Vietnam J. Math. (2017) 45:519–539
DOI 10.1007/s10013-016-0238-3

Forward-Backward Splitting with Bregman Distances

Quang Van Nguyen[1,2]

Received: 28 May 2015 / Accepted: 13 October 2016 / Published online: 28 January 2017

Abstract We propose a forward-backward splitting algorithm based on Bregman distances for composite minimization problems in general reflexive Banach spaces. The convergence is established using the notion of variable quasi-Bregman monotone sequences. Various examples are discussed, including some in Euclidean spaces, where new algorithms are obtained.

Keywords Banach space · Bregman distance · Forward-backward algorithm · Legendre function · Multivariate minimization · Variable quasi-Bregman monotonicity

Mathematics Subject Classification (2010) 90C25

1 Introduction

In this paper, we propose a forward-backward splitting algorithm to solve the following composite convex minimization problem considered in Banach spaces.

Problem 1 Let $\mathcal{X}$ be a reflexive real Banach space, let $\varphi\colon \mathcal{X} \to]-\infty, +\infty]$ and $\psi\colon \mathcal{X} \to]-\infty, +\infty]$ be proper lower semi-continuous convex functions, and suppose that ψ is Gâteaux differentiable on interior of its domain. The problem is to

$$\underset{x \in \mathcal{X}}{\text{minimize}}\ \varphi(x) + \psi(x). \tag{1}$$

✉ Quang Van Nguyen
quangnv@hnue.edu.vn; quang.nguyen@epfl.ch

1 Department of Mathematics, Hanoi National University of Education, 136 Xuan Thuy Street, Cau Giay dist., Hanoi, Vietnam

2 Laboratory for Information Theory and Inference Systems (LIONS), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

The set of solutions to (1) is denoted by $\mathcal{S}$.

A particular instance of (1) when ψ is the Bregman distance associated to a differentiable convex function f, i.e.,

$$D^f : \mathcal{X} \times \mathcal{X} \to [0, +\infty]$$
$$(x, y) \mapsto \begin{cases} f(x) - f(y) - \langle x - y, \nabla f(y) \rangle & \text{if } y \in \operatorname{int} \operatorname{dom} f, \\ +\infty & \text{otherwise,} \end{cases} \tag{2}$$

where $\operatorname{dom} f = \{x \in \mathcal{X} \mid f(x) < +\infty\}$ and $\operatorname{int} \operatorname{dom} f$ is its interior, provides a framework for many problems arising in applied mathematics. For instance, when $\mathcal{X}$ is a Euclidean space and f is Boltzmann–Shannon entropy, it captures many problems in information theory and signal recovery [9].

It was shown in [14] that if $\mathcal{X}$ is Hilbertian and ψ possesses a β^{-1}-Lipschitz continuous gradient for some $\beta \in]0, +\infty[$, then Problem 1 can be solved by the standard forward-backward algorithm

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{prox}_{\gamma_n \varphi}\big(x_n - \gamma \nabla \psi(x_n)\big), \qquad \text{where } 0 < \gamma < 2\beta. \tag{3}$$

Here, prox is Moreau proximity operator [19]. However, many problems in applications do not conform to these hypotheses, for example when $\mathcal{X}$ is a Euclidean space and ψ is Boltzmann–Shannon entropy which appears in many problems in image and signal processing, in statistics, and in machine learning [2, 11, 12, 16–18]. Another difficulty in the implementation of (3) is that the operator prox is not always easy to evaluate.

The objective of the present paper is to propose a forward-backward splitting algorithm to solve Problem 1, which is so far limited to Hilbert spaces, in the general framework of reflexive real Banach spaces. This algorithm, which employs Bregman distance-based proximity operators, provides new algorithms in the framework of Euclidean spaces, which are, in some instances, more favorable than the standard forward-backward splitting algorithm. This framework can be applied in the case when ψ is not everywhere differentiable. The paper is organized as follows. In Section 2, we provide some preliminary results. We present the algorithm and prove its convergence in Section 3. Section 4 is devoted to an application of our result to multivariate minimization problem together with examples.

Notation and Background Throughout this paper, $\mathcal{X}$ is reflexive, $\mathcal{X}^*$ is the dual space of $\mathcal{X}$, $\langle \cdot, \cdot \rangle$ is the duality pairing between $\mathcal{X}$ and $\mathcal{X}^*$ and $\| \cdot \|$ is a norm of $\mathcal{X}$. The symbols $\rightharpoonup$ and $\to$ represent respectively weak and strong convergence. The set of weak sequential cluster points of a sequence $(x_n)_{n\in\mathbb{N}}$ is denoted by $\mathfrak{W}(x_n)_{n\in\mathbb{N}}$. Let $M : \mathcal{X} \to 2^{\mathcal{X}^*}$. The domain of M is $\operatorname{dom} M = \{x \in \mathcal{X} \mid Mx \neq \varnothing\}$ and the range of M is $\operatorname{ran} M = \{x^* \in \mathcal{X}^* \mid (\exists x \in \mathcal{X}) x^* \in Mx\}$. Let $f : \mathcal{X} \to]-\infty, +\infty]$. Then, f is cofinite if $\operatorname{dom} f^* = \mathcal{X}^*$, is coercive if $\lim_{\|x\| \to +\infty} f(x) = +\infty$, is supercoercive if $\lim_{\|x\| \to +\infty} f(x)/\|x\| = +\infty$, and is uniformly convex at $x \in \operatorname{dom} f$ if there exists an increasing function $\phi : [0, +\infty[\to [0, +\infty]$ that vanishes only at 0 such that

$$(\forall y \in \operatorname{dom} f)(\forall \alpha \in]0, 1[) \qquad f(\alpha x + (1-\alpha) y) + \alpha(1-\alpha)\phi(\|x - y\|) \\ \leq \alpha f(x) + (1-\alpha) f(y).$$

Denote by $\Gamma_0(\mathcal{X})$ the class of all lower semicontinuous convex functions $f : \mathcal{X} \to]-\infty, +\infty]$ such that $\operatorname{dom} f = \{x \in \mathcal{X} \mid f(x) < +\infty\} \neq \varnothing$. Let $f \in$

$\Gamma_0(\mathcal{X})$. Denote by $\operatorname{Argmin} f$ the set of global minimizers of f, by $f^*\colon \mathcal{X}^* \to]-\infty, +\infty]\colon x^* \mapsto \sup_{x\in\mathcal{X}}(\langle x, x^*\rangle - f(x))$ the conjugate of f and by

$$\partial f\colon \mathcal{X} \to 2^{\mathcal{X}^*}\colon x \mapsto \{x^* \in \mathcal{X}^* \mid (\forall y \in \mathcal{X})\langle y - x, x^*\rangle + f(x) \le f(y)\}, \tag{4}$$

the Moreau subdifferential of f. In addition, if f is Gâteaux differentiable on $\operatorname{int}\operatorname{dom} f \neq \emptyset$ then

$$\hat{f}\colon \mathcal{X} \to]-\infty, +\infty]\ x \mapsto \begin{cases} f(x) & \text{if } x \in \operatorname{int}\operatorname{dom} f, \\ +\infty & \text{otherwise.} \end{cases} \tag{5}$$

We denote

$$\mathcal{F}(f) = \{g \in \Gamma_0(\mathcal{X}) \mid g \text{ is Gâteaux differentiable on } \operatorname{dom} g = \operatorname{int}\operatorname{dom} f\}.$$

Moreover, if g_1 and g_2 are in $\mathcal{F}(f)$, then

$$g_1 \succcurlyeq g_2 \quad \Leftrightarrow \quad (\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{int}\operatorname{dom} f) \quad D^{g_1}(x, y) \ge D^{g_2}(x, y).$$

For every $\alpha \in [0, +\infty[$, set

$$\mathcal{P}_\alpha(f) = \{g \in \mathcal{F}(f) \mid g \succcurlyeq \alpha f\}.$$

Finally, $\ell^1_+(\mathbb{N})$ is the set of all summable sequences in $[0, +\infty[$.

2 Preliminary Results

In this section, we give some preliminary results on Legendre function, Bregman monotonicity, and Bregman distance-based proximity operator that will be used in the next section.

Definition 1 [5, 6] Let $f \in \Gamma_0(\mathcal{X})$ be Gâteaux differentiable on $\operatorname{int}\operatorname{dom} f \neq \emptyset$. We say that 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 $\operatorname{dom}\partial f$. Let C be a closed convex subset of $\mathcal{X}$ such that $C \cap \operatorname{int}\operatorname{dom} f \neq \emptyset$. The *Bregman projector* onto C induced by f is

$$\begin{aligned} P^f_C\colon \operatorname{int}\operatorname{dom} f &\to C \cap \operatorname{int}\operatorname{dom} f \\ y &\mapsto \operatorname{argmin}_{x\in C} D^f(x, y), \end{aligned}$$

and the D^f*-distance* to C is the function

$$\begin{aligned} D^f_C\colon \mathcal{X} &\to [0, +\infty] \\ y &\mapsto \inf D^f(C, y). \end{aligned}$$

Definition 2 [20] Let $f \in \Gamma_0(\mathcal{X})$ be Gâteaux differentiable on $\operatorname{int}\operatorname{dom} f \neq \emptyset$, let $(f_n)_{n\in\mathbb{N}}$ be in $\mathcal{F}(f)$, let $(x_n)_{n\in\mathbb{N}} \in (\operatorname{int}\operatorname{dom} f)^{\mathbb{N}}$, and let $C \subset \mathcal{X}$ be such that $C \cap \operatorname{dom} f \neq \emptyset$. Then $(x_n)_{n\in\mathbb{N}}$ is:

1. *quasi-Bregman monotone* with respect to C relative to $(f_n)_{n\in\mathbb{N}}$ if

$$\begin{aligned} &(\exists(\eta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N}))(\forall x \in C \cap \operatorname{dom} f)(\exists(\varepsilon_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N}))(\forall n \in \mathbb{N}) \\ &\qquad D^{f_{n+1}}(x, x_{n+1}) \le (1 + \eta_n) D^{f_n}(x, x_n) + \varepsilon_n; \end{aligned}$$

2. *stationarily quasi-Bregman monotone* with respect to C relative to $(f_n)_{n\in\mathbb{N}}$ if

$$(\exists(\varepsilon_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N}))(\exists(\eta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N}))(\forall x \in C \cap \operatorname{dom} f)(\forall n \in \mathbb{N})$$
$$D^{f_{n+1}}(x, x_{n+1}) \le (1+\eta_n)D^{f_n}(x, x_n) + \varepsilon_n.$$

Condition 1 [6, Condition 4.4] Let $f \in \Gamma_0(\mathcal{X})$ be Gâteaux differentiable on $\operatorname{int}\operatorname{dom} f \neq \emptyset$. For every bounded sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ in $\operatorname{int}\operatorname{dom} f$,

$$D^f(x_n, y_n) \to 0 \quad \Rightarrow \quad x_n - y_n \to 0.$$

Proposition 1 ([20]) *Let $f \in \Gamma_0(\mathcal{X})$ be Gâteaux differentiable on* $\operatorname{int}\operatorname{dom} f \neq \emptyset$*, let $\alpha \in]0, +\infty[$, let $(f_n)_{n\in\mathbb{N}}$ be in $\mathcal{P}_\alpha(f)$, let $(x_n)_{n\in\mathbb{N}} \in (\operatorname{int}\operatorname{dom} f)^{\mathbb{N}}$, let $C \subset \mathcal{X}$ be such that $C \cap \operatorname{int}\operatorname{dom} f \neq \emptyset$, and let $x \in C \cap \operatorname{int}\operatorname{dom} f$. Suppose that $(x_n)_{n\in\mathbb{N}}$ is quasi-Bregman monotone with respect to C relative to $(f_n)_{n\in\mathbb{N}}$. Then the following hold.*

1. *$(D^{f_n}(x, x_n))_{n\in\mathbb{N}}$ converges.*
2. *Suppose that $D^f(x, \cdot)$ is coercive. Then $(x_n)_{n\in\mathbb{N}}$ is bounded.*

Proposition 2 ([20]) *Let $f \in \Gamma_0(\mathcal{X})$ be Gâteaux differentiable on* $\operatorname{int}\operatorname{dom} f \neq \emptyset$*, let $(x_n)_{n\in\mathbb{N}} \in (\operatorname{int}\operatorname{dom} f)^{\mathbb{N}}$, let $C \subset \mathcal{X}$ be such that $C \cap \operatorname{int}\operatorname{dom} f \neq \emptyset$, let $(\eta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$, let $\alpha \in]0, +\infty[$, and let $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{P}_\alpha(f)$ be such that $(\forall n \in \mathbb{N})\ (1+\eta_n)f_n \succcurlyeq f_{n+1}$. Suppose that $(x_n)_{n\in\mathbb{N}}$ is quasi-Bregman monotone with respect to C relative to $(f_n)_{n\in\mathbb{N}}$, that there exists $g \in \mathcal{F}(f)$ such that for every $n \in \mathbb{N}$, $g \succcurlyeq f_n$, and that, for every $y_1 \in \mathcal{X}$ and every $y_2 \in \mathcal{X}$,*

$$\begin{cases} y_1 \in \mathfrak{W}(x_n)_{n\in\mathbb{N}} \cap C, \\ y_2 \in \mathfrak{W}(x_n)_{n\in\mathbb{N}} \cap C, \\ \big(\langle y_1 - y_2, \nabla f_n(x_n)\rangle\big)_{n\in\mathbb{N}} \quad \text{converges} \end{cases} \quad \Rightarrow \quad y_1 = y_2.$$

Moreover, suppose that $(\forall x \in \operatorname{int}\operatorname{dom} f)\ D^f(x, \cdot)$ is coercive. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $C \cap \operatorname{int}\operatorname{dom} f$ if and only if $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset C \cap \operatorname{int}\operatorname{dom} f$.

Proposition 3 ([20]) *Let $f \in \Gamma_0(\mathcal{X})$ be a Legendre function, let $\alpha \in]0, +\infty[$, let $(f_n)_{n\in\mathbb{N}}$ be in $\mathcal{P}_\alpha(f)$, let $(x_n)_{n\in\mathbb{N}} \in (\operatorname{int}\operatorname{dom} f)^{\mathbb{N}}$, and let C be a closed convex subset of $\mathcal{X}$ such that $C \cap \operatorname{int}\operatorname{dom} f \neq \emptyset$. Suppose that $(x_n)_{n\in\mathbb{N}}$ is stationarily quasi-Bregman monotone with respect to C relative to $(f_n)_{n\in\mathbb{N}}$, that f satisfies Condition 1, and that $(\forall x \in \operatorname{int}\operatorname{dom} f)$ $D^f(x, \cdot)$ is coercive. In addition, suppose that there exists $\beta \in]0, +\infty[$ such that $(\forall n \in \mathbb{N})\ \beta \hat{f} \succcurlyeq f_n$. Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to a point in $C \cap \overline{\operatorname{dom}} f$ if and only if $\underline{\lim} D^f_C(x_n) = 0$.*

Our framework uses the Bregman distance-based proximity operators whose definition and properties are discussed in the following proposition.

Proposition 4 *Let $f \in \Gamma_0(\mathcal{X})$ be Gâteaux differentiable on* $\operatorname{int}\operatorname{dom} f \neq \emptyset$*, let $\varphi \in \Gamma_0(\mathcal{X})$, and let*

$$\begin{aligned} \operatorname{Prox}^f_\varphi \colon \mathcal{X}^* &\to 2^{\mathcal{X}} \\ x^* &\mapsto \{x \in \mathcal{X} \mid \varphi(x) + f(x) - \langle x, x^*\rangle = \min(\varphi + f - x^*)(\mathcal{X}) < +\infty\} \end{aligned} \quad (6)$$

be f-proximity operator of φ. Then the following hold.

(1) $\operatorname{ran}\operatorname{Prox}_\varphi^f \subset \operatorname{dom} f \cap \operatorname{dom}\varphi$ *and* $\operatorname{Prox}_\varphi^f = (\partial(f+\varphi))^{-1}$.

(2) *Suppose that* $\operatorname{dom}\varphi \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$ *and that* $\operatorname{dom}\partial f \cap \operatorname{dom}\partial\varphi \subset \operatorname{int}\operatorname{dom} f$. *Then the following hold.*

  (a) $\operatorname{ran}\operatorname{Prox}_\varphi^f \subset \operatorname{int}\operatorname{dom} f$ *and* $\operatorname{Prox}_\varphi^f = (\nabla f + \partial\varphi)^{-1}$.

  (b) $\operatorname{int}(\operatorname{dom} f^* + \operatorname{dom}\varphi^*) \subset \operatorname{dom}\operatorname{Prox}_\varphi^f$.

  (c) *Suppose that* $f|_{\operatorname{int}\operatorname{dom} f}$ *is strictly convex. Then* $\operatorname{Prox}_\varphi^f$ *is single-valued on its domain.*

Proof Let us fix $x^* \in \mathcal{X}^*$ and define $f_{x^*}\colon \mathcal{X} \to]-\infty,+\infty]\colon x \mapsto f(x) - \langle x, x^*\rangle + f^*(x^*)$. Then $\operatorname{dom} f_{x^*} = \operatorname{dom} f$ and $\varphi + f_{x^*} \in \Gamma_0(\mathcal{X})$. Moreover, $\partial(\varphi + f_{x^*}) = \partial(\varphi + f) - x^*$.

(1): By definition, $\operatorname{ran}\operatorname{Prox}_\varphi^f \subset \operatorname{dom} f \cap \operatorname{dom}\varphi$. For the second assertion, it is sufficient to prove for the case $\operatorname{dom} f \cap \operatorname{dom}\varphi \neq \varnothing$ since otherwise both sides of the desired identity reduce to the trivial operator $x^* \mapsto \varnothing$. Now let $x \in \operatorname{dom} f \cap \operatorname{dom}\varphi$. Then

$$\begin{aligned} x \in \operatorname{Prox}_\varphi^f x^* &\Leftrightarrow 0 \in \partial(\varphi + f_{x^*})(x) \\ &\Leftrightarrow 0 \in \partial(\varphi + f)(x) - x^* \\ &\Leftrightarrow x^* \in \partial(\varphi + f)(x) \\ &\Leftrightarrow x \in \big(\partial(\varphi + f)\big)^{-1}(x^*). \end{aligned} \tag{7}$$

(2): Suppose that $x^* \in \operatorname{int}(\operatorname{dom} f^* + \operatorname{dom}\varphi^*)$. Since $\operatorname{dom}\varphi \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$, it follows from [1, Theorem 1.1] and [23, Theorem 2.1.3(ix)] that

$$x^* \in \operatorname{int}(\operatorname{dom} f^* + \operatorname{dom}\varphi^*) = \operatorname{int}\operatorname{dom}(f+\varphi)^*. \tag{8}$$

(2a): Since $\operatorname{dom}\varphi \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$, $\partial(\varphi + f) = \partial\varphi + \partial f$ by [1, Corollary 2.1], and hence 1) yields

$$\operatorname{ran}\operatorname{Prox}_\varphi^f = \operatorname{dom}\partial(f+\varphi) = \operatorname{dom}(\partial f + \partial\varphi) = \operatorname{dom}\partial f \cap \operatorname{dom}\partial\varphi \subset \operatorname{int}\operatorname{dom} f.$$

In turn, $\operatorname{ran}\operatorname{Prox}_\varphi^f \subset \operatorname{dom}\varphi \cap \operatorname{int}\operatorname{dom} f$. We now prove that $\operatorname{Prox}_\varphi^f = (\nabla f + \partial\varphi)^{-1}$. Note that $\operatorname{dom}(\nabla f + \partial\varphi) \subset \operatorname{dom}\varphi \cap \operatorname{int}\operatorname{dom} f$. Let $x \in \operatorname{dom}\varphi \cap \operatorname{int}\operatorname{dom} f$. Then $\partial(f+\varphi)(x) = \partial f(x) + \partial\varphi(x) = \nabla f(x) + \partial\varphi(x)$ and therefore,

$$x \in \operatorname{Prox}_\varphi^f x^* \Leftrightarrow x^* \in \partial(f+\varphi)(x) = \nabla f(x) + \partial\varphi(x) \Leftrightarrow x \in (\nabla f + \partial\varphi)^{-1}(x^*).$$

(2b): We derive from (8) and [5, Fact 3.1] that $\varphi + f_{x^*}$ is coercive. Hence, by [23, Theorem 2.5.1], $\varphi + f_{x^*}$ admits at least one minimizer, i.e., $x^* \in \operatorname{dom}\operatorname{Prox}_\varphi^f$.

(2c): Since $f|_{\operatorname{int}\operatorname{dom} f}$ is strictly convex, so is $(\varphi + f_{x^*})|_{\operatorname{int}\operatorname{dom} f}$ and thus, in view of 2b), $\varphi + f_{x^*}$ admits a unique minimizer on $\operatorname{int}\operatorname{dom} f$. However, since

$$\operatorname{Argmin}(\varphi + f_{x^*}) = \operatorname{ran}\operatorname{Prox}_\varphi^f \subset \operatorname{int}\operatorname{dom} f,$$

it follows that $\varphi + f_{x^*}$ admits a unique minimizer and that $\operatorname{Prox}_\varphi^f$ is therefore single-valued. □

Proposition 5 *Let m be a strictly positive integer, let $(\mathcal{X}_i)_{1\le i\le m}$ be reflexive real Banach spaces, and let $\mathcal{X}$ be the vector product space $\times_{i=1}^m \mathcal{X}_i$ equipped with the norm $x =$*

$(x_i)_{1\leq i\leq m} \mapsto \sqrt{\sum_{i=1}^m \|x_i\|^2}$. *For every* $i \in \{1,\ldots,m\}$, *let* $f_i \in \Gamma_0(\mathcal{X}_i)$ *be a Legendre function and let* $\varphi_i \in \Gamma_0(\mathcal{X}_i)$ *be such that* $\operatorname{dom}\varphi_i \cap \operatorname{int}\operatorname{dom} f_i \neq \emptyset$. *Set* $f\colon \mathcal{X} \to]-\infty,+\infty]\colon x \mapsto \sum_{i=1}^m f_i(x_i)$ *and* $\varphi\colon \mathcal{X} \to]-\infty,+\infty]\colon x \mapsto \sum_{i=1}^m \varphi_i(x_i)$. *Then*

$$\left(\forall x^* = (x_i^*)_{1\leq i\leq m} \in \bigtimes_{i=1}^m \operatorname{int}(\operatorname{dom} f_i^* + \operatorname{dom}\varphi_i^*)\right) \quad \operatorname{Prox}_\varphi^f x^* = \left(\operatorname{Prox}_{\varphi_i}^{f_i} x_i^*\right)_{1\leq i\leq m}.$$

Proof First, we observe that $\mathcal{X}^*$ is the vector product space $\bigtimes_{i=1}^m \mathcal{X}_i^*$ equipped with the norm $x^* = (x_i^*)_{1\leq i\leq m} \mapsto \sqrt{\sum_{i=1}^m \|x_i^*\|^2}$. Next, we derive from the definition of f that $\operatorname{dom} f = \bigtimes_{i=1}^m \operatorname{dom} f_i$ and that

$$\partial f\colon \mathcal{X} \to 2^{\mathcal{X}^*}\colon (x_i)_{1\leq i\leq m} \mapsto \bigtimes_{i=1}^m \partial f_i(x_i).$$

Thus, ∂f is single-valued on

$$\operatorname{dom}\partial f = \bigtimes_{i=1}^m \operatorname{dom}\partial f_i = \bigtimes_{i=1}^m \operatorname{int}\operatorname{dom} f_i = \operatorname{int}\left(\bigtimes_{i=1}^m \operatorname{dom} f_i\right) = \operatorname{int}\operatorname{dom} f.$$

Likewise, since

$$f^*\colon \mathcal{X}^* \to]-\infty,+\infty]\colon (x_i^*)_{1\leq i\leq m} \mapsto \sum_{i=1}^m f_i^*(x_i^*),$$

we deduce that ∂f^* is single-valued on $\operatorname{dom}\partial f^* = \operatorname{int}\operatorname{dom} f^*$. Consequently, [5, Theorems 5.4 and 5.6] assert that

$$f \text{ is a Legendre function.} \tag{9}$$

In addition,

$$\begin{aligned}\operatorname{dom}\varphi \cap \operatorname{int}\operatorname{dom} f &= \left(\bigtimes_{i=1}^m \operatorname{dom}\varphi_i\right) \cap \left(\bigtimes_{i=1}^m \operatorname{int}\operatorname{dom} f_i\right)\\ &= \bigtimes_{i=1}^m (\operatorname{dom}\varphi_i \cap \operatorname{int}\operatorname{dom} f_i) \neq \emptyset.\end{aligned} \tag{10}$$

Hence, Proposition 4(2b) and (2c) assert that $\operatorname{int}(\operatorname{dom} f^* + \operatorname{dom}\varphi^*) \subset \operatorname{dom}\operatorname{Prox}_\varphi^f$ and $\operatorname{Prox}_\varphi^f$ is single-valued on its domain. Now set $x = \operatorname{Prox}_\varphi^f x^*$ and $q = (\operatorname{Prox}_{\varphi_i}^{f_i} x_i^*)_{1\leq i\leq m}$. We derive from Proposition 4(2a) that

$$x = \operatorname{Prox}_\varphi^f x^* \quad \Leftrightarrow \quad x = (\nabla f + \partial\varphi)^{-1}(x^*) \quad \Leftrightarrow \quad x^* - \nabla f(x) \in \partial\varphi(x).$$

Consequently, by invoking (4), we get

$$(\forall z \in \operatorname{dom}\varphi) \quad \langle z - x, x^* - \nabla f(x)\rangle + \varphi(x) \leq \varphi(z). \tag{11}$$

Upon setting $z = q$ in (11), we obtain

$$\langle q - x, x^* - \nabla f(x)\rangle + \varphi(x) \leq \varphi(q). \tag{12}$$

For every $i \in \{1,\ldots,m\}$, let us set $q_i = \operatorname{Prox}_{\varphi_i}^{f_i} x_i^*$. The same characterization as in (11) yields

$$(\forall i \in \{1,\ldots,m\})(\forall z_i \in \operatorname{dom}\varphi_i) \quad \langle z_i - q_i, x_i^* - \nabla f_i(q_i)\rangle + \varphi_i(q_i) \leq \varphi_i(z_i).$$

By summing these inequalities over $i \in \{1, \ldots, m\}$, we obtain

$$(\forall z \in \operatorname{dom} \varphi) \quad \langle z - q, x^* - \nabla f(q) \rangle + \varphi(q) \leq \varphi(z). \tag{13}$$

Upon setting $z = x$ in (13), we get

$$\langle x - q, \nabla f(x) - \nabla f(q) \rangle + \varphi(q) \leq \varphi(x). \tag{14}$$

Adding (12) and (14) yields

$$\langle x - q, \nabla f(x) - \nabla f(q) \rangle \leq 0.$$

Now suppose that $x \neq q$. Since $f|_{\operatorname{int} \operatorname{dom} f}$ is strictly convex, it follows from [23, Theorem 2.4.4(ii)] that ∇f is strictly monotone, i.e.,

$$\langle x - q, \nabla f(x) - \nabla f(q) \rangle > 0,$$

and we reach a contradiction. □

In Hilbert spaces, the operator defined in (6) reduces to the Moreau's usual proximity operator $\operatorname{prox}_\varphi$ [19] if $f = \|\cdot\|^2/2$. We provide illustrations of such instances in the standard Euclidean space $\mathbb{R}^m$.

Example 1 Let $\gamma \in]0, +\infty[$, let $\phi \in \Gamma_0(\mathbb{R})$ be such that $\operatorname{dom} \phi \cap]0, +\infty[\neq \emptyset$, and let ϑ be Boltzmann–Shannon entropy, i.e.,

$$\vartheta : \xi \mapsto \begin{cases} \xi \ln \xi - \xi & \text{if } \xi \in]0, +\infty[, \\ 0 & \text{if } \xi = 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Set $\varphi \colon (\xi_i)_{1 \leq i \leq m} \mapsto \sum_{i=1}^m \phi(\xi_i)$ and $f \colon (\xi_i)_{1 \leq i \leq m} \mapsto \sum_{i=1}^m \vartheta(\xi_i)$. Note that f is a supercoercive Legendre function [4, Sections 5 and 6], and hence, Proposition 4(2b) asserts that $\operatorname{dom} \operatorname{Prox}_\varphi^f = \mathbb{R}^m$. Let $(\xi_i)_{1 \leq i \leq m} \in \mathbb{R}^m$, set $(\eta_i)_{1 \leq i \leq m} = \operatorname{Prox}_{\gamma\varphi}^f (\xi_i)_{1 \leq i \leq m}$, let W be the Lambert function [15], 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.

1. Let $\omega \in \mathbb{R}$ and suppose that

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

   Then $\eta_i = e^{(\xi_i + \omega - 1)/(\gamma + 1)}$.
2. Let $p \in [1, +\infty[$ and suppose that either $\phi = |\cdot|^p/p$ or

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

   Then

$$\eta_i = \begin{cases} \left(\dfrac{W(\gamma(p-1)e^{(p-1)\xi_i})}{\gamma(p-1)} \right)^{\frac{1}{p-1}} & \text{if } p \in]1, +\infty[, \\ e^{\xi_i - \gamma} & \text{if } p = 1. \end{cases}$$

3. Let $p \in [1, +\infty[$ and suppose that

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

Then

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

4. Let $p \in]0,1[$ and suppose that

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

Then

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

Example 2 Let $\phi \in \Gamma_0(\mathbb{R})$ be such that $\operatorname{dom}\phi \cap]0,1[\neq \emptyset$ and let ϑ be Fermi–Dirac entropy, i.e.,

$$\vartheta : \xi \mapsto \begin{cases} \xi \ln \xi - (1-\xi)\ln(1-\xi) & \text{if } \xi \in]0,1[, \\ 0 & \text{if } \xi \in \{0,1\}, \\ +\infty & \text{otherwise.} \end{cases}$$

Set $\varphi\colon (\xi_i)_{1\le i\le m} \mapsto \sum_{i=1}^m \phi(\xi_i)$ and $f\colon (\xi_i)_{1\le i\le m} \mapsto \sum_{i=1}^m \vartheta(\xi_i)$. Note that f is a cofinite Legendre function [4, Sections 5 and 6], and hence Proposition 4(2b) asserts that $\operatorname{dom}\operatorname{Prox}_\varphi^f = \mathbb{R}^m$. Let $(\xi_i)_{1\le i\le m} \in \mathbb{R}^m$, set $(\eta_i)_{1\le i\le m} = \operatorname{Prox}_\varphi^f(\xi_i)_{1\le i\le m}$, and let $i \in \{1,\ldots,m\}$. Then η_i can be computed as follows.

1. Let $\omega \in \mathbb{R}$ and suppose that

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

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

2. Suppose that

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

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

Example 3 Let $f\colon (\xi_i)_{1\le i\le m} \mapsto \sum_{i=1}^m \vartheta(\xi_i)$, where ϑ is Hellinger-like function, i.e.,

$$\vartheta : \xi \mapsto \begin{cases} -\sqrt{1-\xi^2} & \text{if } \xi \in [-1,1], \\ +\infty & \text{otherwise,} \end{cases}$$

let $\gamma \in]0,+\infty[$, and let $\varphi = f$. Since f is a cofinite Legendre function [4, Sections 5 and 6], Proposition 4(2b) asserts that $\operatorname{dom}\operatorname{Prox}_{\gamma\varphi}^f = \mathbb{R}^m$. Let $(\xi_i)_{1\le i\le m} \in \mathbb{R}^m$, and set $(\eta_i)_{1\le i\le m} = \operatorname{Prox}_{\gamma\varphi}^f(\xi_i)_{1\le i\le m}$. Then $(\forall i \in \{1,\ldots,m\})$ $\eta_i = \xi_i/\sqrt{(\gamma+1)^2 + \xi_i^2}$.

Example 4 Let $\gamma \in]0,+\infty[$, let $\phi \in \Gamma_0(\mathbb{R})$ be such that $\operatorname{dom}\phi \cap]0,+\infty[\neq \emptyset$, and let ϑ be Burg entropy, i.e.,

$$\vartheta : \xi \mapsto \begin{cases} -\ln \xi & \text{if } \xi \in]0,+\infty[, \\ +\infty & \text{otherwise.} \end{cases}$$

Set $\varphi\colon (\xi_i)_{1\leq i\leq m} \mapsto \sum_{i=1}^m \phi(\xi_i)$ and $f\colon (\xi_i)_{1\leq i\leq m} \mapsto \sum_{i=1}^m \vartheta(\xi_i)$, let $(\xi_i)_{1\leq i\leq m} \in \mathbb{R}^m$, and set $(\eta_i)_{1\leq i\leq m} = \operatorname{Prox}^f_{\gamma\varphi}(\xi_i)_{1\leq i\leq m}$. Let $i \in \{1,\ldots,m\}$. Then η_i can be computed as follows.

1. Suppose that $\phi = \vartheta$ and $\xi_i \in]-\infty, 0]$. Then $\eta_i = -(1+\gamma)^{-1}\xi_i$.
2. Suppose that $\phi\colon \xi \mapsto \alpha|\xi|$ and $\xi_i \in]-\infty, \gamma\alpha]$. Then $\eta_i = (\gamma\alpha - \xi_i)^{-1}$.

The following result will be used subsequently.

Lemma 1 *Let $\mathcal{X}$ be a reflexive real Banach space, let $f \in \Gamma_0(\mathcal{X})$ be a Legendre function, let $x \in \operatorname{int}\operatorname{dom} f$, and let $(x_n)_{n\in\mathbb{N}} \in (\operatorname{int}\operatorname{dom} f)^{\mathbb{N}}$. Suppose that $(D^f(x, x_n))_{n\in\mathbb{N}}$ is bounded, that* dom f^* *is open, and that ∇f^* is weakly sequentially continuous. Then $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \operatorname{int}\operatorname{dom} f$.*

Proof [20, Proof of Theorem 4.1]. □

3 Forward-Backward Splitting in Banach Spaces

The main result in this section is a version of the forward-backward splitting algorithm in reflexive real Banach spaces which employs different Bregman distance-based proximity operators over the iterations.

Theorem 1 *Consider the setting of Problem 1 and let $f \in \Gamma_0(\mathcal{X})$ be a Legendre function such that $\mathcal{S} \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$, $\operatorname{int}\operatorname{dom} f \subset \operatorname{int}\operatorname{dom}\psi$, and $f \succcurlyeq \beta\psi$ for some $\beta \in]0,+\infty[$. Let $(\eta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$, let $\alpha \in]0,+\infty[$, and let $(f_n)_{n\in\mathbb{N}}$ be Legendre functions in $\mathcal{P}_\alpha(f)$ such that*

$$(\forall n \in \mathbb{N}) \quad (1+\eta_n) f_n \succcurlyeq f_{n+1}. \tag{15}$$

Suppose that either $-\operatorname{ran}\nabla\psi \subset \operatorname{dom}\varphi^*$ or $(\forall n \in \mathbb{N})$ f_n is cofinite. Let $\varepsilon \in]0, \alpha\beta/(\alpha\beta+1)[$ and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ such that

$$(\forall n \in \mathbb{N}) \quad \varepsilon \leq \gamma_n \leq \alpha\beta(1-\varepsilon) \quad \text{and} \quad (1+\eta_n)\gamma_n - \gamma_{n+1} \leq \alpha\beta\eta_n. \tag{16}$$

Furthermore, let $x_0 \in \operatorname{int}\operatorname{dom} f$ and iterate

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{Prox}^{f_n}_{\gamma_n\varphi}\left(\nabla f_n(x_n) - \gamma_n \nabla\psi(x_n)\right). \tag{17}$$

Suppose in addition that $(\forall x \in \operatorname{int}\operatorname{dom} f)$ $D^f(x,\cdot)$ is coercive. Then $(x_n)_{n\in\mathbb{N}}$ is a bounded sequence in $\operatorname{int}\operatorname{dom} f$ and $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathcal{S}$. Moreover, there exists $\overline{x} \in \mathcal{S}$ such that the following hold.

(1) *Suppose that $\mathcal{S} \cap \overline{\operatorname{dom}}\, f$ is a singleton. Then $x_n \rightharpoonup \overline{x}$.*
(2) *Suppose that there exists $g \in \mathcal{F}(f)$ such that for every $n \in \mathbb{N}$, $g \succcurlyeq f_n$, and that, for every $y_1 \in \mathcal{X}$ and every $y_2 \in \mathcal{X}$,*

$$\begin{cases} y_1 \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}, \\ y_2 \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}, \\ \left(\langle y_1 - y_2, \nabla f_n(x_n) - \gamma_n\nabla\psi(x_n)\rangle\right)_{n\in\mathbb{N}} \quad \text{converges} \end{cases} \Rightarrow \quad y_1 = y_2. \tag{18}$$

In addition, suppose that one of the following holds.

(a) $\mathcal{S} \subset \operatorname{int}\operatorname{dom} f$.
(b) dom f^* *is open and ∇f^* is weakly sequentially continuous.*

Then $x_n \rightharpoonup \overline{x}$.

(3) *Suppose that* f *satisfies Condition 1 and that one of the following holds.*

(a) *Either* φ *or* ψ *is uniformly convex at* $\overline{x}$.

(b) $\underline{\lim} D_{\mathrm{S}}^{f}(x_n) = 0$ *and there exists* $\mu \in]0, +\infty[$ *such that* $(\forall n \in \mathbb{N})\ \mu \hat{f} \succcurlyeq f_n$.

Then $x_n \to \overline{x}$.

Proof We first derive from Proposition 4(2c) that the operators $(\mathrm{Prox}_{\gamma_n\varphi}^{f})_{n\in\mathbb{N}}$ are single-valued on their domains. We also note that $x_0 \in \mathrm{int}\,\mathrm{dom}\,f$. Suppose that $x_n \in \mathrm{int}\,\mathrm{dom}\,f$ for some $n \in \mathbb{N}$. If f_n is cofinite then Proposition 4(2b) yields

$$\nabla f_n(x_n) - \gamma_n \nabla\psi(x_n) \in \mathcal{X}^* = \mathrm{dom}\,\mathrm{Prox}_{\gamma_n\varphi}^{f_n}. \tag{19}$$

Otherwise,

$$\begin{aligned}\nabla f_n(x_n) - \gamma_n \nabla\psi(x_n) &\in \mathrm{int}\,\mathrm{dom}\,f_n^* + \gamma_n \mathrm{dom}\,\varphi^* = \mathrm{int}(\mathrm{int}\,\mathrm{dom}\,f_n^* + \gamma_n \mathrm{dom}\,\varphi^*)\\ &\subset \mathrm{int}(\mathrm{dom}\,f_n^* + \gamma_n \mathrm{dom}\,\varphi^*) = \mathrm{int}(\mathrm{dom}\,f_n^* + \mathrm{dom}(\gamma_n\varphi^*)).\end{aligned} \tag{20}$$

Since $\mathrm{int}(\mathrm{dom}\,f_n^* + \mathrm{dom}\,(\gamma_n\varphi^*)) \subset \mathrm{dom}\,\mathrm{Prox}_{\gamma_n\varphi}^{f}$ by Proposition 4(2b), we deduce from (17), (19), (20), and Proposition 4(2a) that x_{n+1} is a well-defined element in $\mathrm{ran}\mathrm{Prox}_{\gamma\varphi}^{f_n} = \mathrm{dom}\,\partial\varphi \cap \mathrm{int}\,\mathrm{dom}\,f_n = \mathrm{dom}\,\partial\varphi \cap \mathrm{int}\,\mathrm{dom}\,f \subset \mathrm{int}\,\mathrm{dom}\,f$. By reasoning by induction, we conclude that

$$(x_n)_{n\in\mathbb{N}} \in (\mathrm{int}\,\mathrm{dom}\,f)^{\mathbb{N}} \quad \text{is well-defined.}$$

Next, let us set $\Phi = \varphi + \psi$ and

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad g_n \colon \mathcal{X} &\to]-\infty, +\infty]\\ x &\mapsto \begin{cases} f_n(x) - \gamma_n\psi(x) & \text{if } x \in \mathrm{int}\,\mathrm{dom}\,f,\\ +\infty & \text{otherwise.}\end{cases}\end{aligned} \tag{21}$$

Since $\mathrm{int}\,\mathrm{dom}\,f \subset \mathrm{int}\,\mathrm{dom}\,\psi$, it follows from (21) that $(\forall n \in \mathbb{N})$ g_n is Gâteaux differentiable on $\mathrm{dom}\,g_n = \mathrm{int}\,\mathrm{dom}\,g_n = \mathrm{int}\,\mathrm{dom}\,f$. Since ψ is continuous on $\mathrm{int}\,\mathrm{dom}\,\psi \supset \mathrm{int}\,\mathrm{dom}\,f$ and the functions $(f_n)_{n\in\mathbb{N}}$ are continuous on $\mathrm{int}\,\mathrm{dom}\,f$ [21, Proposition 3.3], we deduce that $(\forall n \in \mathbb{N})$ g_n is continuous on $\mathrm{dom}\,g_n$. In addition,

$$(\forall n \in \mathbb{N}) \quad g_n - \varepsilon\alpha f = (1-\varepsilon)(f_n - \alpha\beta\psi) + \varepsilon(f_n - \alpha f) + \big(\alpha\beta(1-\varepsilon) - \gamma_n\big)\psi. \tag{22}$$

Note that $f \succcurlyeq \beta\psi$ and $(\forall n \in \mathbb{N})$ $f_n \succcurlyeq \alpha f$. Hence, (22) yields

$$(\forall n \in \mathbb{N}) \quad f_n \succcurlyeq \alpha\beta\psi, \tag{23}$$

and hence, we deduce from (16) and (22) that $(\forall n \in \mathbb{N})$ $g_n \succcurlyeq \varepsilon\alpha f$. In turn,

$$\begin{aligned}&(\forall n \in \mathbb{N})(\forall x \in \mathrm{dom}\,g_n)(\forall y \in \mathrm{dom}\,g_n)\\ &\langle x - y, \nabla g_n(x) - \nabla g_n(y)\rangle = D^{g_n}(x,y) + D^{g_n}(y,x) \geq \varepsilon\alpha\big(D^f(x,y) + D^f(y,x)\big) \geq 0,\end{aligned}$$

and it therefore follows from [23, Theorem 2.1.11] that $(\forall n \in \mathbb{N})$ g_n is convex. Consequently,

$$(\forall n \in \mathbb{N}) \quad g_n \in \mathcal{P}_{\varepsilon\alpha}(f). \tag{24}$$

Set $\omega = 1 + 1/\varepsilon$. Then

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad (1+\omega\eta_n)g_n - g_{n+1} &= (1+\omega\eta_n)(f_n - \gamma_n\psi) - (f_{n+1} - \gamma_{n+1}\psi)\\ &= (1+\eta_n)f_n - f_{n+1} + \eta_n\varepsilon^{-1}(f_n - (\gamma_n + \varepsilon\alpha\beta)\psi)\\ &\quad + \big(\alpha\beta\eta_n + \gamma_{n+1} - (1+\eta_n)\gamma_n\big)\psi.\end{aligned}$$

We thus derive from (15), (16) and (23) that

$$(\forall n \in \mathbb{N}) \quad (1+\omega\eta_n)g_n \succcurlyeq g_{n+1}. \tag{25}$$

By invoking (17) and Proposition 4(2a), we get

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

and therefore,

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad \nabla f_n(x_n) - \gamma_n \nabla\psi(x_n) \in\ & \nabla f_n(x_{n+1}) - \gamma_n \nabla\psi(x_{n+1}) \\ & + \gamma_n\big(\partial\varphi(x_{n+1}) + \nabla\psi(x_{n+1})\big).\end{aligned} \tag{26}$$

Since [23, Theorem 2.4.2(vii)–(viii)] yield

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad \partial\varphi(x_{n+1}) + \nabla\psi(x_{n+1}) &\subset \partial\varphi(x_{n+1}) + \partial\psi(x_{n+1}) \\ &\subset \partial(\varphi+\psi)(x_{n+1}) = \partial\Phi(x_{n+1}),\end{aligned}$$

we deduce from (26) that

$$(\forall n \in \mathbb{N}) \quad \nabla g_n(x_n) - \nabla g_n(x_{n+1}) \in \gamma_n \partial\Phi(x_{n+1}). \tag{27}$$

By appealing to (4) and (27), we get

$$\begin{aligned}&(\forall x \in \operatorname{dom}\Phi \cap \operatorname{dom} f)(\forall n \in \mathbb{N}) \\ &\gamma_n^{-1}\langle x - x_{n+1}, \nabla g_n(x_n) - \nabla g_n(x_{n+1})\rangle + \Phi(x_{n+1}) \leq \Phi(x),\end{aligned} \tag{28}$$

and hence, by [6, Proposition 2.3(ii)],

$$\begin{aligned}&(\forall x \in \operatorname{dom}\Phi \cap \operatorname{dom} f)(\forall n \in \mathbb{N}) \\ &\gamma_n^{-1}\big(D^{g_n}(x, x_{n+1}) + D^{g_n}(x_{n+1}, x_n) - D^{g_n}(x, x_n)\big) + \Phi(x_{n+1}) \leq \Phi(x).\end{aligned} \tag{29}$$

In particular,

$$(\forall x \in \mathcal{S} \cap \operatorname{dom} f)(\forall n \in \mathbb{N}) \quad D^{g_n}(x, x_{n+1}) + D^{g_n}(x_{n+1}, x_n) - D^{g_n}(x, x_n) \leq 0. \tag{30}$$

By using (25), we deduce from (30) that

$$\begin{aligned}&(\forall x \in \mathcal{S} \cap \operatorname{dom} f)(\forall n \in \mathbb{N}) \\ &D^{g_{n+1}}(x, x_{n+1}) + (1+\omega\eta_n)D^{g_n}(x_{n+1}, x_n) \leq (1+\omega\eta_n)D^{g_n}(x, x_n),\end{aligned} \tag{31}$$

and therefore,

$$(\forall x \in \mathcal{S} \cap \operatorname{dom} f)(\forall n \in \mathbb{N}) \quad D^{g_{n+1}}(x, x_{n+1}) \leq (1+\omega\eta_n)D^{g_n}(x, x_n). \tag{32}$$

This shows that $(x_n)_{n\in\mathbb{N}}$ is stationarily quasi-Bregman monotone with respect to $\mathcal{S}$ relative to $(g_n)_{n\in\mathbb{N}}$. Hence, we deduce from Proposition 1(2) that

$$(x_n)_{n\in\mathbb{N}} \in (\operatorname{int}\operatorname{dom} f)^{\mathbb{N}} \quad \text{is bounded} \tag{33}$$

and, since $\mathcal{X}$ is reflexive,

$$\mathfrak{W}(x_n)_{n\in\mathbb{N}} \neq \varnothing. \tag{34}$$

In addition, we derive from (32) and Proposition 1(1) that

$$(\forall x \in \mathcal{S} \cap \operatorname{int}\operatorname{dom} f) \quad \big(D^{g_n}(x, x_n)\big)_{n\in\mathbb{N}} \quad \text{converges,} \tag{35}$$

and thus, since (31) yields

$$\begin{aligned}(\forall x \in \mathcal{S} \cap \operatorname{int}\operatorname{dom} f)(\forall n \in \mathbb{N}) \quad 0 &\leq D^{g_n}(x_{n+1}, x_n) \\ &\leq (1+\omega\eta_n)D^{g_n}(x_{n+1}, x_n) \\ &\leq (1+\omega\eta_n)D^{g_n}(x, x_n) - D^{g_{n+1}}(x, x_{n+1}),\end{aligned}$$

and since $\eta_n \to 0$, we obtain

$$D^{g_n}(x_{n+1}, x_n) \to 0. \tag{36}$$

On the other hand, it follows from (24) that

$$(\forall n \in \mathbb{N}) \quad \varepsilon\alpha D^f(x_{n+1}, x_n) \le D^{g_n}(x_{n+1}, x_n),$$

and hence, (36) yields

$$D^f(x_{n+1}, x_n) \to 0. \tag{37}$$

Now, it follows from (29) that

$$(\forall n \in \mathbb{N}) \quad \Phi(x_{n+1}) \le \gamma_n^{-1}\big(D^{g_n}(x_n, x_{n+1}) + D^{g_n}(x_{n+1}, x_n)\big) + \Phi(x_{n+1}) \le \Phi(x_n),$$

which shows that $(\Phi(x_n))_{n\in\mathbb{N}}$ is decreasing and hence, since it is bounded from below by $\inf \Phi(\mathcal{X})$, it is convergent. However, (29) and (32) yield

$$\begin{aligned}
&(\forall x \in \mathcal{S} \cap \operatorname{int}\operatorname{dom} f)(\forall n \in \mathbb{N})\\
&\varepsilon^{-1}\left(\frac{1}{1+\omega\eta_n} D^{g_{n+1}}(x, x_{n+1}) + D^{g_n}(x_{n+1}, x_n) - D^{g_n}(x, x_n)\right) + \Phi(x_{n+1})\\
&\le \gamma_n^{-1}\left(\frac{1}{1+\omega\eta_n} D^{g_{n+1}}(x, x_{n+1}) + D^{g_n}(x_{n+1}, x_n) - D^{g_n}(x, x_n)\right) + \Phi(x_{n+1})\\
&\le \Phi(x).
\end{aligned} \tag{38}$$

Since $\eta_n \to 0$, by taking the limit in (38) and then using (35) and (36), we get

$$\inf \Phi(\mathcal{X}) \le \lim \Phi(x_n) \le \inf \Phi(\mathcal{X}),$$

and thus,

$$\Phi(x_n) \to \inf\Phi(\mathcal{X}). \tag{39}$$

We now show that

$$\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathcal{S}. \tag{40}$$

To this end, suppose that $x \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}$, i.e., $x_{k_n} \rightharpoonup x$. Since Φ is weakly lower semicontinuous [23, Theorem 2.2.1], by (39),

$$\inf \Phi(\mathcal{X}) \le \Phi(x) \le \underline{\lim}\, \Phi(x_{k_n}) = \lim \Phi(x_n) = \inf\Phi(\mathcal{X}).$$

This yields $\Phi(x) = \inf \Phi(\mathcal{X})$, i.e., $x \in \operatorname{Argmin} \Phi = \mathcal{S}$.

(1) Let $\overline{x} \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}$. Since (33) and (40) imply that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathcal{S} \cap \overline{\operatorname{dom}} f$, we obtain $\mathfrak{W}(x_n)_{n\in\mathbb{N}} = \{\overline{x}\}$, and in turn, (34) yields $x_n \rightharpoonup \overline{x}$.

(2) In view of (40) and Proposition 2, it suffices to show that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \operatorname{int}\operatorname{dom} f$.

(2a) We have $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathcal{S} \subset \operatorname{int}\operatorname{dom} f$.

(2b) This follows from Lemma 1.

(3) Let $\overline{x} \in \mathcal{S} \cap \operatorname{int}\operatorname{dom} f$. Since f satisfies Condition 1, (37) yields

$$x_{n+1} - x_n \to 0. \tag{41}$$

Now set

$$(\forall n \in \mathbb{N}) \quad y_n = x_{n+1} \quad \text{and} \quad y_n^* = \gamma_n^{-1}\big(\nabla g_n(x_n) - \nabla g_n(y_n)\big).$$

Then (27) and (41) imply that

$$(\forall n \in \mathbb{N}) \quad y_n^* \in \partial\Phi(y_n) \quad \text{and} \quad y_n - x_n \to 0. \tag{42}$$

Since (31) yields

$$(\forall n\in\mathbb{N})\quad D^{g_{n+1}}(\overline{x},x_{n+1}) = D^{g_{n+1}}(\overline{x},y_n) \leq (1+\omega\eta_n)D^{g_n}(\overline{x},y_n) = (1+\omega\eta_n)D^{g_n}(\overline{x},x_{n+1}) \leq (1+\omega\eta_n)D^{g_n}(\overline{x},x_n),$$

we deduce that

$$(\forall n\in\mathbb{N})\quad (1+\omega\eta_n)^{-1}D^{g_{n+1}}(\overline{x},x_{n+1}) \leq D^{g_n}(\overline{x},y_n) \leq D^{g_n}(\overline{x},x_n). \tag{43}$$

Altogether, (35) and (43) yield

$$D^{g_n}(\overline{x},y_n) - D^{g_n}(\overline{x},x_n) \to 0. \tag{44}$$

In (28), by setting $x=\overline{x}$, we get

$$(\forall n\in\mathbb{N})\quad 0 \leq \gamma_n\langle y_n-\overline{x},y_n^*\rangle = \langle y_n-\overline{x},\nabla g_n(x_n)-\nabla g_n(y_n)\rangle = D^{g_n}(\overline{x},x_n)-D^{g_n}(\overline{x},y_n)-D^{g_n}(y_n,x_n) \leq D^{g_n}(\overline{x},x_n)-D^{g_n}(\overline{x},y_n). \tag{45}$$

By taking to the limit in (45) and using (44), we get

$$\langle y_n-\overline{x},y_n^*\rangle \to 0. \tag{46}$$

(3a) In this case $\mathcal{S}=\{\overline{x}\}$. Since φ is uniformly convex at $\overline{x}$, Φ is likewise and hence, there exists an increasing function $\phi\colon [0,+\infty[\to[0,+\infty]$ that vanishes only at 0 such that

$$(\forall n\in\mathbb{N})(\forall\tau\in]0,1[)\qquad \Phi(\tau\overline{x}+(1-\tau)y_n)+\tau(1-\tau)\phi(\|y_n-\overline{x}\|) \leq \tau\Phi(\overline{x})+(1-\tau)\Phi(y_n).$$

It therefore follows from [23, Page 201] that $\partial\Phi$ is uniformly monotone at $\overline{x}$ and its modulus of convexity is ϕ, i.e,

$$(\forall n\in\mathbb{N})\quad \langle y_n-\overline{x},y_n^*\rangle \geq \phi(\|y_n-\overline{x}\|) \geq 0. \tag{47}$$

Altogether, (46) and (47) yield $\phi(\|y_n-\overline{x}\|)\to 0$, and thus, $y_n\to\overline{x}$. In turn, (42) yields $x_n\to\overline{x}$. The case when ψ is uniformly convex at $\overline{x}$ is similar.

3b) First, we observe that $\mathcal{S}$ is closed and convex since $\Phi\in\Gamma_0(\mathcal{X})$. Next, for every $n\in\mathbb{N}$, since $\mu\hat{f}\succcurlyeq f_n$, we derive from (21) that $\mu\hat{f}\succcurlyeq g_n$. Finally, the strong convergence follows from Proposition 3. □

In Theorem 1, when $(\forall n\in\mathbb{N})$ $f_n=f$, condition (18) is satisfied when both ∇f and $\nabla\psi$ are weakly sequentially continuous. More precisely, we have the following result.

Theorem 2 *Consider the setting of Problem 1 and let* $f\in\Gamma_0(\mathcal{X})$ *be a Legendre function such that* $\mathcal{S}\cap\operatorname{int}\operatorname{dom} f\neq\varnothing$, $\operatorname{int}\operatorname{dom} f\subset\operatorname{int}\operatorname{dom}\psi$, *and* $f\succcurlyeq\beta\psi$ *for some* $\beta\in]0,+\infty[$. *Suppose that either* f *is cofinite or* $-\operatorname{ran}\nabla\psi\subset\operatorname{dom}\varphi^*$, *and that* $(\forall x\in\operatorname{int}\operatorname{dom} f)$ $D^f(x,\cdot)$ *is coercive. Let* $\varepsilon\in]0,\beta/(\beta+1)[$, *let* $(\eta_n)_{n\in\mathbb{N}}\in\ell_+^1(\mathbb{N})$, *and let* $(\gamma_n)_{n\in\mathbb{N}}$ *be a sequence in* $\mathbb{R}$ *such that*

$$(\forall n\in\mathbb{N})\quad \varepsilon\leq\gamma_n\leq\beta(1-\varepsilon)\quad\textit{and}\quad (1+\eta_n)\gamma_n-\gamma_{n+1}\leq\beta\eta_n. \tag{48}$$

Furthermore, let $x_0\in\operatorname{int}\operatorname{dom} f$ *and iterate*

$$(\forall n\in\mathbb{N})\quad x_{n+1}=\operatorname{Prox}^f_{\gamma_n\varphi}\big(\nabla f(x_n)-\gamma_n\nabla\psi(x_n)\big). \tag{49}$$

Then there exists $\overline{x} \in \mathcal{S}$ *such that the following hold.*

(1) *Suppose that one of the following holds.*

(a) $\mathcal{S} \cap \overline{\operatorname{dom} f}$ *is a singleton.*
(b) ∇f *and* $\nabla \psi$ *are weakly sequentially continuous and* $\mathcal{S} \subset \operatorname{int} \operatorname{dom} f$.
(c) $\operatorname{dom} f^*$ *is open and* ∇f, ∇f^*, *and* $\nabla \psi$ *are weakly sequentially continuous.*

Then $x_n \rightharpoonup \overline{x}$.

(2) *Suppose that* f *satisfies Condition 1 and that one of the following holds.*

(a) *Either* φ *or* ψ *is uniformly convex at* $\overline{x}$.
(b) $\underline{\lim}\, D^f_{\mathcal{S}}(x_n) = 0$.

Then $x_n \to \overline{x}$.

Proof Set $(\forall n \in \mathbb{N})\ f_n = f$. Then

$$(\forall n \in \mathbb{N}) \quad \begin{cases} f_n \in \mathcal{P}_1(f), \\ f \succcurlyeq f_n, \\ (1+\eta_n) f_n \succcurlyeq f_{n+1}. \end{cases} \tag{50}$$

(1a): This is a corollary of Theorem 1(1).

(1b)–(1c): Firstly, the proof of Theorem 1(2a) and (2b) shows that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \operatorname{int} \operatorname{dom} f$. Next, in view of Theorem 1(2), it suffices to show that (18) holds. To this end, suppose that y_1 and y_2 are two weak sequential cluster points of $(x_n)_{n\in\mathbb{N}}$ such that

$$\big(\langle y_1 - y_2, \nabla f(x_n) - \gamma_n \nabla \psi(x_n)\rangle\big)_{n\in\mathbb{N}} \quad \text{converges.} \tag{51}$$

Then, there exist two strictly increasing sequences $(k_n)_{n\in\mathbb{N}}$ and $(l_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that $x_{k_n} \rightharpoonup y_1$ and $x_{l_n} \rightharpoonup y_2$. We derive from (48) and [22, Lemma 2.2.2] that there exists $\theta \in [\varepsilon, \beta(1-\varepsilon)]$ such that $\gamma_n \to \theta$. Since ∇f and $\nabla \psi$ are weakly sequentially continuous, after taking the limit in (51) along the subsequences $(x_{k_n})_{n\in\mathbb{N}}$ and $(x_{l_n})_{n\in\mathbb{N}}$, respectively, we get

$$\langle y_1 - y_2, \nabla f(y_1) - \theta \nabla \psi(y_1)\rangle = \langle y_1 - y_2, \nabla f(y_2) - \theta \nabla \psi(y_2)\rangle. \tag{52}$$

Let us define

$$\begin{aligned} h \colon \mathcal{X} &\to \,]-\infty, +\infty] \\ x &\mapsto \begin{cases} f(x) - \theta \psi(x) & \text{if } x \in \operatorname{int} \operatorname{dom} f, \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Then h is Gâteaux differentiable on $\operatorname{int} \operatorname{dom} h = \operatorname{int} \operatorname{dom} f$ and (52) yields

$$\langle y_1 - y_2, \nabla h(y_1) - \nabla h(y_2)\rangle = 0. \tag{53}$$

On the other hand,

$$h - \varepsilon f = f - \theta\psi - \varepsilon f = (1-\varepsilon)(f - \beta\psi) + \big(\beta(1-\varepsilon) - \theta\big)\psi.$$

In turn, since $f \succcurlyeq \beta\psi$ and $\theta \le \beta(1-\varepsilon)$, we obtain $h \succcurlyeq \varepsilon f$, and hence,

$$D^h(y_1, y_2) \ge \varepsilon D^f(y_1, y_2) \quad \text{and} \quad D^h(y_2, y_1) \ge \varepsilon D^f(y_2, y_1).$$

Therefore, (53) yields

$$\begin{aligned} 0 = \langle y_1 - y_2, \nabla h(y_1) - \nabla h(y_2)\rangle &= D^h(y_1, y_2) + D^h(y_2, y_1) \\ &\ge \varepsilon\Big(D^f(y_1, y_2) + D^f(y_2, y_1)\Big) \\ &= \varepsilon\langle y_1 - y_2, \nabla f(y_1) - \nabla f(y_2)\rangle. \end{aligned}$$

Suppose that $y_1 \neq y_2$. Since $f|_{\operatorname{int}\operatorname{dom} f}$ is strictly convex, ∇f is strictly monotone [23, Theorem 2.4.4(ii)], i.e.,

$$\langle y_1 - y_2, \nabla f(y_1) - \nabla f(y_2)\rangle > 0$$

and we reach a contradiction.

(2): The conclusions follow from (50) and Theorem 1(3). □

Remark 1 In condition (48), if we take $(\forall n \in \mathbb{N})\ \eta_n = 0$ then we get the forward-backward splitting algorithm with monotonic step size whose particular case is forward-backward splitting algorithm with constant step-size.

Remark 2 Let us rewrite algorithm (49) as follows

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname*{argmin}_{x\in\mathcal{X}} \Big(\varphi(x) + \langle x - x_n, \nabla\psi(x_n)\rangle + \psi(x_n) + \gamma_n^{-1} D^f(x, x_n)\Big). \tag{54}$$

Another method to solve Problem 1 was proposed in [10]. In that method, instead of solving (54), the authors solve

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname*{argmin}_{x\in\mathcal{X}} \Big(\varphi(x) + \langle x - x_n, \nabla\psi(x_n)\rangle + \psi(x_n) + \gamma_n^{-1}\|x - x_n\|^p\Big) \tag{55}$$

for some $1 < p \leq 2$. The weak convergence is established under the assumptions that Problem 1 admits a unique solution, $\nabla\psi$ is $(p-1)$-Hölder continuous with constant β, and $0 < \inf_{n\in\mathbb{N}} \gamma_n \leq \sup_{n\in\mathbb{N}} \gamma_n \leq (1-\delta)/\beta$, where $0 < \delta < 1$. The high nonlinearity of the regularization in (55) compared to (54) makes the numerical implementation of this method difficult in general. Furthermore, since (55) yields

$$(\forall n \in \mathbb{N}) \quad 0 \in \partial\varphi(x_{n+1}) + \nabla\psi(x_n) + \gamma_n^{-1}\partial\big(\|x_{n+1} - x_n\|^p\big),$$

and since $(\forall n \in \mathbb{N})\ \partial\big(\|x_{n+1} - x_n\|^p\big)$ is not separable, this method is not a splitting method.

Remark 3 We can reformulate Problem 1 as the following joint minimization problem

$$\operatorname*{minimize}_{(x,y)\in V} \varphi(x) + \psi(y),$$

where $V = \{(x, y) \in \mathcal{X} \times \mathcal{X} \mid y = x\}$. This constrained problem is equivalent to the following unconstrained problem

$$\operatorname*{minimize}_{(x,y)\in\mathcal{X}\times\mathcal{X}} \varphi(x) + \psi(y) + \iota_V(x, y).$$

In [8], a different coupling term between the variables x and y was considered and the problem considered there was

$$\operatorname*{minimize}_{(x,y)\in\mathcal{X}\times\mathcal{X}} \varphi(x) + \psi(y) + D^f(x, y),$$

in Euclidean spaces. Their method activates φ and ψ via their so-called left and right Bregman proximity operators alternatively (see also [7] for the projection setting). This method does not require the smoothness of ψ but it requires the computation of Bregman distance-based proximity operator of ψ.

Next, we provide a particular instance of Theorem 2 in finite-dimensional spaces.

Corollary 1 *In the setting of Problem 1, suppose that $\mathcal{X}$ and $\mathcal{Y}$ are finite-dimensional. Let $f \in \Gamma_0(\mathcal{X})$ be a Legendre function such that $\mathcal{S} \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$, $\operatorname{int}\operatorname{dom} f \subset \operatorname{int}\operatorname{dom}\psi$, $f \succcurlyeq \beta\psi$ for some $\beta \in]0,+\infty[$, and $\operatorname{dom} f^*$ is open. Suppose that either f is cofinite or $-\operatorname{ran}\nabla\psi \subset \operatorname{dom}\varphi^*$. Let $\varepsilon \in]0, \beta/(\beta+1)[$, let $(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ such that*

$$(\forall n \in \mathbb{N}) \quad \varepsilon \leq \gamma_n \leq \beta(1-\varepsilon) \quad \text{and} \quad (1+\eta_n)\gamma_n - \gamma_{n+1} \leq \beta\eta_n.$$

Furthermore, let $x_0 \in \operatorname{int}\operatorname{dom} f$ and iterate

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{Prox}^f_{\gamma_n\varphi}\big(\nabla f(x_n) - \gamma_n \nabla\psi(x_n)\big).$$

Then there exists $\overline{x} \in \mathcal{S}$ such that $x_n \to \overline{x}$.

Proof Since $\operatorname{dom} f^*$ is open, [5, Lemma 7.3(ix)] asserts that $(\forall x \in \operatorname{int}\operatorname{dom} f)$ $D^f(x,\cdot)$ is coercive. Hence, the claim follows from Theorem 2(1c). □

4 Application to Multivariate Minimization

In this section, we apply Theorem 2 to solve the following multivariate minimization problem.

Problem 2 Let m and p be strictly positive integers, let $(\mathcal{X}_i)_{1\leq i\leq m}$ and $(\mathcal{Y}_k)_{1\leq k\leq p}$ be reflexive real Banach spaces. For every $i \in \{1,\ldots,m\}$ and every $k \in \{1,\ldots,p\}$, let $\varphi_i \in \Gamma_0(\mathcal{X}_i)$, let $\psi_k \in \Gamma_0(\mathcal{Y}_k)$ be Gâteaux differentiable on $\operatorname{int}\operatorname{dom}\psi_k \neq \varnothing$, and let $L_{ik}\colon \mathcal{X}_i \to \mathcal{Y}_k$ be linear and bounded. The problem is to

$$\underset{x_1\in\mathcal{X}_1,\ldots,x_m\in\mathcal{X}_m}{\text{minimize}} \sum_{i=1}^m \varphi_i(x_i) + \sum_{k=1}^p \psi_k\left(\sum_{i=1}^m L_{ik}x_i\right). \tag{56}$$

Denote by $\mathcal{S}$ the set of solutions to (56).

We derive from Theorem 2 the following result.

Proposition 6 *Consider the setting of Problem 2. For every $k \in \{1,\ldots,p\}$, suppose that there exists $\sigma_k \in]0,+\infty[$ such that for every $(y_{ik})_{1\leq i\leq m} \in \operatorname{int}\operatorname{dom}\psi_k$ and every $(v_{ik})_{1\leq i\leq m} \in \operatorname{int}\operatorname{dom}\psi_k$ satisfying $\sum_{i=1}^m y_{ik} \in \operatorname{int}\operatorname{dom}\psi_k$ and $\sum_{i=1}^m v_{ik} \in \operatorname{int}\operatorname{dom}\psi_k$, one has*

$$D^{\psi_k}\left(\sum_{i=1}^m y_{ik}, \sum_{i=1}^m v_{ik}\right) \leq \sigma_k \sum_{i=1}^m D^{\psi_k}(y_{ik}, v_{ik}). \tag{57}$$

For every $i \in \{1,\ldots,m\}$, let $f_i \in \Gamma_0(\mathcal{X}_i)$ be a Legendre function such that $(\forall x_i \in \operatorname{int}\operatorname{dom} f_i) D^{f_i}(x_i,\cdot)$ is coercive. For every $k \in \{1,\ldots,p\}$, suppose that $\sum_{i=1}^m L_{ik}(\operatorname{int}\operatorname{dom} f_i) \subset \operatorname{int}\operatorname{dom}\psi_k$, that, for every $i \in \{1,\ldots,m\}$, there exists $\beta_{ik} \in]0,+\infty[$ such that $f_i \succcurlyeq \beta_{ik}\psi_k \circ L_{ik}$, and set $\beta_k = \min_{1\leq i\leq m}\beta_{ik}$. In addition, suppose that $\mathcal{S} \cap \bigtimes_{i=1}^m \operatorname{int}\operatorname{dom} f_i \neq \varnothing$ and that either $(\forall i \in \{1,\ldots,m\}) f_i$ is cofinite or $(\forall i \in \{1,\ldots,m\})$ φ_i is cofinite. Let $\varepsilon \in \big]0, 1/\big(1+\sum_{k=1}^p \sigma_k\beta_k^{-1}\big)\big[$, let $(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ such that

$$(\forall n \in \mathbb{N}) \quad \varepsilon \leq \gamma_n \leq \frac{1-\varepsilon}{\sum_{k=1}^p \sigma_k\beta_k^{-1}} \quad \text{and} \quad (1+\eta_n)\gamma_n - \gamma_{n+1} \leq \frac{\eta_n}{\sum_{k=1}^p \sigma_k\beta_k^{-1}}.$$

Furthermore, let $(x_{i,0})_{1\leq i\leq m} \in \times_{i=1}^{m} \operatorname{int}\operatorname{dom} f_i$ *and iterate*

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor \begin{array}{l} \text{for } i = 1, \ldots, m \\ \left\lfloor x_{i,n+1} = \operatorname{Prox}_{\gamma_n \varphi_i}^{f_i}\left(\nabla f_i(x_{i,n}) - \gamma_n \sum_{k=1}^{p} L_{ik}^* \nabla \psi_k \left(\sum_{j=1}^{m} L_{jk} x_{j,n}\right)\right). \right. \end{array} \right.
\end{array}
\tag{58}
$$

Then there exists $(\overline{x}_i)_{1\leq i\leq m} \in \mathcal{S}$ *such that the following hold.*

(1) *Suppose that* $\mathcal{S} \cap \times_{i=1}^{m} \overline{\operatorname{dom}} f_i$ *is a singleton. Then* $(\forall i \in \{1, \ldots, m\})$ $x_{i,n} \rightharpoonup \overline{x}_i$.

(2) *For every* $i \in \{1, \ldots, m\}$ *and every* $k \in \{1, \ldots, p\}$*, suppose that* ∇f_i *and* $\nabla \psi_k$ *are weakly sequentially continuous, and that one of the following holds.*

  (a) $\operatorname{dom} \varphi_i \subset \operatorname{int}\operatorname{dom} f_i$.

  (b) $\operatorname{dom} f_i^*$ *is open and* ∇f_i^* *is weakly sequentially continuous.*

  Then $(\forall i \in \{1, \ldots, m\})$ $x_{i,n} \rightharpoonup \overline{x}_i$.

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

$$
\begin{cases}
\varphi \colon \mathcal{X} \to]-\infty, +\infty] \colon & x \mapsto \sum_{i=1}^{m} \varphi_i(x_i), \\
f \colon \mathcal{X} \to]-\infty, +\infty] \colon & x \mapsto \sum_{i=1}^{m} f_i(x_i), \\
\psi \colon \mathcal{Y} \to]-\infty, +\infty] \colon & y \mapsto \sum_{k=1}^{p} \psi_k(y_k), \\
L \colon \mathcal{X} \to \mathcal{Y} \colon & x \mapsto \left(\sum_{i=1}^{m} L_{ik} x_i\right)_{1\leq k\leq p}.
\end{cases}
\tag{59}
$$

Then ψ is Gâteaux differentiable on $\operatorname{int}\operatorname{dom} \psi = \times_{k=1}^{p} \operatorname{int}\operatorname{dom} \psi_k$ and Problem 2 is a special case of Problem 1. Since (59) yields $\operatorname{dom} f^* = \times_{i=1}^{m} \operatorname{dom} f_i^*$ and $\operatorname{dom} \varphi^* = \times_{i=1}^{m} \operatorname{dom} \varphi_i^*$, we deduce from our assumptions that either f is cofinite or φ is cofinite. As in (9) and (10), f is a Legendre function and $\operatorname{dom} \varphi \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$. In addition,

$$
L(\operatorname{int}\operatorname{dom} f) = \mathop{\times}_{k=1}^{p} \sum_{i=1}^{m} L_{ki}(\operatorname{int}\operatorname{dom} f_i) \subset \mathop{\times}_{k=1}^{p} \operatorname{int}\operatorname{dom} \psi_k = \operatorname{int}\operatorname{dom} \psi.
$$

Now set $\psi_L = \psi \circ L$ and let $x \in \operatorname{int}\operatorname{dom} f$. Then ψ is Gâteaux differentiable at Lx and hence ψ_L is Gâteaux differentiable at x. This implies that $x \in \operatorname{intdom} \psi_L$ and thus $\operatorname{intdom} f \subset \operatorname{intdom} \psi_L$. To show that $D^f(x, \cdot)$ is coercive, we fix $\rho \in \mathbb{R}$. On one hand,

$$
\{z = (z_i)_{1\leq i\leq m} \in \mathcal{X} \mid D^f(x, z) \leq \rho\} \subset \mathop{\times}_{i=1}^{m} \{z_i \in \mathcal{X}_i \mid D^{f_i}(x_i, z_i) \leq \rho\}.
\tag{60}
$$

On the other hand, for every $i \in \{1, \ldots, m\}$, since $D^{f_i}(x_i, \cdot)$ is coercive, we deduce that

$$
\{z_i \in \mathcal{X}_i \mid D^{f_i}(x_i, z_i) \leq \rho\} \quad \text{is bounded.}
$$

Hence, (60) implies that $\{z \in \mathcal{X} \mid D^f(x,z) \leq \rho\}$ is bounded and $D^f(x,\cdot)$ is therefore coercive. Next, set $\beta = 1/\sum_{k=1}^{p} \sigma_k \beta_k^{-1}$. We shall show that $f \succcurlyeq \beta\psi_L$. To this end, fix $z = (z_i)_{1\leq i\leq m} \in \operatorname{int}\operatorname{dom} f$. We have

$$\begin{aligned} D^{\psi_L}(x,z) = D^{\psi}(Lx, Lz) &= \sum_{k=1}^{p} D^{\psi_k}\left(\sum_{i=1}^{m} L_{ik}x_i, \sum_{i=1}^{m} L_{ik}z_i\right) \\ &\leq \sum_{k=1}^{p}\sum_{i=1}^{m} \sigma_k D^{\psi_k}(L_{ik}x_i, L_{ik}z_i) \\ &\leq \sum_{k=1}^{p}\sum_{i=1}^{m} \sigma_k \beta_{ik}^{-1} D^{f_i}(x_i, z_i) \\ &\leq \sum_{k=1}^{p} \sigma_k \beta_k^{-1} D^f(x,z). \end{aligned}$$

Now let us set $(\forall n \in \mathbb{N})\ x_n = (x_{i,n})_{1\leq i\leq m}$. By virtue of Proposition 5, (58) is a particular case of (49).

(1) Since $\mathcal{S} \cap \overline{\operatorname{dom} f}$ is a singleton, the claim follows from Theorem 2(1a).

(2) Our assumptions on $(f_i)_{1\leq i\leq m}$ and $(\psi_k)_{1\leq k\leq p}$ imply that ∇f and $\nabla\psi$ are weakly sequentially continuous.

(2a) Since $\mathcal{S} \cap \times_{i=1}^{m} \operatorname{dom}\varphi_i \subset \times_{i=1}^{m} \operatorname{int}\operatorname{dom} f_i = \operatorname{int}\operatorname{dom} f$, the claim follows from Theorem 2(1b).

(2b) Since, for every $i \in \{1,\ldots,m\}$, $\operatorname{dom} f_i^*$ is open and ∇f_i^* is weakly sequentially continuous, we deduce that $\operatorname{dom} f^*$ is open and ∇f^* is weakly sequentially continuous. The assertion therefore follows from Theorem 2(1c). □

Example 5 In Problem 2, suppose that $m = 1$, that $\mathcal{X}_1$ and $(\mathcal{Y}_k)_{1\leq k\leq p}$ are Hilbert spaces, and that, for every $k \in \{1,\ldots,p\}$, $\varphi_k = \omega_k\|\cdot - r_k\|^2/2$, where $(\omega_k)_{1\leq k\leq p} \in]0,+\infty[^p$ and let $(r_k)_{1\leq k\leq p} \in \times_{k=1}^{p}\mathcal{Y}_k$. Then the weak convergence result in [13, Proposition 6.3] without errors is a particular instance of Proposition 6 with $f_1 = \|\cdot\|^2/2$.

Example 6 Let m and p be strictly positive integers. For every $i \in \{1,\ldots,m\}$ and every $k \in \{1,\ldots,p\}$, let $\omega_{ik} \in]0,+\infty[$, let $\varrho_k \in]0,+\infty[$, and let $\varphi_i \in \Gamma_0(\mathbb{R})$ be cofinite. The problem is to

$$\underset{(\xi_1,\ldots,\xi_m)\in]0,+\infty[^m}{\text{minimize}} \sum_{i=1}^{m}\varphi_i(\xi_i) + \sum_{k=1}^{p}\left(-\ln\frac{\sum_{i=1}^{m}\omega_{ik}\xi_i}{\varrho_k} + \frac{\sum_{i=1}^{m}\omega_{ik}\xi_i}{\varrho_k} - 1\right). \tag{61}$$

Denote by $\mathcal{S}$ the set of solutions to (61) and suppose that $\mathcal{S}\cap]0,+\infty[^m \neq \varnothing$. Let

$$\vartheta\colon \mathbb{R} \to]-\infty,+\infty]\colon \xi \mapsto \begin{cases} -\ln\xi & \text{if } \xi > 0,\\ +\infty & \text{otherwise}\end{cases}$$

be Burg entropy, let $\varepsilon \in]0, 1/(1+p)[$, let $(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ such that

$$(\forall n \in \mathbb{N})\quad \varepsilon \leq \gamma_n \leq p^{-1}(1-\varepsilon) \quad\text{and}\quad (1+\eta_n)\gamma_n - \gamma_{n+1} \leq p^{-1}\eta_n.$$

Let $(\xi_{i,0})_{1\leq i\leq m} \in]0,+\infty[^m$ and iterate

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \text{for } i = 1, \ldots, m \\ \left\lfloor \xi_{i,n+1} = \operatorname{Prox}^{\vartheta}_{\gamma_n\varphi_i} \left(\frac{-1}{\xi_{i,n}} - \gamma_n \sum_{k=1}^{p} \omega_{ik} \left(\frac{-1}{\sum_{j=1}^{m} \omega_{jk}\xi_{j,n}} + \frac{1}{\varrho_k} \right) \right). \right. \end{array} \right. \end{array}$$

Then there exists $(\overline{\xi}_i)_{1\leq i\leq m} \in \mathcal{S}$ such that $(\forall i \in \{1,\ldots,m\})\ \xi_{i,n} \to \overline{\xi}_i$.

Proof For every $i \in \{1,\ldots,m\}$ and every $k \in \{1,\ldots,p\}$, let us set $\mathcal{X}_i = \mathbb{R}$, $\mathcal{Y}_k = \mathbb{R}$, $\psi_k = D^{\vartheta}(\cdot,\varrho_k)$, and $L_{ik} : \xi_i \mapsto \omega_{ik}\xi_i$. Then (61) is a particular case of (56). Since ψ is not differentiable on $\mathbb{R}^p$, the standard forward-backward algorithm is inapplicable. We show that the problem can be solved by using Proposition 6. First, let $(\xi_i)_{1\leq i\leq m}$ and $(\eta_i)_{1\leq i\leq m}$ be in $]0,+\infty[^m$, and consider

$$\phi : \mathbb{R} \to]-\infty,+\infty] : \xi \mapsto \begin{cases} -\ln\xi + \xi - 1 & \text{if } \xi \in]0,+\infty[, \\ +\infty & \text{otherwise.} \end{cases}$$

We see that ϕ is convex and positive. Thus,

$$\phi\left(\frac{\sum_{i=1}^m \xi_i}{\sum_{i=1}^m \eta_i}\right) = \phi\left(\sum_{i=1}^m \frac{\eta_i}{\sum_{j=1}^m \eta_j}\frac{\xi_i}{\eta_i}\right) \leq \sum_{i=1}^m \frac{\eta_i}{\sum_{j=1}^m \eta_j}\phi\left(\frac{\xi_i}{\eta_i}\right) \leq \sum_{i=1}^m \phi\left(\frac{\xi_i}{\eta_i}\right),$$

and hence,

$$-\ln\frac{\sum_{i=1}^m \xi_i}{\sum_{i=1}^m \eta_i} + \frac{\sum_{i=1}^m \xi_i}{\sum_{i=1}^m \eta_i} - 1 \leq \sum_{i=1}^m \left(-\ln\frac{\xi_i}{\eta_i} + \frac{\xi_i}{\eta_i} - 1\right).$$

In turn,

$$D^{\vartheta}\left(\sum_{i=1}^m \xi_i, \sum_{i=1}^m \eta_i\right) \leq \sum_{i=1}^m D^{\vartheta}(\xi_i,\eta_i).$$

This shows that (57) is satisfied with $(\forall k \in \{1,\ldots,p\})\ \sigma_k = 1$. Next, let us set $(\forall i \in \{1,\ldots,m\})\ f_i = \vartheta$. Fix $i \in \{1,\ldots,m\}$ and $k \in \{1,\ldots,p\}$, and let ξ_i and η_i be in $]0,+\infty[$. Then

$$D^{\psi_k}(L_{ik}\xi_i, L_{ik}\eta_i) = D^{\vartheta}(\omega_{ik}\xi_i,\omega_{ik}\eta_i) = D^{\vartheta}(\xi_i,\eta_i) = D^{f_i}(\xi_i,\eta_i),$$

which implies that $f_i \succcurlyeq \psi_k \circ L_{ik}$. In addition, since $\operatorname{dom} f_i^* =]-\infty,0[$ is open, [5, Lemma 7.3(ix)] asserts that $D^{f_i}(\xi_i,\cdot)$ is coercive. We therefore deduce the convergence result from Proposition 6(2b). □

Example 7 Let m and p be strictly positive integers. For every $i \in \{1,\ldots,m\}$ and every $k \in \{1,\ldots,p\}$, let $\omega_{ik} \in]0,+\infty[$, let $\varrho_k \in]0,+\infty[$, and let $\varphi_i \in \Gamma_0(\mathbb{R})$. The problem is to

$$\underset{(\xi_1,\ldots,\xi_m)\in[0,+\infty[^m}{\text{minimize}} \sum_{i=1}^m \varphi_i(\xi_i) + \sum_{k=1}^p \left(\left(\sum_{i=1}^m \omega_{ik}\xi_i\right)\ln\frac{\sum_{i=1}^m \omega_{ik}\xi_i}{\varrho_k} - \sum_{i=1}^m \omega_{ik}\xi_i + \varrho_k\right). \quad (62)$$

Denote by $\mathcal{S}$ the set of solutions to (62) and suppose that $\mathcal{S}\cap]0,+\infty[^m \neq \varnothing$. Let

$$\vartheta : \mathbb{R} \to]-\infty,+\infty] : \xi \mapsto \begin{cases} \xi\ln\xi - \xi & \text{if } \xi \in]0,+\infty[, \\ 0 & \text{if } \xi = 0, \\ +\infty & \text{otherwise} \end{cases}$$

be Boltzmann–Shannon entropy, let $\beta = \max_{1\leq k\leq p}\max_{1\leq i\leq m}\omega_{ik}$, let $\varepsilon \in]0,1/(1+\beta)[$, let $(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ such that

$$(\forall n \in \mathbb{N})\quad \varepsilon \leq \gamma_n \leq (p\beta)^{-1}(1-\varepsilon) \quad\text{and}\quad (1+\eta_n)\gamma_n - \gamma_{n+1} \leq (p\beta)^{-1}\eta_n.$$

Let $(\xi_{i,0})_{1\leq i\leq m}\in]0,+\infty[^m$ and iterate

$$\begin{array}{l}\text{for } n=0,1,\ldots\\ \left\lfloor\begin{array}{l}\text{for } i=1,\ldots,m\\ \left\lfloor \xi_{i,n+1}=\operatorname{Prox}^{\vartheta}_{\gamma_n\varphi_i}\left(\ln\xi_{i,n}-\gamma_n\sum_{k=1}^{p}\omega_{ik}\left(\ln\left(\sum_{j=1}^{m}\omega_{jk}\xi_{j,n}\right)-\ln\varrho_k\right)\right).\right.\end{array}\right.\end{array}$$

Then, there exists $(\overline{\xi}_i)_{1\leq i\leq m}\in\mathcal{S}$ such that $(\forall i\in\{1,\ldots,m\})$ $\xi_{i,n}\to\overline{\xi}_i$.

Proof For every $i\in\{1,\ldots,m\}$ and every $k\in\{1,\ldots,p\}$, let us set $\mathcal{X}_i=\mathbb{R}$, $\mathcal{Y}_k=\mathbb{R}$, $\psi_k=D^{\vartheta}(\cdot,\varrho_k)$, and $L_{ik}\colon\xi_i\mapsto\omega_{ik}\xi_i$. Then (62) is a particular case of (56). We cannot apply the standard forward-backward algorithm here since ψ is not differentiable on $\mathbb{R}^p$. We shall verify the assumptions of Proposition 6. First, let $(\xi_i)_{1\leq i\leq m}$ and $(\eta_i)_{1\leq i\leq m}$ be in $]0,+\infty[^m$. Since

$$\phi\colon\mathbb{R}\to]-\infty,+\infty]\colon\xi\mapsto\begin{cases}\xi\ln\xi & \text{if } \xi\in]0,+\infty[,\\ 0 & \text{if } \xi=0,\\ +\infty & \text{otherwise}\end{cases}$$

is convex, we have

$$\phi\left(\frac{\sum_{i=1}^m\xi_i}{\sum_{i=1}^m\eta_i}\right)=\phi\left(\sum_{i=1}^m\frac{\eta_i}{\sum_{j=1}^m\eta_j}\frac{\xi_i}{\eta_i}\right)\leq\sum_{i=1}^m\frac{\eta_i}{\sum_{j=1}^m\eta_j}\phi\left(\frac{\xi_i}{\eta_i}\right),$$

and hence,

$$\frac{\sum_{i=1}^m\xi_i}{\sum_{i=1}^m\eta_i}\ln\frac{\sum_{i=1}^m\xi_i}{\sum_{i=1}^m\eta_i}\leq\sum_{i=1}^m\frac{\eta_i}{\sum_{j=1}^m\eta_j}\frac{\xi_i}{\eta_i}\ln\frac{\xi_i}{\eta_i}=\frac{\sum_{i=1}^m\xi_i\ln\frac{\xi_i}{\eta_i}}{\sum_{i=1}^m\eta_i}.$$

In turn,

$$\left(\sum_{i=1}^m\xi_i\right)\ln\frac{\sum_{i=1}^m\xi_i}{\sum_{i=1}^m\eta_i}\leq\sum_{i=1}^m\xi_i\ln\frac{\xi_i}{\eta_i},$$

which implies that

$$\begin{aligned}D^{\vartheta}\left(\sum_{i=1}^m\xi_i,\sum_{i=1}^m\eta_i\right)&=\left(\sum_{i=1}^m\xi_i\right)\ln\frac{\sum_{i=1}^m\xi_i}{\sum_{i=1}^m\eta_i}-\sum_{i=1}^m\xi_i+\sum_{i=1}^m\eta_i\\&\leq\sum_{i=1}^m\left(\xi_i\ln\frac{\xi_i}{\eta_i}-\xi_i+\eta_i\right)\\&=\sum_{i=1}^m D^{\vartheta}(\xi_i,\eta_i).\end{aligned}$$

This shows that (57) is satisfied with $(\forall k\in\{1,\ldots,p\})$ $\sigma_k=1$. Next, let us set $(\forall i\in\{1,\ldots,m\})$ $f_i=\vartheta$. Fix $i\in\{1,\ldots,m\}$ and $k\in\{1,\ldots,p\}$, and let ξ_i and η_i be in $]0,+\infty[$. Then

$$D^{\psi_k}(L_{ik}\xi_i,L_{ik}\eta_i)=D^{\vartheta}(\omega_{ik}\xi_i,\omega_{ik}\eta_i)=\omega_{ik}D^{\vartheta}(\xi_i,\eta_i)\leq\beta D^{\vartheta}(\xi_i,\eta_i),$$

which implies that $f_i\succcurlyeq\beta^{-1}\psi_k\circ L_{ik}$. In addition, since f_i is supercoercive, f_i is cofinite and [5, Lemma 7.3(viii)] asserts that $D^{f_i}(\xi_i,\cdot)$ is coercive. Therefore, the claim follows from Proposition 6(2b). □

Remark 4 The Bregman distance associated with Burg entropy, i.e., the Itakura–Saito divergence, is used in linear regression [3, Section 3]. The Bregman distance associated with

Boltzmann–Shannon entropy, i.e., the Kullback–Leibler divergence, is used in information theory [3, Section 3] and image processing [11].

Acknowledgments I would like to thank my doctoral advisor Professor Patrick L. Combettes for bringing this problem to my attention and for helpful discussions. The contributions of the referees to the article are important and I sincerely thank them for those.

References

1. Attouch, H., Brezis, H.: Duality for the sum of convex functions in general Banach spaces. In: Barroso, J.A. (ed.) Aspects of Mathematics and its Applications. North-Holland Mathematics Library, vol 34, pp 125–133, North-Holland, Amsterdam (1986)
2. Banerjee, A., Basu, S., Merugu, S.: Multi-way clustering on relation graphs. In: Apte, C., et al. (eds.) Proceedings of the 2007 SIAM International Conference on Data Mining, pp. 145–156. SIAM, Philadelphia (2007)
3. Basseville, M.: Divergence measures for statistical data processing—An annotated bibliography. Signal Process. **93**, 621–633 (2013)
4. Bauschke, H.H., Borwein, J.M.: Legendre functions and the method of random Bregman projections. J. Convex Anal. **4**, 27–67 (1997)
5. Bauschke, H.H., Borwein, J.M., Combettes, P.L.: Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces. Commun. Contemp. Math. **3**, 615–647 (2001)
6. Bauschke, H.H., Borwein, J.M., Combettes, P.L.: Bregman monotone optimization algorithms. SIAM J. Control Optim. **42**, 596–636 (2003)
7. Bauschke, H.H., Combettes, P.L.: Iterating Bregman retractions. SIAM. J. Optim **13**, 1159–1173 (2003)
8. Bauschke, H.H., Combettes, P.L., Noll, D.: Joint minimization with alternating Bregman proximity operators. Pac. J. Optim. **2**, 401–424 (2006)
9. Bertero, M., Boccacci, P., Desiderà, G., Vicidomini, G.: Image deblurring with Poisson data: from cells to galaxies. Inverse Probl. **25**, 123006 (2009). 26 pages
10. Bredies, K.: A forward-backward splitting algorithm for the minimization of non-smooth convex functionals in Banach space. Inverse Probl. **25**, 015005 (2009). 20 pages
11. Byrne, C.L.: Iterative image reconstruction algorithms based on cross-entropy minimization. IEEE Trans. Image Process. **2**, 96–103 (1993)
12. Chandrasekaran, V., Parrilo, P.A., Willsky, A.S.: Latent variable graphical model selection via convex optimization. Ann. Stat. **40**, 1935–1967 (2012)
13. Combettes, P.L., Vũ, B.C.: Variable metric quasi-Fejér monotonicity. Nonlinear Anal. **78**, 17–31 (2013)
14. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. Multiscale Model Simul. **4**, 1168–1200 (2005)
15. Corless, R.M., Gonnet, G.H., Hare, D.E.G., Jeffrey, D.J., Knuth, D.E.: On the Lambert *W* function. Adv. Comput. Math. **5**, 329–359 (1996)
16. Kivinen, J., Warmuth, M.K.: Relative loss bounds for multidimensional regression problems. Mach. Learn. **45**, 301–329 (2001)
17. Lantéri, H., Roche, M., Aime, C.: Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms. Inverse Probl. **18**, 1397–1419 (2002)
18. Markham, J., Conchello, J.-A.: Fast maximum-likelihood image-restoration algorithms for three-dimensional fluorescence microscopy. J. Opt. Soc. Am. A **18**, 1062–1071 (2001)
19. Moreau, J.-J.: Fonctions convexes duales et points proximaux dans un espace hilbertien. C.R. Acad. Sci. Paris **255**, 2897–2899 (1962)
20. Nguyen, V.Q.: Variable quasi-Bregman monotone sequences. Numer. Algor. doi:10.1007/s11075-016-0132-9 (2016)
21. Phelps, R.R. Convex Functions, Monotone Operators and Differentiability. Lecture Notes in Mathematics, 2nd edn., vol. 1364. Springer-Verlag, Berlin (1993)
22. Polyak, B.T.: Introduction to Optimization. Translations Series in Mathematics and Engineering. Optimization Software, Inc. Publications Division, New York (1987)
23. Zălinescu, C.: Convex Analysis in General Vector Spaces. World Scientific Publishing Co., Inc., River Edge (2002)