

# The Bregman proximal average

Xianfu Wang\*and Heinz H. Bauschke†

Submitted August 25, 2021

Revision January 20, 2022

## Abstract

We provide a proximal average with respect to a 1-coercive Legendre function. In the sense of Bregman distance, the Bregman envelope of the proximal average is a convex combination of Bregman envelopes of individual functions. The Bregman proximal mapping of the average is a convex combination of convexified proximal mappings of individual functions. Techniques from variational analysis provide the keys for the Bregman proximal average.

**2000 Mathematics Subject Classification:** Primary 49J53, 49J52; Secondary 47H05, 90C26, 52A01.

**Keywords:** Bregman distance, Bregman envelope, Bregman proximal mapping, Bregman proximal average, Combettes-Reyes anisotropic envelope, Combettes-Reyes proximal mapping, epi-convergence, Legendre function,  $\phi$ -prox-bounded function.

## 1 Introduction

Starting from the Bauschke, Matoušková and Reich [15], proximal averages have been further studied in [14, 25, 10], and found many applications and generalizations; see, e.g., [43, 39, 30, 4, 38, 3, 29, 33, 42]. Bregman proximal mappings play important roles in the theory of optimization, best approximation, and the design of optimization algorithms; see, e.g., [6, 22, 23, 11, 8, 12, 13, 34, 26, 32, 21, 24]. An open problem in the literature is to extend the proximal average to the framework of Bregman distances. In this paper, we propose a Bregman proximal average, which unifies and significantly broadens the realm of proximal averages. It generalizes the classical proximal average from two perspectives: First the individual functions are not necessarily convex; second, the proximal mappings are considerably more general. It is surprising that the Bregman proximal average has many desirable properties in this generality. Our main results state that a convex combination of convexified Bregman proximal mappings is a Bregman proximal mapping, and that a convex combination of Bregman envelopes is a Bregman envelope. This extends [14, 25, 15, 36] to the framework of Bregman distances. Potential algorithmic consequences can be drawn from [8, 12, 24, 34].

---

\*Mathematics, Irving K. Barber School, The University of British Columbia Okanagan, Kelowna, B.C. V1V 1V7, Canada. Email: [shawn.wang@ubc.ca](mailto:shawn.wang@ubc.ca).

†Mathematics, Irving K. Barber School, The University of British Columbia Okanagan, Kelowna, B.C. V1V 1V7, Canada. Email: [heinz.bauschke@ubc.ca](mailto:heinz.bauschke@ubc.ca).

**Outline of the paper.** The paper is organized as follows. In the remainder of this section we make our setting precise. In Section 2, we collect a few basic facts and preliminary results on  $\phi$ -prox-bounded functions, the Bregman envelopes and proximal maps for possible nonconvex functions,  $\phi$ -proximal-hulls, and Combettes-Reyes anisotropic envelopes and proximal mappings. In Section 3, we propose an  $\alpha$ -weighted Bregman proximal average with parameter  $\mu$  (Bregman proximal average for short) for  $\phi$ -prox-bounded proper lower semicontinuous functions, and provide its key properties. One important consequence is that a convex combination of convexified Bregman proximal mappings is a Bregman proximal mapping. For a general Legendre function  $\phi$ , even when both functions are proper lower semicontinuous and convex, their Bregman proximal average need not be convex. Section 4 gives conditions under which the Bregman proximal average is convex. To accomplish this we provide a Bregman version of the Baillon-Haddad theorem and introduce  $\nabla\phi$ -firmly nonexpansive mappings. In Section 5, we study Fenchel duality properties of Bregman proximal averages by using Combettes and Reyes' anisotropic envelopes and proximity operators. Section 6 focuses on the relationships among arithmetic average, epi-average, and the Bregman proximal average. It is shown that the proximal hulls of individual functions are the epi-limiting instances of the Bregman proximal average when  $\alpha \downarrow 0$  or  $\alpha \uparrow 1$ . It is also shown that the arithmetic average and epi-average of convexified individual functions are the limiting instances of the Bregman proximal average for functions with  $+\infty$ -prox-bound when  $\lambda \downarrow 0$  or  $\lambda \uparrow +\infty$ .

**Notation and standing assumptions.** The notation that we employ is for the most part standard and can be found, for example, in [9, 41, 18, 31, 35]; however, a partial list is provided for the reader's convenience. Throughout,  $\mathbb{R}^n$  is the standard Euclidean space with inner product  $\langle \cdot, \cdot \rangle$  and induced norm  $\|\cdot\|$ . The set of proper lower semicontinuous convex functions from  $\mathbb{R}^n$  to  $] -\infty, +\infty]$  is denoted by  $\Gamma_0(\mathbb{R}^n)$ . For a set  $C \subseteq \mathbb{R}^n$ , its closure, convex hull, closed convex hull, interior and relative interior are denoted by  $\text{cl } C$ ,  $\text{conv } C$ ,  $\text{cl conv } C$ ,  $\text{int } C$  and  $\text{ri } C$ , respectively. The indicator function of  $C$  is  $\iota_C : \mathbb{R}^n \rightarrow ] -\infty, +\infty]$  given by  $\iota_C(x) = 0$  if  $x \in C$ , and  $+\infty$  if  $x \notin C$ . For a function  $f : \mathbb{R}^n \rightarrow [ -\infty, +\infty]$ , its lower semicontinuous hull, convex hull, and closed convex hull are denoted by  $\text{cl } f$ ,  $\text{conv } f$  and  $\text{cl conv } f$ , respectively. The effective domain of  $f$  is  $\text{dom } f := \{x \in \mathbb{R}^n \mid f(x) < +\infty\}$ . The Fenchel conjugate of  $f$  is  $f^*(y) = \sup_{x \in \mathbb{R}^n} (\langle y, x \rangle - f(x))$  for every  $y \in \mathbb{R}^n$ . The epi-multiplication of  $f$  by  $\lambda \in [0, +\infty[$  is defined by

$$(1) \quad \lambda \star f := \begin{cases} \lambda f(\cdot/\lambda), & \text{if } \lambda > 0; \\ \iota_{\{0\}}, & \text{if } \lambda = 0. \end{cases}$$

**Definition 1.1** *Let  $\phi \in \Gamma_0(\mathbb{R}^n)$  be differentiable on  $U := \text{int dom } \phi \neq \emptyset$ . The Bregman distance associated with  $\phi$  is defined by*

$$(2) \quad D_\phi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, +\infty] : (x, y) \mapsto \begin{cases} \phi(x) - \phi(y) - \langle \nabla\phi(y), x - y \rangle, & \text{if } y \in U; \\ +\infty, & \text{otherwise.} \end{cases}$$

In this paper, our standing assumptions on  $\phi$  are:

**A1**  $\phi \in \Gamma_0(\mathbb{R}^n)$  is of Legendre type, i.e.,  $\phi$  is essentially smooth and essentially strictly convex in the sense of [40, Section 26].

**A2**  $\phi$  is 1-coercive, i.e.,  $\lim_{\|x\| \rightarrow +\infty} \phi(x)/\|x\| = +\infty$ . An equivalent requirement is  $\text{dom } \phi^* = \mathbb{R}^n$  (see, e.g., [41, Theorem 11.8(d)]).

Let  $f : \mathbb{R}^n \rightarrow ] -\infty, +\infty]$  be proper and lower semicontinuous. We shall need two types of envelopes and proximal mappings of  $f$ : Bregman envelopes and proximal mappings [32, 13], and Combettes-Reyes anisotropic envelopes and proximal mappings [28].

**Definition 1.2** For  $\lambda \in ]0, +\infty[$ , the left Bregman envelope function to  $f$  is defined by

$$(3) \quad \overleftarrow{\text{env}}_{\lambda}^{\phi} f : \mathbb{R}^n \rightarrow [-\infty, +\infty] : y \mapsto \inf_{x \in \mathbb{R}^n} \left( f(x) + \frac{1}{\lambda} D_{\phi}(x, y) \right),$$

and the left Bregman proximal map of  $f$  is

$$(4) \quad \overleftarrow{\text{prox}}_{\lambda}^{\phi} f : U \rightrightarrows U : y \mapsto \underset{x \in \mathbb{R}^n}{\text{argmin}} \left( f(x) + \frac{1}{\lambda} D_{\phi}(x, y) \right).$$

The right Bregman envelope and right Bregman proximal mapping of  $f$  are defined analogously and denoted by  $\overrightarrow{\text{env}}_{\lambda}^{\phi} f$  and  $\overrightarrow{\text{prox}}_{\lambda}^{\phi} f$ , respectively.

**Definition 1.3** The Combettes-Reyes anisotropic envelope of  $f$  is defined by

$$(5) \quad f \square \phi : \mathbb{R}^n \rightarrow [-\infty, +\infty] : x \mapsto \inf_{y \in \mathbb{R}^n} (f(y) + \phi(x - y)),$$

and the Combettes-Reyes anisotropic proximal map of  $f$  is

$$\text{aprox}_f^{\phi} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n : x \mapsto \underset{y \in \mathbb{R}^n}{\text{argmin}} (f(y) + \phi(x - y)).$$

When  $\phi(x) = (1/2)\|x\|^2$ ,  $D_{\phi}(x, y) = (1/2)\|x - y\|^2$ , both types of envelopes reduce to the classical Moreau envelope [41]. For a general  $\phi$ , even if  $f \in \Gamma_0(\mathbb{R}^n)$ , the Bregman envelope  $\overleftarrow{\text{env}}_{\lambda}^{\phi} f$  might not be convex, although the anisotropic envelope  $f \square \phi$  is always convex.

**Example 1.4** Let  $\lambda := 1$ ,  $f := \iota_{\{1\}}$  on  $\mathbb{R}$ .

- (i) For  $\phi(x) = |x|^3$ , we have  $(\forall y > 0)$   $\overleftarrow{\text{env}}_1^{\phi} f(y) = 1/3 + 2y^3/3 - y^2$ , which is not convex on  $(0, +\infty)$ .
- (ii) For  $\phi(x) = -\ln x + x^2/2$  if  $x > 0$  and  $+\infty$  otherwise, we have  $(\forall y > 0)$   $\overleftarrow{\text{env}}_1^{\phi} f(y) = \ln y + 1/y + (1 - y)^2/2 - 1$ , which is not convex.

## 2 Auxiliary results on envelopes and proximal mappings

In this section, we will collect some key facts and preliminary results of Bregman envelopes and proximal mappings, as well as Combettes-Reyes anisotropic envelope and proximal mappings. Throughout this section,  $f : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  is proper lower semicontinuous and satisfies  $\text{dom } f \cap \text{dom } \phi \neq \emptyset$ .

### 2.1 $\phi$ -prox-boundedness

**Definition 2.1** A function  $f : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  is  $\phi$ -prox-bounded (prox-bounded for short) if there exists  $\lambda > 0$  such that  $\overleftarrow{\text{env}}_{\lambda}^{\phi} f(x) > -\infty$  for some  $x \in \mathbb{R}^n$ . The supremum of all such  $\lambda$  is the threshold  $\lambda_f$  of the prox-boundedness.

Prox-boundedness is crucial to ensure pleasant properties for both the Bregman envelope and proximal mapping.

**Fact 2.2** Let  $f : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  be proper lower semicontinuous with prox-bound  $\lambda_f > 0$ , and let  $0 < \lambda < \lambda_f$ . Then

- (i)  $\overleftarrow{\text{env}}_\lambda^\phi f$  is proper lower semicontinuous on  $\mathbb{R}^n$ , and continuous on  $U$ .
- (ii)  $\overleftarrow{\text{prox}}_\lambda^\phi f$  is nonempty compact valued and upper semicontinuous on  $U$ .

*Proof.* (i)&(ii): See [32, Theorem 2.2, Corollary 2.2], [26, Theorem 3.10, 3.16]. ■

The following result extends [32, Theorem 2.5], in which Kan and Song proved the result on  $\text{dom } f \cap U$  when  $\phi$  is strictly convex. As in [19], an essentially strictly convex function need not be strictly convex.

**Proposition 2.3** Let  $f : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  be proper lower semicontinuous with prox-bound  $\lambda_f > 0$ , and let  $0 < \lambda < \lambda_f$ . Then  $(\forall x \in U) \lim_{\lambda \downarrow 0} \overleftarrow{\text{env}}_\lambda^\phi f(x) = f(x)$ .

*Proof.* In view of [6, Theorem 3.7(iv)], for  $y \in U$ ,  $D_\phi(x, y) = 0 \Leftrightarrow x = y$ . When  $y \in \text{dom } f \cap U$ , the same arguments as in the proof of [32, Theorem 2.5] shows that  $\lim_{\lambda \downarrow 0} \overleftarrow{\text{env}}_\lambda^\phi f(x) = f(x)$ . When  $y \in U \setminus \text{dom } f$ ,  $f(y) = +\infty$ , it suffices to show that for every sequence  $(\lambda_k)_{k \in \mathbb{N}}$  with  $\lambda_k \downarrow 0$  we have

$$(6) \quad \lim_{k \rightarrow \infty} \overleftarrow{\text{env}}_{\lambda_k}^\phi f(y) = +\infty.$$

Indeed, following the proof of [32, Theorem 2.5] we have a sequence  $(w_k)_{k \in \mathbb{N}}$  such that  $w_k \rightarrow \bar{w}$  and  $f(w_k) + \frac{1}{\lambda_k} D_\phi(w_k, y) = \overleftarrow{\text{env}}_{\lambda_k}^\phi f(y)$ . If  $\bar{w} \neq y$ , then  $D_\phi(\bar{w}, y) > 0$  and

$$(7) \quad \liminf_{k \rightarrow \infty} \overleftarrow{\text{env}}_{\lambda_k}^\phi f(y) \geq \liminf_{k \rightarrow \infty} f(w_k) + \liminf_{k \rightarrow \infty} \frac{1}{\lambda_k} D_\phi(w_k, y)$$

$$(8) \quad \geq f(\bar{w}) + D_\phi(\bar{w}, y)/0^+ = +\infty.$$

If  $\bar{w} = y$ , then  $\liminf_{k \rightarrow \infty} \overleftarrow{\text{env}}_{\lambda_k}^\phi f(y) \geq \liminf_{k \rightarrow \infty} f(w_k) \geq f(\bar{w}) = +\infty$ . Hence, (6) holds. ■

The threshold of prox-boundedness has the following useful characterization, which complements [34, Proposition 3.1].

**Proposition 2.4** The following hold:

- (i) If  $f$  is prox-bounded with threshold  $\lambda_f > 0$ , then for every  $\lambda \in ]0, \lambda_f[$  the function  $f + \frac{1}{\lambda} \phi$  is bounded below. Consequently, for every  $\lambda \in ]0, \lambda_f[$  the function  $f + \frac{1}{\lambda} \phi$  is 1-coercive.
- (ii) If there exists  $\ell > 0$  such that for every  $\lambda \in ]0, \ell[$  the function  $f + \frac{1}{\lambda} \phi$  is bounded below, then  $\lambda_f \geq \ell$ .
- (iii) Define  $\ell_f := \sup \left\{ \ell > 0 : (\forall \lambda \in ]0, \ell[) \inf \left( f + \frac{1}{\lambda} \phi \right) > -\infty \right\}$ . Then  $\ell_f = \lambda_f$ .

*Proof.* We follow the proof idea of [34, Proposition 3.5]. Because  $\phi$  is 1-coercive and Legendre, we have  $\nabla \phi^*(0) \in U$ .

(i): For every  $\lambda \in ]0, \lambda_f[$ , one has  $\overleftarrow{\text{env}}_\lambda^\phi f(\nabla \phi^*(0)) > -\infty$ . This gives

$$(\forall w \in \mathbb{R}^n) f(w) + \frac{1}{\lambda} \phi(w) \geq \frac{1}{\lambda} \phi(\nabla \phi^*(0)) + \overleftarrow{\text{env}}_\lambda^\phi f(\nabla \phi^*(0)),$$

which implies  $f + \frac{1}{\lambda}\phi$  is bounded below. Now every  $\tilde{\lambda} \in ]0, \lambda_f[$  and take  $\lambda \in ]\tilde{\lambda}, \lambda_f[$ . Since  $f + \frac{1}{\lambda}\phi$  is bounded below,  $1/\lambda < 1/\tilde{\lambda}$ ,  $\phi$  is 1-coercive, and  $f + \frac{1}{\lambda}\phi = f + \frac{1}{\tilde{\lambda}}\phi + \left(\frac{1}{\lambda} - \frac{1}{\tilde{\lambda}}\right)\phi$ , we conclude that  $f + \frac{1}{\lambda}\phi$  is 1-coercive.

(ii): For every  $\lambda \in ]0, \ell[$ , we have  $\overleftarrow{\text{env}}_{\lambda}^{\phi} f(\nabla\phi^*(0)) = \inf_{w \in \mathbb{R}^n} \left( f(w) + \frac{1}{\lambda}\phi(w) \right) - \frac{1}{\lambda}\phi(\nabla\phi^*(0)) > -\infty$  by the assumption. Hence  $\lambda_f \geq \ell$ .

(iii): Combine (i) and (ii). ■

**Corollary 2.5** *If a function  $f : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  is bounded below by a linear function, then  $\lambda_f = +\infty$ . In particular, this holds when  $f \in \Gamma_0(\mathbb{R}^n)$ .*

*Proof.* This is because that  $\phi$  is 1-coercive. When  $f \in \Gamma_0(\mathbb{R}^n)$ ,  $f$  is bounded below by a linear functional by the Brondsted-Rockafellar theorem, see, e.g., [9, Theorem 16.58]. ■

## 2.2 Properties of the Bregman envelope and proximal mapping

The following is a slightly refined version of [32, Theorem 2.4].

**Fact 2.6** *Let  $f : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  be proper lower semicontinuous with prox-bound  $\lambda_f > 0$ , and let  $0 < \lambda < \lambda_f$ . Then the following hold:*

$$(i) \quad \overleftarrow{\text{env}}_{\lambda}^{\phi} f = \left( \frac{\phi^* - (\lambda f + \phi)^*}{\lambda} \right) \circ \nabla\phi, \text{ and}$$

$$(9) \quad (\lambda f + \phi)^* = \phi^* - \lambda \overleftarrow{\text{env}}_{\lambda}^{\phi} f \circ \nabla\phi^*.$$

(ii) *If  $\nabla\phi$  is locally Lipschitz on  $U$ , then  $\overleftarrow{\text{env}}_{\lambda}^{\phi} f$  is locally Lipschitz on  $U$ .*

*Proof.* (i): The calculation given in [32, Theorem 2.4] applies to every function  $f$ . (ii): This is given by [32, Theorem 2.4]. ■

**Remark 2.7** When  $\lambda = 1$  and  $f \in \Gamma_0(\mathbb{R}^n)$ , in [28] Combettes and Reyes used the notation  $f \diamond \phi$  for  $\overleftarrow{\text{env}}_{\lambda}^{\phi} f$ , and [28, Theorem 1(i)] coincides with (9).

**Corollary 2.8** *Let  $f : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  be proper lower semicontinuous with prox-bound  $\lambda_f > 0$ , and let  $0 < \lambda < \lambda_f$ . If  $\lambda f + \phi$  is convex, then  $\lambda f + \phi = (\phi^* - \lambda \overleftarrow{\text{env}}_{\lambda}^{\phi} f \circ \nabla\phi^*)^*$ . Consequently,*

$$f = \frac{(\phi^* - \lambda \overleftarrow{\text{env}}_{\lambda}^{\phi} f \circ \nabla\phi^*)^* - \phi}{\lambda} \text{ on } \text{dom } \phi.$$

Let  $\hat{\partial}$ ,  $\partial$ , and  $\partial_C$  denote the Fréchet subdifferential, Mordukhovich limiting subdifferential, and Clarke subdifferential, respectively; see, e.g., [41, 35, 27]. While  $\hat{\partial}$ ,  $\partial$  and  $\partial_C$  are different in general, it is well-known that they coincide for proper lower semicontinuous convex functions. The following fact by Kan and Song shows that the Fréchet, limiting, and Clarke subdifferential coincide for  $-\overleftarrow{\text{env}}_{\lambda}^{\phi} f$  and they can be found by using the convex hull of the Bregman proximal mapping of  $f$ .

**Fact 2.9** [32, Theorem 3.1] *Let  $f : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  be proper lower semicontinuous with prox-bound  $\lambda_f > 0$ , and let  $0 < \lambda < \lambda_f$ . Suppose  $\phi$  is second-order continuously differentiable on  $U$ . Then on  $U$  the function  $-\overleftarrow{\text{env}}_\lambda^\phi f$  is Clarke regular, and satisfies*

$$(\forall x \in U) \hat{\partial}(-\overleftarrow{\text{env}}_\lambda^\phi f)(x) = \partial_C(-\overleftarrow{\text{env}}_\lambda^\phi f)(x) = \frac{1}{\lambda} \nabla^2 \phi(x) [\text{conv}(\overleftarrow{\text{prox}}_\lambda^\phi f(x)) - x].$$

The following result establishes the relationship between the Bregman proximal mapping of  $f$  and the limiting subdifferential of  $f$ .

**Proposition 2.10** *Let  $f : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  be proper lower semicontinuous with prox-bound  $\lambda_f > 0$ , and let  $0 < \lambda < \lambda_f$ . Then the following hold:*

(i)  $\overleftarrow{\text{prox}}_\lambda^\phi f \subseteq [\partial(\phi + \lambda f)]^{-1} \circ \nabla \phi$ . If

$$(10) \quad \partial^\infty f(y) \cap -N_{\text{dom } \phi}(y) = \{0\} \text{ for every } y \in \text{dom } \phi,$$

then  $\overleftarrow{\text{prox}}_\lambda^\phi f \subseteq (\nabla \phi + \lambda \partial f)^{-1} \circ \nabla \phi$ .

(ii) *If  $\lambda f + \phi$  is convex, then  $(\forall x \in \mathbb{R}^n)$   $\overleftarrow{\text{prox}}_\lambda^\phi f(x)$  is convex and closed, and  $\overleftarrow{\text{prox}}_\lambda^\phi f = [\partial(\phi + \lambda f)]^{-1} \circ \nabla \phi$ . If, in addition, (10) holds and  $f$  is Clarke regular, then  $\overleftarrow{\text{prox}}_\lambda^\phi f = (\nabla \phi + \lambda \partial f)^{-1} \circ \nabla \phi$ .*

(iii) *If  $f$  is convex, and  $(\text{dom } f) \cap U \neq \emptyset$ , then*

$$(11) \quad \overleftarrow{\text{prox}}_\lambda^\phi f = (\nabla \phi + \lambda \partial f)^{-1} \circ \nabla \phi = \left( \frac{1}{\lambda} \nabla \phi + \partial f \right)^{-1} \circ \left( \frac{1}{\lambda} \nabla \phi \right).$$

Moreover,  $\overleftarrow{\text{prox}}_\lambda^\phi f$  is continuous on  $U$ .

*Proof.* Consider the function  $x \mapsto \frac{1}{\lambda}(\lambda f(x) + \phi(x) - \phi(y) - \langle \nabla \phi(y), x - y \rangle)$ .

(i):  $x \in \overleftarrow{\text{prox}}_\lambda^\phi f(y)$  implies  $0 \in \partial(\lambda f + \phi)(x) - \nabla \phi(y)$ , so  $x \in [\partial(\lambda f + \phi)]^{-1}(\nabla \phi(y))$ . When (10) holds,  $\partial(\lambda f + \phi) \subseteq \lambda \partial f + \nabla \phi$ .

(ii): The convexity of  $\lambda f + \phi$  ensures that  $x \in \overleftarrow{\text{prox}}_\lambda^\phi f(y)$  if and only if  $0 \in \partial(\lambda f + \phi)(x) - \nabla \phi(y)$ , which implies  $\overleftarrow{\text{prox}}_\lambda^\phi f(y) = [\partial(\lambda f + \phi)]^{-1}(\nabla \phi(y))$ . For each fixed  $y \in U$ , being the set of minimizers of convex function  $x \mapsto \lambda f(x) + \phi(x) - \langle \nabla \phi(y), x - y \rangle$ ,  $\overleftarrow{\text{prox}}_\lambda^\phi f(y)$  is convex and closed. When (10) holds and  $f$  is Clarke regular,  $\partial(\lambda f + \phi) = \lambda \partial f + \nabla \phi$  by [41, Proposition 8.12, Corollary 10.9].

(iii): Under the assumption  $(\text{dom } f) \cap U \neq \emptyset$  (instead of (10)) the calculus rule  $\partial(\phi + \lambda f) = \partial \phi + \lambda \partial f$  holds for convex functions  $\phi$  and  $f$ ; see, e.g., [9, Corollary 16.48(ii)]. Hence (11) follows from (ii). Because  $\phi + \lambda f$  is essentially strictly convex and 1-coercive, the conjugate  $(\phi + \lambda f)^*$  is full domain and differentiable, so  $\nabla(\phi + \lambda f)^* = (\nabla \phi + \lambda \partial f)^{-1}$  is continuous on  $\mathbb{R}^n$ , see, e.g., [40, Corollary 25.5.1]. As  $\nabla \phi$  is continuous on  $U$ , we obtain that  $\overleftarrow{\text{prox}}_\lambda^\phi f$  is continuous on  $U$ . ■

**Remark 2.11** Proposition 2.10(i) is a pointwise version reformulation of [34, Lemma 3.3]. See also [8, 13] for  $\overleftarrow{\text{env}}_\lambda^\phi f$  and  $\overleftarrow{\text{prox}}_\lambda^\phi f$  when  $f \in \Gamma_0(\mathbb{R}^n)$ . In [20],  $\overleftarrow{\text{prox}}_\lambda^\phi f$  is called as a warped proximity operator.

Our next result provides a connection between  $\partial(\lambda f + \phi)^*$  and  $\overleftarrow{\text{prox}}_\lambda^\phi f$ .

**Proposition 2.12** *Let  $f : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  be proper lower semicontinuous with prox-bound  $\lambda_f > 0$ , let  $0 < \lambda < \lambda_f$ , and let  $\nabla^2\phi(x)$  be invertible for every  $x \in U$ . Then*

$$(12) \quad \partial(\lambda f + \phi)^* = \text{conv } \overleftarrow{\text{prox}}_\lambda^\phi f \circ \nabla\phi^* \text{ on } U.$$

Hence,  $\text{conv } \overleftarrow{\text{prox}}_\lambda^\phi f \circ \nabla\phi^*$  is always maximally monotone. If, in addition,  $\lambda f + \phi$  is convex, then  $\partial(\lambda f + \phi)^* = \overleftarrow{\text{prox}}_\lambda^\phi f \circ \nabla\phi^*$ .

*Proof.* By Fact 2.6, we get  $(\forall x \in U) [(\lambda f + \phi)^* - \phi^*](\nabla\phi(x)) = -\lambda \overleftarrow{\text{env}}_\lambda^\phi f(x)$ . Taking subdifferential both sides, by the chain rule [41, Theorem 10.6] and Fact 2.9, we have

$$(\forall x \in U) \nabla^2\phi(x) \partial[(\lambda f + \phi)^* - \phi^*](\nabla\phi(x)) = \nabla^2\phi(x) [\text{conv } \overleftarrow{\text{prox}}_\lambda^\phi f(x) - x]$$

from which

$$(13) \quad (\forall x \in U) \partial[(\lambda f + \phi)^* - \phi^*](\nabla\phi(x)) = \text{conv } \overleftarrow{\text{prox}}_\lambda^\phi f(x) - x,$$

because  $\nabla^2\phi(x)$  is invertible by the assumption. By the sum rule [41, Exercise 10.10],

$$\partial[(\lambda f + \phi)^* - \phi^*] = \partial(\lambda f + \phi)^* - \nabla\phi^* = \partial(\lambda f + \phi)^* - (\nabla\phi)^{-1}.$$

Thus,  $(\forall x \in U) \partial(\lambda f + \phi)^*(\nabla\phi(x)) = \text{conv } \overleftarrow{\text{prox}}_\lambda^\phi f(x)$  by (13). When  $\lambda f + \phi$  is convex,  $\overleftarrow{\text{prox}}_\lambda^\phi f$  is convex-valued by Proposition 2.10(ii), so conv is superfluous in (12). ■

### 2.3 $\lambda$ - $\phi$ -proximal hull

The  $\lambda$ - $\phi$ -proximal hull defined below extends the classical proximal hull [41, Example 1.44] ( $\phi(x) = (1/2)\|x\|^2$ ), which is a special case of the Lasry-Lions envelope [1], [41, Example 1.46].

**Definition 2.13** *For a function  $f : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  and  $\lambda > 0$ , the  $\lambda$ - $\phi$ -proximal hull ( $\lambda$ -proximal hull for short) of  $f$  is the function  $\overleftarrow{\text{hul}}_\lambda^\phi f : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  defined as the pointwise supremum of the collection of all the functions of the form  $x \mapsto c - \frac{1}{\lambda}D_\phi(x, w)$  that are majorized by  $f$ , where  $c \in \mathbb{R}, w \in U$ .*

**Proposition 2.14** *The following hold:*

- (i)  $\overleftarrow{\text{hul}}_\lambda^\phi f = -\overleftarrow{\text{env}}_\lambda^\phi(-\overleftarrow{\text{env}}_\lambda^\phi f)$ , i.e.,  $(\forall x \in \mathbb{R}^n) \overleftarrow{\text{hul}}_\lambda^\phi f(x) = \sup_{w \in U} \left( \overleftarrow{\text{env}}_\lambda^\phi f(w) - \frac{1}{\lambda}D_\phi(x, w) \right)$ . Moreover,  $\overleftarrow{\text{env}}_\lambda^\phi(\overleftarrow{\text{hul}}_\lambda^\phi f) = \overleftarrow{\text{env}}_\lambda^\phi f$ .
- (ii)  $\overleftarrow{\text{hul}}_\lambda^\phi f = (f + \frac{1}{\lambda}\phi)^{**} - \frac{1}{\lambda}\phi$ , where we use the convention  $\infty - \infty = \infty$ . If, in addition,  $f + \frac{1}{\lambda}\phi \in \Gamma_0(\mathbb{R}^n)$ , then  $\overleftarrow{\text{hul}}_\lambda^\phi f = f + \iota_{\text{dom } \phi}$ .
- (iii)  $f \geq \overleftarrow{\text{hul}}_\lambda^\phi f \geq \overleftarrow{\text{env}}_\lambda^\phi f$  on  $U$ .

*Proof.* (i): Denote  $\phi_{c,w} = c - \frac{1}{\lambda}D_\phi(\cdot, w)$ . Then  $\phi_{c,w} \leq f$  if and only if  $(\forall x \in \mathbb{R}^n) c \leq f(x) + \frac{1}{\lambda}D_\phi(x, w)$ , which means  $c \leq \overleftarrow{\text{env}}_\lambda^\phi f(w)$ . Therefore,  $\overleftarrow{\text{hul}}_\lambda^\phi f$  can be viewed as the pointwise supremum of the collection of

the functions of the form  $\overleftarrow{\text{env}}_\lambda^\phi f(w) - \frac{1}{\lambda} D_\phi(x, w)$  with  $w \in U$ . The collection of  $\phi_{c,w}$  with  $\phi_{c,w} \leq f$  is the same as the collection of all  $\phi_{c,w}$  with  $\phi_{c,w} \leq \overleftarrow{\text{hul}}_\lambda^\phi f$ . Since

$$\overleftarrow{\text{env}}_\lambda^\phi f(w) = \sup \left\{ c \mid (\forall x \in \mathbb{R}^n) c \leq f(x) + \frac{1}{\lambda} D_\phi(x, w) \right\},$$

$$\overleftarrow{\text{env}}_\lambda^\phi(\overleftarrow{\text{hul}}_\lambda^\phi f)(w) = \sup \left\{ c \mid (\forall x \in \mathbb{R}^n) c \leq \overleftarrow{\text{hul}}_\lambda^\phi f(x) + \frac{1}{\lambda} D_\phi(x, w) \right\},$$

this reveals that  $\overleftarrow{\text{env}}_\lambda^\phi f = \overleftarrow{\text{env}}_\lambda^\phi(\overleftarrow{\text{hul}}_\lambda^\phi f)$ .

(ii): By Fact 2.6 and (i), we have  $\overleftarrow{\text{hul}}_\lambda^\phi f(x) =$

$$(14) \quad \sup_{w \in \mathbb{R}^n} \left[ \left( \frac{1}{\lambda} \phi^* - \frac{1}{\lambda} (\lambda f + \phi)^* \right) \circ \nabla \phi(w) - \frac{1}{\lambda} D_\phi(x, w) \right]$$

$$(15) \quad = \sup_{w \in U} \left[ \left( \frac{1}{\lambda} \phi^* - \frac{1}{\lambda} (\lambda f + \phi)^* \right) (\nabla \phi(w)) + \frac{1}{\lambda} \phi(w) + \frac{1}{\lambda} \langle \nabla \phi(w), x - w \rangle \right] - \frac{1}{\lambda} \phi(x)$$

$$(16) \quad = \frac{1}{\lambda} \sup_{w \in U} [ -(\lambda f + \phi)^* (\nabla \phi(w)) + \phi^* (\nabla \phi(w)) + \phi(w) - \langle \nabla \phi(w), w \rangle + \langle \nabla \phi(w), x \rangle ] - \frac{1}{\lambda} \phi(x)$$

$$(17) \quad = \frac{1}{\lambda} \sup_{w \in U} [ -(\lambda f + \phi)^* (\nabla \phi(w)) + \langle \nabla \phi(w), x \rangle ] - \frac{1}{\lambda} \phi(x)$$

$$(18) \quad = \frac{1}{\lambda} (\lambda f + \phi)^{**}(x) - \frac{1}{\lambda} \phi(x) = \left( f + \frac{1}{\lambda} \phi \right)^{**}(x) - \frac{1}{\lambda} \phi(x),$$

in which we used  $\phi^* (\nabla \phi(w)) + \phi(w) = \langle \nabla \phi(w), w \rangle$  in (16), and  $\text{ran } \nabla \phi = \mathbb{R}^n$  in (17). When  $f + \frac{1}{\lambda} \phi \in \Gamma_0(\mathbb{R}^n)$ , the Fenchel-Moreau biconjugate theorem [9, Theorem 13.37] gives  $(f + \frac{1}{\lambda} \phi)^{**} = f + \frac{1}{\lambda} \phi$ .

(iii): This follows from (i) and (ii).  $\blacksquare$

## 2.4 Properties of the Combettes-Reyes envelope and proximal mapping

The following result refines and complements some results of [28].

**Proposition 2.15** *Let  $f \in \Gamma_0(\mathbb{R}^n)$ . Then the following hold:*

(i)  $\text{dom } f \square \phi = \text{dom } f + \text{dom } \phi$ , and  $f \square \phi \in \Gamma_0(\mathbb{R}^n)$  is essentially smooth, so continuously differentiable on  $\text{int } \text{dom}(f \square \phi) = \text{dom } f + U$ .

(ii)  $\text{dom } \text{aprox}_f^\phi = \text{dom } f + \text{dom } \phi$ . For every  $x \in \text{dom } f + \text{dom } \phi$ ,  $\text{aprox}_f^\phi(x)$  is single-valued.

(iii)  $\text{aprox}_f^\phi$  is continuous on  $\text{dom } f + U$ . Moreover,

$$(19) \quad (\forall x \in \text{dom } f + U) \text{aprox}_f^\phi(x) = (\text{Id} + \nabla \phi^* \circ \partial f)^{-1}(x).$$

(iv)  $\text{argmin } f \cap U = \{x \in U : \text{aprox}_{f^*}^{\phi^*}(\nabla \phi(x)) = 0\}$ .

(v) If  $\phi$  is nonnegative, and  $\phi(0) = 0$ , then

$$(20) \quad f \geq f \square \phi, \quad \inf f = \inf(f \square \phi), \quad \text{and}$$

$$(21) \quad \operatorname{argmin} f = \operatorname{argmin}(f \square \phi).$$

*Proof.* (i): Apply [9, Proposition 12.6(ii)] for  $\operatorname{dom} f \square \phi$ . Because  $f \in \Gamma_0(\mathbb{R}^n)$  and  $\phi$  is essentially smooth with  $\operatorname{dom} \phi^* = \mathbb{R}^n$ , [40, Corollary 26.3.2] shows that  $f \square \phi \in \Gamma_0(\mathbb{R}^n)$  is essentially smooth. Moreover,  $\operatorname{int} \operatorname{dom}(f \square \phi) = \operatorname{dom} f + U$  because  $\operatorname{dom} f + U \subseteq \operatorname{ri} \operatorname{dom}(f \square \phi) = \operatorname{ri} \operatorname{dom} f + \operatorname{ri} \operatorname{dom} \phi \subseteq \operatorname{dom} f + U$ .

(ii): For every  $x \in \operatorname{dom} f + \operatorname{dom} \phi$ , the function  $y \mapsto f(y) + \phi(x - y)$  is in  $\Gamma_0(\mathbb{R}^n)$ , essentially strictly convex and 1-coercive, so it has a unique minimizer.

(iii): Let  $x \in \operatorname{dom} f + U$ . We show that  $\operatorname{aprox}_f^\phi$  is continuous at  $x$ . Let  $(x_k)_{k \in \mathbb{N}}$  be an arbitrary sequence in  $\operatorname{dom} f + U$  such that  $x_k \rightarrow x$ , and let  $y_k := \operatorname{aprox}_f^\phi(x_k)$ . It suffices to show  $y_k \rightarrow \operatorname{aprox}_f^\phi(x)$ . First we show that  $(y_k)_{k \in \mathbb{N}}$  is bounded. Suppose not, after passing to a subsequence and relabelling, we can assume  $\|y_k\| \rightarrow \infty$ . Now  $f \in \Gamma_0(\mathbb{R}^n)$  ensures that  $f$  possesses a continuous minorant, say,  $f \geq \langle u, \cdot \rangle + \eta$  for some  $u \in \mathbb{R}^n$  and  $\eta \in \mathbb{R}$ . By (i) and  $(f \square \phi)(x_k) = f(y_k) + \phi(x_k - y_k)$ , we get

$$\begin{aligned} (f \square \phi)(x) &\leftarrow (f \square \phi)(x_k) = f(y_k) + \phi(x_k - y_k) \\ &\geq \langle u, y_k \rangle + \eta + \phi(x_k - y_k) \geq \|y_k\|(-\|u\| + \phi(x_k - y_k)/\|y_k\|) + \eta \\ &\rightarrow +\infty, \end{aligned}$$

which is impossible. Hence,  $(y_k)_{k \in \mathbb{N}}$  is bounded. Next we show that  $(y_k)_{k \in \mathbb{N}}$  has a unique subsequential limit, namely,  $\operatorname{aprox}_f^\phi(x)$ . Indeed, let  $(y_{k_l})_{l \in \mathbb{N}}$  be a convergent subsequence of  $(y_k)_{k \in \mathbb{N}}$  with a limit  $y \in \mathbb{R}^n$ . Since  $f \square \phi$  is continuous on  $\operatorname{dom} f + U$  by (i), we have  $(f \square \phi)(x) = \lim_{l \rightarrow \infty} (f \square \phi)(x_{k_l}) = \lim_{l \rightarrow \infty} (f(y_{k_l}) + \phi(x_{k_l} - y_{k_l})) \geq \liminf_{l \rightarrow \infty} f(y_{k_l}) + \liminf_{l \rightarrow \infty} \phi(x_{k_l} - y_{k_l}) \geq f(y) + \phi(x - y) \geq (f \square \phi)(x)$ , from which  $f(y) + \phi(x - y) = (f \square \phi)(x)$ , and so  $y = \operatorname{aprox}_f^\phi(x)$  by (ii). We conclude that  $\operatorname{aprox}_f^\phi$  is continuous at  $x$ . In turn, (19) follows from [28, Proposition 6].

(iv): We have  $0 \in \partial f(x) \Leftrightarrow x \in \partial f^*(0) \Leftrightarrow \nabla \phi(x) \in \nabla \phi \circ \partial f^*(0) \Leftrightarrow 0 \in (\operatorname{Id} + \nabla \phi \circ \partial f^*)^{-1}(\nabla \phi(x)) \Leftrightarrow 0 = (\operatorname{Id} + \nabla \phi \circ \partial f^*)^{-1}(\nabla \phi(x)) = \operatorname{aprox}_{f^*}^{\phi^*}(\nabla \phi(x))$ , because  $(\operatorname{Id} + \nabla \phi \circ \partial f^*)^{-1}$  is single-valued and (iii).

(v): (20) follows from (5). To see (21), let  $x \in \operatorname{argmin} f$ . By  $\phi \geq 0$  and (20), we have  $\inf(f \square \phi) = \inf f = f(x) \geq (f \square \phi)(x)$ , so  $x \in \operatorname{argmin}(f \square \phi)$ . Conversely, let  $x \in \operatorname{argmin}(f \square \phi)$ . Because  $y \mapsto f(y) + \phi(x - y)$  is 1-coercive, there exists  $y \in \mathbb{R}^n$  such that  $\inf f = \inf(f \square \phi) = (f \square \phi)(x) = f(y) + \phi(x - y) \geq \inf f$ , which implies  $f(y) = \inf f$  and  $\phi(x - y) = 0$ . Because  $\phi \geq 0$ ,  $\phi(0) = 0$ ,  $\phi$  is essentially strictly convex,  $\phi$  must have a unique minimizer at 0, so  $x = y$ . Hence  $x \in \operatorname{argmin} f$ . Altogether,  $\operatorname{argmin} f = \operatorname{argmin}(f \square \phi)$ .  $\blacksquare$

Our last result in this subsection expresses proximal mappings by anisotropic proximal mappings.

**Proposition 2.16** *Suppose that  $f \in \Gamma_0(\mathbb{R}^n)$  and  $(\operatorname{ri} \operatorname{dom} f) \cap U \neq \emptyset$ . Then for  $\lambda > 0$  one has*

$$(\forall x \in U) \quad \overleftarrow{\operatorname{prox}}_\lambda^\phi f(x) = \nabla \phi^* \left( \nabla \phi(x) - \lambda \operatorname{aprox}_{f^*}^{1/\lambda \star \phi^*}(\nabla \phi(x)/\lambda) \right).$$

Consequently,  $(\forall x \in U) \quad \nabla \phi(\overleftarrow{\operatorname{prox}}_\lambda^\phi f(x)) + \lambda \operatorname{aprox}_{f^*}^{1/\lambda \star \phi^*}(\nabla \phi(x)/\lambda) = \nabla \phi(x)$ .

*Proof.* By Proposition 2.10(iii),

$$(22) \quad \overleftarrow{\operatorname{prox}}_\lambda^\phi f = (\nabla \phi + \lambda \partial f)^{-1} \circ \nabla \phi.$$

As  $(\text{ri dom } f) \cap U \neq \emptyset$  and  $\phi^*$  essentially smooth, we have that  $(\lambda f)^* \square \phi^* = (\phi + \lambda f)^*$  is essentially smooth, see, e.g., [40, Corollary 26.3.2], so differentiable because  $\text{dom } \phi^* = \mathbb{R}^n$ . Then

$$(23) \quad (\nabla \phi + \lambda \partial f)^{-1} = \nabla(\phi + \lambda f)^*.$$

Now [40, Theorem 16.4] implies  $(\phi + \lambda f)^* = \lambda(f + \phi/\lambda)^*(\cdot/\lambda) = \lambda(f^* \square (\phi/\lambda)^*)(\cdot/\lambda)$  and  $\square$  is exact. By [9, Proposition 16.61(i)], for every  $y \in \mathbb{R}^n$ ,

$$(24) \quad \begin{aligned} \nabla(\phi + \lambda f)^*(y) &= \nabla(f^* \square (\phi/\lambda)^*)(y/\lambda) = \nabla(\phi/\lambda)^*(y/\lambda - \text{aprox}_{f^*}^{(\phi/\lambda)^*}(y/\lambda)) \\ &= \nabla \phi^*(\lambda(y/\lambda) - \text{aprox}_{f^*}^{(\phi/\lambda)^*}(y/\lambda)) = \nabla \phi^*(y - \lambda \text{aprox}_{f^*}^{1/\lambda \star \phi^*}(y/\lambda)). \end{aligned}$$

It follows from (22), (23) and (24) that for  $x \in U$ ,

$$\overleftarrow{\text{prox}}_{\lambda}^{\phi} f(x) = \nabla(\phi + \lambda f)^*(\nabla \phi(x)) = \nabla \phi^*(\nabla \phi(x) - \lambda \text{aprox}_{f^*}^{1/\lambda \star \phi^*}(\nabla \phi(x)/\lambda)),$$

as required.  $\blacksquare$

**Corollary 2.17** *Suppose that  $f \in \Gamma_0(\mathbb{R}^n)$  and  $(\text{ri dom } f) \cap U \neq \emptyset$ . Then for  $\lambda > 0$  one has*

$$(\forall x \in U) \quad x = \nabla \phi^*(\overleftarrow{\text{prox}}_{\lambda}^{\phi} f^*(\nabla \phi(x))) + \lambda \text{aprox}_f^{1/\lambda \star \phi}(x/\lambda).$$

*Proof.* In view of  $\text{ran } \nabla \phi = \mathbb{R}^n$ , Proposition 2.16 gives  $(\forall y \in \mathbb{R}^n) \quad y = \nabla \phi(\overleftarrow{\text{prox}}_{\lambda}^{\phi} f(\nabla \phi^*(y))) + \lambda \text{aprox}_{f^*}^{1/\lambda \star \phi^*}(y/\lambda)$ . The result follows by using this identity for  $f^*$  and  $\phi^*$ .  $\blacksquare$

**Remark 2.18** When  $\lambda = 1$ , Corollary 2.17 recovers [28, Theorem 1(ii)].

### 3 The Bregman proximal average

Let  $f_1, f_2 : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$ . In the rest of the paper our standing assumptions on  $f_1, f_2, \alpha$  and  $\lambda$  are:

**A3** Both  $f_1$  and  $f_2$  are proper lower semicontinuous and prox-bounded with thresholds  $\lambda_{f_1}, \lambda_{f_2} > 0$  respectively, and  $\bar{\lambda} := \min\{\lambda_{f_1}, \lambda_{f_2}\}$ .

**A4**  $\text{dom } f_i \cap \text{dom } \phi \neq \emptyset$  for  $i = 1, 2$ ,  $\alpha \in [0, 1]$ , and  $\lambda \in ]0, \bar{\lambda}]$ .

We define the  $\alpha$ -weighted Bregman proximal average with parameter  $\lambda$  of  $f_1, f_2$  with respect to the Legendre function  $\phi$  by

$$(25) \quad \mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha) := \left[ \alpha \left( f_1 + \frac{1}{\lambda} \phi \right)^* + (1 - \alpha) \left( f_2 + \frac{1}{\lambda} \phi \right)^* \right]^* - \frac{1}{\lambda} \phi,$$

with the convention that  $+\infty - (+\infty) = +\infty$ ,  $+\infty - r = +\infty$  for every  $r \in \mathbb{R}$ . As we shall see later that  $\text{dom} \left[ \alpha \left( f_1 + \frac{1}{\lambda} \phi \right)^* + (1 - \alpha) \left( f_2 + \frac{1}{\lambda} \phi \right)^* \right]^* \subseteq \text{dom } \phi$ , so (25) means that

$$(26) \quad \mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha)(x) = \begin{cases} \left[ \alpha \left( f_1 + \frac{1}{\lambda} \phi \right)^* + (1 - \alpha) \left( f_2 + \frac{1}{\lambda} \phi \right)^* \right]^*(x) - \frac{1}{\lambda} \phi(x), & \text{if } x \in \text{dom } \phi; \\ +\infty, & \text{if } x \notin \text{dom } \phi. \end{cases}$$

Therefore, it is possible that  $\mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha)(x) = +\infty$  when  $x \in \text{dom } \phi$ .

**Lemma 3.1** (i) *The function  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  is always lower semicontinuous on  $U$ .*

(ii) *If  $\text{dom } \phi$  is closed, and  $\phi$  is relatively continuous on  $\text{dom } \phi$ , then  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  is lower semicontinuous on  $\mathbb{R}^n$ . Suppose one of the following holds:*

- (a)  *$\text{dom } \phi$  is polyhedral.*
- (b)  *$\text{dom } \phi$  is locally simplicial.*

*Then  $\phi$  is relatively continuous on  $\text{dom } \phi$ .*

*Proof.* (i): This is because that  $\phi$  is continuous on  $U$  and  $\left[ \alpha \left( f_1 + \frac{1}{\lambda} \phi \right)^* + (1 - \alpha) \left( f_2 + \frac{1}{\lambda} \phi \right)^* \right]^*$  is lower semicontinuous on  $U$ .

(ii): On the open set  $\mathbb{R}^n \setminus \text{dom } \phi$ ,  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) \equiv +\infty$ , so  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  is lower semicontinuous on  $\mathbb{R}^n \setminus \text{dom } \phi$ . Now let  $x_0 \in \text{dom } \phi$ . Then

$$(27) \quad \liminf_{x \rightarrow x_0} \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x) = \liminf_{x \rightarrow x_0, x \in \text{dom } \phi} \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x)$$

$$(28) \quad = \liminf_{x \rightarrow x_0, x \in \text{dom } \phi} \left[ \alpha \left( f_1 + \frac{1}{\lambda} \phi \right)^* + (1 - \alpha) \left( f_2 + \frac{1}{\lambda} \phi \right)^* \right]^*(x) - \lim_{x \rightarrow x_0, x \in \text{dom } \phi} \frac{1}{\lambda} \phi(x)$$

$$(29) \quad \geq \left[ \alpha \left( f_1 + \frac{1}{\lambda} \phi \right)^* + (1 - \alpha) \left( f_2 + \frac{1}{\lambda} \phi \right)^* \right]^*(x_0) - \frac{1}{\lambda} \phi(x_0) = \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x_0).$$

Since  $x_0 \in \text{dom } \phi$  was arbitrary,  $f$  is lower semicontinuous on  $\text{dom } \phi$ . Altogether,  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  is lower semicontinuous on  $\mathbb{R}^n$ . Under (ii)(a) or (ii)(b), the relative continuity of  $\phi$  on  $\text{dom } \phi$  follows from [40, Theorem 10.2] or [41, Theorem 2.35]. ■

**Lemma 3.2** *The following holds:*

$$\frac{1}{\lambda} [\alpha(\lambda f_1 + \phi)^* + (1 - \alpha)(\lambda f_2 + \phi)^*]^* = \left[ \alpha \left( f_1 + \frac{1}{\lambda} \phi \right)^* + (1 - \alpha) \left( f_2 + \frac{1}{\lambda} \phi \right)^* \right]^*.$$

*Proof.* Indeed, this is a simple calculation:

$$\begin{aligned} & \frac{1}{\lambda} [\alpha(\lambda f_1 + \phi)^* + (1 - \alpha)(\lambda f_2 + \phi)^*]^* = \\ & \left[ \frac{1}{\lambda} \left( \alpha(\lambda f_1 + \phi)^* + (1 - \alpha)(\lambda f_2 + \phi)^* \right) (\lambda \cdot) \right]^* = \left[ \alpha \frac{1}{\lambda} (\lambda f_1 + \phi)^* (\lambda \cdot) + (1 - \alpha) \frac{1}{\lambda} (\lambda f_2 + \phi)^* (\lambda \cdot) \right]^* \\ & = \left[ \alpha \left( \frac{\lambda f_1 + \phi}{\lambda} \right)^* + (1 - \alpha) \left( \frac{\lambda f_2 + \phi}{\lambda} \right)^* \right]^* = \left[ \alpha \left( f_1 + \frac{1}{\lambda} \phi \right)^* + (1 - \alpha) \left( f_2 + \frac{1}{\lambda} \phi \right)^* \right]^*. \end{aligned}$$

■

Because of Lemma 3.1, in the rest of the paper our additional standing assumption on  $\phi$  is:

**A5**  $\text{dom } \phi$  is closed,  $\phi$  is relatively continuous on  $\text{dom } \phi$ , and  $\phi$  is twice continuously differentiable on  $U$  with  $\nabla^2 \phi(u)$  being positive definite for every  $u \in U$ .

We are now ready for the main result of this section.

**Theorem 3.3 (Bregman proximal average)** *Suppose that A1–A5 hold. Then the following hold:*

- (i)  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = [\alpha \star \text{conv}(f_1 + \frac{1}{\lambda}\phi)] \square [(1 - \alpha) \star \text{conv}(f_2 + \frac{1}{\lambda}\phi)] - \frac{1}{\lambda}\phi$ , where the infimal convolution  $\square$  is exact.
- (ii)  $\text{dom } \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \alpha \text{conv}(\text{dom } f_1 \cap \text{dom } \phi) + (1 - \alpha) \text{conv}(\text{dom } f_2 \cap \text{dom } \phi) \subseteq \text{dom } \phi$ .
- (iii)  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  is proper lower semicontinuous on  $\mathbb{R}^n$ .
- (iv)  $\lambda \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) + \phi \in \Gamma_0(\mathbb{R}^n)$ .
- (v) The function  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  is prox-bounded below with its prox-bound  $\lambda_f \geq \bar{\lambda}$ .
- (vi)  $\overleftarrow{\text{env}}_\lambda^\phi \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \alpha \overleftarrow{\text{env}}_\lambda^\phi f_1 + (1 - \alpha) \overleftarrow{\text{env}}_\lambda^\phi f_2$ .
- (vii)  $(\forall x \in U) \overleftarrow{\text{prox}}_\lambda^\phi \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x) = \alpha \text{conv}(\overleftarrow{\text{prox}}_\lambda^\phi f_1(x)) + (1 - \alpha) \text{conv}(\overleftarrow{\text{prox}}_\lambda^\phi f_2(x))$ .
- (viii) When  $\alpha = 0$ ,  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \overleftarrow{\text{hul}}_\lambda^\phi f_2$ ; when  $\alpha = 1$ ,  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \overleftarrow{\text{hul}}_\lambda^\phi f_1$ ; when  $f_1 = f_2 = f$ ,  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \overleftarrow{\text{hul}}_\lambda^\phi f$ .

*Proof.* (i): Since  $\text{dom}(f_1 + 1/\lambda\phi)^* = \mathbb{R}^n = \text{dom}(f_2 + 1/\lambda\phi)^*$ , by [40, Theorem 16.4],

$$(30) \quad \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \left[ \alpha \left( f_1 + \frac{1}{\lambda}\phi \right)^{**} \left( \frac{\cdot}{\alpha} \right) \right] \square \left[ (1 - \alpha) \left( f_2 + \frac{1}{\lambda}\phi \right)^{**} \left( \frac{\cdot}{(1 - \alpha)} \right) \right] - \frac{1}{\lambda}\phi,$$

and the infimal convolution  $\square$  is exact. Because  $f_1 + 1/\lambda\phi$  and  $f_2 + 1/\lambda\phi$  are 1-coercive by Proposition 2.14, [17, Lemma 3.3] gives

$$\left( f_1 + \frac{1}{\lambda}\phi \right)^{**} = \text{conv} \left( f_1 + \frac{1}{\lambda}\phi \right), \quad \left( f_2 + \frac{1}{\lambda}\phi \right)^{**} = \text{conv} \left( f_2 + \frac{1}{\lambda}\phi \right).$$

Hence (i) holds.

(ii): Because  $\text{dom} [\text{conv}(f_i + \frac{1}{\lambda}\phi)] = \text{conv}(\text{dom } f_i \cap \text{dom } \phi)$  with  $i = 1, 2$ , by [9, Proposition 12.6(ii)] and (i) we obtain

$$(31) \quad \text{dom } \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = [\alpha \text{conv}(\text{dom } f_1 \cap \text{dom } \phi) + (1 - \alpha) \text{conv}(\text{dom } f_2 \cap \text{dom } \phi)] \cap \text{dom } \phi$$

$$(32) \quad = \alpha \text{conv}(\text{dom } f_1 \cap \text{dom } \phi) + (1 - \alpha) \text{conv}(\text{dom } f_2 \cap \text{dom } \phi),$$

where the second “=” follows from the convexity of  $\text{dom } \phi$ .

(iii): By (ii),  $\text{dom } \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) \neq \emptyset$ ; by (i),  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) > -\infty$ ; by Lemma 3.1(ii),  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  lower semicontinuous. Therefore, (iii) is verified.

(iv): By (25) and (ii), we have

$$\lambda \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) + \phi = \lambda \left[ \alpha \left( f_1 + \frac{1}{\lambda}\phi \right)^* + (1 - \alpha) \left( f_2 + \frac{1}{\lambda}\phi \right)^* \right]^*,$$

so  $\lambda \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) + \phi \in \Gamma_0(\mathbb{R}^n)$ .

(v): Let  $0 < \lambda < \tilde{\lambda} < \bar{\lambda}$ . By Proposition 2.4, there exists  $c \in \mathbb{R}$  such that  $f_i + \frac{1}{\lambda}\phi \geq c$  for  $i = 1, 2$ . This implies

$$(33) \quad f_i + \frac{1}{\lambda}\phi = f_i + \frac{1}{\lambda}\phi + \left(\frac{1}{\lambda} - \frac{1}{\tilde{\lambda}}\right)\phi \geq c + \left(\frac{1}{\lambda} - \frac{1}{\tilde{\lambda}}\right)\phi,$$

so  $\left(f_i + \frac{1}{\lambda}\phi\right)^{**} \geq c + \left(\frac{1}{\lambda} - \frac{1}{\tilde{\lambda}}\right)\phi$  because  $\phi \in \Gamma_0(\mathbb{R}^n)$ . In view of (30),  $\forall x \in \text{dom } \phi$  we have  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x)$

$$(34) \quad \geq \left[ \alpha \left( c + \left( \frac{1}{\lambda} - \frac{1}{\tilde{\lambda}} \right) \phi \right) \left( \frac{\cdot}{\alpha} \right) \right] \square \left[ (1 - \alpha) \left( c + \left( \frac{1}{\lambda} - \frac{1}{\tilde{\lambda}} \right) \phi \right) \left( \frac{\cdot}{1 - \alpha} \right) \right] (x) - \frac{1}{\lambda}\phi(x)$$

$$(35) \quad = \inf_{u \in \mathbb{R}^n} \left[ c + \alpha \left( \frac{1}{\lambda} - \frac{1}{\tilde{\lambda}} \right) \phi \left( \frac{u}{\alpha} \right) + (1 - \alpha) \left( \frac{1}{\lambda} - \frac{1}{\tilde{\lambda}} \right) \phi \left( \frac{x - u}{1 - \alpha} \right) \right] - \frac{1}{\lambda}\phi(x)$$

$$(36) \quad = c + \left( \frac{1}{\lambda} - \frac{1}{\tilde{\lambda}} \right) \inf_{u \in \mathbb{R}^n} \left[ \alpha \phi \left( \frac{u}{\alpha} \right) + (1 - \alpha) \phi \left( \frac{x - u}{1 - \alpha} \right) \right] - \frac{1}{\lambda}\phi(x)$$

$$(37) \quad = c + \left( \frac{1}{\lambda} - \frac{1}{\tilde{\lambda}} \right) \phi(x) - \frac{1}{\lambda}\phi(x) = c - \frac{1}{\tilde{\lambda}}\phi(x),$$

where from (36) to (37) we use the convexity of  $\phi$ . Hence  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) + \frac{1}{\lambda}\phi \geq c$  on  $\text{dom } \phi$ , and so  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) + \frac{1}{\lambda}\phi \geq c$  on  $\mathbb{R}^n$ . Because  $\tilde{\lambda} \in ]0, \bar{\lambda}[$  was arbitrary, we conclude that  $\lambda_f \geq \bar{\lambda}$  by Proposition 2.4.

(vi): Since  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  is proper lower semicontinuous by (iii), it follows from Corollary 2.8 and Proposition 2.10 that

$$(38) \quad \lambda \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) + \phi = (\phi^* - \lambda \overleftarrow{\text{env}}_\lambda^\phi \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) \circ \nabla \phi^*)^*, \text{ and} \\ \overleftarrow{\text{prox}}_\lambda^\phi \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) \text{ is convex-valued.}$$

Using Lemma 3.2, we obtain

$$(39) \quad \lambda \left[ \alpha \left( f_1 + \frac{1}{\lambda}\phi \right)^* + (1 - \alpha) \left( f_2 + \frac{1}{\lambda}\phi \right)^* \right]^* = [\alpha(\lambda f_1 + \phi)^* + (1 - \alpha)(\lambda f_2 + \phi)^*]^*.$$

Fact 2.6 gives

$$(40) \quad (\lambda f_i + \phi)^* = \phi^* - \lambda \overleftarrow{\text{env}}_\lambda^\phi f_i \circ \nabla \phi^*,$$

which implies that  $\phi^* - \lambda \overleftarrow{\text{env}}_\lambda^\phi f_i \circ \nabla \phi^*$  is convex. Combining equations (25) and (38)–(40) yields

$$(41) \quad (\phi^* - \lambda \overleftarrow{\text{env}}_\lambda^\phi \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) \circ \nabla \phi^*)^* = \left[ \alpha(\phi^* - \lambda \overleftarrow{\text{env}}_\lambda^\phi f_1 \circ \nabla \phi^*) + (1 - \alpha)(\phi^* - \lambda \overleftarrow{\text{env}}_\lambda^\phi f_2 \circ \nabla \phi^*) \right]^*$$

$$(42) \quad = \left[ -\alpha \lambda \overleftarrow{\text{env}}_\lambda^\phi f_1 \circ \nabla \phi^* - (1 - \alpha) \lambda \overleftarrow{\text{env}}_\lambda^\phi f_2 \circ \nabla \phi^* + \phi^* \right]^*.$$

Because  $\phi$  is coercive,  $\phi^*$  is real-valued on  $\mathbb{R}^n$ . Taking conjugate both sides, followed by subtracting both sides by  $\phi^*$ , and using the fact that  $\nabla \phi^*$  is an isomorphism lead to

$$\overleftarrow{\text{env}}_\lambda^\phi \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \alpha \overleftarrow{\text{env}}_\lambda^\phi f_1 + (1 - \alpha) \overleftarrow{\text{env}}_\lambda^\phi f_2 \text{ on } U.$$

(vii): By (vi), the sum rule of Clarke subdifferential [41, Corollary 10.9] or [27, Proposition 2.3.3, Corollary 3] gives

$$\partial_C(-\overleftarrow{\text{env}}_\lambda^\phi \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)) = \alpha \partial_C(-\overleftarrow{\text{env}}_\lambda^\phi f_1) + (1 - \alpha) \partial_C(-\overleftarrow{\text{env}}_\lambda^\phi f_2),$$

in which “=” holds because both  $-\overleftarrow{\text{env}}_\lambda^\phi f_1$  and  $-\overleftarrow{\text{env}}_\lambda^\phi f_2$  are locally Lipschitz and Clarke regular. Because of (v), we can apply Fact 2.9 to obtain

$$(43) \quad \frac{1}{\lambda} \nabla^2 \phi(x) [\text{conv}(\overleftarrow{\text{prox}}_\lambda^\phi \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x)) - x]$$

$$(44) \quad = \alpha \frac{1}{\lambda} \nabla^2 \phi(x) [\text{conv}(\overleftarrow{\text{prox}}_\lambda^\phi f_1(x)) - x] + (1 - \alpha) \frac{1}{\lambda} \nabla^2 \phi(x) [\text{conv}(\overleftarrow{\text{prox}}_\lambda^\phi f_2(x)) - x].$$

Multiplying both sides by  $(\nabla^2 \phi(x))^{-1}$  and simplifications give

$$\text{conv}(\overleftarrow{\text{prox}}_\lambda^\phi \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x)) = \alpha [\text{conv}(\overleftarrow{\text{prox}}_\lambda^\phi f_1(x))] + (1 - \alpha) [\text{conv}(\overleftarrow{\text{prox}}_\lambda^\phi f_2(x))].$$

Since  $\overleftarrow{\text{prox}}_\lambda^\phi \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x)$  is convex by (iv) and Fact 2.10(ii), (vii) is proved.

(viii): Apply Proposition 2.14(ii).  $\blacksquare$

**Corollary 3.4** *Suppose that **A1–A5** hold, and that  $f_i \in \Gamma_0(\mathbb{R}^n)$  with  $\text{dom } f_i \cap U \neq \emptyset$  for  $i = 1, 2$ . Then for  $\lambda \in ]0, +\infty[$ ,*

$$(45) \quad \left( \partial \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) + \frac{1}{\lambda} \nabla \phi \right)^{-1} = \alpha \left( \partial f_1 + \frac{1}{\lambda} \nabla \phi \right)^{-1} + (1 - \alpha) \left( \partial f_2 + \frac{1}{\lambda} \nabla \phi \right)^{-1}.$$

In particular,  $\forall x \in U$ ,  $\partial \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x) = \hat{\partial} \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x) =$

$$(46) \quad \left[ \alpha \left( \partial f_1 + \frac{1}{\lambda} \nabla \phi \right)^{-1} + (1 - \alpha) \left( \partial f_2 + \frac{1}{\lambda} \nabla \phi \right)^{-1} \right]^{-1} (x) - \frac{1}{\lambda} \nabla \phi(x).$$

*Proof.* By Corollary 2.5,  $\bar{\lambda} = +\infty$ . To see (45), apply Theorem 3.3(vii) and Fact 2.10(ii)&(iii). Next, (46) follows from (45) and that  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \left( \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) + \frac{1}{\lambda} \phi \right) - \frac{1}{\lambda} \phi$  being a difference of a convex function and a  $C^1$  function is Clarke regular.  $\blacksquare$

**Remark 3.5** Note that while  $\partial f_i$  is monotone,  $\partial \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  may be not monotone; see, e.g., Example 4.8.

Let us give a special case when both  $f_1, f_2$  are indicator functions of closed subsets. This highlights the connection to averaged Bregman projections, which solve feasibility problems. As in [11], we define Bregman nearest distance function and nearest-point map.

**Definition 3.6** *The left Bregman nearest-distance function to  $C$  is defined by*

$$(47) \quad \overleftarrow{D}_C : U \rightarrow [0, +\infty] : y \mapsto \inf_{x \in C} D_\phi(x, y),$$

and the left Bregman nearest-point map (i.e., the classical Bregman projector) onto  $C$  is

$$\overleftarrow{P}_C : U \rightrightarrows U : y \mapsto \underset{x \in C}{\text{argmin}} D_\phi(x, y) = \{x \in C : D_\phi(x, y) = \overleftarrow{D}_C(y)\}.$$

Using Lemma 3.2 and Fact 2.6, we can write the proximal average:

$$\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \frac{1}{\lambda} [\phi^* - \alpha \lambda \overleftarrow{\text{env}}_\lambda^\phi f_1 \circ \nabla \phi^* - (1 - \alpha) \lambda \overleftarrow{\text{env}}_\lambda^\phi f_2 \circ \nabla \phi^*]^* - \frac{1}{\lambda} \phi.$$

In view of  $\overleftarrow{\text{env}}_\lambda^\phi \iota_C = 1/\lambda \overleftarrow{D}_C$ ,  $\overleftarrow{\text{prox}}_\lambda^\phi \iota_C = \overleftarrow{P}_C$ , we obtain the following result.

**Corollary 3.7** *Suppose that **A1–A5** hold, and that  $f_i := \iota_{C_i}$  with  $C_i \subseteq \mathbb{R}^n$  being nonempty and closed for  $i = 1, 2$ . Then for  $\lambda \in ]0, +\infty[$  the following hold:*

- (i)  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \frac{1}{\lambda}[\phi^* - \alpha \overleftarrow{D}_{C_1} \circ \nabla \phi^* - (1 - \alpha) \overleftarrow{D}_{C_2} \circ \nabla \phi^*]^* - \frac{1}{\lambda} \phi.$
- (ii)  $\text{dom } \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \alpha \text{conv}(C_1 \cap \text{dom } \phi) + (1 - \alpha) \text{conv}(C_2 \cap \text{dom } \phi) \subseteq \text{dom } \phi.$
- (iii)  $\overleftarrow{\text{env}}_\lambda^\phi \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \alpha \overleftarrow{D}_{C_1} + (1 - \alpha) \overleftarrow{D}_{C_2}.$
- (iv)  $(\forall x \in U) \overleftarrow{\text{prox}}_\lambda^\phi \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x) = \alpha \text{conv } \overleftarrow{P}_{C_1}(x) + (1 - \alpha) \text{conv } \overleftarrow{P}_{C_2}(x).$

If, in addition,  $C_1, C_2$  are convex, then

- (a)  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x) = \frac{1}{\lambda} \inf\{\alpha D_\phi(y_1, x) + (1 - \alpha) D_\phi(y_2, x) : y_i \in C_i \cap \text{dom } \phi, i = 1, 2, \alpha y_1 + (1 - \alpha) y_2 = x\},$  and
- (b) the “conv” operations in (ii) and (iv) are superfluous.

*Proof.* (i)-(iv) follow from Theorem 3.3. To see (a), we consider

$$\begin{aligned}
(48) \quad \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x) &= \left[ \alpha \star \left( \iota_{C_1} + \frac{1}{\lambda} \phi \right) \right] \square \left[ (1 - \alpha) \star \left( \iota_{C_2} + \frac{1}{\lambda} \phi \right) \right] (x) - \frac{1}{\lambda} \phi(x) \\
(49) \quad &= \inf_{x_1 + x_2 = x} \left( \iota_{C_1}(x_1/\alpha) + \alpha \frac{1}{\lambda} \phi(x_1/\alpha) + \iota_{C_2}(x_2/(1 - \alpha)) + (1 - \alpha) \frac{1}{\lambda} \phi(x_2/(1 - \alpha)) \right) - \frac{1}{\lambda} \phi(x) \\
(50) \quad &= \frac{1}{\lambda} \inf\{\alpha \phi(y_1) + (1 - \alpha) \phi(y_2) - \phi(x) : y_i \in C_i \cap \text{dom } \phi, i = 1, 2, \alpha y_1 + (1 - \alpha) y_2 = x\}.
\end{aligned}$$

The proof is complete by using that when  $\alpha y_1 + (1 - \alpha) y_2 = x$ , one has

$$\begin{aligned}
(51) \quad &\alpha \phi(y_1) + (1 - \alpha) \phi(y_2) - \phi(x) \\
(52) \quad &= \alpha(\phi(y_1) - \phi(x) - \langle \nabla \phi(x), y_1 - x \rangle) + (1 - \alpha)(\phi(y_2) - \phi(x) - \langle \nabla \phi(x), y_2 - x \rangle) \\
(53) \quad &= \alpha D_\phi(y_1, x) + (1 - \alpha) D_\phi(y_2, x).
\end{aligned}$$

■

## 4 When is the Bregman proximal average convex?

We shall need a Bregman version of the Baillon-Haddad theorem, see, e.g., [2, 9]. To this end, we introduce  $\nabla \phi$ -firmly nonexpansive mappings. Define the symmetrized Bregman distance  $S_\phi : U \times U \rightarrow \mathbb{R}$  by  $S_\phi(x, y) = D_\phi(x, y) + D_\phi(y, x) = \langle \nabla \phi(x) - \nabla \phi(y), x - y \rangle$ .

**Definition 4.1** *Let  $T : U \subseteq \mathbb{R}^n \rightarrow U$ . We say that  $T$  is  $\nabla \phi$ -firmly nonexpansive on  $U$  if*

$$(\forall u \in U)(\forall v \in U) \langle u - v, Tu - Tv \rangle \geq \langle \nabla \phi(Tu) - \nabla \phi(Tv), Tu - Tv \rangle = S_\phi(Tu, Tv).$$

When  $\phi(x) = 1/2\|x\|^2$ , a  $\nabla\phi$ -firmly nonexpansive mapping is the usual firmly nonexpansive mapping; see, e.g., [9, Proposition 4.4].

**Lemma 4.2** *Suppose that  $g \in \Gamma_0(\mathbb{R}^n)$ ,  $\text{dom } g \subseteq \text{dom } \phi$ , and  $(\text{ri dom } g) \cap U \neq \emptyset$ . Then the following are equivalent:*

- (i)  $g : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  is  $\phi$ -strongly convex, i.e.,  $g = f + \phi$  for a convex function  $f \in \Gamma_0(\mathbb{R}^n)$ .
- (ii)  $g^*$  is a  $\phi^*$ -anisotropic envelope of  $f^*$  with  $f \in \Gamma_0(\mathbb{R}^n)$ , i.e.,  $g^* = f^* \square \phi^*$ .
- (iii)  $g^*$  is differentiable with  $\nabla g^*$  being  $\nabla\phi$ -firmly nonexpansive on  $\mathbb{R}^n$ .
- (iv)  $(\phi^* - g^*) \circ \nabla\phi = \lambda \overleftarrow{\text{env}}_{\lambda}^{\phi} f$  for a convex function  $f \in \Gamma_0(\mathbb{R}^n)$  and  $\lambda > 0$ .
- (v)  $g^*$  is differentiable on  $\mathbb{R}^n$  with  $\nabla g^* \circ \nabla\phi = \overleftarrow{\text{prox}}_1^{\phi} f$  for some  $f \in \Gamma_0(\mathbb{R}^n)$ .

*Proof.* (i) $\Rightarrow$ (ii): Since  $\emptyset \neq \text{ri dom } g = \text{ri}[(\text{dom } f) \cap (\text{dom } \phi)] = (\text{ri dom } f) \cap (\text{ri dom } \phi) \subseteq (\text{dom } f) \cap U$ , we have  $\text{dom } f \cap \text{int dom } \phi \neq \emptyset$ . Apply the Attouch-Brezis theorem [9, Theorem 15.3]. (ii) $\Rightarrow$ (i): Take the conjugation both sides to obtain  $g = g^{**} = f^{**} + \phi^{**} = f + \phi$ ; see, e.g., [9, Theorem 13.37].

(i) $\Rightarrow$ (iii): Since  $\phi$  is 1-coercive, so is  $g$  and hence  $\text{ran } \partial g = \mathbb{R}^n$ . Because  $\text{dom } g \cap \text{int dom } \phi \neq \emptyset$  implies  $\text{dom } f \cap \text{int dom } \phi \neq \emptyset$ , we have  $\partial g = \partial f + \partial\phi$ , so  $\text{dom } \partial g \subseteq \text{dom } \partial\phi$ . As  $f$  is convex,  $\phi$  is essentially strictly convex, we see that  $g$  is essentially strictly convex, so  $g^*$  is essentially smooth. Using  $u \in \partial g(x), v \in \partial g(y)$  if and only if  $x = \nabla g^*(u), y = \nabla g^*(v)$ , we obtain

$$(54) \quad \langle \partial g(x) - \partial g(y), x - y \rangle \geq \langle \nabla\phi(x) - \nabla\phi(y), x - y \rangle$$

$$(55) \quad \Leftrightarrow \langle u - v, \nabla g^*(u) - \nabla g^*(v) \rangle \geq \langle \nabla\phi(\nabla g^*(u)) - \nabla\phi(\nabla g^*(v)), \nabla g^*(u) - \nabla g^*(v) \rangle$$

for all  $u, v \in \mathbb{R}^n$ .

(iii) $\Rightarrow$ (i): Since

$$(56) \quad (\forall u, v \in \mathbb{R}^n) \langle u - v, \nabla g^*(u) - \nabla g^*(v) \rangle \geq \langle \nabla\phi(\nabla g^*(u)) - \nabla\phi(\nabla g^*(v)), \nabla g^*(u) - \nabla g^*(v) \rangle$$

$$(57) \quad \Leftrightarrow (\forall x, y \in \text{dom } \partial g \cap U) \langle \partial g(x) - \partial g(y), x - y \rangle \geq \langle \nabla\phi(x) - \nabla\phi(y), x - y \rangle,$$

the function  $g - \phi$  is convex on convex subsets of  $(\text{dom } \partial g) \cap U \supseteq (\text{ri dom } g) \cap U = \text{ri}(\text{dom } g \cap \text{dom } \phi) = \text{ri dom } g$ . Define  $\tilde{f}(x) = g(x) - \phi(x)$  if  $x \in \text{ri dom } g$ , and  $+\infty$  otherwise. Since  $\tilde{f}$  is proper and convex, by [41, Theorem 2.35], the lower semicontinuous hull  $f = \text{cl } \tilde{f}$  is proper, so  $f \in \Gamma_0(\mathbb{R}^n)$ . We claim that  $g = f + \phi$  on  $\text{dom } g$ . Indeed, as  $g - \phi$  is relatively continuous on  $\text{ri dom } g$ ,  $f = \text{cl}(g - \phi) = g - \phi$ , which gives  $g = f + \phi$  on  $\text{ri dom } g$ . Take  $x_0 \in \text{ri dom } g \cap U$ , which is possible by the assumption, and let  $x \in \text{dom } g$ . Then, by [41, Theorem 2.36],

$$f(x) = \lim_{\tau \uparrow 1} f((1 - \tau)x_0 + \tau x) = \lim_{\tau \uparrow 1} (g((1 - \tau)x_0 + \tau x) - \phi((1 - \tau)x_0 + \tau x)) = g(x) - \phi(x)$$

because both  $g, \phi \in \Gamma_0(\mathbb{R}^n)$ . Therefore,  $f = g - \phi$  on  $\text{dom } g$ . As  $\text{dom } g \subset \text{dom } \phi$ , we get  $g = f + \phi$  on  $\text{dom } g$  and  $f \in \Gamma_0(\mathbb{R}^n)$ . However, at this stage, we do not know whether  $g = f + \phi$  on  $\mathbb{R}^n \setminus \text{dom } g$ . Now write  $g = (f + \iota_{\text{dom } g}) + \phi$ . Because  $\text{dom}(f + \iota_{\text{dom } g}) = \text{dom } g$ ,  $\text{ri dom } g \cap U \neq \emptyset$  and both  $(f + \iota_{\text{dom } g})$  and  $\phi$  are proper convex, [40, Theorem 9.3] gives

$$g = \text{cl } g = \text{cl}(f + \iota_{\text{dom } g}) + \text{cl } \phi = \text{cl}(f + \iota_{\text{dom } g}) + \phi$$

and  $\text{cl}(f + \iota_{\text{dom } g}) \in \Gamma_0(\mathbb{R}^n)$ . This proves (i).

(iv) $\Leftrightarrow$ (i): We have

$$(58) \quad (iv) \Leftrightarrow (\phi^* - g^*) \circ \nabla\phi = \lambda \overleftarrow{\text{env}}_{\lambda}^{\phi} f \Leftrightarrow \phi^* - g^* = \lambda \overleftarrow{\text{env}}_{\lambda}^{\phi} f \circ \nabla\phi^*$$

$$(59) \quad \Leftrightarrow \phi^* - \lambda \overleftarrow{\text{env}}_{\lambda}^{\phi} f \circ \nabla\phi^* = g^* \Leftrightarrow (\lambda f + \phi)^* = g^* \text{ (Fact 2.6)} \Leftrightarrow g = \lambda f + \phi,$$

and  $\lambda f \in \Gamma_0(\mathbb{R}^n)$ .

(ii) $\Rightarrow$ (v): (ii) gives  $\text{dom } g^* = \mathbb{R}^n$  and  $(\forall x^* \in \mathbb{R}^n) \nabla g^*(x^*) = \nabla\phi^*(x^* - \text{aprox}_{f^*}^{\phi^*}(x^*))$ . Put  $x^* = \nabla\phi(x)$  for  $x \in U$  to obtain

$$\nabla g^*(\nabla\phi(x)) = \nabla\phi^*(\nabla\phi(x) - \text{aprox}_{f^*}^{\phi^*}(\nabla\phi(x))) = \overleftarrow{\text{prox}}_1^{\phi} f(x)$$

by Proposition 2.16.

(v) $\Rightarrow$ (ii): (v) gives  $(\forall x \in U) \nabla g^*(\nabla\phi(x)) = \overleftarrow{\text{prox}}_1^{\phi} f(x) = \nabla\phi^*(\nabla\phi(x) - \text{aprox}_{f^*}^{\phi^*}(\nabla\phi(x)))$ . In view of  $\text{ran } \nabla\phi = \mathbb{R}^n$ , replacing  $\nabla\phi(x)$  by  $x^*$  gives

$$(\forall x^* \in \mathbb{R}^n) \nabla g^*(x^*) = \nabla\phi^*(x^* - \text{aprox}_{f^*}^{\phi^*}(x^*)) = \nabla(f^* \square \phi^*)(x^*),$$

which implies  $g^* = (f^* \square \phi^*) + c = (f^* + c) \square \phi^*$  for a constant  $c \in \mathbb{R}$ . Hence (ii) holds.  $\blacksquare$

**Remark 4.3** The above is an extended version of Baillon-Haddad Theorem; see [9, Theorem 18.15, Corollary 18.17], [2].  $\phi$ -strongly convex functions have been used in [5] for studying Bregman gradient algorithms.

**Lemma 4.4** *Let  $S_{\phi}$  be convex. Suppose that  $T_1, T_2$  are  $\nabla\phi$ -firmly nonexpansive on  $U$ . Then  $\alpha T_1 + (1 - \alpha)T_2$  is  $\nabla\phi$ -firmly nonexpansive on  $U$ .*

*Proof.* This follows from the following calculations:  $\forall u, v \in U$ ,

$$\begin{aligned} & \langle \nabla\phi(\alpha T_1 u + (1 - \alpha)T_2 u) - \nabla\phi(\alpha T_1 v + (1 - \alpha)T_2 v), (\alpha T_1 u + (1 - \alpha)T_2 u) - (\alpha T_1 v + (1 - \alpha)T_2 v) \rangle \\ &= S_{\phi}(\alpha T_1 u + (1 - \alpha)T_2 u, \alpha T_1 v + (1 - \alpha)T_2 v) = S_{\phi}(\alpha(T_1 u, T_1 v) + (1 - \alpha)(T_2 u, T_2 v)) \\ &\leq \alpha S_{\phi}(T_1 u, T_1 v) + (1 - \alpha)S_{\phi}(T_2 u, T_2 v) \quad (S_{\phi} \text{ being convex}) \\ &\leq \alpha \langle u - v, T_1 u - T_1 v \rangle + (1 - \alpha) \langle u - v, T_2 u - T_2 v \rangle \quad (T_i \text{ being } \nabla\phi\text{-firmly nonexpansive}) \\ &= \langle u - v, \alpha T_1 u + (1 - \alpha)T_2 u - (\alpha T_1 v + (1 - \alpha)T_2 v) \rangle. \end{aligned}$$

$\blacksquare$

Here is the main result of this section.

**Theorem 4.5 (convexity of Bregman proximal average)** *Let **A1–A5** hold, and let  $S_{\phi}$  be convex. Suppose that  $f_i \in \Gamma_0(\mathbb{R}^n)$  and  $(\text{ri dom } f_i) \cap U \neq \emptyset$  for  $i = 1, 2$ . Then  $\mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha)$  is convex.*

*Proof.* Recall that

$$(60) \quad \mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha) = \left[ \alpha \left( f_1 + \frac{1}{\lambda} \phi \right)^* + (1 - \alpha) \left( f_2 + \frac{1}{\lambda} \phi \right)^* \right]^* - \frac{1}{\lambda} \phi.$$

Since  $f_i + \frac{1}{\lambda} \phi$  is  $\phi/\lambda$ -strongly convex, by Lemma 4.2(iii), each  $T_i = \nabla \left( f_i + \frac{1}{\lambda} \phi \right)^*$  is  $\nabla\phi/\lambda$ -firmly nonexpansive. Lemma 4.4 implies  $\alpha \nabla \left( f_1 + \frac{1}{\lambda} \phi \right)^* + (1 - \alpha) \nabla \left( f_2 + \frac{1}{\lambda} \phi \right)^*$  is  $\nabla\phi/\lambda$ -firmly nonexpansive. Because

$$\text{dom} \left[ \alpha \left( f_1 + \frac{1}{\lambda} \phi \right)^* + (1 - \alpha) \left( f_2 + \frac{1}{\lambda} \phi \right)^* \right]^* = \alpha(\text{dom } f_1 \cap \text{dom } \phi) + (1 - \alpha)(\text{dom } f_2 \cap \text{dom } \phi),$$

by the assumption, we have  $\text{ri}[\alpha(\text{dom } f_1 \cap \text{dom } \phi) + (1 - \alpha)(\text{dom } f_2 \cap \text{dom } \phi)] \cap U \neq \emptyset$ . Apply Lemma 4.2(iii) again to obtain that

$$\left[ \alpha \left( f_1 + \frac{1}{\lambda} \phi \right)^* + (1 - \alpha) \left( f_2 + \frac{1}{\lambda} \phi \right)^* \right]^*$$

is  $\phi/\lambda$ -strongly convex. Hence  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  is convex by Lemma 4.2(i). ■

**Remark 4.6** Clearly, the joint convexity of  $D_\phi$  implies the convexity of  $S_\phi$ . For conditions on joint convexity of  $D_\phi$ , see [7].

**Corollary 4.7** Let **A1–A5** hold, and let  $S_\phi$  be convex. Suppose that  $f_i \in \Gamma_0(\mathbb{R}^n)$  and  $(\text{ri dom } f_i) \cap U \neq \emptyset$  for  $i = 1, 2$ . Then  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  is convex, and  $(\forall x \in U) \overleftarrow{\text{prox}}_\lambda^\phi \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x) = \alpha \overleftarrow{\text{prox}}_\lambda^\phi f_1(x) + (1 - \alpha) \overleftarrow{\text{prox}}_\lambda^\phi f_2(x)$ .

*Proof.* Apply Theorem 4.5(vii) and Proposition 2.10(iii). ■

The example below illustrates that Theorem 4.5 fails without the convexity of  $S_\phi$ .

**Example 4.8** For  $\phi(x) = |x|^3$ , simple calculus shows that  $S_\phi(x, y) = (3|x|x - 3|y|y)(x - y)$  is not convex on  $[0, +\infty]^2$ . Let  $\lambda = 1$ , and let  $a > 0$ ,  $f_1 := \iota_{\{a\}}$ ,  $f_2 := 0$  on  $\mathbb{R}$ . Then

$$(61) \quad \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x) = \alpha|a|^3 + \frac{|x - \alpha a|^3}{(1 - \alpha)^2} - |x|^3,$$

and  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  is not convex.

*Proof.* Because  $f_1, f_2 \in \Gamma_0(\mathbb{R}^n)$  and Theorem 3.3(i), we have

$$(62) \quad \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \left[ \alpha \left( f_1 + \frac{1}{\lambda} \phi \right) \left( \frac{\cdot}{\alpha} \right) \right] \square \left[ (1 - \alpha) \left( f_2 + \frac{1}{\lambda} \phi \right) \left( \frac{\cdot}{(1 - \alpha)} \right) \right] - \frac{1}{\lambda} \phi.$$

As  $\alpha(f_1 + \phi) \left( \frac{\cdot}{\alpha} \right) = \iota_{\{\alpha a\}} + \alpha\phi(a)$  and  $(1 - \alpha)(f_2 + \phi) \left( \frac{\cdot}{(1 - \alpha)} \right) = (1 - \alpha)\phi \left( \frac{\cdot}{(1 - \alpha)} \right)$ , by (62) we have

$$(63) \quad \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x) = \inf_y \left\{ \iota_{\{\alpha a\}}(y) + \alpha\phi(a) + (1 - \alpha)\phi \left( \frac{x - y}{1 - \alpha} \right) \right\} - \phi(x)$$

$$(64) \quad = \alpha\phi(a) + (1 - \alpha)\phi \left( \frac{x - \alpha a}{1 - \alpha} \right) - \phi(x).$$

Equations (61) is immediate from (64).

When  $x \geq \alpha a$ ,  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x) = \frac{(x - \alpha a)^3}{(1 - \alpha)^2} - x^3$ , so  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)''(x) = \frac{6(x - \alpha a)}{(1 - \alpha)^2} - 6x$ . As  $x \rightarrow \alpha a$ ,  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)''(x) < 0$ , so  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  is not convex. ■

It is naturally to ask: If  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  is convex for all  $f_1, f_2 \in \Gamma_0(\mathbb{R}^n)$  and  $\alpha \in ]0, 1[$ , what can we say about the Legendre function  $\phi$  or  $D_\phi$ ? This is partially answered by the following result on  $\mathbb{R}$ .

**Proposition 4.9** Let **A1–A5** hold. Suppose that  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  is convex for every  $\alpha \in ]0, 1[$ ,  $f_1, f_2 \in \Gamma_0(\mathbb{R})$ . Then  $D_\phi$  is separably convex on  $\mathbb{R}^2$ .

*Proof.* Note that

$$(65) \quad \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \left[ \alpha \left( f_1 + \frac{1}{\lambda} \phi \right) \left( \frac{\cdot}{\alpha} \right) \right] \square \left[ (1 - \alpha) \left( f_2 + \frac{1}{\lambda} \phi \right) \left( \frac{\cdot}{(1 - \alpha)} \right) \right] - \frac{1}{\lambda} \phi.$$

Let  $f_1 = \iota_{\{p\}}$  where  $p \in \text{dom } \phi$ , and  $f_2 \equiv 0$ . (65) gives

$$(\forall y \in U) \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(\alpha p + (1 - \alpha)y) = \frac{1}{\lambda} \left( \alpha \phi(p) + (1 - \alpha) \phi(y) - \phi(\alpha p + (1 - \alpha)y) \right).$$

Put  $g(y) = \alpha \phi(p) + (1 - \alpha) \phi(y) - \phi(\alpha p + (1 - \alpha)y)$ . By the assumption,  $g$  is convex for every  $\alpha \in ]0, 1[$ , so  $(\forall y \in U) g''(y) = (1 - \alpha) \phi''(y) - (1 - \alpha)^2 \phi''(\alpha p + (1 - \alpha)y) \geq 0$ . This implies  $\phi''(y) \geq (1 - \alpha) \phi''(\alpha p + (1 - \alpha)y)$ , from which

$$\begin{aligned} \phi''(y) - (1 - \alpha) \phi''(y) &\geq (1 - \alpha) [\phi''(\alpha p + (1 - \alpha)y) - \phi''(y)], \\ \phi''(y) &\geq (1 - \alpha) \frac{\phi''(y + \alpha(p - y)) - \phi''(y)}{\alpha}. \end{aligned}$$

When  $\alpha \downarrow 0$ , we obtain  $\phi''(y) \geq \phi'''(y)(p - y)$ , whence  $D_\phi$  is separably convex by [7, Theorem 3.3(ii)].  $\blacksquare$

## 5 Duality via Combettes and Reyes' anisotropic envelope and proximity operator

The Combettes-Reyes anisotropic envelope and proximity operator are essential in the study of the Fenchel conjugate of Bregman proximal averages.

**Theorem 5.1 (Duality of Bregman proximal average)** *Let **A1–A5** hold, and let  $f_i \in \Gamma_0(\mathbb{R}^n)$  for  $i = 1, 2$ . Then the following hold:*

- (i) *Suppose that  $(\forall i)$   $(\text{ri dom } f_i) \cap U \neq \emptyset$ , and that  $D_\phi$  is jointly convex. Then the anisotropic envelope and proximal mapping of  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)^*$  satisfy*

$$(66) \quad \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)^* \square (1/\lambda \star \phi^*) = \alpha f_1^* \square (1/\lambda \star \phi^*) + (1 - \alpha) f_2^* \square (1/\lambda \star \phi^*),$$

and  $\forall x^* \in \mathbb{R}^n$ ,

$$(67) \quad \begin{aligned} \nabla \phi^* \left( \lambda(x^* - \text{aprox}_{\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)^*}^{1/\lambda \star \phi^*}(x^*)) \right) \\ = \alpha \nabla \phi^* \left( \lambda(x^* - \text{aprox}_{f_1^*}^{1/\lambda \star \phi^*}(x^*)) \right) + (1 - \alpha) \nabla \phi^* \left( \lambda(x^* - \text{aprox}_{f_2^*}^{1/\lambda \star \phi^*}(x^*)) \right). \end{aligned}$$

- (ii) *Suppose that  $D_{\phi^*}$  is jointly convex. Then the anisotropic envelope and proximal mapping of  $\mathcal{P}_{1/\lambda}^{\phi^*}(f_1^*, f_2^*, \alpha)^*$  satisfy*

$$(68) \quad \mathcal{P}_{1/\lambda}^{\phi^*}(f_1^*, f_2^*, \alpha)^* \square (\lambda \star \phi) = \alpha f_1 \square (\lambda \star \phi) + (1 - \alpha) f_2 \square (\lambda \star \phi),$$

and  $\forall x \in [\alpha(\text{dom } f_1^*) + (1 - \alpha)(\text{dom } f_2^*) + \lambda U]$ ,

$$(69) \quad \begin{aligned} \nabla \phi \left( (x - \text{aprox}_{\mathcal{P}_{1/\lambda}^{\phi^*}(f_1^*, f_2^*, \alpha)^*}^{\lambda \star \phi}}(x))/\lambda \right) \\ = \alpha \nabla \phi \left( (x - \text{aprox}_{f_1^*}^{\lambda \star \phi}}(x))/\lambda \right) + (1 - \alpha) \nabla \phi \left( (x - \text{aprox}_{f_2^*}^{\lambda \star \phi}}(x))/\lambda \right). \end{aligned}$$

*Proof.* (i): By Fact 2.6,  $\phi^* = (\lambda f_i + \phi)^* + \lambda \overleftarrow{\text{env}}_{\lambda}^{\phi} f_i \circ \phi^*$ . Multiplying both sides by  $\alpha$  for  $i = 1$ , and  $(1 - \alpha)$  for  $i = 2$ , followed by adding both equations, we have

$$\phi^* - \lambda(\alpha \overleftarrow{\text{env}}_{\lambda}^{\phi} f_1 \circ \phi^* + (1 - \alpha) \overleftarrow{\text{env}}_{\lambda}^{\phi} f_2 \circ \phi^*) = \alpha(\lambda f_1 + \phi)^* + (1 - \alpha)(\lambda f_2 + \phi)^*.$$

Theorem 3.3(vi) gives  $\phi^* - \lambda \overleftarrow{\text{env}}_{\lambda}^{\phi} \mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha) \circ \phi^* = \alpha(\lambda f_1 + \phi)^* + (1 - \alpha)(\lambda f_2 + \phi)^*$ . Use Fact 2.6 again to obtain

$$(70) \quad (\lambda \mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha) + \phi)^* = \alpha(\lambda f_1 + \phi)^* + (1 - \alpha)(\lambda f_2 + \phi)^*.$$

Since  $(\text{ri dom } f_i) \cap U \neq \emptyset$  for  $i = 1, 2$ , by [40, Theorem 16.4] we can write

$$(71) \quad (\lambda f_i + \phi)^* = \lambda \star (f_i^* \square (1/\lambda \star \phi^*)),$$

where the  $\square$  is exact. Moreover, as  $\text{dom } \mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha) = \alpha \text{ dom } f_1 \cap \text{dom } \phi + (1 - \alpha) \text{ dom } f_2 \cap \text{dom } \phi$  by Theorem 3.3(ii), in view of [40, Theorems 6.5, 6.6] we have

$$(72) \quad \begin{aligned} \text{ri dom } \mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha) &= \alpha \text{ ri}(\text{dom } f_1 \cap \text{dom } \phi) + (1 - \alpha) \text{ ri}(\text{dom } f_2 \cap \text{dom } \phi) \\ &= \alpha(\text{ri dom } f_1) \cap U + (1 - \alpha)(\text{ri dom } f_2) \cap U \subseteq U. \end{aligned}$$

Because  $D_{\phi}$  is jointly convex,  $\mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha)$  is convex by Theorem 4.5. In view of (72), it follows from [40, Theorem 16.4] that

$$(73) \quad (\lambda \mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha) + \phi)^* = \lambda \star (\mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha)^* \square (1/\lambda \star \phi^*)),$$

and  $\square$  is exact. Combining (70), (71), and (73) gives (66).

Since  $\phi^*$  is differentiable, [9, Proposition 16.61(i)] or [37, Lemma 2.1] gives

$$(74) \quad \nabla[f_i^* \square (1/\lambda \star \phi^*)](x^*) = \nabla(1/\lambda \star \phi^*) \left( x^* - \text{aprox}_{f_i^*}^{1/\lambda \star \phi^*}(x^*) \right)$$

$$(75) \quad = \nabla \phi^* \left( \lambda(x^* - \text{aprox}_{f_i^*}^{1/\lambda \star \phi^*}(x^*)) \right), \text{ and}$$

$$(76) \quad \nabla[\mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha)^* \square (1/\lambda \star \phi^*)](x^*) = \nabla(1/\lambda \star \phi^*) \left( x^* - \text{aprox}_{\mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha)^*}^{1/\lambda \star \phi^*}(x^*) \right)$$

$$(77) \quad = \nabla \phi^* \left( \lambda(x^* - \text{aprox}_{\mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha)^*}^{1/\lambda \star \phi^*}(x^*)) \right).$$

Hence, (67) follows from (66) by taking derivatives both sides.

(ii): Note that  $\text{dom } \phi^* = \mathbb{R}^n$ . Apply (i) with  $f_i$  replaced by  $f_i^*$ ,  $\phi$  by  $\phi^*$  and  $\lambda$  by  $1/\lambda$ , followed by using Theorem 3.3(ii) and Proposition 2.15(i).  $\blacksquare$

**Remark 5.2** (1).  $D_{\phi}$  jointly convex does not mean  $D_{\phi^*}$  jointly convex. For example, for  $\phi(x) = x \ln x - x$  if  $x \geq 0$  and  $+\infty$  otherwise, and  $\phi^*(x) = \exp(x)$ ,  $D_{\phi}$  is jointly convex, but  $D_{\phi^*}$  is not. (2). In general,  $\mathcal{P}_{1/\lambda}^{\phi^*}(f_1^*, f_2^*, \alpha)^* \neq \mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha)$  because the latter might not be convex. While the anisotropic envelope of  $\mathcal{P}_{1/\lambda}^{\phi^*}(f_1^*, f_2^*, \alpha)^*$  is the convex combination of anisotropic envelopes of  $f_i^*$ 's, the Bregman envelope of  $\mathcal{P}_{\lambda}^{\phi}(f_1, f_2, \alpha)$  is the convex combination of Bregman envelopes of  $f_i$ 's.

**Remark 5.3** Note that  $(\forall f \in \Gamma_0(\mathbb{R}^n))(\forall x^* \in \mathbb{R}^n) \nabla \phi^*(x^* - \text{aprox}_{f^*}^{\phi^*}(x^*)) = \overleftarrow{\text{prox}}_1^{\phi} f(\nabla \phi^*(x^*))$  by Proposition 2.16. Thus, (67) is essentially an identity for proximal mappings, and the same can be said for (69).

## 6 Epi-continuity

This section is devoted to the epi-convergence behaviors of  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  when parameters  $\lambda$  and  $\alpha$  vary.

**Definition 6.1** *A sequence of functions  $(f_k)_{k \in \mathbb{N}}$  from  $\mathbb{R}^n \rightarrow ]-\infty, +\infty]$  epi-converges to  $f$  at a point  $x \in \mathbb{R}^n$  if both of the following conditions are satisfied:*

- (i) *whenever  $(x_k)_{k \in \mathbb{N}}$  converges to  $x$ , we have  $f(x) \leq \liminf_{k \rightarrow \infty} f_k(x_k)$ ;*
- (ii) *there exists a sequence  $(x_k)_{k \in \mathbb{N}}$  converges to  $x$  with  $f(x) = \lim_{k \rightarrow \infty} f_k(x_k)$ .*

If  $(f_k)_{k \in \mathbb{N}}$  epi-converges to  $f$  at every  $x \in C \subseteq \mathbb{R}^n$ , we say  $(f_k)_{k \in \mathbb{N}}$  epi-converges to  $f$  on  $C$ . In the case of  $C = \mathbb{R}^n$ , the functions  $f_k$  are said to epi-converge to  $f$ , denoted by  $f_k \xrightarrow{e} f$ .

See [41, pages 241-243] or [16, page 159] for further details on epi-convergence.

**Theorem 6.2 (epi-continuity I of Bregman proximal average)** *Let **A1–A5** hold. Then the following hold:*

- (i) *As  $\alpha \downarrow 0$ ,  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) \xrightarrow{e} \overleftarrow{\text{hul}}_\lambda^\phi f_2$  on  $U$ .*
- (ii) *As  $\alpha \uparrow 1$ ,  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) \xrightarrow{e} \overleftarrow{\text{hul}}_\lambda^\phi f_1$  on  $U$ .*

In particular, when  $f_1, f_2 \in \Gamma_0(\mathbb{R}^n)$ , we have

- (a) *As  $\alpha \downarrow 0$ ,  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) \xrightarrow{e} f_2$  on  $U$ .*
- (b) *As  $\alpha \uparrow 1$ ,  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) \xrightarrow{e} f_1$  on  $U$ .*

*Proof.* (i): By Proposition 2.4(i), each  $f_i + \frac{1}{\lambda}\phi$  is 1-coercive so that its Fenchel conjugate  $(f_i + \frac{1}{\lambda}\phi)^*$  has a full domain. When  $\alpha \downarrow 0$ ,

$$\left[ \alpha \left( f_1 + \frac{1}{\lambda}\phi \right)^* + (1 - \alpha) \left( f_2 + \frac{1}{\lambda}\phi \right)^* \right] \rightarrow \left( f_2 + \frac{1}{\lambda}\phi \right)^*$$

pointwise, so epi-converges by [41, Theorem 7.17]. By [41, Theorem 11.34],

$$\left[ \alpha \left( f_1 + \frac{1}{\lambda}\phi \right)^* + (1 - \alpha) \left( f_2 + \frac{1}{\lambda}\phi \right)^* \right]^*$$

epi-converges to  $(f_2 + \frac{1}{\lambda}\phi)^{**}$  on  $\mathbb{R}^n$ , so epi-converges at every point of  $U$ . Since  $\phi$  is continuous on  $U$ , in view of [41, Exercise 7.8],  $\left[ \alpha \left( f_1 + \frac{1}{\lambda}\phi \right)^* + (1 - \alpha) \left( f_2 + \frac{1}{\lambda}\phi \right)^* \right]^* - \frac{1}{\lambda}\phi$  epi-converges to  $(f_2 + \frac{1}{\lambda}\phi)^{**} - \frac{1}{\lambda}\phi$  on  $U$ , when  $\alpha \downarrow 0$ .

(ii): The proof is analogous to that of (i). Finally, (a)&(b) hold because Proposition 2.14(ii) implies  $\overleftarrow{\text{hul}}_\lambda^\phi f_i = f_i$  on  $U$  when  $f_i \in \Gamma_0(\mathbb{R}^n)$ . ■

The next result shows that the Bregman proximal average lies between the epi-average of convexified individual functions and the arithmetic average of individual functions.

**Theorem 6.3** *Let A1–A5 hold. Then the following hold:*

- (i)  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) \geq [\alpha \operatorname{conv} f_1 \left(\frac{\cdot}{\alpha}\right)] \square [(1 - \alpha) \operatorname{conv} f_2 \left(\frac{\cdot}{1 - \alpha}\right)]$ .
- (ii)  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) \leq \alpha f_1 + (1 - \alpha)f_2$  on  $\operatorname{dom} \phi$ . In particular,  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) \leq \alpha f_1 + (1 - \alpha)f_2$  if  $\operatorname{dom} f_1 \cap \operatorname{dom} f_2 \subseteq \operatorname{dom} \phi$ .

*Proof.* (i): Because  $\phi$  is convex, we have  $[\alpha \phi \left(\frac{\cdot}{\alpha}\right)] \square [(1 - \alpha) \phi \left(\frac{\cdot}{1 - \alpha}\right)] = \phi$  and  $\operatorname{conv} \phi = \phi$ . It follows from Theorem 3.3(i) that  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$

$$\begin{aligned}
&= \left[ \alpha \operatorname{conv} \left( f_1 + \frac{1}{\lambda} \phi \right) \left( \frac{\cdot}{\alpha} \right) \right] \square \left[ (1 - \alpha) \operatorname{conv} \left( f_2 + \frac{1}{\lambda} \phi \right) \left( \frac{\cdot}{1 - \alpha} \right) \right] - \frac{1}{\lambda} \phi \\
&\geq \left[ \alpha \left( \operatorname{conv} f_1 + \frac{1}{\lambda} \phi \right) \left( \frac{\cdot}{\alpha} \right) \right] \square \left[ (1 - \alpha) \left( \operatorname{conv} f_2 + \frac{1}{\lambda} \phi \right) \left( \frac{\cdot}{1 - \alpha} \right) \right] - \frac{1}{\lambda} \phi \\
&\geq \left[ \alpha \operatorname{conv} f_1 \left( \frac{\cdot}{\alpha} \right) \right] \square \left[ (1 - \alpha) \operatorname{conv} f_2 \left( \frac{\cdot}{1 - \alpha} \right) \right] + \left[ \alpha \frac{1}{\lambda} \phi \left( \frac{\cdot}{\alpha} \right) \right] \square \left[ (1 - \alpha) \frac{1}{\lambda} \phi \left( \frac{\cdot}{1 - \alpha} \right) \right] - \frac{1}{\lambda} \phi \\
&= \left[ \alpha \operatorname{conv} f_1 \left( \frac{\cdot}{\alpha} \right) \right] \square \left[ (1 - \alpha) \operatorname{conv} f_2 \left( \frac{\cdot}{1 - \alpha} \right) \right].
\end{aligned}$$

(ii): For every  $x \in \operatorname{dom} \phi$ , we have  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x)$

$$\begin{aligned}
&\leq \alpha \operatorname{conv} \left( f_1 + \frac{1}{\lambda} \phi \right) \left( \frac{\alpha x}{\alpha} \right) + (1 - \alpha) \operatorname{conv} \left( f_2 + \frac{1}{\lambda} \phi \right) \left( \frac{(1 - \alpha)x}{1 - \alpha} \right) - \frac{1}{\lambda} \phi(x) \\
&\leq \alpha \left( f_1 + \frac{1}{\lambda} \phi \right) \left( \frac{\alpha x}{\alpha} \right) + (1 - \alpha) \left( f_2 + \frac{1}{\lambda} \phi \right) \left( \frac{(1 - \alpha)x}{1 - \alpha} \right) - \frac{1}{\lambda} \phi(x) \\
&= \alpha f_1(x) + \alpha \frac{1}{\lambda} \phi(x) + (1 - \alpha) f_2(x) + (1 - \alpha) \frac{1}{\lambda} \phi(x) - \frac{1}{\lambda} \phi(x) = \alpha f_1(x) + (1 - \alpha) f_2(x).
\end{aligned}$$

■

**Theorem 6.4 (epi-continuity II of Bregman proximal average)** *Let A1–A5 hold. Define  $\tilde{f}_i := f_i + \iota_{\operatorname{dom} \phi}$  for  $i = 1, 2$ . Then the following hold:*

- (i) *For every  $x \in \mathbb{R}^n$ , the function  $\lambda \mapsto \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x)$  is monotonically decreasing on  $]0, \bar{\lambda}[$ .*
- (ii)  $\lim_{\lambda \uparrow \bar{\lambda}} \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \left[ \alpha \star \operatorname{conv} \left( f_1 + \frac{1}{\lambda} \phi \right) \right] \square \left[ (1 - \alpha) \star \operatorname{conv} \left( f_2 + \frac{1}{\lambda} \phi \right) \right] - \frac{1}{\lambda} \phi$  pointwise. In particular, for  $\bar{\lambda} = +\infty$  one has  $\lim_{\lambda \uparrow \infty} \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \left[ \alpha \star \operatorname{conv} \tilde{f}_1 \right] \square \left[ (1 - \alpha) \star \operatorname{conv} \tilde{f}_2 \right]$  pointwise; consequently,  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) \xrightarrow{e} \operatorname{cl} \left[ (\alpha \star \operatorname{conv} \tilde{f}_1) \square ((1 - \alpha) \star \operatorname{conv} \tilde{f}_2) \right]$  as  $\lambda \uparrow \infty$ .
- (iii)  $\lim_{\lambda \downarrow 0} \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \alpha f_1 + (1 - \alpha) f_2$  pointwise on  $U$ . Consequently, when  $\operatorname{dom} f_i \subseteq U$  for  $i = 1, 2$ ,  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) \xrightarrow{e} \alpha f_1 + (1 - \alpha) f_2$  as  $\lambda \downarrow 0$ .

*Proof.* We have  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x) =$

$$\begin{aligned}
& \inf_{u+v=x} \left[ \alpha \inf_{\substack{\sum_i \alpha_i x_i = \frac{u}{\alpha} \\ \sum_i \alpha_i = 1, \alpha_i \geq 0}} \sum_i \alpha_i \left( f_1(x_i) + \frac{1}{\lambda} \phi(x_i) \right) + (1-\alpha) \inf_{\substack{\sum_j \beta_j y_j = \frac{v}{1-\alpha} \\ \sum_j \beta_j = 1, \beta_j \geq 0}} \sum_j \beta_j \left( f_2(y_j) + \frac{1}{\lambda} \phi(y_j) \right) \right] \\
& - \frac{1}{\lambda} \phi(x) \\
& = \inf_{\substack{\alpha \sum_i \alpha_i x_i + (1-\alpha) \sum_j \beta_j y_j = x \\ \sum_i \alpha_i = 1, \sum_j \beta_j = 1, \alpha_i \geq 0, \beta_j \geq 0}} \left[ \alpha \sum_i \alpha_i f_1(x_i) + (1-\alpha) \sum_j \beta_j f_2(y_j) + \right. \\
(78) \quad & \left. \frac{1}{\lambda} \left( \alpha \sum_i \alpha_i \phi(x_i) + (1-\alpha) \sum_j \beta_j \phi(y_j) - \phi \left( \alpha \sum_i \alpha_i x_i + (1-\alpha) \sum_j \beta_j y_j \right) \right) \right].
\end{aligned}$$

The underbraced part is nonnegative because  $\phi$  is convex,  $\sum_i \alpha_i = 1, \sum_j \beta_j = 1, \alpha_i, \beta_j \geq 0$ .

(i): By (78),  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  is monotonically decreasing with respect to  $\lambda$  on  $]0, \bar{\lambda}[$ .

(ii): From (i) we obtain  $\lim_{\lambda \uparrow \bar{\lambda}} \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x) = \inf_{\bar{\lambda} > \lambda > 0} \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)(x) =$

$$\begin{aligned}
(79) \quad & \inf_{\bar{\lambda} > \lambda > 0} \inf_{\substack{\alpha \sum_i \alpha_i x_i + (1-\alpha) \sum_j \beta_j y_j = x \\ \sum_i \alpha_i = 1, \sum_j \beta_j = 1, \alpha_i \geq 0, \beta_j \geq 0}} \left[ \alpha \sum_i \alpha_i f_1(x_i) + (1-\alpha) \sum_j \beta_j f_2(y_j) + \right. \\
& \left. \frac{1}{\lambda} \left( \alpha \sum_i \alpha_i \phi(x_i) + (1-\alpha) \sum_j \beta_j \phi(y_j) - \phi \left( \alpha \sum_i \alpha_i x_i + (1-\alpha) \sum_j \beta_j y_j \right) \right) \right]
\end{aligned}$$

$$\begin{aligned}
(80) \quad & = \inf_{\substack{\alpha \sum_i \alpha_i x_i + (1-\alpha) \sum_j \beta_j y_j = x \\ \sum_i \alpha_i = 1, \sum_j \beta_j = 1, \alpha_i \geq 0, \beta_j \geq 0}} \inf_{\bar{\lambda} > \lambda > 0} \left[ \alpha \sum_i \alpha_i f_1(x_i) + (1-\alpha) \sum_j \beta_j f_2(y_j) + \right. \\
& \left. \frac{1}{\lambda} \left( \alpha \sum_i \alpha_i \phi(x_i) + (1-\alpha) \sum_j \beta_j \phi(y_j) - \phi \left( \alpha \sum_i \alpha_i x_i + (1-\alpha) \sum_j \beta_j y_j \right) \right) \right]
\end{aligned}$$

$$\begin{aligned}
(81) \quad & = \inf_{\substack{\alpha \sum_i \alpha_i x_i + (1-\alpha) \sum_j \beta_j y_j = x \\ \sum_i \alpha_i = 1, \sum_j \beta_j = 1, \alpha_i \geq 0, \beta_j \geq 0}} \left[ \alpha \sum_i \alpha_i f_1(x_i) + (1-\alpha) \sum_j \beta_j f_2(y_j) + \right. \\
& \left. \frac{1}{\lambda} \left( \alpha \sum_i \alpha_i \phi(x_i) + (1-\alpha) \sum_j \beta_j \phi(y_j) - \phi \left( \alpha \sum_i \alpha_i x_i + (1-\alpha) \sum_j \beta_j y_j \right) \right) \right] \\
& = \left[ \alpha \operatorname{conv} \left( f_1 + \frac{1}{\lambda} \phi \right) \left( \frac{\cdot}{\alpha} \right) \square (1-\alpha) \operatorname{conv} \left( f_2 + \frac{1}{\lambda} \phi \right) \left( \frac{\cdot}{1-\alpha} \right) \right] (x) - \frac{1}{\lambda} \phi(x).
\end{aligned}$$

The above arguments also apply for  $\bar{\lambda} = +\infty$ . The epi-convergence follows from [41, Proposition 7.4(c)].

(iii): By Theorem 3.3(vi), Proposition 2.14(iii) and Theorem 6.3, on  $U$  we have

$$\alpha f_1 + (1-\alpha) f_2 \geq \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) \geq \alpha \overleftarrow{\operatorname{env}}_\lambda^\phi f_1 + (1-\alpha) \overleftarrow{\operatorname{env}}_\lambda^\phi f_2.$$

The result follows by sending  $\lambda$  to 0 and applying Proposition 2.3.

When  $\operatorname{dom} f_i \subseteq U$  for  $i = 1, 2$ , we have  $\operatorname{dom} \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) \subseteq U$  by Theorem 3.3(ii). Then  $\lim_{\lambda \downarrow 0} \mathcal{P}_\lambda^\phi(f_1, f_2, \alpha) = \alpha f_1 + (1-\alpha) f_2$  on  $\mathbb{R}^n$ . Because  $\mathcal{P}_\lambda^\phi(f_1, f_2, \alpha)$  is increasing as  $\lambda \downarrow 0$ , the  $\xrightarrow{e}$  follows from [41, Theorem 7.4(d)]. ■

## Acknowledgments

The authors thank the editor M. Teboulle and two anonymous reviewers for helpful suggestions and constructive feedback. Xianfu Wang and Heinz Bauschke were partially supported by the Natural Sciences and Engineering Research Council of Canada.

## References

- [1] H. Attouch and D. Azé, Approximation and regularization of arbitrary functions in Hilbert spaces by the Lasry-Lions method, *Ann. Inst. H. Poincaré Anal. Non Linéaire* 10 (1993), 289–312.
- [2] J.-B. Baillon and G. Haddad, Quelques propriétés des opérateurs angle-bornés et  $n$ -cycliquement monotones, *Israel J. Math.* 26 (1977), 137–150.
- [3] S. Bartz, H. H. Bauschke, S. M. Moffat, and X. Wang, The resolvent average of monotone operators: dominant and recessive properties, *SIAM J. Optim.* 26 (2016), 602–634.
- [4] H. H. Bauschke and X. Wang, The kernel average for two convex functions and its application to the extension and representation of monotone operators, *Trans. Amer. Math. Soc.* 361 (2009), 5947–5965.
- [5] H.H. Bauschke, J. Bolte, J. Chen, M. Teboulle, and X. Wang, On linear convergence of non-Euclidean gradient methods without strong convexity and Lipschitz gradient continuity, *J. Optim. Theory Appl.* 182 (2019), 1068–1087.
- [6] H. H. Bauschke and J. M. Borwein, Legendre functions and the method of random Bregman projections, *J. Convex Anal.* 4 (1997), 27–67.
- [7] H. H. Bauschke and J. M. Borwein, Joint and separate convexity of the Bregman distance, *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications* (Haifa, 2000), 23–36, Stud. Comput. Math., 8, North-Holland, Amsterdam, 2001.
- [8] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, Bregman monotone optimization algorithms, *SIAM J. Control Optim.* 42 (2003), 596–636.
- [9] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed., Springer, Cham, 2017.
- [10] H. H. Bauschke, Y. Lucet, and M. Trienis, How to transform one convex function continuously into another, *SIAM Rev.* 50 (2008), 115–132.
- [11] H. H. Bauschke, X. Wang, J. Ye, and X. Yuan, Bregman distances and Chebyshev sets, *J. Approx. Theory* 159 (2009), 3–25.
- [12] H. H. Bauschke, P. L. Combettes, and D. Noll, Joint minimization with alternating Bregman proximity operators, *Pac. J. Optim.* 2 (2006), no. 3, 401–424.
- [13] H. H. Bauschke, M. N. Dao, and S. B. Lindstrom, Regularizing with Bregman-Moreau envelopes, *SIAM J. Optim.* 28 (2018), 3208–3228.
- [14] H. H. Bauschke, R. Goebel, Y. Lucet, and X. Wang, The proximal average: basic theory, *SIAM J. Optim.* 19 (2008), 766–785.
- [15] H. H. Bauschke, E. Matoušková, and S. Reich, Projection and proximal point methods: convergence results and counterexamples, *Nonlinear Anal.* 56 (2004), 715–738.
- [16] G. Beer, *Topologies on Closed and Closed Convex Sets*, Mathematics and Its Applications, Kluwer, Dordrecht, 1993.
- [17] J. Benoist and J.-B. Hiriart-Urruty, What is the subdifferential of the closed convex hull of a function? *SIAM J. Math. Anal.* 27 (1996), 1661–1679.
- [18] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization*, 2nd ed., Springer, New York, 2006.

- [19] J. M. Borwein and J. D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples*, Cambridge University Press, Cambridge, 2010.
- [20] M. N. Bui and P. L. Combettes, Warped proximal iterations for monotone inclusions, *J. Math. Anal. Appl.* 491 (2020), 124315, 21 pp.
- [21] D. Butnariu, Y. Censor, and S. Reich, Iterative averaging of entropic projections for solving stochastic convex feasibility problems, *Comput. Optim. Appl.* 8 (1997), 21–39.
- [22] D. Butnariu and A. N. Iusem, *Totally Convex Functions for Fixed Point Computation in Infinite Dimensional Optimization*, Kluwer, Dordrecht, 2000.
- [23] Y. Censor and S. A. Zenios, Proximal minimization algorithm with D-functions, *J. Optim. Theory Appl.* 73 (1992), 451–464.
- [24] G. Chen and M. Teboulle, Convergence analysis of a proximal-like minimization algorithm using Bregman functions, *SIAM J. Optim.* 3 (1993), 538–543.
- [25] J. Chen, X. Wang, and C. Planiden, A proximal average for prox-bounded functions, *SIAM J. Optim.* 30 (2020), 1366–1390.
- [26] Y. Chen, C. Kan, and W. Song, The Moreau envelope function and proximal mapping with respect to the Bregman distances in Banach spaces, *Vietnam J. Math.* 40 (2012), 181–199.
- [27] F. H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM, Philadelphia, 1990.
- [28] P. L. Combettes and N. N. Reyes, Moreau’s decomposition in Banach spaces, *Math. Program.* 139 (2013), Ser. B, 103–114.
- [29] R. Goebel, The proximal average for saddle functions and its symmetry properties with respect to partial and saddle conjugacy, *J. Nonlinear Convex Anal.* 11 (2010), 1–11.
- [30] W. L. Hare, A proximal average for nonconvex functions: a proximal stability perspective, *SIAM J. Optim.* 20 (2009), 650–666.
- [31] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms II*, Springer, New York, 1996.
- [32] C. Kan and W. Song, The Moreau envelope function and proximal mapping in the sense of the Bregman distance, *Nonlinear Anal.* 75 (2012), 1385–1399.
- [33] S. Kum and Y. Lim, The resolvent average on symmetric cones, *Linear Algebra Appl.* 438 (2013), 1159–1169.
- [34] E. Laude, P. Ochs, and D. Cremers, Bregman proximal mappings and Bregman-Moreau envelopes under relative prox-regularity, *J. Optim. Theory Appl.* 184 (2020), 724–761. <https://arxiv.org/abs/1907.04306>
- [35] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation I*, Springer-Verlag, Berlin, 2006.
- [36] J.-J. Moreau, Proximité et dualité dans un espace hilbertien, *Bull. Soc. Math. France* 93 (1965) 273–299.
- [37] J.-P. Penot, Proximal mappings, *J. Approx. Theory* 94 (1998), 203–221.
- [38] C. Planiden and X. Wang, Strongly convex functions, Moreau envelopes, and the generic nature of convex functions with strong minimizers, *SIAM J. Optim.* 26 (2016), 1341–1364.
- [39] M. D. Reid and R. C. Williamson, Information, divergence and risk for binary experiments, *J. Mach. Learn. Res.* 12 (2011), 731–817.
- [40] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.
- [41] R. T. Rockafellar and R. J-B Wets, *Variational Analysis*, Springer-Verlag, New York, 1998.
- [42] S. Simons, *Minimax and Monotonicity*, Lecture Notes in Mathematics, vol. 1693, Springer-Verlag, 1998.
- [43] Y. L. Yu, Better approximation and faster algorithm using the proximal average, *Advances in Neural Information Processing Systems (NeurIPS)*, 2013.
- [44] K. Zhang, E. Crooks, and A. Orlando, Compensated convexity methods for approximations and interpolations of sampled functions in Euclidean spaces: theoretical foundations, *SIAM J. Math. Anal.* 48 (2016), 4126–4154.