_n)_{n \in \mathbb{N}} \cap C$, $y_{k_n} \rightharpoonup y \Rightarrow \nabla f_{k_n}(y_{k_n}) \rightharpoonup \nabla g(y)$.

Then the following hold:

(i) $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in $\mathcal{S}$.

(ii) $(x_n)_{n \in \mathbb{N}}$ is a monotone minimizing sequence: $\varphi(x_n) + \psi(x_n) \downarrow \min(\varphi + \psi)(\mathcal{X})$.

(iii) $\sum_{n \in \mathbb{N}} ((\varphi + \psi)(x_n) - \min(\varphi + \psi)(\mathcal{X})) < +\infty$ and $(\varphi + \psi)(x_n) - \min(\varphi + \psi)(\mathcal{X}) = o(1/n)$.

(iv) $\sum_{n \in \mathbb{N}} n(D_{f_n}(x_{n+1}, x_n) + D_{f_n}(x_n, x_{n+1})) < +\infty$.

Proof. (i): We shall derive this result from Theorem 2.8 with $A = \partial\varphi$, $B = \partial\psi$, $\delta_1 = 0$, and $\delta_2 = 1$. First, appealing to [24, Theorem 2.4.4(i)], B is single-valued on $\text{int dom } B = \text{int dom } \psi$ and $B = \nabla\psi$ on $\text{int dom } B$. Next, set $\theta = \varphi + \psi$. Since $\emptyset \neq (\text{int dom } f) \cap \text{dom } \partial\varphi \subset \text{int dom } \psi$, we have $\text{dom } \varphi \cap \text{int dom } \psi \neq \emptyset$. Hence, [9, Theorem 4.1.19] yields $A+B = \partial\theta$. Therefore, $\text{Argmin } \theta = \text{zer } \partial\theta = \text{zer}(A+B)$ and $\mathcal{S} = (\text{int dom } f) \cap \text{zer}(A+B)$. Next, in view of Proposition 2.1(iii), (1.1) is fulfilled. On the other hand, conditions [a] and [b] in Algorithm 2.4 are trivially satisfied. To verify condition [c] in Algorithm 2.4, it suffices to show that, for every $n \in \mathbb{N}$, $(\nabla f_n - \gamma_n B)(C) \subset \text{ran}(\nabla f_n + \gamma_n A)$, i.e., since $C \subset \text{int dom } B$ and $B = \nabla\psi$ on $\text{int dom } B$, that $(\nabla f_n - \gamma_n \nabla\psi)(C) \subset \text{ran}(\nabla f_n + \gamma_n A)$. To do so, fix temporarily $n \in \mathbb{N}$, let $x \in C$, and set

$$A_n = \nabla f_n + \gamma_n A - \nabla f_n(x) + \gamma_n \nabla\psi(x). \quad (3.22)$$

Then, since $\text{dom } \partial f_n \cap \text{dom } A = (\text{int dom } f_n) \cap \text{dom } A = (\text{int dom } f) \cap \text{dom } A \neq \emptyset$ by condition [c] in Algorithm 3.8, it results from [6, Proposition 3.12] that A_n is maximally monotone. Next, we deduce from condition [a] in Algorithm 3.8 and (3.21) that

$$(\forall u \in C)(\forall v \in C) \quad \gamma_n D_\psi(u, v) \leq \alpha(1-\varepsilon) D_\psi(u, v)/\kappa \leq \alpha(1-\varepsilon) D_f(u, v) \leq (1-\varepsilon) D_{f_n}(u, v). \quad (3.23)$$

In turn,

$$\begin{aligned} (\forall u \in C)(\forall v \in C) \quad \gamma_n \langle u - v, \nabla\psi(u) - \nabla\psi(v) \rangle &= \gamma_n (D_\psi(u, v) + D_\psi(v, u)) \\ &\leq (1-\varepsilon) (D_f(u, v) + D_f(v, u)) \\ &= (1-\varepsilon) \langle u - v, \nabla f_n(u) - \nabla f_n(v) \rangle. \end{aligned} \quad (3.24)$$

However, by coercivity of θ , there exists $\rho \in]0, +\infty[$ such that

$$(\forall y \in \mathcal{X}) \quad \|y\| \geq \rho \quad \Rightarrow \quad \inf \langle y, (A+B)(y+x) \rangle = \inf \langle y, \partial\theta(y+x) \rangle \geq \theta(y+x) - \theta(x) \geq 0. \quad (3.25)$$

Now suppose that $(y, y^*) \in \text{gra } A_n(\cdot + x)$ satisfies $\|y\| \geq \rho$. Then $y+x \in \text{dom } \nabla f_n \cap \text{dom } A = (\text{int dom } f_n) \cap \text{dom } A = C$ and $y^* - \nabla f_n(y+x) + \gamma_n \nabla\psi(y+x) + \nabla f_n(x) - \gamma_n \nabla\psi(x) \in \gamma_n (A+B)(y+x)$. Thus, it follows from (3.25) and (3.24) that

$$0 \leq \langle y, y^* \rangle - \langle (y+x) - x, (\nabla f_n - \gamma_n \nabla\psi)(y+x) - (\nabla f_n - \gamma_n \nabla\psi)(x) \rangle \leq \langle y, y^* \rangle. \quad (3.26)$$

Therefore, in view of [22, Proposition 2] and the maximal monotonicity of $A_n(\cdot + x)$, there exists $\bar{y} \in \mathcal{X}$ such that $0 \in A_n(\bar{y} + x)$. Hence $(\nabla f_n - \gamma_n \nabla \psi)(x) \in \nabla f_n(\bar{y} + x) + \gamma_n A(\bar{y} + x) \subset \text{ran}(\nabla f_n + \gamma_n A)$, as desired. Since $(x_{n+1}, \gamma_n^{-1}(\nabla f_n(x_n) - \nabla f_n(x_{n+1})) - \nabla \psi(x_n))$ lies in $\text{gra } \partial \varphi$ by construction, we derive from [6, Proposition 2.3(ii)] that

$$\begin{aligned} (\forall x \in C) \quad \varphi(x) &\geq \varphi(x_{n+1}) - \langle x - x_{n+1}, \nabla \psi(x_n) \rangle + \gamma_n^{-1} \langle x - x_{n+1}, \nabla f_n(x_n) - \nabla f_n(x_{n+1}) \rangle \\ &\geq \varphi(x_{n+1}) - \langle x - x_{n+1}, \nabla \psi(x_n) \rangle \\ &\quad + \gamma_n^{-1} (D_{f_n}(x, x_{n+1}) + D_{f_n}(x_{n+1}, x_n) - D_{f_n}(x, x_n)). \end{aligned} \quad (3.27)$$

On the other hand, (3.23) and the convexity of ψ entail that

$$\begin{aligned} (\forall x \in C) \quad \psi(x_{n+1}) &\leq \psi(x_n) + \langle x_{n+1} - x_n, \nabla \psi(x_n) \rangle + (1 - \varepsilon) \gamma_n^{-1} D_{f_n}(x_{n+1}, x_n) \\ &= \psi(x_n) + \langle x - x_n, \nabla \psi(x_n) \rangle + \langle x_{n+1} - x, \nabla \psi(x_n) \rangle \\ &\quad + (1 - \varepsilon) \gamma_n^{-1} D_{f_n}(x_{n+1}, x_n) \\ &\leq \psi(x) + \langle x_{n+1} - x, \nabla \psi(x_n) \rangle + (1 - \varepsilon) \gamma_n^{-1} D_{f_n}(x_{n+1}, x_n). \end{aligned} \quad (3.28)$$

Altogether, upon adding (3.27) and (3.28), we obtain

$$(\forall x \in C) \quad \theta(x_{n+1}) + \gamma_n^{-1} D_{f_n}(x, x_{n+1}) + \varepsilon \gamma_n^{-1} D_{f_n}(x_{n+1}, x_n) \leq \theta(x) + \gamma_n^{-1} D_{f_n}(x, x_n). \quad (3.29)$$

In particular, since $x_n \in C$,

$$\theta(x_{n+1}) + \gamma_n^{-1} (D_{f_n}(x_n, x_{n+1}) + \varepsilon D_{f_n}(x_{n+1}, x_n)) \leq \theta(x_n). \quad (3.30)$$

This shows that

$$(\theta(x_n))_{n \in \mathbb{N}} \text{ decreases.} \quad (3.31)$$

In turn, using the coercivity of θ , we infer that $(x_n)_{n \in \mathbb{N}}$ is bounded, which secures [a] in Theorem 2.8. It remains to verify that Algorithm 3.8 is focusing. Towards this end, let $z \in \mathcal{S}$ and suppose that

$$(D_{f_n}(z, x_n))_{n \in \mathbb{N}} \text{ converges} \quad (3.32)$$

and

$$\varepsilon \sum_{n \in \mathbb{N}} D_{f_n}(x_{n+1}, x_n) \leq \sum_{n \in \mathbb{N}} (1 - \kappa \gamma_n / \alpha) D_{f_n}(x_{n+1}, x_n) < +\infty. \quad (3.33)$$

Set $\gamma = \inf_{n \in \mathbb{N}} \gamma_n$ and $\ell = \lim D_{f_n}(z, x_n)$. It follows from (3.29) applied to $z \in C$ that

$$(\forall n \in \mathbb{N}) \quad \gamma(\theta(x_{n+1}) - \min \theta(\mathcal{X})) + D_{f_n}(z, x_{n+1}) + \varepsilon D_{f_n}(x_{n+1}, x_n) \leq D_{f_n}(z, x_n) \quad (3.34)$$

and therefore from condition [b] in Algorithm 3.8 that

$$\begin{aligned} (\forall n \in \mathbb{N}) \quad \gamma(\theta(x_{n+1}) - \min \theta(\mathcal{X})) + D_{f_{n+1}}(z, x_{n+1}) + \varepsilon D_{f_n}(x_{n+1}, x_n) \\ \leq (1 + \eta_n) \left(\gamma(\theta(x_{n+1}) - \min \theta(\mathcal{X})) + D_{f_n}(z, x_{n+1}) + \varepsilon D_{f_n}(x_{n+1}, x_n) \right) \\ \leq (1 + \eta_n) D_{f_n}(z, x_n). \end{aligned} \quad (3.35)$$

Hence, $\overline{\lim} \gamma(\theta(x_{n+1}) - \min \theta(\mathcal{X})) + \ell \leq \ell$ and therefore $\overline{\lim}(\theta(x_{n+1}) - \min \theta(\mathcal{X})) = 0$. Thus

$$\theta(x_n) \rightarrow \min \theta(\mathcal{X}). \quad (3.36)$$

Now take $x \in \mathfrak{W}(x_n)_{n \in \mathbb{N}}$, say $x_{k_n} \rightharpoonup x$. By weak lower semicontinuity of θ , $\min \theta(\mathcal{X}) \leq \theta(x) \leq \liminf \theta(x_{k_n}) = \min \theta(\mathcal{X})$ and it follows that $x \in \text{Argmin } \theta = \text{zer}(A + B)$. Consequently, Theorem 2.8 asserts that $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in $\mathcal{S}$.

(ii): Combine (3.31) and (3.36).

(iii)&(iv): Fix $z \in \mathcal{S}$ and set $\gamma = \inf_{n \in \mathbb{N}} \gamma_n$. Arguing along the same lines as above, we obtain

$$(\forall n \in \mathbb{N}) \quad \gamma(\theta(x_{n+1}) - \min \theta(\mathcal{X})) + D_{f_{n+1}}(z, x_{n+1}) + \varepsilon D_{f_n}(x_{n+1}, x_n) \leq (1 + \eta_n) D_{f_n}(z, x_n) \quad (3.37)$$

and therefore [7, Lemma 5.31] guarantees that $\sum_{n \in \mathbb{N}} (\theta(x_n) - \min \theta(\mathcal{X})) < +\infty$. In addition, $(\theta(x_n) - \min \theta(\mathcal{X}))_{n \in \mathbb{N}}$ is decreasing by virtue of (3.31). However, recall that if $(\alpha_n)_{n \in \mathbb{N}}$ is a decreasing sequence in $[0, +\infty[$ such that $\sum_{n \in \mathbb{N}} \alpha_n < +\infty$, then

$$\alpha_n = o\left(\frac{1}{n}\right) \quad \text{and} \quad \sum_{n \in \mathbb{N}} n(\alpha_n - \alpha_{n+1}) < +\infty. \quad (3.38)$$

Hence, $\theta(x_n) - \min \theta(\mathcal{X}) = o(1/n)$ and $\sum_{n \in \mathbb{N}} n(\theta(x_n) - \theta(x_{n+1})) < +\infty$. Consequently, since (3.29) yields

$$(\forall n \in \mathbb{N}) \quad \gamma_n^{-1} D_{f_n}(x_n, x_{n+1}) + \varepsilon \gamma_n^{-1} D_{f_n}(x_{n+1}, x_n) \leq \theta(x_n) - \theta(x_{n+1}), \quad (3.39)$$

we infer that $\sum_{n \in \mathbb{N}} n(D_{f_n}(x_{n+1}, x_n) + D_{f_n}(x_n, x_{n+1})) < +\infty$. $\square$

Remark 3.10 Let us relate Theorem 3.9 to the literature.

- (i) The conclusions of items (i) and (ii) are obtained in [18, Theorem 1(2)] under more restrictive conditions on the sequences $(\gamma_n)_{n \in \mathbb{N}}$ and $(f_n)_{n \in \mathbb{N}}$. Thus, we do not require in Theorem 3.9 the additional condition $(\forall n \in \mathbb{N}) (1 + \eta_n) \gamma_n - \gamma_{n+1} \leq \alpha \eta_n / \kappa$. Furthermore, we do not suppose either that $-\text{ran } \nabla \psi \subset \text{dom } \varphi^*$ or that the functions $(f_n)_{n \in \mathbb{N}}$ are cofinite.
- (ii) Items (iii) and (iv) are new even in Euclidean spaces. In the finite-dimensional setting, partial results can be found in [3], where:
 - (a) A single convex function is used: $(\forall n \in \mathbb{N}) f_n = f$.
 - (b) The viability of the sequence $(x_n)_{n \in \mathbb{N}}$ is a blanket assumption, while it is guaranteed in Theorem 3.9.
 - (c) Only the rates $\sum_{n \in \mathbb{N}} D_f(x_{n+1}, x_n) < +\infty$ and $(\varphi + \psi)(x_n) - \min(\varphi + \psi)(\mathcal{X}) = O(1/n)$ are obtained.

3.4 Further applications

Theorems 2.8 and 3.9 operate under broad assumptions which go beyond those of the existing forward-backward methods of [6, 15, 18, 20] described in (1.4)–(1.7). Here are two examples which do not fit the existing scenarios and exploit this generality.

Example 3.11 Consider the setting of Problem 1.1. Suppose, in addition, that the following hold:

- [a] A is uniformly monotone on bounded sets.
- [b] There exist $\psi \in \Gamma_0(\mathcal{X})$ and $\kappa \in]0, +\infty[$ such that $B = \partial \psi$ and $(\forall x \in C)(\forall y \in C) D_\psi(x, y) \leq \kappa D_f(x, y)$.

[c] f is supercoercive.

[d] $\text{zer}(A + B) \subset \text{int dom } f$.

Let $(\gamma_n)_{n \in \mathbb{N}}$ be a sequence in $]0, +\infty[$ such that $0 < \inf_{n \in \mathbb{N}} \gamma_n \leq \sup_{n \in \mathbb{N}} \gamma_n < 1/\kappa$, take $x_0 \in C$, and set $(\forall n \in \mathbb{N}) x_{n+1} = (\nabla f + \gamma_n A)^{-1}(\nabla f(x_n) - \gamma_n \nabla \psi(x_n))$. Then $(x_n)_{n \in \mathbb{N}}$ converges strongly to the unique zero of $A + \nabla \psi$.

The next example concerns variational inequalities.

Example 3.12 Let $\varphi \in \Gamma_0(\mathcal{X})$, let $B: \mathcal{X} \rightarrow 2^{\mathcal{X}^*}$ be maximally monotone, let $f \in \Gamma_0(\mathcal{X})$ be essentially smooth, and set $C = (\text{int dom } f) \cap \text{dom } \partial \varphi$. Suppose that $C \subset \text{int dom } B$ and B is single-valued on $\text{int dom } B$. Consider the problem of finding a point in

$$\mathcal{S} = \{x \in C \mid (\forall y \in \mathcal{X}) \langle x - y, Bx \rangle + \varphi(x) \leq \varphi(y)\}, \quad (3.40)$$

which is assumed to be nonempty. This is a special case of Problem 1.1 with $A = \partial \varphi$ and, given $x_0 \in C$, Algorithm 2.4 produces the iterations $(\forall n \in \mathbb{N}) x_{n+1} = \text{prox}_{\gamma_n \varphi}^{f_n}(\nabla f_n(x_n) - \gamma_n Bx_n)$. The weak convergence of $(x_n)_{n \in \mathbb{N}}$ to a point in $\mathcal{S}$ is discussed in Theorem 2.8. Even in Euclidean spaces, this scheme is new and of interest since, as shown in [3, 13, 18], the Bregman proximity operator $\text{prox}_{\gamma_n \varphi}^{f_n}$ may be easier to compute for a particular f_n than for the standard kernel $\|\cdot\|^2/2$. Altogether, our framework makes it possible to solve variational inequalities by forward-backward splitting with non-cocoercive operators and/or outside of Hilbert spaces.

References

- [1] J.-B. Baillon and G. Haddad, Quelques propriétés des opérateurs angle-bornés et n -cycliquement monotones, *Israel J. Math.*, vol. 26, pp. 137–150, 1977.
- [2] S. Banach, *Théorie des Opérations Linéaires*. Seminar. Matem. Univ. Warszawa, 1932.
- [3] H. H. Bauschke, J. Bolte, and M. Teboulle, A descent lemma beyond Lipschitz gradient continuity: First-order methods revisited and applications, *Math. Oper. Res.*, vol. 42, pp. 330–348, 2017.
- [4] H. H. Bauschke and J. M. Borwein, Legendre functions and the method of random Bregman projections, *J. Convex Anal.*, vol. 4, pp. 27–67, 1997.
- [5] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces, *Commun. Contemp. Math.*, vol. 3, pp. 615–647, 2001.
- [6] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, Bregman monotone optimization algorithms, *SIAM J. Control Optim.*, vol. 42, pp. 596–636, 2003.
- [7] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed. Springer, New York, 2017.
- [8] H. H. Bauschke, M. N. Dao, and S. B. Lindstrom, Regularizing with Bregman–Moreau envelopes, *SIAM J. Optim.*, vol. 28, pp. 3208–3228, 2018.
- [9] J. M. Borwein and J. D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples*. Cambridge University Press, 2010.
- [10] N. Bourbaki, *Espaces Vectoriels Topologiques, Chapitres 1 à 5*. Masson, Paris, 1981. English translation: *Topological Vector Spaces, Chapters 1–5*. Springer-Verlag, New York, 1987.
- [11] H. Brézis and A. Haraux, Image d’une somme d’opérateurs monotones et applications, *Israel J. Math.*, vol. 23, pp. 165–186, 1976.

- [12] Y. Censor and S. A. Zenios, *Parallel Optimization – Theory, Algorithms and Applications*. Oxford University Press, New York, 1997.
- [13] P. L. Combettes and Q. V. Nguyen, Solving composite monotone inclusions in reflexive Banach spaces by constructing best Bregman approximations from their Kuhn-Tucker set, *J. Convex Anal.*, vol. 23, pp. 481–510, 2016.
- [14] P. L. Combettes and B. C. Vũ, Variable metric quasi-Fejér monotonicity, *Nonlinear Anal.*, vol. 78, pp. 17–31, 2013.
- [15] P. L. Combettes and B. C. Vũ, Variable metric forward-backward splitting with applications to monotone inclusions in duality, *Optimization*, vol. 63, pp. 1289–1318, 2014.
- [16] J. Frecon, S. Salzo, and M. Pontil, Bilevel learning of the group lasso structure, *Adv. Neural Inform. Process. Syst.*, vol. 31, pp. 8301–8311, 2018.
- [17] B. Mercier, *Topics in Finite Element Solution of Elliptic Problems* (Lectures on Mathematics, no. 63). Tata Institute of Fundamental Research, Bombay, 1979.
- [18] Q. V. Nguyen, Forward-backward splitting with Bregman distances, *Vietnam J. Math.*, vol. 45, pp. 519–539, 2017.
- [19] G. Ortiz-Jiménez, M. El Gheche, E. Simou, H. Petric Maretić, and P. Frossard, Forward-backward splitting for optimal transport based problems, *Proc. Intl. Conf. Acoust., Speech, Signal Process.*, pp. 5405–5409, 2020.
- [20] A. Renaud and G. Cohen, An extension of the auxiliary problem principle to nonsymmetric auxiliary operators, *ESAIM Control Optim. Calc. Var.*, vol. 2, pp. 281–306, 1997.
- [21] R. T. Rockafellar, Local boundedness of nonlinear, monotone operators, *Michigan Math. J.*, vol. 16, pp. 397–407, 1969.
- [22] R. T. Rockafellar, On the maximality of sums of nonlinear monotone operators, *Trans. Amer. Math. Soc.*, vol. 149, pp. 75–88, 1970.
- [23] S. Salzo, The variable metric forward-backward splitting algorithm under mild differentiability assumptions, *SIAM J. Optim.*, vol. 27, pp. 2153–2181, 2017.
- [24] C. Zălinescu, *Convex Analysis in General Vector Spaces*. World Scientific Publishing, River Edge, NJ, 2002.