CrossMark

On Bregman-Type Distances for Convex Functions and Maximally Monotone Operators

Regina S. Burachik[1] · Juan E. Martínez-Legaz[2]

Abstract Given two point to set operators, one of which is maximally monotone, we introduce a new distance in their graphs. This new concept reduces to the classical Bregman distance when both operators are the gradient of a convex function. We study the properties of this new distance and establish its continuity properties. We derive its formula for some particular cases, including the case in which both operators are linear monotone and continuous. We also characterize all bi-functions D for which there exists a convex function h such that D is the Bregman distance induced by h.

1 Introduction

This paper focuses on an extension of the concept of Bregman distances to the framework of point-to-set operators, one of which is maximally monotone. In the same way as classical

✉ Juan E. Martínez-Legaz
JuanEnrique.Martinez.Legaz@uab.cat

Regina S. Burachik
Regina.Burachik@unisa.edu.au

[1] School of Information Technology and Mathematical Sciences, University of South Australia, Mawson Lakes Campus SA 5095, Adelaide, Australia

[2] Departament d'Economia i d'Història Econòmica, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

🖄 Springer

Bregman distances are induced by convex functions, our new distance is induced by convex representations of one of the maps; however, the way we associate a Bregman distance with a convex representation is completely different from the association of a classical Bregman distance with a differentiable convex function.

Classical Bregman distances have proved to be useful in devising algorithms for convex optimization problems, as well as for variational inequalities, in which the distance plays a penalization role. It is then natural to investigate whether one could introduce a more general notion of Bregman distance which could be useful in algorithms for solving monotone variational inequalities. Our new distance also provides a new interpretation of solutions of variational inequalities. The variational inequality can be formulated as follows. Let X be a Banach space and X^* its dual. Given a maximally monotone operator $S : X \rightrightarrows X^*$ and a closed and convex set $C \subseteq X$, a solution of the variational inequality problem $VIP(S, C)$ is a pair $(x, v) \in X \times X^*$ such that $x \in C$, $v \in Sx$ and

$$\langle y - x, v \rangle \geq 0, \text{ for all } y \in C. \tag{1.1}$$

The minimization of a convex function f constrained to the set C is a particular instance of the $VIP(S, C)$, with $S = \partial f$. The variational inequality problem, in turn, is a particular instance of the more general inclusion problem $0 \in T(x)$, with T being the sum of ∂f plus the normal cone N_C, where $N_C(x) := \{v \in X^* : \langle y - x, v \rangle \leq 0, \text{ for all } y \in C\}$. Combining this definition with (1.1) shows that solutions of the $VIP(S, C)$ are those elements in the graph of S which intersect the graph of $-N_C$. We will show that our distance vanishes at solutions of $VIP(S, C)$, when applied to the maps S and $-N_C$, which gives a new interpretation to solutions of $VIP(S, C)$. Moreover, our notion of Bregman distance extends the classical one. Namely, when both maps are ∂f, it reduces to the Bregman distance induced by f, in the sense of [16]. When f is convex and differentiable, it reduces to the classical Bregman distance, in the sense of [6, Section 6.2].

In the present paper, we study the basic properties of this distance and show some specific examples. We also study classical Bregman distances. We provide two axiomatic characterizations, that is, we give necessary and sufficient conditions for a bifunction defined on the product of a Banach space with itself to be the Bregman distance associated to some convex function. Moreover, we study the correspondence that assigns to each differentiable convex function f its associated Bregman distance

$$D_f(x, y) := f(x) - f(y) - \langle x - y, \nabla f(y) \rangle.$$

The rest of the paper is organized as follows. Section 2 presents some preliminaries on convex functions and maximally monotone operators. In particular, we recall the basic ideas on the representability of monotone operators by convex functions as well as the related notion of enlargement of a maximally monotone operator and its main properties. In Section 2 we introduce and study our new notion of Bregman distance. It contains two subsections: in the first one we consider the particular case when the monotone operators are linear, and the second one is devoted to the study of the lower semicontinuity properties of the newly introduced Bregman distances. Section 3 contains our characterizations of classical Bregman distances and studies the mapping $f \longmapsto D_f$ defined above.

2 Preliminaries

Let $(X, \|\cdot\|)$ and $(X^*, \|\cdot\|_*)$ be a Banach space and its dual, respectively. Given a point-to-set operator $T : X \rightrightarrows X^*$, the set $\mathcal{D}(T) := \{x \in X : T(x) \neq \emptyset\}$ is called the domain of T,

while $\mathcal{G}(T) := \{(x, x^*) \in X \times X^* : x^* \in T(x)\}$ is the graph of T. Fix C a subset of a vector space Z. The *indicator function of C* is the function $\delta_C : Z \to \mathbb{R} \cup \{+\infty\} =: \mathbb{R}_\infty$ defined as $\delta_C(z) := 0$ for $z \in C$ and $\delta_C(z) := +\infty$ for $z \notin C$. We denote by int C and bdry C the interior and the boundary of C, respectively.

Let $f : X \to \mathbb{R}_\infty$. Then dom $f := f^{-1}(\mathbb{R})$ is the *domain* (or *effective domain*) of f. We say that f is proper if dom $f \neq \varnothing$. The *Fenchel conjugate* of a proper function f is $f^* : X^* \to \mathbb{R}_\infty : x^* \mapsto \sup_{x \in X}\{\langle x, x^* \rangle - f(x)\}$. Recall that $f : X \to \mathbb{R}_\infty$ is *lower-semicontinuous* at $\bar{x}$ (lsc at $\bar{x}$) if and only if for all $\lambda \in \mathbb{R}$ such that $f(\bar{x}) > \lambda$, there exists a neighborhood U of $\bar{x}$ such that $f(x) > \lambda$ for all $x \in U$. Similarly, f is *upper-semicontinuous* at $\bar{x}$ (usc at $\bar{x}$) if and only if for all $\lambda \in \mathbb{R}$ such that $f(\bar{x}) < \lambda$, there exists a neighborhood U of $\bar{x}$ such that $f(x) < \lambda$ for all $x \in U$.

Given a function $f : X \to \mathbb{R}_\infty$, the *subdifferential* of f is the point-to-set mapping $\partial f : X \rightrightarrows X^*$ defined by

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

If C is a closed and convex set, then $\partial \delta_C =: N_C$, the normal cone to the set C. Namely,

$$N_C(x) := \{v \in X^* : \langle v, y - x \rangle \leq 0 \text{ for all } y \in C\}.$$

Given $\varepsilon \geq 0$, the $\varepsilon-$*subdifferential* of f is the point-to-set mapping $\partial_\varepsilon f : X \rightrightarrows X^*$ defined by

$$\partial_\varepsilon f(x) := \begin{cases} \{x^* \in X^* \mid (\forall y \in X)\, \langle y - x, x^* \rangle + f(x) \leq f(y) + \varepsilon\} & \text{if } x \in \text{dom } f; \\ \varnothing & \text{otherwise.} \end{cases} \tag{2.2}$$

For future use we recall the *Fenchel-Young inequality* for a convex and lower semicontinuous function $f : X \to \mathbb{R}_\infty$,

$$f^*(v) + f(x) \geq \langle x, v \rangle, \quad \text{for all } (x, v) \in X \times X^*, \tag{2.3}$$

and

$$f^*(v) + f(x) = \langle x, v \rangle \quad \text{if and only if } v \in \partial f(x). \tag{2.4}$$

If Y is a vector space and $x, y \in Y$ with $x \neq y$, we denote by $[x, y]$, $]x, y[$ and $]y, x_{+\infty}[$ the sets of points $\lambda x + (1-\lambda)y$, with $\lambda \in [0, 1]$, $\lambda \in\ ]0, 1[$ and $\lambda \in\ ]0, +\infty[$, respectively.

In our analysis, we will make use of the concept of enlargement of a maximally monotone operator, which we define next. Since these objects approximate the graph of the operator, it is not surprising that they are useful in analyzing the distance induced by the graphs of these operators. The definition below was introduced in [22].

Definition 2.1 Let $T : X \rightrightarrows X^*$. We say that $E : X \times \mathbb{R}_+ \rightrightarrows X^*$ is an *enlargement of T* when the following hold.

(E_1) $T(x) \subseteq E(x, \varepsilon)$ for all $\varepsilon \geq 0$, $x \in X$.

(E_2) If $0 \leq \varepsilon_1 \leq \varepsilon_2$, then $E(x, \varepsilon_1) \subseteq E(x, \varepsilon_2)$ for all $x \in X$.

(E_3) The *transportation formula holds* for E: Let $v^1 \in E(x^1, \varepsilon_1)$, $v^2 \in E(x^2, \varepsilon_2)$, and $\alpha \in [0, 1]$. Define

$$\hat{x} := \alpha x^1 + (1 - \alpha)x^2,$$
$$\hat{v} := \alpha v^1 + (1 - \alpha)v^2,$$
$$\hat{\varepsilon} := \alpha \varepsilon_1 + (1 - \alpha)\varepsilon_2 + \alpha\langle x^1 - \hat{x}, v^1 - \hat{v} \rangle + (1 - \alpha)\langle x^2 - \hat{x}, v^2 - \hat{v} \rangle.$$

Then $\hat{\varepsilon} \geq 0$ and $\hat{v} \in E(\hat{x}, \hat{\varepsilon})$.

When E verifies $(E_1) - (E_3)$, we write $E \in \mathbb{E}(T)$.

Assume $T = \partial f$, with f a convex and lower semicontinuous function. In this case the ε-subdifferential $\partial_{\{\cdot\}} f(\cdot) : X \times \mathbf{R}_+ \rightrightarrows X^*$, which maps (x, ε) to the set $\partial_\varepsilon f(x)$, is a fundamental example of enlargement.

Another important example of an enlargement is defined as follows. Given an arbitrary maximally monotone operator $S : X \rightrightarrows X^*$, denote by $S^e : X \times \mathbb{R}_+ \rightrightarrows X^*$ the set valued map defined as

$$S^e(x, \varepsilon) := \{v \in X^* : \langle y - x, u - v \rangle \geq -\varepsilon \text{ for all } y \in X, \, u \in S(y)\} \qquad (2.5)$$

(the set $S^e(x, \varepsilon)$ was called $S^\epsilon(x)$ in [11]).

The following fact collects properties of enlargements that we will need in the sequel.

Fact 2.2 The enlargement S^e was introduced in [7] for the finite-dimensional case, and extended first to Hilbert spaces in [9, 10] and then to Banach spaces in [11]. The following facts (whose proofs can be found in the aforementioned references) hold.

(i)   The set $S^e(x, \varepsilon)$ is weak*-closed for every fixed x and ε.

(ii)   If $x \in \text{int } \mathcal{D}(S)$, then the set $S^e(x, \varepsilon)$ is weak*-compact (see [11] and [6, Theorem 5.3.4]).

(iii)   The mapping S^e is the biggest element in the family $\mathbb{E}(S)$ (see [22] and [6, Theorem 5.4.2]). This means that $E \subseteq S^e$ for every $E \in \mathbb{E}(S)$.

(iv)   Denote by $\mathbb{E}_C(S)$ the subset of $\mathbb{E}(S)$ consisting of all $E \in \mathbb{E}(S)$ such that $E(x, \varepsilon)$ is weak*-closed for every x and every $\varepsilon \geq 0$. Then for every $E \in \mathbb{E}_C(S)$ we have that $E(x, \varepsilon)$ is weak*-compact for every $\varepsilon \geq 0$ and every $x \in \text{int } \mathcal{D}(S)$.

(v) $S^e(\cdot, \varepsilon)$ is *locally bounded* in $\text{int } \mathcal{D}(S)$. Namely, for every $x \in \text{int } \mathcal{D}(S)$ there exists a neighbourhood V of x such that $S^e(V, \varepsilon)$ is bounded (see [6, Theorem 5.3.4]). Since $E(\cdot, \varepsilon) \subset S^e(\cdot, \varepsilon)$ for every $E \in \mathbb{E}(S)$, local boundedness in $\text{int } \mathcal{D}(S)$ is inherited by all $E \in \mathbb{E}(S)$.

Our distance will make use of a family of convex functions associated with maximally monotone operators. We define this family next.

Definition 2.3 Let $S : X \rightrightarrows X^*$ be a maximally monotone operator. We say that $h : X \times X^* \to \mathbb{R}_\infty$ *represents* S if the following three conditions hold:

(i) h is convex and norm $\times$ weak* lower semicontinuous in $X \times X^*$,

(ii) $h(x, v) \geq \langle x, v \rangle$, $\forall (x, v) \in X \times X^*$, and

(iii) $h(x, v) = \langle x, v \rangle \Longleftrightarrow (x, v) \in G(S)$.

We denote this situation as $h \in \mathcal{H}(S)$.

Remark 2.4 Fix $S : X \rightrightarrows X^*$ a maximally monotone operator. It is well known (see, e.g., [12]) that $\mathcal{H}(S)$ has a smallest element and a biggest one. The smallest element is the *Fitzpatrick function associated to S*:

$$\mathcal{F}_S(x, v) := \sup_{(z, w) \in G(S)} \langle z - x, v - w \rangle + \langle x, v \rangle.$$

The biggest element is $\sigma_S := \mathcal{F}_S^* = cl \, conv(\pi + \delta_{G(S)})$, where $\pi : X \times X^* \to \mathbb{R}$ is defined as $\pi(x, v) := \langle x, v \rangle$. For more details on the family $\mathcal{H}(S)$, see [5, 12, 13, 15].

Remark 2.5 An operator $T : X \rightrightarrows X^*$ admitting a representing function h satisfying conditions (i), (ii) and (iii) of Definition 2.3 is necessarily monotone [17, Theorem 5], but it may not be maximally monotone. Such monotone operators are called representable. According to [17, Proposition 32], in finite-dimensional spaces, the monotone representable operators are the intersections of arbitrary families of maximally monotone operators. In infinite dimensional Banach spaces, such intersections are still representable, as easily follows from [17, Corollary 10] and the representability of maximally monotone operators, but a representable operator which cannot be expressed as an intersection of maximally monotone operators was presented in [21]. Some further results on representable monotone operators were given in [4].

We will need the following fact. For its proof, see [12, Propositions 2.6 and 3.5].

Fact 2.6 Let $S : X \rightrightarrows X^*$ a maximally monotone operator. Fix $x \in \mathcal{D}(S)$ and $h \in \mathcal{H}(S)$, if $v \in X^*$ is such that

$$h(x, v) \leq \langle x, v \rangle + \varepsilon, \tag{2.6}$$

then $v \in S^e(x, \varepsilon)$. From the latter inclusion and the definition of $\mathcal{D}_T^{\flat,h}$ (see Definition 3.1 below) we derive that

$$\mathcal{D}_T^{\flat,h}(x, y) = 0 \Leftrightarrow Ty \cap S^e(x, \varepsilon) \neq \emptyset, \text{ for all } \varepsilon > 0.$$

Fact 2.6 motivates the following definition of enlargement.

Remark 2.7 Recall from [12, 13] that to a given maximally monotone operator $S : X \rightrightarrows X^*$ and a fixed $h \in \mathcal{H}(S)$, one can associate the enlargement L^h of S defined as follows:

$$L^h(x, \varepsilon) := \{v \in X^* : h(x, v) \leq \langle x, v \rangle + \varepsilon\}.$$

The norm-weak* lower semicontinuity of h implies that the graph of $L^h(\cdot, \cdot)$ is closed w.r.t. the strong-weak* convergence . From the minimality of the Fitzpatrick function, it can be seen that one has $L^{\mathcal{F}_S} = S^e$; in other words, for $v \in X^*$ one has

$$v \in S^e(x, \varepsilon) \Leftrightarrow \mathcal{F}_S(x, v) \leq \langle x, v \rangle + \varepsilon.$$

3 A Bregman Distance for Maximally Monotone Operators

We will consider the following notion, which generalizes the concept of Bregman distance as given in [16] (see Proposition 3.5 below).

Definition 3.1 Let $S : X \rightrightarrows X^*$ be a maximally monotone operator, and let $T : X \rightrightarrows X^*$. Assume that $h \in \mathcal{H}(S)$. For fixed $(x, y) \in \mathcal{D}(S) \times \mathcal{D}(T)$, define

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

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

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

Remark 3.2 From Remark 2.4 we have that every $h \in \mathcal{H}(S)$ satisfies the inequalities

$$\mathcal{F}_S \leq h \leq \sigma_S,$$

hence, we have directly from the definition that

$$0 \leq \mathcal{D}_T^{\sharp, \mathcal{F}_S} \leq \mathcal{D}_T^{\sharp, h} \leq \mathcal{D}_T^{\sharp, \sigma_S}. \tag{3.7}$$

Analogous inequalities hold for $\mathcal{D}_T^{\flat, h}$.

Remark 3.3 Fix $(x, y) \in \mathcal{D}(S) \times \mathcal{D}(T)$. From the definitions, we readily obtain the following facts.

(a) $\mathcal{D}_T^{\sharp, h}(x, y) \geq \mathcal{D}_T^{\flat, h}(x, y) \geq 0$.
(b)   If $Ty \cap Sx \neq \emptyset$, then $\mathcal{D}_T^{\flat, h}(x, y) = 0$ for every $h \in \mathcal{H}(S)$.
(c)   If $Ty \subset Sx$, then $\mathcal{D}_T^{\sharp, h}(x, y) = \mathcal{D}_T^{\flat, h}(x, y) = 0$ for every $h \in \mathcal{H}(S)$.

Remark 3.4 For Definition 3.1 to make sense one does not need the operator S to be maximally monotone; it is enough for it to be *representable*. Namely, when there exists a function h verifying conditions (i)-(iii) in Definition 2.3. In fact, several results in this paper, namely Remark 3.3, Proposition 3.7(b)(c), implication (ii) $\Longrightarrow$ (i) in Corollary 3.8 , and Lemmas 3.17(b) and 3.18 still hold true if S is assumed to be just representable instead of maximally monotone.

Recall from [16] that, to a given strictly convex function $f : X \to \mathbb{R}_\infty$, we can associate two Bregman distances, defined as follows.

$$D_f^\flat(x, y) := f(x) - f(y) + \inf_{v \in \partial f(y)} \langle y - x, v \rangle,$$

$$D_f^\sharp(x, y) := f(x) - f(y) + \sup_{v \in \partial f(y)} \langle y - x, v \rangle.$$

When f is differentiable at y, then we clearly have

$$D_f^\flat(x, y) = D_f^\sharp(x, y) = D_f(x, y) := f(x) - f(y) - \langle x - y, \nabla f(y) \rangle, \tag{3.8}$$

which is the classical definition of Bregman distance. We prove next that our distances reduce to $D_f^\flat, D_f^\sharp$ when $T = \partial f$.

Proposition 3.5 *Fix a lsc and strictly convex function* $f : X \to \mathbb{R}_\infty$. *Take* $T := S := \partial f$ *and* $h_f(x, v) := f(x) + f^*(v)$ *(note that* $h_f \in \mathcal{H}(S) = \mathcal{H}(T)$*) and we have*

$$\mathcal{D}_T^{\flat, h_f} = D_f^\flat,$$

$$\mathcal{D}_T^{\sharp, h_f} = D_f^\sharp,$$

In particular, when f *is differentiable at* y, *for every* $x \in X$ *we have*

$$\mathcal{D}_T^{\sharp, h_f}(x, y) = \mathcal{D}_T^{\flat, h_f}(x, y) = D_f(x, y) := f(x) - f(y) - \langle x - y, \nabla f(y) \rangle,$$

as in the classical definition of Bregman distances.

Proof For all $v \in \partial f(y)$, we have $f(y) + f^*(v) = \langle v, y \rangle$. Hence we can write

$$\mathcal{D}_T^{\flat, h_f}(x, y) = \inf_{v \in \partial f(y)} \left[f(x) + f^*(v) - \langle x, v \rangle \right]$$

$$= f(x) - f(y) + \inf_{v \in \partial f(y)} \langle y - x, v \rangle = D_f^\flat(x, y),$$

as wanted. The statement for $D_f^\sharp$ follows the same steps. The last statement is a direct consequence of the definitions. $\qquad \square$

The following example shows that our distance can become the classical Bregman distance even when h_f does not represent T.

Example 3.6 Let X be a Hilbert space and fix $\lambda > 0$. Consider the operators $S := \nabla f$ and $T_\lambda := \nabla f + \lambda I$, for $f : X \to \mathbb{R}_\infty$ a convex, coercive, and differentiable function with open domain. Under these assumptions, we have that $S = \nabla f$ is surjective (see., e.g., [2, Theorem 3.3]). Consider $h_f(x, v) := f(x) + f^*(v)$. Then $h_f \in \mathcal{H}(S)$. Call $u_{y,\lambda} := \nabla f(y) + \lambda y$ and $w_{y,\lambda}$ be such that $\nabla f(w_{y,\lambda}) = u_{y,\lambda}$. We have

$$\mathcal{D}_{T_\lambda}^{h_f}(x, y) = f(x) + f^*(u_{y,\lambda}) - \langle x, u_{y,\lambda} \rangle$$

$$= f(x) - f(w_{y,\lambda}) - \langle x - w_{y,\lambda}, \nabla f(w_{y,\lambda}) \rangle = D_f(x, w_{y,\lambda}),$$

where we used Fenchel-Young equality (2.4) in the second equality, and the definition of Bregman distance in the last one. In this way, we can express the distance induced by the operators as a classical Bregman distance.

We have seen in Remark 3.3 that different levels of "overlap" between the sets Sx and Ty imply that the distances $\mathcal{D}_T^{\flat, h}$ and $\mathcal{D}_T^{\sharp, h}$ vanish at (x, y). The next result studies the converse situation, i.e., under which conditions the fact that the distance is zero implies the corresponding "overlap" between the sets Sx and Ty.

Proposition 3.7 *Let $S : X \rightrightarrows X^*$ be a maximally monotone operator, $h \in \mathcal{H}(S)$. Let $(x, y) \in \mathcal{D}(S) \times \mathcal{D}(T)$. The following properties hold:*

(a) *Assume that $T : X \rightrightarrows X^*$ is locally bounded in $\operatorname{int} \mathcal{D}(T)$ and weak*-closed valued (i.e., Tz is weakly* closed for all $z \in \mathcal{D}(T)$). Assume also that $(x, y) \notin \operatorname{bdry} \mathcal{D}(S) \times \operatorname{bdry} \mathcal{D}(T)$. If $\mathcal{D}_T^{\flat, h}(x, y) = 0$ then $Ty \cap Sx \neq \emptyset$.*

(b) *If $\mathcal{D}_T^{\sharp, h}(x, y) = 0$ then $Ty \subset Sx$. Consequently, we have that*

$$\mathcal{D}_T^{\sharp, h}(x, y) = 0 \iff Ty \subset Sx.$$

(c) *Assume that T is point to point. If $\mathcal{D}_T^h(x, y) = 0$ then $Ty \in Sx$. Consequently, if both T and S are point to point, then $Ty = Sx$ if $\mathcal{D}_T^h(x, y) = 0$.*

Proof Let us prove part (a). Assume that $\mathcal{D}_T^{\flat, h}(x, y) = 0$. The assumption on (x, y) implies that either x or y must be in the interior of the corresponding domain. We consider each case separately. If $y \in \operatorname{int} \mathcal{D}(T)$ then Ty is weak*-compact, so the infimum for $v \in Ty$ in Definition 3.1 is attained at some $\bar{v} \in Ty$. This attainment, combined with the fact that $\mathcal{D}_T^{\flat, h}(x, y) = 0$ yields

$$h(x, \bar{v}) - \langle x, \bar{v} \rangle = \mathcal{D}_T^{\flat, h}(x, y) = 0,$$

and we deduce that $\bar{v} \in Sx$. Hence $Ty \cap Sx \neq \emptyset$. This proves the claim in the case that $y \in \text{int}\mathcal{D}(T)$. Assume now that $x \in \text{int}\mathcal{D}(S)$. By Fact 2.2(ii) this implies that the set $S^e(x, \varepsilon)$ is weak*-compact for every $\varepsilon \geq 0$. Since $\mathcal{D}_T^{\flat,h}(x, y) = 0$, by Fact 2.6, we have that the weak*-compact sets $Ty \cap S^e(x, \varepsilon)$ are nonempty for every $\epsilon > 0$; hence the family $\{Ty \cap S^e(x, \varepsilon)\}_{\epsilon>0}$ has the finite intersection property, which implies that $Ty \cap Sx = Ty \cap \bigcap_{\epsilon>0} S^e(x, \varepsilon) = \bigcap_{\epsilon>0} (Ty \cap S^e(x, \varepsilon))$ is nonempty. Let us prove now part (b). Assume that $\mathcal{D}_T^{\sharp,h}(x, y) = 0$. The fact that $\mathcal{D}_T^{\sharp,h}(x, y) = 0$ yields

$$0 \leq h(x, v) - \langle x, v \rangle \leq 0,$$

for all $v \in Ty$, and we deduce that $v \in Sx$ without any additional hypothesis. For part (c), note that the equality $\mathcal{D}_T^h(x, y) = 0$ yields

$$h(x, Ty) = \langle x, Ty \rangle,$$

which in turn gives $Ty \in Sx$ because $h \in \mathcal{H}(S)$. $\qquad\square$

The next result characterizes solutions of $VIP(S, C)$ in terms of the new distance.

Corollary 3.8 *Let $S : X \rightrightarrows X^*$ be maximally monotone, $h \in \mathcal{H}(S)$, and fix $C \subseteq X$ a closed and convex set. Fix $x \in \text{int}\,\mathcal{D}(S) \cap C$. The following properties are equivalent.*

(i) $\mathcal{D}_{-N_C}^{\flat,h}(x, x) = 0.$
(ii) x *solves* $VIP(S, C)$.

Proof The implication (ii)→(i) follows from Remark 3.3(b) for $T = -N_C$, $x = y$ and the fact that (ii) entails the existence of $v \in -N_C(x)$ such that $v \in S(x)$. The converse follows from Proposition 3.7(a) for $T = -N_C$ and $x = y$. $\qquad\square$

Remark 3.9 We see from Proposition 3.7 that, when $x \notin \text{bdry}\mathcal{D}(S)$, having $\mathcal{D}_T^{\flat,h}(x, x) = 0$ results in a nonempty intersection of the sets Sx and Tx. Can we say something more when these distances vanish on some open set? A possible way to address this question is by using Theorem 3.10 below.

In the following theorem the maps E_T and E_S belong to $\mathbb{E}_C(T)$ and $\mathbb{E}_C(S)$, respectively. (see Fact 2.2(iv)).

Theorem 3.10 [8, Corollary 2.4, (iii)⇔(iv)] *Let $T, S : X \rightrightarrows X^*$ be two maximal monotone operators and $D \subseteq \mathcal{D}(T)$ be a nonempty open set. Then the following statements are equivalent:*

(i) $T(x) = S(x)$ *for all* $x \in D$;
(ii) $D \subseteq \text{int}\mathcal{D}(S)$ *and* $E_T(x, \varepsilon) \cap E_S(x, \varepsilon) \neq \emptyset$ *for every* $x \in D$, $\varepsilon > 0$.

We will use this theorem to establish the coincidence result between the operators.

Proposition 3.11 *Let $T, S : X \rightrightarrows X^*$ be two maximal monotone operators and let $D \subset \mathcal{D}(S) \cap \mathcal{D}(T)$ be an open set. Fix $h \in \mathcal{H}(S)$. If $\mathcal{D}_T^{\flat,h}(x, x) = 0$ for all $x \in D$, then $T = S$ in D.*

Proof From Proposition 3.7(a) we see that $Tx \cap Sx \neq \emptyset$. This fact, together with condition (E_1) in Definition 2.1 directly imply condition (ii) in Theorem 3.10. Therefore, we have $S = T$ in the open set D. □

Remark 3.12 According to [3, Theorem 9.7.2, Exercise 9.7.3], if $S : X \rightrightarrows X^*$ is a maximally monotone operator of type (NI) (in particular, if the space is reflexive), for every $h \in \mathcal{H}(S)$ we have

$$h(x, v) - \langle x, v \rangle \geq \frac{1}{4} d^2((x, v), G(S));$$

here d denotes the distance on $X \times X^*$ defined by $d((x, v), (y, w)) := \|x - y\|^2 + \|v - w\|_*^2$. Combining this fact with Definition 3.1 we obtain

$$\mathcal{D}_T^{\flat, h}(x, y) \geq \frac{1}{4} \inf_{v \in Ty} d^2((x, v), G(S)) = \frac{1}{4} d^2(\{x\} \times Ty, G(S)).$$

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 $G(S)$. This result gives an alternative proof of Proposition 3.7(a).

3.1 The Linear Case

When the operators are point to point and linear the distances can be explicitly computed. Let H be a Hilbert space and assume that $A : H \rightarrow H$ and $B : H \rightarrow H$ are linear, monotone and continuous. It is well known that such operators are automatically maximally monotone [20, p. 30]. Following [1], for a given linear monotone operator A, we define $q_A : H \rightarrow \mathbb{R}$ as

$$q_A(x) := \frac{\langle x, Ax \rangle}{2}, \tag{3.9}$$

the quadratic function associated to A. Recall that the *conjugate* of A is the linear map $A^* : H \rightarrow H$ defined by $\langle x, A^*y \rangle = \langle Ax, y \rangle$ for every $x, y \in H$.

Remark 3.13 The quadratic form q_A can be used to compute the Fitzpatrick function of a continuous linear and monotone operator.

$$\mathcal{F}_A(x, u) = \sup_{z \in H} \{\langle x, Az \rangle + \langle z, u \rangle - \langle z, Az \rangle\}$$

$$= 2 \sup_{z \in H} \left\{ \langle z, \frac{A^*x + u}{2} \rangle - q_A(z) \right\} \tag{3.10}$$

$$= 2 q_A^*(\frac{A^*x + u}{2}),$$

where the first equality follows from the definition of Fitzpatrick function.

In this section, the maps are point to point. So when the function $h \in \mathcal{H}(A)$ is chosen as the Fitzpatrick function the distance induced by two monotone linear mappings A and B will be

$$\mathcal{D}_B^{\mathcal{F}_A}(x, y) = \mathcal{F}_A(x, By) - \langle x, By \rangle. \tag{3.11}$$

We now collect some results taken from [1, Fact 2.2 and Theorem 2.3].

Fact 3.14 Let $A : H \rightarrow H$ be continuous, linear and symmetric. Then q_A is convex if and only if A is monotone. In this case, the following facts hold.

(i) $q_A^* \circ A = q_A$,

(ii) $\operatorname{ran}(A) \subset \operatorname{dom} q_A^* \subset \operatorname{cl}(\operatorname{ran} A)$,

(iii)   For all $(x, u) \in H \times H$ we have $\mathcal{F}_A^*(u, x) = \delta_{G(A)}(x, u) + \langle x, Ax \rangle$.

In the next result we compute our distance induced by A and B.

Proposition 3.15 *Let $A, B : H \to H$ be linear, monotone and continuous. Then the following hold.*

(a) $\mathcal{D}_B^{\mathcal{F}_A}(x, y) = 2q_A^*\left(\frac{A^*x + By}{2}\right) - \langle x, By \rangle$

(b)   *If A is symmetric and $\operatorname{ran} A$ is closed, then*

$$\mathcal{D}_B^{\mathcal{F}_A}(x, y) = \begin{cases} +\infty & \text{if } By \notin \operatorname{ran} A \\[2mm] 2q_A(z_0) - \langle x, By \rangle & \text{if } By \in \operatorname{ran} A, \end{cases}$$

where $z_0 \in A^{-1}\left(\frac{Ax + By}{2}\right) := \{z \in H : Az = \frac{Ax + By}{2}\}$.

(c) $\mathcal{D}_B^{\sigma_A}(x, y) = \delta_{\{0\}}(By - Ax)$.

Proof (a) This follows from (1) and (3). Part (b) follows directly from part (a) and Fact 3.14(i) and (ii) for the operator A_+ instead of A. Part (c) follows directly from Fact 3.14(iii). □

3.2 Continuity Properties

In this section we assume that X is a reflexive Banach space. Our aim is to establish lower semicontinuity properties of our distances. We show that $\mathcal{D}_T^{\flat,h}(\cdot, y)$ and $\mathcal{D}_T^{\flat,h}(x, \cdot)$ are lsc w.r.t. the strong topology in the interior of the domains. On the other hand, $\mathcal{D}_T^{\sharp,h}(\cdot, y)$ is lsc w.r.t. the weak topology at every $x \in \mathcal{D}(S)$. We also provide two examples: one showing that $\mathcal{D}_T^{\flat,h}(x, \cdot)$ is not usc in general, and the other showing that $\mathcal{D}_T^{\sharp,h}(x, \cdot)$ is not lsc in general.

Remark 3.16 In the next result, we use the Eberlein-Šmulian theorem, which states that a subset of a Banach space is weakly compact if and only if it is weakly sequentially compact (see [14, Chapter III, page 18]). We also use the fact that enlargements are locally bounded at a point which is in the interior of their domains. This provides a neighbourhood of the reference point which is norm-closed and bounded, and hence weakly compact (by Bourbaki-Alaoglu's theorem and reflexivity). We then use the Eberlein-Šmulian theorem to deduce that the given neighbourhood is in fact weakly sequentially compact. Since Lemmas 3.17 and 3.18 involve the strong topology in X, we can use sequences instead of nets.

Lemma 3.17 *Assume that $S : X \rightrightarrows X^*$ is maximally monotone and $h \in \mathcal{H}(S)$, and fix $y \in \mathcal{D}(T)$.*

(a)   *Let $T : X \rightrightarrows X^*$ be such that Tz is weakly closed for any z in its domain. Then the function $\mathcal{D}_T^{\flat,h}(\cdot, y) : X \to \mathbb{R}_\infty$ is lsc at every $x \in \operatorname{int}\mathcal{D}(S)$ with respect to the strong topology in X.*

(b)   *The function $\mathcal{D}_T^{\sharp,h}(\cdot, y) : X \to \mathbb{R}_\infty$ is lsc at every $x \in \mathcal{D}(S)$ with respect to the strong topology in X.*

Proof Assume (a) is not true. This means that there exists $a \in \mathbb{R}$ and a sequence x_n converging strongly to x such that $\mathcal{D}_T^{\flat,h}(x, y) > a$ and $\mathcal{D}_T^{\flat,h}(x_n, y) \leq a$. For n_0 large enough we have that

$$\mathcal{D}_T^{\flat,h}(x_n, y) < a + \frac{1}{n} < \mathcal{D}_T^{\flat,h}(x, y),$$

for all $n \geq n_0$. The definition of $\mathcal{D}_T^{\flat,h}$, together with the left hand side of the above expression, imply that for each fixed $n \geq n_0$, there exists $v_n \in Ty$ such that

$$h(x_n, v_n) - \langle x_n, v_n \rangle < a + \frac{1}{n}. \tag{3.12}$$

By Remark 2.7, this implies that

$$v_n \in L^h(x_n, a + \frac{1}{n}) \subset L^h(x_n, a + 1), \ \forall n \geq n_0. \tag{3.13}$$

Since $x \in \text{int}\mathcal{D}(S)$, we can use Fact 2.2(v) to deduce that the enlargement $E(\cdot, a + 1) := L^h(\cdot, a + 1)$ is locally bounded at x. This implies the existence of two closed balls, denoted by $B(x, r) \subset X$ and $B_0 \subset X^*$, respectively, such that $L^h(B(x, r), a+1) \subset B_0$. By Remark 3.16, B_0 is weakly sequentially compact. The latter fact, (3.13), and the weak sequential compactness of B_0 imply that $\{v_n\} \subset B_0$ for n large enough, and hence there is a subsequence of $\{v_n\}$ converging weakly to some vector v. Recalling now that $\{v_n\} \subset Ty$ and the set Ty is weakly closed, we deduce that $v \in Ty$. By reflexivity and Remark 2.7 , the graph of L^h is closed for the strong-weak convergence. Taking limit for n tending to infinity in (3.13) yields $v \in L^h(x, a)$. Taking limits in (3.12) (for the corresponding strong-weak convergent subsequence), and using the definition of $\mathcal{D}_T^{\flat,h}$ gives

$$\mathcal{D}_T^{\flat,h}(x, y) \leq h(x, v) - \langle x, v \rangle \leq a,$$

where we used the fact that $v \in L^h(x, a)$ in the rightmost inequality. The above expression contradicts our assumption on a, completing the proof of (a).

Assume now that $x \in \mathcal{D}(S)$ and (b) is not true. For simplicity, write $\psi(x) := \mathcal{D}_T^{\sharp,h}(x, y)$. The statement that ψ is not (strongly) lower semicontinuous at x means that there exists $a < \psi(x)$ and a sequence x_n converging strongly to x such that $\psi(x_n) \leq a$. Using the definition, this inequality implies that for every fixed $v \in Ty$ and all n we have

$$h(x_n, v) - \langle x_n, v \rangle \leq a.$$

Since $h(\cdot, v)$ is strongly lsc and $\{x_n\}$ converges strongly to x the above inequality yields

$$h(x, v) - \langle x, v \rangle \leq a.$$

Since we can do this for every $v \in Ty$ we deduce that $\mathcal{D}_T^{\sharp,h}(x, y) \leq a$, contradicting our assumptions. Hence (b) holds. $\qquad\square$

The following result establishes lower semicontinuity of $\mathcal{D}_T^{\flat,h}(x, \cdot)$. This fact is not true for $\mathcal{D}_T^{\sharp,h}(x, \cdot)$, as will be shown in Example 3.21.

Lemma 3.18 *Assume that $S : X \rightrightarrows X^*$ is maximally monotone, $h \in \mathcal{H}(S)$, and $T : X \rightrightarrows X^*$ is locally bounded in the interior of its domain. Suppose also that the graph of T is closed w.r.t. the strong-weak topology. Fix $y \in \text{int}\mathcal{D}(T)$ and $x \in \mathcal{D}(S)$. 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.*

Proof Assume the claim is not true. Since we consider here the norm topology, this means that there exists $a \in \mathbb{R}$ and a sequence y_n converging strongly to y such that $\mathcal{D}_T^{\flat,h}(x, y) > a$ and $\mathcal{D}_T^{\flat,h}(x, y_n) \leq a$. From the second inequality for all n we deduce the existence of $v_n \in Ty_n$ such that

$$h(x, v_n) - \langle x, v_n \rangle < a + \frac{1}{n}. \tag{3.14}$$

Since $y \in \mathrm{int}\mathcal{D}(T)$, T is locally bounded at y. Using now a similar argument as the one used in the proof of Lemma 3.17(a), we obtain a subsequence of $\{v_n\}$ converging weakly to some vector v. By the strong-weak closedness of the graph of T we deduce that $v \in Ty$. Using the (strong-weak) lsc of h we can write

$$a < \mathcal{D}_T^{\flat,h}(x, y) \leq h(x, v) - \langle x, v \rangle \leq \liminf_n [h(x, v_n) - \langle x, v_n \rangle] \leq a,$$

where we used also (3.14) in the last inequality. This expression entails a contradiction and therefore the claim on lower semicontinuity is true. □

Example 3.20 below shows that $\mathcal{D}_T^{\flat,h}(x, \cdot)$ may fail to be usc. In both of the next examples, we make use of the following fact (for a proof, see [19]).

Fact 3.19 Assume that X is a Banach space and $g : X \to \mathbb{R}$ is defined by $g(x) := \|x\|$. Then,

$$\partial g(0) = B,$$

where B is the closed unit ball. For every $y \neq 0$ we have

$$\partial g(y) = \{z \in X^* : \langle y, z \rangle = \|y\|\}.$$

If X is a Hilbert space, then for all $y \neq 0$ we have

$$\partial g(y) = \{\nabla g(y)\} = \{y/\|y\|\}. \tag{3.15}$$

Example 3.20 Let X be a Hilbert space with dimension at least two, and let $S := T := \partial g$, with g as in Fact 3.19. It was proved in [5] (see also [18, Example 5]) that the set $\mathcal{H}(S)$ has only one element, which is then necessarily the Fitzpatrick function, given by

$$\mathcal{F}_S(x, v) = \|x\| + \delta_B(v),$$

with δ_B denoting the indicator function of the closed unit ball of X. Thus, for $x, y \in X$ we have

$$\mathcal{D}_T^{\flat,\mathcal{F}_S}(x, y) = \inf_{v \in \partial g(y)} \{\|x\| + \delta_B(v) - \langle x, v \rangle\} = \inf_{v \in \partial g(y)} \{\|x\| - \langle x, v \rangle\}$$

$$= \|x\| - \sup_{v \in \partial g(y)} \langle x, v \rangle = \begin{cases} 0 & \text{if } y = 0 \\ \|x\| - \left\langle x, \frac{y}{\|y\|} \right\rangle & \text{if } y \neq 0 \end{cases},$$

where we used (3.15) in the last equality. If $x \neq 0$ then $\mathcal{D}_{S,T}^{\flat}(x, \cdot)$ is not usc at 0, since for every sequence $y_n \neq 0$ orthogonal to x and strongly converging to 0 one has

$$\mathcal{D}_T^{\flat}(x, y_n) = \|x\| > 0 = \mathcal{D}_{S,T}^{\flat}(x, 0).$$

Example 3.21 Assume X is a Hilbert space and fix a nonzero $x \in X$. Take $S = Id$ and $T = \partial g$, for g as in Fact 3.19. Using the second equality in (3.10) for $A = Id$ we can write

$$\mathcal{D}_T^{\sharp, \mathcal{F}_S}(x, 0) = \sup_{v \in B} \left\{ 2 \sup_{z \in X} \left\{ \langle z, \tfrac{x+v}{2} \rangle - \tfrac{\|z\|^2}{2} \right\} - \langle x, v \rangle \right\}$$

$$= \sup_{v \in B} \left\{ \|\tfrac{x+v}{2}\|^2 - \langle x, v \rangle \right\} = \tfrac{1}{4} \sup_{v \in B} \|x - v\|^2$$

Computing the supremum in the right hand side, we obtain,

$$\mathcal{D}_T^{\sharp, \mathcal{F}_S}(x, 0) = \frac{1}{4}(\|x\| + 1)^2.$$

Take now a nonzero sequence $\{y_n\}$ converging to 0. Since y_n is never zero we have from Fact 3.19 that $T y_n = \{\frac{y_n}{\|y_n\|}\}$ and hence we can write

$$\mathcal{D}_T^{\sharp, \mathcal{F}_S}(x, y_n) = 2 \sup_{z \in X} \left\{ \langle z, (x + \tfrac{y_n}{\|y_n\|})/2 \rangle - \|z\|^2/2 \right\} - \langle x, \tfrac{y_n}{\|y_n\|} \rangle.$$

As in the previous example, we take again the sequence $\{y_n\}$ orthogonal to x and tending to zero, so the expression above yields

$$\mathcal{D}_T^{\sharp, \mathcal{F}_S}(x, y_n) = \|(x + \tfrac{y_n}{\|y_n\|})/2\|^2$$

$$= \tfrac{\|x\|^2}{4} + \tfrac{1}{4}.$$

Noting that for every nonzero x we have

$$\mathcal{D}_T^{\sharp, \mathcal{F}_S}(x, y_n) = \frac{\|x\|^2}{4} + \frac{1}{4} < \frac{1}{4}(\|x\| + 1)^2 = \mathcal{D}_T^{\sharp, \mathcal{F}_S}(x, 0),$$

we conclude that $\mathcal{D}_T^{\sharp, \mathcal{F}_S}(x, \cdot)$ is not lsc at $y = 0$.

4 A Characterization of Bregman Distances

In this section we focus on the classical Bregman distance as in (3.8). Our aim is to characterize the bifunctions $D(\cdot, \cdot)$ for which there exists a convex differentiable function h such that $D = D_h$.

We say that a bifunction $G : C \times C \to X^*$, with $C \subseteq X$, is *additively separable* when there exist two functions R and P such that $G(x, y) = R(x) + P(y)$ for every $x, y \in C$.

Theorem 4.1 *Let $C \subseteq X$ be a nonempty convex set. A function $D : C \times C \to \mathbb{R}$ is a Bregman distance, that is, there exists a differentiable convex function $h : C \to \mathbb{R}$ such that $D = D_h$ if and only if D satisfies the following conditions:*

(a) D is convex and differentiable in its first argument,
(b) $D(x, x) = 0$ for every $x \in C$,
(c)   The mapping $(x, y) \mapsto \nabla_1 D(x, y)$ is additively separable,
(d) $\nabla_1 D(x, x) = 0$ for every $x \in C$.

Proof Clearly, properties (a)-(d) are satisfied by every Bregman distance. For (c), notice that $\nabla_1 D_h(x, y) = \nabla h(x) - \nabla h(y)$. Conversely, assume that properties (a)-(d) hold. By (c), there exist two mappings $R, U : C \to X^*$ such that

$$\nabla_1 D(x, y) = R(x) + U(y) \qquad \text{for every } x, y \in C. \tag{4.16}$$

From (d) it follows that $R(x) + U(x) = 0$ for every $x \in C$, hence (4.16) reduces to

$$\nabla_1 D(x, y) = R(x) - R(y) \qquad \text{for every } x, y \in C. \tag{4.17}$$

Fix $\overline{y} \in C$ and define $h := D(\cdot, \overline{y}) + \langle \cdot, R(\overline{y}) \rangle$. By (a), the function h is convex and differentiable. By (4.17), one has $\nabla h = R$, and the expression

$$\nabla(D(\cdot, y) + (\cdot, R(y)))(x) = \nabla_1 D(x, y) + R(y) = R(x) \tag{4.18}$$

depends only on x. Therefore the difference $D(x, y) + \langle x, R(y) \rangle - h(x)$ depends only on y. Indeed, by (4.18) and the equality $\nabla h = R$, we have

$$\nabla_x [D(x, y) + \langle x, R(y) \rangle - h(x)] = R(x) - R(x) = 0.$$

Hence, $D(x, y) + \langle x, R(y) \rangle - h(x) = k(y)$ for some function $k : C \to \mathbb{R}$. Using this function k, we can write for every $x, y \in C$

$$\begin{aligned} &h(x) - h(y) - \langle x - y, \nabla h(y) \rangle \\ &= D(x, y) + \langle x, R(y) \rangle - k(y) - (D(y, y) + \langle y, R(y) \rangle - k(y)) - \langle x - y, \nabla h(y) \rangle, \end{aligned}$$

which, in view of (b) and the fact that $\nabla h = R$, reduces to

$$h(x) - h(y) - \langle x - y, \nabla h(y) \rangle = D(x, y).$$

Hence $D = D_h$, as claimed. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 4.2 One can easily prove that h is determined by D up to addition with an affine function, that is, $D_{h_1} = D_{h_2}$ if and only if $h_1 - h_2$ is affine.

In view of Theorem 4.1 and Remark 4.2, one obtains

Corollary 4.3 *Let $C \subseteq X$ be a nonempty convex set, γ_C be the set of differentiable convex functions on C, and $\sim$ be the equivalence relation on γ_C defined by*

$$h_1 \sim h_2 \Leftrightarrow h_1 - h_2 \text{ is affine.}$$

Then the mapping $[h] \mapsto D_h$ is a well defined bijection from the quotient set $\gamma_C / \sim$ onto the set of functions $D : C \times C \to \mathbb{R}$ satisfying conditions (a)-(d) of Theorem 4.1.

We now give an alternative characterization.

Theorem 4.4 *Let $C \subseteq X$ be a convex set. A function $D : C \times C \to \mathbb{R}$ is a Bregman distance if and only if it satisfies conditions (a) and (b) of Theorem 4.1 and*
(e) $\nabla_1 D(x, y) + \nabla_1 D(y, z) = \nabla_1 D(x, z)$ for every $x, y, z \in C$.

Proof Clearly, in view of the equality $\nabla_1 D_h(x, y) = \nabla h(x) - \nabla h(y)$, condition (e) holds. Conversely, assume that properties (a), (b) and (e) hold. Fix $z \in C$ and define $h := D(\cdot, z)$. By (a), the function h is convex and differentiable. Using (e), for $x, y \in C$ one obtains

$$\begin{aligned} \nabla h(x) &= \nabla_1 D(x, z) = \nabla_1 D(x, y) + \nabla_1 D(y, z) = \nabla_1 D(x, y) + \nabla h(y) \\ &= \nabla(D(\cdot, y) + \langle \cdot, \nabla h(y) \rangle)(x). \end{aligned}$$

The above equality implies that

$$\nabla_x (h(x) - D(x, y) - \langle x, \nabla h(y) \rangle) = 0.$$

In other words, there exists a function k depending only on y such that

$$k(y) := h(x) - D(x, y) - \langle x, \nabla h(y) \rangle.$$

Because k does not depend on x, we can replace x for y in the preceding expression:

$$k(y) = h(y) - D(y, y) - \langle y, \nabla h(y) \rangle = h(y) - \langle y, \nabla h(y) \rangle,$$

where we also used (b). We thus have

$$
\begin{aligned}
D(x, y) &= h(x) - \langle x, \nabla h(y) \rangle - k(y) \\
&= h(x) - \langle x, \nabla h(y) \rangle - (h(y) - \langle y, \nabla h(y) \rangle),
\end{aligned}
$$

which shows that $D = D_h$. $\qquad\square$

Corollary 4.5 *Let $C \subseteq X$ be a convex set, γ_C the set of differentiable convex functions on C, and $\sim$ the equivalence relation on γ_C defined by*

$$h_1 \sim h_2 \Leftrightarrow h_1 - h_2 \text{ is affine.}$$

Then the mapping $[h] \mapsto D_h$ is a well defined bijection from the quotient set $\gamma_C / \sim$ onto the set of functions $D : C \times C \to \mathbb{R}$ satisfying conditions (a) and (b) of Theorem 4.1 and (e) of Theorem 4.4.

References

1. Bauschke, H.H., Borwein, J.M., Wang, X.: Fitzpatrick function and continuous linear operators. SIAM J. Optim. **18**(3), 789–809 (2007)
2. Borwein, J.M., Fitzpatrick, S., Vanderwerff, J.D.: Examples of convex functions and classifications of normed spaces. J. Convex Anal. **1**(1), 61–73 (1994)
3. Borwein, J.M., Vanderwerff, J.D.: Convex functions: Constructions, characterizations and counterexamples. Cambridge University Press, Cambridge (2010)
4. Bueno, O., Martínez-Legaz, J.E., Svaiter, B.F.: On the monotone polar and representable closures of monotone operators. J. Convex Anal **21**(2), 495–505 (2014)
5. Burachik, R.S., Fitzpatrick, S.: On the Fitzpatrick family associated to some subdifferentials. J. Nonlinear Convex Anal. **6**(1), 165–171 (2005)
6. Burachik, R.S., Iusem, A.N.: Set valued mappings and enlargements of monotone operators. Springer Optimization and Its Applications, vol. 8. Springer, New York (2008)
7. Burachik, R.S., Iusem, A.N., Svaiter, B.F.: Enlargements of maximal monotone operators with application to variational inequalities. Set-Valued Anal. **5**, 159–180 (1997)
8. Burachik, R., Martínez-Legaz, J.E., Rocco, M.: On a sufficient condition for equality of two maximal monotone operators. Set-Valued Variational Anal. **18**(3-4), 327–335 (2010)
9. Burachik, R.S., Sagastizábal, C.A., Svaiter, B.F.: ε-enlargements of maximal monotone operators: theory and applications. In: Fukushima, M., Qi, L. (eds.) Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, pp. 25–43. Kluwer, Dordrecht (1997)
10. Burachik, R.S., Sagastizábal, C.A., Svaiter, B.F.: Bundle methods for maximal monotone operators. In: Tichatschke, R., Théra, M. (eds.) Ill–Posed Variational Problems and Regularization Techniques, pp. 49–64. Springer, Berlin (1999)
11. Burachik, R.S., Svaiter, B.F.: ε-enlargements of maximal monotone operators in Banach spaces. Set-Valued Anal. **7**, 117–132 (1999)
12. Burachik, R.S., Svaiter, B.F.: Maximal monotone operators, convex functions and a special family of enlargements. Set-Valued Anal. **10**(4), 297–316 (2002)
13. Burachik, R.S., Svaiter, B.F.: Enlargements of monotone operators: new connection with convex functions. Pacific Journal of Optimization **2**(3), 425–445 (2006)
14. Diestel, J.: Sequences and series in banach spaces, graduate texts in mathematics, Springer-Verlag (1984)

15. Fitzpatrick, S.: Representing monotone operators by convex functions, pp. 59–65. Functional Analysis and Optimization, Workshop and Miniconference, Australia (1988). Proc. Center Math. Anal. Australian Nat. Univ. 20
16. Kiwiel, K.: Proximal minimization methods with generalized bregman functions. SIAM J. Control Optim. **35**(4), 1142–1168 (1997)
17. Martínez-Legaz, J.E., Svaiter, B.F.: Monotone operators representable by l.s.c. convex functions. Set-Valued Anal. **13**(1), 21–46 (2005)
18. Penot, J.P.: The relevance of convex analysis for the study of monotonicity. Nonlinear Anal. **58**(7-8), 855–871 (2004)
19. Phelps, R.R.: Convex functions, monotone operators and differentiability. Springer, New York (2013)
20. Simons, S.: Minimax and monotonicity. Springer, Berlin (1998)
21. Simons, S.: Positive sets and monotone sets. J. Convex Anal. **14**(2), 297–317 (2007)
22. Svaiter, B.F.: A family of enlargements of maximal monotone operators. Set-Valued Anal. **8**, 311–328 (2000)