THE GENERALIZED BREGMAN DISTANCE*

REGINA S. BURACHIK†, MINH N. DAO‡, AND SCOTT B. LINDSTROM§

Abstract. Recently, a new kind of distance has been introduced for the graphs of two point-to-set operators, one of which is maximally monotone. When both operators are the subdifferential of a proper lower semicontinuous convex function, this kind of distance specializes under modest assumptions to the classical Bregman distance. We name this new kind of distance the *generalized Bregman distance*, and we shed light on it with examples that utilize the other two most natural representative functions: the Fitzpatrick function and its conjugate. We provide sufficient conditions for convexity, coercivity, and supercoercivity: properties which are essential for implementation in proximal point type algorithms. We establish these results for both the left and right variants of this new kind of distance. We construct examples closely related to the Kullback–Leibler divergence, which was previously considered in the context of Bregman distances and whose importance in information theory is well known. In so doing, we demonstrate how to compute a difficult Fitzpatrick conjugate function, and we discover natural occurrences of the Lambert $\mathcal{W}$ function, whose importance in optimization is of growing interest.

Key words. Bregman distance, generalized Bregman distance, convex function, Fitzpatrick distance, Fitzpatrick function, representative function, regularization

AMS subject classifications. 90C25, 49K40, 47H05

DOI. 10.1137/19M1288140

1. Introduction. Throughout, unless stated otherwise, $(X, \|\cdot\|)$ is a real Banach space with dual $(X^*, \|\cdot\|_*)$, and $\Gamma_0(X)$ is the set of all proper lower semicontinuous convex functions from X to $\mathbb{R}_\infty := \,]-\infty, \infty]$.

In 1967, Bregman introduced the distance constructed for a differentiable convex function f,

(1)

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

which now bears his name [14] and whose corresponding envelopes and proximity operators specify to the Moreau proximity operator [1, 2, 36] and envelope when f is the energy $\|\cdot\|^2/2$. The study of Bregman distances has become popular, following their 1981 reintroduction by Censor and Lent [27]. Bregman functions were introduced to solve feasibility problems and then used for producing more general versions of the classical proximal-point algorithm, both for convex minimization and for monotone variational inequality problems. The use of Bregman distances in prox-like methods

---

†Mathematics, UniSA STEM, University of South Australia, Mawson Lakes, SA 5095, Australia (regina.burachik@unisa.edu.au).

‡School of Engineering, Information Technology and Physical Sciences, Federation University Australia, Ballarat, VIC 3353, Australia (m.dao@federation.edu.au).

§Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong (lindstrom. scott@gmail.com).

for convex minimization is found in [15, 6, 24, 25, 28, 29, 30, 35, 38, 39], while the use for monotone variational inequalities can be found in [16, 17, 18, 20, 22, 32]. Properties of Bregman functions have been the focus of research since the late 1990s; see [3, 4, 26]. A good, brief bibliographic overview of their history is found in [33, page 1233], and the book of Censor and Zenios is also instructive [29]. When f is not the energy, the distance may fail to be symmetric, and so one is led to consider the left and right versions of envelopes and their associated proximity operators. The asymptotic properties of Bregman envelopes with respect to a parameter were explored in [7].

From [3, Proposition 3.2], we have a dual characterization of the Bregman distance for a differentiable convex function f:

$$\mathcal{D}_f(x,y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle \tag{2a}$$

$$= f(x) + f^*(\nabla f(y)) - \langle \nabla f(y), x \rangle. \tag{2b}$$

Here (2a) is the definition of the Bregman distance and (2b) uses the Fenchel–Young equality: $(\forall v \in \nabla f(y))\ f(y) + f^*(v) = \langle y, v \rangle$. From (2b), an observation is that

$$\mathcal{D}_f(x,y) = 0 \iff \nabla f(x) = \nabla f(y), \tag{3}$$

and so $\mathcal{D}_f$ has the dual characterization of serving as a distance between gradients.[1] Based on this characterization, Burachik and Martínez-Legaz [21] introduced a new kind of distance based on the representative function h of a monotone operator:

$$\mathcal{D}_T^{\flat,h}(x,y) := \inf_{v \in Ty} (h(x,v) - \langle x,v \rangle), \tag{4a}$$

$$\mathcal{D}_T^{\sharp,h}(x,y) := \sup_{v \in Ty} (h(x,v) - \langle x,v \rangle). \tag{4b}$$

This kind of distance generalizes the Bregman distance, specializing—under the mild domain conditions in section 2.2—thereto when h is the *Fenchel–Young representative* for $T = \nabla f$, which is defined by $f \oplus f^*(x,y) := f(x) + f^*(y)$.

Naturally, we name this more general distance the *generalized Bregman distance (GBD)*. In lieu of the Fenchel–Young representative, the Fitzpatrick function and its conjugate are the two other functions that are most natural to consider. As with the Bregman distance, we obtain left and right variants; these admit new left and right coercivity and supercoercivity properties, along with envelopes and corresponding proximity operators, when the GBD replaces the Bregman distance in the construction of envelopes.

1.1. Outline and contributions. This article is outlined as follows. In section 2, we recall the GBD as introduced in [21]. We provide its basic properties and clarify the domain conditions under which the Fenchel–Young representative case specializes to the Bregman distance. We also introduce the closed variant, which may specialize to the Bregman distance at more points on the boundary of the domain.

In section 3, we show how to compute the new GBDs. We first illustrate with the energy, which specifies to the Moreau case when the Fenchel–Young representative is used. We also illustrate with the Boltzmann–Shannon entropy, whose derivative

---

[1]To show $\impliedby$, simply substitute $\nabla f(x)$ for $\nabla f(y)$ and apply Fenchel–Young once more, this time for the variable x instead of y.

is the logarithm and whose Bregman distance, the Kullback–Leibler divergence, is commonly used as a measure of the difference between positive vectors in information theory and elsewhere. We compare the Kullback–Leibler divergence with the similar Fenchel–Young representative GBD for the logarithm, and also with its closed version. We also illustrate how to compute with the two other most natural representatives to consider: the book-end cases for the representative function set. These are the smallest representative, named the *Fitzpatrick function*, and the biggest representative, which is obtained using the conjugate of the Fitzpatrick function.

Interestingly, while the Fitzpatrick function for the logarithm was discovered in [9], the present work contains the first computation of its conjugate. The discovery and proof rely upon the graphical characterizations of representative functions, and the special function Lambert $\mathcal{W}$ plays an important role in the computational aspects of discovery. The way that we tackle this problem is very prototypical of the approach that one might need to use when computing other representative functions and GBDs. We furnish a full discussion of the process in section 3.1 so that it may serve as a tutorial for other researchers.

Section 4 contains our most important results. We provide a framework of sufficient conditions for coercivity and supercoercivity of the left and right GBDs. This framework uses the fact that the GBDs majorize a set distance. We illustrate with figures in 2 dimensions, and we provide examples of what may go wrong when the sufficient conditions provided by our framework are not satisfied. In section 4.3, we explain how these coercivity and supercoercivity conditions may be used to guarantee the coercivity of the sum of the distance together with a Legendre function.

Such sums are the basis of the corresponding envelope functions and their proximity operators. In the Bregman case, these coercivity conditions admit the further analysis of the envelopes and proximity operators, including their asymptotic behavior as the scalar parameter varies [7]. Our work lays the necessary foundation for such an analysis in the case of envelopes built from GBDs. The study of envelopes is important, because many optimization algorithms may be viewed as special cases of gradient descent applied to envelopes; see, for example, [37, 41]. We conclude in section 5.

2. Preliminaries on GBDs. Given $A \subset X$, we denote by $\overline{A}$ the closure of A with respect to the strong topology. Given a function $f: X \to \mathbb{R}_\infty$, its *domain* (or *effective domain*) is defined by $\operatorname{dom} f := \{x \in X : f(x) < \infty\}$ and its *lower level set* at height $\xi \in \mathbb{R}$ by $\operatorname{lev}_{\leq \xi} f := \{x \in X : f(x) \leq \xi\}$. The function f is said to be *proper* if $\operatorname{dom} f \neq \varnothing$; *lower semicontinuous (lsc)* at $\bar{x}$ if $f(\bar{x}) \leq \liminf_{x \to \bar{x}} f(x)$; *convex* if

$$(5) \qquad \forall x, y \in X, \ \forall \lambda \in [0,1], \quad f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y);$$

coercive if $\lim_{\|x\| \to \infty} f(x) = \infty$; and *supercoercive* if $\lim_{\|x\| \to \infty} f(x)/\|x\| = \infty$.

Let $f: X \to \mathbb{R}_\infty$ be proper. The *subdifferential* of f is the point-to-set mapping $\partial f : X \rightrightarrows X^*$ defined by

$$(6) \qquad \partial f(x) := \begin{cases} \{v \in X^* : \forall y \in X, \ \langle y - x, v \rangle + f(x) \leq f(y)\} & \text{if } x \in \operatorname{dom} f, \\ \varnothing & \text{otherwise.} \end{cases}$$

The *Fenchel conjugate* of f is the mapping

$$(7) \qquad\qquad f^*: X^* \to \mathbb{R}_\infty: v \mapsto \sup_{x \in X}\{\langle x, v \rangle - f(x)\}.$$

From the definition, we have the *Fenchel–Young inequality*

$$(8) \qquad \forall (x,v) \in X \times X^*, \quad f(x) + f^*(v) \geq \langle x, v \rangle,$$

and if f is convex, then

$$(9) \qquad f(x) + f^*(v) = \langle x, v \rangle \iff v \in \partial f(x).$$

Given a point-to-set operator $T \colon X \rightrightarrows X^*$, its *domain* is $\operatorname{dom} T := \{x \in X : Tx \neq \varnothing\}$, its *range* is $\operatorname{ran} T := T(X)$, and its *graph* is $\mathcal{G}(T) := \{(x, x^*) \in X \times X^* : x^* \in Tx\}$. Additionally, T is said to be maximally monotone if

$$(10) \qquad (x,u) \in \mathcal{G}(T) \iff (\forall (y,v) \in \mathcal{G}(T)) \quad \langle x - y, u - v \rangle \geq 0.$$

A detailed study of maximally monotone operators can be found in [5, Chapters 20 and 21] for Hilbert spaces and in [19, Chapter 4] for the Banach space case.

2.1. Representative functions. Let $S \colon X \rightrightarrows X^*$ be a maximally monotone operator. We recall from [21, Definition 2.3] that $h \colon X \times X^* \to \mathbb{R}_\infty$ *represents* S and denote $h \in \mathcal{H}(S)$ if the following three conditions hold:
   (a) h is convex and norm $\times$ weak* lsc in $X \times X^*$.
   (b) $\forall (x,v) \in X \times X^*$, $h(x,v) \geq \langle x, v \rangle$.
   (c) $h(x,v) = \langle x, v \rangle \iff (x,v) \in \mathcal{G}(S)$.
   We will make use, in particular, of several representative functions. These are as follows.
   (i) The Fitzpatrick function $F_S \colon (x,y) \mapsto \sup_{(z,w) \in \mathcal{G}(S)} (\langle z - x, y - w \rangle + \langle x, y \rangle)$ is the smallest member of $\mathcal{H}(S)$; see Fitzpatrick's 1998 paper [34, Theorem 3.7].
   (ii) The largest member of $\mathcal{H}(S)$ we denote by σ_S; it may be computed using the identities in Fact 2.1(iii).
   (iii) The Fenchel–Young representative $f \oplus f^* \in \mathcal{H}(\partial f)$, where $f \in \Gamma_0(X)$ and $f \oplus f^* \colon X \times X^* \to \mathbb{R}_\infty$ is defined by

$$(11) \qquad \forall (x,v) \in X \times X^*, \quad f \oplus f^*(x,v) := f(x) + f^*(v).$$

The Fitzpatrick function has proven quite useful in monotone operator theory; see, for example, [5, 12].

FACT 2.1. *Let $S \colon X \rightrightarrows X^*$ be a maximally monotone operator and X a real Banach space. We have the following characterizations of σ_S and F_S^*.*
   (i) *From [23, equation (33)], we have that*

$$(12) \qquad \operatorname{epi} \sigma_S = \overline{\operatorname{co}} \left(\operatorname{epi} \left(\pi + \iota_{G(S)} \right) \right), \quad \text{where } \pi \colon (p, p^*) \mapsto \langle p, p^* \rangle$$

   is the duality product defined in $X \times X^*$ and $\overline{\operatorname{co}}(A)$ is an abbreviation for the closure of the convex hull of a set A.
   (ii) *From [5, Proposition 10.56][2], we have that $F_S = \left(\iota_{G(S^{-1})} + \pi \right)^*$.*
   (iii) *From [23, equation (39)], we have that $\sigma_S(x,y) = F_S^*(y,x)$.*
   (iv) *From [23, Corollary 4.2], if h is convex and lsc on $X \times X^*$ and $F_S \leq h \leq \sigma_S$, then $h \in \mathcal{H}(S)$.*

---

[2]Note that the setting in the exposition [5] is a Hilbert space, although the veracity of (ii) in a Banach space follows from (i) and (iii). The setting of [23] is a Banach space.

The astute reader will notice that (iii) *may be obtained by combining* (i) *and* (ii)*, since*

$$\left(\iota_{\mathcal{G}(S)} + \pi\right)(y,x) = \begin{cases} \infty & \text{if } (y,x) \notin \mathcal{G}(S), \\ \langle y,x \rangle & \text{otherwise} \end{cases}$$

$$= \begin{cases} \infty & \text{if } (x,y) \notin \mathcal{G}(S^{-1}), \\ \langle x,y \rangle & \text{otherwise} \end{cases}$$

$$= \left(\iota_{\mathcal{G}(S^{-1})} + \pi\right)(x,y).$$

Additionally, (iv) *is quite pleasing, because it admits as representative functions the convex combinations of other representative functions.*

2.2. A new GBD between point-to-set operators. From now on, we assume that $S\colon X \rightrightarrows X^*$ is a maximally monotone operator, $h \in \mathcal{H}(S)$, and $T\colon X \rightrightarrows X^*$. Following [21, Definition 3.1], for fixed $(x,y) \in \operatorname{dom} S \times \operatorname{dom} T$, we define

$$(13\text{a}) \qquad \mathcal{D}_T^{\flat,h}(x,y) := \inf_{v \in Ty} \left(h(x,v) - \langle x,v \rangle\right),$$

$$(13\text{b}) \qquad \mathcal{D}_T^{\sharp,h}(x,y) := \sup_{v \in Ty} \left(h(x,v) - \langle x,v \rangle\right).$$

If $y \notin \operatorname{dom} T$, then $\mathcal{D}_T^{\flat,h}(x,y) = \mathcal{D}_T^{\sharp,h}(x,y) := \infty$ for every $x \in X$. If $x \notin \operatorname{dom} S$, then $\mathcal{D}_T^{\flat,h}(x,y) = \mathcal{D}_T^{\sharp,h}(x,y) := \infty$ for every $y \in X$. When T is point-to-point, we simply write $\mathcal{D}_T^h := \mathcal{D}_T^{\flat,h} = \mathcal{D}_T^{\sharp,h}$.

For our examples, $T = S = \partial f$ is point-to-point on $\operatorname{int} \operatorname{dom} f$ in which case we simply write $\mathcal{D}_h$. Additionally, when employing a specific representative function, we will use the name of the representative function used in place of h. If a distance is of the form (13b) or (13a) we call it a *GBD*. The GBD specializes to the Bregman distance under certain circumstances, which we now recall. To a proper and convex function $f\colon X \to \mathbb{R}_\infty$, we associate two Bregman distances (see [35]) defined by

$$(14\text{a}) \qquad \mathcal{D}_f^{\flat}(x,y) := f(x) - f(y) + \inf_{v \in \partial f(y)} \langle y - x, v \rangle$$

$$(14\text{b}) \qquad \text{and} \quad \mathcal{D}_f^{\sharp}(x,y) := f(x) - f(y) + \sup_{v \in \partial f(y)} \langle y - x, v \rangle.$$

Burachik and Martínez-Legaz observed that the GBD specializes to the Bregman distance in the case where the Fenchel–Young representative is used as h. The following proposition fills a minor omission from [21, Proposition 3.5], namely, that the Fitzpatrick distance specializes to the Bregman distance under the mild condition that $(x,y) \notin (\operatorname{dom} f \setminus \operatorname{dom} \partial f) \times \operatorname{dom} \partial f$. In the case when $\operatorname{dom} f \setminus \operatorname{dom} \partial f = \varnothing$, they are everywhere equal. We will see later in an example that when f is the Boltzmann–Shannon entropy (21), we have $\operatorname{dom} f \setminus \operatorname{dom} \partial f = \{0\}$, and the two distances fail to be equal on the set $\{(0,y) \mid y > 0\}$.

PROPOSITION 2.2 (the GBD generalizes the Bregman distance). *Let* $f \in \Gamma_0(X)$. *Then*

$$(15) \qquad \mathcal{D}_{f \oplus f^*}^{\flat}(x,y) = \mathcal{D}_f^{\flat}(x,y) \quad \text{and} \quad \mathcal{D}_{f \oplus f^*}^{\sharp}(x,y) = \mathcal{D}_f^{\sharp}(x,y)$$

whenever $(x,y) \notin (\operatorname{dom} f \setminus \operatorname{dom} \partial f) \times \operatorname{dom} \partial f$.

Proof. If $y \notin \operatorname{dom} \partial f$, then $\mathcal{D}_{f \oplus f^*}^\flat(x, y) = \mathcal{D}_f^\flat(x, y) = \infty$. If $y \in \operatorname{dom} \partial f$ and $x \notin \operatorname{dom} f$, we also have that $\mathcal{D}_{f \oplus f^*}^\flat(x, y) = \mathcal{D}_f^\flat(x, y) = \infty$.

It suffices to assume that $y \in \operatorname{dom} \partial f$ and $x \in \operatorname{dom} \partial f \subseteq \operatorname{dom} f$. Then, since $f(y) + f^*(v) = \langle y, v \rangle$ for all $v \in \partial f(y)$, we derive that

$$\text{(16a)} \qquad \mathcal{D}_{f \oplus f^*}^\flat(x, y) = \inf_{v \in \partial f(y)} (f(x) + f^*(v) - \langle x, v \rangle)$$

$$\text{(16b)} \qquad = \inf_{v \in \partial f(y)} (f(x) - f(y) + \langle y, v \rangle - \langle x, v \rangle)$$

$$\text{(16c)} \qquad = f(x) - f(y) + \inf_{v \in \partial f(y)} \langle y - x, v \rangle = \mathcal{D}_f^\flat(x, y).$$

Similarly, $\mathcal{D}_{f \oplus f^*}^\sharp(x, y) = \mathcal{D}_f^\sharp(x, y)$. The conclusion follows. $\qquad\square$

We recall now the following results regarding the lower semicontinuity of the left and right distances; these apply to each of our computed examples.

LEMMA 2.3 (see [21, Lemma 3.17]). *Let $y \in \operatorname{dom} T$. Then the following hold:*
  (i) *The function $\mathcal{D}_T^{\flat, h}(\cdot, y) : X \to \mathbb{R}_\infty$ is lsc at every $x \in \operatorname{int} \operatorname{dom}(S)$ with respect to the strong topology in X provided that Tz is weakly closed for any z in its domain.*
  (ii) *The function $\mathcal{D}_T^{\sharp, h}(\cdot, y) : X \to \mathbb{R}_\infty$ is lsc at every $x \in \operatorname{dom}(S)$ with respect to the strong topology in X.*

LEMMA 2.4 (see [21, Lemma 3.18]). *Suppose that T is locally bounded in the interior of its domain and that the graph of T is closed with respect to the strong-weak topology. Fix $y \in \operatorname{int} \operatorname{dom} T$ and $x \in \operatorname{dom} S$. Then the function $\mathcal{D}_T^{\flat, h}(x, \cdot) : X \to \mathbb{R}_\infty$ is lsc at y with respect to the strong topology in X.*

Remark 2.5 (The lower closed distance). Notice that in Lemma 2.3(ii), $\mathcal{D}_T^{\star, h}(\cdot, y)$ may not be lsc at $x \in \overline{\operatorname{dom}} \partial f \setminus \operatorname{dom} \partial f$, a case we will encounter in our examples. Notice also that in Lemma 2.4, for $y \notin \operatorname{int} \operatorname{dom} T$, the distance may *not* be lsc with respect to the second variable, a phenomenon we will encounter in our examples.

For these two reasons, we also introduce the notion of the *lower closed GBD*, denoted by $\overline{\mathcal{D}}_T^{\star, h}$, which satisfies

$$\text{(17)} \qquad \operatorname{epi} \overline{\mathcal{D}}_T^{\star, h} = \overline{\operatorname{epi}} \mathcal{D}_T^{\star, h},$$

where $\star$ may be either $\flat$ or $\sharp$. The lower closed GBD is the lsc regularization of the function $\mathcal{D}_T^{\star, h}$, as described in [40]; its direct formula is given by

$$\overline{\mathcal{D}}_T^{\star, h}(x, y) := \liminf_{(x', y') \to (x, y)} \mathcal{D}_T^{\star, h}(x', y').$$

The lower closed distances $\overline{\mathcal{D}}_{F_S}$, $\overline{\mathcal{D}}_{\sigma_S}$, and $\overline{\mathcal{D}}_{f \oplus f^*}$ are defined analogously.

3. How to compute GBDs.

Example 3.1 (energy). Let $f : x \mapsto \frac{1}{2} x^2$ be the *energy*. If we have h_{Id} as the Fitzpatrick function for $\partial f = \mathrm{Id}$, then our GBD distance is

$$\text{(18)} \qquad \mathcal{D}_{F_{\mathrm{Id}}}(x, y) = \frac{1}{4}(x - y)^2,$$

which is equivalent to a scaled version of the usual Moreau distance. On the other hand, the largest element of $\mathcal{H}(\mathrm{Id})$ is just

$$\text{(19)} \qquad \sigma_{\mathrm{Id}}(x, y) = \begin{cases} x^2 & x = y, \\ \infty & \text{otherwise.} \end{cases}$$

One can obtain this result by computing F_{Id}^* straight from the definition of the conjugate and using Fact 2.1(iii). One can also obtain this result by using Fact 2.1(i), because the graph of Id is simply the diagonal. The corresponding distance is

$$(20) \qquad \mathcal{D}_{\sigma_{\mathrm{Id}}}(x, y) = \begin{cases} 0 & x = y, \\ \infty & \text{otherwise.} \end{cases}$$

In [7], the asymptotic properties of Bregman envelopes are illustrated using Bregman distances constructed from three functions. One of these was the *energy* from Example 3.1, for which the Bregman proximity operator and envelope specialize to the Moreau case. While the choice of representative function F_{Id} is equivalent to the Moreau case up to a change in parameter, notice that the example $\mathcal{D}_{\sigma_{\mathrm{Id}}}$ illustrates that this is *not* the case for *any* choice of representative function. This is an important distinction in our context.

3.1. Boltzmann–Shannon entropy. Another function whose translated version was early considered by Censor and Lent [27], and whose Bregman envelopes are studied in [7], is the (negative) Boltzmann–Shannon entropy:

$$(21) \qquad \mathrm{ent} : \mathbb{R} \to \mathbb{R} : \quad x \mapsto \begin{cases} x \log x - x & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ \infty & \text{otherwise.} \end{cases}$$

The Boltzmann–Shannon entropy is particularly important and natural to consider, because its derivative is log, its conjugate is $\mathrm{ent}^* = \exp$, and its associated Bregman distance is the Kullback–Leibler divergence,

$$(22) \qquad \mathcal{D}_{\mathrm{ent}} : (x, y) \mapsto \begin{cases} x(\log(x) - \log(y)) - x + y & \text{if } y > 0, \\ y & \text{if } y > 0 \text{ and } x = 0, \\ \infty & \text{otherwise,} \end{cases}$$

which is frequently used as a measure of distance between positive vectors in information theory, statistics, and portfolio selection. The GBD associated with the Fenchel–Young representative $\mathrm{ent} \oplus \mathrm{ent}^* \in \mathcal{H}(\log)$ is

$$(23) \qquad \mathcal{D}_{\mathrm{ent} \oplus \mathrm{ent}^*} : (x, y) \mapsto \begin{cases} x(\log(x) - \log(y)) - x + y & \text{if } x, y > 0, \\ \infty & \text{otherwise.} \end{cases}$$

Thus it may be seen that the *Bregman* distance of the *Boltzmann–Shannon entropy* is the special case of the *GBD* for the Fenchel–Young representative of the *logarithm* function, except on the set $(\mathrm{dom}\, f \setminus \mathrm{dom}\, \partial f) \times \mathrm{dom}\, \partial f = \{0\} \times]0, \infty[$ (see Proposition 2.2 and Remark 2.5). Its lower closure is given by[3]

$$(24) \qquad \overline{\mathcal{D}}_{\mathrm{ent} \oplus \mathrm{ent}^*} : (x, y) \mapsto \begin{cases} x(\log(x) - \log(y)) - x + y & \text{if } y > 0, x > 0, \\ y & \text{if } y > 0, x = 0, \\ 0 & \text{if } y = x = 0, \\ \infty & \text{otherwise} \end{cases}$$

and is shown in Figure 1(b), while the Fenchel–Young representative $\mathrm{ent} \oplus \mathrm{ent}^*$ is shown in Figure 2(b).

---

[3]One may rewrite the first case as $y > 0, x \geq 0$ and omit writing the separate case $x = 0, y > 0$ in (24), as long as one remembers to use the convention that $0 \log(0) = 0$.
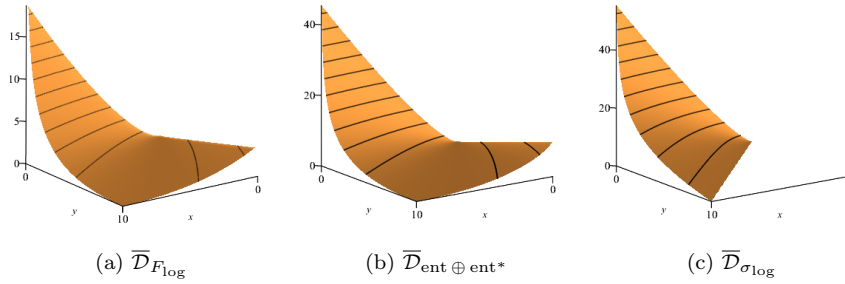
(a) $\overline{\mathcal{D}}_{F_{\log}}$ (b) $\overline{\mathcal{D}}_{\mathrm{ent} \oplus \mathrm{ent}^*}$ (c) $\overline{\mathcal{D}}_{\sigma_{\log}}$

FIG. 1. *Distances constructed from $\mathcal{H}(\log)$.*

(a) $F_{\log}$ (b) $\mathrm{ent} \oplus \mathrm{ent}^*$ (c) $\sigma_{\log}$
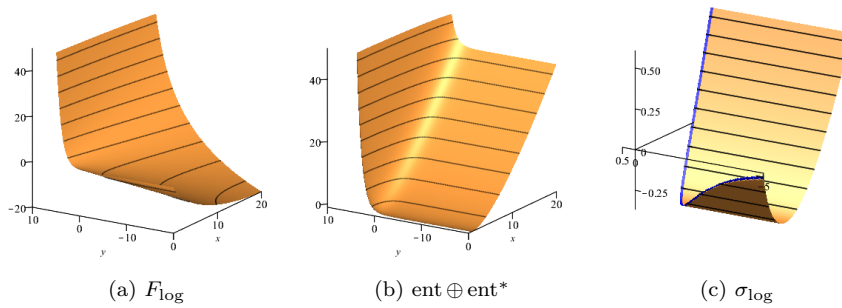
FIG. 2. *Constructing representative functions for the logarithm.*

We will consider *new* distances built from the maximally monotone operator log and compare these to the known special case of the Bregman distance for the Boltzmann-Shannon entropy. The corresponding Fitzpatrick function (as computed in [9]) and shown in Figure 2(a) is

$$(25) \qquad F_{\log} : (x,y) \mapsto \begin{cases} \infty & \text{if } x < 0, \\ \exp(y-1) & \text{if } x = 0, \\ xy + x \left(\mathcal{W}(xe^{1-y}) + \frac{1}{\mathcal{W}(xe^{1-y})} - 2 \right) & \text{if } x > 0, \end{cases}$$

where $\mathcal{W}$ is the real principal branch of the Lambert $\mathcal{W}$ function that satisfies $\mathcal{W}(x)e^{W(x)} = x$ on $[-1/e, \infty[$. See, for example, [31]. Its occurrences in convex analysis and its relationship to the Boltzmann–Shannon entropy have been discussed in, for example, [8, 10, 11].

Example 3.2 (GBD $\overline{\mathcal{D}}_{F_{\log}}$). The corresponding (closed) GBD is

$$(26) \quad \overline{\mathcal{D}}_{F_{\log}} : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+ \cup \{\infty\}$$

$$(x,y) \mapsto \begin{cases} \infty & \text{if } x < 0 \text{ or } y < 0 \text{ or } (x > 0 \text{ and } y = 0), \\ ye^{-1} & \text{if } x = 0 \text{ and } y \geq 0, \\ x \left(\mathcal{W}\left(\frac{xe}{y}\right) + \frac{1}{\mathcal{W}\left(\frac{xe}{y}\right)} - 2 \right) & \text{otherwise.} \end{cases}$$

This distance is shown in Figure 1(a).

Proof. Combining Definitions 13 and 26 with the fact that $\operatorname{dom} \log = \,]0, \infty[$, we have

(27a)
$$\mathcal{D}_{F_{\log}}(x, y) = \begin{cases} F_{\log}(x, \log(y)) - \langle x, \log(y) \rangle & \text{if } x, y \in \operatorname{dom} \log, \\ \infty & \text{otherwise} \end{cases}$$

(27b)
$$= \begin{cases} \infty & \text{if } x \le 0 \text{ or } y \le 0, \\ x \log(y) + x \left(\mathcal{W}(x e^{1 - \log(y)}) + \frac{1}{\mathcal{W}(x e^{1 - \log(y)})} - 2 \right) - x \log(y) & \text{otherwise.} \end{cases}$$

This simplifies, by a bit of arithmetic, to the form in (26), except on the set $\{0\} \times [0, \infty[$. Taking the closure of the epigraph admits $\overline{\mathcal{D}}_{F_{\log}}(0, y) = y e^{-1}$. This example is particularly interesting, because we see the loss of the *left* lower semicontinuity property at 0 because $0 \in \operatorname{dom} f \setminus \operatorname{dom} \log$, and we also see the loss of the *right* lower semicontinuity property because $0 \notin \operatorname{int} \operatorname{dom} \log$; see Remark 2.5. □

3.2. Computation of a difficult representative function and the conjugate of a Fitzpatrick function. Next we consider the case where $S = \log$ is the logarithm function on $]0, \infty[$. Even though we know the form of $F_{\log}$, it is not straightforward to compute $\sigma_{\log}$ using the equality $\sigma_{\log}(x, y) = F_{\log}^*(y, x)$ from Fact 2.1(iii) by subdifferentiating with the latter and solving. Instead, we use the characterization from Fact 2.1(i).

Recall that, for an arbitrary function g and its *convex hull* function $\operatorname{co}(g)$, the lsc regularization or *lower closure*, denoted as $\overline{\operatorname{co}}(g)$, has the property

(28)
$$\operatorname{epi}(\overline{\operatorname{co}}(g)) = \overline{\operatorname{co}}(\operatorname{epi}(g)).$$

See, for example, [40, Chapter 1].

THEOREM 3.3 (The representative $\sigma_{\log}$). *Let $T = S = \log$. Then*

(29)
$$\sigma_{\log} : (x, y) \mapsto \begin{cases} x \log(x) & \text{if } y \le \log(x), \\ \infty & \text{otherwise,} \end{cases}$$

whose graph is shown in Figure 2(c).

Proof. Using Fact 2.1 together with the fact that $\mathcal{G}(S) = \{(z_1, \log(z_1)) | z_1 \in \,]0, \infty[\}$ and the fact that $\log$ is a concave function, we have that $z_2 > \log(z_1)$ implies $\sigma_{\partial f}(z_1, z_2) = \infty$. Indeed, let $g : \mathbb{R}^2 \to \mathbb{R}_\infty$ be defined as $g := \pi + \iota_{\mathcal{G}(S)}$; the graph of g is shown as the dark curve at the boundary of the surface in Figure 2(c). By Fact 2.1 we have that $\sigma_{\partial f} = \overline{\operatorname{co}}(g)$. If $\overline{\operatorname{co}}(g)(z_1, z_2) < \infty$, then there exists $a \in \mathbb{R}$ such that $(z_1, z_2, a) \in \operatorname{epi} \overline{\operatorname{co}}(g)$. By (28) this is equivalent to $(z_1, z_2, a) \in \overline{\operatorname{co}} \operatorname{epi}(g)$. The last inclusion means that there exists a sequence $w_n := (z_1^n, z_2^n, a_n) \in \operatorname{conv} \operatorname{epi}(g)$ such that $(z_1, z_2, a) = \lim_{n \to \infty}(z_1^n, z_2^n, a_n)$. Note that we can assume that

$$(z_1^n, z_2^n, a_n) = \sum_{i=1}^{4} \lambda_{n,i}(z_{1,i}^n, z_{2,i}^n, a_{i,n}), \text{ with } (z_{1,i}^n, z_{2,i}^n, a_{i,n}) \in \operatorname{epi}(g),$$

thanks to Carathéodory's theorem. Using the fact that $(z_{1,i}^n, z_{2,i}^n, a_{i,n}) \in \operatorname{epi}(g)$, we have that

$$\textstyle\sum_{i=1}^{4} \lambda_{n,i} = 1, \ \lambda_{n,i} \geq 0 \quad \forall i = 1, \ldots, 4,$$

$$z_1^n = \textstyle\sum_{i=1}^{4} \lambda_{n,i} z_{1,i}^n, \qquad z_2^n = \textstyle\sum_{i=1}^{4} \lambda_{n,i} z_{2,i}^n,$$

$$z_{1,i}^n > 0, \qquad\qquad z_{2,i}^n = \log z_{1,i}^n \ \forall i = 1, \ldots, 4,$$

Using the above expression for z_2^n and the fact that $z_{2,i}^n = \log z_{1,i}^n$, we can write

$$z_2^n = \sum_{i=1}^{4} \lambda_{n,i} z_{2,i}^n = \sum_{i=1}^{4} \lambda_{n,i} \log z_{1,i}^n \leq \log\left(\sum_{i=1}^{4} \lambda_{n,i} z_{1,i}^n\right) = \log z_1^n,$$

where we used the fact that $\log(\cdot)$ is concave. Taking limits and using the continuity of the $\log(\cdot)$ we deduce that $z_2 \leq \log(z_1)$. This implies that, when $z_2 > \log(z_1)$ we must have $\sigma_{\partial f}(z_1, z_2) = \infty$. This shows the second part of the definition in the statement of the theorem. We proceed now to prove the first part of the definition of $\sigma_{\partial f}$. Let $z \in \mathbb{R}^2$ be such that $z_1 > 0$ and $z_2 \leq \log(z_1)$. For any $x, y \in \mathcal{G}(S)$ that satisfy $\lambda x + (1 - \lambda)y = z$ for some $\lambda \in [0, 1]$, we have that $g(x) = \langle x_1, x_2 \rangle = x_1 x_2$ and that $g(y) = \langle y_1, y_2 \rangle = y_1 y_2$. Thus

(30a) $$\qquad\qquad (x_1, x_2, \langle x_1 x_2 \rangle) = (x_1, x_2, x_1 x_2) \in \operatorname{epi} g$$

(30b) $$\qquad\text{and}\quad (y_1, y_2, \langle y_1 y_2 \rangle) = (y_1, y_2, y_1 y_2) \in \operatorname{epi} g.$$

From the definition of convexity,

(31) $$\qquad\qquad \lambda(x_1, x_2, x_1 x_2) + (1 - \lambda)(y_1, y_2, y_1 y_2) \in \operatorname{conv}(\operatorname{epi} g).$$

This is just

(32) $$\qquad\qquad (z_1, z_2, \lambda x_1 x_2 + (1 - \lambda) y_1 y_2) \in \operatorname{conv}(\operatorname{epi} g).$$

Using the fact that $x, y \in \mathcal{G}(S)$, this is just

(33) $$\qquad\qquad (z_1, z_2, \lambda x_1 \log(x_1) + (1 - \lambda) y_1 \log(y_1)) \in \operatorname{conv}(\operatorname{epi} g).$$

Let $(\varphi_n)_{n \in \mathbb{N}} \subset \,]0, \pi/2[$ be a sequence that satisfies $\lim_{n \to \infty} \varphi_n = \pi/2$. For any φ_n, we may find a line in $\mathbb{R}^2$ that goes through z and has slope $\tan(\varphi_n)$, which is given by

(34) $$\qquad\qquad L_n := \{u \in \mathbb{R}^2 \mid u_2 = \tan(\varphi_n)(u_1 - z_1) + z_2\}.$$

Now $\varphi_n \in \,]0, \pi/2[$ and $z_2 \leq \log(z_1)$ guarantees that $L_n \cap \mathcal{G}(S)$ is a doubleton $\{x^n, y^n\}$ where $x_1^n < z_1$ and $y_1^n > z_1$ and $z = p_n x^n + q_n y^n$ with $q_n + p_n = 1, p_n, q_n \in [0, 1]$. The construction of this sequence is shown in Figure 3. As $\varphi_n \to \pi/2$, the slope $\tan(\varphi_n)$ of L_n goes to infinity, and so we have that

(35a) $$\qquad\qquad \lim_{n \to \infty} y^n = (z_1, \log(z_1)) \quad\text{and}\quad \lim_{n \to \infty} x^n = (0, -\infty),$$

(35b) $$\qquad\text{and so}\quad \lim_{n \to \infty} x^n \log(x^n) = 0 \quad\text{and}\quad \lim_{n \to \infty} p_n = 0 \quad\text{and}\quad \lim_{n \to \infty} q_n = 1.$$

Thus we have that

(36)
$$\lim_{n \to \infty} (z_1, z_2, p_n x^n \log(x^n) + q_n y^n \log(y^n)) = (z_1, z_2, z_1 \log(z_1)) \in \overline{\operatorname{co}}(\operatorname{epi} g) = \operatorname{epi} \sigma_{\log}.$$

Thus $z_1 \log(z_1) \geq \sigma_{\partial f}(z)$ for every (z_1, z_2) such that $z_1 > 0$ and $z_2 \leq \log(z_1)$. For the converse inequality, define the function

$$(37) \qquad w(t_1, t_2) := \begin{cases} t_1 \log t_1 & \text{if } t_1 > 0, \\ 0 & \text{if } t_1 = 0. \end{cases}$$

The function w is convex and lsc. It is easy to check that $w \leq g$ in $\mathbb{R}^2$. Therefore,

$$(38) \qquad \operatorname{epi} w \supset \operatorname{epi} g.$$

Using the fact that w is convex and lsc we deduce that

$$(39) \qquad \operatorname{epi} w = \overline{\operatorname{co}}(\operatorname{epi} w) \supset \overline{\operatorname{co}}(\operatorname{epi} g) = \operatorname{epi} \sigma_{\log};$$

equivalently, $w \leq \sigma_{\log}$. This implies that $z_1 \log(z_1) \leq \sigma_{\log}(z)$ for every (z_1, z_2) such that $z_1 > 0$ and $z_2 \leq \log(z_1)$. Consequently, we showed that

$$(40) \qquad \sigma_{\log}(z) = z_1 \log(z_1) \ \forall (z_1, z_2) \text{ such that } z_1 > 0 \text{ and } z_2 \leq \log(z_1),$$

which is the claim of the theorem. $\qquad\square$
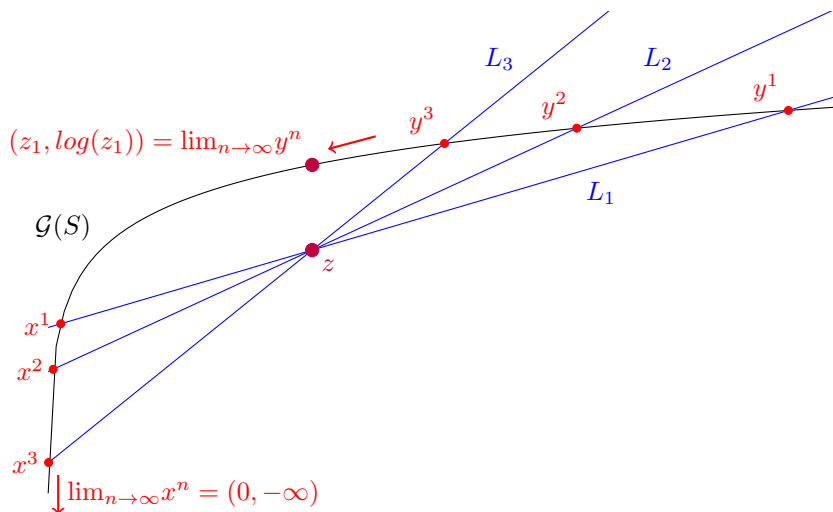

FIG. 3. *Construction of sequence in proof of Theorem* 3.3.

COROLLARY 3.4 (the conjugate of the Fitzpatrick function $F_{\log}$). *We have that*

$$(41) \qquad F_{\log}^* : (x, y) \mapsto \begin{cases} y \log(y) & \text{if } x \leq \log(y), \\ \infty & \text{otherwise.} \end{cases}$$

Proof. This immediately follows from Theorem 3.3 together with Fact 2.1(iii). $\square$

Remark 3.5. In the proof of Theorem 3.3, the sequences x^n, y^n may be given explicitly by

$$(42a) \qquad x_1^n = \exp\left(-\mathcal{W}_0\left(-\tan(\varphi_n)\exp\left(-z_1\tan(\varphi_n) + z_2\right)\right)\right),$$
$$(42b) \qquad y_1^n = \exp\left(-\mathcal{W}_{-1}\left(-\tan(\varphi_n)\exp\left(-z_1\tan(\varphi_n) + z_2\right)\right)\right),$$

where $\mathcal{W}_0$ and $\mathcal{W}_{-1}$ are the principal and secondary real branches of the Lambert $\mathcal{W}$ function.

Most of the analysis of Lambert $\mathcal{W}$ in the context of convex optimization has focused on its principal branch. However, in order to experimentally discover the true form for $\sigma_{\partial f}$ from Theorem 3.3, we had to make use of both real branches. The reason for this is that our attempts to explicitly solve the systems

(43a)
$$\sup_{y \in \mathbb{R}^2} \{\langle x, y \rangle - F_{\partial f}(y)\}$$

(43b)
$$\text{or} \quad \inf_{\substack{z = \lambda x + (1-\lambda)y \\ x, y \in G(\partial f), \ \lambda \in [0,1]}} \{\lambda x_1 x_2 + (1 - \lambda) y_1 y_2\}$$

were not successful. Seeking to compute $\sigma_{\log}$ numerically, we constructed a numerical procedure that evaluated $\lambda x_1^n x_2^n + (1-\lambda) y_1^n y_2^n$ for a finite sequence $(\varphi_n)_{n=1}^N$ and chose the smallest value to represent $\sigma_{\log}(z)$. The fast evaluation of Lambert $\mathcal{W}$ obviated the implementation of slower numerical routines to solve the equation system

(44)
$$\log(\eta) = \tan(\varphi_n)(\eta - z_1) + z_2.$$

We observed that the smallest value was always the last value, corresponding to φ_n nearest to $\pi/2$. Once we observed that the values were consistently approaching $z_1 \log(z_1)$ for any z chosen, we "knew" the true form. Upon further scrutiny of the geometry, we realized that the sequences that led to the discovery also yielded the proof.

COROLLARY 3.6 (GBD for $\sigma_{\log}$). *The corresponding (closed) conjugate Fitzpatrick distance is*

(45)
$$\overline{\mathcal{D}}_{\sigma_{\log}} : (x, y) \mapsto \begin{cases} x \log(x) - x \log(y) & \text{if } 0 < y \leq x, \\ 0 & \text{if } x = y = 0, \\ \infty & \text{otherwise,} \end{cases}$$

which is shown in Figure 1(c).

Proof. From the definition and Theorem 3.3,

(46a)
$$\mathcal{D}_{\sigma_{\log}} : (x, y) \mapsto \sigma_{\log}(x, \log(y)) - \langle x, \log(y) \rangle$$

(46b)
$$= -x \log(y) + \begin{cases} x \log(x) & \text{if } \log(y) \leq \log(x), \\ \infty & \text{otherwise,} \end{cases}$$

which may be recognized as the form in (45) except at the point $(0, 0)$. Taking the closure of the epigraph of $\mathcal{D}_{\sigma_{\log}}$, we obtain $\overline{\mathcal{D}}_{\sigma_{\log}}(0, 0) = 0$. This example is illustrative because lower semicontinuity is lost, but only at the point $(0, 0)$, since $0 \notin \text{dom} \log$; see Remark 2.5. $\square$

4. A coercivity framework for the GBD. In this section, we will establish important properties of the GBD. From now on, X is a *reflexive* real Banach space. For the sake of simplicity, when we make use of the norms $\| \cdot \|_{X \oplus X^*}$, $\| \cdot \|_X$, and $\| \cdot \|_{X^*}$, we allow context to make clear which norm is being used.

4.1. Convexity.

PROPOSITION 4.1. *Let $x \in \text{dom} \, S$ and $y \in \text{dom} \, T$. Then the following hold:*
 (i) $\mathcal{D}_T^{\sharp, h}(\cdot, y)$ *is convex.*
 (ii) *If T is an affine mapping, then $\mathcal{D}_T^h(x, \cdot)$ is convex.*

Proof. We first note that h is convex on $X \times X^*$ by definition.

(i): As $h(\cdot, v) - \langle \cdot, v \rangle$ is convex for all $v \in Ty$, it follows from [19, Proposition 3.4.3(iii)] that $\mathcal{D}_T^{\sharp,h}(\cdot, y)$ is convex.

(ii): Since T is single-valued, we have for all $y \in \operatorname{dom} T$ that

$$(47) \qquad \mathcal{D}_T^h(x, y) = \mathcal{D}_T^{\flat,h}(x, y) = \mathcal{D}_T^{\sharp,h}(x, y) = h(x, Ty) - \langle x, Ty \rangle.$$

Therefore, $\mathcal{D}_T^h(x, \cdot) = (h(x, \cdot) - \langle x, \cdot \rangle) \circ T$ is convex because it is the composition of a convex function with an affine function; see, e.g., [13, Lemma 2.1.8(b)]. □

4.2. Coercivity and supercoercivity. We now turn our attention to coercivity and supercoercivity. These properties of distances are important, because they are essential to the analysis of associated envelopes and proximity operators. After first providing a framework for verifying these properties of the GBDs, we will show in section 4.3 how these properties admit corresponding coercivity properties for the sum of the GBDs together with Legendre functions. These results on sums are the key to analysing the envelopes; see [6, Lemma 2.12] and [7].

From now on, as mentioned, we assume our spaces to be reflexive so that we may make use of the following fact from [21, Remark 3.12].

FACT 4.2 (see [21, Remark 3.12]). *When X is a reflexive space, it holds that*

$$(48) \qquad \forall x, y \in X, \quad \mathcal{D}_T^{\flat,h}(x, y) \geq \frac{1}{4} \inf_{v \in Ty} d^2\left((x, v), \mathcal{G}(S)\right) = \frac{1}{4} d^2\left(\{x\} \times Ty, \mathcal{G}(S)\right),$$

where d denotes the distance on $X \times X^*$ defined by

$$d((x, v), (y, w)) := \sqrt{\|x - y\|^2 + \|v - w\|^2}.$$

Consequently, we can see $\mathcal{D}_T^{\flat,h}(x, y)$ as providing us with an upper estimate of the distance between the sets $\{x\} \times Ty$ and $\mathcal{G}(S)$.

Throughout this section, we exploit the fact that the GBD is minorized by the distance between the sets $\{x\} \times Ty$ and $\mathcal{G}(S)$ in order to establish left and right coercivity and supercoercivity of the distance. The intuition behind the results is shown in Figure 4.

The following elementary lemma will be useful for our analysis.

LEMMA 4.3. *Let $(x_n)_{n \in \mathbb{N}}$, $(y_n)_{n \in \mathbb{N}}$, and $(z_n)_{n \in \mathbb{N}}$ be sequences in X such that $\|x_n\| \to \infty$ as $n \to \infty$. Then the following hold:*

(i) *Suppose that $\|z_{k_n}\| \to \infty$ whenever $(y_{k_n})_{n \in \mathbb{N}}$ is a subsequence of $(y_n)_{n \in \mathbb{N}}$ with $\|y_{k_n}\| \to \infty$. Then, for all $\alpha \in \mathbb{R}_{++}$,*

$$(49) \qquad\qquad \|x_n - y_n\|^\alpha + \|z_n\|^\alpha \to \infty \quad as\ n \to \infty.$$

(ii) *Suppose that $\|z_{k_n}\|^2 / \|y_{k_n}\| \to \infty$ whenever $(y_{k_n})_{n \in \mathbb{N}}$ is a subsequence of $(y_n)_{n \in \mathbb{N}}$ with $\|y_{k_n}\| \to \infty$. Then*

$$(50) \qquad\qquad \frac{\|x_n - y_n\|^2 + \|z_n\|^2}{\|x_n\|} \to \infty \quad as\ n \to \infty.$$

Proof. (i): Suppose, to the contrary, that there exist subsequences $(x_{k_n})_{n \in \mathbb{N}}$, $(y_{k_n})_{n \in \mathbb{N}}$, and $(z_{k_n})_{n \in \mathbb{N}}$ such that the sequence $(\|x_{k_n} - y_{k_n}\|^\alpha + \|z_{k_n}\|^\alpha)_{n \in \mathbb{N}}$ is bounded. Then both $(x_{k_n} - y_{k_n})_{n \in \mathbb{N}}$ and $(z_{k_n})_{n \in \mathbb{N}}$ are bounded. By assumption, passing to another subsequence if necessary, we obtain that the sequence $(y_{k_n})_{n \in \mathbb{N}}$ is also bounded, and so is $(x_{k_n})_{n \in \mathbb{N}}$ since
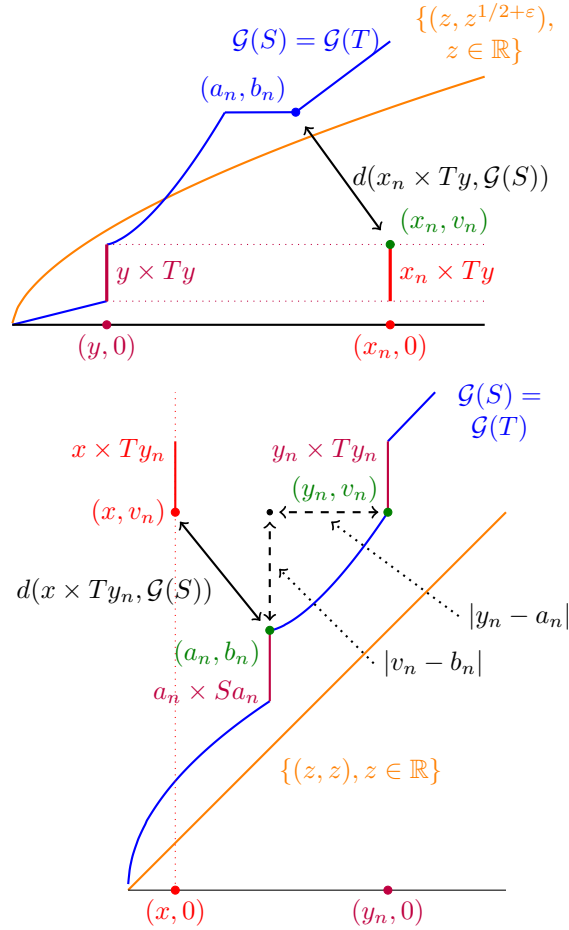
FIG. 4. *Top: the motivation for Theorem* 4.5 *is exemplified by Example* 4.9. *Bottom: the motivation for Theorem* 4.6 *is exemplified by Example* 4.11.

$$(51) \qquad \forall n \in \mathbb{N}, \quad \|x_{k_n}\| \leq \|x_{k_n} - y_{k_n}\| + \|y_{k_n}\|.$$

This contradicts the assumption that $\|x_n\| \to \infty$.

(ii): Suppose that there exist subsequences $(x_{k_n})_{n \in \mathbb{N}}$, $(y_{k_n})_{n \in \mathbb{N}}$, $(z_{k_n})_{n \in \mathbb{N}}$ and a constant $\mu > 0$ such that

$$(52) \qquad \forall n \in \mathbb{N}, \quad \frac{\|x_{k_n} - y_{k_n}\|^2 + \|z_{k_n}\|^2}{\|x_{k_n}\|} < \mu.$$

Then, by the Cauchy–Schwarz inequality,

$$(53\text{a}) \qquad 2\|y_{k_n}\| \geq \frac{2\langle x_{k_n}, y_{k_n}\rangle}{\|x_{k_n}\|} = \frac{\|x_{k_n}\|^2 + \|y_{k_n}\|^2 - \|x_{k_n} - y_{k_n}\|^2}{\|x_{k_n}\|}$$

$$(53\text{b}) \qquad \geq \|x_{k_n}\| - \frac{\|x_{k_n} - y_{k_n}\|^2}{\|x_{k_n}\|} > \|x_{k_n}\| - \mu.$$

As $n \to \infty$, since $\|x_n\| \to \infty$, it follows from (53) that $\|y_{k_n}\| \to \infty$ and, by assumption, $\|z_{k_n}\|^2/\|y_{k_n}\| \to \infty$. On the other hand, combining (52) with (53) yields

$$(54) \qquad \frac{\|z_{k_n}\|^2}{\|y_{k_n}\|} < \mu \frac{\|x_{k_n}\|}{\|y_{k_n}\|} \leq \mu \frac{2\|y_{k_n}\| + \mu}{\|y_{k_n}\|} = 2\mu + \frac{\mu^2}{\|y_{k_n}\|} \to 2\mu.$$

A contradiction is thus obtained, and we complete the proof. $\qquad\square$

Remark 4.4. (i) Since $\mathcal{D}_T^{\sharp,h}(x,y) \geq \mathcal{D}_T^{\flat,h}(x,y)$ for all $(x,y) \in \operatorname{dom} S \times \operatorname{dom} T$, if $\mathcal{D}_T^{\flat,h}$ is coercive or supercoercive with respect to the first or second variable, then so is $\mathcal{D}_T^{\sharp,h}$. We will thus focus on the coercivity and supercoercivity of $\mathcal{D}_T^{\flat,h}$.

(ii) Assume that Ty is compact with respect to the strong topology. Then $\{x\} \times Ty$ is also compact. This, together with Fact 4.2 and the fact that $\mathcal{G}(S)$ is closed, allows us to choose $v \in Ty$ and $(a,b) \in \mathcal{G}(S)$ such that

$$(55)$$
$$\|x - a\|^2 + \|v - b\|^2 = d^2\big((x,v),(a,b)\big) = d^2(\{x\} \times Ty, \mathcal{G}(S)) \leq 4\mathcal{D}_T^{\flat,h}(x,y).$$

THEOREM 4.5 (left coercivity and left supercoercivity of $\mathcal{D}_T^{\flat,h}$). *Let $y \in \operatorname{dom} T$. Then*

(i) *If $\operatorname{dom} S$ is bounded, then $\mathcal{D}_T^{\flat,h}(\cdot,y)$ is supercoercive and hence coercive.*

Suppose further that Ty is compact with respect to the strong topology. Then the following hold:

(ii) *If S is coercive in the sense that $(a_n,b_n) \in \mathcal{G}(S)$ and $\|a_n\| \to \infty$ imply $\|b_n\| \to \infty$, then $\mathcal{D}_T^{\flat,h}(\cdot,y)$ is coercive.*

(iii) *If $(a_n,b_n) \in \mathcal{G}(S)$ and $\|a_n\| \to \infty$ imply $\|b_n\|^2/\|a_n\| \to \infty$, then $\mathcal{D}_T^{\flat,h}(\cdot,y)$ is supercoercive.*

Proof. Let $(x_n)_{n\in\mathbb{N}}$ satisfy $\|x_n\| \to \infty$ as $n \to \infty$.

(i): As $\operatorname{dom} S$ is bounded, there exists $N \in \mathbb{N}$ such that for $n \geq N$ we have $x_n \notin \operatorname{dom} S$. Fixing an arbitrary $n \geq N$, by definition, $\mathcal{D}_T^{\flat,h}(x_n,y) = \infty$, so $\mathcal{D}_T^{\flat,h}(x_n,y)/\|x_n\| = \infty$, and we are done.

To prove (ii) and (iii), we derive from Remark 4.4(ii) that, since Ty is compact, there exist $v_n \in Ty$ and $(a_n,b_n) \in \mathcal{G}(S)$ such that

$$(56) \qquad \forall n \in \mathbb{N}, \quad 4\mathcal{D}_T^{\flat,h}(x_n,y) \geq \|x_n - a_n\|^2 + \|v_n - b_n\|^2.$$

Here, we note that $\|x_n\| \to \infty$ as $n \to \infty$ and that $(v_n)_{n\in\mathbb{N}}$ is bounded due to compactness of Ty.

(ii): If $(a_{k_n})_{n\in\mathbb{N}}$ is a subsequence of $(a_n)_{n\in\mathbb{N}}$ with $\|a_{k_n}\| \to \infty$, then by assumption, $\|b_{k_n}\| \to \infty$, which implies that $\|v_{k_n} - b_{k_n}\| \to \infty$. Applying Lemma 4.3(i) to the sequences $(x_n)_{n\in\mathbb{N}}$, $(a_n)_{n\in\mathbb{N}}$, and $(v_n-b_n)_{n\in\mathbb{N}}$, we obtain that $\|x_n-a_n\|^2+\|v_n-b_n\|^2 \to \infty$ and, by (56), $\mathcal{D}_T^{\flat,h}(x_n,y) \to \infty$ as $n \to \infty$.

(iii): If $(a_{k_n})_{n\in\mathbb{N}}$ is a subsequence of $(a_n)_{n\in\mathbb{N}}$ with $\|a_{k_n}\| \to \infty$, then by assumption, $\|b_{k_n}\|^2/\|a_{k_n}\| \to \infty$, so $\|b_{k_n}\| \to \infty$ and

$$(57) \qquad \frac{\|v_{k_n} - b_{k_n}\|^2}{\|a_{k_n}\|} \geq \frac{\|b_{k_n}\|^2 - 2\|v_{k_n}\|\|b_{k_n}\|}{\|a_{k_n}\|} = \frac{\|b_{k_n}\|^2}{\|a_{k_n}\|}\left(1 - \frac{2\|v_{k_n}\|}{\|a_{k_n}\|\|b_{k_n}\|}\right) \to \infty.$$

Now, using Lemma 4.3(ii) yields

$$(58) \qquad \frac{\|x_n - a_n\|^2 + \|v_n - b_n\|^2}{\|x_n\|} \to \infty \quad \text{as } n \to \infty,$$

which together with (56) completes the proof. $\qquad\square$

THEOREM 4.6 (right coercivity and right supercoercivity of $\mathcal{D}_T^{\flat,h}$). *Let $x \in \operatorname{dom} S$. Then*

(i) *If $\operatorname{dom} T$ is bounded, then $\mathcal{D}_T^{\flat,h}(x,\cdot)$ is supercoercive and hence coercive.*

Suppose further that T has strongly compact images. Then the following hold:

(ii) *If $(a_n, b_n) \in \mathcal{G}(S)$, $(y_n, v_n) \in \mathcal{G}(T)$, and $\|y_n - a_n\| \to \infty$ imply $\|v_n - b_n\| \to \infty$, then $\mathcal{D}_T^{\flat,h}(x,\cdot)$ is coercive.*

(iii) *If $(a_n, b_n) \in \mathcal{G}(S)$, $(y_n, v_n) \in \mathcal{G}(T)$, and $\|y_n - a_n\| \to \infty$ imply $\|v_n - b_n\|^2/\|y_n - a_n\| \to \infty$, then $\mathcal{D}_T^{\flat,h}(x,\cdot)$ is supercoercive.*

Proof. Let $(y_n)_{n \in \mathbb{N}}$ satisfy $\|y_n\| \to \infty$ as $n \to \infty$.

(i): By the boundedness of $\operatorname{dom} T$, there exists $N \in \mathbb{N}$ such that for $n \geq N$, $y_n \notin \operatorname{dom} T$. Fixing $n \geq N$, the definition of $\mathcal{D}_T^{\flat,h}$ yields $\mathcal{D}_T^{\flat,h}(x, y_n) = \infty$, which implies that $\mathcal{D}_T^{\flat,h}(x, y_n)/\|y_n\| = \infty$, and we are done.

(ii) and (iii): Since T has strongly compact images, Ty_n is compact with respect to the strong topology. By Remark 4.4(ii), there exist $v_n \in Ty_n$ and $(a_n, b_n) \in \mathcal{G}(S)$ such that

(59)
$$\forall n \in \mathbb{N}, \quad 4\mathcal{D}_T^{\flat,h}(x, y_n) \geq \|x - a_n\|^2 + \|v_n - b_n\|^2 = \|(y_n - x) - (y_n - a_n)\|^2 + \|v_n - b_n\|^2.$$

As $n \to \infty$, $\|y_n - x\| \to \infty$ since $\|y_n\| \to \infty$.

Now we show (ii). Suppose that $(a_n, b_n) \in \mathcal{G}(S)$, $(y_n, v_n) \in \mathcal{G}(T)$, and $\|y_n - a_n\| \to \infty$ imply $\|v_n - b_n\| \to \infty$. We have that if $(y_{k_n} - a_{k_n})_{n \in \mathbb{N}}$ is a subsequence of $(y_n - a_n)_{n \in \mathbb{N}}$ with $\|y_{k_n} - a_{k_n}\| \to \infty$, then $\|v_{k_n} - b_{k_n}\| \to \infty$. Applying Lemma 4.3(i) to the sequences $(y_n - x)_{n \in \mathbb{N}}$, $(y_n - a_n)_{n \in \mathbb{N}}$, and $(v_n - b_n)_{n \in \mathbb{N}}$, we obtain that $\|(y_n - x) - (y_n - a_n)\|^2 + \|v_n - b_n\|^2 \to \infty$, and so $\mathcal{D}_T^{\flat,h}(x, y_n) \to \infty$ as $n \to \infty$. This shows (ii).

Now we show (iii). Suppose that $(a_n, b_n) \in \mathcal{G}(S)$, $(y_n, v_n) \in \mathcal{G}(T)$, and $\|y_n - a_n\| \to \infty$ imply $\|v_n - b_n\|^2/\|y_n - a_n\| \to \infty$. We derive that if $(y_{k_n} - a_{k_n})_{n \in \mathbb{N}}$ is a subsequence of $(y_n - a_n)_{n \in \mathbb{N}}$ with $\|y_{k_n} - a_{k_n}\| \to \infty$, then $\|v_{k_n} - b_{k_n}\|^2/\|y_{k_n} - a_{k_n}\| \to \infty$. Now, Lemma 4.3(ii) completes the proof. This shows (iii). $\square$

As we will see in the following example, the conditions in Theorem 4.5 (resp., Theorem 4.6) are not necessary conditions for the left (resp., right) coercivity or supercoercivity of $\mathcal{D}_T^{\flat,h}$.

Example 4.7. Suppose that $X = \mathbb{R}$. Let $f = \operatorname{Id}: \mathbb{R} \to \mathbb{R}$, $S = \nabla f = 1$, and $h: \mathbb{R} \times \mathbb{R} \to \mathbb{R}_\infty$ given by

(60)
$$h(x, v) = f(x) + f^*(v) = x + \iota_{\{1\}}(v).$$

Let also $T = 0$. Then

(61)
$$\forall (x, y) \in \mathbb{R}^2, \quad \mathcal{D}_T^h(x, y) = h(x, 0) - \langle x, 0 \rangle = h(x, 0) = \infty.$$

Therefore, both $\mathcal{D}_T^h(\cdot, y)$ and $\mathcal{D}_T^h(x, \cdot)$ are supercoercive and hence coercive (for all $x, y \in \mathbb{R}$), while S and T do not satisfy the assumptions in Theorem 4.5 or Theorem 4.6.

COROLLARY 4.8 (left supercoercivity of $\mathcal{D}_h$). *Let $f \in \Gamma_0(X)$ be such that ∂f is point-to-point, $T = S = \partial f$, $h \in \mathcal{H}(S)$, and $y \in \operatorname{dom} S$. Then the following hold:*

(i) *If $\operatorname{dom} S$ is bounded, $\mathcal{D}_h(\cdot, y)$ is supercoercive and hence coercive.*

(ii) *If S satisfies the property that $(a_n, b_n) \in \mathcal{G}(S)$ and $\|a_n\| \to \infty$ implies $\|b_n\| \to \infty$, then $\mathcal{D}_h(\cdot, y)$ is coercive.*

(iii) *If $(a_n, b_n) \in \mathcal{G}(S)$ and $\|a_n\| \to \infty$ imply $\|b_n\|^2/\|a_n\| \to \infty$, then $\mathcal{D}_h(\cdot, y)$ is supercoercive.*

Proof. Apply Theorem 4.5 with $T = S = \partial f$. Because $f \in \Gamma_0(\mathcal{H})$, we have that ∂f is maximally monotone. The compactness of $\partial f(y)$ comes from the fact that ∂f is point-to-point. $\square$

Example 4.9 (left supercoercive distances on $\mathbb{R}$). Let $X = \mathbb{R}$ and $f := x \mapsto |x|^{3/2+\varepsilon}$ for some $\varepsilon > 0$, and let $h \in \mathcal{H}(\nabla f)$. Then $\mathcal{D}_h$ is left supercoercive.

To check, we need only show that f satisfies the criteria for Corollary 4.8.

Since $\nabla f : x \mapsto (3/2 + \varepsilon)\mathrm{sign}(x)|x|^{1/2+\varepsilon}$, we have that

$$(62) \qquad \frac{\|\nabla f(x)\|^2}{\|x\|} \geq \frac{|x|^{1+2\varepsilon}}{|x|} = |x|^{2\varepsilon} \to \infty \ \text{ as } \ |x| \to \infty,$$

showing the sufficient conditions for Corollary 4.8. This example is illustrated in Figure 4, which shows the geometric intuition underpinning Theorem 4.5.

COROLLARY 4.10 (right supercoercivity of $\mathcal{D}_h$). *Let $f \in \Gamma_0(X)$ be such that ∂f is point-to-point, $T = S = \partial f$, $h \in \mathcal{H}(\partial f)$, and $x \in \mathrm{dom}\, \partial f$. Then the following hold:*
  (i) *If ∂f is bounded, then $\mathcal{D}_h$ is supercoercive and hence coercive.*
  (ii) *If $(a_n, b_n), (y_n, v_n) \in \mathcal{G}(\partial f)$ and $\|y_n - a_n\| \to \infty$ imply $\|v_n - b_n\| \to \infty$, then $\mathcal{D}_h$ is coercive.*
  (iii) *If $(a_n, b_n), (y_n, v_n) \in \mathcal{G}(\partial f)$ and $\|y_n - a_n\| \to \infty$ imply $\|v_n - b_n\|^2/\|y_n - a_n\| \to \infty$, then $\mathcal{D}_h$ is supercoercive.*

Proof. Apply Theorem 4.6 with $T = S = \partial f$. Here ∂f automatically has compact images because it is point-to-point. $\square$

Example 4.11 (right supercoercive distances on $\mathbb{R}$). Let $X = \mathbb{R}$ and $f := x \mapsto |x|^p$ for some $p \geq 2$ and $h \in \mathcal{H}(\partial f)$. Then $\mathcal{D}_h$ is right supercoercive.

To check, we need only show that f satisfies the criteria for Corollary 4.10.

Since $\nabla f : x \to \mathrm{sign}(x)p|x|^{p-1}$ and $p \geq 2$ we have that $(|x - y| \geq 2) \implies \|\nabla f(x) - \nabla f(y)\| \geq |x - y|$. Thus, for $|x - y| \geq 2$,

$$(63) \qquad \frac{\|\nabla f(x) - \nabla f(y)\|^2}{\|y - x\|} \geq \frac{|y - x|^2}{|y - x|} = |y - x| \to \infty \ \text{ as } \ |y - x| \to \infty,$$

showing the sufficient conditions for Corollary 4.10. This example is illustrated at bottom in Figure 4.

Example 4.12 (functions on $\mathbb{R}$ which fail the assumptions of Theorems 4.5 and 4.6). Let $X = \mathbb{R}$ and $f := x \mapsto |x|^{3/2}$.

Let $x_n := n$ and $y_n := 0$. Then, since $\nabla f : x \mapsto \mathrm{sign}(x)|x|^{1/2}$, we have that

$$(64) \quad d((x_n, \nabla f(y_n)), \mathcal{G}(\nabla f)) = d((n, 0), \mathcal{G}(\nabla f)) \leq d((n, 0), (n, n^{1/2})) = n^{1/2} = x_n^{1/2},$$

and so $d((x_n, \nabla f(y_n)), \mathcal{G}(\nabla f))^2 = x_n$ for all n.

Example 4.13 (theorem conditions sufficient but not necessary). The conditions of Theorem 4.5 are sufficient but not necessary. Let f be the Boltzmann–Shannon entropy, and we have that for $x > 1$,

$$(65) \qquad \frac{\|\nabla f(x)\|^2}{\|x\|} = \frac{\log(x)^2}{x} \leq \frac{x}{x} = 1 \to 1 \ \text{ as } \ x \to \infty,$$

so the sufficient conditions from Theorem 4.5 fail.

The form of $\mathcal{D}_{\sigma_{\log}}$ is given in (45). If x or y is less than or equal zero, $\mathcal{D}_{\sigma_{\log}}(x, y) = \infty$. Fixing $y > 0$, we have that for $x > 1$,

$$\tag{66} \frac{\mathcal{D}_{\sigma_{\partial f}}(x, y)}{\|x\|} = \frac{x(\log(x) - \log(y))}{x} = \log(x) - \log(y) \to \infty \ \text{ as } \ x \to \infty,$$

and so $\mathcal{D}_{\sigma_{\log}}$ is left supercoercive.

4.3. Coercivity of the sum of $\mathcal{D}$ and a convex function. The following propositions and their accompanying proofs extend and follow the template of Bauschke, Combettes, and Noll in [6, Lemma 2.12], with modifications necessary in order to handle the greater generality of $\mathcal{D}_T^{\flat, h}$. In the following, X is assumed to be a real Hilbert space, $U_S := \operatorname{int} \operatorname{dom} S$, and $U_T := \operatorname{int} \operatorname{dom} T$.

PROPOSITION 4.14 (left coercivity of the sum of $\mathcal{D}_T^{\flat, h}$ and a convex function). *Let $\theta \in \Gamma_0(X)$ be such that $U_S \cap \operatorname{dom} \theta \neq \varnothing$, and let $\gamma \in \mathbb{R}_{++}$. Suppose that one of the following holds:*
   (a) *$U_S \cap \operatorname{dom} \theta$ is bounded, and for all $y \in U_T$, $\mathcal{D}_T^{\flat, h}(\cdot, y)$ is coercive.*
   (b) *$\inf \theta(U_S) > -\infty$, and for all $y \in U_T$, $\mathcal{D}_T^{\flat, h}(\cdot, y)$ is coercive.*
   (c) *For all $y \in U_T$, $\mathcal{D}_T^{\flat, h}(\cdot, y)$ is supercoercive.*
Then

$$\tag{67} \forall y \in U_T, \quad \theta(\cdot) + \frac{1}{\gamma} \mathcal{D}_T^{\flat, h}(\cdot, y) \text{ is coercive.}$$

Proof. We will show that (a) $\implies$ (b) $\implies$ (67) and that (c) $\implies$ (67). First, we have from [5, Theorem 9.20] that there exists $(u, \alpha) \in X \times \mathbb{R}$ such that

$$\tag{68} \theta \geq \langle u, \cdot \rangle + \alpha.$$

(a) $\implies$ (b): By the Cauchy–Schwarz inequality,

$$\tag{69} \forall x \in X, \quad \theta(x) \geq \langle u, x \rangle + \alpha \geq -\|u\|\|x\| + \alpha,$$

which yields

$$\tag{70} \inf \theta(U_S) = \inf \theta(U_S \cap \operatorname{dom} \theta) \geq -\|u\| \sup_{x \in U_S \cap \operatorname{dom} \theta} \|x\| + \alpha > -\infty$$

since $U_S \cap \operatorname{dom} \theta$ is bounded. Hence, (a) $\implies$ (b).

(b) $\implies$ (67): Let $y \in U_T$. Suppose for a contradiction that there exist a sequence $(x_n)_{n \in \mathbb{N}}$ in X and a constant $\mu \in \mathbb{R}_{++}$ such that $\|x_n\| \to \infty$ and

$$\tag{71} \forall n \in \mathbb{N}, \quad \theta(x_n) + \frac{1}{\gamma} \mathcal{D}_T^{\flat, h}(x_n, y) \leq \mu.$$

For each $n \in \mathbb{N}$, since $\theta(x_n) > -\infty$, it follows that $\mathcal{D}_T^{\flat, h}(x_n, y) < \infty$, and so $x_n \in \operatorname{dom} S$. Next, according to [5, Proposition 11.1(iv)], $\inf \theta(\operatorname{dom} S) = \inf \theta(U_S)$, which implies that $\theta(x_n) \geq \inf \theta(U_S) > -\infty$ for all $n \in \mathbb{N}$. Combining with (71), we obtain that

$$\tag{72} \forall n \in \mathbb{N}, \quad \mathcal{D}_T^{\flat, h}(x_n, y) \leq \gamma(\mu - \inf \theta(U_S)) < \infty,$$

which contradicts the coercivity of $\mathcal{D}_T^{\flat, h}(\cdot, y)$.

(c) $\implies$ (67): Notice that

$$(73) \qquad \theta(\cdot) + \frac{1}{\gamma}\mathcal{D}_T^{\flat,h}(\cdot, y) \geq \langle u, \cdot \rangle + \alpha + \frac{1}{\gamma}\mathcal{D}_T^{\flat,h}(\cdot, y).$$

The right-hand side is the sum of a supercoercive function and an affine function, and hence a coercive function due to [5, Corollary 16.21]. Since $\theta(\cdot) + \frac{1}{\gamma}\mathcal{D}_T^{\flat,h}(\cdot, y)$ is bounded from below by a coercive function, it is coercive. □

PROPOSITION 4.15 (right coercivity of the sum of $\mathcal{D}_T^{\flat,h}$ and a convex function). *Let $\theta \in \Gamma_0(X)$ be such that $U_T \cap \operatorname{dom}\theta \neq \varnothing$ and let $\gamma \in \mathbb{R}_{++}$. Suppose that one of the following holds:*
  (a) *$U_T \cap \operatorname{dom}\theta$ is bounded, and for all $x \in U_S$, $\mathcal{D}_T^{\flat,h}(x, \cdot)$ is coercive.*
  (b) *$\inf\theta(U_T) > -\infty$, and for all $x \in U_S$, $\mathcal{D}_T^{\flat,h}(x, \cdot)$ is coercive.*
  (c) *For all $x \in U_S$, $\mathcal{D}_T^{\flat,h}(x, \cdot)$ is supercoercive.*
Then

$$(74) \qquad \forall x \in U_S, \quad \theta(\cdot) + \frac{1}{\gamma}\mathcal{D}_T^{\flat,h}(x, \cdot) \text{ is coercive.}$$

Proof. This is analogous to the proof of Proposition 4.14. □

5. Conclusion. In section 2, we illuminated the similarities between Bregman distances and the new GBDs, explaining the domain conditions under which they are equal when the Fenchel–Young representative is employed. We also introduced the lower closed GBD, a variant whose advantages we motivated in sections 3 and 4.

In section 3, we provided detailed examples of how to compute the new GBDs, illustrating with the energy and the Boltzmann–Shannon entropy, whose Bregman distances, respectively, correspond to the classical Moreau case and the Kullback–Leibler divergence. We compared the Fenchel–Young representative case with the two cases of the Fitzpatrick representative and its conjugate. These are the two other most natural representative functions to consider because they serve as book-ends for the representative set $\mathcal{H}(S)$, as motivated in section 2.

In section 3.2 we answered the open question of finding the conjugate for the Fitzpatrick function of the logarithm. In so doing, we demonstrated how to use the graphical characterizations of representative functions in order to compute GBDs, and we illustrated the role that special functions like Lambert $\mathcal{W}$ play in computational discovery. The method of computational discovery that we used is prototypical of what one might employ in similar situations where the symbolic computation poses a challenge.

Section 4 contains the most important theoretical contribution of this work: a framework for verifying the coercivity and supercoercivity of the left and right distances, as well as the coercivity of the sum of these distances together with a Legendre function. We have also illustrated how this framework for sufficiency possesses a useful geometric interpretation, because the GBDs provide an upper estimate on a set distance. In our examples, we illustrated what might go wrong when sufficient criteria do not hold. These coercivity properties are important because of the role they play in establishing asymptotic properties for envelopes and proximity operators in the classical Bregman case, and also in establishing existence of minimizers of regularized problems; see, for example, [7, 16, 17, 18, 22]. Such properties are important because many optimization algorithms may be viewed as special cases of gradient descent applied to envelope functions.

Future work. The coercivity framework we have established makes possible several new avenues of inquiry. While the conditions we provide for verifying coercivity and supercoercivity in section 4 are sufficient, they are not always necessary. An important future work is to catalog useful (computable) distances for which the coercivity results hold. In particular, by establishing the aforementioned coercivity framework, we have set the table for a study of the left and right envelopes, along with their corresponding proximity operators. A much more interesting question is whether certain optimization algorithms might be viewed as gradient descent applied to GBD envelopes other than already known Fenchel–Young cases. Another natural question is, What do the dual characterizations of such algorithms look like?

REFERENCES

[1] H. Attouch, *Convergence de fonctions convexes, des sous-différentiels et semi-groupes associés*, C. R. Acad. Sci. Paris, 284 (1977), pp. 539–542.

[2] H. Attouch, *Variational Convergence for Functions and Operators*, Pitman, Boston, MA, 1984.

[3] H. H. Bauschke and J. M. Borwein, *Legendre functions and the method of random Bregman projections*, J. Convex Anal., 4 (1997), pp. 27–67.

[4] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, *Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces*, Commun. Contemp. Math., 3 (2001), pp. 615–647.

[5] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed., Springer, Cham, 2017.

[6] H. H. Bauschke, P. L. Combettes, and D. Noll, *Joint minimization with alternating Bregman proximity operators*, Pac. J. Optim., 2 (2006), pp. 401–424.

[7] H. H. Bauschke, M. N. Dao, and S. B. Lindstrom, *Regularizing with Bregman–Moreau envelopes*, SIAM J. Optim., 28 (2018), pp. 3208–3228.

[8] H. H. Bauschke and S. B. Lindstrom, *Proximal averages for minimization of entropy functionals*, Pure Appl. Funct. Anal., 5 (2020), pp. 505–531.

[9] H. H. Bauschke, D. A. McLaren, and H. S. Sendov, *Fitzpatrick functions: Inequalities, examples, and remarks on a problem by S. Fitzpatrick*, J. Convex Anal., 13 (2006), pp. 499–523.

[10] J. M. Borwein and S. B. Lindstrom, *Meetings with Lambert W and other special functions in optimization and analysis*, Pure Appl. Funct. Anal., 1 (2017), pp. 361–396.

[11] J. M. Borwein, S. Reich, and S. Sabach, *A characterization of Bregman firmly nonexpansive operators using a new monotonicity concept*, J. Nonlinear Convex Anal., 12 (2011), pp. 161–184.

[12] S. Bartz, H. H. Bauschke, J. M. Borwein, S. Reich, and X. Wang, *Fitzpatrick functions, cyclic monotonicity and Rockafellar's antiderivative*, Nonlinear Anal., 66 (2007), pp. 1198–1223.

[13] J. M. Borwein and J. D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples*, Cambridge University Press, Cambridge, UK, 2010.

[14] 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*, Comput. Math. Math. Phys., 7 (1967), pp. 200–217.

[15] M. N. Bùi and P. L. Combettes, *Bregman Forward-Backward Operator Splitting*, preprint, arXiv:1908.03878, 2019.

[16] R. S. Burachik and J. Dutta, *Inexact proximal point methods for variational inequality problems*, SIAM J. Optim., 20 (2010), pp. 2653–2678.

[17] R. S. BURACHIK AND A. N. IUSEM, *A generalized proximal point algorithm for the variational inequality problem in a Hilbert space*, SIAM J. Optim., 8 (1998), pp. 197–216.

[18] R. S. BURACHIK AND A. N. IUSEM, *A generalized proximal point algorithm for the nonlinear complementarity problem*, RAIRO Oper. Res., 33 (1999), pp. 447–479.

[19] R. S. BURACHIK AND A. N. IUSEM, *Set Valued Mappings and Enlargements of Monotone Operators*, Springer, New York, 2008.

[20] R. S. BURACHIK AND G. KASSAY, *On a generalized proximal point method for solving equilibrium problems in Banach spaces*, Nonlinear Anal., 75 (2012), pp. 6456–6464.

[21] R. S. BURACHIK AND J. E. MARTÍNEZ-LEGAZ, *On Bregman-type distances for convex functions and maximally monotone operators*, Set-Valued Var. Anal., 26 (2018), pp. 369–384.

[22] R. S. BURACHIK AND S. SCHEIMBERG, *A proximal point method for the variational inequality problem in Banach spaces*, SIAM J. Control Optim., 39 (2001), pp. 1633–1649.

[23] R. S. BURACHIK AND B. F. SVAITER, *Maximal monotone operators, convex functions and a special family of enlargements*, Set-Valued Anal., 10 (2002), pp. 297–316.

[24] C. BYRNE AND Y. CENSOR, *Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback–Leibler distance minimization*, Ann. Oper. Res., 105 (2001), pp. 77–98.

[25] Y. CENSOR AND G. T. HERMAN, *Block-iterative algorithms with underrelaxed Bregman projections*, SIAM J. Optim., 13 (2002), pp. 283–297.

[26] Y. CENSOR, A. N. IUSEM, AND S. A. ZENIOS, *An interior point method with Bregman functions for the variational inequality problem with paramonotone operators*, Math. Program., 81 (1998), pp. 373–400.

[27] Y. CENSOR AND A. LENT, *An iterative row-action method for interval convex programming*, J. Optim. Theory Appl., 34 (1981), pp. 321–353.

[28] Y. CENSOR AND S. A. ZENIOS, *Proximal minimization algorithm with D-functions*, J. Optim. Theory Appl., 73 (1992), pp. 451–464.

[29] Y. CENSOR AND S. A. ZENIOS, *Parallel Optimization: Theory, Algorithms, and Applications*, Oxford University Press, New York, 1997.

[30] G. CHEN AND M. TEBOULLE, *Convergence analysis of a proximal-like minimization algorithm using Bregman functions*, SIAM J. Optim., 3 (1993), pp. 538–543.

[31] R. M. CORLESS, G. H. GONNET, D. E. HARE, D. J. JEFFREY, AND D. E. KNUTH, *On the Lambert W function*, Adv. Comput. Math., 5 (1996), pp. 329–359.

[32] J. ECKSTEIN, *Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming*, Math. Oper. Res., 18 (1993), pp. 202–226.

[33] F. FACCHINEI AND J. S. PANG, *Finite-dimensional Variational Inequalities and Complementarity Problems*, volume II, Springer Science & Business Media, New York, 2007.

[34] S. FITZPATRICK, *Representing monotone operators by convex functions*, in Functional Analysis and Optimization, Workshop and Miniconference, Canberra, Australia, Proc. Center Math. Anal. Australian Nat. Univ. 20, 1988, 1988, pp. 59–65.

[35] K. C. KIWIEL, *Proximal minimization methods with generalized Bregman functions*, SIAM J. Control Optim., 35 (1997), pp. 1142–1168.

[36] J.-J. MOREAU, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.

[37] P. PATRINOS, L. STELLA, AND A. BEMPORAD, *Douglas–Rachford splitting: Complexity estimates and accelerated variants*, in Proceedings of the 53rd IEEE Conference on Decision and Control, 2014, pp. 4234–4239.

[38] S. REICH AND S. SABACH, *Two strong convergence theorems for a proximal method in reflexive Banach spaces*, Numer. Funct. Anal. Optim., 31 (2010), pp. 22–44.

[39] D. REEM, DANIEL, S. REICH, AND A. DE PIERRO, *Re-examination of Bregman functions and new properties of their divergences*, Optimization, 68 (2019), pp. 279–348.

[40] R. T. ROCKAFELLAR AND R. J. B. WETS, *Variational Analysis*, Springer, Cham, 2009.

[41] A. THEMELIS, L. STELLA, AND P. PATRINOS, *Forward–backward envelope for the sum of two nonconvex functions: Further properties and nonmonotone linesearch algorithms*, SIAM J. Optim., 28 (2018), pp. 2274–2303.