# Dualities for Non-Euclidean Smoothness and Strong Convexity under the Light of Generalized Conjugacy

Laude, Emanuel Department of Electrical Engineering (ESAT-STADIUS), KU Leuven

Themelis, Andreas Faculty of Information Science and Electrical Engineering (ISEE), Kyushu University

Patrinos, Panagiotis Department of Electrical Engineering (ESAT-STADIUS), KU Leuve

https://hdl.handle.net/2324/7151986

出版情報:SIAM Journal on Optimization. 33 (4), pp.2721-2749, 2023-12-31. Society for Industrial and Applied Mathematics バージョン: 権利関係:© 2022 Society for Industrial and Applied Mathematics

# DUALITIES FOR NON-EUCLIDEAN SMOOTHNESS AND STRONG CONVEXITY UNDER THE LIGHT OF GENERALIZED CONJUGACY\*

EMANUEL LAUDE<sup>†</sup>, ANDREAS THEMELIS<sup>‡</sup>, AND PANAGIOTIS PATRINOS<sup>†</sup>

**Abstract.** Relative smoothness and strong convexity have recently gained considerable attention in optimization. These notions are generalizations of the classical Euclidean notions of smoothness and strong convexity that are known to be dual to each other. However, conjugate dualities for non-Euclidean relative smoothness and strong convexity remain an open problem, as noted earlier by Lu, Freund, and Nesterov [*SIAM J. Optim.*, 28 (2018), pp. 333–354]. In this paper, we address this question by introducing the notions of anisotropic strong convexity and smoothness as the respective dual counterparts. The dualities are developed under the light of generalized conjugacy, which leads us to embed the anticipated dual notions within the superclasses of certain upper and lower envelopes. In contrast to the Euclidean case, these inclusions are proper in general, as showcased by means of counterexamples.

Key words. Bregman distance, generalized conjugacy, duality,  $\Phi$ -convexity

MSC codes. 26B25, 49J52, 49J53, 49M29, 49N15, 90C46

DOI. 10.1137/21M1465913

## 1. Introduction.

1.1. Motivation. An important duality result in optimization is that a convex function is differentiable with Lipschitz continuous (in the Euclidean sense) gradient mapping if and only if its convex conjugate is strongly convex, i.e., it remains convex after the subtraction of a positive multiple of the squared Euclidean norm; see Figure 1 (upper row). The notions of Euclidean smoothness and strong convexity can be generalized to *relative smoothness* [15, 5, 31] and *relative strong convexity* [31, 4], which have recently received considerable attention in the optimization literature [15, 5, 31, 22, 4, 1, 36, 16, 24]. However, it remains an open problem whether the known conjugate duality between Euclidean Lipschitz smoothness and Euclidean strong convexity can be generalized to the non-Euclidean case, as noted earlier in [31, section 3.4].

The goal of this paper is to close this gap by developing a full<sup>1</sup> conjugate duality for non-Euclidean smoothness and strong convexity. Following [37], we examine the duality under the light of  $\Phi$ -convexity [35], which typically appears in the context of eliminating duality gaps [40, 37, 3, 18] in nonconvex and nonsmooth optimization and optimal transport theory; see, e.g., [42]. We shall see that the sought duality

<sup>\*</sup>Received by the editors December 17, 2021; accepted for publication (in revised form) May 4, 2023; published electronically October 13, 2023.

https://doi.org/10.1137/21M1465913

**Funding:** The work of the first and third authors was supported by FWO projects G081222N, G033822N, and G0A0920N and the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant agreement 953348. The first author is supported by the postdoctoral mandate PDMt1/22/023. The second author is supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI grant JP21K17710.

<sup>&</sup>lt;sup>†</sup>Department of Electrical Engineering (ESAT-STADIUS), KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium (emanuel.laude@esat.kuleuven.be, panos.patrinos@esat.kuleuven.be).

<sup>&</sup>lt;sup>‡</sup>Faculty of Information Science and Electrical Engineering (ISEE), Kyushu University, 744 Motooka, Nishi-ku, 819-0395 Fukuoka, Japan (andreas.themelis@ees.kyushu-u.ac.jp).

<sup>&</sup>lt;sup>1</sup>Here, "full" means that we do not exclude nonsmoothness.

f L-Lipschitz smooth		$f^* \frac{1}{L}$ -strongly convex
[41, Prop. 12.60]		1[19, Thm. 7.2(b)]
$f = h \Box \tfrac{L}{2} \  \cdot \ ^2$	$\iff$	$f^* = (-g) \diamondsuit \tfrac{1}{2L} \  \cdot \ ^2$

FIG. 1. Euclidean conjugate dualities for proper, l.s.c., convex functions f with L > 0. Here,  $h \Box \frac{L}{2} \| \cdot \|^2 := \inf_{y \in \mathbb{R}^n} h(y) + \frac{L}{2} \| \cdot -y \|^2$  and  $(-g) \diamond \frac{1}{2L} \| \cdot \|^2 := \sup_{y \in \mathbb{R}^n} \frac{1}{2L} \| \cdot -y \|^2 - g(y)$  denote infimal and supremal convolutions of  $h : \mathbb{R}^n \to \overline{\mathbb{R}}$  (resp., -g) for  $g : \mathbb{R}^n \to \overline{\mathbb{R}}$  and a positive multiple of the squared Euclidean norm, respectively.

relation is not found between relative smoothness and relative strong convexity. Instead, it involves different notions of relative smoothness and strong convexity which are developed in this paper.

Overall, our contribution is twofold:

- (i) We furnish the dual counterparts of relative smoothness and relative strong convexity in terms of lower and upper (sub)gradient inequalities that we refer to, respectively, as *anisotropic strong convexity* and *anisotropic smoothness*. The latter is closely related to anisotropic prox-regularity introduced in [28, Definition 2.13]. Restricting to convex functions, we prove conjugate dualities between relative smoothness and anisotropic strong convexity as well as relative strong convexity and anisotropic smoothness.
- (ii) Following [37], the duality correspondences are developed under the light of generalized conjugacy. This leads us to study certain superclasses of the anticipated non-Euclidean notions of smoothness and strong convexity, examining the properness of these inclusions. In the primal, these superclasses correspond to the Bregman–Moreau [10, 9, 12, 27, 21, 29] and Bregman–Klee envelopes [11, 13] and in the dual to infimal and supremal convolutions, aka Klee envelopes [19, 26, 20]. The known equivalence between the notion of Euclidean weak convexity (aka paraconvexity, semiconvexity, or hypoconvexity) and the representation in terms of a negative Moreau envelope (see, e.g., [44]) also motivates the study of non-Euclidean generalizations of the class of weakly convex functions.

In the convex case, the aforementioned equivalences and implications are synopsized in Figure 2. Further restricting to the Euclidean setting, the relations simplify into the ones depicted in Figure 1.

**1.2.** Paper organization. The remainder of the paper is organized as follows. Section 2 offers an overview of basic notions of generalized conjugacy and convexity and lists some related results which will be used in what follows. Section 3 will then revise the classes of relatively smooth and relatively strongly convex functions under the lens of generalized conjugacy and introduce their *anisotropic* counterparts, which in the subsequent section 4 will be shown to be the sought dual classes. The shortcomings of  $\Phi$ -convexity in the duality picture will be ultimately showcased in section 5 by means of counterexamples. Section 6 concludes the paper.

**1.3. Preliminaries and notation.** With  $\mathbb{R}$  and  $\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty, -\infty\}$ , we denote the real and the extended real line, respectively. We denote by  $\langle \cdot, \cdot \rangle$  the standard Euclidean inner product on  $\mathbb{R}^n$  and by  $||x|| := \sqrt{\langle x, x \rangle}$  for any  $x \in \mathbb{R}^n$  the standard Euclidean norm on  $\mathbb{R}^n$ . The closed ball of radius r > 0 centered at  $x \in \mathbb{R}^n$  is denoted

2723

(a) Conjugate duality correspondences between  $B_{\phi}$ -smoothness and  $a_{\phi^*}$ -strong convexity as well as Bregman–Moreau envelopes and supremal convolutions for proper, lsc, convex functions  $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ . The implications are equivalences if, in addition, f is essentially strictly convex or, equivalently,  $f^*$  is essentially smooth, cf. Theorem 4.3.

$f$ a <sub><math>\phi</math></sub> -smooth	$\stackrel{\text{Thm. 4.6}}{\Longleftrightarrow}$	$f^* \operatorname{B}_{\phi^*}$ -strongly convex $\stackrel{\operatorname{Ex. 3.6}}{\not\Longrightarrow} f \operatorname{B}_{\phi}$ -smooth	
Thm. 4.6 $\bigoplus$			
$f=h \ \Box \ \phi$	$\iff$	$f^* = \overrightarrow{\mathrm{klee}}^{\phi^*} g$	

(b) Conjugate duality correspondences between  $B_{\phi^*}$ -strong convexity and  $a_{\phi}$ -smoothness as well as Bregman–Klee envelopes and infimal convolutions for proper, lsc, convex functions  $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ .

FIG. 2. Schematics of the interrelations between smoothness, strong convexity, and certain pointwise min/max representations for proper, lower semicontinuous (l.s.c.), convex functions.  $B_{\phi^-}$ and  $B_{\phi^*}$  - are short for relative to  $\phi$ , as in Assumptions (A1)–(A3), and to its conjugate  $\phi^*$ , respectively, in the classical Bregman sense; see Definition 3.1. Similarly,  $a_{\phi^-}$  and  $a_{\phi^*}$  - refer to the anisotropic counterparts defined in Definition 3.8. Here,  $\overrightarrow{env}^{\phi}g := \inf_{x \in \mathbb{R}^n} D_{\phi}(x, \cdot) + g(x)$  and  $\overrightarrow{klee}^{\phi}g := \sup_{x \in \mathbb{R}^n} D_{\phi}(x, \cdot) - g(x)$  are (right) Bregman–Moreau and Bregman–Klee envelopes of  $g : \mathbb{R}^n \to \mathbb{R}$ , respectively, and  $D_{\phi}(x, y) := \phi(x) - \phi(y) - \langle \nabla \phi(y), x - y \rangle$  is the Bregman distance generated by  $\phi$ .

as  $\overline{\mathbb{B}}(x;r)$ . For a set  $S \subseteq \mathbb{R}^n$ , int  $S \subseteq S$  denotes its interior, con  $S \supseteq S$  its convex hull, and  $\delta_S : \mathbb{R}^n \to \overline{\mathbb{R}}$  its indicator function, namely,  $\delta_S(x) = 0$  if  $x \in S$  and  $\infty$  otherwise.

The notation  $T: \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  indicates a set-valued mapping, whose domain, graph, and range are defined as dom  $T = \{x \in \mathbb{R}^n : T(x) \neq \emptyset\}$ , gph  $T = \{(x, y) : y \in T(x)\}$ , and range  $T = \bigcup_{x \in \mathbb{R}^n} T(x)$ , respectively. We say that T is locally bounded at  $\bar{x}$  if there exists a neighborhood  $\mathcal{N}_{\bar{x}}$  of  $\bar{x}$  such that  $\bigcup_{x \in \mathcal{N}_{\bar{x}}} T(x)$  is bounded.

The effective domain of an extended real-valued function  $f: \mathbb{R}^n \to \overline{\mathbb{R}}$  is denoted by dom  $f := \{x \in \mathbb{R}^n : f(x) < \infty\}$ , and we say that f is proper if  $f(x) > -\infty$  for all  $x \in \mathbb{R}^n$  and dom  $f \neq \emptyset$ , lower semicontinuous (l.s.c.) if  $f(\overline{x}) \leq \liminf_{x \to \overline{x}} f(x)$  for all  $\overline{x} \in \mathbb{R}^n$ , and supercoercive if  $f(x)/||x|| \to \infty$  as  $||x|| \to \infty$ . The convex conjugate of fis denoted as  $f^* := \sup_{y \in \mathbb{R}^n} \{\langle \cdot, y \rangle - f(y)\}$ . We say that f is strictly continuous at a point  $\overline{x}$  where f is finite if  $\limsup_{x,x' \to \overline{x}} \frac{|f(x) - f(x')|}{||x - x'||} < \infty$ . With  $\mathcal{C}^k(\mathbb{R}^n)$ , we indicate the set of functions from  $\mathbb{R}^n$  to  $\mathbb{R}$  which are k times continuously differentiable while with  $\mathbb{E}_{-}(\mathbb{R}^n)$  the extended real valued areas which are proper labels of a super-

with  $\Gamma_0(\mathbb{R}^n)$  the extended real-valued ones which are proper, l.s.c., and convex. The identity function  $x \mapsto x$  on  $\mathbb{R}^n$  is denoted by id. The set-valued mappings  $\widehat{\partial} f \, \partial f : \mathbb{R}^n \to \mathbb{R}^n$  are the regular and the limiting

The set-valued mappings  $\widehat{\partial}f, \partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  are the regular and the limiting subdifferential of f, where  $\bar{v} \in \widehat{\partial}f(\bar{x})$  if  $\liminf_{\substack{x \to \bar{x} \\ x \neq \bar{x}}} \frac{f(x) - f(\bar{x}) - \langle \bar{v}, x - \bar{x} \rangle}{\|x - \bar{x}\|} \ge 0$ , while  $\bar{v} \in \partial f(\bar{x})$  if  $\bar{x} \in \text{dom } f$  and there exists a sequence  $\text{gph} \widehat{\partial}f \ni (x^k, v^k) \to (\bar{x}, \bar{v})$  such that

Copyright (c) by SIAM. Unauthorized reproduction of this article is prohibited.

 $f(x^k) \to f(\bar{x})$ . If  $g \in \mathcal{C}^1(\mathbb{R}^n)$ , then  $\widehat{\partial}(f+g) = \widehat{\partial}f + \nabla g$  and  $\partial(f+g) = \partial f + \nabla g$  [41, Exercise 8.8(c)]. The inclusion  $0 \in \widehat{\partial}f(\bar{x})$  is necessary for local minimality of f at  $\bar{x}$  [41, Theorem 10.1]. We define  $\varepsilon$ -arg min  $f := \{x : f(x) \leq \inf f + \varepsilon\}$ . We adopt the notions of essential smoothness, essential strict convexity, and Legendre-type functions from [39, section 26]. We say that a function  $f \in \Gamma_0(\mathbb{R}^n)$  is essentially smooth if  $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$ , f is differentiable on  $\operatorname{int}(\operatorname{dom} f)$  and  $\|\nabla f(x^{\nu})\| \to \infty$  whenever  $\operatorname{int}(\operatorname{dom} f) \ni x^{\nu} \to x \in$ bdry dom f, essentially strictly convex if f is strictly convex on every convex subset of dom  $\partial f$ , and Legendre if f is both essentially smooth and essentially strictly convex.

**2. Generalized conjugacy.** In this section, we recapitulate the notions of  $\Phi$ -convexity and  $\Phi$ -conjugacy [35, 2, 23, 41, 43], which will prove valuable tools in what follows. Throughout this section, X and Y are nonempty sets, and  $\Phi: X \times Y \to \mathbb{R}$  is a real-valued function called pairing (or coupling). Although this setting is general enough for our purposes, we emphasize that all the results are true even for extended real-valued couplings, up to the adoption of suitable extended arithmetics to account for indeterminate forms; see, e.g., [34, 41].

DEFINITION 2.1 ( $\Phi$ -convexity). We say that  $f: X \to \overline{\mathbb{R}}$  is (left)  $\Phi$ -convex if there is  $\{(y_i, \beta_i) \in Y \times \overline{\mathbb{R}} : i \in \mathcal{I}\}$  for some index set  $\mathcal{I}$  such that

(2.1) 
$$f(x) = \sup_{i \in \mathcal{I}} \Phi(x, y_i) - \beta_i \quad \forall x \in X.$$

When  $\mathcal{I} = \emptyset$ , we define  $f \equiv -\infty$ . Likewise, we say that  $g: Y \to \overline{\mathbb{R}}$  is (right)  $\Phi$ -convex if there is  $\{(x_j, \alpha_j) \in X \times \overline{\mathbb{R}} : j \in \mathcal{J}\}$  for some index set  $\mathcal{J}$  such that for all  $y \in Y$ ,

(2.2) 
$$g(y) = \sup_{j \in \mathcal{J}} \Phi(x_j, y) - \alpha_j \quad \forall y \in Y.$$

When  $\mathcal{J} = \emptyset$ , we define  $g \equiv -\infty$ .

2724

Sometimes  $\Phi$ -convex functions are also referred to as  $\Phi$ -envelopes. Note that if  $X = Y = \mathbb{R}^n$  and  $\Phi = \langle \cdot, \cdot \rangle$  is the Euclidean inner product, one recovers the class of proper convex l.s.c. functions. This condition, however, differs from the classical secant line definition of convexity, which also includes functions which are not l.s.c. More generally, a typical choice for the coupling is the negative squared Euclidean norm  $\Phi(x, y) = -\frac{1}{2\lambda} ||x - y||^2$  for  $\lambda > 0$ , which leads to the proximal hull [41, Example 1.44]. Another interesting choice is  $\Phi(x, y) := \langle v, x \rangle - \frac{r}{2} ||x||^2$  for y = (v, r). Under this coupling, all proper l.s.c. and prox-bounded functions are  $\Phi$ -convex [41, Example 11.66]. Further examples can be found in [2, 23, 43].

DEFINITION 2.2 ( $\Phi$ -conjugate functions). The (left)  $\Phi$ -conjugate of  $f: X \to \overline{\mathbb{R}}$  on Y is  $f^{\Phi}: Y \to \overline{\mathbb{R}}$ , defined by

(2.3) 
$$f^{\Phi}(y) := \sup_{x \in X} \Phi(x, y) - f(x),$$

and the (left)  $\Phi$ -biconjugate of f back on X is  $f^{\Phi\Phi}: X \to \overline{\mathbb{R}}$ , defined by

(2.4) 
$$f^{\Phi\Phi}(x) := \sup_{y \in Y} \Phi(x, y) - f^{\Phi}(y).$$

Likewise, the (right)  $\Phi$ -conjugate of  $g: Y \to \overline{\mathbb{R}}$  on X is defined by

(2.5) 
$$g^{\Phi}(x) := \sup_{y \in Y} \Phi(x, y) - g(y),$$

(

and the (right)  $\Phi$ -biconjugate of g back on Y is given by

(2.6) 
$$g^{\Phi\Phi}(y) := \sup_{x \in X} \Phi(x, y) - g^{\Phi}(x)$$

Once again specializing the definition to  $X = Y = \mathbb{R}^n$  and  $\Phi = \langle \cdot, \cdot \rangle$ , one recovers the familiar convex conjugate and biconjugate operations. If  $\Phi(x,y) = -\frac{1}{2\lambda} \|x-y\|^2$ for  $\lambda > 0$  is the negative squared Euclidean norm,  $\Phi$ -conjugacy yields the proximal transform [41, Example 11.64]. Another choice for  $\Phi$  is -||x - y||, which leads to the negative Pasch–Hausdorff envelopes; see, e.g., [41, Example 9.11]. The corresponding class of  $\Phi$ -conjugates is the class of globally Lipschitz functions and plays an important role in the Kantorovich–Rubinstein duality in optimal transport theory; see, e.g., [42, Remark 6.5]. Convex subgradients can similarly be captured as a special case of  $\Phi$ subgradients, defined next; see, e.g., [2, Remark 3.3].

DEFINITION 2.3 ( $\Phi$ -subgradients and  $\Phi$ -subdifferential). We say that  $\bar{y} \in Y$  is a (right)  $\Phi$ -subgradient of  $f: X \to \mathbb{R}$  at  $\bar{x} \in \text{dom } f$ , denoted by  $\bar{y} \in \partial_{\Phi} f(\bar{x})$ , if the generalized subgradient inequality

(2.7) 
$$f(x) \ge f(\bar{x}) + \Phi(x, \bar{y}) - \Phi(\bar{x}, \bar{y}) \quad \forall x \in X$$

holds or, equivalently, if  $\bar{x} \in \arg \max_{x \in X} \{\Phi(x, \bar{y}) - f(x)\}$ . We call the set  $\partial_{\Phi} f(\bar{x})$  of all subgradients  $\bar{y}$  of f at  $\bar{x}$  the (right)  $\Phi$ -subdifferential of f at  $\bar{x}$ .

For  $g: Y \to \mathbb{R}$ , the definition of a (left)  $\Phi$ -subgradient and (left)  $\Phi$ -subdifferential is parallel.

The remainder of this section collects a list of useful properties of generalized convexity which will be invoked later in the manuscript. Although quite standard (see, e.g., [2, 23, 41, 43]), these results are instrumental in our analysis, and for completeness, we will thus detail the simple proofs. We confine the discussion to "left" functions  $f: X \to \overline{\mathbb{R}}$ ; the results for "right" functions  $g: Y \to \overline{\mathbb{R}}$  are parallel.

The following result can be regarded as a generalization of the *Fenchel-Moreau* theorem (see, e.g., [33, Theorem 2.78]).

LEMMA 2.4. For any  $f: X \to \overline{\mathbb{R}}$ , the following properties hold:

- (i)  $f^{\Phi}: Y \to \overline{\mathbb{R}}$  is right  $\Phi$ -convex.
- (ii)  $f^{\Phi\Phi}: X \to \overline{\mathbb{R}}$  is the largest left  $\Phi$ -convex function majorized by f; namely, for

$$\mathcal{A}_f := \{ (y, \beta) \in Y \times \overline{\mathbb{R}} : \Phi(x, y) - \beta \le f(x) \quad \forall x \in X \}$$

we have

$$f(\bar{x}) \ge f^{\Phi\Phi}(\bar{x}) = \sup_{(y,\beta) \in \mathcal{A}_f} \left\{ \Phi(\bar{x}, y) - \beta \right\} \quad \forall \bar{x} \in X,$$

where the inequality holds with equality if, in addition, f is  $\Phi$ -convex.

- (iii)  $(f^{\Phi\Phi})^{\Phi} = f^{\Phi}$ .
- (iv) ( $\Phi$ -Fenchel-Young inequality)  $f(x) + f^{\Phi}(y) \ge \Phi(x, y)$  for every  $(x, y) \in X \times Y$ . Proof.

2.4(i) Trivial (compare (2.3) and (2.2)).

2.4(ii)  $f^{\Phi\Phi}$  is left  $\Phi$ -convex by assertion 2.4(i), its being the right  $\Phi$ -conjugate of  $f^{\Phi}$ . Moreover, for  $\bar{x} \in X$ , one has

$$\begin{split} f^{\Phi\Phi}(\bar{x}) &\stackrel{(def)}{=} \sup_{y \in Y} \left\{ \Phi(\bar{x}, y) - f^{\Phi}(y) \right\} = \sup_{(y, \beta) \in Y \times \overline{\mathbb{R}}} \left\{ \Phi(\bar{x}, y) - \beta : f^{\Phi}(y) \leq \beta \right\} \\ &= \sup_{(y, \beta) \in \mathcal{A}_f} \left\{ \Phi(\bar{x}, y) - \beta \right\}, \end{split}$$

where the last identity holds since  $f^{\Phi}(y) \stackrel{(def)}{=} \sup_{x \in X} \Phi(x, y) - f(x) \leq \beta$  if and only if  $(y, \beta) \in \mathcal{A}_f$ . In particular, this implies the inequality  $f(\bar{x}) \geq f^{\Phi\Phi}(\bar{x})$ . If, in addition, f is  $\Phi$ -convex, there is  $\{(y_i, \beta_i) \in Y \times \mathbb{R} : i \in \mathcal{I}\}$  for some index set  $\mathcal{I}$  such that  $f = \sup_{i \in \mathcal{I}} \Phi(\cdot, y_i) - \beta_i$ . Therefore, for any  $i \in \mathcal{I}$ , we have  $f(x) \geq \Phi(x, y_i) - \beta_i$  for all  $x \in X$  and thus  $\mathcal{I} \subseteq \mathcal{A}_f$ . This implies that  $f(\bar{x}) \leq \sup_{(y,\beta)\in\mathcal{A}_f} \{\Phi(\bar{x}, y) - \beta\} = f^{\Phi\Phi}(\bar{x})$  and thus  $f(\bar{x}) = f^{\Phi\Phi}(\bar{x})$  as claimed.

2.4(iii) By definition,  $f^{\Phi\Phi} = (f^{\Phi})^{\Phi}$ . Thus, we have  $(f^{\Phi})^{\Phi\Phi} = ((f^{\Phi})^{\Phi})^{\Phi} = (f^{\Phi\Phi})^{\Phi}$ , and the claim follows from assertions 2.4(i) and 2.4(ii) since  $f^{\Phi}$  is right  $\Phi$ -convex and as such coincides with its biconjugate.

2.4(iv) Follows from the definition of the left  $\Phi$ -conjugate; cf. (2.3).

LEMMA 2.5. For any  $f: X \to \overline{\mathbb{R}}$ , the following are equivalent:

- (i) f is left  $\Phi$ -convex.
- (ii)  $f^{\Phi\Phi} = f$ .
- (iii)  $f = g^{\Phi}$  for some  $g: Y \to \overline{\mathbb{R}}$ .

Proof.

2726

 $2.5(i) \Rightarrow 2.5(ii)$  Follows from Lemma 2.4(ii).

 $2.5(ii) \Rightarrow 2.5(iii)$  Holds with  $q = f^{\Phi}$ .

 $2.5(\text{iii}) \Rightarrow 2.5(\text{i})$  Trivial since  $g^{\Phi}$  is left  $\Phi$ -convex by Lemma 2.4(i).

LEMMA 2.6 (Fenchel  $\Phi$ -duality). Let  $f: X \to \overline{\mathbb{R}}$  be proper. For any  $\overline{x} \in X$  and  $\overline{y} \in Y$ , the following statements are equivalent:

(i)  $\bar{y} \in \partial_{\Phi} f(\bar{x})$ .

- (ii)  $f(\bar{x}) + f^{\Phi}(\bar{y}) = \Phi(\bar{x}, \bar{y}) \in \mathbb{R}.$
- (iii)  $\bar{x} \in \operatorname{arg\,min}_{x \in X} \{ f(x) \Phi(x, \bar{y}) \}.$
- (iv)  $\bar{x} \in \partial_{\Phi} f^{\Phi}(\bar{y}) \cap \operatorname{dom} \partial_{\Phi} f.$

Any one of these equivalent conditions implies that  $f(\bar{x}) = f^{\Phi\Phi}(\bar{x}) \in \mathbb{R}$  and  $\bar{y} \in \partial_{\Phi} f^{\Phi\Phi}(\bar{x})$ . Moreover, when f is left  $\Phi$ -convex, the intersection with dom  $\partial_{\Phi} f$  is superfluous in the last condition.

*Proof.* The equivalence of the first three assertions follows from the definitions of  $\Phi$ -conjugate and  $\Phi$ -subdifferential (cf. (2.3) and (2.7)), and in particular, each implies the inclusion  $\bar{x} \in \text{dom } \partial_{\Phi} f$ . Suppose now that 2.6(ii) holds. Then we have

$$f(\bar{x}) \ge f^{\Phi\Phi}(\bar{x}) \stackrel{(def)}{=} \sup_{y \in Y} \Phi(\bar{x}, y) - f^{\Phi}(y) \ge \Phi(\bar{x}, \bar{y}) - f^{\Phi}(\bar{y}) = f(\bar{x}),$$

where the first inequality holds due to Lemma 2.4(ii) and the last equality is assertion 2.6(ii). All inequalities thus hold as equalities showing that  $f(\bar{x}) = f^{\Phi\Phi}(\bar{x})$  and  $\bar{y} \in \arg \max \Phi(\bar{x}, \cdot) - f^{\Phi}$ , which means by definition that  $\bar{x} \in \partial_{\Phi} f^{\Phi}(\bar{y})$ . This concludes the proof of implications 2.6(i)  $\Leftrightarrow$  2.6(ii)  $\Leftrightarrow$  2.6(iii)  $\Rightarrow$  2.6(iv).

Suppose now that assertion 2.6(iv) holds. Since  $\bar{x} \in \text{dom} \partial_{\Phi} f$ , there exists  $\eta \in \partial_{\Phi} f(\bar{x})$ , which by the equivalence of the first two assertions implies that  $f(\bar{x}) = \Phi(\bar{x},\eta) - f^{\Phi}(\eta) \in \mathbb{R}$ . Similarly, the inclusion  $\bar{x} \in \partial_{\Phi} f^{\Phi}(\bar{y})$  implies, through the same equivalence applied to  $f^{\Phi}: Y \to \overline{\mathbb{R}}$ , that  $f^{\Phi}(\bar{y}) + f^{\Phi\Phi}(\bar{x}) = \Phi(\bar{x},\bar{y})$ . Then,

$$f(\bar{x}) = \Phi(\bar{x},\eta) - f^{\Phi}(\eta) \le \sup\left\{\Phi(\bar{x}, \cdot) - f^{\Phi}\right\} \stackrel{(def)}{=} f^{\Phi\Phi}(\bar{x}) \stackrel{2.4(ii)}{\le} f(\bar{x}),$$

Therefore,  $\Phi(\bar{x}, \bar{y}) - f^{\Phi}(\bar{y}) = f^{\Phi\Phi}(\bar{x}) = f(\bar{x}) \in \mathbb{R}$ , which is assertion 2.6(ii).

To conclude, observe that  $\bar{y} \in \partial_{\Phi} f(\bar{x}) \Rightarrow \bar{x} \in \partial_{\Phi} f^{\Phi}(\bar{y}) \Rightarrow \bar{y} \in \partial_{\Phi} f^{\Phi\Phi}(\bar{x})$  due to the shown equivalence of the first and last assertions, applied first to  $f: X \to \overline{\mathbb{R}}$  and then to  $f^{\Phi}: Y \to \overline{\mathbb{R}}$ . If f is left  $\Phi$ -convex, then  $f = f^{\Phi\Phi}$  by Lemma 2.5, and the above

chain of implications then reduces to  $\bar{y} \in \partial_{\Phi} f(\bar{x}) \Leftrightarrow \bar{x} \in \partial_{\Phi} f^{\Phi}(\bar{y})$ , without domain inclusion.

#### 3. Bregman and anisotropic convexity and smoothness.

**3.1. Standing assumptions.** The notions considered in the remainder of this manuscript involve a *reference function*  $\phi$ , which complies with the following standing requirements and will be henceforth assumed without further mention:

(A1)  $\phi \in \Gamma_0(\mathbb{R}^m)$  is Legendre.

 $(\mathbf{A2}) \operatorname{dom} \phi = \mathbb{R}^n.$ 

(A3)  $\phi$  is supercoercive.

Assumptions (A2) and (A3) are made for simplicity so as to avoid treatment of complicating boundary cases and extended arithmetics. In view of [6, Proposition 2.16], the convex conjugate  $\phi^*$  of  $\phi$  also complies with the same requirements, and the gradients  $\nabla \phi, \nabla \phi^* : \mathbb{R}^n \to \mathbb{R}^n$  are mutually inverse bijections:  $(\nabla \phi)^{-1} = \nabla \phi^*$ . In particular, Assumptions (A1)–(A3) are equivalent to  $\phi$  and  $\phi^*$  both being strictly convex and differentiable (and supercoercive) with full domain.

**3.2. Bregman convexity and smoothness.** In this subsection, we revisit the existing notions of relative strong/weak convexity [21, 31, 4] and relative smoothness [15, 5, 31] and collect some basic properties. In addition, we provide alternative characterizations in terms of  $\Phi$ -convexity. To begin with, we introduce the Bregman distance [17]  $D_{\phi}$  generated by the reference function  $\phi$ , which, thanks to the full domain property of  $\phi$ , is defined as

$$D_{\phi}(x,y) := \phi(x) - \phi(y) - \langle \nabla \phi(y), x - y \rangle$$

for any  $x, y \in \mathbb{R}^n$ . Next, we recall the definitions of relative weak convexity [21], relative strong convexity [31, 4], and a two-sided version [16] of relative smoothness [15, 5, 31]. Since our dual notions involve a reference function as well, we will refine the existing terminology and henceforth refer to relative weak/strong convexity and smoothness as Bregman weak/strong convexity and smoothness ( $B_{\phi}$ -weak/ $B_{\phi}$ -strong convexity and  $B_{\phi}$ -smoothness).

DEFINITION 3.1 (Bregman weak/strong convexity and smoothness). A function  $f: \mathbb{R}^n \to \overline{\mathbb{R}}$  is said to be

- (i)  $B_{\phi}$ -weakly convex if  $f + \phi$  is convex;
- (ii)  $B_{\phi}$ -strongly convex if  $f \phi$  is convex;
- (iii)  $B_{\phi}$ -smooth if both f and -f are proper and  $B_{\phi}$ -weakly convex.

A source of practical examples for  $B_{\phi}$ -weakly convex functions is the class of compositions  $c \circ F : \mathbb{R}^n \to \mathbb{R}$  for a convex and globally Lipschitz function  $c : \mathbb{R}^m \to \mathbb{R}$  and a mapping  $F : \mathbb{R}^n \to \mathbb{R}^m$  whose component functions  $F_i : \mathbb{R}^n \to \mathbb{R}$  are  $B_{\phi}$ -smooth; see, e.g., [29, 36].

Next we provide basic equivalent characterizations of  $B_{\phi}$ -weak/strong convexity and  $B_{\phi}$ -smoothness in terms of lower and upper (sub)gradient inequalities. The characterization of  $B_{\phi}$ -smoothness is a specialization of the two-sided *full extended descent lemma* [16, Lemma 2.1] to  $\phi$  with full domain; see [5, 31] for the one-sided version. For a joint treatment, we introduce a flag  $r \in \{\pm 1\}$  such that convexity of  $f - r\phi$ means either  $B_{\phi}$ -weak (r = -1) or  $B_{\phi}$ -strong (r = +1) convexity.

PROPOSITION 3.2. Let  $f : \mathbb{R}^n \to \overline{\mathbb{R}}$  be proper and l.s.c., and let r = -1 (resp., r = +1). The following are equivalent:

(i) f is  $B_{\phi}$ -weakly convex (resp.,  $B_{\phi}$ -strongly convex); that is,  $f - r\phi$  is convex.

Downloaded 10/13/23 to 133.5.36.93 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

(ii) For every  $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$ , it holds that

(3.1) 
$$f(x) \ge f(\bar{x}) + \langle \bar{v}, x - \bar{x} \rangle + r \operatorname{D}_{\phi}(x, \bar{x}) \quad \forall x \in \mathbb{R}^{n}.$$

(iii)  $\langle x - x', v - v' \rangle \ge r \langle x - x', \nabla \phi(x) - \nabla \phi(x') \rangle$  holds for all  $(x, v), (x', v') \in \operatorname{gph} \partial f$ . Moreover, f is  $B_{\phi}$ -smooth if and only if it is continuously differentiable and satisfies

(3.2) 
$$|f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x} \rangle| \le \mathcal{D}_{\phi}(x, \bar{x}) \quad \forall x \in \mathbb{R}^n.$$

*Proof.* We start by observing that since  $\phi$  is smooth,  $\partial(f - r\phi) = \partial f - r\nabla\phi$ . 3.2(i)  $\Rightarrow$  3.2(ii) For  $(\bar{x}, \bar{v}) \in \text{gph} \partial f$ , convexity of  $f - r\phi$  yields

$$(f - r\phi)(x) \ge (f - r\phi)(\bar{x}) + \langle \bar{v} - r\nabla\phi(\bar{x}), x - \bar{x} \rangle \quad \forall x \in \mathbb{R}^n,$$

which after reordering is (3.1).

 $3.2(ii) \Rightarrow 3.2(iii)$  For  $(x, v), (x', v') \in \operatorname{gph} \partial f$ , (3.1) yields

$$f(x') \ge f(x) + \langle v, x' - x \rangle + r \operatorname{D}_{\phi}(x', x) \text{ and } f(x) \ge f(x') + \langle v', x - x' \rangle + r \operatorname{D}_{\phi}(x, x').$$

By summing the two inequalities and reordering, assertion 3.2(iii) is obtained.

 $3.2(\text{iii}) \Rightarrow 3.2(\text{i})$  By assumption, for all  $(x, v), (x', v') \in \text{gph} \partial f$ , it holds that  $\langle x - x', v - v' \rangle \geq r \langle x - x', \nabla \phi(x) - \nabla \phi(x') \rangle$ . By suitably rearranging, this can equivalently be written as  $\langle x - x', v - r \nabla \phi(x) - (v' - r \nabla \phi(x')) \rangle \geq 0$ . Let now two pairs  $(x, y), (x', y') \in \text{gph} \partial (f - r\phi)$  be fixed. Since  $\partial (f - r\phi) = \partial f - r \nabla \phi$ , this means that  $y = v - r \nabla \phi(x)$  for some  $v \in \partial f(x)$  and  $y' = v' - r \nabla \phi(x)$  for some  $v' \in \partial f(x')$ . In light of the previous inequality, we obtain  $\langle x - x', y - y' \rangle \geq 0$ . This implies that  $\partial (f - r\phi)$  is monotone, and convexity of  $f - r\phi$  then follows from [41, Theorem 12.17].

We now prove the final claim. If f is  $B_{\phi}$ -smooth, since both f and -f are proper, f must be finite-valued. It then follows from [1, Proposition 2.5] that f is continuously differentiable. Since, by definition, f and -f are both  $B_{\phi}$ -weakly convex, in light of the shown equivalence among assertions 3.2(i) and 3.2(ii), (3.1) holds for both f and -f with r = -1. By combining the two resulting inequalities and using the fact that  $\partial f = \nabla f$ , (3.2) is obtained. Conversely, if (3.2) holds and f is continuously differentiable, then both f and -f satisfy (3.1) with r = -1. By invoking again the shown equivalence of the first two assertions, we conclude that both f and -f are (proper and)  $B_{\phi}$ -weakly convex, and hence  $B_{\phi}$ -smooth by definition.

Note that, in contrast to [16, Lemma 2.1], we do not assume smoothness in Definition 3.1(iii), as this is implied automatically; see also [1, Propostion 2.5]. Under the restriction that f is finite-valued and locally Lipschitz, the equivalence between Propositions 3.2(i) and 3.2(ii) in the  $B_{\phi}$ -weakly convex case, i.e., for r = -1, was provided in [22, Lemma 2.2] for more general Legendre functions  $\phi$  that possibly do not have full domain.

In addition, we provide an alternative definition of  $B_{\phi}$ -weak/strong convexity and  $B_{\phi}$ -smoothness involving an *envelope* representation that identifies a function as a hull of supporting elementary functions. This observation naturally leads to framing these notions within the scope of  $\Phi$ -convexity (cf. Definition 2.1), and in doing so, it enables the possibility to employ its general yet very powerful tools. In this perspective, note that with  $\Phi = -D_{\phi}$ , left and right  $\Phi$ -conjugates amount to left and right (negative) Bregman–Moreau envelopes  $f^{\Phi} = -\overleftarrow{\operatorname{env}}^{\phi} f$  and  $g^{\Phi} = -\overrightarrow{\operatorname{env}}^{\phi} g$ , where

3.3) 
$$\operatorname{env}^{\phi} f(y) := \inf \left\{ f + \mathcal{D}_{\phi}(\cdot, y) \right\} \quad \text{and} \quad \operatorname{env}^{\phi} g(x) := \inf \left\{ g + \mathcal{D}_{\phi}(x, \cdot) \right\}.$$

Copyright (c) by SIAM. Unauthorized reproduction of this article is prohibited.

Likewise, the attainment sets, which constitute the respective left and right Bregman proximal mappings, are the inverse  $\Phi$ -subdifferentials. By changing the sign of the coupling function, negative Bregman–Moreau envelopes are replaced by their Bregman–Klee counterparts: For  $\Phi = D_{\phi}$ , one has that  $f^{\Phi} = k e^{\phi} f$  and  $g^{\Phi} = k e^{\phi} g$ , where

(3.4) 
$$\overleftarrow{\operatorname{klee}}^{\phi} f(y) := \sup \{ \mathcal{D}_{\phi}(\cdot, y) - f \} \text{ and } \overrightarrow{\operatorname{klee}}^{\phi} g(x) := \sup \{ \mathcal{D}_{\phi}(x, \cdot) - g \},$$

and once again the attainment sets coincide with the inverse  $\Phi$ -subdifferentials.

First, we provide some useful formulas that express these Bregman envelopes in terms of convex conjugates. To this end, we introduce the following convenient notation for the reflection:

$$h_{-}(x) := h(-x).$$

We point out that reflection and convex conjugation commute, namely,

$$(3.5) (h_{-})^* = (h^*)_{-}$$

holds for any  $h: \mathbb{R}^n \to \overline{\mathbb{R}}$ ; see, e.g., [19, Lemma 7.1] or [8, Proposition 13.23(v)]. This fact allows us to unambiguously omit the brackets in favor of a lighter notation  $h_{-}^*$ .

LEMMA 3.3. For any  $f, g: \mathbb{R}^n \to \overline{\mathbb{R}}$ , the following identities hold:

- (i)  $\overrightarrow{env}^{\phi}g = \phi (g \circ \nabla \phi^* + \phi^*)^*.$ (ii)  $\overleftarrow{env}^{\phi}f \circ \nabla \phi^* = \phi^* - (f + \phi)^*.$ (iii)  $-\overrightarrow{env}^{\phi}(-\overleftarrow{env}^{\phi}f) = (f + \phi)^{**} - \phi.$
- (iv)  $\overrightarrow{\mathrm{klee}}^{\phi}g = \phi + (g \circ \nabla \phi^* \phi^*)^*_{-}.$
- (v) klee<sup> $\phi$ </sup>  $f \circ \nabla \phi^* = \phi^* + (f \phi)^*_-$ .

(vi) 
$$\overrightarrow{\text{klee}}^{\phi}(\overleftarrow{\text{klee}}^{\phi}f) = \phi + (f - \phi)^{**}.$$

/

*Proof.* The first three identities follow by [10, Proposition 2.4(ii)], [27, Theorem 2.4], and [45, Proposition 2.14], respectively. Thus, it suffices to show the identities involving the Klee envelopes. Thanks to the fact that  $\nabla \phi(\nabla \phi^*(v)) = v$  and  $\langle v, \nabla \phi^*(v) \rangle = \phi^*(v) + \phi(\nabla \phi^*(v))$ , we have that

$$(3.6) \qquad \mathcal{D}_{\phi}(\ \cdot \ ,\nabla\phi^{*}(v)) = \phi - \phi(\nabla\phi^{*}(v)) - \langle v, \ \cdot \ -\nabla\phi^{*}(v)\rangle = \phi + \phi^{*}(v) - \langle v, \ \cdot \ \rangle$$

holds for every  $v \in \mathbb{R}^n$ . By using the fact that  $\nabla \phi^* : \mathbb{R}^n \to \mathbb{R}^n$  is surjective, for  $x \in \mathbb{R}^n$ , we have

$$\overrightarrow{\operatorname{klee}}^{\phi}g(x) = \sup_{v \in \mathbb{R}^n} \left\{ \mathcal{D}_{\phi}(x, \nabla \phi^*(v)) - g(\nabla \phi^*(v)) \right\}$$

$$\stackrel{(3.6)}{=} \sup_{v \in \mathbb{R}^n} \left\{ \phi(x) + \langle -x, v \rangle - \left( g(\nabla \phi^*(v)) - \phi^*(v) \right) \right\},$$

and assertion 3.3(iv) follows from the definition of the convex conjugate. Similarly, for any  $v \in \mathbb{R}^n$ ,

$$\begin{aligned} (\hat{\mathbf{k}} \mathrm{lee}^{\phi} f \circ \nabla \phi^*)(v) &= \sup_{x \in \mathbb{R}^n} \left\{ \mathbf{D}_{\phi}(x, \nabla \phi^*(v)) - f(x) \right\} \\ &\stackrel{(3.6)}{=} \sup_{x \in \mathbb{R}^n} \left\{ \phi^*(v) + \langle x, -v \rangle - \left( f(x) - \phi(x) \right) \right\} \end{aligned}$$

Copyright (c) by SIAM. Unauthorized reproduction of this article is prohibited.

which yields assertion 3.3(v). As a consequence of 3.3(v), we have  $(f - \phi)^*_{-} = \overleftarrow{\mathsf{klee}}^{\phi} f \circ \nabla \phi^* - \phi^*$ . Taking conjugates on both sides yields  $((f - \phi)^*_{-})^* = (\overleftarrow{\mathsf{klee}}^{\phi} f \circ \nabla \phi^* - \phi^*)^*$ . Thus, we have the following equalities:

$$((f-\phi)_{-}^{*})_{-}^{*} + \phi = (\overleftarrow{\mathsf{klee}}^{\phi} f \circ \nabla \phi^{*} - \phi^{*})_{-}^{*} + \phi = \overrightarrow{\mathsf{klee}}^{\phi} (\overleftarrow{\mathsf{klee}}^{\phi} f),$$

where the last equality follows from 3.3(iv). In view of (3.5), we have  $((f - \phi)^*_{-})^*_{-} = (f - \phi)^{**}$ , and 3.3(vi) is obtained.

Acquainted with these preliminaries, we can detail the anticipated equivalent formulation of Definition 3.1 in terms of  $\Phi$ -convexity.

PROPOSITION 3.4 ( $\Phi$ - and hull characterization of  $B_{\phi}$ -convexity). Let r = -1 (resp., r = +1), and let  $\Phi = r D_{\phi}$ . For any proper  $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ , the following are equivalent:

- (i) f is l.s.c. and  $B_{\phi}$ -weakly convex (resp.,  $B_{\phi}$ -strongly convex).
- (ii) f is  $\Phi$ -convex.
- (iii)  $f = -\overrightarrow{env}^{\phi}(-\overleftarrow{env}^{\phi}f)$  (resp.,  $f = \overrightarrow{klee}^{\phi}(\overleftarrow{klee}^{\phi}f)$ ).
- (iv) There exists a function  $q: \mathbb{R}^n \to \overline{\mathbb{R}}$  such that  $f = -\overrightarrow{env}^{\phi}q$  (resp.,  $f = \overrightarrow{klee}^{\phi}q$ ).
- (v) f is the pointwise supremum over all functions of the form  $r\phi + \langle v, \cdot \rangle + \delta$ , with  $(v, \delta) \in \mathbb{R}^n \times \overline{\mathbb{R}}$ , that are majorized by f.
- (vi) f is l.s.c. and  $\partial_{\Phi} f = \nabla \phi^* \circ (\nabla \phi r \partial f)$ .

Proof.

2730

 $3.4(iv) \Rightarrow 3.4(i)$  By assumption, there is a function  $g : \mathbb{R}^n \to \mathbb{R}$  such that  $f = -\overrightarrow{env}^{\phi}g$  (resp.,  $f = k\overrightarrow{lee}^{\phi}g$ ). It follows from Lemma 3.3(i) (resp., Lemma 3.3(iv)) that  $f = -\overrightarrow{env}^{\phi}g = (g \circ \nabla \phi^* + \phi^*)^* - \phi$  (resp.,  $f = k\overrightarrow{lee}^{\phi}g = \phi + (g \circ \nabla \phi^* - \phi^*)^*_-$ ). Since convex conjugates, as pointwise suprema over affine functions, are convex l.s.c., and since  $\phi$  is smooth, f is l.s.c. and  $B_{\phi}$ -weakly convex (resp.,  $B_{\phi}$ -strongly convex).

 $3.4(i) \Rightarrow 3.4(iii)$  If  $f - r\phi$  is proper l.s.c. and convex, then  $(f - r\phi)^{**} = f - r\phi$ , and the claim follows from Lemma 3.3(iii) (resp., Lemma 3.3(vi)).

 $3.4(\text{iii}) \Rightarrow 3.4(\text{iv})$  Consider  $g = -\overleftarrow{\text{env}}^{\phi} f$  (resp.,  $g = \overleftarrow{\text{klee}}^{\phi} f$ ).

3.4(ii)  $\Leftrightarrow$  3.4(iii) This follows from Lemma 2.5 via the choice  $\Phi := r D_{\phi}$ , noting that  $f^{\Phi\Phi} \stackrel{(def)}{=} -\overrightarrow{env}^{\phi}(-\overleftarrow{env}^{\phi}f)$  (resp.,  $f^{\Phi\Phi} \stackrel{(def)}{=} \overrightarrow{klee}^{\phi}(\overleftarrow{klee}^{\phi}f)$ ).

 $3.4(\text{ii}) \Leftrightarrow 3.4(\text{v})$  In light of Lemma 2.5,  $\Phi$ -convexity is equivalent to the identity  $f = f^{\Phi\Phi}$ . By using the surjectivity of the map  $v \mapsto \nabla \phi^*(-rv)$ , we have (recall that  $\Phi = r D_{\phi}$ ), invoking Lemma 2.4(ii),

$$(3.7) \qquad f^{\Phi\Phi} = \sup_{(v,\beta)\in\mathbb{R}^n\times\overline{\mathbb{R}}} \left\{ r \operatorname{D}_{\phi}(\ \cdot\ ,\nabla\phi^*(-rv)) - \beta : r \operatorname{D}_{\phi}(\ \cdot\ ,\nabla\phi^*(-rv)) - \beta \le f \right\}.$$

Since  $r D_{\phi}(\cdot, \nabla \phi^*(-rv)) = r\phi + r\phi^*(-rv) + \langle v, \cdot \rangle$  by (3.6), the change of variable  $\delta = r\phi^*(-rv) - \beta$  in (3.7) results in the supremum over the functions  $r\phi + \langle v, \cdot \rangle + \delta$  majorized by f, as in assertion 3.4(v).

 $3.4(i) \Rightarrow 3.4(vi)$  Let f be l.s.c. and  $B_{\phi}$ -weakly convex, and for  $\bar{x} \in \text{dom}(\nabla \phi^* \circ (\nabla \phi - r\partial f))$ , let  $\bar{y} \in \nabla \phi^*(\nabla \phi(\bar{x}) - r\partial f(\bar{x}))$ , so that there exists  $\bar{v} \in \partial f(\bar{x})$  such that

$$\bar{y} = \nabla \phi^* (\nabla \phi(\bar{x}) - r\bar{v}).$$

From Proposition 3.2, we obtain (recall that  $\Phi := r D_{\phi}$ )

$$f(x) \ge f(\bar{x}) + \langle \bar{v}, x - \bar{x} \rangle + r \operatorname{D}_{\phi}(x, \bar{x})$$
  
=  $f(\bar{x}) + r\phi(x) - r\phi(\bar{x}) - r \langle \nabla \phi(\bar{x}) - r\bar{v}, x - \bar{x} \rangle$ 

$$\stackrel{(3.6)}{=} f(\bar{x}) + r \operatorname{D}_{\phi}(x, \nabla \phi^* (\nabla \phi(\bar{x}) - r\bar{v})) - r \operatorname{D}_{\phi}(\bar{x}, \nabla \phi^* (\nabla \phi(\bar{x}) - r\bar{v}))$$
  
=  $f(\bar{x}) + \Phi(x, \bar{y}) - \Phi(\bar{x}, \bar{y}) \quad \forall x \in \mathbb{R}^n,$ 

and thus  $\bar{y} \in \partial_{\Phi} f(\bar{x})$  by definition. Now choose  $\bar{y} \in \partial_{\Phi} f(\bar{x})$ . In view of Lemma 2.6, this means that  $\bar{x} \in \arg\min_{x \in \mathbb{R}^n} \{f(x) - r \operatorname{D}_{\phi}(x, \bar{y})\}$ , and therefore  $0 \in \partial f(\bar{x}) - r \nabla \phi(\bar{x}) + r \nabla \phi(\bar{y})$ . Reordering yields  $\bar{y} \in \nabla \phi^* (\nabla \phi(\bar{x}) - r \partial f(\bar{x}))$ .

 $3.4(\text{vi}) \Rightarrow 3.4(\text{i})$  Conversely, let f be l.s.c., and assume that  $\partial_{\Phi} f = \nabla \phi^* \circ (\nabla \phi - r\partial f)$ . Let  $\bar{x} \in \text{dom} \partial f$ , and for  $\bar{v} \in \partial f(\bar{x})$  let  $\bar{y} := \nabla \phi^* (\nabla \phi(\bar{x}) - r\bar{v})$ . Thus, we have  $\bar{y} \in \partial_{\Phi} f(\bar{x})$ , which means by definition (again recall that  $\Phi := r D_{\phi}$ ) that

$$f(x) \ge f(\bar{x}) + r \operatorname{D}_{\phi}(x, \bar{y}) - r \operatorname{D}_{\phi}(\bar{x}, \bar{y}) \quad \forall x \in \mathbb{R}^n.$$

Applying the calculation above using (3.6) we obtain (3.1), and thus, in view of Proposition 3.2, f is  $B_{\phi}$ -convex.

The result above can be seen as a generalization of [41, Example 11.26(d)] for the Euclidean proximal hull. We remark that the equivalence among Propositions 3.4(i), 3.4(iii), and 3.4(v) was previously shown in [45, Proposition 2.14] for the weakly convex case (i.e., r = -1) for more general reference functions  $\phi$  that need not have full domain.

In light of Proposition 3.4, we have the following result.

COROLLARY 3.5 ( $B_{\phi}$ -smooth functions as pointwise infima). For any  $f \in \Gamma_0(\mathbb{R}^n)$ , the following are equivalent:

- (i) f is  $B_{\phi}$ -smooth.
- (ii)  $f = \overrightarrow{env}^{\phi}g$  for some function  $g: \mathbb{R}^n \to \overline{\mathbb{R}}$ .

Proof.

 $3.5(i) \Rightarrow 3.5(ii)$  Let  $f \in \Gamma_0(\mathbb{R}^n)$  be  $B_{\phi}$ -smooth. This means that both -f and f are  $B_{\phi}$ -weakly convex and proper. This implies that f is finite-valued and, by convexity, continuous, implying that -f is l.s.c. In light of Proposition 3.4, this means that there is some  $g: \mathbb{R}^n \to \overline{\mathbb{R}}$  such that  $-f = -\overline{\operatorname{env}}^{\phi}g$ , implying that  $f = \overline{\operatorname{env}}^{\phi}g$ .

3.5(ii)  $\Rightarrow$  3.5(i) Let  $f \in \Gamma_0(\mathbb{R}^n)$  be such that  $f = \overrightarrow{env}^{\phi}g$  for some function  $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ . By properness of f, we have that  $f > -\infty$ , and there is some  $x \in \mathbb{R}^n$  such that  $\infty > f(x) \stackrel{(def)}{=} \inf_{y \in \mathbb{R}^n} \{g(y) + \mathcal{D}_{\phi}(x, y)\}$ , implying that  $\infty > g(y)$  for some  $y \in \mathbb{R}^n$  and thus  $\infty > g(y) + \mathcal{D}_{\phi}(\cdot, y) \ge f$ . This implies that f is finite-valued and thus  $-f = -\overrightarrow{env}^{\phi}g$  is proper. Thanks to Proposition 3.4, we have that -f is  $\mathcal{B}_{\phi}$ -weakly convex and, due to convexity and properness of f, that f is  $\mathcal{B}_{\phi}$ -smooth.

**3.3.** Anisotropic convexity and smoothness. The motivation of this work stems from the observation that a conjugate duality between  $B_{\phi}$ -smoothness and  $B_{\phi^*}$ -strong convexity in general does not hold, as the following counterexample reveals.

Example 3.6. Let  $\phi(x) = \frac{1}{3}|x|^3$ . Choose  $f(x) = \phi(x) + x$ . Clearly, f is both  $B_{\phi}$ -strongly convex and  $B_{\phi}$ -smooth. The convex conjugate  $f^*$  amounts to  $f^*(x) = \phi^*(x-1)$ . A simple calculation shows that the second-order derivative  $(f^* - \phi^*)''(x) < 0$  is not bounded from below for x near 0 and that  $(\phi^* - f^*)''(x) < 0$  is not bounded from below for x near 1. As a consequence,  $f^*$  is neither  $B_{\phi^*}$ -smooth nor  $B_{\phi^*}$ -strongly convex.

Instead, we shall introduce different notions of relative smoothness and strong convexity which we call *anisotropic smoothness* and *anisotropic strong convexity*, respectively. The characterization of  $\Phi$ -convexity in Proposition 3.4(v) and Corollary 3.5 related the notions of  $B_{\phi}$ -strong convexity and  $B_{\phi}$ -smoothness to expressions involving pointwise suprema and infima over a family of *tilt*-parametrized functions  $q_i := \phi + \langle y_i, \cdot \rangle + \beta_i$  for some  $(y_i, \beta_i) \in \mathbb{R}^n \times \overline{\mathbb{R}}$  and with *i* ranging over some index set  $\mathcal{I}$ . The calculus of [41, Theorem 11.23(d)] reveals that convex conjugates of such functions are to be found in the space of infima and suprema over *shift*-parametrized families  $q_i^* = \phi^*(\cdot - y_i) - \beta_i$ . This observation motivates the study of  $\Phi$ -convexity with a coupling  $\Phi(x, y) = r\phi(x - y)$  for  $r \in \{\pm 1\}$ .

Remark 3.7 ( $\Phi$ -convexity and infimal/supremal convolutions). As shown in Lemma 2.5, for any coupling  $\Phi$  it holds that a function is (left or right)  $\Phi$ -convex if and only if it is a (right or left)  $\Phi$ -conjugate. When  $\Phi(x, y) = -\phi(x - y)$ , the right conjugate of a function  $g: \mathbb{R}^n \to \mathbb{R}$  amounts to the negative *infimal convolution* of g and  $\phi$ ,

$$g^{\Phi}(x) = \sup_{y \in \mathbb{R}^n} \{-\phi(x-y) - g(y)\} = -\inf_{y \in \mathbb{R}^n} \{\phi(x-y) + g(y)\}$$
$$= -\inf_{\xi \in \mathbb{R}^n} \{g(x-\xi) + \phi(\xi)\} =: -(g \Box \phi)(x),$$

and similarly the left conjugate of  $f : \mathbb{R}^n \to \overline{\mathbb{R}}$  amounts to the negative infimal convolution of f and  $\phi_-$ ,

$$\begin{split} f^{\Phi}(y) &= \sup_{x \in \mathbb{R}^n} \left\{ -\phi(x-y) - f(x) \right\} = -\inf_{x \in \mathbb{R}^n} \left\{ \phi_-(y-x) + f(x) \right\} \\ &= -\inf_{\xi \in \mathbb{R}^n} \left\{ f(y-\xi) + \phi_-(\xi) \right\} =: -(f \Box \phi_-)(y). \end{split}$$

When  $\Phi(x, y) = \phi(x - y)$ , the right conjugate of a function  $g : \mathbb{R}^n \to \overline{\mathbb{R}}$  amounts to the supremal convolution of -g and  $\phi$ ,

$$g^{\Phi}(x) = \sup_{y \in \mathbb{R}^n} \left\{ \phi(x - y) - g(y) \right\} = \sup_{\xi \in \mathbb{R}^n} \left\{ (-g)(x - \xi) + \phi(\xi) \right\} =: ((-g) \Diamond \phi)(x) = 0$$

and similarly the left conjugate of  $f : \mathbb{R}^n \to \overline{\mathbb{R}}$  amounts to the supremal convolution of -f and  $\phi_-$ ,

$$f^{\Phi}(y) = \sup_{x \in \mathbb{R}^n} \{\phi(x-y) - f(x)\} = \sup_{x \in \mathbb{R}^n} \{\phi_-(y-x) - f(x)\}$$
$$= \sup_{\xi \in \mathbb{R}^n} \{(-f)(y-\xi) + \phi_-(\xi)\} =: ((-f) \Diamond \phi_-)(y).$$

As we shall see, while the conjugate of a  $B_{\phi}$ -smooth function is a supremal convolution, the conjugate of a supremal convolution is not necessarily  $B_{\phi}$ -smooth. Instead, the anticipated duality of  $B_{\phi}$ -smoothness and  $B_{\phi}$ -strong convexity involves their *anisotropic* counterparts, defined next, which in general constitute proper subclasses of the classes of  $\Phi$ -convex functions with  $\Phi(x,y) = \pm \phi(x-y)$ , as will be demonstrated in Example 5.4.

DEFINITION 3.8 (anisotropic weak/strong convexity and smoothness). A proper l.s.c. function  $f : \mathbb{R}^n \to \overline{\mathbb{R}}$  is said to be

(i)  $a_{\phi}$ -weakly convex if for every  $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$  it holds that

(3.8) 
$$f(x) \ge f(\bar{x}) - \phi(x - \bar{x} + \nabla \phi^*(-\bar{v})) + \phi(\nabla \phi^*(-\bar{v})) \quad \forall x \in \mathbb{R}^n;$$

(ii)  $a_{\phi}$ -strongly convex if for every  $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$  it holds that

(

3.9) 
$$f(x) \ge f(\bar{x}) + \phi(x - \bar{x} + \nabla \phi^*(\bar{v})) - \phi(\nabla \phi^*(\bar{v})) \quad \forall x \in \mathbb{R}^n;$$

(iii)  $a_{\phi}$ -smooth if  $f \in \mathcal{C}^{1}(\mathbb{R}^{n})$  and both f and -f are  $a_{\phi}$ -weakly convex.

Note that  $a_{\phi}$ -weak convexity is a globalized version of anisotropic prox-regularity introduced in [28, Definition 2.13]. Geometrically, the  $a_{\phi}$ -weak convexity inequality (3.8) (resp., the  $a_{\phi}$ -strong convexity inequality (3.9)) ensures that every limiting subgradient  $\bar{v} \in \partial f(\bar{x})$  generates a lower approximation to f that supports the graph of f at  $\bar{x}$  and takes the form  $x \mapsto -\phi(x - \bar{y}) - \bar{\beta}$  for  $\bar{y} = \bar{x} - \nabla \phi^*(-\bar{v})$ and  $\bar{\beta} = -\phi(\nabla \phi^*(-\bar{v})) - f(\bar{x})$  (resp.,  $x \mapsto \phi(x - \bar{y}) - \bar{\beta}$  for  $\bar{y} = \bar{x} - \nabla \phi^*(\bar{v})$  and  $\bar{\beta} = \phi(\nabla \phi^*(\bar{v})) - f(\bar{x})$ ).

Examples will be presented in section 4 alongside the anticipated conjugate dualities. As will be shown in section 5, another source of examples in the univariate case is the class of pointwise maxima/minima where the index sets in the definition of  $\Phi$ -convexity are finite.

Remark 3.9. As in the case of  $B_{\phi}$ -convexity and specifically in the characterization of Proposition 3.2(ii), the definition of  $a_{\phi}$ -weak and  $a_{\phi}$ -strong convexity can be combined into a single inequality up to a sign difference. Namely, a proper and l.s.c. function  $f : \mathbb{R}^n \to \mathbb{R}$  is  $a_{\phi}$ -weakly convex (resp.,  $a_{\phi}$ -strongly convex) if for every  $(\bar{x}, \bar{v}) \in \text{gph} \partial f$  it holds that

$$(3.10) f(x) \ge f(\bar{x}) + r\phi(x - \bar{x} + \nabla\phi^*(r\bar{v})) - r\phi(\nabla\phi^*(r\bar{v})) \quad \forall x \in \mathbb{R}^n$$

with r = -1 (resp., r = 1). In particular, if  $\phi = \frac{1}{2\lambda} \| \cdot \|^2$  is a positive multiple of the squared Euclidean norm with  $\lambda > 0$ , by expanding the square in (3.10), we obtain

$$f(x) \ge f(\bar{x}) + \langle \bar{v}, x - \bar{x} \rangle + \frac{r}{2\lambda} \|x - \bar{x}\|^2,$$

so that the inequalities specialize to the Euclidean weak (r = -1) and strong (r = +1) convexity inequalities, respectively, and thus also coincide with the Bregman version (3.1) due to the identity  $\frac{r}{2\lambda} ||x - \bar{x}||^2 = r D_{\phi}(x, \bar{x})$ .

Remark 3.10 (anisotropic descent inequality and algorithmic implications). In the same way that Euclidean weak convexity of -f entails a quadratic upper bound in the likes of the Euclidean descent lemma (see, e.g., [14, Propostion A.24]),  $a_{\phi}$ -weak convexity of -f for  $f \in \mathcal{C}^1(\mathbb{R}^n)$  can be read as the anisotropic descent inequality

$$(3.11) f(x) \le f(\bar{x}) + \phi(x - \bar{x} + \nabla \phi^*(\nabla f(\bar{x})) - \phi(\nabla \phi^*(\nabla f(\bar{x}))) \quad \forall x \in \mathbb{R}^n,$$

holding at any  $\bar{x} \in \mathbb{R}^n$ . This bound is simplified if one considers  $\phi$  such that  $\phi(0) = 0$ and  $\nabla \phi(0) = 0$ . Given a function f satisfying (3.11) and a current iterate  $x^t \in \mathbb{R}^n$ , by minimizing the upper bound (3.11) at  $\bar{x} = x^t$ , we recast a recent anisotropic generalization of gradient descent, called *dual space preconditioning* [32],

$$\begin{split} x^{t+1} &:= \mathop{\arg\min}_{x \in \mathbb{R}^n} f(x^t) + \phi(x - x^t + \nabla \phi^*(\nabla f(x^t)) - \phi(\nabla \phi^*(\nabla f(x^t))) \\ &= x^t - \nabla \phi^*(\nabla f(x^t)), \end{split}$$

where the second equality follows from the fact that  $\nabla \phi(0) = 0$  together with the first-order optimality condition  $0 = \nabla \phi(x^{t+1} - x^t + \nabla \phi^*(\nabla f(x^t)))$ . Indeed, inserting this recurrence into (3.11) generates a sufficient descent

$$f(x^{t+1}) - f(x^t) \le \phi(0) - \phi(\nabla \phi^*(\nabla f(x^t))) = -\phi(\nabla \phi^*(\nabla f(x^t))) = -\phi(x^t - x^{t+1}),$$

where the second equality holds since  $\phi(0) = 0$ . Note that this argument differs from the one in [32], where the descent is obtained by using a different criterion (see [32, Lemma 3.8]). In contrast to [32], which focuses on the convex case only, the anisotropic descent inequality also holds for possibly nonconvex smooth functions.

Copyright (C) by SIAM. Unauthorized reproduction of this article is prohibited.

LEMMA 3.11. Let  $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ , and consider the following statements:

- (i) f is  $a_{\phi}$ -strongly convex.
- (ii) f is proper, l.s.c. and convex.
- (iii) f is  $a_{\phi}$ -weakly convex.

Then it holds that  $3.11(i) \Rightarrow 3.11(ii) \Rightarrow 3.11(iii)$ .

*Proof.* If f is  $a_{\phi}$ -strongly convex, then it is proper and l.s.c. by definition. For any  $(x, v), (x', v') \in \operatorname{gph} \partial f$ , it follows from (3.9) that

$$f(x') \ge f(x) + \phi(x' - x + \nabla \phi^*(v)) - \phi(\nabla \phi^*(v)) \ge f(x) + \langle v, x' - x \rangle,$$

where the second inequality is due to the convexity of  $\phi$  [41, Theorem 2.14] and the fact that  $\nabla \phi(\nabla \phi^*(v)) = v$ . Similarly,

$$f(x) \ge f(x') + \langle v', x - x' \rangle.$$

By summing the two inequalities, it readily follows that the mapping  $\partial f$  is monotone, and convexity of f then follows from [41, Theorem 12.17].

Suppose now that f is proper, l.s.c., and convex, and let  $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$ . Then

$$f(x) \ge f(\bar{x}) + \langle \bar{v}, x - \bar{x} \rangle \ge f(\bar{x}) - \phi(x - \bar{x} + \nabla \phi^*(-\bar{v})) + \phi(\nabla \phi^*(-\bar{v})) \quad \forall x \in \mathbb{R}^n$$

where again we used convexity of  $\phi$  and the fact that  $\nabla \phi(\nabla \phi^*(-\bar{v})) = -\bar{v}$ . This shows that f is  $a_{\phi}$ -weakly convex.

As explained above, the anisotropic subgradient inequality (3.10) ensures that every classical subgradient  $\bar{v} \in \partial f(\bar{x})$  generates a supporting lower approximation  $x \mapsto r\phi(x-\bar{y}) - \bar{\beta}$  to f. This means that  $\bar{y} = \bar{x} - \nabla \phi^*(r\bar{v}) \in \partial_{\Phi} f(\bar{x})$  is a  $\Phi$ -subgradient of f at  $\bar{x}$ . As we shall see, in this case the converse is also true; that is, every  $\Phi$ subgradient  $\bar{y} \in \partial_{\Phi} f(\bar{x})$  takes the form  $\bar{y} \in (\mathrm{id} - \nabla \phi^* \circ r \partial f)(\bar{x})$ .

PROPOSITION 3.12. Let r = -1 (resp., r = +1), and let  $\Phi(x, y) = r\phi(x - y)$ . For any proper l.s.c.  $f : \mathbb{R}^n \to \overline{\mathbb{R}}$  and  $\overline{x} \in \mathbb{R}^n$ , it holds that  $\partial_{\Phi} f(\overline{x}) \subseteq \overline{x} - \nabla \phi^*(r \partial f(\overline{x}))$ . In addition, the following conditions are equivalent:

- (i) f satisfies the a<sub>φ</sub>-weak convexity inequality (3.8) (resp., the a<sub>φ</sub>-strong convexity inequality (3.9)) at x̄ for every v̄ ∈ ∂f(x̄).
- (ii)  $\partial_{\Phi} f(\bar{x}) = \bar{x} \nabla \phi^* (r \partial f(\bar{x})).$

Under any of the above equivalent conditions, one has that  $\partial f(\bar{x}) = \widehat{\partial} f(\bar{x})$ . In particular, if f is either  $a_{\phi}$ -weakly or  $a_{\phi}$ -strongly convex, then  $\partial f = \widehat{\partial} f$  and dom  $\partial_{\Phi} f =$ dom  $\widehat{\partial} f$ .

*Proof.* Let a pair  $(\bar{x}, \bar{y}) \in \text{gph} \partial_{\Phi} f$  be fixed. In view of Lemma 2.6, we know that  $\bar{x} \in \operatorname{arg\,min} \{f - r\phi(\cdot - \bar{y})\}$ ; hence,  $0 \in \widehat{\partial} f(\bar{x}) - r\nabla\phi(\bar{x} - \bar{y})$ , and therefore  $\bar{y} \in \bar{x} - \nabla\phi^*(r\widehat{\partial} f(\bar{x})) \subseteq \bar{x} - \nabla\phi^*(r\partial f(\bar{v}))$ . In particular, if assertion 3.12(ii) holds, then

$$\bar{x} - \nabla \phi^*(r\partial f(\bar{x})) = \partial_{\Phi} f(\bar{x}) \subseteq \bar{x} - \nabla \phi^*(r\partial f(\bar{x})) \subseteq \bar{x} - \nabla \phi^*(r\partial f(\bar{x})),$$

which by the injectivity of  $\nabla \phi^*$  yields the claimed identity  $\widehat{\partial} f(\bar{x}) = \partial f(\bar{x})$ . In what follows, fix  $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$ , and let  $\bar{y} := \bar{x} - \nabla \phi^*(r\bar{v})$ .

 $3.12(i) \Rightarrow 3.12(ii)$  Substituting  $\bar{y}$  into the anisotropic subgradient inequality (3.10) gives

$$f(x) \ge r\phi(x-\bar{y}) - r\phi(\bar{x}-\bar{y}) + f(\bar{x}) \quad \forall x \in \mathbb{R}^n,$$

Copyright (c) by SIAM. Unauthorized reproduction of this article is prohibited.

that is,  $\bar{y} \in \partial_{\Phi} f(\bar{x})$ . In combination with  $\partial_{\Phi} f(\bar{x}) \subseteq \bar{x} - \nabla \phi^*(r \partial f(\bar{x}))$ , we obtain the desired result.

 $3.12(ii) \Rightarrow 3.12(i)$  By assumption,  $\bar{y} \in \partial_{\Phi} f(\bar{x})$ , and therefore

$$f(x) \ge r\phi(x - \bar{y}) - r\phi(\bar{x} - \bar{y}) + f(\bar{x}) = r\phi(x - \bar{x} + \nabla\phi^*(r\bar{v})) - r\phi(\nabla\phi^*(r\bar{v})) + f(\bar{x})$$

follows from the  $\Phi$ -subgradient inequality (2.7).

The relation  $\partial_{\Phi} f = \mathrm{id} - \nabla \phi^* \circ r \partial f$  sets up a certain one-to-one correspondence between limiting subgradients  $\bar{v} \in \partial f(\bar{x})$  and  $\Phi$ -subgradients  $\bar{y} = \bar{x} - \nabla \phi^*(r\bar{v})$ , and thus the anisotropic subgradient inequality can be understood in terms of a  $\Phi$ -subgradient inequality. The next result shows that for  $\Phi$ -convex f, the  $\Phi$ -subdifferential is nonempty and compact at points at which f is finite and strictly continuous.

PROPOSITION 3.13. Let  $r \in \{\pm 1\}$ , and choose  $\Phi(x, y) = r\phi(x - y)$ . Let  $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be  $\Phi$ -convex and strictly continuous at  $\overline{x} \in \mathbb{R}^n$ , a point where f is finite. Then  $\partial_{\Phi} f(\overline{x})$ is nonempty and compact.

*Proof.* Since f is  $\Phi$ -convex, the identity  $f(\bar{x}) = \sup_{y \in \mathbb{R}^n} \left\{ \Phi(\bar{x}, y) - f^{\Phi}(y) \right\}$  follows from Lemma 2.5. This means that for any  $\varepsilon > 0$  there exists  $\bar{y}_{\varepsilon}$  such that

$$(3.12) \qquad \qquad -\infty < f(\bar{x}) - \varepsilon \le \Phi(\bar{x}, \bar{y}_{\varepsilon}) - f^{\Phi}(\bar{y}_{\varepsilon}) \le f(\bar{x}) < +\infty.$$

In addition, in view of Lemma 2.4(iv), for any  $x \in \mathbb{R}^n$  we have that

(3.13) 
$$f(x) \ge \Phi(x, \bar{y}_{\varepsilon}) - f^{\Phi}(\bar{y}_{\varepsilon}).$$

Summing the inequalities (3.12) and (3.13) yields

(3.14) 
$$f(\bar{x}) - r\phi(\bar{x} - \bar{y}_{\varepsilon}) \le f(x) - r\phi(x - \bar{y}_{\varepsilon}) + \varepsilon \quad \forall x \in \mathbb{R}^n.$$

Then  $\bar{x} \in \varepsilon$ -arg min $_{x \in \mathbb{R}^n} \{f(x) - r\phi(x - \bar{y}_{\varepsilon})\}$ . Ekeland's variational principle with  $\delta := \sqrt{\varepsilon}$  (see [41, Proposition 1.43]) yields the existence of  $\bar{x}_{\varepsilon} \in \overline{B}(\bar{x}; \sqrt{\varepsilon})$  with  $f(\bar{x}_{\varepsilon}) - r\phi(\bar{x}_{\varepsilon} - \bar{y}_{\varepsilon}) \leq f(\bar{x}) - r\phi(\bar{x} - \bar{y}_{\varepsilon})$  and  $\bar{x}_{\varepsilon} = \arg\min\{f - r\phi(\cdot - \bar{y}_{\varepsilon}) + \sqrt{\varepsilon}\|\cdot - \bar{x}_{\varepsilon}\|\}$ . Minimality of  $\bar{x}_{\varepsilon}$  implies that  $0 \in \partial f(\bar{x}_{\varepsilon}) - r\nabla\phi(\bar{x}_{\varepsilon} - \bar{y}_{\varepsilon}) + \overline{B}(0; \sqrt{\varepsilon})$ , and therefore

(3.15) 
$$r\nabla\phi(\bar{x}_{\varepsilon}-\bar{y}_{\varepsilon})-u_{\varepsilon}=:\bar{v}_{\varepsilon}\in\partial f(\bar{x}_{\varepsilon})$$

holds for some  $u_{\varepsilon}$  with  $||u_{\varepsilon}|| \leq \sqrt{\varepsilon}$ . Since f, as a pointwise supremum over continuous functions, is l.s.c. and by assumption strictly continuous at  $\bar{x}$ , in view of [41, Theorem 9.13],  $x \mapsto \partial f(x)$  is locally bounded at  $\bar{x}$ . Without loss of generality, up to extracting a subnet,  $\bar{v}_{\varepsilon} \to \bar{v}$  for some  $\bar{v} \in \mathbb{R}^n$  as  $\varepsilon \searrow 0$ . It follows from (3.15) that  $\bar{y}_{\varepsilon} = \bar{x}_{\varepsilon} \nabla \phi^*(r\bar{v}_{\varepsilon} + ru_{\varepsilon})$ , and due to continuity of  $\nabla \phi^*$ , we obtain that  $\bar{y}_{\varepsilon} \to \bar{y} := \bar{x} - \nabla \phi^*(r\bar{v})$ . Passing to the limit in (3.14) yields

$$f(\bar{x}) + r\phi(x - \bar{y}) - r\phi(\bar{x} - \bar{y}) \le f(x) \quad \forall x \in \mathbb{R}^n$$

and therefore  $\bar{y} \in \partial_{\Phi} f(\bar{x})$ .

In view of Lemma 2.6, since f is  $\Phi$ -convex, we have  $\bar{y} \in \partial_{\Phi} f(\bar{x}) \Leftrightarrow \bar{x} \in \partial_{\Phi} f^{\Phi}(\bar{y}) \Leftrightarrow \bar{y} \in \arg\min\{f^{\Phi} - r\phi(\bar{x} - \cdot)\}$  and thus  $\partial_{\Phi} f(\bar{x}) = \arg\min\{f^{\Phi} - r\phi(\bar{x} - \cdot)\}$ . Since  $f^{\Phi}$  as a pointwise supremum over continuous functions is l.s.c.,  $\partial_{\Phi} f(\bar{x})$  is closed. By Proposition 3.12,  $\partial_{\Phi} f(\bar{x}) \subseteq \bar{x} - \nabla \phi^*(r\partial f(\bar{x}))$ . Since f is strictly continuous,  $\partial f(\bar{x})$  is bounded, and thus so is  $\nabla \phi^*(r\partial f(\bar{x}))$  due to continuity of  $\nabla \phi^*$ . As a result,  $\partial_{\Phi} f(\bar{x})$  too is bounded.

### 4. Anisotropic and Bregman conjugate dualities.

**4.1.**  $\mathbf{B}_{\phi}$ -smooth and  $\mathbf{a}_{\phi^*}$ -strongly convex conjugate duality. In this subsection, we will study the conjugate duality between  $\mathbf{a}_{\phi}$ -strong convexity as in Definition 3.8(ii) and  $\mathbf{B}_{\phi^*}$ -smoothness for a convex function  $f \in \Gamma_0(\mathbb{R}^n)$ ; the main result is Theorem 4.3. To begin with, we show that essential strict convexity is implied by  $\mathbf{a}_{\phi}$ -strong convexity.

PROPOSITION 4.1. Let  $f : \mathbb{R}^n \to \overline{\mathbb{R}}$  be  $a_{\phi}$ -strongly convex. Then f is essentially strictly convex.

*Proof.* Invoking Lemma 3.11, we know that f is convex, proper l.s.c. and in particular has a convex domain. Fix a nonempty convex set  $K \subseteq \text{dom} \partial f$ . To avoid trivialities, we may assume that K is not a singleton. Let  $x, x' \in K$  with  $x \neq x'$  be fixed. Choose  $\tau \in (0,1)$ . Since K is convex,  $x_{\tau} := \tau x + (1-\tau)x' \in K$ . By  $a_{\phi}$ -strong convexity of f and Proposition 3.12,

$$(4.1) \qquad \qquad \operatorname{dom} \partial_{\Phi} f = \operatorname{dom} \partial f \supseteq K,$$

and thus there exists  $y_{\tau} \in \partial_{\Phi} f(x_{\tau})$ . Lemma 2.6 implies that  $f(x_{\tau}) + f^{\Phi}(y_{\tau}) = \phi(x_{\tau} - y_{\tau})$ . Since  $\phi$  is strictly convex and thus so is  $\phi(\cdot - y_{\tau}) - f^{\Phi}(y_{\tau})$ , we have

$$\begin{aligned} f(\tau x + (1 - \tau)x') &= \phi(\tau x + (1 - \tau)x' - y_{\tau}) - f^{\Phi}(y_{\tau}) \\ &< \tau(\phi(x - y_{\tau}) - f^{\Phi}(y_{\tau})) + (1 - \tau)(\phi(x' - y_{\tau}) - f^{\Phi}(y_{\tau})) \\ &\leq \tau \sup_{y \in \mathbb{R}^{n}} \left\{ \phi(x - y) - f^{\Phi}(y) \right\} + (1 - \tau) \sup_{y \in \mathbb{R}^{n}} \left\{ \phi(x' - y) - f^{\Phi}(y) \right\} \\ &= \tau f(x) + (1 - \tau) f(x'), \end{aligned}$$

where the last equality follows by definition of  $f^{\Phi\Phi}$  and the fact that  $f \equiv f^{\Phi\Phi}$  on  $\operatorname{dom} \partial_{\Phi} f \supseteq K$ ; see Lemma 2.6. This proves that f is essentially strictly convex.

We state a generalization of Pshenichnyi's and Hiriart–Urruty's formula [38, 25] specialized from [19, Theorem 7.1] to our simplified setting. We remind the reader that  $h_{-} := h(-\cdot)$  and that  $(h_{-})^* = (h^*)_{-} =: h^*_{-}$ . In addition, we introduce upper and lower subtraction, i.e.,  $\infty - \infty := \infty$  and  $\infty - \infty := -\infty$ ; see also [19, section 2].

LEMMA 4.2. Let  $g: \mathbb{R}^n \to \overline{\mathbb{R}}$  and  $h \in \Gamma_0(\mathbb{R}^n)$ . Then the following identity holds:

(4.2) 
$$\sup_{y \in \mathbb{R}^n} \{h(x-y) - g_-(y)\} = (h^* - g^*)^*(x).$$

Next we show the main result of this subsection: the anticipated duality between the classes of  $a_{\phi}$ -strongly convex and  $B_{\phi^*}$ -smooth functions. The inclusion of the former in the superclass of  $\Phi$ -convex functions for  $\Phi(x, y) = \phi(x - y)$  will also be demonstrated. It will become apparent that the culprit of the properness of such inclusion is to be found in a certain saddle-point property. Further investigation on this property will be developed in subsection 5.2, where special cases guaranteeing its validity will be presented and its sharpness in more general contexts ultimately showcased with Example 5.4.

THEOREM 4.3 (a<sub> $\phi$ </sub>-strong and B<sub> $\phi^*$ </sub>-smooth duality). Let  $\Phi(x,y) := \phi(x-y)$ . For any proper  $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ , the following conditions are equivalent:

- (i) f is  $a_{\phi}$ -strongly convex.
- (ii) f is l.s.c. and convex, and  $f^*$  is  $B_{\phi^*}$ -smooth (i.e.,  $\phi^* f^* \in \Gamma_0(\mathbb{R}^n)$ ).

- (iii) f is  $\Phi$ -convex (equivalently,  $f = (-h) \Diamond \phi$  for some  $h \in \Gamma_0(\mathbb{R}^n)$ ), and for all  $\bar{v} \in \mathbb{R}^n$ , the following saddle-point property holds:
  - (4.3)  $\sup_{x \in \mathbb{R}^n} \inf_{y \in \mathbb{R}^n} \left\{ \langle x, \bar{v} \rangle - \Phi(x, y) + f^{\Phi}(y) \right\} = \inf_{y \in \mathbb{R}^n} \sup_{x \in \mathbb{R}^n} \left\{ \langle x, \bar{v} \rangle - \Phi(x, y) + f^{\Phi}(y) \right\}.$

If, in addition, f is essentially smooth, the saddle-point property in 4.3(iii) is super-fluous.

Proof.

 $4.3(i) \Rightarrow 4.3(ii)$  Let f be  $a_{\phi}$ -strongly convex. In view of Lemma 3.11,  $f \in \Gamma_0(\mathbb{R}^n)$ . Fix  $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$ , and define  $\bar{y} := \bar{x} - \nabla \phi^*(\bar{v}) \in \partial_{\Phi} f(\bar{x})$ , where the inclusion holds by Proposition 3.12. Invoking Lemma 2.6, we have  $f(\bar{x}) + f^{\Phi}(\bar{y}) = \phi(\bar{x} - \bar{y})$  and  $\bar{x} \in \partial_{\Phi} f^{\Phi}(\bar{y})$ , where the latter means by definition that  $\bar{y} \in \operatorname{arg\,max} \phi(\bar{x} - \cdot) - f^{\Phi}$ . Combined, these yield

(4.4) 
$$\sup_{y \in \mathbb{R}^n} \left\{ \phi(\bar{x} - y) - f^{\Phi}(y) \right\} = \phi(\bar{x} - \bar{y}) - f^{\Phi}(\bar{y}) = f(\bar{x}).$$

We have

$$\phi(\bar{x}-y) = \phi^{**}(\bar{x}-y) = \sup_{v \in \mathbb{R}^n} \left\{ \langle \bar{x}-y, v \rangle - \phi^*(v) \right\}.$$

Define  $q(y) := \sup_{v \in \mathbb{R}^n} h(y, v)$  for

$$h(y,v) := \langle \bar{x} - y, v \rangle - \phi^*(v) - f^{\Phi}(y).$$

Then we can rewrite the supremum in (4.4) in terms of the joint supremum

(4.5) 
$$f(\bar{x}) = \sup_{y \in \mathbb{R}^n} q(y),$$

where  $\bar{y} \in \arg \max q$ . Since  $\bar{y} = \bar{x} - \nabla \phi^*(\bar{v})$ , we have that  $\bar{v} = \nabla \phi(\bar{x} - \bar{y})$ . Using the fact that  $\bar{v} \in \partial \phi(\bar{x} - \bar{y}) \Leftrightarrow \bar{x} - \bar{y} \in \partial \phi^*(\bar{v})$ , this means that  $\bar{v} \in \arg \max \langle \bar{x} - \bar{y}, \cdot \rangle - \phi^*$ , showing that  $\bar{v} \in \arg \max h(\bar{y}, \cdot)$ . Let  $p(v) := \sup_{y \in \mathbb{R}^n} h(y, v)$ . Then [41, Proposition 1.35] yields that  $(\bar{y}, \bar{v}) \in \arg \max h$  as well as

$$(4.6) \qquad \qquad \bar{v} \in \arg\max p.$$

Overall, this yields

$$f(\bar{x}) \stackrel{(4.5)}{=} \sup_{y \in \mathbb{R}^n} q(y) = \sup_{v \in \mathbb{R}^n} p(v) \stackrel{(4.6)}{=} p(\bar{v})$$

$$\stackrel{(def)}{=} \langle \bar{x}, \bar{v} \rangle - \phi^*(\bar{v}) + \sup_{y \in \mathbb{R}^n} \left\{ \langle y, -\bar{v} \rangle - f^{\Phi}(y) \right\} \stackrel{(def)}{=} \langle \bar{x}, \bar{v} \rangle - \phi^*(\bar{v}) + (f^{\Phi})^*(-\bar{v})$$

which, combined with the fact that  $\bar{v} \in \partial f(\bar{x})$ , results in

$$f^*(\bar{v}) = \langle \bar{x}, \bar{v} \rangle - f(\bar{x}) = \phi^*(\bar{v}) - (f^{\Phi})^*_{-}(\bar{v}) \in \mathbb{R}.$$

Since  $f \ge \phi(\cdot - \bar{y}) - f^{\Phi}(\bar{y})$  holds by Lemma 2.4(iv) and since  $\phi$  is supercoercive, f too is supercoercive, and as such, dom  $f^* = \mathbb{R}^n = \operatorname{dom} \partial f^*$  in view of [6, Proposition 2.16]. Since  $\phi^*(\bar{v}) - f^*(\bar{v}) = (f^{\Phi})^*_{-}(\bar{v}) \in \mathbb{R}$  holds for all  $\bar{v} \in \operatorname{rge} \partial f = \operatorname{dom} \partial f^* = \mathbb{R}^n$ ,

Copyright (C) by SIAM. Unauthorized reproduction of this article is prohibited.

we have that  $\phi^* - f^*$  is finite-valued, convex, and l.s.c., and thus  $\phi^* - f^* \in \Gamma_0(\mathbb{R}^n)$  as claimed.

 $4.3(ii) \Rightarrow 4.3(i)$  By assumption,

(4.7) 
$$g := \phi^* - f^* \in \Gamma_0(\mathbb{R}^n).$$

It follows from Proposition 3.2 that  $f^* \in \mathcal{C}^1(\mathbb{R}^n)$ , and as such, so does g. Fix  $(\bar{x}, \bar{v}) \in$  gph  $\partial f$ , and let  $\bar{y} := \bar{x} - \nabla \phi^*(\bar{v})$ , so that  $\nabla \phi(\bar{x} - \bar{y}) = \bar{v} \in \partial f(\bar{x})$ . Since  $f \in \Gamma_0(\mathbb{R}^n)$  by assumption, by using the fact that  $\bar{v} \in \partial f(\bar{x}) \Leftrightarrow \bar{x} \in \partial f^*(\bar{v})$  and the smoothness of  $f^*$ , we obtain

(4.8) 
$$\bar{x} = \nabla f^*(\bar{v}) = \nabla f^*(\nabla \phi(\bar{x} - \bar{y})).$$

We have

(4.9)  
$$f(x) = f^{**}(x) = (\phi^* - g)^*(x) = \sup_{y \in \mathbb{R}^n} \left\{ \phi(x - y) - g^*_-(y) \right\}$$
$$\geq \phi(x - \bar{y}) - g^*_-(\bar{y}),$$

where the last equality follows from Lemma 4.2. Smoothness of g yields

$$\bar{x} \stackrel{(4.8)}{=} \nabla f^* (\nabla \phi(\bar{x} - \bar{y})) \stackrel{(4.7)}{=} \nabla (\phi^* - g) (\nabla \phi(\bar{x} - \bar{y})) = \nabla \phi^* (\nabla \phi(\bar{x} - \bar{y})) - \nabla g (\nabla \phi(\bar{x} - \bar{y})) = \bar{x} - \bar{y} - \nabla g (\nabla \phi(\bar{x} - \bar{y})),$$

and therefore

(4.10) 
$$\nabla g(\nabla \phi(\bar{x} - \bar{y})) = -\bar{y}.$$

We thus have

$$\begin{split} f(\bar{x}) &\stackrel{(4.8)}{=} \langle \nabla \phi(\bar{x} - \bar{y}), \bar{x} \rangle - f^* (\nabla \phi(\bar{x} - \bar{y})) \\ &\stackrel{(4.7)}{=} \langle \nabla \phi(\bar{x} - \bar{y}), \bar{x} \rangle - \phi^* (\nabla \phi(\bar{x} - \bar{y})) + g(\nabla \phi(\bar{x} - \bar{y})) \\ &= \langle \nabla \phi(\bar{x} - \bar{y}), \bar{x} - \bar{y} \rangle - \phi^* (\nabla \phi(\bar{x} - \bar{y})) - \left( \langle \nabla \phi(\bar{x} - \bar{y}), -\bar{y} \rangle - g(\nabla \phi(\bar{x} - \bar{y})) \right) \\ &\stackrel{(4.10)}{=} \phi(\bar{x} - \bar{y}) - g^*(-\bar{y}), \end{split}$$

where the last equality uses the fact that  $g \in \Gamma_0(\mathbb{R}^n)$ . By combining the above equality with (4.9), we obtain via the definition of  $\bar{y} = \bar{x} - \nabla \phi^*(\bar{v})$  that

$$f(x) \ge f(\bar{x}) + \phi(x - \bar{y}) - \phi(\bar{x} - \bar{y}) = f(\bar{x}) + \phi(x - \bar{x} + \nabla \phi^*(\bar{v})) - \phi(\nabla \phi^*(\bar{v})) \quad \forall x \in \mathbb{R}^n,$$

which is the  $a_{\phi}$ -strong convexity inequality (3.9).

 $4.3(\text{iii}) \Rightarrow 4.3(\text{ii})$  Since f is  $\Phi$ -convex, due to Lemma 2.5 and Remark 3.7,  $f = f^{\Phi\Phi} = (-f^{\Phi}) \Diamond \phi$ . Thus, both f and  $h := f^{\Phi}$  are pointwise suprema over convex, l.s.c. functions and thus convex l.s.c. Since f is proper,  $h(y) = \sup_{x \in \mathbb{R}^n} \{\phi(x-y) - f(x)\} > -\infty$  for all  $y \in \mathbb{R}^n$ . Suppose that  $h \equiv +\infty$ . This means that for any  $x \in \mathbb{R}^n$ , we have  $f(x) = f^{\Phi\Phi}(x) = h^{\Phi}(x) = \sup_{y \in \mathbb{R}^n} \{\phi(x-y) - h(y)\} = -\infty$ , which contradicts properness of f. Thus, h must be proper as well. For any  $\bar{v} \in \mathbb{R}^n$ , it holds that

$$\begin{split} \phi^*(\bar{v}) - (f^{\Phi})^*_{-}(\bar{v})^{(def)} &= \inf_{y \in \mathbb{R}^n} \left\{ \langle y, \bar{v} \rangle + f^{\Phi}(y) \right\} + \sup_{\xi \in \mathbb{R}^n} \left\{ \langle \xi, \bar{v} \rangle - \phi(\xi) \right\} \\ &= \inf_{y \in \mathbb{R}^n} \left\{ \sup_{\xi \in \mathbb{R}^n} \left\{ \langle \xi + y, \bar{v} \rangle - \phi(\xi) \right\} + f^{\Phi}(y) \right\} \end{split}$$

Copyright (c) by SIAM. Unauthorized reproduction of this article is prohibited.

where the last equality is due to Lemma 2.5. Therefore, and since  $h \in \Gamma_0(\mathbb{R}^n)$ , we have  $\phi^* - f^* = (f^{\Phi})^*_{-} = h^*_{-} \in \Gamma_0(\mathbb{R}^n)$ . 4.3(ii)  $\Rightarrow 4.3(iii)$  Let  $g := \phi^* - f^* \in \Gamma_0(\mathbb{R}^n)$ . We have

$$f = f^{**} = (\phi^* - g)^* = \sup_{y \in \mathbb{R}^n} \left\{ \phi(\ \cdot \ -y) - g_-^*(y) \right\} = (-g_-^*) \Diamond \phi,$$

where the third equality follows from Lemma 4.2, and thus f is  $\Phi$ -convex. By applying Lemma 4.2 again to  $g = \phi^* - f^*$ , we also have

$$g^* = (\phi^* - f^*)^* = \sup_{x \in \mathbb{R}^n} \left\{ \phi(\cdot - x) - f(-x) \right\} = \sup_{x \in \mathbb{R}^n} \left\{ \phi(x - (-\cdot)) - f(x) \stackrel{(def)}{=} (f^{\Phi})_{-} \right\}$$

Since  $g \in \Gamma_0(\mathbb{R}^n)$ , by taking conjugates on both sides, we obtain that  $g = (f^{\Phi})^*_{-}$ . In addition, it holds that

$$f^{\Phi\Phi}(\stackrel{(def)}{=} \sup_{y \in \mathbb{R}^n} \left\{ \phi(\ \cdot \ -y) - f^{\Phi}(y) \right\} = (\phi^* - (f^{\Phi})^*_-)^*,$$

where the second identity again uses Lemma 4.2. By taking conjugates on both sides, since  $\phi^* - (f^{\Phi})^*_{-} = \phi^* - g = f^*$  is convex, proper, and l.s.c., we obtain that  $(f^{\Phi\Phi})^* = \phi^* - (f^{\Phi})^*_{-}$ . As derived in (4.11), this is precisely the sought saddle-point identity (4.3).

Suppose now that f is  $\Phi$ -convex and essentially smooth. Let  $\bar{x} \in \text{dom} \partial f$ . Due to the essential smoothness of f, in view of [39, Theorem 26.1], we have  $\bar{x} \in \text{int}(\text{dom} f) = \text{dom} \partial f$  and f is differentiable at  $\bar{x}$ . In particular, it is strictly continuous at  $\bar{x}$ . Invoking Proposition 3.13, we know that

$$\emptyset \neq \partial_{\Phi} f(\bar{x}) \subseteq \bar{x} - \nabla \phi^*(\partial f(\bar{x})),$$

where the inclusion follows from Proposition 3.12. Since  $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$ , the righthand side of the inclusion is single-valued. Since the left-hand side is nonempty, the inclusion holds with equality. Proposition 3.12 then implies that the anisotropic strong convexity inequality (3.9) holds at every  $(\bar{x}, \bar{v}) \in \text{gph} \partial f$ , and thus f is  $a_{\phi}$ -strongly convex.

Remark 4.4. Let  $\Phi(x, y) = \phi(x - y)$ , and recall that the definition of the (left)  $\Phi$ conjugate  $f^{\Phi}(y) = \sup_{x \in \mathbb{R}^n} \{\phi(x - y) - f(x)\}$ . If  $\phi = \frac{1}{2} \|\cdot\|^2$  is quadratic. By expanding

(4.11)

the square, we obtain  $f^{\Phi}(y) = \frac{1}{2} ||y||^2 + \sup_{x \in \mathbb{R}^n} \left\{ \langle x, -y \rangle + \frac{1}{2} ||x||^2 - f(x) \right\} = \frac{1}{2} ||y||^2 + (f - \frac{1}{2} || \cdot ||^2)_-^*(y)$ , and hence  $f^{\Phi} - \frac{1}{2} || \cdot ||^2$  is convex. This shows that the saddle-point problem (4.3) becomes convex-concave with a bilinear coupling:

$$f^*(\bar{v}) = \sup_{x \in \mathbb{R}^n} \inf_{y \in \mathbb{R}^n} \left\{ \langle x, \bar{v} \rangle - \frac{1}{2} \|x\|^2 + \langle x, y \rangle + f^{\Phi}(y) - \frac{1}{2} \|y\|^2 \right\}.$$

As a consequence, the saddle-point property in Theorem 4.3(iii) is superfluous in the Euclidean case.

In subsection 5.2, we will illustrate the sharpness of the saddle-point property in the non-Euclidean case.

**4.2.**  $\mathbf{B}_{\phi}$ -strongly convex and  $\mathbf{a}_{\phi^*}$ -smooth conjugate duality. The main result of this subsection is Theorem 4.6, which, unlike the main result of the previous subsection, is derived via an existing conjugate duality between infimal convolution and pointwise addition [41, Theorem 11.23(a)]. We first prove the following result, which shows that at least for smooth functions,  $\Phi$ -convexity for  $\Phi(x,y) = -\phi(x-y)$  and  $\mathbf{a}_{\phi}$ -weak convexity are equivalent. As will be shown in Example 5.5, this equivalence does not hold in general.

PROPOSITION 4.5. Let  $\Phi(x, y) = -\phi(x-y)$ , and let  $f \in \mathcal{C}^1(\mathbb{R}^n)$ . Then the following conditions are equivalent:

(i) f is  $a_{\phi}$ -weakly convex.

(ii) f is  $\Phi$ -convex.

Proof.

 $4.5(i) \Rightarrow 4.5(ii)$  The  $a_{\phi}$ -weak convexity bound (3.8) in combination with Proposition 3.12 implies that  $\partial_{\Phi} f(\bar{x}) \neq \emptyset$  for all  $\bar{x} \in \mathbb{R}^n$ . It then follows from Lemma 2.6 that  $f^{\Phi\Phi} \equiv f$  on  $\mathbb{R}^n$ , and as such, f is  $\Phi$ -convex by Lemma 2.5.

 $4.5(ii) \Rightarrow 4.5(i)$  Let  $\bar{x} \in \mathbb{R}^n$ . Invoking Proposition 3.13 for r = -1, we have

$$\emptyset \neq \partial_{\Phi} f(\bar{x}) \subseteq \bar{x} - \nabla \phi^*(-\partial f(\bar{x})) = \{ \bar{x} - \nabla \phi^*(-\nabla f(\bar{x})) \},\$$

where the inclusion follows from Proposition 3.12. Since the right-hand side is singlevalued, the inclusion holds with equality. Proposition 3.12 then implies that f satisfies the  $a_{\phi}$ -weak convexity bound (3.8) at  $\bar{x}$  for all  $\bar{v} \in \partial f(\bar{x})$ .

Further restricting to convex f, we can eventually deduce the conjugate duality between  $a_{\phi}$ -smooth and  $B_{\phi^*}$ -strongly convex functions invoking an existing conjugate duality between infimal convolution and pointwise addition [41, Theorem 11.23(a)].

THEOREM 4.6. Let  $\Phi(x,y) = -\phi(x-y)$ . For any  $f \in \Gamma_0(\mathbb{R}^n)$ , the following conditions are equivalent:

(i) f is  $a_{\phi}$ -smooth.

(ii) -f is  $\Phi$ -convex.

(iii)  $f^*$  is  $B_{\phi^*}$ -strongly convex.

(iv)  $f = g \Box \phi$  for some  $g \in \Gamma_0(\mathbb{R}^n)$ .

(v) -f is  $a_{\phi}$ -weakly convex.

Proof.

 $4.6(iv) \Rightarrow 4.6(iii)$  Assume that  $f = g \Box \phi$  for some  $g \in \Gamma_0(\mathbb{R}^n)$ . By taking the convex conjugate, we obtain via [41, Theorem 11.23(a)] that  $f^* = (g \Box \phi)^* = g^* + \phi^*$ , showing that  $f^* - \phi^* = g^*$  is convex.

 $4.6(\text{iii}) \Rightarrow 4.6(\text{iv})$  Suppose that  $h := f^* - \phi^*$  is convex. Since  $f^*$  is proper l.s.c. and  $\phi^*$  is smooth, we deduce that h is also proper and l.s.c. Then, since f is proper,

convex, and l.s.c. and  $f^* = h + \phi^*$ , we have that  $f = f^{**} = (h + \phi^*)^*$ , and since  $\phi^*$  has full domain, in view of [41, Theorem 11.23(a)], we obtain that  $(h + \phi^*)^* = h^* \Box \phi$ . The claim follows by considering  $g = h^* \in \Gamma_0(\mathbb{R}^n)$ .

 $4.6(i) \Rightarrow 4.6(ii)$  Let f be  $a_{\phi}$ -smooth. In particular, this means that -f is  $a_{\phi}$ -weakly convex and  $f \in \mathcal{C}^1(\mathbb{R}^n)$ . Invoking Proposition 4.5, this implies that -f is  $\Phi$ -convex.

 $4.6(ii) \Rightarrow 4.6(iv)$  In view of Lemma 2.5,  $\Phi$ -convexity of -f is equivalent to the identity  $(-f)^{\Phi\Phi} = -f$ . Specialized to  $\Phi(x, y) = -\phi(x - y)$ , we have in light of Remark 3.7 that

(4.12) 
$$f = -(-f)^{\Phi \Phi} = (-f)^{\Phi} \Box \phi.$$

It remains to show that  $(-f)^{\Phi} \in \Gamma_0(\mathbb{R}^n)$ . Thanks to Lemma 4.2, since  $f \in \Gamma_0(\mathbb{R}^n)$ , we have

$$(-f)^{\Phi}(y) \stackrel{(def)}{=} \sup_{x \in \mathbb{R}^n} \left\{ -\phi(x-y) + f(x) \right\} = \sup_{\xi \in \mathbb{R}^n} \left\{ f(y-\xi) - \phi_-(\xi) \right\} = (f^* - \phi^*)^*(y),$$

and thus, in particular,  $(-f)^{\Phi}$  is convex, l.s.c. It is also proper since f is proper.

 $4.6(iv) \Rightarrow 4.6(ii)$  Follows by Lemma 2.5 using the identity (4.12).

 $4.6(iv) \Rightarrow 4.6(i)$  Let  $f = g \Box \phi$  for some  $g \in \Gamma_0(\mathbb{R}^n)$ . As shown above, this implies that -f is  $\Phi$ -convex. Since  $\phi \in \mathcal{C}^1(\mathbb{R}^n)$  is supercoverve, [8, Corollary 18.8] implies that  $f \in \mathcal{C}^1(\mathbb{R}^n)$ . Thanks to Proposition 4.5, we deduce that -f is  $a_{\phi}$ -weakly convex. Since  $f \in \Gamma_0(\mathbb{R}^n)$ , in view of Lemma 3.11, f is  $a_{\phi}$ -weakly convex too. Overall, this means that f is  $a_{\phi}$ -smooth.

 $4.6(i) \Rightarrow 4.6(v)$  Follows by definition (cf. Definition 3.8(iii))

 $4.6(\mathbf{v}) \Rightarrow 4.6(\mathbf{ii})$  Assume that -f is  $a_{\phi}$ -weakly convex. This implies that -f is proper by definition. Since f is proper convex, this means that f is finite-valued and thus strictly continuous in view of [41, Example 9.14]. This implies that also -f is strictly continuous, and thus, in view of [41, Theorem 9.13] and [41, Corollary 8.10],  $\operatorname{dom} \partial(-f) = \mathbb{R}^n$ .  $A_{\phi}$ -weak convexity of -f thus implies via Proposition 3.12 that  $\operatorname{dom} \partial_{\Phi}(-f) = \mathbb{R}^n$ . In view of Lemma 2.6, this means that  $-f = (-f)^{\Phi\Phi}$  and hence  $\Phi$ -convex by Lemma 2.5.

We remark that the equivalence between Theorems 4.6(iii) and 4.6(iv) was previously shown in [45, Lemma 4.2] for more general Legendre functions  $\phi$  that need not be supercoercive.

**4.3. Examples and a\_{\phi}-smooth calculus.** We conclude this section by providing examples of  $a_{\phi}$ -smooth and  $a_{\phi}$ -strongly convex functions and their calculus. By exploiting the conjugate duality between  $B_{\phi}$ -smooth and  $a_{\phi^*}$ -strongly convex functions from Theorem 4.3, we provide an example for an  $a_{\phi^*}$ -strongly convex function on  $\mathbb{R}^2$ .

Example 4.7. Let  $A = [1 - 1] \in \mathbb{R}^{1 \times 2}$  and b = 5. Consider  $f : \mathbb{R}^2 \to \mathbb{R}$  defined by  $f(x) = \frac{1}{4} |Ax - b|^4$ . In view of [31, Proposition 2.1], f is  $B_{\phi}$ -smooth with reference function

$$\phi(x) = \frac{L}{4} \|x\|_2^4 + \frac{L}{2} \|x\|_2^2$$

for  $L = 3||A||^4 + 6||A||^3|b| + 3||A||^2|b| = 42 + 60\sqrt{2}$ . Invoking [41, Theorem 11.23(b)], we have

$$f^*(y) = \inf \left\{ (3/4)|x|^{4/3} + bx : A^\top x = y \right\} = \begin{cases} 5y_1 + \frac{3}{4}|y_1|^{4/3} & y_1 + y_2 = 0, \\ \infty & \text{otherwise.} \end{cases}$$

In light of Theorem 4.3,  $f^*$  is  $a_{\phi^*}$ -strongly convex, where

$$\phi^*(y) = \frac{3}{4L^{1/3}} \|y\|_2^{4/3} \Box \frac{1}{2L} \|y\|_2^2.$$

Next we provide an example for a function which is  $a_{\phi}$ -smooth but, due to the lack of shift-invariance in Bregman distances, not  $B_{\phi}$ -smooth. This complements Example 3.6.

Example 4.8 ( $a_{\phi}$ -smoothness versus  $B_{\phi}$ -smoothness). Let  $\phi = \frac{2}{3} |\cdot|^{3/2}$ , and let  $f : \mathbb{R} \to \mathbb{R}$  be defined as  $f = \delta_{[-1,1]} \Box \phi$ , namely,

$$f(x) = \begin{cases} 0 & x \in [-1,1], \\ (2/3)|x+1|^{3/2} & x < -1, \\ (2/3)|x-1|^{3/2} & x > 1. \end{cases}$$

By Theorem 4.6, f is  $a_{\phi}$ -smooth. The convex conjugate  $f^* = |\cdot| + (1/3)|\cdot|^3$  is  $B_{\phi^*}$ -strongly convex, where  $\phi^* = \frac{1}{3}|\cdot|^3$ . We prove that f is not  $B_{L\phi}$ -smooth for any L > 0. Note that  $f''(x) = 1/\sqrt{4(x-1)}$  for x > 1, while  $L\phi''(x) = L/\sqrt{4x}$  for x > 0. For  $x \searrow 1$ , we have that  $L\phi''(x) \searrow L/2$ , while  $f''(x) \to \infty$ . This contradicts the convexity of  $L\phi - f$ , which would require  $L\phi'' - f'' \ge 0$  on  $(1,\infty)$ .

Moreover, f is not  $B_{L\phi(.-b)}$ -smooth for any L > 0 and  $b \in \mathbb{R}^n$  since f has an unbounded second-order derivative at the points  $x \in \{\pm 1\}$ , i.e.,  $f''(x) \to \infty$  for  $x \nearrow -1$  (resp.,  $x \searrow +1$ ).

By invoking Theorem 4.6, we can provide a simple rule of calculus for  $a_{\phi}$ -smooth functions. For a > 0 and some function  $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ , the epi-scaling  $a \star f$  is defined by  $(a \star f)(x) := af(a^{-1}x)$ .

COROLLARY 4.9 (epi-calculus for  $a_{\phi}$ -smoothness). Let  $f_1, f_2 \in \Gamma_0(\mathbb{R}^n)$  be  $a_{\phi}$ smooth relative to  $\phi_1$  and  $\phi_2$ , respectively, and let  $a_1, a_2 > 0$ . Assume that one function among  $f_1, f_2$  is lower bounded and the other is coercive or that one of the two functions is supercoercive. Then  $a_1 \star f_1 \square a_2 \star f_2$  is  $a_{\phi}$ -smooth with reference function  $\phi :=$  $a_1 \star \phi_1 \square a_2 \star \phi_2$ .

*Proof.* In view of Theorem 4.6, there exist  $g_i \in \Gamma_0(\mathbb{R}^n)$  such that  $f_i = g_i \Box \phi_i$ , i = 1, 2. By [8, Proposition 12.14], we have that  $a_1 \star f_1 \Box a_2 \star f_2 \in \Gamma_0(\mathbb{R}^n)$  and thus

$$a_1 \star f_1 \square a_2 \star f_2 = (a_1 \star f_1 \square a_2 \star f_2)^{**} = (a_1 f_1^* + a_2 f_2^*)^*$$
$$= (a_1 g_1^* + a_2 g_2^* + a_1 \phi_1^* + a_2 \phi_2^*)^*.$$

Since  $\phi_i$  is Legendre and supercoercive and has full domain, this also holds for  $a_1\phi_1^* + a_2\phi_2^*$ , which is thus compliant with Assumptions (A1)–(A3) and by [41, Theorem 11.23(a)]  $(a_1\phi_1^* + a_2\phi_2^*)^* = a_1 \star \phi_1 \Box a_2 \star \phi_2 = \phi$ . Therefore,  $a_1g_1^* + a_2g_2^* + a_1\phi_1^* + a_2\phi_2^*$  is  $B_{\phi^*}$ -strongly convex. Invoking Theorem 4.6,  $(a_1g_1^* + a_2g_2^* + a_1\phi_1^* + a_2\phi_2^*)^* = a_1 \star f_1 \Box a_2 \star f_2$  is  $a_{\phi}$ -smooth as claimed.

In contrast to  $B_{\phi}$ -smoothness, the sum of  $a_{\phi}$ -smooth functions may fail to be  $a_{\phi}$ smooth. In fact,  $a_{\phi}$ -smoothness is not even preserved by addition of a linear function. In combination with Example 3.6, this illustrates that, in general, shift- and tiltinvariance are mutually exclusive properties of  $a_{\phi}$ - and  $B_{\phi}$ -smoothness, respectively.

Example 4.10 (lack of tilt-invariance for  $a_{\phi}$ -smooth functions). Let  $f = \phi = \frac{1}{3} |\cdot|^3$ and g(x) = 5x. Clearly, both f and g are  $a_{\phi}$ -smooth since  $g^* = \delta_{\{5\}}$  is  $B_{\phi^*}$ -strongly convex, where  $\phi^* = \frac{2}{3} |\cdot|^{3/2}$ . In view of [41, Theorem 11.23(a)], we have

Copyright (c) by SIAM. Unauthorized reproduction of this article is prohibited.

$$f + g = (f + g)^{**} = (f^* \Box g^*)^*,$$

where  $f^* \Box g^* = \frac{2}{3} |\cdot|^{3/2} \Box \delta_{\{5\}} = \frac{2}{3} |\cdot|^{5/2}$ . For  $x \searrow 0$ , we have  $(f^* \Box g^*)''(x) \rightarrow 1/\sqrt{20}$ , while  $(\phi^*)''(x) = 1/\sqrt{4x} \rightarrow \infty$ . Thus,  $f^* \Box g^*$  is not  $B_{\phi^*}$ -strongly convex. From Theorem 4.6, we infer that f + g is not  $a_{\phi}$ -smooth.

If the conjugate reference function  $\phi^*$  generates a jointly convex Bregman distance  $D_{\phi^*}$ ,  $a_{\phi}$ -smoothness is closed under pointwise average as recently shown in the context of the Bregman proximal average [45, Theorem 5.1(ii)]. Unfortunately, however, joint convexity of  $D_{\phi^*}$  under Assumption (A3), i.e., full domain of  $\phi^*$ , implies that  $\phi^*$  is quadratic [7, Remark 3.6]. The more general case has been recently explored in [30] along with a practical algorithm and more practical examples complementing the ones provided in [32].

## 5. Anisotropic and generalized convexity gap.

**5.1.** Univariate pointwise maxima and minima. In this subsection, we investigate whether  $\Phi$ -convexity implies anisotropic convexity, i.e., whether  $\Phi$ -convex functions for  $\Phi(x, y) = r\phi(x - y)$  with  $r \in \{\pm 1\}$  satisfy the anisotropic subgradient inequality (3.10). We have already seen in Proposition 4.5 (resp., Theorem 4.3) that  $\Phi$ -convexity is equivalent to  $a_{\phi}$ -weak convexity (resp.,  $a_{\phi}$ -strong convexity) under (essential) smoothness of f. Instead, in this subsection, our focus is on nonsmooth  $\Phi$ -envelopes. To this end, we confine ourselves to  $\Phi$ -convex functions where the index set  $\mathcal{I}$  as in Definition 2.1 is finite. This, in general, leads to functions which have downward-pointing cusps at points at which multiple pieces intersect and the limiting subdifferential is multivalued. We first show that at least in the univariate case, such functions satisfy the anisotropic subgradient inequality. To this end, we first show that this holds whenever the index set  $\mathcal{I}$  is a doubleton. The situation is depicted in Figure 3 (upper row).

LEMMA 5.1. Let n = 1. Let r = +1 (resp., r = -1), and choose  $\Phi(x, y) = r\phi(x - y)$ . Let  $f : \mathbb{R} \to \mathbb{R}$  with  $f(x) = \max_{i \in \{1,2\}} \{\Phi(x, y_i) - \beta_i\}$  for some  $y_i, \beta_i \in \mathbb{R}$  with  $i \in \{1,2\}$ . Then f satisfies the  $a_{\phi}$ -strong convexity inequality (3.9) (resp., the  $a_{\phi}$ -weak convexity inequality (3.8)) at any  $(\bar{x}, \bar{v}) \in \text{gph} \partial f$ .

*Proof.* Let  $\bar{x} \in \mathbb{R}$  be fixed, and define  $h_i := r\phi(\cdot - y_i) - \beta_i$  for  $i \in \{1, 2\}$ . Fix  $\bar{v} \in \partial f(\bar{x}) = \operatorname{con} \{h'_i(\bar{x}) : h_i(\bar{x}) = f(\bar{x}), i \in \{1, 2\}\}$ ; cf. [41, Exercise 8.31]. If  $\bar{v} = h'_j(\bar{x}) = r\phi'(\bar{x} - y_j)$  for some  $j \in \{1, 2\}$  with  $r\phi(\bar{x} - y_j) - \beta_j \stackrel{(def)}{=} h_j(\bar{x}) = f(\bar{x})$ , the anisotropic subgradient inequality (3.10) holds at  $(\bar{x}, y_j)$  with  $y_j = \bar{x} - (\phi^*)'(r\bar{v})$ , having

$$f(x) \stackrel{(def)}{=} \max \{ r\phi(x - y_i) - \beta_i : i \in \{1, 2\} \} \ge f(\bar{x}) + r\phi(x - y_j) - r\phi(\bar{x} - y_j)$$

in this case. Therefore, we may assume that  $h_1(\bar{x}) = h_2(\bar{x}) = f(\bar{x})$  and  $\bar{v} = \tau h'_1(\bar{x}) + (1-\tau)h'_2(\bar{x})$  for some  $\tau \in (0,1)$ . Let  $\bar{y} := \bar{x} - (\phi^*)'(r\bar{v})$ . To arrive at a contradiction, suppose that there exists  $\hat{x}$  at which the anisotropic subgradient inequality (3.10) is violated:

$$r\phi(\hat{x} - y_i) - \beta_i < f(\bar{x}) + r\phi(\hat{x} - \bar{y}) - r\phi(\bar{x} - \bar{y}), \quad i \in \{1, 2\}.$$

Since  $\beta_i = r\phi(\bar{x} - y_i) - f(\bar{x})$ , the above inequality can be equivalently rewritten as

$$q_{\bar{u}}(\hat{x} + (\bar{y} - y_i)) - q_{\bar{u}}(\hat{x}) < q_{\bar{u}}(\bar{x} + (\bar{y} - y_i)) - q_{\bar{u}}(\bar{x}), \quad i \in \{1, 2\},$$

Copyright (C) by SIAM. Unauthorized reproduction of this article is prohibited.



FIG. 3. Upper row: illustration of  $a_{\phi}$ -strong convexity (left) and  $a_{\phi}$ -weak convexity (right) of univariate pointwise maxima as in Lemma 5.1. Lower row: failure of  $a_{\phi}$ -strong convexity (left) and  $a_{\phi}$ -weak convexity (right) of multivariate pointwise maxima plotted along the line  $t \mapsto \bar{x} + t(\hat{x} - \bar{x})$ as described in Examples 5.4 and 5.5. The dashed lines correspond to the lower approximations  $x \mapsto r\phi(x - \bar{y}) - \bar{\beta}$  with  $\bar{y} = \bar{x} - \nabla \phi^*(r\bar{v})$  and  $\bar{\beta} = r\phi(\nabla \phi^*(r\bar{v})) - f(\bar{x})$  as in the anisotropic subgradient inequality (3.10), where  $\bar{v} \in \partial f(\bar{x})$  is an element of the multivalued subdifferential of f at  $\bar{x}$ , a point where the graph of f has a cusp. While in the upper row these approximations are global lower bounds, in the lower row, the dashed curves are not below the solid ones.

where  $q_{\bar{y}} := r\phi(\cdot -\bar{y})$ . Without loss of generality, assume that r = 1. Since  $q_{\bar{y}} \in \mathcal{C}^1(\mathbb{R}^n)$ is strictly convex (strictly concave if r = -1) and thus  $(q'_{\bar{y}}(x + (\bar{y} - y_i)) - q'_{\bar{y}}(x))(\bar{y} - y_i) > 0$ , the function  $x \mapsto q_{\bar{y}}(x + (\bar{y} - y_i)) - q_{\bar{y}}(x)$  is strictly monotone (either increasing or decreasing depending on the sign of  $\bar{y} - y_i$ ). The above inequality then implies that

$$q_{\bar{y}}(x + (\bar{y} - y_i)) - q_{\bar{y}}(x) < q_{\bar{y}}(\bar{x} + (\bar{y} - y_i)) - q_{\bar{y}}(\bar{x}), \quad i \in \{1, 2\}\,,$$

holds for all x between  $\hat{x}$  and  $\bar{x}$ . This means that  $f(x) < r\phi(x-\bar{y}) - r\phi(\bar{x}-\bar{y}) + f(\bar{x})$ for all such x and with equality holding for  $x = \bar{x}$ . Without loss of generality, assume that  $\hat{x} > \bar{x}$ . Let  $j \in \{1, 2\}$  be such that  $h'_j(\bar{x}) > \bar{v}$  (which exists since  $\bar{v}$  lies strictly between  $h'_1(\bar{x})$  and  $h'_2(\bar{x})$ ). Since  $\bar{v} = r\phi'(\bar{x}-\bar{y})$ , by linearizing  $r\phi(\cdot -\bar{y})$  at  $\bar{x}$ , one obtains

$$f(x) < r\phi(x - \bar{y}) - r\phi(\bar{x} - \bar{y}) + f(\bar{x}) = \bar{v}(x - \bar{x}) + o(|x - \bar{x}|) + f(\bar{x})$$

for all  $x \in (\bar{x}, \hat{x})$ . Similarly, by linearizing  $r\phi(\cdot -y_i)$  at  $\bar{x}$ , we also have

$$f(x) \ge r\phi(x-y_j) - r\phi(\bar{x}-y_j) + f(\bar{x}) = h'_j(\bar{x})(x-\bar{x}) + o(|x-\bar{x}|) + f(\bar{x})$$

for all  $x \in (\bar{x}, \hat{x})$ . By combining the two inequalities, we arrive to

$$h'_{i}(\bar{x})(x-\bar{x}) + o(|x-\bar{x}|) < \bar{v}(x-\bar{x}) + o(|x-\bar{x}|)$$

whence the contradiction  $h'_i(\bar{x}) < \bar{v}$ .

PROPOSITION 5.2. Let n = 1. Let r = +1 (resp., r = -1), and choose  $\Phi(x, y) = r\phi(x - y)$ . Let  $f : \mathbb{R} \to \mathbb{R}$  with  $f(x) = \max_{i \in \mathcal{I}} \Phi(x, y_i) - \beta_i$  for some  $y_i, \beta_i \in \mathbb{R}$  with  $i \in \mathcal{I}$  finite. Then f is  $a_{\phi}$ -strongly convex (resp.,  $a_{\phi}$ -weakly convex).

*Proof.* Let  $\bar{x} \in \mathbb{R}$ . Since  $\Phi(\cdot, y_i) - \beta_i \in \mathcal{C}^1(\mathbb{R}^n)$  and  $\mathcal{I}$  is finite, [41, Exercise 8.31] gives

$$\partial f(\bar{x}) = \operatorname{con} \left\{ r\phi'(\bar{x} - y_i) : r\phi(\bar{x} - y_i) - \beta_i = f(\bar{x}), \, i \in \mathcal{I} \right\}.$$

Let  $\bar{v} \in \partial f(\bar{x})$  be fixed. Then there exist  $i_1, i_2 \in \mathcal{I}$  (not necessarily distinct) with  $r\phi(\bar{x} - y_{i_1}) - \beta_{i_1} = r\phi(\bar{x} - y_{i_2}) - \beta_{i_2} = f(\bar{x})$  such that  $r\phi'(\bar{x} - y_{i_1}) \leq \bar{v} \leq r\phi'(\bar{x} - y_{i_2})$ . Define  $h := \max\{r\phi(\cdot - y_{i_j}) - \beta_{i_j} : j \in \{1, 2\}\}$ . In view of [41, Exercise 8.31],  $\bar{v} \in \partial h(\bar{x}) = \operatorname{con}\{r\phi'(\bar{x} - y_{i_j}) : r\phi(\bar{x} - y_{i_j}) - \beta_{i_j} = h(\bar{x}), j \in \{1, 2\}\}$ . Lemma 5.1 applied to h then yields

$$h(x) \ge h(\bar{x}) + r\phi(x - \bar{x} + (\phi^*)'(r\bar{v})) - r\phi((\phi^*)'(r\bar{v})) \quad \forall x \in \mathbb{R}^n.$$

By using the fact that  $h(\bar{x}) = f(\bar{x})$ , we have for  $\bar{y} := \bar{x} - (\phi^*)'(r\bar{v})$ 

$$f(x) \stackrel{(def)}{=} \max_{i \in \mathcal{I}} \Phi(x, y_i) - \beta_i \ge h(x) \ge f(\bar{x}) + r\phi(x - \bar{y}) - r\phi(\bar{x} - \bar{y}) \quad \forall x \in \mathbb{R}.$$

Since  $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$  was arbitrary, the claim follows.

**5.2.** Anisotropic strong convexity. In the univariate case, in light of Proposition 5.2 for r = +1,  $\Phi$ -convexity with finite  $\mathcal{I}$  implies  $a_{\phi}$ -strong convexity. The situation is depicted in Figure 3 (upper left). Since finite-valued  $\Phi$ -convex functions for r = +1 are in particular convex and thus strictly continuous, we can derive the following refinement of Proposition 5.2 for  $\Phi$ -convex functions with  $\mathcal{I}$  not necessarily being finite.

PROPOSITION 5.3 ( $a_{\phi}$ -strong convexity of supremal convolutions on  $\mathbb{R}$ ). Let f:  $\mathbb{R} \to \mathbb{R}$  be  $\Phi$ -convex for  $\Phi(x,y) = \phi(x-y)$  (equivalently,  $f = (-g) \Diamond \phi$  for some  $g \in \Gamma_0(\mathbb{R})$ ). Then f is  $a_{\phi}$ -strongly convex.

*Proof.* Since f is finite-valued, in view of [41, Example 9.14], f is strictly continuous on  $\mathbb{R}$ . By Lemma 2.5, we have that

(5.1) 
$$f(x) = f^{\Phi\Phi}(x) = \sup_{y \in \mathbb{R}} \phi(x-y) - f^{\Phi}(y)$$

holds for any  $x \in \mathbb{R}$ , and Lemma 2.6 yields the equivalences  $y \in \partial_{\Phi} f(x) \Leftrightarrow x \in \partial_{\Phi} f^{\Phi}(y) \Leftrightarrow y \in \arg \max \{\phi(x - \cdot) - f^{\Phi}\}$ . In light of Propositions 3.12 and 3.13,  $\emptyset \neq \partial_{\Phi} f(x) \subseteq x - (\phi^*)'(\partial f(x))$ , and hence the supremum in (5.1) is attained on  $\partial_{\Phi} f(x) \subseteq x - (\phi^*)'(\partial f(x))$ . Let  $\bar{x} \in \mathbb{R}$ . Because of strict continuity,  $\partial f$  is locally bounded at  $\bar{x}$ , and as such, so is  $\partial_{\Phi} f$ . Thus, there exists a compact set Y so that for

Copyright (c) by SIAM. Unauthorized reproduction of this article is prohibited.

Π

any x sufficiently near  $\bar{x}$ , we can restrict the maximization over y in (5.1) to Y, i.e.,  $f(x) = \sup_{y \in Y} \phi(x - y) - f^{\Phi}(y)$  for x near  $\bar{x}$ . This implies that f is lower- $C^1$  in the sense of [41, Definition 10.29]. In light of [41, Theorem 10.31], we have that

$$\partial f(\bar{x}) \subseteq \operatorname{con} \left\{ \phi'(\bar{x} - y) : y \in \partial_{\Phi} f(\bar{x}) \right\}$$

Let  $\bar{v} \in \partial f(\bar{x})$  be fixed. Then there exist  $y_1, y_2 \in \partial_{\Phi} f(\bar{x}) = \arg \max \left\{ \phi(\bar{x} - \cdot) - f^{\Phi} \right\}$ (not necessarily distinct) such that  $\phi'(\bar{x} - y_1) \leq \bar{v} \leq \phi'(\bar{x} - y_2)$ . In particular, we have  $\phi(\bar{x} - y_i) - f^{\Phi}(y_i) = f(\bar{x})$  for  $i \in \{1, 2\}$ . Define  $h := \max\{\phi(\cdot - y_i) - f^{\Phi}(y_i) : i \in \{1, 2\}\}$ . In view of [41, Exercise 8.31],

$$\bar{v} \in \partial h(\bar{x}) = \operatorname{con} \left\{ \phi'(\bar{x} - y_i) : \phi(\bar{x} - y_i) - f^{\Phi}(y_i) = h(\bar{x}), \, i \in \{1, 2\} \right\}.$$

Lemma 5.1 applied to h then yields

2746

$$h(x) \ge h(\bar{x}) + \phi(x - \bar{x} + (\phi^*)'(\bar{v})) - \phi((\phi^*)'(\bar{v})) \quad \forall x \in \mathbb{R}^n.$$

By using the fact that  $\phi(\bar{x} - y_i) - f^{\Phi}(y_i) = f(\bar{x})$  for  $i \in \{1, 2\}$  and thus  $h(\bar{x}) = f(\bar{x})$ , we have for  $\bar{y} := \bar{x} - (\phi^*)'(\bar{v})$ 

$$f(x) = f^{\Phi\Phi}(x) = \sup_{y \in \mathbb{R}} \phi(x-y) - f^{\Phi}(y) \ge h(x) \ge f(\bar{x}) + \phi(x-\bar{y}) - \phi(\bar{x}-\bar{y}) \quad \forall x \in \mathbb{R}.$$

Since  $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$  was arbitrary, the claim follows.

In light of Theorem 4.3, this shows that the saddle-point property (4.3) holds automatically in the finite-valued univariate case. In the multivariate case, however, the situation is different:  $\Phi$ -convex functions on  $\mathbb{R}^n$  in general fail to be  $a_{\phi}$ -strongly convex. This is verified by the following counterexample.

*Example* 5.4 (failure of  $a_{\phi}$ -strong convexity in higher dimensions). Choose  $\phi$  :  $\mathbb{R}^2 \to \mathbb{R}$  as  $\phi(x) = x_1^2 + x_2^4$ , and for y = (-1, 1), let

$$f(x) := \max \left\{ \phi(x), \phi(x-y) \right\}$$

The graphs of the two pieces  $\phi, \phi(\cdot - y)$  intersect at points  $x \in \mathbb{R}^2$  at which  $\phi(x) = \phi(x - y)$ , i.e., which satisfy  $x_1 = 2x_2^3 - 3x_2^2 + 2x_2 - 1$ . At these points, the graph of f has a downward-pointing cusp. We have  $\nabla \phi(x) = (2x_1, 4x_2^3)$  and  $\nabla \phi(x - y) = (2x_1 + 2, 4(x_2 - 1)^3)$ . Take  $\bar{x} = (-1, 0)$ , at which  $\phi(\bar{x}) = \phi(\bar{x} - y)$  holds in particular, and

$$\bar{v} := \begin{pmatrix} -1 \\ -2 \end{pmatrix} \in \partial f(\bar{x}) = \operatorname{con}\left\{ \begin{pmatrix} -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -4 \end{pmatrix} \right\},\$$

so that  $\nabla \phi(\bar{x} - \bar{y}) = \bar{v}$  and  $\phi(\bar{x} - \bar{y}) - \bar{\beta} = f(\bar{x})$  for  $\bar{y} = \binom{-1/2}{2^{-1/3}}$  and  $\bar{\beta} = \phi(\bar{x} - \bar{y}) - f(\bar{x}) = \frac{1}{4} + 2^{-4/3} - 1$ . At  $\hat{x} = (-8, -1)$ , we have  $f(\hat{x}) = 65$  and  $\phi(\hat{x} - \bar{y}) - \bar{\beta} > 65 = f(\hat{x})$ , and thus  $\phi(\cdot - \bar{y}) - \bar{\beta}$  is not a lower bound of f. As a consequence, f is not  $a_{\phi}$ -strongly convex. This is shown in Figure 3 (lower left): It can be seen that the dashed line, the graph of  $\phi(\cdot - \bar{y}) - \bar{\beta}$ , is above the solid line, the graph of f, when restricted to the line  $\bar{x} + t(\hat{x} - \bar{x})$  and t is close to 1.

In light of Theorem 4.3, this shows that the saddle-point property (4.3) does not hold in general.

**5.3.** Anisotropic weak convexity. In light of Proposition 5.2, for r = -1, univariate pointwise maxima over finite collections of functions  $\Phi(\cdot, y_i) - \beta_i$  are  $a_{\phi}$ -weakly convex.

The following example reveals that this is no longer true in the multivariate case.

Example 5.5 (failure of  $a_{\phi}$ -weak convexity in higher dimensions). Choose  $\phi : \mathbb{R}^2 \to \mathbb{R}$  as  $\phi(x) = x_1^2 + x_2^4$ , and for y = (1, -1), let

$$f(x) := \max\left\{-\phi(x), -\phi(x-y)\right\} = \begin{cases} -\phi(x) & x_1 \le 2x_2^3 + 3x_2^2 + 2x_2 + 1, \\ -\phi(x-y) & \text{otherwise.} \end{cases}$$

The graphs of the two pieces  $-\phi, -\phi(\cdot - y)$  intersect at points  $x \in \mathbb{R}^2$  at which  $\phi(x) = \phi(x - y)$ . Take  $\bar{x} = (1, 0)$ , for which this holds in particular, and

$$\bar{v} := \begin{pmatrix} -1 \\ -2 \end{pmatrix} \in \partial f(\bar{x}) = \operatorname{con} \left\{ \begin{pmatrix} -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -4 \end{pmatrix} \right\},\$$

so that  $-\nabla \phi(\bar{x} - \bar{y}) = \bar{v}$  and  $-\phi(\bar{x} - \bar{y}) - \bar{\beta} = f(\bar{x})$  for

$$\bar{y} = \begin{pmatrix} 1/2 \\ -2^{-1/3} \end{pmatrix}$$
 and  $\bar{\beta} = \phi(\bar{x} - \bar{y}) - f(\bar{x}) = \frac{3 - 2^{2/3}}{4}.$ 

At  $\hat{x} = (0, -1)$ , we have  $-\phi(\hat{x} - \bar{y}) - \bar{\beta} = 2^{1/3}(2^{4/3} - 3) > -1 = f(\hat{x})$ , and thus  $-\phi(\cdot -\bar{y}) - \bar{\beta}$  is not a lower bound of f. As a consequence, f is not  $a_{\phi}$ -weakly convex. This is shown in Figure 3 (lower right): It can be seen that the dashed line, the graph of  $-\phi(\cdot -\bar{y}) - \bar{\beta}$ , is above the solid line, the graph of f, when restricted to the line  $\bar{x} + t(\hat{x} - \bar{x})$  and t is close to 1.

6. Conclusion. In this paper, we provided the dual counterparts to relative smoothness and strong convexity ( $B_{\phi}$ -smoothness and  $B_{\phi}$ -strong convexity), called anisotropic smoothness and strong convexity ( $a_{\phi^*}$ -smoothness and  $a_{\phi^*}$ -strong convexity). In the context of generalized convexity, both notions can be seen as  $\Phi$ -subgradient inequalities or as anisotropic generalizations of the descent lemma and strong convexity inequality which can be fruitful in the development of first-order algorithms. Unlike the Euclidean case, we have shown a gap between the classes of functions that satisfy the anisotropic strong convexity subgradient inequality and the class of supremal convolutions, the latter being a proper subset of the former for which a certain saddlepoint property holds. In spite of the duality relation being limited to convex functions, the viewpoint of  $\Phi$ -convexity naturally motivated the study of weak convexity counterparts, which led to the discovery of yet another unexpected gap. While  $B_{\phi}$ -weak convexity can be characterized in terms of  $\Phi$ -convexity, there exist  $\Phi$ -convex (with the corresponding  $\Phi$ ) functions that are not  $a_{\phi}$ -weakly convex; whether a converse inclusion holds is an open problem. It is also unclear whether differentiability follows from  $a_{\phi}$ -weak convexity of both f and -f in the same way it does for the Bregman notions. In general, we believe that a more detailed investigation of anisotropic weak convexity is an interesting direction for future research. Algorithmic implications and more practical examples of the newly introduced anisotropic notions are under investigation in [30].

Acknowledgments. We are grateful for the enlightening and encouraging feedback of three anonymous referees which led to substantial improvements of this manuscript.

#### E. LAUDE, A. THEMELIS, AND P. PATRINOS

#### REFERENCES

- M. AHOOKHOSH, A. THEMELIS, AND P. PATRINOS, A Bregman forward-backward linesearch algorithm for nonconvex composite optimization: Superlinear convergence to nonisolated local minima, SIAM J. Optim., 31 (2021), pp. 653–685.
- E. J. BALDER, An extension of duality-stability relations to nonconvex optimization problems, SIAM J. Control Optim., 15 (1977), pp. 329–343.
- [3] H. BAUERMEISTER, E. LAUDE, T. MÖLLENHOFF, M. MOELLER, AND D. CREMERS, Lifting the convex conjugate in Lagrangian relaxations: A tractable approach for continuous Markov random fields, SIAM J. Imaging Sci., 15 (2022), pp. 1253–1281.
- [4] 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), pp. 1068–1087.
- [5] H. H. BAUSCHKE, J. BOLTE, AND M. TEBOULLE, A descent lemma beyond Lipschitz gradient continuity: First-order methods revisited and applications, Math. Oper. Res., 42 (2017), pp. 330–348.
- [6] H. H. BAUSCHKE AND J. M. BORWEIN, Legendre functions and the method of random Bregman projections, J. Convex Anal., 4 (1997), pp. 27–67.
- [7] H. H. BAUSCHKE AND J. M. BORWEIN, Joint and separate convexity of the Bregman distance, in Studies in Computational Mathematics, Vol. 8, Elsevier, New York, 2001, pp. 23–36.
- [8] H. H. BAUSCHKE AND P. L. COMBETTES, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, Springer Science & Business Media, New York, 2011.
- [9] H. H. BAUSCHKE, P. L. COMBETTES, AND D. NOLL, Joint minimization with alternating Bregman proximity operators, Pac. J. Optim., 2 (2006), pp. 401–424.
- [10] H. H. BAUSCHKE, M. DAO, AND S. LINDSTROM, Regularizing with Bregman-Moreau envelopes, SIAM J. Optim., 28 (2018), pp. 3208–3228.
- [11] H. H. BAUSCHKE, M. S. MACKLEM, AND X. WANG, Chebyshev Sets, Klee Sets, and Chebyshev Centers with Respect to Bregman Distances: Recent Results and Open Problems, Springer, New York, 2011.
- [12] H. H. BAUSCHKE, X. WANG, J. YE, AND X. YUAN, Bregman distances and Chebyshev sets, J. Approx. Theory, 159 (2009), pp. 3–25.
- [13] H. H. BAUSCHKE, X. WANG, J. YE, AND X. YUAN, Bregman distances and Klee sets, J. Approx. Theory, 158 (2009), pp. 170–183.
- [14] D. P. BERTSEKAS, Nonlinear Programming, Athena Scientific Optimization and Computation Series, Athena Scientific, Nashua, NH, 1995.
- [15] B. BIRNBAUM, N. R. DEVANUR, AND L. XIAO, Distributed algorithms via gradient descent for Fisher markets, in Proceedings of the 12th ACM Conference on Electronic Commerce, 2011, pp. 127–136.
- [16] J. BOLTE, S. SABACH, M. TEBOULLE, AND Y. VAISBOURD, First order methods beyond convexity and Lipschitz gradient continuity with applications to quadratic inverse problems, SIAM J. Optim., 28 (2018), pp. 2131–2151.
- [17] L. M. BREGMAN, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, USSR Comput. Math. Math. Phys., 7 (1967), pp. 200–217.
- [18] H. T. BUI, R. S. BURACHIK, A. Y. KRUGER, AND D. T. YOST, Zero duality gap conditions via abstract convexity, Optimization, 71 (2022), pp. 811–847.
- [19] A. CABOT, A. JOURANI, AND L. THIBAULT, Envelopes for sets and functions: Regularization and generalized conjugacy, Mathematika, 63 (2017), pp. 383–432.
- [20] A. CABOT, A. JOURANI, L. THIBAULT, AND D. ZAGRODNY, The attainment set of the φ-envelope and genericity properties, Math. Scand., 124 (2019), pp. 203–246.
- [21] Y. Y. CHEN, C. KAN, AND W. SONG, The Moreau envelope function and proximal mapping with respect to the Bregman distance in Banach spaces, Vietnam J. Math., 40 (2012), pp. 181–199.
- [22] D. DAVIS, D. DRUSVYATSKIY, AND K. J. MACPHEE, Stochastic Model-Based Minimization under High-Order Growth, preprint, arXiv:1807.00255, 2018.
- [23] S. DOLECKI AND S. KURCYUSZ, On Φ-convexity in extremal problems, SIAM J. Control Optim., 16 (1978), pp. 277–300.
- [24] R.-A. DRAGOMIR, A. B. TAYLOR, A. D'ASPREMONT, AND J. BOLTE, Optimal complexity and certification of Bregman first-order methods, Math. Program., 194 (2022), pp. 41–83.
- [25] J.-B. HIRIART-URRUTY, A general formula on the conjugate of the difference of functions, Canad. Math. Bull., 29 (1986), pp. 482–485.
- [26] A. JOURANI, L. THIBAULT, AND D. ZAGRODNY, The NSLUC property and Klee envelope, Math. Ann., 365 (2016), pp. 923–967.

- [27] C. KAN AND W. SONG, The Moreau envelope function and proximal mapping in the sense of the Bregman distance, Nonlinear Anal. Theory Methods Appl., 75 (2012), pp. 1385–1399.
- [28] E. LAUDE, Lower Envelopes and Lifting for Structured Nonconvex Optimization, Ph.D. thesis, Technical University of Munich, 2021.
- [29] E. LAUDE, P. OCHS, AND D. CREMERS, Bregman proximal mappings and Bregman-Moreau envelopes under relative prox-regularity, J. Optim. Theory Appl., 184 (2020), pp. 724-761.
- [30] E. LAUDE AND P. PATRINOS, Anisotropic Proximal Gradient, preprint, arXiv:2210.15531, 2022.
- [31] H. LU, R. M. FREUND, AND Y. NESTEROV, Relatively smooth convex optimization by first-order methods, and applications, SIAM J. Optim., 28 (2018), pp. 333–354.
- [32] C. J. MADDISON, D. PAULIN, Y. W. TEH, AND A. DOUCET, Dual space preconditioning for gradient descent, SIAM J. Optim., 31 (2021), pp. 991–1016.
- [33] B. S. MORDUKHOVICH AND N. M. NAM, An easy path to convex analysis and applications, Synth. Lect. Math. Stat., 6 (2013), pp. 1–218.
- [34] J.-J. MOREAU, Fonctionnelles convexes, Séminaire Jean Leray, (1966), pp. 1–108.
- [35] J.-J. MOREAU, Inf-convolution, sous-additivité, convexité des fonctions numériques, J. Math. Pures Appl., 49 (1970), pp. 33–41.
- [36] M. C. MUKKAMALA, J. FADILI, AND P. OCHS, Global convergence of model function based Bregman proximal minimization algorithms, J. Global Optim., 83 (2022), pp. 753–781.
- [37] J.-P. PENOT AND M. VOLLE, On strongly convex and paraconvex dualities, in Generalized Convexity and Fractional Programming with Economic Applications, Springer, New York, 1990, pp. 198–218.
- [38] B. N. PSHENICHNYI, Leçons sur les jeux différentiels, Contrôle optimal et jeux différentiels, Cahiers de l'IRIA, no. 4 (1971).
- [39] R. T. ROCKAFELLAR, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.
- [40] R. T. ROCKAFELLAR, Augmented Lagrange multiplier functions and duality in nonconvex programming, SIAM J. Optim., 12 (1974), pp. 268–285.
- [41] R. T. ROCKAFELLAR AND R. J. WETS, Variational Analysis, Springer, New York, 1998.
- [42] C. VILLANI, Optimal Transport: Old and New, Springer, New York, 2008.
- [43] C. VILLANI, Topics in Optimal Transportation, Vol. 58, American Mathematical Society, Providence, RI, 2021.
- [44] X. WANG, On Chebyshev functions and Klee functions, J. Math. Anal. Appl., 368 (2010), pp. 293–310.
- [45] X. WANG AND H. H. BAUSCHKE, The Bregman proximal average, SIAM J. Optim., 32 (2022), pp. 1379–1401.