Joint minimization with alternating Bregman proximity operators

Heinz H. Bauschke[*], Patrick L. Combettes[†], and Dominikus Noll[‡]

December 22, 2004

Abstract

A systematic study of the proximity properties of Bregman distances is carried out. This investigation leads to the introduction of a new type of proximity operator which complements the usual Bregman proximity operator. We establish key properties of these operators and utilize them to devise a new alternating procedure for solving a broad class of joint minimization problems. We provide a comprehensive convergence analysis of this algorithm. Our framework is shown to capture and extend various optimization methods.

2000 Mathematics Subject Classification: Primary 90C25; Secondary 49M05, 49M45, 65K05, 90C30.

Keywords: Alternating minimization, alternating projections, Bregman distance, Bregman projection, left proximity operator, Moreau envelope, proximal point algorithm, prox operator, right proximity operator.

1 Introduction

1.1 Standing assumptions

Throughout, $X = \mathbb{R}^J$ is the standard Euclidean space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$, and

(1) $$f \colon X \to \left]-\infty, +\infty\right] \text{ is convex and differentiable on } U = \operatorname{int} \operatorname{dom} f \neq \varnothing.$$

Recall that (see [20])

(2) $$D_f \colon X \times X \to [0, +\infty] : (x, y) \mapsto \begin{cases} f(x) - f(y) - \langle f'(y), x - y \rangle, & \text{if } y \in U; \\ +\infty, & \text{otherwise} \end{cases}$$

[*]Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario N1G 2W1, Canada. E-mail: hbauschk@uoguelph.ca.

[†]Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie – Paris 6, 75005 Paris, France. E-mail: plc@math.jussieu.fr.

[‡]Université Paul Sabatier, Laboratoire MIP, CNRS UMR 5640, 118 route de Narbonne, 31062 Toulouse Cedex, France. E-mail: noll@mip.ups-tlse.fr.

is the *Bregman distance* associated with f, also denoted by D for brevity. Let $\Gamma_0(X)$ be the set of all proper lower semicontinuous convex functions from X to $]-\infty,+\infty]$. In addition, f satisfies the following standard properties:

A1 $f \in \Gamma_0(X)$ is a convex function of Legendre type, i.e., f is essentially smooth and essentially strictly convex in the sense of [41, Section 26];

A2 f'' exists and is continuous on U;

A3 D is jointly convex, i.e., convex on $X \times X$;

A4 $(\forall x \in U)$ $D(x,\cdot)$ is strictly convex on U;

A5 $(\forall x \in U)$ $D(x,\cdot)$ is coercive, i.e., the lower level set $\{y \in X \colon D(x,y) \leq \eta\}$ is bounded, for every $\eta \in \mathbb{R}$.

These assumptions allow us to encompass several important scenarios, see Example 2.5. Finally, φ and ψ are two functions such that

$$(3) \qquad \begin{cases} \varphi \in \Gamma_0(X), \\ (\forall y \in U)\ \ \varphi(\cdot) + D(\cdot, y) \ \text{ is coercive}, \\ \operatorname{dom}\varphi \cap U \neq \varnothing, \end{cases} \qquad \text{and} \qquad \begin{cases} \psi \in \Gamma_0(X), \\ (\forall x \in U)\ \ \psi(\cdot) + D(x, \cdot) \ \text{ is coercive}, \\ \operatorname{dom}\psi \cap U \neq \varnothing. \end{cases}$$

1.2 Problem statement

Bregman distances were introduced in [12] as an extension to the usual discrepancy measure $(x,y) \mapsto \|x-y\|^2$ and have since found numerous applications in optimization, convex feasibility, convex inequalities, variational inequalities, monotone inclusions, equilibrium problems; see [7, 15, 20] and the references therein. The problem under consideration in the present paper is the joint minimization problem

$$(4) \qquad \text{minimize} \quad \Lambda \colon (x,y) \mapsto \varphi(x) + \psi(y) + D(x,y) \quad \text{over} \quad U \times U.$$

The optimal value of (4) and its set of solutions will be denoted by

$$(5) \qquad p = \inf \Lambda(U \times U) \ \text{ and } \ S = \big\{(x,y) \in U \times U \colon \Lambda(x,y) = p\big\},$$

respectively.

The objective function Λ in (4) consists of a separable term $(x,y) \mapsto \varphi(x) + \psi(y)$ and of a coupling term D. This structure arises explicitly or implicitly in a variety of problems, for instance in the areas of image processing [2, 44], signal recovery [23], statistics [17, 25, 30], mechanics [36], and wavelet synthesis [39]. Further applications will be described in Section 5.

Let $\Delta = \{(x,x) \colon x \in X\}$. Then it follows from Lemma 2.4(i) and **A1** that

$$(6) \qquad (\forall (x,y) \in U \times U) \quad D(x,y) = 0 \quad \Leftrightarrow \quad (x,y) \in \Delta.$$

Therefore, Problem (4) can be viewed as a relaxation of

$$\text{minimize} \quad (x,y) \mapsto \varphi(x) + \psi(y) + \iota_\Delta(x,y) \quad \text{over} \quad U \times U, \tag{7}$$

which, in turn, is equivalent to the standard problem

$$\text{minimize} \quad \varphi + \psi \quad \text{over} \quad U. \tag{8}$$

For the sake of illustration, let us consider the case when $f = \frac{1}{2}\|\cdot\|^2$, so that $U = X$ and $D\colon (x,y) \mapsto \frac{1}{2}\|x-y\|^2$. If φ and ψ are the indicator functions of two nonempty closed convex sets A and B, respectively, then (8) corresponds to the *convex feasibility problem* of finding a point in $A \cap B$. When no such point exists, a sensible alternative is to look for a pair $(x,y) \in A \times B$ such that $\|x-y\| = \inf \|A - B\|$. This formulation, which corresponds to (4), was proposed in [22] and has found many applications in engineering [23, 36, 39]. The algorithm devised in [22] to solve this joint best approximation problem is the alternating projections method

$$\text{fix } x_0 \in X \quad \text{and set} \quad (\forall n \in \mathbb{N}) \quad y_n = P_B(x_n) \text{ and } x_{n+1} = P_A(y_n). \tag{9}$$

More generally, let $\operatorname{prox}_\theta\colon x \mapsto \operatorname{argmin}_y \theta(y) + \frac{1}{2}\|x-y\|^2$ be the proximity operator [37, 38] associated with a function $\theta \in \Gamma_0(X)$. In [1], (9) was extended to the algorithm

$$\text{fix } x_0 \in X \quad \text{and set} \quad (\forall n \in \mathbb{N}) \quad y_n = \operatorname{prox}_\psi(x_n) \text{ and } x_{n+1} = \operatorname{prox}_\varphi(y_n) \tag{10}$$

in order to solve

$$\text{minimize} \quad (x,y) \mapsto \varphi(x) + \psi(y) + \tfrac{1}{2}\|x-y\|^2 \quad \text{over} \quad X \times X. \tag{11}$$

The purpose of this paper is to introduce and analyze a proximal-like method to solve (4) under the assumptions stated above. The lack of symmetry of D prompts us to consider two single-valued operators defined on U, namely

$$\overleftarrow{\operatorname{prox}}_\varphi\colon y \mapsto \operatorname*{argmin}_{x \in U} \varphi(x) + D(x,y) \quad \text{and} \quad \overrightarrow{\operatorname{prox}}_\psi\colon x \mapsto \operatorname*{argmin}_{y \in U} \psi(y) + D(x,y). \tag{12}$$

The operators $\overleftarrow{\operatorname{prox}}_\varphi$ and $\overrightarrow{\operatorname{prox}}_\psi$ will be called the *left* and the *right proximity operator*, respectively. While left proximity operators have already been used in the literature (see [7] and the references therein), the notion of a right proximity operator at this level of generality appears to be new. We note that [28, p. 26f] observes (but does not exploit) a superficial similarity between the iterative step of a multiplicative algorithm and the application of the right proximity operator $\overrightarrow{\operatorname{prox}}_\psi$ in the Kullback-Leibler divergence setting (see Example 2.5(ii)), where ψ is assumed to be the sum of a continuous convex function and the indicator function of the nonnegative orthant in X.

In this paper, we shall provide a detailed analysis of these operators and establish key properties. With these tools in place, we shall be in a position to tackle (4) by alternating minimizations of Λ. We thus obtain the following algorithm

$$\text{fix } x_0 \in U \quad \text{and set} \quad (\forall n \in \mathbb{N}) \quad y_n = \overrightarrow{\operatorname{prox}}_\psi(x_n) \text{ and } x_{n+1} = \overleftarrow{\operatorname{prox}}_\varphi(y_n). \tag{13}$$

It is important to realize that it is quite nontrivial to see that this iteration is even well defined. The difficulty lies in guaranteeing that every iterate formally defined in (13) lies again in U, so that the iterative update can be carried out. The crucial details of our analysis rely on various results on the interplay between the Bregman distance and the assumptions **A1**–**A5** imposed on f. Armed with those results, we shall analyze the asymptotic behavior of this algorithm and, in particular, we shall establish convergence to a solution of (4). In the special case when $\psi = 0$, we recover the classical Bregman proximal method proposed in [19] (see also [15] and [20]). Moreover, if we let $\varphi = 0$, we obtain a novel proximal point method. We shall also recover and extend various parallel decomposition algorithms, including least-squares techniques for inconsistent feasibility problems with finitely many sets. Let us also note that if φ is an indicator function and ψ is the sum of an indicator function and a differentiable convex function, then problem (4) reduces to a setting discussed in [29, Remark 2.18]. However, the proofs in that manuscript are somewhat sketchy as several details are omitted. For instance, [29] does not explain why the iteration (13) is well defined. Algorithm (13) may also be interpreted as a cyclic descent or nonlinear Gauss-Seidel method. However, the typical general convergence results for the latter methods (see, e.g., [11, Proposition 3.3.9]) fail to cover our main result (Theorem 4.4).

The paper is organized as follows. In Section 2, we collect the technical results required by our analysis. Left and right Bregman proximity operators are introduced and studied in Section 3. The asymptotic properties of Algorithm (13) are investigated in Section 4. Finally, various applications and connections with previous works are described in Section 5.

2 Auxiliary results

Lemma 2.1 *Let $g\colon X \to \left]-\infty, +\infty\right]$ be a proper convex function with $V = \operatorname{int}\operatorname{dom} g$.*

(i) *If $V \neq \emptyset$ and g is differentiable on V, then g' is continuous on V.*

(ii) *The function g admits an affine minorant.*

Proof. (i): [41, Theorem 25.5]. (ii): [41, Corollary 12.1.2]. ■

Lemma 2.2 [41, Corollary 14.2.2] *Let $g \in \Gamma_0(X)$. Then g is coercive if and only if $0 \in \operatorname{int}\operatorname{dom} g^*$.*

Lemma 2.3 *Let C be an open convex subset of X and let $g \in \Gamma_0(X)$ be such that $C \cap \operatorname{dom} g \neq \emptyset$. Then $\inf g(C) = \inf g(\overline{C})$.*

Proof. The inequality $\inf g(C) \geq \inf g(\overline{C})$ is clear. Since $C \cap \operatorname{dom} g \neq \emptyset$, the convexity of $\operatorname{dom} g$ and [41, Corollary 6.3.2] imply that there exists $c \in C \cap \operatorname{ri}\operatorname{dom} g$. Now fix $x \in \overline{C} \cap \operatorname{dom} g$ and note that, by [41, Theorem 6.1],

$$\left]x, c\right] \subset C \cap \operatorname{ri}\operatorname{dom} g. \tag{14}$$

Next, we define, for every $\alpha \in]0,1]$, $x_\alpha = (1-\alpha)x + \alpha c \in C \cap \operatorname{ri}\operatorname{dom} g$. It follows from the segment continuity property [41, Theorem 7.5] that $g(x) = \lim_{\alpha\to 0^+} g(x_\alpha)$. Thus, $g(x) \geq \inf g(C)$. We conclude that $\inf g(\overline{C}) \geq \inf g(C)$. ■

Lemma 2.4

(i) $(\forall x \in X)(\forall y \in U)\ \ D(x,y) = 0 \ \ \Leftrightarrow \ \ x = y$.

(ii) $(\forall y \in U)\ D(\cdot, y)$ *is coercive.*

(iii) *If* $x \in U$ *and* $(y_n)_{n\in\mathbb{N}}$ *is a sequence in* U *such that* $y_n \to y \in \operatorname{bdry} U$, *then* $D(x, y_n) \to +\infty$.

Proof. (i): [6, Theorem 3.7.(iv)]. (ii): [6, Theorem 3.7.(iii)]. (iii): [6, Theorem 3.8.(i)]. ■

Example 2.5 [10, Example 2.16] Assumptions **A1**–**A5** hold in the following cases, where $x = (\xi_j)_{1\leq j\leq J}$ and $y = (\eta_j)_{1\leq j\leq J}$ are two generic points in $\mathbb{R}^J$.

(i) *Energy:* If $f\colon x \mapsto \frac{1}{2}\|x\|^2$, then $U = X$ and

$$D(x,y) = \frac{1}{2}\|x-y\|^2.$$

(ii) *Boltzmann-Shannon entropy:* If $f\colon x \mapsto \sum_{j=1}^J \xi_j \ln(\xi_j) - \xi_j$, then $U = \{x \in X\colon x > 0\}$ and one obtains the *Kullback-Leibler divergence*

$$D(x,y) = \begin{cases} \sum_{j=1}^J \xi_j \ln(\xi_j/\eta_j) - \xi_j + \eta_j, & \text{if } x \geq 0 \text{ and } y > 0;\\ +\infty, & \text{otherwise.}\end{cases}$$

(iii) *Fermi-Dirac entropy:* If $f\colon x \mapsto \sum_{j=1}^J \xi_j \ln(\xi_j) + (1-\xi_j)\ln(1-\xi_j)$, then $U = \{x \in X\colon 0 < x < 1\}$ and

$$D(x,y) = \begin{cases} \sum_{j=1}^J \xi_j \ln(\xi_j/\eta_j) + (1-\xi_j)\ln\big((1-\xi_j)/(1-\eta_j)\big), & \text{if } 0 \leq x \leq 1 \text{ and } 0 < y < 1;\\ +\infty, & \text{otherwise.}\end{cases}$$

(iv) *Log-quad function* [3]: If $f\colon x \mapsto \frac{1}{2}\|x\|^2 + \sum_{j=1}^J \xi_j \ln(\xi_j) - \xi_j$, then $U = \{x \in X\colon x > 0\}$ and

$$D(x,y) = \begin{cases} \frac{1}{2}\|x-y\|^2 + \sum_{j=1}^J \xi_j \ln(\xi_j/\eta_j) - \xi_j + \eta_j, & \text{if } x \geq 0 \text{ and } y > 0;\\ +\infty, & \text{otherwise.}\end{cases}$$

Lemma 2.6 *Suppose that* $x \in X$ *and* $\{u, v\} \subset U$. *Then:*

$$D(x,v) = D(x,u) + D(u,v) + \big\langle f'(v) - f'(u), u - x\big\rangle. \tag{15}$$

Moreover, D *is continuous on* $U \times U$ *and* $D(u,\cdot) \in \Gamma_0(X)$.

Proof. The proof of the identity (15) is clear from (2). The continuity of D on $U \times U$ follows from Lemma 2.1(i) The function $D(u, \cdot)$ is convex by **A3**, and proper since $u \in U$. To verify lower semicontinuity of $D(u, \cdot)$, it suffices — in view of (2) — to take a sequence $(y_n)_{n\in\mathbb{N}}$ in U that converges to $y \in \mathrm{cl}(U)$ and to show that $D(u, y) \leq \underline{\lim}\, D(u, y_n)$. If $y \in U$, then $D(u, y_n) \to D(u, y)$ by continuity of D on $U \times U$. If $y \in \mathrm{bdry}(U)$, then Lemma 2.4(iii) implies that $D(u, y_n) \to +\infty = D(u, y)$. ■

Remark 2.7 The simple yet useful identity (15) is also known as "three points identity", see [21, Lemma 3.1].

Lemma 2.8 *Take $z \in U$ and $h \in X$. Then:*

$$\lim_{t\to 0^+} \frac{D(z, z + th)}{t} = 0 = \lim_{t\to 0^+} \frac{D(z + th, z)}{t}. \tag{16}$$

Proof. Expanding according to (2) shows that for all positive t sufficiently small one has

$$\frac{D(z, z + th)}{t} = \frac{f(z) - f(z + th)}{t} + \langle f'(z + th), h\rangle \tag{17}$$

and

$$\frac{D(z + th, z)}{t} = \frac{f(z + th) - f(z)}{t} - \langle f'(z), h\rangle. \tag{18}$$

The result now follows by letting tend t to 0 from above. ■

Because of **A2** and **A3**, the function D_f conforms to (1) and therefore its Bregman distance D_{D_f}, which will play a central role in our analysis, is well-defined.

Lemma 2.9 [10, Lemma 2.9] *Take $\{x, y, u, v\} \subset U$. Then:*

$$D_{D_f}\big((x, y), (u, v)\big) = D_f(x, y) + D_f(x, u) - D_f(x, v) + \big\langle f''(v)(u - v), y - v\big\rangle. \tag{19}$$

Moreover, D_{D_f} is continuous on U^4.

Note that D_f itself does not satisfy the counterparts of properties **A1**–**A5**; for instance, strict convexity fails as we shall see shortly. However, the expression for D_{D_f} becomes simpler when we deal with the energy or the Boltzmann-Shannon entropy (which are defined in Example 2.5):

Example 2.10 [10, Example 2.12] Take $\{x, y, u, v\} \subset U$. Then:

(i) If f is the energy, then $D_{D_f}\big((x, y), (u, v)\big) = D_f\big(x, y + (u - v)\big)$.

(ii) If f is the Boltzmann-Shannon entropy, then $D_{D_f}\big((x, y), (u, v)\big) = D_f\big(x, yu/v\big)$, where the product and the quotient is taken coordinate-wise.

We do not know whether a similar simplification can be obtained for the Fermi-Dirac entropy.

Lemma 2.11 *Let $\theta\colon X \to \,]-\infty,+\infty]$ be convex and $x \in X$ be such that $\operatorname{dom}\theta \cap U \neq \emptyset$ and $\theta(\cdot) + D(x,\cdot)$ is coercive. Suppose $(y_n)_{n\in\mathbb{N}}$ is a sequence in U such that $\big(\theta(y_n) + D(x,y_n)\big)_{n\in\mathbb{N}}$ is bounded. Then $(y_n)_{n\in\mathbb{N}}$ is bounded and all its cluster points belong to U.*

Proof. The coercivity assumption implies the boundedness of $(y_n)_{n\in\mathbb{N}}$. Now let y be a cluster point of $(y_n)_{n\in\mathbb{N}}$, say $y_{k_n} \to y$. We argue by contradiction and assume that $y \in \operatorname{bdry} U$. By Lemma 2.1(ii), the function θ has an affine minorant, say a. On the other hand, Lemma 2.4(iii) implies that $D(x,y_{k_n}) \to +\infty$. Hence $-\infty \leftarrow \theta(y_{k_n}) \geq a(y_{k_n}) \to a(y) > -\infty$, which is contradictory. ■

Lemma 2.12 [10, Lemma 2.20] or [7, Section 4.1] *Suppose that $\emptyset \neq C \subset U$ and $(y_n)_{n\in\mathbb{N}}$ is a sequence in U which is* Bregman monotone *with respect to C, i.e.,*

$$(\forall x \in C)(\forall n \in \mathbb{N}) \quad D(x,y_{n+1}) \leq D(x,y_n). \tag{20}$$

Then $(y_n)_{n\in\mathbb{N}}$ converges to a point in C if and only if all cluster points of $(y_n)_{n\in\mathbb{N}}$ lie in C.

Lemma 2.13 *Take $\theta \in \Gamma_0(X)$ such that $\operatorname{dom}\theta \cap U \neq \emptyset$. Consider the following properties:*

(a) *$\operatorname{dom}\theta \cap U$ is bounded.*

(b) *$\inf\theta(U) > -\infty$.*

(c) *f is supercoercive, i.e., $\lim\limits_{\|x\|\to+\infty} f(x)/\|x\| = +\infty$.*

(d) *$(\forall x \in U)$ $D(x,\cdot)$ is supercoercive.*

Then:

(i) *If any of the conditions (a), (b), or (c) holds, then*

$$(\forall y \in U) \quad \theta(\cdot) + D(\cdot,y) \ \text{ is coercive} \tag{21}$$

or, equivalently,

$$\operatorname{ran} f' \subset \operatorname{int}\operatorname{dom}(f+\theta)^*. \tag{22}$$

(ii) *If any of the conditions (a), (b), or (d) holds, then*

$$(\forall x \in U) \quad \theta(\cdot) + D(x,\cdot) \ \text{ is coercive.} \tag{23}$$

Proof. By Lemma 2.1(ii), there exists $(z^*,\alpha) \in X \times \mathbb{R}$ such that $\theta(\cdot) \geq \langle z^*,\cdot\rangle + \alpha$. (a) $\Rightarrow$ (b): By Cauchy-Schwarz, we have $\inf\theta(U) = \inf\theta(\operatorname{dom}\theta \cap U) \geq -\|z^*\| \cdot \sup\|\operatorname{dom}\theta \cap U\| + \alpha > -\infty$ as $\operatorname{dom}\theta \cap U$ is bounded. (b) $\Rightarrow$ (21): Suppose to the contrary that there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in $\operatorname{dom} f$ such that $\|x_n\| \to +\infty$ and $(\theta(x_n) + D(x_n,y))_{n\in\mathbb{N}}$ is bounded. Now observe that

$\mu = \inf \theta(\operatorname{dom} f) = \inf \theta(U) > -\infty$ by (b) and Lemma 2.3. Then we arrive at the contradiction $+\infty > \sup_{n\in\mathbb{N}} \theta(x_n) + D(x_n, y) \geq \mu + \sup_{n\in\mathbb{N}} D(x_n, y) = +\infty$ since, by Lemma 2.4(ii), $D(\cdot, y)$ is coercive. (21) $\Leftrightarrow$ (22): It follows from Lemma 2.2 that (21) $\Leftrightarrow$ $(\forall y \in U)\ f'(y) \in \operatorname{int} \operatorname{dom} (f + \theta)^*$ $\Leftrightarrow$ (22). (b) $\Rightarrow$ (23): Arguing by contradiction as above, we get a sequence $(y_n)_{n\in\mathbb{N}}$ in U such that $\|y_n\| \to +\infty$ and $+\infty > \sup_{n\in\mathbb{N}} \theta(y_n) + D(x, y_n) \geq \mu + \sup_{n\in\mathbb{N}} D(x, y_n) = +\infty$ by virtue of **A5**. (c) $\Rightarrow$ (21): Letting $\|x\| \to +\infty$, we obtain

$$\theta(x) + D(x, y) \geq \big(\alpha - f(y) + \langle f'(y), y\rangle\big) + \|x\| \left(\frac{f(x)}{\|x\|} - \|z^*\| - \|f'(y)\|\right) \to +\infty. \tag{24}$$

(d) $\Rightarrow$ (23): Letting $\|y\| \to +\infty$, we obtain

$$\theta(y) + D(x, y) \geq \alpha + \|y\| \left(\frac{D(x, y)}{\|y\|} - \|z^*\|\right) \to +\infty. \quad \blacksquare \tag{25}$$

Lemma 2.14 *Let $g\colon X \to \,]-\infty, +\infty]$ be proper, coercive, and convex. Then* $\inf g(X) > -\infty$.

Proof. Set $\mu = \inf g(X)$ and take a sequence $(x_n)_{n\in\mathbb{N}}$ in X such that $g(x_n) \to \mu$. Since g is coercive, $(x_n)_{n\in\mathbb{N}}$ is bounded and therefore it has a cluster point, say $x_{k_n} \to x$. By Lemma 2.1(ii), there exists an affine minorant of g, say a. Then $\mu \leftarrow g(x_{k_n}) \geq a(x_{k_n}) \to a(x) > -\infty$. $\blacksquare$

3 Bregman envelopes and proximity operators

Definition 3.1 *Take $\theta\colon X \to \,]-\infty, +\infty]$. The* left Bregman envelope *of θ is*

$$\overleftarrow{\operatorname{env}}_\theta\colon X \to [-\infty, +\infty] : y \mapsto \inf_{x\in X} \theta(x) + D(x, y), \tag{26}$$

and the right Bregman envelope *of θ is*

$$\overrightarrow{\operatorname{env}}_\theta\colon X \to [-\infty, +\infty] : x \mapsto \inf_{y\in X} \theta(y) + D(x, y). \tag{27}$$

Let us provide two illustrations of these definitions.

Example 3.2 Suppose $f = \frac{1}{2}\|\cdot\|^2$ and take $\theta\colon X \to \,]-\infty, +\infty]$. Then $D\colon (x, y) \mapsto \frac{1}{2}\|x - y\|^2$ and $\overleftarrow{\operatorname{env}}_\theta = \overrightarrow{\operatorname{env}}_\theta = \theta \,\square\, (\frac{1}{2}\|\cdot\|^2)$ is the *Moreau envelope* of θ [42, Section 1.G].

Example 3.3 Let C be a subset of X. The *left Bregman distance* to C is defined by

$$\overleftarrow{D}_C = \overleftarrow{\operatorname{env}}_{\iota_C} : y \mapsto \inf_{x\in C} D(x, y), \tag{28}$$

and the *right Bregman distance* to C is defined by

$$\overrightarrow{D}_C = \overrightarrow{\operatorname{env}}_{\iota_C} : x \mapsto \inf_{y\in C} D(x, y). \tag{29}$$

The following propositions collect some basic properties of Bregman envelopes.

Proposition 3.4 *Let* $\theta\colon X \to \left]-\infty,+\infty\right]$ *be such that* $\operatorname{dom}\theta \cap U \neq \emptyset$. *Then:*

(i) $\operatorname{dom} \overleftarrow{\operatorname{env}}_\theta = U$ *and* $(\forall y \in U)\ \overleftarrow{\operatorname{env}}_\theta(y) \leq \theta(y)$.

(ii) $\operatorname{dom} \overrightarrow{\operatorname{env}}_\theta = \operatorname{dom} f$ *and* $(\forall x \in U)\ \overrightarrow{\operatorname{env}}_\theta(x) \leq \theta(x)$.

(iii) *Suppose that* θ *is convex. Then* $\overleftarrow{\operatorname{env}}_\theta$ *is convex and continuous on* U. *If, in addition,* (21) *holds, i.e.,*

$$(\forall y \in U) \quad \theta(\cdot) + D(\cdot, y) \ \text{ is coercive,} \tag{30}$$

then $\overleftarrow{\operatorname{env}}_\theta$ *is proper.*

(iv) *Suppose that* θ *is convex. Then* $\overrightarrow{\operatorname{env}}_\theta$ *is convex and continuous on* U. *If, in addition,* (23) *holds, i.e.,*

$$(\forall x \in U) \quad \theta(\cdot) + D(x, \cdot) \ \text{ is coercive,} \tag{31}$$

then $\overrightarrow{\operatorname{env}}_\theta$ *is proper.*

Proof. (i) and (ii) follow at once from Definition 3.1 and (2). (iii): **A3** asserts that the function $(y,z) \mapsto \theta(z) + D(z,y)$ is convex. Hence, it follows from [42, Proposition 2.22.(a)] that the marginal function $\overleftarrow{\operatorname{env}}_\theta$ is also convex. The continuity of $\overleftarrow{\operatorname{env}}_\theta$ on U then follows from (i) and the fact that every convex function on X is continuous on the interior of its domain [41, Theorem 10.1]. It is clear from (i) that $\overleftarrow{\operatorname{env}}_\theta \not\equiv +\infty$. On the other hand it follows from (21) and Lemma 2.14 that $-\infty \notin \overleftarrow{\operatorname{env}}_\theta(X)$. (iv): Similar to (iii). ■

We now provide additional conditions guaranteeing that the infima in Definition 3.1 are uniquely attained in U.

Proposition 3.5 *Let* $\theta \in \Gamma_0(X)$ *such that* $\operatorname{dom}\theta \cap U \neq \emptyset$.

(i) *Suppose that* $y \in U$ *and that* $\theta(\cdot) + D(\cdot, y)$ *is coercive. Then there exists a unique point* $z \in U$ *such that* $\overleftarrow{\operatorname{env}}_\theta(y) = \theta(z) + D(z,y)$.

(ii) *Suppose that* $x \in U$ *and that* $\theta(\cdot) + D(x, \cdot)$ *is coercive. Then there exists a unique point* $z \in U$ *such that* $\overrightarrow{\operatorname{env}}_\theta(x) = \theta(z) + D(x,z)$.

Proof. (i): Apply [7, Proposition 3.21.(ii)], [7, Proposition 3.23.(v)(b)], and [7, Proposition 3.22.(ii)(d)]. (ii): Set $\mu = \overrightarrow{\operatorname{env}}_\theta(x)$. By Lemma 2.14, $\mu \in \mathbb{R}$. Take $(z_n)_{n\in\mathbb{N}}$ in U such that $\theta(z_n) + D(x, z_n) \to \mu$. Then by Lemma 2.11, $(z_n)_{n\in\mathbb{N}}$ has a cluster point in U, say $z_{k_n} \to z \in U$. However, by Lemma 2.6, $\theta(\cdot) + D(x,\cdot)$ is lower semicontinuous at z and therefore $\mu \leq \theta(z) + D(x,z) \leq \underline{\lim}(\theta(z_{k_n}) + D(x, z_{k_n})) = \mu$. Furthermore, **A4** implies that $\theta(\cdot) + D(x,\cdot)$ is strictly convex, which secures the uniqueness of z. ■

Remark 3.6 Suppose that $U \neq X$ (as happens for the two entropies in Example 2.5) and set $C = \operatorname{cl}(U)$ in Example 3.3. Now pick $z \in \operatorname{bdry} U$ and $(z_n)_{n\in\mathbb{N}}$ in U such that $z_n \to z$. Then $\overleftarrow{D}_C(z_n) \equiv 0$, but $\overleftarrow{D}_C(z) = +\infty$. Hence $\overleftarrow{D}_C$ is not lower semicontinuous at z. Therefore, left Bregman envelopes need not belong to $\Gamma_0(X)$.

Proposition 3.5 allows us to define the following operators on U.

Definition 3.7 *Let $\theta \in \Gamma_0(X)$ be such that* $\operatorname{dom}\theta \cap U \neq \emptyset$. *If* (21) *holds, then the* left proximity operator *associated with θ is*

$$\overleftarrow{\operatorname{prox}}_\theta \colon U \to U \colon y \mapsto \operatorname*{argmin}_{x\in X}\ \theta(x) + D(x,y). \tag{32}$$

If (23) *holds, then the* right proximity operator *associated with θ is*

$$\overrightarrow{\operatorname{prox}}_\theta \colon U \to U \colon x \mapsto \operatorname*{argmin}_{y\in X}\ \theta(y) + D(x,y). \tag{33}$$

Left (i.e., classical Bregman) proximity operators have already been used in several works, see e.g., [7], [19], [20, Chapter 3], [21], [27], and [31]. On the other hand, the notion of a right proximity operator appears to be new. We note that [28, p. 26f] discusses the formal resemblance between a multiplicative algorithm and the right proximity operator in the Kullback-Leibler divergence setting when θ is the sum of a continuous convex function and the indicator function of the nonnegative orthant in X. Further, we point out below (see Example 3.9) that in the special case of indicator functions, the right proximity operator was previously considered in [10].

Example 3.8 Suppose that $f = \frac{1}{2}\|\cdot\|^2$ and take $\theta \in \Gamma_0(X)$. Since f is supercoercive, it follows from Lemma 2.13(i) that (21) is satisfied, and we obtain *Moreau's proximity operator* [37, 38, 42]: $\overleftarrow{\operatorname{prox}}_\theta = \overrightarrow{\operatorname{prox}}_\theta = (\operatorname{Id} + \partial\theta)^{-1}$.

Example 3.9 Let $C \subset X$ be a closed convex set such that $C \cap U \neq \emptyset$. Since ι_C is bounded below, Lemma 2.13 guarantees that (21) and (23) hold; furthermore, $\overleftarrow{\operatorname{prox}}_{\iota_C} = \overleftarrow{P}_C$ is the *(left, i.e.,) classical Bregman projector* onto C [6, 12, 17, 18] and $\overrightarrow{\operatorname{prox}}_{\iota_C} = \overrightarrow{P}_C$ is the *right Bregman projector* onto C [8, 10]. Note that in the last two references, the left and right Bregman projector are called backward and forward Bregman projector. However, because of possible ambiguity in the context of splitting methods, the notions of left and right Bregman projector are preferable.

The following properties will be needed later.

Proposition 3.10 *Let $\theta \in \Gamma_0(X)$ be such that* $\operatorname{dom}\theta \cap U \neq \emptyset$.

(i) *Suppose that* (21) *holds. Then for every $(x,y) \in U^2$, the following conditions are equivalent:*

(a) $x = \overleftarrow{\operatorname{prox}}_\theta(y)$;

(b) $0 \in \partial\theta(x) + f'(x) - f'(y)$;

(c) $(\forall z \in X) \quad \langle f'(y) - f'(x), z - x\rangle + \theta(x) \leq \theta(z)$.

Moreover,

$$\overleftarrow{\operatorname{prox}}_\theta = (f' + \partial\theta)^{-1} \circ f' \tag{34}$$

is continuous on U.

(ii) *Suppose that* (23) *holds. Then for every $(x, y) \in U^2$, the following conditions are equivalent:*

(a) $y = \overrightarrow{\operatorname{prox}}_\theta(x)$;

(b) $0 \in \partial\theta(y) + f''(y)(y - x)$;

(c) $(\forall z \in X) \quad \langle f''(y)(x - y), z - y\rangle + \theta(y) \leq \theta(z)$.

Moreover, $\overrightarrow{\operatorname{prox}}_\theta$ is continuous on U.

Proof. (i): We verify only continuity as the equivalence of (a)–(c), as well as the identity (34) are known, see e.g. [7, Section 3.4]. Since f is Legendre by **A1**, it is essentially strictly convex and so is $f + \theta$. By [41, Theorem 26.3], $(f + \theta)^*$ is essentially smooth. Now Lemma 2.1(i) implies that $\nabla(f + \theta)^*$ is continuous on $\operatorname{int}\operatorname{dom}(f + \theta)^*$ and that f' is continuous on U. Therefore, $\nabla(f + \theta)^* \circ f' = \big(\partial(f + \theta)\big)^{-1} \circ f' = (f' + \partial\theta)^{-1} \circ f' = \overleftarrow{\operatorname{prox}}_\theta$ is continuous on U.

(ii): The equivalence of (a)–(c) is clear from (33) and convex calculus. To establish the continuity of $\overrightarrow{\operatorname{prox}}_\theta$ on U, pick a sequence $(x_n)_{n\in\mathbb{N}}$ in U converging to $x \in U$ and set $y_n = \overrightarrow{\operatorname{prox}}_\theta(x_n)$, for all $n \in \mathbb{N}$. Take $q \in \operatorname{dom}\theta \cap U$. Then, using Lemma 2.6, Lemma 2.9, and item (ii)(c), we obtain

$$\begin{aligned} D(x, q) &\leftarrow D(x, q) + D(x, x_n) \\ &= D_{D_f}\big((x, q), (x_n, y_n)\big) + D(x, y_n) - \big\langle f''(y_n)(x_n - y_n), q - y_n\big\rangle \\ &\geq D(x, y_n) - \big\langle f''(y_n)(x_n - y_n), q - y_n\big\rangle \\ &\geq D(x, y_n) + \theta(y_n) - \theta(q). \end{aligned} \tag{35}$$

It follows that $\big(\theta(y_n) + D(x, y_n)\big)_{n\in\mathbb{N}}$ is bounded. By (23) and Lemma 2.11, the sequence $(y_n)_{n\in\mathbb{N}}$ is bounded and its cluster points belong to U. Let us extract a converging subsequence, say $y_{k_n} \to y \in U$. In view of item (ii)(c), we have

$$(\forall z \in X)(\forall n \in \mathbb{N}) \quad \big\langle f''(y_{k_n})(x_{k_n} - y_{k_n}), z - y\big\rangle + \theta(y_{k_n}) \leq \theta(z). \tag{36}$$

We let n tend to $+\infty$ in (36), use continuity of f'' (see **A2**) and lower semicontinuity of θ to obtain

$$(\forall z \in X) \quad \big\langle f''(y)(x - y), z - y\big\rangle + \theta(y) \leq \theta(z). \tag{37}$$

The equivalence between items (ii)(a) and (ii)(c) now results in $y = \overrightarrow{\operatorname{prox}}_\theta(x)$. ■

Remark 3.11 The proof of continuity of $\overleftarrow{\operatorname{prox}}_\theta$ presented above extends Lewis' unpublished proof [33] of continuity of $\overleftarrow{P}_C$, where $C \subset X$ is a closed convex set such that $C \cap U \neq \emptyset$.

Proposition 3.12 *Let $\theta \in \Gamma_0(X)$ be such that* $\operatorname{dom}\theta \cap U \neq \emptyset$.

(i) *If* (21) *holds, then* $\overleftarrow{\operatorname{env}}_\theta$ *is differentiable on* U *and*

$$(\forall y \in U) \quad \nabla\, \overleftarrow{\operatorname{env}}_\theta(y) = f''(y)(y - \overleftarrow{\operatorname{prox}}_\theta(y)). \tag{38}$$

(ii) *If* (23) *holds, then* $\overrightarrow{\operatorname{env}}_\theta$ *is differentiable on* U *and*

$$(\forall x \in U) \quad \nabla\, \overrightarrow{\operatorname{env}}_\theta(x) = f'(x) - f'(\overrightarrow{\operatorname{prox}}_\theta(x)). \tag{39}$$

Proof. (i): Fix $y \in U$, $h \in X$, and $t \in \,]0,+\infty[$ such that $y + th \in U$. For the sake of brevity, set $P = \overleftarrow{\operatorname{prox}}_\theta$. Using Lemma 2.6 twice, we estimate

$$\begin{aligned}
&D(y, y+th) + \big\langle f'(y+th) - f'(y), y - P(y+th)\big\rangle \\
&\qquad = D(P(y+th), y+th) - D(P(y+th), y) \\
&\qquad = \theta(P(y+th)) + D(P(y+th), y+th) - \theta(P(y+th)) - D(P(y+th), y) \\
&\qquad \leq \theta(P(y+th)) + D(P(y+th), y+th) - \theta(P(y)) - D(P(y), y) \\
&\qquad = \overleftarrow{\operatorname{env}}_\theta(y+th) - \overleftarrow{\operatorname{env}}_\theta(y) \\
&\qquad \leq \theta(P(y)) + D(P(y), y+th) - \theta(P(y)) - D(P(y), y) \\
&\qquad = D(P(y), y+th) - D(P(y), y) \\
&\qquad = D(y, y+th) + \langle f'(y+th) - f'(y), y - P(y)\rangle.
\end{aligned}$$

After dividing this chain of inequalities by t, we take the limit as $t \to 0^+$. Lemma 2.8, **A2**, and the continuity of P (see Proposition 3.10(i)) imply that the leftmost limit is the same as the rightmost limit, namely $\langle f''(y)(h), y - P(y)\rangle$. It follows that

$$\lim_{t\to 0^+} \frac{\overleftarrow{\operatorname{env}}_\theta(y+th) - \overleftarrow{\operatorname{env}}_\theta(y)}{t} = \langle f''(y)(h), y - P(y)\rangle. \tag{40}$$

(ii): Fix $x \in U$, $h \in X$, and $t \in \,]0,+\infty[$ such that $x + th \in U$. We set $P = \overrightarrow{\operatorname{prox}}_\theta$ and obtain, using Lemma 2.6 twice,

$$\begin{aligned}
&D(x+th, x) + \langle f'(x) - f'(P(x+th)), th\rangle \\
&\qquad = D(x+th, P(x+th)) - D(x, P(x+th)) \\
&\qquad = \theta(P(x+th)) + D(x+th, P(x+th)) - \theta(P(x+th)) - D(x, P(x+th)) \\
&\qquad \leq \theta(P(x+th)) + D(x+th, P(x+th)) - \theta(P(x)) - D(x, P(x)) \\
&\qquad = \overrightarrow{\operatorname{env}}_\theta(x+th) - \overrightarrow{\operatorname{env}}_\theta(x) \\
&\qquad \leq \theta(P(x)) + D(x+th, P(x)) - \theta(P(x)) - D(x, P(x)) \\
&\qquad = D(x+th, P(x)) - D(x, P(x)) \\
&\qquad = D(x+th, x) + \big\langle f'(x) - f'(P(x)), th\big\rangle
\end{aligned}$$

Let us divide this chain of inequalities by t, and then take the limit as $t \to 0^+$. Lemma 2.8 and the continuity of P (see Proposition 3.10(ii)) imply that the leftmost limit is the same as the rightmost limit, namely $\langle f'(x) - f'(P(x)), h\rangle$. Thus

$$\lim_{t\to 0^+} \frac{\overrightarrow{\mathrm{env}}_\theta(x+th) - \overrightarrow{\mathrm{env}}_\theta(x)}{t} = \langle f'(x) - f'(P(x)), h\rangle. \quad \blacksquare \tag{41}$$

Remark 3.13 Special cases of Proposition 3.12(i) have been observed previously in the literature; see, for instance, [43, Theorem 4.1(b)] when $\inf \theta(U) > -\infty$ (so that (21) holds by Lemma 2.13(i)) and [20, Proposition 3.2.3] when $\operatorname{dom} \theta = X$.

Let us now provide two illustrations of Proposition 3.12.

Example 3.14 If $\theta \in \Gamma_0(X)$ and $f = \frac{1}{2}\|\cdot\|^2$, then Proposition 3.12 reduces to Moreau's gradient formula [38, Proposition 7.d], namely $\nabla\left(\theta \,\square\, (\frac{1}{2}\|\cdot\|^2)\right) = \mathrm{Id} - (\mathrm{Id} + \partial\theta)^{-1}$.

Example 3.15 Let $C \subset X$ be a closed convex set such that $C \cap U \neq \emptyset$ and take $\{x, y\} \subset U$. In view of Examples 3.3 and 3.9, setting $\theta = \iota_C$ in Proposition 3.12 yields $\nabla \overleftarrow{D}_C(y) = f''(y)(y - \overleftarrow{P}_C(y))$ and $\nabla \overrightarrow{D}_C(x) = f'(x) - f'(\overrightarrow{P}_C(x))$.

As the following proposition shows, left and right envelopes and prox operators arise naturally in connection with our basic problem (4). Let us introduce the two auxiliary relaxed problems

$$\text{minimize} \quad \overleftarrow{\mathrm{env}}_\varphi + \psi \quad \text{over} \quad U \tag{42}$$

and

$$\text{minimize} \quad \varphi + \overrightarrow{\mathrm{env}}_\psi \quad \text{over} \quad U. \tag{43}$$

Their solution sets will be denoted by

$$F = \left\{y \in U \mid \overleftarrow{\mathrm{env}}_\varphi(y) + \psi(y) = \inf(\overleftarrow{\mathrm{env}}_\varphi + \psi)(U)\right\} \tag{44}$$

and

$$E = \left\{x \in U \mid \varphi(x) + \overrightarrow{\mathrm{env}}_\psi(x) = \inf(\varphi + \overrightarrow{\mathrm{env}}_\psi)(U)\right\}, \tag{45}$$

respectively. In the standard metric setting, i.e., $f = \frac{1}{2}\|\cdot\|^2$, (42) and (43) are the two classical partial Moreau regularizations of (8), e.g., [34]. We now relate the sets E and F to the set S, defined in (5), as well as to the operators $\overleftarrow{\mathrm{prox}}_\varphi$ and $\overrightarrow{\mathrm{prox}}_\psi$.

Proposition 3.16 *The following properties hold:*

(i) *E and F are convex.*

(ii) $(\forall (x,y) \in U \times U) \quad (x,y) \in S \;\Leftrightarrow\; \big(x = \overleftarrow{\operatorname{prox}}_\varphi(y) \;\; \textit{and} \;\; y = \overrightarrow{\operatorname{prox}}_\psi(x)\big)$.

(iii) $E = \operatorname{Fix}\big(\overleftarrow{\operatorname{prox}}_\varphi \circ \overrightarrow{\operatorname{prox}}_\psi\big) \;\; \textit{and} \;\; F = \operatorname{Fix}\big(\overrightarrow{\operatorname{prox}}_\psi \circ \overleftarrow{\operatorname{prox}}_\varphi\big)$.

(iv) $(\forall (x,y) \in E \times F) \quad \big(x, \overrightarrow{\operatorname{prox}}_\psi(x)\big) \in S \;\; \textit{and} \;\; \big(\overleftarrow{\operatorname{prox}}_\varphi(y), y\big) \in S$.

Proof. (i): In view of (3) and Proposition 3.4(iii)&(iv), E and F are convex, as sets of minimizers of convex functions. For the remainder of the proof, we fix $(x,y) \in U \times U$. (ii): Since

$$\nabla D(x,y) = \big(f'(x) - f'(y), f''(y)(y-x)\big), \tag{46}$$

we obtain by standard convex calculus and by invoking items (i)(b) and (ii)(b) of Proposition 3.10 the chain of equivalences

$$\begin{aligned} (x,y) \in S \;&\Leftrightarrow\; (0,0) \in \partial\Lambda(x,y) = \big(\partial\varphi(x) + f'(x) - f'(y), \partial\psi(y) + f''(y)(y-x)\big) \\ &\Leftrightarrow\; 0 \in \partial\varphi(x) + f'(x) - f'(y) \quad \text{and} \quad 0 \in \partial\psi(y) + f''(y)(y-x) \\ &\Leftrightarrow\; x = \overleftarrow{\operatorname{prox}}_\varphi(y) \quad \text{and} \quad y = \overrightarrow{\operatorname{prox}}_\psi(x). \end{aligned} \tag{47}$$

(iii): It follows from Proposition 3.12(ii) and Proposition 3.10(i) that

$$\begin{aligned} x \in E \;&\Leftrightarrow\; 0 \in \partial(\varphi + \overrightarrow{\operatorname{env}}_\psi)(x) = \partial\varphi(x) + \nabla\, \overrightarrow{\operatorname{env}}_\psi(x) = \partial\varphi(x) + f'(x) - f'(\overrightarrow{\operatorname{prox}}_\psi(x)) \\ &\Leftrightarrow\; x = \overleftarrow{\operatorname{prox}}_\varphi \circ \overrightarrow{\operatorname{prox}}_\psi(x). \end{aligned} \tag{48}$$

Likewise, it follows from Proposition 3.12(i) and Proposition 3.10(ii) that

$$\begin{aligned} y \in F \;&\Leftrightarrow\; 0 \in \partial(\overleftarrow{\operatorname{env}}_\varphi + \psi)(y) = \nabla\, \overleftarrow{\operatorname{env}}_\varphi(y) + \partial\psi(y) = f''(y)(y - \overleftarrow{\operatorname{prox}}_\varphi(y)) + \partial\psi(y) \\ &\Leftrightarrow\; y = \overrightarrow{\operatorname{prox}}_\psi \circ \overleftarrow{\operatorname{prox}}_\varphi(y). \end{aligned} \tag{49}$$

(iv) follows at once from (ii) and (iii). ■

4 Alternating left and right proximity operators

Recall that the standing assumptions on our problem (4) are described by (3), that its solution set S and its optimal value p are defined in (5).

In (13), we proposed the following algorithm to solve (4).

$$\text{fix } x_0 \in U \quad \text{and set} \quad (\forall n \in \mathbb{N}) \quad y_n = \overrightarrow{\operatorname{prox}}_\psi(x_n) \;\text{ and }\; x_{n+1} = \overleftarrow{\operatorname{prox}}_\varphi(y_n). \tag{50}$$

In view of (3) and Definition 3.7, the sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are well-defined. We now study the asymptotic behavior of this algorithm, starting with two key monotonicity properties.

Proposition 4.1 *Let* $\big((x_n,y_n)\big)_{n\in\mathbb{N}}$ *be generated by* (50). *Then:*

$$(\forall n \in \mathbb{N}) \quad \Lambda(x_{n+1}, y_{n+1}) \leq \Lambda(x_{n+1}, y_n) \leq \Lambda(x_n, y_n). \tag{51}$$

Proof. This is a direct consequence of Definition 3.7 and (50). ■

Proposition 4.2 *Let* $\big((x_n,y_n)\big)_{n\in\mathbb{N}}$ *be generated by* (50) *and take* $\{x,y\}\subset U$. *Then:*

$$(\forall n\in\mathbb{N})\ \ D(x,x_{n+1})\leq D(x,x_n)-\Lambda(x_{n+1},y_n)+\Lambda(x,y)-D_{D_f}\big((x,y),(x_n,y_n)\big). \tag{52}$$

Proof. Fix $n\in\mathbb{N}$. If $x\notin\operatorname{dom}\varphi$ or $y\notin\operatorname{dom}\psi$, then (52) is clear. Otherwise, it follows from Lemma 2.9, Lemma 2.6, Proposition 3.10 that

$$\begin{aligned}
D(x,y)+D(x,x_n)&=D(x,y_n)+D_{D_f}\big((x,y),(x_n,y_n)\big)+\big\langle f''(y_n)(x_n-y_n),y_n-y\big\rangle\\
&=D(x,x_{n+1})+D(x_{n+1},y_n)+\big\langle f'(y_n)-f'(x_{n+1}),x_{n+1}-x\big\rangle\\
&\quad+D_{D_f}\big((x,y),(x_n,y_n)\big)+\big\langle f''(y_n)(x_n-y_n),y_n-y\big\rangle\\
&=D(x,x_{n+1})+D(x_{n+1},y_n)+D_{D_f}\big((x,y),(x_n,y_n)\big)\\
&\quad+\big\langle f'(y_n)-f'(x_{n+1}),x_{n+1}-x\big\rangle+\varphi(x)-\varphi(x_{n+1})\\
&\quad+\big\langle f''(y_n)(x_n-y_n),y_n-y\big\rangle+\psi(y)-\psi(y_n)\\
&\quad+\big(\varphi(x_{n+1})+\psi(y_n)\big)-\big(\varphi(x)+\psi(y)\big)\\
&\geq D(x,x_{n+1})+D(x_{n+1},y_n)+D_{D_f}\big((x,y),(x_n,y_n)\big)\\
&\quad+\big(\varphi(x_{n+1})+\psi(y_n)\big)-\big(\varphi(x)+\psi(y)\big).
\end{aligned}\tag{53}$$

Hence

$$\begin{aligned}
D(x,x_{n+1})-D(x,x_n)&\leq -D(x_{n+1},y_n)+D(x,y)-D_{D_f}\big((x,y),(x_n,y_n)\big)\\
&\quad-\big(\varphi(x_{n+1})+\psi(y_n)\big)+\big(\varphi(x)+\psi(y)\big)\\
&=-\Lambda(x_{n+1},y_n)+\Lambda(x,y)-D_{D_f}\big((x,y),(x_n,y_n)\big).
\end{aligned}\tag{54}$$

■

Corollary 4.3 *Let* $\big((x_n,y_n)\big)_{n\in\mathbb{N}}$ *be generated by* (50) *and suppose that* p *in* (5) *is finite. Then*

$$\lim\Lambda(x_n,y_n)=\lim\Lambda(x_{n+1},y_n)=p. \tag{55}$$

Proof. In view of Proposition 4.1, let $\lambda=\lim\Lambda(x_n,y_n)=\lim\Lambda(x_{n+1},y_n)$. Clearly, $\lambda\in[p,+\infty[$. Let us assume that $\lambda>p$ and we shall obtain a contradiction. Take $\{x,y\}\subset U$ and $\varepsilon\in\,]0,+\infty[$ such that $\lambda=\Lambda(x,y)+\varepsilon$. Then Proposition 4.2 yields

$$(\forall n\in\mathbb{N})\quad D(x,x_n)-D(x,x_{n+1})\geq\lambda-\Lambda(x,y)=\varepsilon. \tag{56}$$

It follows that $(\forall n\in\mathbb{N})\ 0\leq D(x,x_n)\leq D(x,x_0)-n\varepsilon$, which is contradictory for $n>D(x,x_0)/\varepsilon$. Therefore $\lambda=p$. ■

Our main convergence result can now be stated and proved.

Theorem 4.4 *Let* $\big((x_n,y_n)\big)_{n\in\mathbb{N}}$ *be a sequence generated by algorithm* (50) *and suppose that* S *is nonempty (and hence* p *in* (5) *is finite). Then*

$$\sum_{n\in\mathbb{N}}\big(\Lambda(x_{n+1},y_n)-p\big)<+\infty\quad\text{and}\quad\big(\forall(x,y)\in S\big)\quad\sum_{n\in\mathbb{N}}D_{D_f}\big((x,y),(x_n,y_n)\big)<+\infty. \tag{57}$$

Moreover, $\big((x_n,y_n)\big)_{n\in\mathbb{N}}$ *converges to a point in* S.

Proof. Take $(x,y) \in S$. It follows from (52) that

$$(\forall n \in \mathbb{N}) \;\; 0 \leq \big(\Lambda(x_{n+1}, y_n) - p\big) + D_{D_f}\big((x,y),(x_n,y_n)\big) \leq D(x,x_n) - D(x,x_{n+1}). \tag{58}$$

Therefore, (57) holds. Moreover, (58) and Proposition 3.16 imply that

$$(x_n)_{n\in\mathbb{N}} \text{ is Bregman monotone with respect to } E \subset U. \tag{59}$$

In view of Lemma 2.11 (with $\theta = 0$) and **A5**, the sequence $(x_n)_{n\in\mathbb{N}}$ is bounded and all its cluster points lie in U. Let us consider a convergent subsequence, say $x_{k_n} \to \widetilde{x} \in U$. Using Proposition 3.10, let us set $\widetilde{y} = \overrightarrow{\operatorname{prox}}_\psi(\widetilde{x}) = \lim \overrightarrow{\operatorname{prox}}_\psi(x_{k_n})$. Continuity of D on $U \times U$ (Lemma 2.6), lower semicontinuity of φ and ψ, and Corollary 4.3 now yield

$$\begin{aligned}
\Lambda(\widetilde{x},\widetilde{y}) &= \varphi(\widetilde{x}) + \psi(\widetilde{y}) + D(\widetilde{x},\widetilde{y}) \\
&\leq \underline{\lim}\, \varphi(x_{k_n}) + \underline{\lim}\, \psi(y_{k_n}) + \lim D(x_{k_n}, y_{k_n}) \\
&\leq \underline{\lim}\, \big(\varphi(x_{k_n}) + \psi(y_{k_n}) + D(x_{k_n}, y_{k_n})\big) \\
&= \underline{\lim}\, \Lambda(x_{k_n}, y_{k_n}) \\
&= p.
\end{aligned} \tag{60}$$

Hence $(\widetilde{x},\widetilde{y}) \in S$ and thus $\widetilde{x} \in E$ by Proposition 3.16(ii)&(iii). Therefore, every cluster point of $(x_n)_{n\in\mathbb{N}}$ belongs to E. Consequently, utilizing (59) and Lemma 2.12, we conclude that $(x_n)_{n\in\mathbb{N}}$ converges a point in E, say $\bar{x}$. Set $\bar{y} = \overrightarrow{\operatorname{prox}}_\psi(\bar{x})$. Proposition 3.16(iv) shows that $(\bar{x},\bar{y}) \in S$. On the other hand, Proposition 3.10 implies that $y_n = \overrightarrow{\operatorname{prox}}_\psi(x_n) \to \overrightarrow{\operatorname{prox}}_\psi(\bar{x}) = \bar{y}$. Altogether, $(x_n, y_n) \to (\bar{x},\bar{y}) \in S$. ■

Let us illustrate Theorem 4.4 by presenting some immediate applications; further examples will be provided in Section 5.

Corollary 4.5 *Suppose that the solution set S of the problem*

$$\textit{minimize} \quad (x,y) \mapsto \varphi(x) + \psi(y) + \tfrac{1}{2}\|x-y\|^2 \quad \textit{over} \quad X \times X \tag{61}$$

is nonempty. Then the sequence $\big((x_n,y_n)\big)_{n\in\mathbb{N}}$ generated by the algorithm

$$\textit{fix } x_0 \in X \quad \textit{and set} \quad (\forall n \in \mathbb{N}) \qquad y_n = (I+\partial\psi)^{-1}(x_n) \quad \textit{and} \quad x_{n+1} = (I+\partial\varphi)^{-1}(y_n) \tag{62}$$

converges to a point in S.

Proof. This is a consequence of Example 3.8 and Theorem 4.4. ■

Remark 4.6 For an extension of Corollary 4.5, see [1, Théorème 2(ii)] and, for further refinements, [9, Theorem 4.6]. If we specialize Corollary 4.5 to indicator functions or Corollary 4.7 to the energy, then we recover a classical result due to Cheney and Goldstein [22].

Corollary 4.7 *Let A and B be closed convex sets in X such that $A \cap U \neq \emptyset$ and $B \cap U \neq \emptyset$. Suppose that the solution set S of the problem*

$$\text{minimize} \quad D \quad \text{over} \quad (A \times B) \cap (U \times U) \tag{63}$$

is nonempty. Then the sequence $\big((x_n, y_n)\big)_{n\in\mathbb{N}}$ generated by the alternating left-right projections *algorithm*

$$\text{fix } x_0 \in U \quad \text{and set} \quad (\forall n \in \mathbb{N}) \quad y_n = \overrightarrow{P}_B(x_n) \quad \text{and} \quad x_{n+1} = \overleftarrow{P}_A(y_n). \tag{64}$$

converges to a point in S.

Proof. This is a consequence of Example 3.9 and Theorem 4.4. ■

Remark 4.8 Corollary 4.7 corresponds to Csiszár and Tusnády's classical alternating minimization procedure (see their seminal work [25]) which, in turn, covers the *expectation-maximization* method for a specific Poisson model [30]. For an alternative proof of Corollary 4.7 when $A \cap B \neq \emptyset$, see [8, Application 5.5].

Corollary 4.9 *Take $\theta \in \Gamma_0(X)$ and suppose that its set M of minimizers over U is nonempty.*

(i) *If θ satisfies* (21)*, then the sequence $(z_n)_{n\in\mathbb{N}}$ generated by the* left proximal point algorithm

$$\text{fix } z_0 \in U \quad \text{and set} \quad (\forall n \in \mathbb{N}) \quad z_{n+1} = \overleftarrow{\operatorname{prox}}_\theta(z_n) \tag{65}$$

converges to a point in M.

(ii) *If θ satisfies* (23)*, then the sequence $(z_n)_{n\in\mathbb{N}}$ generated by the* right proximal point algorithm

$$\text{fix } z_0 \in U \quad \text{and set} \quad (\forall n \in \mathbb{N}) \quad z_{n+1} = \overrightarrow{\operatorname{prox}}_\theta(z_n) \tag{66}$$

converges to a point in M.

Proof. (i): Set $\varphi = \theta$ and $\psi = 0$ in Theorem 4.4. (ii): Set $\varphi = 0$ and $\psi = \theta$ in Theorem 4.4. ■

Remark 4.10 Item (i) in Corollary 4.9 goes back to [19]. A special case of item (ii) in the context of the Kullback-Leibler divergence appears in [28], see also [29, Remark 2.18]. If $f = \frac{1}{2}\|\cdot\|^2$, then items (i) and (ii) reduce to a classical result of Martinet [35].

The next two statements concern the invariance of the solution set S.

Corollary 4.11 *Take $\{(x, y), (\widetilde{x}, \widetilde{y})\} \subset S$. Then: $D_{D_f}\big((x, y), (\widetilde{x}, \widetilde{y})\big) = 0$.*

Proof. Consider the iteration (50) with starting point $x_0 = \widetilde{x}$. Using Proposition 3.16, we see that $x_n \equiv \widetilde{x}$ and $y_n \equiv \widetilde{y}$. Hence (57) yields $D_{D_f}\big((x, y), (\widetilde{x}, \widetilde{y})\big) = 0$. ■

Example 4.12 Take $\big\{(x,y),(\widetilde{x},\widetilde{y})\big\} \subset S$. Then:

(i) If f is the energy, then $y-x=\widetilde{y}-\widetilde{x}$.

(ii) If f is the Boltzmann-Shannon entropy, then $y/x=\widetilde{y}/\widetilde{x}$.

Proof. Combine Corollary 4.11, Example 2.10, and (6). ■

We now turn to an optimization problem dual to (4), namely, to determine

$$-\min_{(x^*,y^*)\in X\times X} \varphi^*(x^*)+\psi^*(y^*)+D^*(-x^*,-y^*). \tag{67}$$

This is precisely the standard Fenchel dual of the optimization problem (4) — the minus sign is simply added to ensure that the optimal values of the two problems coincide. Now (3) implies the standard constraint qualification which in the present setting states that $\operatorname{int}\operatorname{dom} D_f = U\times U$ and $\operatorname{dom}\big((x,y)\mapsto \varphi(x)+\psi(y)\big)=\operatorname{dom}\varphi\times\operatorname{dom}\psi$ possess common points. Consequently, by [41, Corollary 6.3.2 and Theorem 31.1], the minimum in (67) is always attained. More information can be obtained when (4) has solutions:

Proposition 4.13 *Suppose that $S\neq\emptyset$. Then the minimum in (67) is attained at a unique point (x^*,y^*) and, for every $(x,y)\in S$, $(x^*,y^*)=\big(f'(y)-f'(x),f''(y)(x-y)\big)$. Consequently, if $\big((x_n,y_n)\big)_{n\in\mathbb{N}}$ is generated by (50), then $\big(f'(y_n)-f'(x_n),f''(y_n)(x_n-y_n)\big)\to(x^*,y^*)$.*

Proof. Let (x^*,y^*) be a solution of (67) and take $(x,y)\in S$, i.e., (x,y) solves (4). Then, using the Fenchel Duality Theorem [41, Theorem 31.1] and the Fenchel-Young inequality, we obtain

$$\begin{aligned} 0 &= \varphi(x)+\psi(y)+D(x,y)+\varphi^*(x^*)+\psi^*(y^*)+D^*(-x^*,-y^*)\\ &\geq \langle x^*,x\rangle+\langle y^*,y\rangle+D(x,y)+D^*(-x^*,-y^*)\\ &\geq 0. \end{aligned} \tag{68}$$

Hence $D(x,y)+D^*(-x^*,-y^*)=\langle(-x^*,-y^*),(x,y)\rangle$ and thus, with the help of (46), we obtain

$$(-x^*,-y^*)=\nabla D(x,y)=\big(f'(x)-f'(y),f''(y)(y-x)\big). \tag{69}$$

The "Consequently" part follows from Theorem 4.4 and the continuity of f' and f'' (see **A2**). ■

Remark 4.14 Take $(x,y)\in S$. Proposition 4.13 asserts that $(x^*,y^*)=\big(f'(y)-f'(x),f''(y)(x-y)\big)$ is the *unique* solution of (67).

(i) If f is the energy, then
$$(x^*,y^*)=(y-x,x-y)=(x^*,-x^*).$$

(ii) If f is the Boltzmann-Shannon entropy, then
$$(x^*,y^*)=\big(\ln(y/x),x/y-1\big)=\big(x^*,\exp(-x^*)-1\big).$$

(iii) If f is the Fermi-Dirac entropy, then

$$(x^*, y^*) = \left(\ln \frac{y/x}{(1-y)/(1-x)}, \frac{x-y}{y(1-y)} \right).$$

Note that (i) and (ii) combined with the uniqueness of (x^*, y^*) lead to an alternative proof of the identities in Example 4.12.

5 Applications and connections with previous works

5.1 Preliminaries

In this section, we discuss various applications of Theorem 4.4 revolving around the basic constrained optimization problem

$$\text{minimize} \quad \theta \quad \text{over} \quad C \cap U, \tag{70}$$

where $\theta \in \Gamma_0(X)$, $\operatorname{dom} \theta \cap U \neq \emptyset$, and C is a closed convex subset of X such that $C \cap U \neq \emptyset$. We are going to consider increasingly specialized realizations of (70). First, suppose that $C = X$, that I is an ordered finite index set, and that θ can be decomposed as $\theta = \sum_{i \in I} \omega_i \theta_i$, where

$$(\forall i \in I) \;\; \theta_i \in \Gamma_0(X) \quad \text{and} \quad \operatorname{dom} \theta_i \cap U \neq \emptyset, \tag{71}$$

and the weights $\{\omega_i\}_{i \in I} \subset]0, 1]$ satisfy $\sum_{i \in I} \omega_i = 1$. Then (70) becomes

$$\text{minimize} \quad \sum_{i \in I} \omega_i \theta_i \quad \text{over} \;\; U. \tag{72}$$

In particular, if we set

$$(\forall i \in I) \;\; \theta_i = \tfrac{1}{2} \max\{0, g_i\}^2, \quad \text{where} \quad g_i \in \Gamma_0(X) \;\; \text{and} \;\; \operatorname{dom} g_i \cap U \neq \emptyset, \tag{73}$$

then (72) reduces to solving a system of convex inequalities, namely,

$$\text{find} \quad x \in U \quad \text{such that} \quad \max_{i \in I} g_i(x) \leq 0. \tag{74}$$

Furthermore, if we set $(g_i)_{i \in I} = (\iota_{S_i})_{i \in I}$, where $(S_i)_{i \in I}$ is a family of closed convex sets such that, for every $i \in I$, $S_i \cap U \neq \emptyset$, then (74) reduces to the basic convex feasibility problem

$$\text{find} \quad x \in U \cap \bigcap_{i \in I} S_i. \tag{75}$$

We shall employ a product space setup initially introduced in [40] for metric projection methods and revisited in [20, Section 5.9] in the context of feasibility problems with Bregman distances.

Denote the standard Euclidean product space X^I by $\mathbf{X}$ and write $\mathbf{x} = (x_i)_{i\in I}$, for $\mathbf{x} \in \mathbf{X}$ (hence, $\|\mathbf{x}\|^2 = \sum_{i\in I} \|x_i\|^2$). Now define

$$\begin{cases} \boldsymbol{\Delta} = \big\{(x,\ldots,x) \in \mathbf{X} \colon x \in X\big\}, \\ \mathbf{f} \colon \mathbf{X} \to \,]-\infty,+\infty] : \mathbf{x} \mapsto \sum_{i\in I} \omega_i f(x_i), \\ \mathbf{U} = U^I = \operatorname{int}\operatorname{dom}\mathbf{f}, \\ \boldsymbol{\theta} \colon \mathbf{X} \to \,]-\infty,+\infty] : \mathbf{x} \mapsto \sum_{i\in I} \omega_i \theta_i(x_i). \end{cases} \tag{76}$$

Then $\mathbf{f}$ induces a Bregman distance $\mathbf{D}$ on $\mathbf{X}$ defined by

$$(\forall (\mathbf{x},\mathbf{y}) \in \mathbf{X}\times\mathbf{X}) \quad \mathbf{D}(\mathbf{x},\mathbf{y}) = \sum_{i\in I} \omega_i D(x_i,y_i). \tag{77}$$

It is straightforward to verify that $\mathbf{f}$ satisfies **A1**–**A5**, that $\operatorname{dom}\boldsymbol{\theta} \cap \mathbf{U} \neq \emptyset$, and that $\boldsymbol{\Delta} \cap \mathbf{U} \neq \emptyset$.

When (70) is not guaranteed to have solutions, one can turn to the Bregman relaxations (42) and (43). Let us now explore these relaxed formulations and derive algorithms to solve them.

5.2 Left Bregman relaxation

Setting $\varphi = \theta$ and $\psi = \iota_C$ in (8) yields (70). Accordingly, the left relaxation of (70) with respect to θ is derived from (42) to be

$$\text{minimize} \quad \overleftarrow{\operatorname{env}}_\theta \quad \text{over} \quad C \cap U. \tag{78}$$

A direct application of Theorem 4.4 and Proposition 3.16 then yields the following result.

Proposition 5.1 *Suppose that* (21) *holds and that the solution set F of* (78) *is nonempty. Then the sequence $(y_n)_{n\in\mathbb{N}}$ generated by*

$$y_0 \in U \quad \textit{and} \quad (\forall n \in \mathbb{N}) \quad y_{n+1} = \big(\overrightarrow{P}_C \circ \overleftarrow{\operatorname{prox}}_\theta\big)(y_n) \tag{79}$$

converges to a point in F.

Next, we consider the problem

$$\text{minimize} \quad \sum_{i\in I} \omega_i \overleftarrow{\operatorname{env}}_{\theta_i} \quad \text{over} \;\; U \tag{80}$$

as a left relaxation of (72) under the assumption that the functions $(\theta_i)_{i\in I}$ satisfy (21) (hence so does $\boldsymbol{\theta}$). The convenience of the product space setup of (76) becomes apparent in the following result.

Proposition 5.2 *Let $(\mathbf{x},\mathbf{y}) \in \mathbf{U}^2$. Then:*

(i) $\overleftarrow{\operatorname{env}}_{\boldsymbol{\theta}}(\mathbf{y}) = \sum_{i\in I}\omega_i\,\overleftarrow{\operatorname{env}}_{\theta_i}(y_i)$.

(ii) $\overrightarrow{P}_{\boldsymbol{\Delta}}(\mathbf{x}) = (z,\ldots,z)$, *where* $z = \sum_{i\in I}\omega_i x_i$.

(iii) $\overleftarrow{\operatorname{prox}}_{\boldsymbol{\theta}}(\mathbf{y}) = \big(\overleftarrow{\operatorname{prox}}_{\theta_i}(y_i)\big)_{i\in I}$.

(iv) $\operatorname{Fix}\overrightarrow{P}_{\boldsymbol{\Delta}}\circ\overleftarrow{\operatorname{prox}}_{\boldsymbol{\theta}} = \big\{(z,\cdots,z)\colon z \text{ solves } (80)\big\}$.

Proof. (i): By definition,

$$\begin{aligned}\overleftarrow{\operatorname{env}}_{\boldsymbol{\theta}}(\mathbf{y}) &= \inf_{\mathbf{x}\in\mathbf{X}}\ \boldsymbol{\theta}(\mathbf{x}) + \mathbf{D}(\mathbf{x},\mathbf{y})\\ &= \inf_{\mathbf{x}\in\mathbf{X}}\ \sum_{i\in I}\omega_i\big(\theta_i(y_i) + D(x_i,y_i)\big)\\ &= \sum_{i\in I}\omega_i\inf_{x_i\in X}\ \theta_i(y_i) + D(x_i,y_i)\\ &= \sum_{i\in I}\omega_i\,\overleftarrow{\operatorname{env}}_{\theta_i}(y_i).\end{aligned} \tag{81}$$

(ii): See [8, Example 3.16(ii)]. (iii): It follows from Proposition 3.10(i) that

$$\begin{aligned}\mathbf{x} = \overleftarrow{\operatorname{prox}}_{\boldsymbol{\theta}}(\mathbf{y}) &\Leftrightarrow \mathbf{0}\in\partial\boldsymbol{\theta}(\mathbf{x}) + \mathbf{f}'(\mathbf{x}) - \mathbf{f}'(\mathbf{y}) = \big(\omega_i\partial\theta_i(x_i)\big)_{i\in I} + \big(\omega_i f'(x_i) - \omega_i f'(y_i)\big)_{i\in I}\\ &\Leftrightarrow (\forall i\in I)\ \ 0\in\partial\theta_i(x_i) + f'(x_i) - f'(y_i)\\ &\Leftrightarrow (\forall i\in I)\ \ x_i = \overleftarrow{\operatorname{prox}}_{\theta_i}(y_i).\end{aligned} \tag{82}$$

(iv): Since $\operatorname{Fix}\overrightarrow{P}_{\boldsymbol{\Delta}}\circ\overleftarrow{\operatorname{prox}}_{\boldsymbol{\theta}}\subset\boldsymbol{\Delta}\cap\mathbf{U}$, let us fix $\mathbf{z}\in\boldsymbol{\Delta}\cap\mathbf{U}$, say $\mathbf{z} = (z,\cdots,z)$, where $z\in U$. Then it follows from (ii), (iii), **A1**, **A2**, and (38) that

$$\begin{aligned}\mathbf{z}\in\operatorname{Fix}\overrightarrow{P}_{\boldsymbol{\Delta}}\circ\overleftarrow{\operatorname{prox}}_{\boldsymbol{\theta}} &\Leftrightarrow z = \sum_{i\in I}\omega_i\overleftarrow{\operatorname{prox}}_{\theta_i}(z)\\ &\Leftrightarrow 0 = f''(z)\left(\sum_{i\in I}\omega_i\big(z - \overleftarrow{\operatorname{prox}}_{\theta_i}(z)\big)\right) = \sum_{i\in I}\omega_i f''(z)\big(z - \overleftarrow{\operatorname{prox}}_{\theta_i}(z)\big)\\ &\Leftrightarrow 0 = \sum_{i\in I}\omega_i\nabla\,\overleftarrow{\operatorname{env}}_{\theta_i}(z) = \nabla\left(\sum_{i\in I}\omega_i\,\overleftarrow{\operatorname{env}}_{\theta_i}\right)(z)\\ &\Leftrightarrow\ z\ \text{ solves (80).}\quad\blacksquare\end{aligned} \tag{83}$$

Important conclusions can be drawn from the above proposition. First, item (i) asserts that Problem (80) in X is equivalent to

$$\text{minimize}\quad \overleftarrow{\operatorname{env}}_{\boldsymbol{\theta}}\quad\text{over}\ \ \boldsymbol{\Delta}\cap\mathbf{U} \tag{84}$$

in $\mathbf{X}$. This is a special case of (78) for which Algorithm (79) becomes

$$\mathbf{y}_0\in\mathbf{U}\quad\text{and}\quad(\forall n\in\mathbb{N})\quad\mathbf{y}_{n+1} = \big(\overrightarrow{P}_{\boldsymbol{\Delta}}\circ\overleftarrow{\operatorname{prox}}_{\boldsymbol{\theta}}\big)(\mathbf{y}_n). \tag{85}$$

A direct application of Propositions 5.1 and 3.16 shows that $(\mathbf{y}_n)_{n\in\mathbb{N}}$ converges to a fixed point of $\overrightarrow{P}_{\boldsymbol{\Delta}} \circ \overleftarrow{\operatorname{prox}}_{\boldsymbol{\theta}}$, provided that such a point exists. In view of Proposition 5.2(ii)–(iv), we therefore obtain the following proposition.

Proposition 5.3 *Suppose that the solution set G of* (80) *is nonempty and let $(y_n)_{n\in\mathbb{N}}$ be a sequence generated by*

$$y_0 \in U \quad \text{and} \qquad (\forall n \in \mathbb{N}) \quad y_{n+1} = \sum_{i\in I} \omega_i \overleftarrow{\operatorname{prox}}_{\theta_i}(y_n). \tag{86}$$

Then $(y_n)_{n\in\mathbb{N}}$ converges to a point in G.

The above result can be applied to the problem of finding relaxed solutions to the inequality problem (74) by choosing $(\theta_i)_{i\in I}$ as in (73). This approach is of special interest when (74) has no solution since the standard subgradient projection techniques that are available to solve this problem [5, 24, 32] all fail in this situation. In the particular case of the convex feasibility problem (75), the relaxed problem (80) becomes (see Example 3.3)

$$\text{minimize} \quad \sum_{i\in I} \omega_i \overleftarrow{D}_{S_i} \quad \text{over } U. \tag{87}$$

Proposition 5.3 now reduces to the following statement.

Proposition 5.4 *Suppose that the solution set G of* (87) *is nonempty and let $(y_n)_{n\in\mathbb{N}}$ be a sequence generated by*

$$y_0 \in U \quad \text{and} \qquad (\forall n \in \mathbb{N}) \quad y_{n+1} = \sum_{i\in I} \omega_i \overleftarrow{P}_{S_i}(y_n) \tag{88}$$

Then $(y_n)_{n\in\mathbb{N}}$ converges to a point in G.

5.3 Right Bregman relaxation

The left relaxation techniques developed in Section 5.2 have natural right counterparts. Since the resulting statements have largely similar proofs, we shall only highlight the main aspects of this approach.

The right relaxation of (70) with respect θ is obtained by setting $\varphi = \iota_C$ and $\psi = \theta$ in (43), which yields

$$\text{minimize} \quad \overrightarrow{\operatorname{env}}_\theta \quad \text{over} \quad C \cap U. \tag{89}$$

We derive at once from Theorem 4.4 and Proposition 3.16 the following result.

Proposition 5.5 *Suppose that* (23) *holds and that the solution set* E *of* (89) *is nonempty. Then the sequence* $(x_n)_{n\in\mathbb{N}}$ *generated by*

$$x_0 \in U \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = \big(\overleftarrow{P}_C \circ \overrightarrow{\operatorname{prox}}_\theta\big)(x_n) \tag{90}$$

converges to a point in E.

Now assume that the functions $(\theta_i)_{i\in I}$ in (71) satisfy (23) (hence so does $\boldsymbol{\theta}$). Then a right relaxation of (72) is

$$\text{minimize} \quad \sum_{i\in I} \omega_i \overrightarrow{\operatorname{env}}_{\theta_i} \quad \text{over} \;\; U. \tag{91}$$

The next two results are the right counterparts of Propositions 5.2 and 5.3.

Proposition 5.6 *Let* $(\mathbf{x},\mathbf{y}) \in \mathbf{U}^2$. *Then:*

(i) $\overrightarrow{\operatorname{env}}_{\boldsymbol{\theta}}(\mathbf{x}) = \sum_{i\in I} \omega_i \overrightarrow{\operatorname{env}}_{\theta_i}(x_i)$.

(ii) $\overleftarrow{P}_{\boldsymbol{\Delta}}(\mathbf{y}) = (z,\ldots,z)$, *where* $z = \nabla f^*\big(\sum_{i\in I} \omega_i \nabla f(y_i)\big)$.

(iii) $\overrightarrow{\operatorname{prox}}_{\boldsymbol{\theta}}(\mathbf{x}) = \big(\overrightarrow{\operatorname{prox}}_{\theta_i}(x_i)\big)_{i\in I}$.

(iv) $\operatorname{Fix} \overleftarrow{P}_{\boldsymbol{\Delta}} \circ \overrightarrow{\operatorname{prox}}_{\boldsymbol{\theta}} = \big\{(z,\cdots,z)\colon z \text{ solves } (91)\big\}$.

Proof. (i): Proceed as in Proposition 5.2(i). (ii): See [8, Example 3.16(i)]. (iii): It follows from Proposition 3.10(ii) that

$$\begin{aligned} \mathbf{y} = \overrightarrow{\operatorname{prox}}_{\boldsymbol{\theta}}(\mathbf{x}) &\Leftrightarrow \mathbf{0} \in \partial\boldsymbol{\theta}(\mathbf{y}) + \mathbf{f}''(\mathbf{y})(\mathbf{y}-\mathbf{x}) = \big(\omega_i \partial\theta_i(y_i)\big)_{i\in I} + \big(\omega_i f''(y_i)(y_i - x_i)\big)_{i\in I} \\ &\Leftrightarrow (\forall i \in I)\;\; 0 \in \partial\theta_i(y_i) + f''(y_i)(y_i - x_i) \\ &\Leftrightarrow (\forall i \in I)\;\; y_i = \overrightarrow{\operatorname{prox}}_{\theta_i}(x_i). \end{aligned} \tag{92}$$

(iv): Take $\mathbf{z} \in \boldsymbol{\Delta} \cap \mathbf{U}$, say $\mathbf{z} = (z,\cdots,z)$. Then it follows from (ii), (iii), and (39) that

$$\begin{aligned} \mathbf{z} \in \operatorname{Fix} \overleftarrow{P}_{\boldsymbol{\Delta}} \circ \overrightarrow{\operatorname{prox}}_{\boldsymbol{\theta}} &\Leftrightarrow f'(z) = \sum_{i\in I} \omega_i f'\big(\overrightarrow{\operatorname{prox}}_{\theta_i}(z)\big) \\ &\Leftrightarrow 0 = \sum_{i\in I} \omega_i \big(f'(z) - f'(\overrightarrow{\operatorname{prox}}_{\theta_i}(z))\big) \\ &\Leftrightarrow 0 = \nabla\left(\sum_{i\in I} \omega_i \overrightarrow{\operatorname{env}}_{\theta_i}\right)(z) \\ &\Leftrightarrow z \;\text{ solves } (91). \quad \blacksquare \end{aligned} \tag{93}$$

Proposition 5.7 *Suppose that the solution set G of* (91) *is nonempty and let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by*

$$x_0 \in U \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = \nabla f^* \left(\sum_{i\in I} \omega_i \, \nabla f\big(\overrightarrow{\operatorname{prox}}_{\theta_i}(x_n)\big) \right). \tag{94}$$

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

We conclude with an application of this proposition to the right Bregman relaxation of (75):

$$\text{minimize} \quad \sum_{i\in I} \omega_i \overrightarrow{D}_{S_i} \quad \text{over } U. \tag{95}$$

Proposition 5.8 *Suppose that the solution set G of* (95) *is nonempty and let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by*

$$x_0 \in U \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = \nabla f^* \left(\sum_{i\in I} \omega_i \nabla f\big(\overrightarrow{P}_{S_i}(x_n)\big) \right). \tag{96}$$

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

5.4 Connections with previous works

We conclude by providing links between the results of Sections 5.2 and 5.3 and previous works.

Remark 5.9

(i) When $f = \frac{1}{2}\|\cdot\|^2$, Algorithms (86) and (94) coincide with [40, Algorithm 3.1] (see also [34]).

(ii) When $f = \frac{1}{2}\|\cdot\|^2$, (87) and (95) reduce to the problem of minimizing a weighted sum of the squares of the distances to the sets whereas (88) and (96) reduce to the method of barycentric metric projections. This framework has been explored from different viewpoints in [4, 23, 26].

(iii) Algorithm (88) has been studied at various levels of generality in [13, 14, 17].

(iv) Algorithm (96) is discussed in [8, 16].

Acknowledgments

H. H. Bauschke's work was partially supported by the Natural Sciences and Engineering Research Council of Canada and by a NATO grant. The work of P. L. Combettes was partially supported by a NATO grant.

References

[1] F. Acker and M. A. Prestel, "Convergence d'un schéma de minimisation alternée," *Annales de la Faculté des Sciences de Toulouse, Série 5*, vol. 2, pp. 1–9, 1980.

[2] J. F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle, "Image decomposition into a bounded variation component and an oscillating component," *Journal of Mathematical Imaging and Vision,* vol. 22, 2005.

[3] A. Auslender, M. Teboulle, and S. Ben-Tiba, "A logarithmic-quadratic proximal method for variational inequalities," *Computational Optimization and Applications,* vol. 12, pp. 31–40, 1999.

[4] H. H. Bauschke and J. M. Borwein, "On the convergence of von Neumann's alternating projection algorithm for two sets," *Set-Valued Analysis,* vol. 1, pp. 185–212, 1993.

[5] H. H. Bauschke and J. M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Review*, vol. 38, pp. 367–426, 1996.

[6] H. H. Bauschke and J. M. Borwein, "Legendre functions and the method of random Bregman projections," *Journal of Convex Analysis*, vol. 4, pp. 27–67, 1997.

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

[8] H. H. Bauschke and P. L. Combettes, "Iterating Bregman retractions," *SIAM Journal on Optimization*, vol. 13, pp. 1159–1173, 2003.

[9] H. H. Bauschke, P. L. Combettes, and S. Reich, "The asymptotic behavior of the composition of two resolvents," *Nonlinear Analysis*, vol. 60, pp. 283–301, 2005.

[10] H. H. Bauschke and D. Noll, "The method of forward projections," *Journal of Nonlinear and Convex Analysis*, vol. 3, pp. 191–205, 2002.

[11] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods,* Prentice-Hall, New Jersey, 1989.

[12] 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," *U.S.S.R. Computational Mathematics and Mathematical Physics*, vol. 7, pp. 200–217, 1967.

[13] D. Butnariu, Y. Censor, and S. Reich, "Iterative averaging of entropic projections for solving stochastic convex feasibility problems," *Computational Optimization and Applications*, vol. 8, pp. 21–39, 1997.

[14] D. Butnariu, A. N. Iusem, and R. S. Burachik, "Iterative methods of solving stochastic convex feasibility problems and applications," *Computational Optimization and Applications*, vol. 15, pp. 269–307, 2000.

[15] D. Butnariu and A. N. Iusem, *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization,* Kluwer, Boston, MA, 2000.

[16] C. Byrne and Y. Censor, "Proximity function minimization using multiple Bregman projections, with applications to entropy optimization and Kullback-Leibler distance minimization," unpublished research report, `http://math.haifa.ac.il/yair/bc29b070699.ps` (June 7, 1999).

[17] C. Byrne and Y. Censor, "Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback-Leibler distance minimization," *Annals of Operations Research*, vol. 105, pp. 77–98, 2001.

[18] Y. Censor and G. T. Herman, "Block-iterative algorithms with underrelaxed Bregman projections," *SIAM Journal on Optimization,* vol. 13, pp. 283–297, 2002.

[19] Y. Censor and S. A. Zenios, "Proximal minimization algorithm with D-functions," *Journal of Optimization Theory and Applications*, vol. 73, pp. 451–464, 1992.

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

[21] G. Chen and M. Teboulle, "Convergence analysis of a proximal-like minimization algorithm using Bregman functions," *SIAM Journal on Optimization*, vol. 3, pp. 538–543, 1993.

[22] W. Cheney and A. A. Goldstein, "Proximity maps for convex sets," *Proceedings of the American Mathematical Society*, vol. 10, pp. 448–450, 1959.

[23] P. L. Combettes, "Inconsistent signal feasibility problems: Least-squares solutions in a product space," *IEEE Transactions on Signal Processing,* vol. 42, pp. 2955–2966, 1994.

[24] P. L. Combettes, "Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections," *IEEE Transactions on Image Processing,* vol. 6, pp. 493–506, 1997.

[25] I. Csiszár and G. Tusnády, "Information geometry and alternating minimization procedures," *Statistics and Decisions* (Supplement 1), pp. 205–237, 1984.

[26] A. R. De Pierro and A. N. Iusem, "A parallel projection method for finding a common point of a family of convex sets," *Pesquisa Operacional,* vol. 5, pp. 1–20, 1985.

[27] J. Eckstein, "Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming," *Mathematics of Operations Research*, vol. 18, pp. 202–226, 1993.

[28] P. P. B. Eggermont, "Multiplicative iterative algorithms for convex programming," *Linear Algebra and its Applications*, vol. 130, pp. 25–42, 1990.

[29] P. P. B. Eggermont and V. N. LaRiccia, "On EM-like algorithms for minimum distance estimation," unpublished research report, March 1998.
`http://www.udel.edu/FREC/eggermont/Preprints/emlike.pdf`.

[30] A. N. Iusem, "A short convergence proof of the EM algorithm for a specific Poisson model," *Revista Brasileira de Probabilidade e Estatística*, vol. 6, pp. 57–67, 1992.

[31] K. C. Kiwiel, "Proximal minimization methods with generalized Bregman functions," *SIAM Journal on Control and Optimization*, vol. 35, pp. 1142–1168, 1997.

[32] K. C. Kiwiel and B. Łopuch, "Surrogate projection methods for finding fixed points of firmly nonexpansive mappings," *SIAM Journal on Optimization*, vol. 7, pp. 1084–1102, 1997.

[33] A. S. Lewis, personal communication, 1997.

[34] P. Mahey and D. T. Pham, "Partial regularization of the sum of two maximal monotone operators," *RAIRO Modélisation Mathématique et Analyse Numérique,* vol. 27, pp. 375–392, 1993.

[35] B. Martinet, "Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox," *Comptes Rendus de l'Académie des Sciences de Paris,* vol. A274, pp. 163–165, 1972.

[36] B. Mercier, *Inéquations Variationnelles de la Mécanique* (Publications Mathématiques d'Orsay, no. 80.01). Orsay, France, Université de Paris-XI, 1980.

[37] J.-J. Moreau, "Propriétés des applications prox," *Comptes Rendus de l'Académie des Sciences de Paris,* vol. A256, pp. 1069–1071, 1963.

[38] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bulletin de la Société Mathématique de France,* vol. 93, pp. 273–299, 1965.

[39] J.-C. Pesquet and P. L. Combettes, "Wavelet synthesis by alternating projections," *IEEE Transactions on Signal Processing,* vol. 44, pp. 728–732, 1996.

[40] G. Pierra, "Decomposition through formalization in a product space," *Mathematical Programming*, vol. 28, pp. 96–115, 1984.

[41] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[42] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis.* New York: Springer-Verlag, 1998.

[43] M. Teboulle, "Entropic proximal mappings with applications to nonlinear programming," *Mathematics of Operations Research,* vol. 17, pp. 670–690, 1992.

[44] L. A. Vese and S. J. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *Journal of Scientific Computing,* vol. 19, pp. 553–572, 2003.