Solving Composite Monotone Inclusions in Reflexive Banach Spaces by Constructing Best Bregman Approximations from Their Kuhn-Tucker Set

Patrick L. Combettes and Quang Van Nguyen

Sorbonne Universités – UPMC Univ. Paris 06
UMR 7598, Laboratoire Jacques-Louis Lions
F-75005 Paris, France
plc@ljll.math.upmc.fr, quangnv@ljll.math.upmc.fr

In memory of Jean Jacques Moreau (1923–2014)

Abstract

We introduce the first operator splitting method for composite monotone inclusions outside of Hilbert spaces. The proposed primal-dual method constructs iteratively the best Bregman approximation to an arbitrary point from the Kuhn-Tucker set of a composite monotone inclusion. Strong convergence is established in reflexive Banach spaces without requiring additional restrictions on the monotone operators or knowledge of the norms of the linear operators involved in the model. The monotone operators are activated via Bregman distance-based resolvent operators. The method is novel even in Euclidean spaces, where it provides an alternative to the usual proximal methods based on the standard distance.

Key words. Best approximation, Banach space, Bregman distance, duality, Legendre function, monotone operator, operator splitting, primal-dual algorithm.

1 Introduction

Let $\mathcal{X}$ be a reflexive real Banach space with norm $\|\cdot\|$ and let $\langle\cdot,\cdot\rangle$ be the duality pairing between $\mathcal{X}$ and its topological dual $\mathcal{X}^*$. A set-valued operator $M\colon \mathcal{X}\to 2^{\mathcal{X}^*}$ with graph $\operatorname{gra} M = \big\{(x,x^*)\in\mathcal{X}\times\mathcal{X}^* \;\big|\; x^*\in Mx\big\}$ is monotone if

$$(\forall (x_1,x_1^*)\in\operatorname{gra} M)(\forall (x_2,x_2^*)\in\operatorname{gra} M)\quad \langle x_1-x_2, x_1^*-x_2^*\rangle \geqslant 0, \tag{1.1}$$

and maximally monotone if, furthermore, there exists no monotone operator from $\mathcal{X}$ to $2^{\mathcal{X}^*}$ the graph of which properly contains $\operatorname{gra} M$. Monotone operator theory emerged in the early 1960s as a well-structured branch of nonlinear analysis [24, 29, 30, 41], and it remains very active [9, 10, 35, 42]. One of the main reasons for the success of the theory is that a significant range of problems in areas such as optimization, economics, variational inequalities, partial differential equations, mechanics, signal and image processing, optimal transportation, machine learning, and traffic theory can be reduced to solving inclusions of the type

$$\text{find}\;\; x\in\mathcal{X}\quad\text{such that}\quad 0\in Mx, \tag{1.2}$$

where $M\colon\mathcal{X}\to 2^{\mathcal{X}^*}$ is maximally monotone. Conceptually, this inclusion can be solved via the Bregman proximal point algorithm, special instances of which go back to [22, 25, 36]. To present its general form [7], we need the following definitions, which revolve around the notion of a Bregman distance pioneered in [13].

Definition 1.1 [6, 7] Let $\mathcal{X}$ be a reflexive real Banach space and let $f\colon\mathcal{X}\to\left]-\infty,+\infty\right]$ be a proper lower semicontinuous convex function, with conjugate $f^*\colon\mathcal{X}^*\to\left]-\infty,+\infty\right]\colon x^*\mapsto \sup_{x\in\mathcal{X}}(\langle x,x^*\rangle - f(x))$ and Moreau subdifferential [32]

$$\partial f\colon\mathcal{X}\to 2^{\mathcal{X}^*}\colon x\mapsto\big\{x^*\in\mathcal{X}^* \;\big|\; (\forall y\in\mathcal{X})\,\langle y-x,x^*\rangle + f(x)\leqslant f(y)\big\}. \tag{1.3}$$

Then 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 $\operatorname{dom}\partial f$. Moreover, f is Gâteaux differentiable on $\operatorname{int}\operatorname{dom} f\neq\varnothing$ and the associated *Bregman distance* is

$$\begin{aligned} D^f\colon \mathcal{X}\times\mathcal{X} &\to [0,+\infty]\\ (x,y)&\mapsto\begin{cases} f(x)-f(y)-\langle x-y,\nabla f(y)\rangle, & \text{if }\; y\in\operatorname{int}\operatorname{dom} f;\\ +\infty, & \text{otherwise.}\end{cases}\end{aligned} \tag{1.4}$$

Let C be a closed convex subset of $\mathcal{X}$ such that $C\cap\operatorname{int}\operatorname{dom} f\neq\varnothing$. The *Bregman projector* onto C induced by f is

$$\begin{aligned} P_C^f\colon \operatorname{int}\operatorname{dom} f &\to C\cap\operatorname{int}\operatorname{dom} f\\ y&\mapsto \underset{x\in C}{\operatorname{argmin}}\, D^f(x,y).\end{aligned} \tag{1.5}$$

The fact that, for every $y \in \operatorname{int}\operatorname{dom} f$, $P_C^f y \in \operatorname{int}\operatorname{dom} f$ exists and is unique is established in [6, Corollary 7.9]. It follows from [7, Theorem 5.18] that, under suitable assumptions on f and M, given a sequence $(\gamma_n)_{n\in\mathbb{N}}$ in $]0,+\infty[$ such that $\inf_{n\in\mathbb{N}} \gamma_n > 0$, the sequence defined by

$$x_0 \in \operatorname{int}\operatorname{dom} f \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = (\nabla f + \gamma_n M)^{-1} \circ \nabla f(x_n) \tag{1.6}$$

converges weakly to a solution to (1.2) (in the case when $\mathcal{X}$ is a Hilbert space and $f = \|\cdot\|^2/2$, $(\nabla f+\gamma_n M)^{-1}\circ\nabla f$ reduces to the standard resolvent $J_{\gamma_n M}$ and we obtain the classical result of [34, Theorem 1]). A strongly convergent variant of (1.6) was proposed in [8]. In applications, however, M is typically too complex for (1.6) to be implementable. For instance, given a real Banach space $\mathcal{Y}$, a typical composite model is $M = A + L^*BL$, where $A\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ and $B\colon \mathcal{Y} \to 2^{\mathcal{Y}^*}$ are monotone, and $L\colon \mathcal{X} \to \mathcal{Y}$ is linear and bounded. In Hilbert spaces, if $\mathcal{X} = \mathcal{Y}$ and $L = \mathrm{Id}$, several well-established splitting methods are available to solve (1.2), i.e., to find a zero of $A + B$ using A and B separately at each iteration [9, 27, 28, 37]. Splitting methods for the more versatile composite model $M = A + L^*BL$ in Hilbert spaces were first proposed in [14] (see [1, 11, 12, 19, 20, 38] for subsequent developments). These methods provide in general only weak convergence to an unspecified solution and, in addition, they require knowledge of $\|L\|$ or potentially costly inversions of linear operators. The recent method primal-dual method of [3] circumvents these limitations and, in addition, converges to the best approximation to a reference point from the Kuhn-Tucker set relative to the underlying hilbertian distance. The objective of this paper is to extend it to reflexive Banach spaces and to best approximation relative to general Bregman distances. Let us stress that the theory of splitting algorithms in Banach spaces is rather scarce as most hilbertian splitting methods cannot be naturally extended to that setting; in particular, to the best of our knowledge there exists at present no splitting algorithm for finding a zero of $M = A + L^*BL$ outside of Hilbert spaces. By contrast, the geometric primal-dual construction of [3], which consists in projecting a reference point onto successive simple outer approximations to the Kuhn-Tucker set of the inclusion $0 \in Ax + L^*BLx$, lends itself to such an extension. Our analysis will borrow tools on Legendre functions and Bregman-based algorithms from [6] and [7], as well as geometric constructs from [3] and [8]. The proposed results will provide not only the first splitting methods for composite inclusions outside of Hilbert spaces, but also new algorithms in Hilbert, and even Euclidean, spaces.

The problem under consideration is the following.

Problem 1.2 Let $\mathcal{X}$ and $\mathcal{Y}$ be reflexive real Banach spaces such that $\mathcal{X} \neq \{0\}$ and $\mathcal{Y} \neq \{0\}$, let $\boldsymbol{\mathcal{X}}$ be the standard product vector space $\mathcal{X} \times \mathcal{Y}^*$ equipped with the norm $(x, y^*) \mapsto \sqrt{\|x\|^2 + \|y^*\|^2}$, and let $\boldsymbol{\mathcal{X}}^*$ be its topological dual, that is, $\mathcal{X}^* \times \mathcal{Y}$ equipped with the norm $(x^*, y) \mapsto \sqrt{\|x^*\|^2 + \|y\|^2}$. Let $A\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ and $B\colon \mathcal{Y} \to 2^{\mathcal{Y}^*}$ be maximally monotone, and let $L\colon \mathcal{X} \to \mathcal{Y}$ be linear and bounded. Consider the inclusion problem

$$\text{find} \ \ x \in \mathcal{X} \ \ \text{such that} \ \ 0 \in Ax + L^*BLx, \tag{1.7}$$

the dual problem

$$\text{find} \ \ y^* \in \mathcal{Y}^* \ \ \text{such that} \ \ 0 \in -LA^{-1}(-L^*y^*) + B^{-1}y^*, \tag{1.8}$$

and let

$$\boldsymbol{Z} = \big\{(x, y^*) \in \boldsymbol{\mathcal{X}} \;\big|\; -L^*y^* \in Ax \quad \text{and} \quad Lx \in B^{-1}y^*\big\} \tag{1.9}$$

be the associated Kuhn-Tucker set. Let $f\colon \mathcal{X} \to \,]-\infty,+\infty]$ and $g\colon \mathcal{Y} \to \,]-\infty,+\infty]$ be Legendre functions, set

$$\boldsymbol{f}\colon \boldsymbol{\mathcal{X}} \to \,]-\infty,+\infty] : (x,y^*) \mapsto f(x) + g^*(y^*), \tag{1.10}$$

let $x_0 \in \operatorname{int}\operatorname{dom} f$, let $y_0^* \in \operatorname{int}\operatorname{dom} g^*$, and suppose that $\boldsymbol{Z} \cap \operatorname{int}\operatorname{dom} \boldsymbol{f} \neq \varnothing$. The problem is to find the best Bregman approximation $(\overline{x},\overline{y}^*) = P^{\boldsymbol{f}}_{\boldsymbol{Z}}(x_0,y_0^*)$ to (x_0,y_0^*) from $\boldsymbol{Z}$.

Notation. The symbols $\rightharpoonup$ and $\to$ denote respectively weak and strong convergence. The set of weak sequential cluster points of a sequence $(x_n)_{n\in\mathbb{N}}$ is denoted by $\mathfrak{W}(x_n)_{n\in\mathbb{N}}$. The closed ball of center $x \in \mathcal{X}$ and radius $\rho \in \,]0,+\infty[$ is denoted by $B(x;\rho)$. Let $M\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be a set-valued operator. The domain of M is $\operatorname{dom} M = \big\{x \in \mathcal{X} \;\big|\; Mx \neq \varnothing\big\}$, the range of M is $\operatorname{ran} M = \big\{x^* \in \mathcal{X}^* \;\big|\; (\exists\, x \in \mathcal{X})\; x^* \in Mx\big\}$, and the set of zeros of M is $\operatorname{zer} M = \big\{x \in \mathcal{X} \;\big|\; 0 \in Mx\big\}$. $\Gamma_0(\mathcal{X})$ is the class of all lower semicontinuous convex functions $f\colon \mathcal{X} \to \,]-\infty,+\infty]$ such that $\operatorname{dom} f = \big\{x \in \mathcal{X} \;\big|\; f(x) < +\infty\big\} \neq \varnothing$. Let $f\colon \mathcal{X} \to \,]-\infty,+\infty]$. Then f is coercive if $\lim_{\|x\|\to+\infty} f(x) = +\infty$ and supercoercive if $\lim_{\|x\|\to+\infty} f(x)/\|x\| = +\infty$.

2 Preliminary results

2.1 Properties of the Kuhn-Tucker set

The following proposition revisits and complements some results of [2] and [14] on the properties of the Kuhn-Tucker set in the more general setting of Problem 1.2.

Proposition 2.1 *Consider the setting of Problem 1.2. Then the following hold:*

(i) *Let $\mathscr{P}$ be the set of solutions to (1.7) and let $\mathscr{D}$ be the set of solutions to (1.8). Then the following hold:*

  (a) *$\boldsymbol{Z}$ is a closed convex subset of $\mathscr{P} \times \mathscr{D}$.*

  (b) *Set $\mathcal{Q}_{\mathcal{X}}\colon \boldsymbol{\mathcal{X}} \to \mathcal{X}\colon (x,y^*) \mapsto x$ and $\mathcal{Q}_{\mathcal{Y}^*}\colon \boldsymbol{\mathcal{X}} \to \mathcal{Y}^*\colon (x,y^*) \mapsto y^*$. Then $\mathscr{P} = \mathcal{Q}_{\mathcal{X}}(\boldsymbol{Z})$ and $\mathscr{D} = \mathcal{Q}_{\mathcal{Y}^*}(\boldsymbol{Z})$.*

  (c) *$\mathscr{P} \neq \varnothing \Leftrightarrow \boldsymbol{Z} \neq \varnothing \Leftrightarrow \mathscr{D} \neq \varnothing$.*

(ii) *For every $\mathsf{a} = (a,a^*) \in \operatorname{gra} A$ and $\mathsf{b} = (b,b^*) \in \operatorname{gra} B$, set $\boldsymbol{s}^*_{\mathsf{a},\mathsf{b}} = (a^* + L^*b^*, b - La)$, $\eta_{\mathsf{a},\mathsf{b}} = \langle a,a^*\rangle + \langle b,b^*\rangle$, and $\boldsymbol{H}_{\mathsf{a},\mathsf{b}} = \big\{\boldsymbol{x} \in \boldsymbol{\mathcal{X}} \;\big|\; \langle \boldsymbol{x}, \boldsymbol{s}^*_{\mathsf{a},\mathsf{b}}\rangle \leqslant \eta_{\mathsf{a},\mathsf{b}}\big\}$. Then the following hold:*

  (a) $(\forall \mathsf{a} \in \operatorname{gra} A)(\forall \mathsf{b} \in \operatorname{gra} B)\ \big[\boldsymbol{s}^*_{\mathsf{a},\mathsf{b}} = \boldsymbol{0} \;\Leftrightarrow\; \boldsymbol{H}_{\mathsf{a},\mathsf{b}} = \boldsymbol{\mathcal{X}} \;\Rightarrow\; (a,b^*) \in \boldsymbol{Z}$ *and* $\eta_{\mathsf{a},\mathsf{b}} = 0\big]$.

  (b) $\boldsymbol{Z} = \bigcap_{\mathsf{a}\in\operatorname{gra} A} \bigcap_{\mathsf{b}\in\operatorname{gra} B} \boldsymbol{H}_{\mathsf{a},\mathsf{b}}$.

(iii) *Let $(a_n,a_n^*)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} A$, let $(b_n,b_n^*)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} B$, and let $(x,y^*) \in \boldsymbol{\mathcal{X}}$. Suppose that $a_n \rightharpoonup x$, $b_n^* \rightharpoonup y^*$, $a_n^* + L^*b_n^* \to 0$, and $La_n - b_n \to 0$. Then $(x,y^*) \in \boldsymbol{Z}$.*

Proof. Set $\boldsymbol{M}\colon \boldsymbol{\mathcal{X}} \to 2^{\boldsymbol{\mathcal{X}}^*}\colon (x, y^*) \mapsto Ax \times B^{-1}y^*$ and $\boldsymbol{S}\colon \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{X}}^*\colon (x, y^*) \mapsto (L^*y^*, -Lx)$. Since A and B^{-1} are maximally monotone, so is $\boldsymbol{M}$. On the other hand, $\boldsymbol{S}$ is linear, bounded, and positive since

$$(\forall (x, y^*) \in \boldsymbol{\mathcal{X}}) \quad \langle \boldsymbol{S}(x, y^*), (x, y^*) \rangle = \langle x, L^*y^* \rangle + \langle -Lx, y^* \rangle = 0. \tag{2.1}$$

Thus, it follows from [35, Section 17] that $\boldsymbol{S}$ is maximally monotone with $\operatorname{dom} \boldsymbol{S} = \boldsymbol{\mathcal{X}}$. In turn, we derive from [35, Theorem 24.1(a)] that

$$\boldsymbol{M} + \boldsymbol{S} \text{ is maximally monotone.} \tag{2.2}$$

(i)(a): Let $(x, y^*) \in \boldsymbol{\mathcal{X}}$. Then

$$\begin{aligned} \mathbf{0} \in \boldsymbol{M}(x, y^*) + \boldsymbol{S}(x, y^*) \quad &\Leftrightarrow \quad 0 \in Ax + L^*y^* \quad \text{and} \quad 0 \in B^{-1}y^* - Lx \\ &\Leftrightarrow \quad -L^*y^* \in Ax \quad \text{and} \quad Lx \in B^{-1}y^* \\ &\Leftrightarrow \quad (x, y^*) \in \boldsymbol{Z}. \end{aligned} \tag{2.3}$$

Therefore, we derive from (2.2) and [15, Lemma 1.1(a)] that $\boldsymbol{Z} = \operatorname{zer}(\boldsymbol{M} + \boldsymbol{S}) = (\boldsymbol{M} + \boldsymbol{S})^{-1}(0)$ is closed and convex.

(i)(b): Let $x \in \mathcal{X}$. Then $x \in \mathscr{P} \Leftrightarrow 0 \in Ax + L^*BLx \Leftrightarrow (\exists y^* \in \mathcal{Y}^*)\ [-L^*y^* \in Ax \text{ and } y^* \in BLx] \Leftrightarrow (\exists y^* \in \mathcal{Y}^*)\ (x, y^*) \in \boldsymbol{Z}$. Hence $\mathscr{P} \neq \varnothing \Leftrightarrow \boldsymbol{Z} \neq \varnothing$. Likewise, let $y^* \in \mathcal{Y}^*$. Then $y^* \in \mathscr{D} \Leftrightarrow 0 \in -LA^{-1}(-L^*y^*) + B^{-1}y^* \Leftrightarrow (\exists x \in \mathcal{X})\ [x \in A^{-1}(-L^*y^*) \text{ and } 0 \in -Lx + B^{-1}y^*] \Leftrightarrow (\exists x \in \mathcal{X})\ [-L^*y^* \in Ax \text{ and } Lx \in B^{-1}y^*] \Leftrightarrow (\exists x \in \mathcal{X})\ (x, y^*) \in \boldsymbol{Z}$.

(i)(c): Clear from (i)(b) (see also [33, Corollary 2.4]).

(ii)(a) : Let $\mathsf{a} \in \operatorname{gra} A$ and $\mathsf{b} \in \operatorname{gra} B$. Then $\boldsymbol{s}^*_{\mathsf{a},\mathsf{b}} = \mathbf{0} \Rightarrow [-L^*b^* = a^* \in Aa \text{ and } La = b \in B^{-1}b^*] \Rightarrow (a, b^*) \in \boldsymbol{Z}$. In addition,

$$\begin{aligned} \boldsymbol{s}^*_{\mathsf{a},\mathsf{b}} = \mathbf{0} \quad \Rightarrow \quad \eta_{\mathsf{a},\mathsf{b}} = \langle a, a^* \rangle + \langle b, b^* \rangle &= \langle a, -L^*b^* \rangle + \langle La, b^* \rangle \\ &= -\langle La, b^* \rangle + \langle La, b^* \rangle = 0. \end{aligned} \tag{2.4}$$

Thus $\boldsymbol{s}^*_{\mathsf{a},\mathsf{b}} = \mathbf{0} \Rightarrow \boldsymbol{H}_{\mathsf{a},\mathsf{b}} = \boldsymbol{\mathcal{X}}$. Conversely, $\boldsymbol{H}_{\mathsf{a},\mathsf{b}} = \boldsymbol{\mathcal{X}} \Rightarrow \boldsymbol{s}^*_{\mathsf{a},\mathsf{b}} = \mathbf{0} \Rightarrow \eta_{\mathsf{a},\mathsf{b}} = 0$.

(ii)(b): First, suppose that $\boldsymbol{x} = (x, y^*) \in \bigcap_{\mathsf{a} \in \operatorname{gra} A} \bigcap_{\mathsf{b} \in \operatorname{gra} B} \boldsymbol{H}_{\mathsf{a},\mathsf{b}}$. Then

$$\begin{aligned} (\forall \mathsf{a} \in \operatorname{gra} A)(\forall \mathsf{b} \in \operatorname{gra} B) \quad &\langle (a, b^*) - (x, y^*), (a^*, b) - (-L^*y^*, Lx) \rangle \\ &= \langle (a - x, b^* - y^*), (a^* + L^*y^*, b - Lx) \rangle \\ &= \langle a - x, a^* + L^*y^* \rangle + \langle b - Lx, b^* - y^* \rangle \\ &= \eta_{\mathsf{a},\mathsf{b}} - \langle \boldsymbol{x}, \boldsymbol{s}^*_{\mathsf{a},\mathsf{b}} \rangle \\ &\geqslant 0. \end{aligned} \tag{2.5}$$

On the other hand, since

$$\big\{ \big((a, b^*), (a^*, b) \big) \ \big|\ \mathsf{a} \in \operatorname{gra} A, \mathsf{b} \in \operatorname{gra} B \big\} = \operatorname{gra} \boldsymbol{M}, \tag{2.6}$$

it follows from (2.2) and (2.5) that $((x,y^*),(-L^*y^*,Lx)) \in \operatorname{gra} \boldsymbol{M}$, i.e., $\boldsymbol{x} \in \boldsymbol{Z}$. Thus

$$\bigcap_{\mathsf{a}\in\operatorname{gra} A}\ \bigcap_{\mathsf{b}\in\operatorname{gra} B} \boldsymbol{H}_{\mathsf{a},\mathsf{b}} \subset \boldsymbol{Z}. \tag{2.7}$$

Conversely, let $\mathsf{a} \in \operatorname{gra} A$, let $\mathsf{b} \in \operatorname{gra} B$, and let $(x,y^*) \in \boldsymbol{Z}$. Then $(x,-L^*y^*) \in \operatorname{gra} A$ and $(Lx,y^*) \in \operatorname{gra} B$. Since A and B are monotone, we obtain

$$\langle a-x, a^*+L^*y^*\rangle \geqslant 0 \quad \text{and} \quad \langle b-Lx, b^*-y^*\rangle \geqslant 0. \tag{2.8}$$

Adding these two inequalities yields

$$\langle x-a, a^*+L^*y^*\rangle + \langle Lx-b, b^*-y^*\rangle \leqslant 0 \tag{2.9}$$

and, therefore,

$$\begin{aligned}
\langle \boldsymbol{x}, \boldsymbol{s}^*_{\mathsf{a},\mathsf{b}}\rangle &= \langle x, a^*+L^*b^*\rangle + \langle b-La, y^*\rangle \\
&= \langle x, a^*+L^*y^*\rangle + \langle Lx, b^*-y^*\rangle + \langle b-Lx, y^*\rangle + \langle x-a, L^*y^*\rangle \\
&= \langle x-a, a^*+L^*y^*\rangle + \langle a, a^*\rangle + \langle La, y^*\rangle \\
&\quad + \langle Lx-b, b^*-y^*\rangle + \langle b, b^*\rangle - \langle b, y^*\rangle + \langle b-Lx, y^*\rangle + \langle x-a, L^*y^*\rangle \\
&\leqslant \langle a, a^*\rangle + \langle b, b^*\rangle + \langle La-b, y^*\rangle + \langle b-Lx, y^*\rangle + \langle x-a, L^*y^*\rangle \\
&= \langle a, a^*\rangle + \langle b, b^*\rangle \\
&= \eta_{\mathsf{a},\mathsf{b}}.
\end{aligned} \tag{2.10}$$

This implies that $(x,y^*) \in \boldsymbol{H}_{\mathsf{a},\mathsf{b}}$. Hence, $\boldsymbol{Z} \subset \boldsymbol{H}_{\mathsf{a},\mathsf{b}}$.

(iii): Set $(\forall n \in \mathbb{N})$ $\boldsymbol{x}_n = (a_n, b_n^*)$ and $\boldsymbol{x}_n^* = (a_n^*+L^*b_n^*, b_n-La_n)$. Then $\boldsymbol{x}_n \rightharpoonup (x,y^*)$, $\boldsymbol{x}_n^* \to \boldsymbol{0}$, and $(\forall n \in \mathbb{N})$ $(\boldsymbol{x}_n, \boldsymbol{x}_n^*) \in \operatorname{gra}(\boldsymbol{M}+\boldsymbol{S})$. However, it follows from (2.2) that $\operatorname{gra}(\boldsymbol{M}+\boldsymbol{S})$ is sequentially closed in $\boldsymbol{\mathcal{X}}^{\mathrm{weak}} \times \boldsymbol{\mathcal{X}}^{*\mathrm{strong}}$ [15, Lemma 1.2]. Hence, $\boldsymbol{0} \in (\boldsymbol{M}+\boldsymbol{S})(x,y^*)$, i.e., by (2.3), $(x,y^*) \in \boldsymbol{Z}$. □

Proposition 2.2 *Consider the setting of Problem 1.2. Then the following hold:*

(i) *$\boldsymbol{f}$ is a Legendre function.*

(ii) *The solution $(\overline{x},\overline{y}^*)$ to Problem 1.2 exists and is unique.*

Proof. (i): Since f and g are Legendre functions, so are f^* and g^* [6, Corollary 5.5]. Therefore, it follows from [6, Theorem 5.6(iii)] that ∂f and ∂g^* are single-valued on $\operatorname{dom}\partial f = \operatorname{int}\operatorname{dom} f$ and $\operatorname{dom}\partial g^* = \operatorname{int}\operatorname{dom} g^*$, respectively. On the other hand, we derive from (1.10) that $\operatorname{dom}\boldsymbol{f} = \operatorname{dom} f \times \operatorname{dom} g^*$ and that

$$\partial\boldsymbol{f} \colon \boldsymbol{\mathcal{X}} \to 2^{\boldsymbol{\mathcal{X}}^*} \colon (x,y^*) \mapsto \partial f(x) \times \partial g^*(y^*). \tag{2.11}$$

Thus, $\partial\boldsymbol{f}$ is single-valued on

$$\operatorname{dom}\partial\boldsymbol{f} = \operatorname{dom}\partial f \times \operatorname{dom}\partial g^* = \operatorname{int}\operatorname{dom} f \times \operatorname{int}\operatorname{dom} g^* = \operatorname{int}(\operatorname{dom} f \times \operatorname{dom} g^*) = \operatorname{int}\operatorname{dom}\boldsymbol{f}. \tag{2.12}$$

Likewise, since

$$\boldsymbol{f}^*\colon \mathcal{X}^* \to \left]-\infty,+\infty\right] : (x^*, y) \mapsto f^*(x^*) + g(y), \tag{2.13}$$

we deduce that $\partial \boldsymbol{f}^*$ is single-valued on $\operatorname{dom} \partial \boldsymbol{f}^* = \operatorname{int} \operatorname{dom} \boldsymbol{f}^*$. Consequently, [6, Theorems 5.4 and 5.6] assert that $\boldsymbol{f}$ is a Legendre function.

(ii): It follows from Proposition 2.1(i)(a) that $\boldsymbol{Z}$ is a closed convex subset of $\boldsymbol{\mathcal{X}}$. Hence, since $\boldsymbol{Z} \cap \operatorname{int} \operatorname{dom} \boldsymbol{f} \neq \varnothing$, we derive from (i) and [6, Corollary 7.9] that $(\overline{x}, \overline{y}^*) = P^{\boldsymbol{f}}_{\boldsymbol{Z}}(x_0, y_0^*)$ is uniquely defined. ☐

2.2 Best Bregman approximation algorithm

The approach we present goes back to Haugazeau's algorithm [23, Théorème 3-2] (see also [9, Theorem 29.3]) for projecting a point onto the intersection of closed convex sets in a Hilbert space using the projections onto the individual sets. The method was extended in [18] to minimize certain convex functions over the intersection of closed convex sets in Banach spaces. The adaptation to the problem of finding the best Bregman approximation from a closed convex set was investigated in [8].

Definition 2.3 [7, Definition 3.1] and [8, Section 3] Let $\boldsymbol{\mathcal{X}}$ be a reflexive real Banach space, let $\boldsymbol{f} \in \Gamma_0(\boldsymbol{\mathcal{X}})$ be a Legendre function, let $\boldsymbol{x}_0 \in \operatorname{int} \operatorname{dom} \boldsymbol{f}$, let $\boldsymbol{x} \in \operatorname{int} \operatorname{dom} \boldsymbol{f}$, and let $\boldsymbol{y} \in \operatorname{int} \operatorname{dom} \boldsymbol{f}$. Then

$$\begin{aligned} H^{\boldsymbol{f}}(\boldsymbol{x}, \boldsymbol{y}) &= \big\{ \boldsymbol{z} \in \boldsymbol{\mathcal{X}} \bigm| \langle \boldsymbol{z} - \boldsymbol{y}, \nabla \boldsymbol{f}(\boldsymbol{x}) - \nabla \boldsymbol{f}(\boldsymbol{y}) \rangle \leqslant 0 \big\} \\ &= \big\{ \boldsymbol{z} \in \boldsymbol{\mathcal{X}} \bigm| D^{\boldsymbol{f}}(\boldsymbol{z}, \boldsymbol{y}) + D^{\boldsymbol{f}}(\boldsymbol{y}, \boldsymbol{x}) \leqslant D^{\boldsymbol{f}}(\boldsymbol{z}, \boldsymbol{x}) \big\} \end{aligned} \tag{2.14}$$

is the closed affine half-space onto which $\boldsymbol{y}$ is the Bregman projection of $\boldsymbol{x}$ if $\boldsymbol{x} \neq \boldsymbol{y}$. Moreover, if $H^{\boldsymbol{f}}(\boldsymbol{x}_0, \boldsymbol{x}) \cap H^{\boldsymbol{f}}(\boldsymbol{x}, \boldsymbol{y}) \cap \operatorname{int} \operatorname{dom} \boldsymbol{f} \neq \varnothing$, then

$$Q^{\boldsymbol{f}}(\boldsymbol{x}_0, \boldsymbol{x}, \boldsymbol{y}) = P^{\boldsymbol{f}}_{H^{\boldsymbol{f}}(\boldsymbol{x}_0, \boldsymbol{x}) \cap H^{\boldsymbol{f}}(\boldsymbol{x}, \boldsymbol{y})} \boldsymbol{x}_0. \tag{2.15}$$

Lemma 2.4 [7, Lemma 3.2] *Let $\boldsymbol{\mathcal{X}}$ be a reflexive real Banach space, and let $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ be convex subsets of $\boldsymbol{\mathcal{X}}$ such that $\boldsymbol{C}_1$ is closed and $\boldsymbol{C}_1 \cap \operatorname{int} \boldsymbol{C}_2 \neq \varnothing$. Then $\overline{\boldsymbol{C}_1 \cap \operatorname{int} \boldsymbol{C}_2} = \boldsymbol{C}_1 \cap \overline{\boldsymbol{C}}_2$.*

Proposition 2.5 *Let $\boldsymbol{\mathcal{X}}$ be a reflexive real Banach space, let $\boldsymbol{f} \in \Gamma_0(\boldsymbol{\mathcal{X}})$ be a Legendre function, let $\boldsymbol{C}$ be a closed convex subset of $\overline{\operatorname{dom} \boldsymbol{f}}$ such that $\boldsymbol{C} \cap \operatorname{int} \operatorname{dom} \boldsymbol{f} \neq \varnothing$, let $\boldsymbol{x}_0 \in \operatorname{int} \operatorname{dom} \boldsymbol{f}$, and set $\overline{\boldsymbol{x}} = P^{\boldsymbol{f}}_{\boldsymbol{C}} \boldsymbol{x}_0$. At every iteration $n \in \mathbb{N}$, find $\boldsymbol{x}_{n+1/2} \in \operatorname{int} \operatorname{dom} \boldsymbol{f}$ such that $\boldsymbol{C} \subset H^{\boldsymbol{f}}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1/2})$ and set*

$$\boldsymbol{x}_{n+1} = Q^{\boldsymbol{f}}\big(\boldsymbol{x}_0, \boldsymbol{x}_n, \boldsymbol{x}_{n+1/2}\big). \tag{2.16}$$

Then the following hold:

(i) $(\forall n \in \mathbb{N})$ $\boldsymbol{C} \subset H^{\boldsymbol{f}}(\boldsymbol{x}_0, \boldsymbol{x}_n)$.

(ii) $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ *is a well-defined bounded sequence in* $\operatorname{int}\operatorname{dom}\boldsymbol{f}$.

(iii) *Suppose that, for some* $n\in\mathbb{N}$, $\boldsymbol{x}_n\in\boldsymbol{C}$. *Then* $(\forall k\in\mathbb{N})\ \boldsymbol{x}_{n+k}=\overline{\boldsymbol{x}}$.

(iv) $\sum_{n\in\mathbb{N}} D^{\boldsymbol{f}}(\boldsymbol{x}_{n+1},\boldsymbol{x}_n)<+\infty$.

(v) $(\forall n\in\mathbb{N})\ D^{\boldsymbol{f}}(\boldsymbol{x}_{n+1/2},\boldsymbol{x}_n)\leqslant D^{\boldsymbol{f}}(\boldsymbol{x}_{n+1},\boldsymbol{x}_n)$.

(vi) $\sum_{n\in\mathbb{N}} D^{\boldsymbol{f}}(\boldsymbol{x}_{n+1/2},\boldsymbol{x}_n)<+\infty$.

(vii) $[\,\boldsymbol{x}_n\rightharpoonup\overline{\boldsymbol{x}}$ *and* $\boldsymbol{f}(\boldsymbol{x}_n)\to\boldsymbol{f}(\overline{\boldsymbol{x}})\,]\Leftrightarrow D^{\boldsymbol{f}}(\boldsymbol{x}_n,\overline{\boldsymbol{x}})\to 0\Leftrightarrow\mathfrak{W}(\boldsymbol{x}_n)_{n\in\mathbb{N}}\subset\boldsymbol{C}$.

Proof. Item (i) is found in [8, Proof of Proposition 3.3]. The first equivalence in (vii) follows from [8, Propositions 2.2(ii)]. To establish the remaining assertions, set

$$(\forall n\in\mathbb{N})\quad T_n=P^{\boldsymbol{f}}_{H^{\boldsymbol{f}}(\boldsymbol{x}_n,\boldsymbol{x}_{n+1/2})}. \tag{2.17}$$

Then, for every $n\in\mathbb{N}$, Definition 2.3 yields $T_n\boldsymbol{x}_n=\boldsymbol{x}_{n+1/2}$, [7, Proposition 3.32(ii)(b)] yields $\operatorname{Fix}T_n=H^{\boldsymbol{f}}(\boldsymbol{x}_n,\boldsymbol{x}_{n+1/2})\cap\operatorname{int}\operatorname{dom}\boldsymbol{f}$, and we derive from Lemma 2.4 that

$$\boldsymbol{C}\subset H^{\boldsymbol{f}}(\boldsymbol{x}_n,\boldsymbol{x}_{n+1/2})\cap\overline{\operatorname{dom}}\,\boldsymbol{f}=\overline{H^{\boldsymbol{f}}(\boldsymbol{x}_n,\boldsymbol{x}_{n+1/2})\cap\operatorname{int}\operatorname{dom}\boldsymbol{f}}=\overline{\operatorname{Fix}}\,T_n. \tag{2.18}$$

On the other hand,

$$(\forall n\in\mathbb{N})\quad \varnothing\neq\boldsymbol{C}\cap\operatorname{int}\operatorname{dom}\boldsymbol{f}\subset H^{\boldsymbol{f}}(\boldsymbol{x}_n,\boldsymbol{x}_{n+1/2})\cap\operatorname{int}\operatorname{dom}\boldsymbol{f}=\operatorname{Fix}T_n \tag{2.19}$$

and, therefore, $\bigcap_{n\in\mathbb{N}}\operatorname{Fix}T_n\neq\varnothing$. Altogether, [8, Condition 3.2] is satisfied and it follows from [8, Propositions 3.3 and 3.4] and [8, Proof of Proposition 3.4(vii)] that the proof is complete. □

2.3 Coercivity and boundedness of monotone operators

Definition 2.6 Let $\mathcal{X}$ be a reflexive real Banach space such that $\mathcal{X}\neq\{0\}$ and let $M\colon\mathcal{X}\to 2^{\mathcal{X}^*}$. Then M is *coercive* if

$$(\exists\, z\in\operatorname{dom}M)\qquad \lim_{\|x\|\to+\infty}\ \inf\frac{\langle x-z,Mx\rangle}{\|x\|}=+\infty, \tag{2.20}$$

and it is *bounded* if it maps bounded sets to bounded sets.

Lemma 2.7 *Let $\mathcal{X}$ be a reflexive real Banach space such that $\mathcal{X}\neq\{0\}$ and let $M\colon\mathcal{X}\to 2^{\mathcal{X}^*}$. Suppose that one of the following holds:*

(i) $\operatorname{dom}M$ *is nonempty and bounded.*

(ii) M *is uniformly monotone at some point* $z\in\operatorname{dom}M$ *with a supercoercive modulus: there exists a strictly increasing function* $\phi\colon[0,+\infty[\,\to[0,+\infty]$ *that vanishes only at* 0 *such that* $\lim_{t\to+\infty}\phi(t)/t=+\infty$ *and*

$$(\forall(x,x^*)\in\operatorname{gra}M)(\forall z^*\in Mz)\quad\langle x-z,x^*-z^*\rangle\geqslant\phi(\|x-z\|). \tag{2.21}$$

(iii) $M = \partial\varphi$, *where* φ *is a supercoercive function in* $\Gamma_0(\mathcal{X})$.

Then M *is coercive.*

Proof. (i): Let $x \in \mathcal{X}$ and let $z \in \operatorname{dom} M$. Then, if $\|x\|$ is sufficiently large, we have $Mx = \varnothing$ and therefore $\inf \langle x - z, Mx\rangle / \|x\| = +\infty$.

(ii): We have

$$(\forall (x, x^*) \in \operatorname{gra} M)(\forall z^* \in Mz) \quad \langle x - z, x^* - z^*\rangle \geqslant \phi(\|x - z\|). \tag{2.22}$$

Hence, for every $x \in \operatorname{dom} M$ such that $\|x\| > \|z\|$, we have

$$\begin{aligned}(\forall x^* \in Mx)(\forall z^* \in Mz) \quad \frac{\langle x - z, x^*\rangle}{\|x\|} &\geqslant \frac{\phi(\|x - z\|) - \|x - z\|\,\|z^*\|}{\|x\|} \\ &= \frac{\|x - z\|}{\|x\|}\left(\frac{\phi(\|x - z\|)}{\|x - z\|} - \|z^*\|\right).\end{aligned} \tag{2.23}$$

Thus,

$$\lim_{\|x\| \to +\infty} \inf \frac{\langle x - z, Mx\rangle}{\|x\|} = +\infty. \tag{2.24}$$

(iii): In view of (i), we suppose that $\operatorname{dom} M$ is unbounded. Let $z \in \operatorname{dom} M$. Then we derive from (1.3) that, for every $x \in \operatorname{dom} M \smallsetminus \{0\}$,

$$(\forall x^* \in Mx) \quad \frac{\varphi(x) - \varphi(z)}{\|x\|} \leqslant \frac{\langle x - z, x^*\rangle}{\|x\|}. \tag{2.25}$$

Hence, the supercoercivity of φ yields

$$\lim_{\|x\| \to +\infty} \inf \frac{\langle x - z, Mx\rangle}{\|x\|} = +\infty \tag{2.26}$$

and M is therefore coercive. ☐

Lemma 2.8 *Let* $\mathcal{X}$ *be a reflexive real Banach space such that* $\mathcal{X} \neq \{0\}$, *let* $M_1 \colon \mathcal{X} \to 2^{\mathcal{X}^*}$, *and let* $M_2 \colon \mathcal{X} \to 2^{\mathcal{X}^*}$ *be monotone. Suppose that there exists* $z \in \operatorname{dom} M_1 \cap \operatorname{dom} M_2$ *such that*

$$\lim_{\|x\| \to +\infty} \inf \frac{\langle x - z, M_1 x\rangle}{\|x\|} = +\infty. \tag{2.27}$$

Then $M_1 + M_2$ *is coercive.*

Proof. Suppose that $x \in (\operatorname{dom} M_1 \cap \operatorname{dom} M_2) \smallsetminus \{0\}$, let $x^* \in (M_1 + M_2)x$, and let $z^* \in (M_1 + M_2)z$. Then there exist $x_1^* \in M_1 x$, $x_2^* \in M_2 x$, $z_1^* \in M_1 z$, and $z_2^* \in M_2 z$ such that $x^* = x_1^* + x_2^*$ and

$z^* = z_1^* + z_2^*$. In turn, the monotonicity of M_2 yields

$$\begin{aligned}\frac{\langle x-z, x^*\rangle}{\|x\|} &= \frac{\langle x-z, x_1^*-z_1^*\rangle}{\|x\|} + \frac{\langle x-z, x_2^*-z_2^*\rangle}{\|x\|} + \frac{\langle x-z, z^*\rangle}{\|x\|}\\ &\geqslant \frac{\langle x-z, x_1^*\rangle}{\|x\|} + \frac{\langle x-z, z_2^*\rangle}{\|x\|}\\ &\geqslant \frac{\langle x-z, x_1^*\rangle - \|x-z\|\,\|z_2^*\|}{\|x\|}\end{aligned} \tag{2.28}$$

and (2.27) implies that $M_1 + M_2$ is coercive. □

Lemma 2.9 *Let $\mathcal{X}$ be a reflexive real Banach space such that $\mathcal{X} \neq \{0\}$, let $M_1\colon \mathcal{X} \to 2^{\mathcal{X}^*}$, let $M_2\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be monotone, let $(x_n^*)_{n\in\mathbb{N}}$ be a bounded sequence in $\mathcal{X}^*$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a bounded sequence in $]0,+\infty[$. Suppose that there exists $z \in \operatorname{dom} M_1 \cap \operatorname{dom} M_2$ such that*

$$\lim_{\|x\|\to+\infty} \inf \frac{\langle x-z, M_1x\rangle}{\|x\|} = +\infty, \tag{2.29}$$

and that

$$(\forall n \in \mathbb{N}) \quad x_n \in (M_1 + \gamma_n M_2)^{-1} x_n^*. \tag{2.30}$$

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

Proof. Set $\beta = \sup_{n\in\mathbb{N}} \|x_n^*\|$ and $\sigma = \sup_{n\in\mathbb{N}} \gamma_n$. It follows from (2.30) that, for every $n \in \mathbb{N}$, there exist $a_n^* \in M_1x_n$ and $b_n^* \in M_2x_n$ such that $x_n^* = a_n^* + \gamma_n b_n^*$. If $(x_n)_{n\in\mathbb{N}}$ is unbounded, there exists a strictly increasing sequence $(k_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that $0 < \|x_{k_n}\| \uparrow +\infty$. Therefore, (2.29) yields

$$\lim_{n\to+\infty} \frac{\langle x_{k_n} - z, a_{k_n}^*\rangle}{\|x_{k_n}\|} = +\infty. \tag{2.31}$$

Now let $z^* \in M_2 z$. By monotonicity of M_2, $(\forall n \in \mathbb{N})$ $\langle x_{k_n} - z, b_{k_n}^* - z^*\rangle \geqslant 0$. Hence, (2.31) implies that

$$\begin{aligned}\beta\left(1 + \frac{\|z\|}{\|x_{k_0}\|}\right) &\geqslant \beta\left(1 + \frac{\|z\|}{\|x_{k_n}\|}\right)\\ &\geqslant \beta\frac{\|x_{k_n} - z\|}{\|x_{k_n}\|}\\ &\geqslant \frac{\langle x_{k_n} - z, x_{k_n}^*\rangle}{\|x_{k_n}\|}\\ &= \frac{\langle x_{k_n} - z, a_{k_n}^*\rangle}{\|x_{k_n}\|} + \gamma_{k_n}\frac{\langle x_{k_n} - z, b_{k_n}^* - z^*\rangle}{\|x_{k_n}\|} + \gamma_{k_n}\frac{\langle x_{k_n} - z, z^*\rangle}{\|x_{k_n}\|}\\ &\geqslant \frac{\langle x_{k_n} - z, a_{k_n}^*\rangle}{\|x_{k_n}\|} - \sigma\frac{\|x_{k_n} - z\|\,\|z^*\|}{\|x_{k_n}\|}\\ &\geqslant \frac{\langle x_{k_n} - z, a_{k_n}^*\rangle}{\|x_{k_n}\|} - \sigma\|z^*\|\left(1 + \frac{\|z\|}{\|x_{k_0}\|}\right)\\ &\to +\infty,\end{aligned} \tag{2.32}$$

and we reach a contradiction. □

Corollary 2.10 *Let $\mathcal{X}$ be a reflexive real Banach space such that $\mathcal{X} \neq \{0\}$ and let $M \colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be coercive. Then M^{-1} is bounded.*

Proof. Take $M_1 = M$ and $M_2 = 0$ in Lemma 2.9. ☐

Proposition 2.11 *Let $\mathcal{X}$ be a reflexive real Banach space such that $\mathcal{X} \neq \{0\}$, let $h \in \Gamma_0(\mathcal{X})$ be essentially smooth, and let $M \colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be such that $\operatorname{dom} M \cap \operatorname{int} \operatorname{dom} h \neq \varnothing$. Suppose that one of the following holds:*

(i) $\operatorname{dom} M \cap \operatorname{int} \operatorname{dom} h$ *is bounded.*

(ii) *There exists $z \in \operatorname{dom} M \cap \operatorname{int} \operatorname{dom} h$ such that*

$$\lim_{\|x\| \to +\infty} \inf \frac{\langle x - z, Mx \rangle}{\|x\|} = +\infty. \tag{2.33}$$

(iii) *M is uniformly monotone at a point $z \in \operatorname{dom} M \cap \operatorname{int} \operatorname{dom} h$ with a supercoercive modulus.*

(iv) *M is monotone and h is supercoercive.*

(v) *M is monotone and h is uniformly convex at a point $z \in \operatorname{dom} M \cap \operatorname{int} \operatorname{dom} h$, i.e., there exists an increasing function $\phi \colon \left[0, +\infty\right[\to [0, +\infty]$ that vanishes only at 0 such that*

$$(\forall y \in \operatorname{dom} h)(\forall \alpha \in \left]0,1\right[) \;\; h(\alpha y + (1-\alpha)z) + \alpha(1-\alpha)\phi(\|y - z\|) \leqslant \alpha h(y) + (1-\alpha)h(z). \tag{2.34}$$

Then $\nabla h + M$ is coercive. If, in addition, M is maximally monotone, then $\operatorname{dom} (\nabla h + M)^{-1} = \mathcal{X}^*$.

Proof. We first observe that [35, Theorem 18.7] and [6, Theorem 5.6] imply that ∇h is maximally monotone and that $\operatorname{dom} \nabla h = \operatorname{int} \operatorname{dom} h$.

(i): Lemma 2.7(i).

(ii): It follows from Lemma 2.8 that $\nabla h + M$ is coercive.

(iii): Since $\nabla h + M$ is uniformly monotone at z with a supercoercive modulus, the claim follows from Lemma 2.7(ii).

(iv): Let $z \in \operatorname{dom} M \cap \operatorname{int} \operatorname{dom} h = \operatorname{dom} M \cap \operatorname{dom} \partial h$. Then we derive from (2.26) that

$$\lim_{\|x\| \to +\infty} \frac{\langle x - z, \nabla h(x) \rangle}{\|x\|} = +\infty. \tag{2.35}$$

Thus, ∇h satisfies (2.27) and it follows from Lemma 2.8 that $\nabla h + M$ is coercive.

(v): It follows from [39, Definition 2.2 and Remark 2.8] that ∇h is uniformly monotone at z with a supercoercive modulus. Hence, $\nabla h + M$ is likewise and Lemma 2.7(ii) implies that $\nabla h + M$ is coercive. Alternatively, this is a special case of (iv).

Finally, suppose that M is maximally monotone. Then [35, Theorem 24.1(a)] asserts that $\nabla h + M$ is maximally monotone. Consequently, since $\nabla h+M$ is coercive, it follows from [42, Corollary II-B.32.35] that $\operatorname{dom}(\nabla h + M)^{-1} = \operatorname{ran}(\nabla h + M) = \mathcal{X}^*$. □

Lemma 2.12 *Let $\mathcal{X}$ and $\mathcal{Y}$ be real Banach spaces, let $D \subset \mathcal{X}$ be a nonempty open set, and let C be a nonempty bounded convex subset of D. Suppose that $T\colon D \to \mathcal{Y}$ is uniformly continuous on C in the sense that*

$$(\forall \varepsilon \in \left]0,+\infty\right[)(\exists\, \delta \in \left]0,+\infty\right[)(\forall x \in C)(\forall y \in C) \quad \|x-y\| \leqslant \delta \;\Rightarrow\; \|Tx-Ty\| \leqslant \varepsilon. \tag{2.36}$$

Then T is bounded on C.

Proof. In view of (2.36), there exists $\delta \in \left]0,+\infty\right[$ such that

$$(\forall x \in C)(\forall y \in C) \quad \|x-y\| \leqslant \delta \quad\Rightarrow\quad \|Tx-Ty\| \leqslant 1. \tag{2.37}$$

Now fix $z \in C$ and $\rho \in \left]0,+\infty\right[$ such that $C \subset \big\{x \in \mathcal{X} \bigm| \|x-z\| \leqslant \rho\big\}$, and take an integer $m \geqslant 1+\rho/\delta$. Let $x \in C$ and set

$$(\forall n \in \{0,\ldots,m\}) \quad x_n = x + \frac{n}{m}(z-x) \in C. \tag{2.38}$$

Then, for every $n \in \{0,\ldots,m-1\}$, $\|x_{n+1}-x_n\| = \|z-x\|/m \leqslant \rho/m \leqslant \delta$ and (2.37) yields $\|Tx_{n+1}-Tx_n\| \leqslant 1$. Hence, $\|Tz-Tx\| \leqslant \sum_{n=0}^{m-1}\|Tx_{n+1}-Tx_n\| \leqslant m$ and therefore $\|Tx\| < \|Tz\|+m$. We conclude that $\sup_{x\in C}\|Tx\| \leqslant \|Tz\|+m$. □

3 Best Bregman approximation algorithm

Proposition 2.1(i)(a) asserts that Problem 1.2 reduces to finding the Bregman projection of a reference point (x_0,y_0^*) onto the closed convex subset $\boldsymbol{C} = \boldsymbol{Z} \cap \overline{\operatorname{dom}\boldsymbol{f}}$ of $\overline{\operatorname{dom}\boldsymbol{f}}$. Our strategy is to employ Proposition 2.5 for this task. The following condition will be used subsequently (see [7, Examples 4.10, 5.11, and 5.13] for special cases).

Condition 3.1 [8, Condition 4.3(ii)] Let $\mathcal{X}$ be a reflexive real Banach space and let $h \in \Gamma_0(\mathcal{X})$ be Gâteaux differentiable on $\operatorname{int}\operatorname{dom} h \neq \varnothing$. For every sequence $(x_n)_{n\in\mathbb{N}}$ in $\operatorname{int}\operatorname{dom} h$ and every bounded sequence $(y_n)_{n\in\mathbb{N}}$ in $\operatorname{int}\operatorname{dom} h$,

$$D^h(x_n,y_n) \to 0 \quad\Rightarrow\quad x_n - y_n \to 0. \tag{3.1}$$

We now derive from Proposition 2.5 our best Bregman approximation algorithm to solve Problem 1.2.

Theorem 3.2 *Consider the setting of Problem 1.2. Let $h \in \Gamma_0(\mathcal{X})$ and $j \in \Gamma_0(\mathcal{Y})$ be Legendre functions such that $\operatorname{int}\operatorname{dom} f \subset \operatorname{int}\operatorname{dom} h$, $L(\operatorname{int}\operatorname{dom} f) \subset \operatorname{int}\operatorname{dom} j$, and there exist ε and δ in $\left]0,+\infty\right[$ such* *that $\nabla h+\varepsilon A$ and $\nabla j+\delta B$ are coercive. Let $\sigma\in[\max\{\varepsilon,\delta\},+\infty[$ and iterate*

$$
\begin{array}{l}
\text{for } n=0,1,\ldots\\
\left\lfloor
\begin{array}{l}
(\gamma_n,\mu_n)\in[\varepsilon,\sigma]\times[\delta,\sigma]\\
a_n=(\nabla h+\gamma_n A)^{-1}\big(\nabla h(x_n)-\gamma_n L^*y_n^*\big)\\
a_n^*=\gamma_n^{-1}\big(\nabla h(x_n)-\nabla h(a_n)\big)-L^*y_n^*\\
b_n=(\nabla j+\mu_n B)^{-1}\big(\nabla j(Lx_n)+\mu_n y_n^*\big)\\
b_n^*=\mu_n^{-1}\big(\nabla j(Lx_n)-\nabla j(b_n)\big)+y_n^*\\
\boldsymbol{H}_n=\big\{(x,y^*)\in\boldsymbol{\mathcal{X}}\mid\langle x,a_n^*+L^*b_n^*\rangle+\langle b_n-La_n,y^*\rangle\leqslant\langle a_n,a_n^*\rangle+\langle b_n,b_n^*\rangle\big\}\\
(x_{n+1/2},y_{n+1/2}^*)=P_{\boldsymbol{H}_n}^{\boldsymbol{f}}(x_n,y_n^*)\\
(x_{n+1},y_{n+1}^*)=Q^{\boldsymbol{f}}\big((x_0,y_0^*),(x_n,y_n^*),(x_{n+1/2},y_{n+1/2}^*)\big).
\end{array}
\right.
\end{array}
\tag{3.2}
$$

Then the following hold:

(i) *Let $n\in\mathbb{N}$. Then the following are equivalent:*

(a) $(x_n,y_n^*)=(\overline{x},\overline{y}^*)$.

(b) $(x_n,y_n^*)\in\boldsymbol{Z}$.

(c) $(x_n,y_n^*)\in\boldsymbol{H}_n$.

(d) $x_n=a_n$ *and* $y_n^*=b_n^*$.

(e) $La_n=b_n$ *and* $a_n^*=-L^*b_n^*$.

(f) $\boldsymbol{H}_n=\boldsymbol{\mathcal{X}}$.

(g) $(x_{n+1/2},y_{n+1/2}^*)=(x_n,y_n^*)$.

(h) $(x_{n+1},y_{n+1}^*)=(x_n,y_n^*)$.

(ii) $\sum_{n\in\mathbb{N}}D^f(x_{n+1},x_n)<+\infty$ *and* $\sum_{n\in\mathbb{N}}D^{g^*}(y_{n+1}^*,y_n^*)<+\infty$.

(iii) $\sum_{n\in\mathbb{N}}D^f(x_{n+1/2},x_n)<+\infty$ *and* $\sum_{n\in\mathbb{N}}D^{g^*}(y_{n+1/2}^*,y_n^*)<+\infty$.

(iv) *Suppose that f, g^*, h, and j satisfy Condition 3.1, and that ∇h and ∇j are uniformly continuous on every bounded subset of* $\operatorname{int}\operatorname{dom}h$ *and* $\operatorname{int}\operatorname{dom}j$*, respectively. Then $x_n\to\overline{x}$ and $y_n^*\to\overline{y}^*$.*

Proof. We apply Proposition 2.5 to

$$\boldsymbol{C}=\boldsymbol{Z}\cap\overline{\operatorname{dom}\boldsymbol{f}}.\tag{3.3}$$

It follows from Proposition 2.1(i)(a) and our assumptions that $\boldsymbol{C}$ is a closed convex subset of $\overline{\operatorname{dom}\boldsymbol{f}}$ and that $\boldsymbol{C}\cap\operatorname{int}\operatorname{dom}\boldsymbol{f}\neq\varnothing$. Moreover, Proposition 2.2(i) asserts that $\boldsymbol{f}$ is a Legendre function. Now let $\gamma\in[\varepsilon,+\infty[$ and let $\mu\in[\delta,+\infty[$. Since h is strictly convex, ∇h is strictly monotone [40, Theorem 2.4.4(ii)] and $\nabla h+\gamma A$ is likewise. Let (x^*,x_1) and (x^*,x_2) be two elements in $\operatorname{gra}(\nabla h+\gamma A)^{-1}$ such that $x_1\neq x_2$. Then (x_1,x^*) and (x_2,x^*) lie in $\operatorname{gra}(\nabla h+\gamma A)$ and the strict monotonicity of $\nabla h+\gamma A$ implies that

$$0=\langle x_1-x_2,x^*-x^*\rangle>0,\tag{3.4}$$

which is impossible. Thus,

$$(\nabla h + \gamma A)^{-1} \text{ is at most single-valued.} \tag{3.5}$$

The same argument shows that

$$(\nabla j + \mu B)^{-1} \text{ is at most single-valued.} \tag{3.6}$$

On the other hand, by assumption, there exists $(x, y^*) \in \boldsymbol{Z} \cap \operatorname{int}\operatorname{dom} \boldsymbol{f}$. It follows from (1.9) that $x \in \operatorname{dom} A$ and $Lx \in \operatorname{dom} B$. Furthermore, (2.12) yields $x \in \operatorname{int}\operatorname{dom} f$. Therefore

$$\begin{cases} x \in \operatorname{dom} A \cap \operatorname{int}\operatorname{dom} f \subset \operatorname{dom} A \cap \operatorname{int}\operatorname{dom} h \\ Lx \in \operatorname{dom} B \cap L(\operatorname{int}\operatorname{dom} f) \subset \operatorname{dom} B \cap \operatorname{int}\operatorname{dom} j. \end{cases} \tag{3.7}$$

Thus, $\operatorname{dom} A \cap \operatorname{int}\operatorname{dom} h \neq \varnothing$ and $\operatorname{dom} B \cap \operatorname{int}\operatorname{dom} j \neq \varnothing$. It therefore follows from Lemma 2.8 that

$$\nabla h + \gamma A = (\nabla h + \varepsilon A) + (\gamma - \varepsilon)A \quad \text{and} \quad \nabla j + \mu B = (\nabla j + \delta B) + (\mu - \delta)B \ \text{ are coercive.} \tag{3.8}$$

Altogether, (3.5), (3.6), (3.8), and Proposition 2.11 assert that the operators

$$\begin{cases} (\nabla h + \gamma A)^{-1} \colon \mathcal{X}^* \to \operatorname{dom} A \cap \operatorname{int}\operatorname{dom} h \\ (\nabla j + \mu B)^{-1} \colon \mathcal{Y}^* \to \operatorname{dom} B \cap \operatorname{int}\operatorname{dom} j \end{cases} \tag{3.9}$$

are well defined and single-valued. Now set

$$(\forall n \in \mathbb{N}) \quad \boldsymbol{x}_n = (x_n, y_n^*) \quad \text{and} \quad \boldsymbol{x}_{n+1/2} = (x_{n+1/2}, y_{n+1/2}^*). \tag{3.10}$$

Since (3.2) yields

$$(\forall n \in \mathbb{N}) \quad (a_n, a_n^*) \in \operatorname{gra} A \quad \text{and} \quad (b_n, b_n^*) \in \operatorname{gra} B, \tag{3.11}$$

it follows from (3.3), Proposition 2.1(ii)(b), (3.2), and Definition 2.3 that

$$(\forall n \in \mathbb{N}) \quad \varnothing \neq \boldsymbol{C} \subset \boldsymbol{Z} \subset \boldsymbol{H}_n = H^{\boldsymbol{f}}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1/2}). \tag{3.12}$$

Hence, appealing to Proposition 2.2(i) and (1.5), we see that

$$(\forall n \in \mathbb{N}) \quad P^{\boldsymbol{f}}_{\boldsymbol{H}_n} \colon \operatorname{int}\operatorname{dom} \boldsymbol{f} \to \boldsymbol{H}_n \cap \operatorname{int}\operatorname{dom} \boldsymbol{f} \tag{3.13}$$

and that

$$(\forall n \in \mathbb{N}) \quad \boldsymbol{x}_{n+1} = Q^{\boldsymbol{f}}\big(\boldsymbol{x}_0, \boldsymbol{x}_n, \boldsymbol{x}_{n+1/2}\big). \tag{3.14}$$

Thus, we derive from (3.10), (3.12), and Proposition 2.5(ii) that $(x_n)_{n\in\mathbb{N}}$ and $(y_n^*)_{n\in\mathbb{N}}$ are well-defined sequences in $\operatorname{int}\operatorname{dom} f$ and $\operatorname{int}\operatorname{dom} g^*$, respectively.

(i): We prove the following implications.

(i)(a)⇒(i)(b): Clear.

(i)(b)⇒(i)(a): Proposition 2.5(iii).

(i)(b)⇒(i)(c): Clear by (3.12).

(i)(c)⇒(i)(d): In view of (3.2),

$$\begin{aligned} 0 &\geqslant \langle x_n, a_n^* + L^* b_n^* \rangle + \langle b_n - La_n, y_n^* \rangle - \langle a_n, a_n^* \rangle - \langle b_n, b_n^* \rangle \\ &= \langle x_n - a_n, a_n^* + L^* y_n^* \rangle + \langle Lx_n - b_n, b_n^* - y_n^* \rangle \\ &= \gamma_n^{-1} \langle x_n - a_n, \nabla h(x_n) - \nabla h(a_n) \rangle + \mu_n^{-1} \langle Lx_n - b_n, \nabla j(Lx_n) - \nabla j(b_n) \rangle. \end{aligned} \tag{3.15}$$

Consequently, the strict monotonicity of ∇h and ∇j yields

$$x_n = a_n \quad \text{and} \quad Lx_n = b_n. \tag{3.16}$$

Furthermore,

$$b_n^* = \mu_n^{-1}\big(\nabla j(Lx_n) - \nabla j(b_n)\big) + y_n^* = \mu_n^{-1}\big(\nabla j(b_n) - \nabla j(b_n)\big) + y_n^* = y_n^*. \tag{3.17}$$

(i)(d)⇒(i)(e): We derive from (3.2) that $a_n^* = \gamma_n^{-1}(\nabla h(x_n) - \nabla h(a_n)) - L^* y_n^* = -L^* y_n^* = -L^* b_n^*$. On the other hand, since

$$\begin{aligned} \langle La_n - b_n, \nabla j(La_n) - \nabla j(b_n) \rangle &= \langle Lx_n - b_n, \nabla j(Lx_n) - \nabla j(b_n) \rangle \\ &= \mu_n \langle Lx_n - b_n, b_n^* - y_n^* \rangle \\ &= 0, \end{aligned} \tag{3.18}$$

the strict monotonicity of ∇j yields $La_n = b_n$.

(i)(e)⇔(i)(f): Proposition 2.1(ii)(a).

(i)(f)⇒(i)(g): Indeed, $\boldsymbol{x}_{n+1/2} = P^{\boldsymbol{f}}_{\boldsymbol{H}_n} \boldsymbol{x}_n = \boldsymbol{x}_n$.

(i)(g)⇒(i)(h): We have

$$\boldsymbol{x}_{n+1} = Q^{\boldsymbol{f}}(\boldsymbol{x}_0, \boldsymbol{x}_n, \boldsymbol{x}_{n+1/2}) = P^{\boldsymbol{f}}_{H^{\boldsymbol{f}}(\boldsymbol{x}_0, \boldsymbol{x}_n) \cap H^{\boldsymbol{f}}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1/2})} \boldsymbol{x}_0 = P^{\boldsymbol{f}}_{H^{\boldsymbol{f}}(\boldsymbol{x}_0, \boldsymbol{x}_n)} \boldsymbol{x}_0 = \boldsymbol{x}_n. \tag{3.19}$$

(i)(h)⇒(i)(g): By Proposition 2.5(v), $0 \leqslant D^{\boldsymbol{f}}(\boldsymbol{x}_{n+1/2}, \boldsymbol{x}_n) \leqslant D^{\boldsymbol{f}}(\boldsymbol{x}_{n+1}, \boldsymbol{x}_n) = 0$. Therefore $D^{\boldsymbol{f}}(\boldsymbol{x}_{n+1/2}, \boldsymbol{x}_n) = 0$ and we derive from [6, Lemma 7.3(vi)] that $\boldsymbol{x}_{n+1/2} = \boldsymbol{x}_n$.

(i)(g)⇒(i)(c): Indeed, $\boldsymbol{x}_n = \boldsymbol{x}_{n+1/2} = P^{\boldsymbol{f}}_{\boldsymbol{H}_n} \boldsymbol{x}_n \in \boldsymbol{H}_n$.

(i)(d)⇒(i)(b): We derive from (3.2) that

$$\nabla h(x_n) - \gamma_n L^* y_n^* \in \nabla h(a_n) + \gamma_n A a_n = \nabla h(x_n) + \gamma_n A x_n. \tag{3.20}$$

Hence $-L^* y_n^* \in Ax_n$. Likewise, as in (3.18), we first obtain $Lx_n = b_n$ and then

$$\nabla j(Lx_n) + \mu_n y_n^* \in \nabla j(b_n) + \mu_n B b_n = \nabla j(Lx_n) + \mu_n B(Lx_n). \tag{3.21}$$

Thus, $y_n^* \in B(Lx_n)$, i.e., $Lx_n \in B^{-1}y_n^*$. In view of (1.9), the implication is proved.

(ii): Proposition 2.5(iv) yields

$$\sum_{n\in\mathbb{N}} D^f(x_{n+1},x_n) + \sum_{n\in\mathbb{N}} D^{g^*}(y_{n+1}^*,y_n^*) = \sum_{n\in\mathbb{N}} D^{\boldsymbol{f}}(\boldsymbol{x}_{n+1},\boldsymbol{x}_n) < +\infty. \tag{3.22}$$

(iii): Proposition 2.5(vi) yields

$$\sum_{n\in\mathbb{N}} D^f(x_{n+1/2},x_n) + \sum_{n\in\mathbb{N}} D^{g^*}(y_{n+1/2}^*,y_n^*) = \sum_{n\in\mathbb{N}} D^{\boldsymbol{f}}(\boldsymbol{x}_{n+1/2},\boldsymbol{x}_n) < +\infty. \tag{3.23}$$

(iv): Proposition 2.5(ii) implies that $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ is a bounded sequence in $\operatorname{int}\operatorname{dom}\boldsymbol{f}$. In turn,

$$(x_n)_{n\in\mathbb{N}} \in (\operatorname{int}\operatorname{dom} f)^{\mathbb{N}} \quad \text{and} \quad (y_n^*)_{n\in\mathbb{N}} \in (\operatorname{int}\operatorname{dom} g^*)^{\mathbb{N}} \quad \text{are bounded.} \tag{3.24}$$

On the other hand, by (3.2),

$$(\forall n\in\mathbb{N})\quad (x_{n+1/2},y_{n+1/2}^*) = \boldsymbol{x}_{n+1/2} = P_{\boldsymbol{H}_n}^{\boldsymbol{f}}\boldsymbol{x}_n \in \boldsymbol{H}_n \tag{3.25}$$

and

$$(\forall n\in\mathbb{N})\quad \langle x_{n+1/2}, a_n^*+L^*b_n^*\rangle + \langle b_n - La_n, y_{n+1/2}^*\rangle = \langle a_n,a_n^*\rangle + \langle b_n,b_n^*\rangle. \tag{3.26}$$

Therefore,

$$\begin{aligned}
(\forall n\in\mathbb{N})\quad & \|x_n - x_{n+1/2}\|\,\|a_n^*+L^*b_n^*\| + \|b_n - La_n\|\,\|y_n^* - y_{n+1/2}^*\| \\
&\geqslant \langle x_n - x_{n+1/2}, a_n^*+L^*b_n^*\rangle + \langle b_n - La_n, y_n^* - y_{n+1/2}^*\rangle \\
&= \langle x_n, a_n^*+L^*b_n^*\rangle + \langle b_n - La_n, y_n^*\rangle - \langle a_n,a_n^*\rangle - \langle b_n,b_n^*\rangle \\
&= \langle x_n - a_n, a_n^*+L^*y_n^*\rangle + \langle Lx_n - b_n, b_n^* - y_n^*\rangle \\
&= \gamma_n^{-1}\langle x_n - a_n, \nabla h(x_n) - \nabla h(a_n)\rangle + \mu_n^{-1}\langle Lx_n - b_n, \nabla j(Lx_n) - \nabla j(b_n)\rangle \\
&\geqslant \sigma^{-1}\big(D^h(x_n,a_n) + D^h(a_n,x_n) + D^j(Lx_n,b_n) + D^j(b_n,Lx_n)\big) \\
&\geqslant \sigma^{-1}\big(D^h(x_n,a_n) + D^j(Lx_n,b_n)\big).
\end{aligned} \tag{3.27}$$

However, since (iii) yields

$$D^f(x_{n+1/2},x_n)\to 0 \quad \text{and} \quad D^{g^*}(y_{n+1/2}^*,y_n^*)\to 0 \tag{3.28}$$

and since f and g^* satisfy Condition 3.1, (3.1) yields

$$x_{n+1/2} - x_n \to 0 \quad \text{and} \quad y_{n+1/2}^* - y_n^* \to 0. \tag{3.29}$$

Since ∇h is uniformly continuous on every bounded subset of $\operatorname{int}\operatorname{dom} h$, Lemma 2.12 asserts that ∇h is bounded on every bounded subset of $\operatorname{int}\operatorname{dom} h$ and hence, since $\operatorname{int}\operatorname{dom} f \subset \operatorname{int}\operatorname{dom} h$ and

L^* is bounded, it follows from (3.24) that $\big(\nabla h(x_n) - \gamma_n L^* y_n^*\big)_{n\in\mathbb{N}}$ is bounded. We therefore deduce from (3.9), (3.2), and Lemma 2.9 that

$$(a_n)_{n\in\mathbb{N}} \in (\operatorname{int}\operatorname{dom} h)^{\mathbb{N}} \quad \text{is bounded.} \tag{3.30}$$

Similarly, since ∇j is uniformly continuous on every bounded subset of $\operatorname{int}\operatorname{dom} j$ and $L(\operatorname{int}\operatorname{dom} f) \subset \operatorname{int}\operatorname{dom} j$, it follows from (3.24) and Lemma 2.12 that $\big(\nabla j(Lx_n) + \mu_n y_n^*\big)_{n\in\mathbb{N}}$ is bounded and hence (3.9), (3.2), and Lemma 2.9 yield

$$(b_n)_{n\in\mathbb{N}} \in (\operatorname{int}\operatorname{dom} j)^{\mathbb{N}} \quad \text{is bounded.} \tag{3.31}$$

Thus, $(\nabla h(x_n))_{n\in\mathbb{N}}$, $(\nabla h(a_n))_{n\in\mathbb{N}}$, $(\nabla j(Lx_n))_{n\in\mathbb{N}}$, and $(\nabla j(b_n))_{n\in\mathbb{N}}$ are bounded and we deduce from (3.2) that

$$(a_n^*)_{n\in\mathbb{N}} \quad \text{and} \quad (b_n^*)_{n\in\mathbb{N}} \quad \text{are bounded.} \tag{3.32}$$

We therefore derive from (3.27), (3.29), (3.30), and (3.31) that

$$D^h(x_n, a_n) \to 0 \quad \text{and} \quad D^j(Lx_n, b_n) \to 0. \tag{3.33}$$

Since h and j satisfy Condition 3.1, we get

$$x_n - a_n \to 0 \quad \text{and} \quad Lx_n - b_n \to 0. \tag{3.34}$$

Therefore, since ∇h is uniformly continuous on every bounded subset of $\operatorname{int}\operatorname{dom} h$ and ∇j is uniformly continuous on every bounded subset of $\operatorname{int}\operatorname{dom} j$,

$$\nabla h(x_n) - \nabla h(a_n) \to 0 \quad \text{and} \quad \nabla j(Lx_n) - \nabla j(b_n) \to 0. \tag{3.35}$$

Hence, using (3.2), we get

$$a_n^* + L^* y_n^* \to 0 \quad \text{and} \quad b_n^* - y_n^* \to 0. \tag{3.36}$$

Now, let $\boldsymbol{x} = (x, y^*) \in \mathfrak{W}(\boldsymbol{x}_n)_{n\in\mathbb{N}}$, say $\boldsymbol{x}_{k_n} \rightharpoonup \boldsymbol{x}$. Then $x_{k_n} \rightharpoonup x$ and $y_{k_n}^* \rightharpoonup y^*$, and we derive from (3.34) and (3.36) that

$$\begin{cases} a_{k_n} \rightharpoonup x \\ b_{k_n}^* \rightharpoonup y^* \end{cases} \quad \text{and} \quad \begin{cases} La_{k_n} - b_{k_n} \to 0 \\ a_{k_n}^* + L^* b_{k_n}^* \to 0. \end{cases} \tag{3.37}$$

It therefore follows from (3.11), Proposition 2.1(iii), and (3.24) that $\boldsymbol{x} \in \boldsymbol{Z} \cap \overline{\operatorname{dom}\boldsymbol{f}} = \boldsymbol{C}$. Hence, we derive from Proposition 2.5(vii) that

$$D^f(x_n, \overline{x}) + D^{g^*}(y_n^*, \overline{y}^*) = D^{\boldsymbol{f}}(\boldsymbol{x}_n, \overline{\boldsymbol{x}}) \to 0, \tag{3.38}$$

where $\overline{\boldsymbol{x}} = (\overline{x}, \overline{y}^*)$. Hence, $D^f(x_n, \overline{x}) \to 0$, $D^{g^*}(y_n^*, \overline{y}^*) \to 0$, and, since f and g^* satisfy Condition 3.1, we conclude that $x_n \to \overline{x}$ and $y_n^* \to \overline{y}^*$. ☐

Remark 3.3 We provide a couple of settings that satisfy the assumptions of Theorem 3.2.

(i) In Problem 1.2, suppose that $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces, that $f = \|\cdot\|^2/2$, and that $g = \|\cdot\|^2/2$. Furthermore, in Theorem 3.2, set $h = f$ and $j = g$, and note that, for any $\varepsilon \in \left]0, +\infty\right[$, $\nabla h + \varepsilon A = \mathrm{Id} + \varepsilon A$ and $\nabla j + \varepsilon B = \mathrm{Id} + \varepsilon B$ are strongly monotone and hence coercive by Lemma 2.7(ii). Then we recover the framework of [3], which has been applied to domain decomposition problems in [4].

(ii) Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be measure spaces, let p and q be in $\left]1, +\infty\right[$, and set $p^* = p/(p-1)$ and $q^* = q/(q-1)$. In Problem 1.2, suppose that $\mathcal{X} = L^p(\Omega_1, \mathcal{F}_1, \mu_1)$, $\mathcal{Y} = L^q(\Omega_2, \mathcal{F}_2, \mu_2)$, $f = \|\cdot\|^p/p$, and $g = \|\cdot\|^q/q$. Then $\mathcal{X}^* = L^{p^*}(\Omega_1, \mathcal{F}_1, \mu_1)$, $\mathcal{Y}^* = L^{q^*}(\Omega_2, \mathcal{F}_2, \mu_2)$, and $g^* = \|\cdot\|^{q^*}/q^*$. Moreover, it follows from Clarkson's theorem [17, Theorem II.4.7] that $\mathcal{X}$, $\mathcal{X}^*$, $\mathcal{Y}$, and $\mathcal{Y}^*$ are uniformly convex and uniformly smooth. Hence, we derive from [6, Corollary 5.5 and Example 6.5] that f, g, and g^* are Legendre functions which are uniformly convex on every bounded set, and which therefore satisfy Condition 3.1 by virtue of [7, Example 4.10(i)]. Now set $h = f$ and $j = g$ in Theorem 3.2. We derive from [17, Theorem II.2.16(i)] that ∇h and ∇j are uniformly continuous on every bounded subset of $\mathcal{X}$ and $\mathcal{Y}$, respectively. In addition, h and j are supercoercive and therefore, for any $\varepsilon \in \left]0, +\infty\right[$, Proposition 2.11(iv) asserts that $\nabla h + \varepsilon A$ and $\nabla j + \varepsilon B$ are coercive. Finally, it follows from [17, Proposition II.4.9] that $\nabla h \colon x \mapsto |x|^{p-1}\mathrm{sign}(x)$ and $\nabla j \colon y \mapsto |y|^{q-1}\mathrm{sign}(y)$.

Remark 3.4 The implementation of algorithm (3.2) requires the evaluation of the operator $(\nabla h + A)^{-1}$. We provide a simple example in the Euclidean plane $\mathcal{X}$ of a maximally monotone operator A for which $(\nabla h + A)^{-1}$ can be computed explicitly, whereas the classical resolvent $(\mathrm{Id} + A)^{-1}$ is difficult to evaluate. Let $\beta \in \left]0, +\infty\right[$ and let $\psi \colon \mathbb{R} \to \mathbb{R}$ be a Legendre function with a β Lipschitz-continuous derivative. Set

$$A \colon \mathbb{R}^2 \to \mathbb{R}^2 \colon (\xi_1, \xi_2) \mapsto \big(\beta\xi_1 - \psi'(\xi_1) - \xi_2, \xi_1 + \beta\xi_2 - \psi'(\xi_2)\big) \tag{3.39}$$

and

$$h \colon \mathbb{R}^2 \to \left]-\infty, +\infty\right] \colon (\xi_1, \xi_2) \mapsto \psi(\xi_1) + \psi(\xi_2). \tag{3.40}$$

Then it follows from [9, Theorem 18.15] that A is the sum of the gradient of the convex function $(\xi_1, \xi_2) \mapsto \beta\xi_1^2/2 - \psi(\xi_1) + \beta\xi_2^2/2 - \psi(\xi_2)$ and of the skew linear operator $(\xi_1, \xi_2) \mapsto (-\xi_2, \xi_1)$. Thus, A is a maximally monotone operator [9, Corollary 24.4] which is not the subdifferential of a convex function. In addition, as in Proposition 2.2(i), h is a Legendre function and

$$\big(\forall (\xi_1, \xi_2) \in \mathbb{R}^2\big) \quad (\nabla h + A)^{-1}(\xi_1, \xi_2) = \left(\frac{\beta\xi_1 + \xi_2}{1 + \beta^2}, \frac{\beta\xi_2 - \xi_1}{1 + \beta^2}\right). \tag{3.41}$$

Remark 3.5 At every iteration n, algorithm (3.2) requires the computation of $\boldsymbol{x}_{n+1/2} = P^{\boldsymbol{f}}_{\boldsymbol{H}_n}(x_n, y_n^*)$ and then of $\boldsymbol{x}_{n+1} = Q^{\boldsymbol{f}}((x_0, y_0^*), (x_n, y_n^*), (x_{n+1/2}, y_{n+1/2}^*))$. Set $\boldsymbol{s}_n^* = (a_n^* + L^*b_n^*, b_n - La_n)$, $\eta_n = \langle a_n, a_n^* \rangle + \langle b_n, b_n^* \rangle$, and $\boldsymbol{x}_n = (x_n, y_n^*)$, Then, if $\boldsymbol{x}_n \notin \boldsymbol{H}_n$, $\boldsymbol{x}_{n+1/2}$ is the Bregman projection of $\boldsymbol{x}_n$ onto the closed affine hyperplane $\{\boldsymbol{x} \in \boldsymbol{\mathcal{X}} \mid \langle \boldsymbol{x}, \boldsymbol{s}_n^* \rangle = \eta_n\}$. Thus, $\boldsymbol{x}_{n+1/2}$ is the solution of the problem

$$\underset{\langle \boldsymbol{p}, \boldsymbol{s}_n^* \rangle = \eta_n}{\text{minimize}} \;\; \boldsymbol{f}(\boldsymbol{p}) - \langle \boldsymbol{p}, \nabla \boldsymbol{f}(\boldsymbol{x}_n) \rangle \tag{3.42}$$

which, using standard first order conditions, is characterized by (see also [5, Remark 6.13] and [16, Lemma 2.2.1])

$$\begin{cases}\nabla \boldsymbol{f}(\boldsymbol{x}_{n+1/2}) = \nabla \boldsymbol{f}(\boldsymbol{x}_n) - \lambda \boldsymbol{s}_n^* \\ \langle \boldsymbol{x}_{n+1/2}, \boldsymbol{s}_n^* \rangle = \eta_n \\ \lambda \in \left]0,+\infty\right[. \end{cases} \tag{3.43}$$

In view of [6, Theorem 5.10], the Lagrange multiplier λ is uniquely determined by the equation $\langle \nabla \boldsymbol{f}^*(\nabla \boldsymbol{f}(\boldsymbol{x}_n) - \lambda \boldsymbol{s}_n^*), \boldsymbol{s}_n^* \rangle = \eta_n$. The problem therefore reduces to finding the solution $\overline{\lambda}$ to this equation in $\left]0,+\infty\right[$ and then setting $\boldsymbol{x}_{n+1/2} = \nabla \boldsymbol{f}^*(\nabla \boldsymbol{f}(\boldsymbol{x}_n) - \overline{\lambda} \boldsymbol{s}_n^*)$. Likewise, it follows from (2.15) that $\boldsymbol{x}_{n+1}$ is the unique solution to the problem

$$\underset{\substack{\langle \boldsymbol{p}-\boldsymbol{x}_n, \nabla \boldsymbol{f}(\boldsymbol{x}_0) - \nabla \boldsymbol{f}(\boldsymbol{x}_n)\rangle \leqslant 0 \\ \langle \boldsymbol{p}-\boldsymbol{x}_{n+1/2}, \nabla \boldsymbol{f}(\boldsymbol{x}_n) - \nabla \boldsymbol{f}(\boldsymbol{x}_{n+1/2})\rangle \leqslant 0}}{\text{minimize}} \quad \boldsymbol{f}(\boldsymbol{p}) - \langle \boldsymbol{p}, \nabla \boldsymbol{f}(\boldsymbol{x}_0) \rangle. \tag{3.44}$$

Depending on the number of active constraints, this problem boils down to determining up to two Lagrange multipliers in $\left]0,+\infty\right[$.

Next, we consider a specialization of Problem 1.2 to multivariate structured minimization.

Problem 3.6 Let m and p be strictly positive integers, let $(\mathcal{X}_i)_{1\leqslant i\leqslant m}$ and $(\mathcal{Y}_k)_{1\leqslant k\leqslant p}$ be reflexive real Banach spaces, and let $\boldsymbol{\mathcal{X}}$ be the standard vector product space $\left(\times_{i=1}^m \mathcal{X}_i\right) \times \left(\times_{k=1}^p \mathcal{Y}_k^*\right)$ equipped with the norm

$$(x, y^*) = \big((x_i)_{1\leqslant i\leqslant m}, (y_k^*)_{1\leqslant k\leqslant p}\big) \mapsto \sqrt{\sum_{i=1}^m \|x_i\|^2 + \sum_{k=1}^p \|y_k^*\|^2}. \tag{3.45}$$

For every $i \in \{1,\ldots,m\}$ and every $k \in \{1,\ldots,p\}$, let $\varphi_i \in \Gamma_0(\mathcal{X}_i)$, let $\psi_k \in \Gamma_0(\mathcal{Y}_k)$, and let $L_{ki} \colon \mathcal{X}_i \to \mathcal{Y}_k$ be linear and bounded. Consider the primal problem

$$\underset{x_1\in\mathcal{X}_1,\ldots,\, x_m\in\mathcal{X}_m}{\text{minimize}} \quad \sum_{i=1}^m \varphi_i(x_i) + \sum_{k=1}^p \psi_k\bigg(\sum_{i=1}^m L_{ki} x_i\bigg), \tag{3.46}$$

the dual problem

$$\underset{y_1^*\in\mathcal{Y}_1^*,\ldots,\, y_p^*\in\mathcal{Y}_p^*}{\text{minimize}} \quad \sum_{i=1}^m \varphi_i^*\bigg(-\sum_{k=1}^p L_{ki}^* y_k^*\bigg) + \sum_{k=1}^p \psi_k^*(y_k^*), \tag{3.47}$$

and let

$$\boldsymbol{Z} = \bigg\{ (x, y^*) \in \boldsymbol{\mathcal{X}} \;\bigg|\; (\forall i \in \{1,\ldots,m\}) \;\; -\sum_{k=1}^p L_{ki}^* y_k^* \in \partial \varphi_i(x_i) \;\text{ and }\; (\forall k \in \{1,\ldots,p\}) \;\; \sum_{i=1}^m L_{ki} x_i \in \partial \psi_k^*(y_k^*) \bigg\} \tag{3.48}$$

be the associated Kuhn-Tucker set. For every $i \in \{1, \dots, m\}$, let $f_i \in \Gamma_0(\mathcal{X}_i)$ be a Legendre function and let $x_{i,0} \in \operatorname{int} \operatorname{dom} f_i$. For every $k \in \{1, \dots, p\}$, let $g_k \in \Gamma_0(\mathcal{Y}_k)$ be a Legendre function and let $y^*_{k,0} \in \operatorname{int} \operatorname{dom} g^*_k$. Set $x_0 = (x_{i,0})_{1 \leqslant i \leqslant m}$, $y^*_0 = (y^*_{k,0})_{1 \leqslant k \leqslant p}$, and

$$\boldsymbol{f} \colon \boldsymbol{\mathcal{X}} \to \left]-\infty, +\infty\right] : (x, y^*) \mapsto \sum_{i=1}^{m} f_i(x_i) + \sum_{k=1}^{p} g^*_k(y^*_k), \tag{3.49}$$

and suppose that $\boldsymbol{Z} \cap \operatorname{int} \operatorname{dom} \boldsymbol{f} \neq \varnothing$. The objective is to find the best Bregman approximation $\big((\overline{x}_i)_{1 \leqslant i \leqslant m}, (\overline{y}^*_k)_{1 \leqslant k \leqslant p}\big) = P^{\boldsymbol{f}}_{\boldsymbol{Z}}(x_0, y^*_0)$ to (x_0, y^*_0) from $\boldsymbol{Z}$.

We derive from Theorem 3.2 the following convergence result for a splitting algorithm to solve Problem 3.6.

Proposition 3.7 *Consider the setting of Problem 3.6. For every $i \in \{1, \dots, m\}$, let $h_i \in \Gamma_0(\mathcal{X}_i)$ be a Legendre function such that $\operatorname{int} \operatorname{dom} f_i \subset \operatorname{int} \operatorname{dom} h_i$ and $h_i + \varepsilon_i \varphi_i$ is supercoercive for some $\varepsilon_i \in \left]0, +\infty\right[$. For every $k \in \{1, \dots, p\}$, let $j_k \in \Gamma_0(\mathcal{Y}_k)$ be a Legendre function such that $\sum_{i=1}^m L_{ki}(\operatorname{int} \operatorname{dom} f_i) \subset \operatorname{int} \operatorname{dom} j_k$ and $j_k + \delta_k \psi_k$ is supercoercive for some $\delta_k \in \left]0, +\infty\right[$. Set $\varepsilon = \max_{1 \leqslant i \leqslant m} \varepsilon_i$ and $\delta = \max_{1 \leqslant k \leqslant p} \delta_k$, let $\sigma \in \left[\max\{\varepsilon, \delta\}, +\infty\right[$, and iterate*

$$\begin{array}{l}
\text{for } n = 0, 1, \dots \\
\left\lfloor \begin{array}{l}
(\gamma_n, \mu_n) \in [\varepsilon, \sigma] \times [\delta, \sigma] \\
\text{for } i = 1, \dots, m \\
\left\lfloor \begin{array}{l}
a_{i,n} = (\nabla h_i + \gamma_n \partial \varphi_i)^{-1} \Big(\nabla h_i(x_{i,n}) - \gamma_n \sum_{k=1}^p L^*_{ki} y^*_{k,n}\Big) \\
a^*_{i,n} = \gamma_n^{-1} \big(\nabla h_i(x_{i,n}) - \nabla h_i(a_{i,n})\big) - \sum_{k=1}^p L^*_{ki} y^*_{k,n}
\end{array} \right. \\
\text{for } k = 1, \dots, p \\
\left\lfloor \begin{array}{l}
b_{k,n} = (\nabla j_k + \mu_n \partial \psi_k)^{-1} \big(\nabla j_k\big(\sum_{i=1}^m L_{ki} x_{i,n}\big) + \mu_n y^*_{k,n}\big) \\
b^*_{k,n} = \mu_n^{-1} \big(\nabla j_k\big(\sum_{i=1}^m L_{ki} x_{i,n}\big) - \nabla j_k(b_{k,n})\big) + y^*_{k,n} \\
t_{k,n} = b_{k,n} - \sum_{i=1}^m L_{ki} a_{i,n}
\end{array} \right. \\
\text{for } i = 1, \dots, m \\
\left\lfloor \begin{array}{l}
s^*_{i,n} = a^*_{i,n} + \sum_{k=1}^p L^*_{ki} b^*_{k,n}
\end{array} \right. \\
\eta_n = \sum_{i=1}^m \langle a_{i,n}, a^*_{i,n} \rangle + \sum_{k=1}^p \langle b_{k,n}, b^*_{k,n} \rangle \\
\boldsymbol{H}_n = \Big\{ (x, y^*) \in \boldsymbol{\mathcal{X}} \;\Big|\; \sum_{i=1}^m \langle x_i, s^*_{i,n} \rangle + \sum_{k=1}^p \langle t_{k,n}, y^*_k \rangle \leqslant \eta_n \Big\} \\
(x_{n+1/2}, y^*_{n+1/2}) = P^{\boldsymbol{f}}_{\boldsymbol{H}_n}(x_n, y^*_n) \\
(x_{n+1}, y^*_{n+1}) = Q^f\big((x_0, y^*_0), (x_n, y^*_n), (x_{n+1/2}, y^*_{n+1/2})\big),
\end{array} \right.
\end{array} \tag{3.50}$$

where we use the notation $(\forall n \in \mathbb{N})$ $x_n = (x_{i,n})_{1 \leqslant i \leqslant m}$ and $y^*_n = (y^*_{k,n})_{1 \leqslant k \leqslant p}$. Suppose that the following hold:

(i) *For every $i \in \{1, \dots, m\}$, f_i and h_i satisfy Condition 3.1 and ∇h_i is uniformly continuous on every bounded subset of $\operatorname{int} \operatorname{dom} h_i$.*

(ii) *For every $k \in \{1, \dots, p\}$, g^*_k and j_k satisfy Condition 3.1 and ∇j_k is uniformly continuous on every bounded subset of $\operatorname{int} \operatorname{dom} j_k$.*

Then

$$(\forall i \in \{1,\ldots,m\}) \quad x_{i,n} \to \overline{x}_i \quad \textit{and} \quad (\forall k \in \{1,\ldots,p\}) \quad y^*_{k,n} \to \overline{y}^*_k. \tag{3.51}$$

Proof. Denote by $\mathcal{X}$ and $\mathcal{Y}$ the standard vector product spaces $\bigtimes_{i=1}^m \mathcal{X}_i$ and $\bigtimes_{k=1}^p \mathcal{Y}_k$ equipped with the norms $x = (x_i)_{1\leqslant i\leqslant m} \mapsto \sqrt{\sum_{i=1}^m \|x_i\|^2}$ and $y = (y_k)_{1\leqslant k\leqslant p} \mapsto \sqrt{\sum_{k=1}^p \|y_k\|^2}$, respectively. Then $\mathcal{X}^*$ is the vector product space $\bigtimes_{i=1}^m \mathcal{X}_i^*$ equipped with the norm $x^* \mapsto \sqrt{\sum_{i=1}^m \|x_i^*\|^2}$ and $\mathcal{Y}^*$ is the vector product space $\bigtimes_{k=1}^p \mathcal{Y}_k^*$ equipped with the norm $y^* \mapsto \sqrt{\sum_{k=1}^p \|y_k^*\|^2}$. Let us introduce the operators

$$\begin{cases} A\colon \mathcal{X} \to 2^{\mathcal{X}^*} : x \mapsto \bigtimes_{i=1}^m \partial\varphi_i(x_i) \\ B\colon \mathcal{Y} \to 2^{\mathcal{Y}^*} : y \mapsto \bigtimes_{k=1}^p \partial\psi_k(y_k) \\ L\colon \mathcal{X} \to \mathcal{Y} \colon x \mapsto \big(\sum_{i=1}^m L_{ki} x_i\big)_{1\leqslant k\leqslant p} \end{cases} \tag{3.52}$$

and the functions

$$\begin{cases} f\colon \mathcal{X} \to \left]-\infty,+\infty\right] : x \mapsto \sum_{i=1}^m f_i(x_i) \\ h\colon \mathcal{X} \to \left]-\infty,+\infty\right] : x \mapsto \sum_{i=1}^m h_i(x_i) \\ \varphi\colon \mathcal{X} \to \left]-\infty,+\infty\right] : x \mapsto \sum_{i=1}^m \varphi_i(x_i) \\ g\colon \mathcal{Y} \to \left]-\infty,+\infty\right] : y \mapsto \sum_{k=1}^p g_k(y_k) \\ j\colon \mathcal{Y} \to \left]-\infty,+\infty\right] : y \mapsto \sum_{k=1}^p j_k(y_k). \end{cases} \tag{3.53}$$

Then it follows from [40, Theorem 3.1.11] that A and B are maximally monotone. In addition, the adjoint of L is $L^*\colon \mathcal{Y}^* \to \mathcal{X}^*\colon y^* \mapsto (\sum_{k=1}^p L_{ki}^* y_k^*)_{1\leqslant i\leqslant m}$, and, as in Proposition 2.2(i), f and g are Legendre functions. Thus, Problem 3.6 is a special case of Problem 1.2. Furthermore, h and j are Legendre functions,

$$\operatorname{int}\operatorname{dom} f = \bigtimes_{i=1}^m \operatorname{int}\operatorname{dom} f_i \subset \bigtimes_{i=1}^m \operatorname{int}\operatorname{dom} h_i = \operatorname{int}\operatorname{dom} h, \tag{3.54}$$

and

$$L\big(\operatorname{int}\operatorname{dom} f\big) = \bigtimes_{k=1}^p \sum_{i=1}^m L_{ki}(\operatorname{int}\operatorname{dom} f_i) \subset \bigtimes_{k=1}^p \operatorname{int}\operatorname{dom} j_k = \operatorname{int}\operatorname{dom} j. \tag{3.55}$$

Next we observe that, for every $i \in \{1,\ldots,m\}$, since $h_i+\varepsilon\varphi_i$ is supercoercive, $(h_i+\varepsilon\varphi_i)^*$ is bounded above on every bounded subset of $\mathcal{X}_i^*$ [6, Theorem 3.3]. As a result, $(h+\varepsilon\varphi)^*\colon x^* \mapsto \sum_{i=1}^m (h_i + \varepsilon\varphi_i)^*(x_i^*)$ is bounded above on every bounded subset of $\mathcal{X}^*$, and it follows from [6, Theorem 3.3] that $h+\varepsilon\varphi$ is supercoercive. In turn since, as in (3.7), $\varnothing \neq \operatorname{dom} A \cap \operatorname{int}\operatorname{dom} f \subset \operatorname{dom}\varphi \cap \operatorname{int}\operatorname{dom} f$, we derive from [40, Theorem 2.8.3] and Lemma 2.7(iii) that

$$\nabla h + \varepsilon A = \nabla h + \varepsilon\partial\varphi = \partial(h+\varepsilon\varphi) \tag{3.56}$$

is coercive. We show in a similar fashion that $\nabla j + \delta B$ is coercive. Now set, for every $n \in \mathbb{N}$, $a_n = (a_{i,n})_{1\leqslant i\leqslant m}$, $a_n^* = (a_{i,n}^*)_{1\leqslant i\leqslant m}$, $b_n = (b_{k,n})_{1\leqslant k\leqslant p}$, and $b_n^* = (b_{k,n}^*)_{1\leqslant k\leqslant p}$. Then, for every $n \in \mathbb{N}$, we have

$$
\begin{aligned}
&(\forall i \in \{1,\ldots,m\}) \quad a_{i,n} = (\nabla h_i + \gamma_n \partial\varphi_i)^{-1}\Big(\nabla h_i(x_{i,n}) - \gamma_n \sum_{k=1}^{p} L_{ki}^* y_{k,n}^*\Big)\\
&\quad\Leftrightarrow (\forall i \in \{1,\ldots,m\}) \quad \nabla h_i(x_{i,n}) - \gamma_n \sum_{k=1}^{p} L_{ki}^* y_{k,n}^* \in \nabla h_i(a_{i,n}) + \gamma_n \partial\varphi_i(a_{i,n})\\
&\quad\Leftrightarrow \nabla h(x_n) - \gamma_n L^* y_n^* \in \nabla h(a_n) + \gamma_n A a_n\\
&\quad\Leftrightarrow a_n = (\nabla h + \gamma_n A)^{-1}\big(\nabla h(x_n) - \gamma_n L^* y_n^*\big).
\end{aligned}
\tag{3.57}
$$

Likewise,

$$
(\forall n \in \mathbb{N}) \quad b_n = (\nabla j + \mu_n B)^{-1}\big(\nabla j(Lx_n) + \mu_n y_n^*\big). \tag{3.58}
$$

Thus, (3.50) is a special case of (3.2). In addition, it follows from our assumptions and (3.53) that f, g^*, h, and j satisfy Condition 3.1, and that ∇h and ∇j are uniformly continuous on every bounded subset of $\operatorname{int}\operatorname{dom} h$ and $\operatorname{int}\operatorname{dom} j$, respectively. Altogether, the conclusions follow from Theorem 3.2(iv), with $\overline{x} = (\overline{x}_i)_{1\leqslant i\leqslant m}$ and $\overline{y}^* = (\overline{y}_k^*)_{1\leqslant k\leqslant p}$. ☐

Remark 3.8 In Problem 3.6, suppose that, for every $i \in \{1,\ldots,m\}$ and every $k \in \{1,\ldots,p\}$, φ_i and ψ_k are supercoercive Legendre functions satisfying Condition 3.1, that $\nabla\varphi_i$ and $\nabla\psi_k$ are uniformly continuous on bounded subset of $\operatorname{int}\operatorname{dom}\varphi_i$ and $\operatorname{int}\operatorname{dom}\psi_k$, respectively, and that $\sum_{i=1}^m L_{ki}(\operatorname{int}\operatorname{dom}\varphi_i) \subset \operatorname{int}\operatorname{dom}\psi_k$. Then, in Proposition 3.7, we can choose, for every $i \in \{1,\ldots,m\}$ and every $k \in \{1,\ldots,p\}$, $h_i = \varphi_i$ and $j_k = \psi_k$, and in (3.50), we obtain

$$
a_{i,n} = \nabla h_i^* \left(\frac{\nabla h_i(x_{i,n}) - \gamma_n \sum_{k=1}^{p} L_{ki}^* y_{k,n}^*}{1+\gamma_n} \right) \tag{3.59}
$$

and

$$
b_{k,n} = \nabla j_k^* \left(\frac{\nabla j_k\Big(\sum_{i=1}^{m} L_{ki} x_{i,n}\Big) + \mu_n y_{k,n}^*}{1+\mu_n} \right). \tag{3.60}
$$

For example, suppose that, for every $i \in \{1,\ldots,m\}$ and every $k \in \{1,\ldots,p\}$, $\mathcal{X}_i = \mathbb{R}$, $\mathcal{Y}_k = \mathbb{R}$, and $\varphi_i = h_i$ is the Hellinger-like function, i.e.,

$$
\varphi_i \colon \mathbb{R} \to \left]-\infty,+\infty\right] : x_i \mapsto \begin{cases} -\sqrt{1-x_i^2}, & \text{if } x_i \in [-1,1]\,;\\ +\infty, & \text{otherwise.} \end{cases} \tag{3.61}
$$

Then (3.59) becomes

$$a_{i,n} = \frac{x_{i,n} - \gamma_n\left(\displaystyle\sum_{k=1}^{p} L_{ki}^* y_{k,n}^*\right)\sqrt{1-x_{i,n}^2}}{\sqrt{(1+\gamma_n)^2(1-x_{i,n}^2) + \left(x_{i,n} - \gamma_n\left(\displaystyle\sum_{k=1}^{p} L_{ki}^* y_{k,n}^*\right)\sqrt{1-x_{i,n}^2}\right)^2}}. \tag{3.62}$$

Furthermore, as shown in the next section, in finite-dimensional spaces, we can remove Condition 3.1 and the assumption on the uniform continuity of $(\nabla\varphi_i)_{1\leqslant i\leqslant m}$ and $(\nabla\psi_k)_{1\leqslant k\leqslant p}$.

4 Finite-dimensional setting

In finite-dimensional spaces, the convergence of algorithm (3.2) can be obtained under more general assumptions. To establish the corresponding results, the following technical facts will be needed.

Lemma 4.1 *Let $\mathcal{X}$ be a finite-dimensional real Banach space and let $f \in \Gamma_0(\mathcal{X})$ be a Legendre function. Then the following hold:*

(i) *f and ∇f are continuous on* $\operatorname{int}\operatorname{dom} f$ [9, Corollaries 8.30(iii), 17.34, and 17.35].

(ii) $\nabla f\colon \operatorname{int}\operatorname{dom} f \to \operatorname{int}\operatorname{dom} f^*$ *is bijective with inverse* $\nabla f^*\colon \operatorname{int}\operatorname{dom} f^* \to \operatorname{int}\operatorname{dom} f$ [6, Theorem 5.10].

(iii) *Let $x \in \operatorname{int}\operatorname{dom} f$, let $y \in \overline{\operatorname{dom}} f$, and let $(y_n)_{n\in\mathbb{N}} \in (\operatorname{int}\operatorname{dom} f)^{\mathbb{N}}$. Suppose that $y_n \to y$ and that $(D^f(x,y_n))_{n\in\mathbb{N}}$ is bounded. Then $y \in \operatorname{int}\operatorname{dom} f$ and $D^f(y,y_n) \to 0$* [5, Theorem 3.8(ii)].

(iv) *Let $x \in \operatorname{int}\operatorname{dom} f$, let $y \in \operatorname{int}\operatorname{dom} f$, let $(x_n)_{n\in\mathbb{N}} \in (\operatorname{dom} f)^{\mathbb{N}}$, and let $(y_n)_{n\in\mathbb{N}} \in (\operatorname{int}\operatorname{dom} f)^{\mathbb{N}}$. Suppose that $D^f(x_n,y_n) \to 0$. Then $x = y$* [5, Theorem 3.9(iii)].

(v) *Let $y \in \operatorname{int}\operatorname{dom} f$. Then $D^f(\cdot,y)$ is coercive* [6, Lemma 7.3(v)].

(vi) *Let $\{x,y\} \subset \operatorname{int}\operatorname{dom} f$. Then $D^f(x,y) = D^{f^*}(\nabla f(y),\nabla f(x))$* [6, Lemma 7.3(vii)].

Proposition 4.2 *In Problem 1.2, suppose that $\mathcal{X}$ and $\mathcal{Y}$ are finite-dimensional. Let $h \in \Gamma_0(\mathcal{X})$ and $j \in \Gamma_0(\mathcal{Y})$ be Legendre functions such that $\operatorname{int}\operatorname{dom} f \subset \operatorname{int}\operatorname{dom} h$, $L(\operatorname{int}\operatorname{dom} f) \subset \operatorname{int}\operatorname{dom} j$, and there exist ε and δ in $]0,+\infty[$ such that $\nabla h + \varepsilon A$ and $\nabla j + \delta B$ are coercive. Let $\sigma \in [\max\{\varepsilon,\delta\},+\infty[$ and execute algorithm (3.2). Then $(x_n, y_n^*) \to (\overline{x}, \overline{y}^*)$.*

Proof. Set $\boldsymbol{C} = \boldsymbol{Z} \cap \overline{\operatorname{dom}}\,\boldsymbol{f}$. We first observe that, as in (3.24),

$$(x_n)_{n\in\mathbb{N}} \in (\operatorname{int}\operatorname{dom} f)^{\mathbb{N}} \quad \text{and} \quad (y_n^*)_{n\in\mathbb{N}} \in (\operatorname{int}\operatorname{dom} g^*)^{\mathbb{N}} \quad \text{are bounded.} \tag{4.1}$$

In addition, we deduce from (3.10), (3.14), and Proposition 2.5(i) that $\overline{\boldsymbol{x}} = (\overline{x}, \overline{y}^*) \in \boldsymbol{C} \subset \bigcap_{n\in\mathbb{N}} H^{\boldsymbol{f}}(\boldsymbol{x}_0, \boldsymbol{x}_n)$, and hence from (2.14) that

$$(\forall n \in \mathbb{N}) \quad D^f(\overline{x}, x_n) + D^{g^*}(\overline{y}^*, y_n^*) = D^{\boldsymbol{f}}\big(\overline{\boldsymbol{x}}, \boldsymbol{x}_n\big) \leqslant D^{\boldsymbol{f}}\big(\overline{\boldsymbol{x}}, \boldsymbol{x}_0\big). \tag{4.2}$$

By virtue of Proposition 2.5(vii) and (4.1), it suffices to show that every cluster point of $(x_n, y_n^*)_{n\in\mathbb{N}}$ belongs to $\boldsymbol{Z}$. To this end, take $x \in \mathcal{X}$, $y^* \in \mathcal{Y}$, and a strictly increasing sequence $(k_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that $x_{k_n} \to x$ and $y_{k_n}^* \to y^*$. Then $Lx_{k_n} \to Lx$, $x \in \overline{\operatorname{dom}}\, f$, and $y^* \in \overline{\operatorname{dom}}\, g^*$. Since $\overline{x} \in \operatorname{int}\operatorname{dom} f$ and since (4.2) implies that $(D^f(\overline{x}, x_{k_n}))_{n\in\mathbb{N}}$ is bounded, it follows from Lemma 4.1(iii) that $x \in \operatorname{int}\operatorname{dom} f$. Analogously, $y^* \in \operatorname{int}\operatorname{dom} g^*$. In turn, Lemma 4.1(i) asserts that

$$\nabla f\big(x_{k_n}\big) \to \nabla f(x) \quad \text{and} \quad \nabla g^*\big(y_{k_n}^*\big) \to \nabla g^*(y^*). \tag{4.3}$$

Furthermore, since $\operatorname{int}\operatorname{dom} f \subset \operatorname{int}\operatorname{dom} h$ and $L(\operatorname{int}\operatorname{dom} f) \subset \operatorname{int}\operatorname{dom} j$, we obtain $x \in \operatorname{int}\operatorname{dom} h$ and $Lx \in \operatorname{int}\operatorname{dom} j$. Thus, there exists $\rho \in \left]0, +\infty\right[$ such that $B(x;\rho) \subset \operatorname{int}\operatorname{dom} h$ and $B(Lx;\rho) \subset \operatorname{int}\operatorname{dom} j$. We therefore assume without loss of generality that

$$(x_{k_n})_{n\in\mathbb{N}} \in B(x;\rho)^{\mathbb{N}} \quad \text{and} \quad (Lx_{k_n})_{n\in\mathbb{N}} \in B(Lx;\rho)^{\mathbb{N}}. \tag{4.4}$$

In view of Lemma 4.1(i), $h(B(x;\rho))$ and $\nabla h(B(x;\rho))$ are therefore compact, which implies that $(h(x_{k_n}))_{n\in\mathbb{N}}$ and $(\nabla h(x_{k_n}))_{n\in\mathbb{N}}$ are bounded. Hence $(D^h(\overline{x}, x_{k_n}))_{n\in\mathbb{N}}$ is bounded and, moreover, it follows from (3.2), (4.1), Lemma 2.9, and (3.9) that $(a_{k_n})_{n\in\mathbb{N}}$ is a bounded sequence in $\operatorname{int}\operatorname{dom} h$. We show likewise that $(D^j(L\overline{x}, Lx_{k_n}))_{n\in\mathbb{N}}$ and $(b_{k_n})_{n\in\mathbb{N}}$ are bounded. Next, since the convexity of h yields

$$\begin{aligned}(\forall n \in \mathbb{N}) \;\; D^h(\overline{x}, a_{k_n}) &= h(\overline{x}) - h(a_{k_n}) - \langle \overline{x} - a_{k_n}, \nabla h(a_{k_n})\rangle \\ &= h(\overline{x}) - h(x_{k_n}) - \langle \overline{x} - x_{k_n}, \nabla h(x_{k_n})\rangle + \big\langle \overline{x} - a_{k_n}, \nabla h(x_{k_n}) - \nabla h(a_{k_n})\big\rangle \\ &\quad - \big(h(a_{k_n}) - h(x_{k_n}) - \langle a_{k_n} - x_{k_n}, \nabla h(x_{k_n})\rangle\big) \\ &\leqslant D^h(\overline{x}, x_{k_n}) + \big\langle \overline{x} - a_{k_n}, \nabla h(x_{k_n}) - \nabla h(a_{k_n})\big\rangle,\end{aligned} \tag{4.5}$$

we derive from (3.2) that

$$\begin{aligned}(\forall n \in \mathbb{N}) \;\; \sigma^{-1} D^h(\overline{x}, a_{k_n}) &\leqslant \gamma_{k_n}^{-1} D^h(\overline{x}, a_{k_n}) \\ &\leqslant \varepsilon^{-1} D^h(\overline{x}, x_{k_n}) + \big\langle \overline{x} - a_{k_n}, \gamma_{k_n}^{-1}\big(\nabla h(x_{k_n}) - \nabla h(a_{k_n})\big)\big\rangle \\ &= \varepsilon^{-1} D^h(\overline{x}, x_{k_n}) + \langle \overline{x} - a_{k_n}, a_{k_n}^* + L^* y_{k_n}^*\rangle \\ &= \varepsilon^{-1} D^h(\overline{x}, x_{k_n}) + \langle \overline{x} - a_{k_n}, a_{k_n}^* + L^* \overline{y}^*\rangle \\ &\quad + \langle L\overline{x} - La_{k_n}, y_{k_n}^* - \overline{y}^*\rangle.\end{aligned} \tag{4.6}$$

Similarly,

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad \sigma^{-1} D^j(L\overline{x}, b_{k_n}) &\leqslant \delta^{-1} D^j(L\overline{x}, Lx_{k_n}) + \langle L\overline{x} - b_{k_n}, b_{k_n}^* - y_{k_n}^*\rangle \\ &\leqslant \delta^{-1} D^j(L\overline{x}, Lx_{k_n}) + \langle L\overline{x} - b_{k_n}, b_{k_n}^* - \overline{y}^*\rangle \\ &\quad + \langle L\overline{x} - b_{k_n}, \overline{y}^* - y_{k_n}^*\rangle.\end{aligned} \tag{4.7}$$

Since (3.12) entails that

$$\overline{\boldsymbol{x}} \in \boldsymbol{C} \subset \bigcap_{n\in\mathbb{N}} \boldsymbol{H}_{k_n} = \bigcap_{n\in\mathbb{N}} \big\{(x,y^*) \in \boldsymbol{\mathcal{X}} \mid \langle x-a_{k_n}, a^*_{k_n}+L^*y^*\rangle + \langle Lx-b_{k_n}, b^*_{k_n}-y^*\rangle \leqslant 0\big\}, \tag{4.8}$$

we deduce from (4.6) and (4.7) that

$$\begin{aligned}
(\forall n\in\mathbb{N})\quad & \sigma^{-1}\big(D^h(\overline{x},a_{k_n}) + D^j(L\overline{x},b_{k_n})\big)\\
&\leqslant \varepsilon^{-1}D^h(\overline{x},x_{k_n}) + \delta^{-1}D^j(L\overline{x},Lx_{k_n}) + \langle L\overline{x}-La_{k_n}, y^*_{k_n}-\overline{y}^*\rangle + \langle L\overline{x}-b_{k_n}, \overline{y}^*-y^*_{k_n}\rangle\\
&= \varepsilon^{-1}D^h(\overline{x},x_{k_n}) + \delta^{-1}D^j(L\overline{x},Lx_{k_n}) + \langle b_{k_n}-La_{k_n}, y^*_{k_n}-\overline{y}^*\rangle\\
&\leqslant \varepsilon^{-1}D^h(\overline{x},x_{k_n}) + \delta^{-1}D^j(L\overline{x},Lx_{k_n}) + (\|b_{k_n}\| + \|L\|\,\|a_{k_n}\|)(\|y^*_{k_n}\| + \|\overline{y}^*\|).
\end{aligned} \tag{4.9}$$

Hence, the boundedness of $(a_{k_n})_{n\in\mathbb{N}}$, $(b_{k_n})_{n\in\mathbb{N}}$, $(y^*_{k_n})_{n\in\mathbb{N}}$, $(D^h(\overline{x},x_{k_n}))_{n\in\mathbb{N}}$, and $(D^j(L\overline{x},Lx_{k_n}))_{n\in\mathbb{N}}$ implies that of $(D^h(\overline{x},a_{k_n}))_{n\in\mathbb{N}}$ and $(D^j(L\overline{x},b_{k_n}))_{n\in\mathbb{N}}$. In turn, by Lemma 4.1(vi),

$$\Big(D^{h^*}\big(\nabla h(a_{k_n}),\nabla h(\overline{x})\big)\Big)_{n\in\mathbb{N}} \quad\text{and}\quad \Big(D^{j^*}\big(\nabla j(b_{k_n}),\nabla j(L\overline{x})\big)\Big)_{n\in\mathbb{N}} \quad\text{are bounded} \tag{4.10}$$

and, since Lemma 4.1(v) asserts that $D^{h^*}(\cdot,\nabla h(\overline{x}))$ and $D^{j^*}(\cdot,\nabla j(L\overline{x}))$ are coercive, it follows from (4.10) that $(\nabla h(a_{k_n}))_{n\in\mathbb{N}}$ and $(\nabla j(b_{k_n}))_{n\in\mathbb{N}}$ are bounded. Thus, since $(y^*_{k_n})_{n\in\mathbb{N}}$, $(\nabla h(x_{k_n}))_{n\in\mathbb{N}}$ and $(\nabla j(Lx_{k_n}))_{n\in\mathbb{N}}$ are bounded, we infer from (3.2) that

$$(a^*_{k_n})_{n\in\mathbb{N}} \quad\text{and}\quad (b^*_{k_n})_{n\in\mathbb{N}} \quad\text{are bounded.} \tag{4.11}$$

On the other hand, since (3.12) yields $\overline{\boldsymbol{x}} \in \boldsymbol{C} \subset \bigcap_{n\in\mathbb{N}} H^{\boldsymbol{f}}(\boldsymbol{x}_{k_n},\boldsymbol{x}_{k_n+1/2})$, (2.14) and (4.2) imply that

$$(\forall n\in\mathbb{N})\quad D^f(\overline{x},x_{k_n+1/2}) + D^{g^*}(\overline{y}^*,y^*_{k_n+1/2}) = D^{\boldsymbol{f}}\big(\overline{\boldsymbol{x}},\boldsymbol{x}_{k_n+1/2}\big) \leqslant D^{\boldsymbol{f}}\big(\overline{\boldsymbol{x}},\boldsymbol{x}_{k_n}\big) \leqslant D^{\boldsymbol{f}}\big(\overline{\boldsymbol{x}},\boldsymbol{x}_0\big). \tag{4.12}$$

Thus, Lemma 4.1(vi) yields

$$\begin{aligned}
(\forall n\in\mathbb{N})\quad D^{f^*}\big(\nabla f(x_{k_n+1/2}),\nabla f(\overline{x})\big) + D^g\big(\nabla g^*(y^*_{k_n+1/2}),\nabla g^*(\overline{y}^*)\big)&\\
= D^f(\overline{x},x_{k_n+1/2}) + D^{g^*}(\overline{y}^*,y^*_{k_n+1/2})&\\
\leqslant D^{\boldsymbol{f}}\big(\overline{\boldsymbol{x}},\boldsymbol{x}_0\big)&
\end{aligned} \tag{4.13}$$

and, since $D^{f^*}(\cdot,\nabla f(\overline{x}))$ and $D^g(\cdot,\nabla g^*(\overline{y}^*))$ are coercive by Lemma 4.1(v), it follows that

$$(\nabla f(x_{k_n+1/2}))_{n\in\mathbb{N}} \quad\text{and}\quad (\nabla g^*(y^*_{k_n+1/2}))_{n\in\mathbb{N}} \quad\text{are bounded.} \tag{4.14}$$

However, as in (3.28), $D^f(x_{k_n+1/2},x_{k_n}) \to 0$ and $D^{g^*}(y^*_{k_n+1/2},y^*_{k_n}) \to 0$, and it therefore follows from Lemma 4.1(vi) that

$$D^{f^*}\big(\nabla f(x_{k_n}),\nabla f(x_{k_n+1/2})\big) \to 0 \quad\text{and}\quad D^g\big(\nabla g^*(y^*_{k_n}),\nabla g^*(y^*_{k_n+1/2})\big) \to 0. \tag{4.15}$$

In view of Lemma 4.1(iv), we infer from (4.3), (4.14), and (4.15) that there exists a strictly increasing sequence $(p_{k_n})_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that

$$\nabla f\big(x_{p_{k_n}+1/2}\big) \to \nabla f(x) \quad\text{and}\quad \nabla g^*\big(y^*_{p_{k_n}+1/2}\big) \to \nabla g^*(y^*). \tag{4.16}$$

Since, by Lemma 4.1(i)–(ii), $(\nabla f)^{-1} = \nabla f^*$ is continuous on $\operatorname{int}\operatorname{dom} f^*$ and $(\nabla g^*)^{-1} = \nabla g$ is continuous on $\operatorname{int}\operatorname{dom} g$, we obtain $x_{p_{k_n}+1/2} \to x$ and $y^*_{p_{k_n}+1/2} \to y^*$. Thus,

$$x_{p_{k_n}+1/2} - x_{p_{k_n}} \to 0 \quad \text{and} \quad y^*_{p_{k_n}+1/2} - y^*_{p_{k_n}} \to 0. \tag{4.17}$$

On the other hand, as in (3.27),

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad \|x_{p_{k_n}} - x_{p_{k_n}+1/2}\|\,\|a^*_{p_{k_n}} + L^* b^*_{p_{k_n}}\| + \|b_{p_{k_n}} - La_{p_{k_n}}\|\,\|y^*_{p_{k_n}} - y^*_{p_{k_n}+1/2}\| \\ \geqslant \sigma^{-1}\big(D^h(x_{p_{k_n}}, a_{p_{k_n}}) + D^j(Lx_{p_{k_n}}, b_{p_{k_n}})\big),\end{aligned} \tag{4.18}$$

and hence, since $(a_{p_{k_n}})_{n\in\mathbb{N}}$ and $(b_{p_{k_n}})_{n\in\mathbb{N}}$ are bounded, we deduce from (4.11) and (4.17) that

$$\begin{cases} D^h\big(x_{p_{k_n}}, a_{p_{k_n}}\big) \to 0 \\ D^j\big(Lx_{p_{k_n}}, b_{p_{k_n}}\big) \to 0 \\ x_{p_{k_n}} \to x \\ Lx_{p_{k_n}} \to Lx \\ (a_{p_{k_n}})_{n\in\mathbb{N}} \text{ has a cluster point} \\ (b_{p_{k_n}})_{n\in\mathbb{N}} \text{ has a cluster point.} \end{cases} \tag{4.19}$$

Consequently, by dropping to a subsequence if necessary and invoking Lemma 4.1(iv), we get

$$a_{p_{k_n}} \to x \quad \text{and} \quad b_{p_{k_n}} \to Lx. \tag{4.20}$$

Hence, using the fact that $\nabla h(x_{p_{k_n}}) \to \nabla h(x)$ and $\nabla j(Lx_{p_{k_n}}) \to \nabla j(Lx)$, we derive that $\nabla h(x_{p_{k_n}}) - \nabla h(a_{p_{k_n}}) \to 0$ and $\nabla j(Lx_{p_{k_n}}) - \nabla j(b_{p_{k_n}}) \to 0$, which, in view of (3.2), yields

$$a^*_{p_{k_n}} + L^* y^*_{p_{k_n}} \to 0 \quad \text{and} \quad b^*_{p_{k_n}} - y^*_{p_{k_n}} \to 0. \tag{4.21}$$

Thus, since $y^*_{p_{k_n}} \to y^*$, it follows that $a^*_{p_{k_n}} \to -L^* y^*$ and $b^*_{p_{k_n}} \to y^*$. In summary,

$$\operatorname{gra} A \ni (a_{p_{k_n}}, a^*_{p_{k_n}}) \to (x, -L^* y^*) \quad \text{and} \quad \operatorname{gra} B \ni (b_{p_{k_n}}, b^*_{p_{k_n}}) \to (Lx, y^*). \tag{4.22}$$

Since $\operatorname{gra} A$ and $\operatorname{gra} B$ are closed [9, Proposition 20.33(iii)], we conclude that $(x, -L^* y^*) \in \operatorname{gra} A$ and $(Lx, y^*) \in \operatorname{gra} B$, and therefore that $(x, y^*) \in \boldsymbol{Z}$. ☐

Let us note that, even in Euclidean spaces, it may be easier to evaluate $(\nabla h + \gamma \partial \varphi)^{-1}$ than the usual proximity operator $\operatorname{prox}_{\gamma\varphi} = (\operatorname{Id} + \gamma\partial\varphi)^{-1}$ introduced by Moreau [31], which is based on $h = \|\cdot\|^2/2$. We provide illustrations of such instances in the standard Euclidean space $\mathbb{R}^m$.

Example 4.3 Let $\gamma \in \left]0, +\infty\right[$, let $\phi \in \Gamma_0(\mathbb{R})$ be such that $\operatorname{dom}\phi \cap \left]0, +\infty\right[\neq \varnothing$, and let ϑ be the Boltzmann-Shannon entropy function, i.e.,

$$\vartheta \colon \xi \mapsto \begin{cases} \xi \ln \xi - \xi, & \text{if } \xi \in \left]0, +\infty\right[; \\ 0, & \text{if } \xi = 0; \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.23}$$

Set $\varphi\colon (\xi_i)_{1\leqslant i\leqslant m} \mapsto \sum_{i=1}^m \phi(\xi_i)$ and $h\colon (\xi_i)_{1\leqslant i\leqslant m} \mapsto \sum_{i=1}^m \vartheta(\xi_i)$. Note that h is a supercoercive Legendre function [5, Sections 5 and 6] and hence Proposition 2.11(iv) asserts that $\nabla h+\gamma\partial\varphi$ is coercive and $\operatorname{dom}(\nabla h+\gamma\partial\varphi)^{-1} = \mathbb{R}^m$. Now let $(\xi_i)_{1\leqslant i\leqslant m} \in \mathbb{R}^m$, set $(\eta_i)_{1\leqslant i\leqslant m} = (\nabla h+\gamma\partial\varphi)^{-1}(\xi_i)_{1\leqslant i\leqslant m}$, let W be the Lambert function [21, 26], i.e., the inverse of $\xi \mapsto \xi e^{\xi}$ on $[0,+\infty[$, and let $i \in \{1,\ldots,m\}$. Then η_i can be computed as follows.

(i) Let $\omega \in \mathbb{R}$ and suppose that

$$\phi\colon \xi \mapsto \begin{cases} \xi\ln\xi - \omega\xi, & \text{if } \ \xi \in \left]0,+\infty\right[; \\ 0, & \text{if } \ \xi = 0; \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.24}$$

Then $\eta_i = e^{(\xi_i+\gamma(\omega-1))/(\gamma+1)}$.

(ii) Let $p \in \left[1,+\infty\right[$ and suppose that either $\phi = |\cdot|^p/p$ or

$$\phi\colon \xi \mapsto \begin{cases} \xi^p/p, & \text{if } \ \xi \in \left[0,+\infty\right[; \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.25}$$

Then

$$\eta_i = \begin{cases} \left(\dfrac{W(\gamma(p-1)e^{(p-1)\xi_i})}{\gamma(p-1)}\right)^{\frac{1}{p-1}}, & \text{if } \ p > 1; \\ e^{\xi_i-\gamma}, & \text{if } \ p = 1. \end{cases} \tag{4.26}$$

(iii) Let $p \in \left[1,+\infty\right[$ and suppose that

$$\phi\colon \xi \mapsto \begin{cases} \xi^{-p}/p, & \text{if } \ \xi \in \left]0,+\infty\right[; \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.27}$$

Then

$$\eta_i = \left(\frac{W(\gamma(p+1)e^{-(p+1)\xi_i})}{\gamma(p+1)}\right)^{\frac{-1}{p+1}}. \tag{4.28}$$

(iv) Let $p \in \left]0,1\right[$ and suppose that

$$\phi\colon \xi \mapsto \begin{cases} -\xi^p/p, & \text{if } \ \xi \in \left[0,+\infty\right[; \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.29}$$

Then

$$\eta_i = \left(\frac{W(\gamma(1-p)e^{(p-1)\xi_i})}{\gamma(1-p)}\right)^{\frac{1}{p-1}}. \tag{4.30}$$

Example 4.4 Let $\phi \in \Gamma_0(\mathbb{R})$ be such that $\operatorname{dom}\phi \cap \left]0,1\right[\neq \varnothing$ and let ϑ be the Fermi-Dirac entropy, i.e.,

$$\vartheta\colon \xi \mapsto \begin{cases} \xi\ln\xi + (1-\xi)\ln(1-\xi), & \text{if } \ \xi \in \left]0,1\right[\,; \\ 0 & \text{if } \ \xi \in \{0,1\}; \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.31}$$

Set $\varphi\colon (\xi_i)_{1\leqslant i\leqslant m} \mapsto \sum_{i=1}^m \phi(\xi_i)$ and $h\colon (\xi_i)_{1\leqslant i\leqslant m} \mapsto \sum_{i=1}^m \vartheta(\xi_i)$. Note that h is a Legendre function [5, Sections 5 and 6] and that $\operatorname{int}\operatorname{dom} h = \left]0,1\right[^m$ is bounded. Therefore, Proposition 2.11(i) asserts that $\nabla h + \partial\varphi$ is coercive and that $\operatorname{dom}(\nabla h + \partial\varphi)^{-1} = \mathbb{R}^m$. Now let $(\xi_i)_{1\leqslant i\leqslant m} \in \mathbb{R}^m$, set $(\eta_i)_{1\leqslant i\leqslant m} = (\nabla h + \partial\varphi)^{-1}(\xi_i)_{1\leqslant i\leqslant m}$, and let $i \in \{1,\ldots,m\}$. Then η_i can be computed as follows.

(i) Let $\omega \in \mathbb{R}$ and suppose that

$$\phi\colon \xi \mapsto \begin{cases} \xi\ln\xi - \omega\xi, & \text{if } \ \xi \in \left]0,+\infty\right[\,; \\ 0, & \text{if } \ \xi = 0; \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.32}$$

Then $\eta_i = -e^{\xi_i+\omega-1}/2 + \sqrt{e^{2(\xi_i+\omega-1)}/4 + e^{\xi_i+\omega-1}}$.

(ii) Suppose that

$$\phi\colon \xi \mapsto \begin{cases} (1-\xi)\ln(1-\xi) + \xi, & \text{if } \ \xi \in \left]-\infty,1\right[\,; \\ 1, & \text{if } \ \xi = 1; \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.33}$$

Then $\eta_i = 1 + e^{-\xi_i}/2 - \sqrt{e^{-\xi_i} + e^{-2\xi_i}/4}$.

References

[1] M. A. Alghamdi, A. Alotaibi, P. L. Combettes, and N. Shahzad, A primal-dual method of partial inverses for composite inclusions, *Optim. Lett.,* vol. 8, pp. 2271–2284, 2014.

[2] A. Alotaibi, P. L. Combettes, and N. Shahzad, Solving coupled composite monotone inclusions by successive Fejér approximations of their Kuhn-Tucker set, *SIAM J. Optim.,* vol. 24, pp. 2076–2095, 2014.

[3] A. Alotaibi, P. L. Combettes, and N. Shahzad, Best approximation from the Kuhn-Tucker set of composite monotone inclusions, *Numer. Funct. Anal. Optim.,* vol. 36, pp. 1513–1532, 2015.

[4] H. Attouch, L. M. Briceño-Arias, and P. L. Combettes, A strongly convergent primal-dual method for nonoverlapping domain decomposition, *Numer. Math.,* published online 2015-07-10.

[5] 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.

[6] 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.

[7] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, Bregman monotone optimization algorithms, *SIAM J. Control Optim.,* vol. 42, pp. 596–636, 2003.

[8] H. H. Bauschke and P. L. Combettes, Construction of best Bregman approximations in reflexive Banach spaces, *Proc. Amer. Math. Soc.,* vol. 131, pp. 3757–3766, 2003.

[9] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* Springer, New York, 2011.

[10] J. M. Borwein, Fifty years of maximal monotonicity, *Optim. Lett.,* vol. 4, pp. 473–490, 2010.

[11] R. I. Boţ, E. R. Csetnek, and A. Heinrich, A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators, *SIAM J. Optim.,* vol. 23, pp. 2011–2036, 2013.

[12] R. I. Boţ and C. Hendrich, A Douglas-Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators, *SIAM J. Optim.,* vol. 23, pp. 2541–2565, 2013.

[13] L. M. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, *USSR Comput. Math. Math. Phys.*, vol 7, pp. 200–217, 1967.

[14] L. M. Briceño-Arias and P. L. Combettes, A monotone+skew splitting model for composite monotone inclusions in duality, *SIAM J. Optim.*, vol. 21, pp. 1230–1250, 2011.

[15] F. E. Browder, Multi-valued monotone nonlinear mappings and duality mappings in Banach spaces, *Trans. Amer. Math. Soc.,* vol. 118, pp. 338–351, 1965.

[16] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithms and Applications.* Oxford University Press, New York, 1997.

[17] I. Cioranescu, *Geometry of Banach spaces, Duality Mappings and Nonlinear Problems.* Kluwer, Dordrecht, 1990.

[18] P. L. Combettes, Strong convergence of block-iterative outer approximation methods for convex optimization, *SIAM J. Control Optim.*, vol. 38, pp. 538–565, 2000.

[19] P. L. Combettes, Systems of structured monotone inclusions: Duality, algorithms, and applications, *SIAM J. Optim.,* vol. 23, pp. 2420–2447, 2013.

[20] 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.

[21] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, On the Lambert W function, *Adv. Comput. Math.,* vol. 5, pp. 329–359, 1996.

[22] J. Eckstein, Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming, *Math. Oper. Res.*, vol. 18, pp. 202–226, 1993.

[23] Y. Haugazeau, *Sur les Inéquations Variationnelles et la Minimisation de Fonctionnelles Convexes.* Thèse, Université de Paris, Paris, France, 1968.

[24] R. I. Kačurovskiĭ, On monotone operators and convex functionals, *Uspekhi Mat. Nauk*, vol. 15, pp. 213–215, 1960.

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

[26] J. H. Lamberti, Observationes variae in mathesin puram, *Acta Helvetica, Physico-Mathematico-Anatomico-Botanico-Medica,* vol. 3, pp. 128–168, 1758.

[27] P. L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.*, vol. 16, pp. 964–979, 1979.

[28] B. Mercier, *Topics in Finite Element Solution of Elliptic Problems* (Lectures on Mathematics, no. 63). Tata Institute of Fundamental Research, Bombay, 1979.

[29] G. J. Minty, On the maximal domain of a "monotone" function, *Michigan Math. J.*, vol. 8, pp. 135–137, 1961.

[30] G. J. Minty, Monotone (nonlinear) operators in Hilbert space, *Duke Math. J.*, vol. 29, pp. 341–346, 1962.

[31] J. J. Moreau, Fonctions convexes duales et points proximaux dans un espace hilbertien, *C. R. Acad. Sci. Paris Sér. A Math.,* vol. 255, pp. 2897–2899, 1962.

[32] J. J. Moreau, Fonctionnelles sous-différentiables, *C. R. Acad. Sci. Paris Sér. A Math.,* vol. 257, pp. 4117–4119, 1963.

[33] T. Pennanen, Dualization of generalized equations of maximal monotone type, *SIAM J. Optim.,* vol. 10, pp. 809–835, 2000.

[34] R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.,* vol. 14, pp. 877–898, 1976.

[35] S. Simons, *From Hahn-Banach to Monotonicity,* Lecture Notes in Math. 1693, Springer-Verlag, New York, 2008.

[36] M. Teboulle, Entropic proximal mappings with applications to nonlinear programming, *Math. Oper. Res.*, vol. 17, pp. 670–690, 1992.

[37] P. Tseng, A modified forward-backward splitting method for maximal monotone mappings, *SIAM J. Control Optim.*, vol. 38, pp. 431–446, 2000.

[38] B. C. Vũ, A splitting algorithm for dual monotone inclusions involving cocoercive operators, *Adv. Comput. Math.*, vol. 38, pp. 667–681, 2013.

[39] C. Zălinescu, On uniformly convex functions, *J. Math. Anal. Appl.,* vol. 95, pp. 344–374, 1983.

[40] C. Zălinescu, *Convex Analysis in General Vector Spaces.* World Scientific Publishing, River Edge, NJ, 2002.

[41] E. H. Zarantonello, *Solving functional equations by contractive averaging*, Tech. Report 160, Math. Res. Center U.S. Army, Madison, University of Wisconsin, June 1960.

[42] E. Zeidler, *Nonlinear Functional Analysis and Its Applications,* vols. I–V. Springer-Verlag, New York, 1985–1993.