# DUALITY, FRÉCHET DIFFERENTIABILITY AND BREGMAN DISTANCES IN HYPERBOLIC SPACES

NICHOLAS PISCHKE

Department of Mathematics, Technische Universität Darmstadt,
Schlossgartenstraße 7, 64289 Darmstadt, Germany,

E-mail: pischke@mathematik.tu-darmstadt.de

Abstract. For general hyperbolic metric spaces, we introduce a new notion of a dual system (extending the influential notion from the context of normed linear spaces) that allows for a uniform study of different notions of duality for these nonlinear spaces. Using this abstract notion of duality, we lift various notions from convex analysis into this nonlinear setting, including Fréchet differentiability and Bregman distances. Further, we introduce a notion of a monotone operator relative to a given dual system and, using the new Fréchet derivatives, we study corresponding resolvents relative to a given gradient, generalizing the seminal notion of Eckstein from the linear setting. These resolvents are then related to corresponding notions of Bregman nonexpansive mappings which are introduced relative to this generalization of the classical Bregman distance and we prove a convergence result of an analogue of the proximal point algorithm. For that, using methods from proof mining, we even provide quantitative results on its convergence in very general settings.

**Keywords:** Hyperbolic Spaces, Bregman Distances, Fréchet Derivatives, Monotone Operators, Proximal Point Algorithm, Proof Mining.
**MSC2010 Classification:** 47J25, 47H09, 47H10, 03F10, 47H05, 54E40

## 1. Introduction

In the context of CAT(0)-spaces, one of the most influential classes of nonlinear spaces, a few approaches for providing a sensible definition for a dual space of such a nonlinear space have been investigated in the recent years (see e.g. [18, 22, 24] among others). Besides of the general intellectual interest of extending the duality theory of linear spaces, one of the cornerstones of functional analysis, to this nonlinear context, one could hope that such a notion would in general also provide the theoretical basis for lifting some of the many influential applied notions from e.g. convex analysis and optimization to such a nonlinear setting and open up new areas and methods for applications. As such, in [18, 24] for example, monotone operators in the sense of Browder [10, 11] have been extended from a linear to this nonlinear setting of CAT(0)-spaces and this, together with a corresponding extended notion of a resolvent as given in [24], gave rise to new fixed point iterations modeled in analogy of the famous proximal point algorithm (originally due to Rockafellar [55] and Martinet [42]) in these nonlinear contexts. Further, these objects provided a new medium by which, for example, the influential notion of proximal mappings as introduced by Jost [21] can be studied (see the discussion in [24]).

However, many of the influential concepts from convex analysis still lack a proper analogue in the context of these duality notions. The prime example is that of the Fréchet derivative of a convex function, as one would thus hope that a suitable lift of this notion would then in particular also allow for the generalization of many of the influential notions associated with

*Date*: July 2, 2024.

it, such as e.g. Bregman distances, to be portable to this nonlinear setting.

In this paper, we at first introduce the notion of a dual system for hyperbolic spaces. These dual systems are inspired by the classical notion of dual systems from topological vector spaces (see e.g. [56] for a standard reference) and allow for a uniform and abstract study of duality in this nonlinear context. In particular, these dual systems also provide a uniform way to study the different notions of dual spaces proposed for CAT(0)-spaces mentioned before. In the context of such a dual system, we then introduce a notion of (uniform) Fréchet differentiability that allows for a wide range of the associated theory from convex analysis over linear spaces (we refer to standard references [54, 61]) to be carried out in these nonlinear spaces. As a prime example, we use this notion of a Fréchet derivative to extend the influential Bregman distance [9] to this hyperbolic context. This seems to be the first notion of a Bregman distance in this nonlinear context and we use it to further generalize corresponding notions from fixed point theory like various notions of Bregman nonexpansive maps (see e.g. [35, 36, 39, 40, 41, 48, 49, 50, 51]). Further, we introduce a notion of monotone operator relative to a dual system, generalizing the notion from CAT(0)-spaces introduced in [24]. In the context of these operators, we use our new notion of Fréchet derivatives to define resolvents relative to a convex function $f$ and its gradient, extending the influential definition from the normed case as introduced by Eckstein [17] and generalizing the previous resolvents for monotone operators on CAT(0)-spaces from [24]. We then use these relativized resolvents to study a proximal point type method for which we prove a "strong" (i.e. in the sense of the metric) convergence theorem in proper hyperbolic spaces and, under a regularity assumption, even such a convergence theorem in the absence of any compactness assumptions. Further, these convergence results are actually derived from more general quantitative results that we establish first which even provide (quasi-)rates for these convergences. The convergence proof for this method uses that the iteration thus defined is in particular Fejér monotone w.r.t. these new Bregman distances. As such, these iterations therefore fit into the recent abstract framework developed for the convergence of generalized Fejér monotone sequences [45] which in particular allows for the treatment of sequences with Fejér-type properties formulated using distance functions which are not metrics and which do not operate in the context of a normed setting (thereby generalizing many of the abstract works on convergence of Fejér methods, see e.g. [4, 33, 34]). Throughout, we show how these abstract notions, all formulated relative to a given dual system, instantiate in the context of the most influential dual space notion in CAT(0)-spaces introduced in [22].

For all of these results, we want to note that the paper stems from recent insights [46] into the logical properties of all these notions from convex analysis in the context of normed spaces together with recent applications [47], as facilitated by the so-called proof mining program, a program in mathematical logic going back to Georg Kreisel's unwinding of proofs [37, 38] and brought to maturity by U. Kohlenbach and his collaborators, that aims to classify and extract the computational content of prima facie 'non-computational' proofs (see [29] for a comprehensive monograph on the subject and [31, 32] for surveys). While this logical background was instrumental for deriving the notions and results given here, the paper itself does not require any knowledge of logic or proof mining and we only comment on logical aspects in small remarks. We however want to emphasize that, in that way, the present paper, further illustrates the usefulness of analyses of notions and proofs provided by the proof mining program for deriving wholly new notions and results.

Lastly, we want to mention that beyond the notions introduced and discussed in this paper, many prevalent and important topics from convex analysis still are left to be investigated in the context of these nonlinear spaces, the newly introduced dual systems and the notion of Fréchet differentiability discussed here. Among them are in particular the notion of a Fenchel conjugate (which could be introduced for general dual systems by following the approach given in [22] for CAT(0)-spaces), the concept of Legendre functions and the relationship between notions like sequential consistency and uniform as well as total convexity in this context (see e.g. [2, 3, 15, 53] for some canonical references for these topics in the context of linear spaces). Also, we only discuss "strongly convergent" iterations, i.e. iterations that converge in the sense of the metric, in this paper but it seems conceivable that some of the results can be generalized to broader contexts if one would introduce a suitable notion of weak convergence relative to dual systems (possibly akin to how a notion of weak convergence in CAT(0)-spaces is introduced in [22]).

## 2. Preliminaries

In this section, we introduce the main spaces from nonlinear analysis relevant for this paper. In that context, we mostly follow the presentation from [27].

**Definition 2.1** ([27])**.** A triple $(X, d_X, W_X)$ is called a hyperbolic space if $(X, d_X)$ is a metric space and $W_X : X \times X \times [0,1] \to X$ is a function satisfying

(i) $\forall x, y, z \in X \forall \lambda \in [0,1] \, (d_X(z, W_X(x,y,\lambda)) \leqslant (1-\lambda) d_X(z,x) + \lambda d_X(z,y))$.
(ii) $\forall x, y \in X \forall \lambda_1, \lambda_2 \in [0,1] \, (d_X(W_X(x,y,\lambda_1), W_X(x,y,\lambda_2)) = |\lambda_1 - \lambda_2| \cdot d_X(x,y))$.
(iii) $\forall x, y \in X \forall \lambda \in [0,1] \, (W_X(x,y,\lambda) = W_X(y,x,1-\lambda))$.
(iv) $\forall x, y, z, w \in X \forall \lambda \in [0,1] \, (d_X(W_X(x,z,\lambda), W_X(y,w,\lambda)) \leqslant (1-\lambda) d_X(x,y) + \lambda d_X(z,w))$.

We refer to [27] for a discussion on the relationship of this notion to other influential definitions of hyperbolic spaces in nonlinear analysis like e.g. Takahashi's convex metric spaces [58], spaces of hyperbolic type in the sense of Goebel and Kirk [20] or the hyberbolic spaces of Reich and Shafrir [52] or Kirk [25].

Before moving on, we fix some notation regarding balls and bounded sets: given a point $o \in X$ and $b > 0$, we define

$$\overline{B}_b(o) := \{x \in X \mid d_X(o,x) \leqslant b\}$$

and we say that a set $A \subseteq X$ is $o$-bounded if there exists a $b > 0$ such that $A \subseteq \overline{B}_b(o)$.

**Definition 2.2.** Let $(X, d_X)$ be a metric space. A geodesic in $X$ is a map $\gamma : [0, I] \to X$ such that

$$d_X(\gamma(t), \gamma(s)) = |t - s| \text{ for all } t, s \in [0, I].$$

Image sets $\gamma([0, I])$ of geodesics $\gamma$ are called geodesic segments and we say that the points $x = \gamma(0)$ and $y = \gamma(I)$ are joined by the geodesic segment (which entails $I = d_X(x,y)$). The space $(X, d_X)$ is a geodesic space if every two points in $X$ are joined by a geodesic segment.

Naturally, any hyperbolic space is a geodesic space with

$$\{W_X(x,y,\lambda) \mid \lambda \in [0,1]\}$$

being a geodesic segment joining $x$ and $y$ which arises from $[0, d_X(x,y)]$ as the image of the geodesic defined by

$$\gamma(\alpha) = W_X\left(x, y, \frac{\alpha}{d_X(x,y)}\right).$$

**Definition 2.3.** A CAT(0)-space is a geodesic space that satisfies the so-called CN-inequality of Bruhat and Tits [13], i.e.

$$\mathrm{CN}: \begin{cases} \forall x, y_0, y_1, y_2 \in X( & d_X(y_0,y_1) = \frac{1}{2}d_X(y_1,y_2) = d_X(y_0,y_2) \text{ implies} \\ & d^2_X(x,y_0) \leqslant \frac{1}{2}d^2_X(x,y_1) + \frac{1}{2}d^2_X(x,y_2) - \frac{1}{4}d^2_X(y_1,y_2) \quad ). \end{cases}$$

Every CAT(0)-space is a uniquely geodesic space, i.e. every two points are joined by a unique geodesic segment. Clearly any hyperbolic space, being a geodesic space, that satisfies the CN-inequality is a CAT(0)-space. Conversely, as every CAT(0)-space is a uniquely geodesic space, any CAT(0)-space is also a hyperbolic space by setting $W_X(x,y,\lambda) = \gamma(\lambda d_X(x,y))$ with $\gamma$ being the unique geodesic with $\gamma(0) = x$ and $\gamma(d_X(x,y)) = y$ (see [26]). Hence being a CAT(0)-space is the same as being a hyperbolic space satisfying the CN-inequality and we shall rely on the latter characterization as the underlying definition of CAT(0)-spaces in this paper.

Another characterization of CAT(0)-spaces arises through the use of the so-called quasi-linearization function introduced by Berg and Nikolaev [8]. This function, emulating an inner product in CAT(0)-spaces (at least in certain ways), is defined on pairs from $X$, denoted by $\overrightarrow{ab}, \overrightarrow{cd} \in X^2$, via the following formula:

$$\langle \overrightarrow{ab}, \overrightarrow{cd} \rangle := \frac{1}{2}\left(d^2_X(a,d) + d^2_X(b,c) - d^2_X(a,c) - d^2_X(b,d)\right).$$

As discussed in [8], this function is the unique function $X^2 \times X^2 \to \mathbb{R}$ in any metric space satisfying the following four properties for all $x, y, u, v \in X$:

(I) $\langle \overrightarrow{xy}, \overrightarrow{xy} \rangle = d^2_X(x,y)$,
(II) $\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle = \langle \overrightarrow{uv}, \overrightarrow{xy} \rangle$,
(III) $\langle \overrightarrow{yx}, \overrightarrow{uv} \rangle = -\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle$,
(IV) $\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle + \langle \overrightarrow{xy}, \overrightarrow{vw} \rangle = \langle \overrightarrow{xy}, \overrightarrow{uw} \rangle$.

As further shown in [8], a geodesic metric space (i.e. in particular a hyperbolic space) is a CAT(0)-space if, and only if, the quasi-linearization function satisfies the following analog of the Cauchy-Schwarz inequality:

$$\langle \overrightarrow{ab}, \overrightarrow{cd} \rangle \leqslant d_X(a,b)d_X(c,d) \text{ for all } a,b,c,d \in X.$$

At last, just a word on a general type of notation that we will use: In the later parts of the paper, we will often have to convert between errors represented by general real numbers $\varepsilon > 0$ and errors of the form $1/(k+1)$ for $k \in \mathbb{N}$. In that context, for a given function $\varphi : (0,\infty) \to (0,\infty)$, we write

$$\widehat{\varphi}(k) := \left\lceil \frac{1}{\varphi\left(\frac{1}{k+1}\right)} \right\rceil.$$

This function then has the property that

$$\frac{1}{\widehat{\varphi}(k)+1} < \varphi\left(\frac{1}{k+1}\right).$$

## 3. Dual systems of metric spaces

In this section, we now define the first main new notion of this paper, the so-called dual systems for metric spaces, which provide an abstract account of duality in this nonlinear context. In particular, we will later discuss concrete examples of dual spaces of hyperbolic spaces and show how they instantiate this abstract notion. The idea of the dual systems for metric space, which is a pair of spaces together with a function representing a sort of application of elements from one space to elements of the other, is conceptually similar to the notion of dual systems known from topological vector spaces (see e.g. [56] for a standard reference).

Throughout, given a metric space $(X, d_X)$, we fix an arbitrary point $o \in X$ acting as a center.

**Definition 3.1.** Let $(X, d_X)$ and $(Y, d_Y)$ be two metric spaces. We call $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ with

$$\langle\cdot,\cdot\rangle : X^2 \times Y \to \mathbb{R}$$

a *dual system* if the following conditions are satisfied:

(1) $\langle\overrightarrow{xy}, x^*\rangle + \langle\overrightarrow{yz}, x^*\rangle = \langle\overrightarrow{xz}, x^*\rangle$ for all $x, y, z \in X$ and $x^* \in Y$.
(2) $\langle\overrightarrow{xy}, x^*\rangle = -\langle\overrightarrow{yx}, x^*\rangle$ for all $x, y \in X$ and $x^* \in Y$.
(3) $|\langle\overrightarrow{xy}, x^*\rangle - \langle\overrightarrow{xy}, y^*\rangle| \leqslant d_X(x, y)d_Y(x^*, y^*)$ for all $x, y \in X$ and $x^*, y^* \in Y$.
(4) There exists an element $\mathcal{O} \in Y$ such that $\langle\overrightarrow{xy}, \mathcal{O}\rangle = 0$ for all $x, y \in X$.

Here, we wrote $\overrightarrow{xy}$ for $(x, y) \in X^2$ (similar as in the preliminaries).

The space $(Y, d_Y)$ serves as an abstract dual for the space $(X, d_X)$, which we sometimes call a dual companion of $X$. Further, we call $\langle\cdot,\cdot\rangle$ an *action* or a *pairing*.

In the context of these dual systems we use the following notation (in similarity to [22]): For a dual system $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$, we write $\operatorname{span} Y$ for the set of all formal sums $\sum_{i=1}^n \alpha_i x_i^*$ where $\alpha_i \in \mathbb{R}$ as well as $x_i^* \in Y$ for $i = 1, \ldots, n$. Then, for these formal sums, we define

$$\langle\overrightarrow{xy}, \sum_{i=1}^n \alpha_i x_i^*\rangle := \sum_{i=1}^n \alpha_i \langle\overrightarrow{xy}, x_i^*\rangle$$

for $\overrightarrow{xy} \in X^2$. With that notation, condition (3) of Definition 3.1 can be equivalently rewritten as

$$|\langle\overrightarrow{xy}, x^* - y^*\rangle| \leqslant d_X(x, y)d_Y(x^*, y^*) \text{ for all } x, y \in X \text{ and } x^*, y^* \in Y.$$

Further, for $x^*, y^* \in \operatorname{span} Y$, we write $x^* =_{\mathcal{D}} y^*$ if

$$\langle\overrightarrow{xy}, x^*\rangle = \langle\overrightarrow{xy}, y^*\rangle \text{ for all } x, y \in X.$$

A feature commonly required in dual systems $(X, Y, \langle\cdot,\cdot\rangle)$ of normed vector spaces $X, Y$ is that of non-degeneracy (see e.g. [56]), i.e. that

$$\langle x, x^*\rangle = 0 \text{ for all } x^* \in Y \text{ implies } x = 0,$$
$$\langle x, x^*\rangle = 0 \text{ for all } x \in X \text{ implies } x^* = 0.$$

An appropriately translated variant of the second property will turn out to be key for some later investigations. Concretely, we in analogy want to introduce the following notion of non-degenerateness for dual systems of metric spaces:

**Definition 3.2.** Let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ be a dual system. We call $\mathcal{D}$ *non-degenerate* if

$$\forall x, y \in X\, (\langle\overrightarrow{xy}, x^* - y^*\rangle = 0) \text{ implies } d_Y(x^*, y^*) = 0$$

for all $x^*, y^* \in Y$. We call $\mathcal{D}$ *uniformly non-degenerate* if for all $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $x^*, y^* \in Y$:

$$\forall x \neq y \in X\, (|\langle\overrightarrow{xy}, x^* - y^*\rangle| < \delta d_X(x, y)) \text{ implies } d_Y(x^*, y^*) < \varepsilon.$$

We call a function $\Delta(\varepsilon)$ bounding (viz. witnessing) such a $\delta$ in terms of $\varepsilon$ a *modulus of uniform non-degenerateness for $\mathcal{D}$*.

*Remark* 3.3. Note that in a dual system $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$, for $x^*, y^* \in Y$ with $x^* = y^*$, it follows by condition (3) that

$$|\langle\overrightarrow{xy}, x^* - y^*\rangle| \leqslant d_X(x,y)d_Y(x^*, y^*) = 0,$$

i.e. $x^* =_{\mathcal{D}} y^*$. Now, non-degenerateness as defined above is just stipulating that the converse of this implication also holds true, i.e. that $x^* =_{\mathcal{D}} y^*$ implies $x^* = y^*$.

As for examples of dual systems, clearly any dual system $(X, Y, \langle\cdot,\cdot\rangle)$ in the usual sense of normed linear spaces $X$ and $Y$ can be reformulated as a dual system of metric spaces by defining

$$\langle\overrightarrow{xy}, x^*\rangle = \langle y - x, x^*\rangle.$$

Further, it encompasses the notion of a dual space in CAT(0)-spaces $X^*$ from [22] as well as its linearization $X^\diamond := \operatorname{span} X^*$ considered e.g. in [18]. The following example discusses this explicitly for $X^*$.

*Example* 3.4. Consider $X^*$ defined for a metric space $(X, d, W)$ as in [22]: Given $t \in \mathbb{R}$ and $a, b \in X$, define $\Theta(t, a, b) \in C(X)$ (the space of continuous functions $X \to \mathbb{R}$) by

$$\Theta(t,a,b)(x) := t\langle\overrightarrow{ab}, \overrightarrow{ax}\rangle$$

where $\langle\cdot,\cdot\rangle$ is now the quasi-linearization function of Berg and Nikolaev as mentioned before. As the quasi-linearization function in a CAT(0)-space satisfies the Cauchy-Schwarz inequality, the function $\Theta$ is a Lipschitz function and if

$$L(\varphi) := \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d(x,y)}$$

is the Lipschitz semi-norm of a function $\varphi : X \to \mathbb{R}$, then $L(\Theta(t, a, b)) = |t|d(a, b)$. On $\mathbb{R} \times X \times X$, one defines the pseudometric

$$D((t,a,b),(s,c,d)) := L(\Theta(t,a,b) - \Theta(s,c,d)).$$

The space $X^*$ is now defined as the set of equivalence classes

$$[t\,\overrightarrow{ab}] := \{(s,c,d) \mid D((t,a,b),(s,c,d)) = 0\}.$$

Writing $x^*$ for $[t\,\overrightarrow{ab}]$, we can define an action of $x^*$ on $X^2$ by

$$\langle\overrightarrow{xy}, x^*\rangle := t\langle\overrightarrow{xy}, \overrightarrow{ab}\rangle$$

which can easily be shown to be well-defined. We now show that $\mathcal{D}^* = (X, X^*, \langle\cdot,\cdot\rangle)$ (where $\langle\cdot,\cdot\rangle$ is the above action) is a dual system. To see that, we have to verify the properties (1) – (4) of Definition 3.1. Property (4) is immediate by taking e.g. $\mathcal{O} = [1\,\overrightarrow{oo}]$ for the arbitrarily designated point $o \in X$. Properties (1) and (2) are immediate from the main properties of $\langle\cdot,\cdot\rangle$ exhibited in [8] (recall also Section 2). So we only discuss property (3). For this, note that for

$x^*, y^* \in X^*$ with say $x^* = [t\,\overrightarrow{x_1^* x_2^*}]$ and $y^* = [s\,\overrightarrow{y_1^* y_2^*}]$, it holds that

$$\begin{aligned}
D(x^*, y^*) &= L(\Theta(t, x_1^*, x_2^*) - \Theta(s, y_1^*, y_2^*)) \\
&= \sup_{a \neq b} \frac{|t\langle \overrightarrow{x_1^* a}, \overrightarrow{x_1^* x_2^*}\rangle - s\langle \overrightarrow{y_1^* a}, \overrightarrow{y_1^* y_2^*}\rangle - t\langle \overrightarrow{x_1^* b}, \overrightarrow{x_1^* x_2^*}\rangle + s\langle \overrightarrow{y_1^* b}, \overrightarrow{y_1^* y_2^*}\rangle|}{d(a,b)} \\
&= \sup_{a \neq b} \frac{|t(\langle \overrightarrow{x_1^* a}, \overrightarrow{x_1^* x_2^*}\rangle + \langle \overrightarrow{b x_1^*}, \overrightarrow{x_1^* x_2^*}\rangle) - s(\langle \overrightarrow{y_1^* a}, \overrightarrow{y_1^* y_2^*}\rangle + \langle \overrightarrow{b y_1^*}, \overrightarrow{y_1^* y_2^*}\rangle)|}{d(a,b)} \\
&= \sup_{a \neq b} \frac{|t\langle \overrightarrow{ba}, \overrightarrow{x_1^* x_2^*}\rangle - s\langle \overrightarrow{ba}, \overrightarrow{y_1^* y_2^*}\rangle|}{d(a,b)} \\
&= \sup_{a \neq b} \frac{|\langle \overrightarrow{ba}, x^*\rangle - \langle \overrightarrow{ba}, y^*\rangle|}{d(a,b)}.
\end{aligned}$$

Therefore, we in particular have

$$|\langle \overrightarrow{ba}, x^*\rangle - \langle \overrightarrow{ba}, y^*\rangle| \leqslant D(x^*, y^*) d(a,b) \text{ for all } a, b \in X$$

and our version of the Cauchy-Schwarz inequality for dual systems in property (3) holds for $X^*$ (which has to be contrasted however to the Cauchy-Schwarz inequality in $X$ in the sense of CAT(0)-spaces).

Lastly, $\mathcal{D}^*$ is even uniformly non-degenerate. To see that note that as before

$$D(x^*, y^*) = \sup_{a \neq b} \frac{|\langle \overrightarrow{ba}, x^*\rangle - \langle \overrightarrow{ba}, y^*\rangle|}{d(a,b)}$$

and so, given $\varepsilon > 0$, there exists $a \neq b$ such that

$$D(x^*, y^*) - \frac{|\langle \overrightarrow{ba}, x^*\rangle - \langle \overrightarrow{ba}, y^*\rangle|}{d(a,b)} < \frac{\varepsilon}{2}.$$

So, if we pick $\Delta(\varepsilon) := \varepsilon/2$ and assume that for $x^*, y^*$, we have

$$\forall x \neq y \in X\, (|\langle \overrightarrow{xy}, x^* - y^*\rangle| < \Delta(\varepsilon) d(x,y))$$

then we clearly get

$$D(x^*, y^*) - \frac{\varepsilon}{2} = D(x^*, y^*) - \frac{\Delta(\varepsilon) d(a,b)}{d(a,b)} < D(x^*, y^*) - \frac{|\langle \overrightarrow{ba}, x^*\rangle - \langle \overrightarrow{ba}, y^*\rangle|}{d(a,b)} < \frac{\varepsilon}{2}$$

and so we have $D(x^*, y^*) < \varepsilon$.

While these notions of dual spaces for nonlinear spaces are defined explicitly in terms on the underlying structure of $X$ using the quasi-linearization function $\langle \cdot, \cdot \rangle$, the general notion of dual systems of course admits more abstractly described duals like e.g. in the following example:

*Example* 3.5. Define a dual $X^\times$ by

$$X^\times := \left\{ f : X^2 \to \mathbb{R} \mid f \text{ satisfies } \begin{cases} \forall x, y \in X\, (f(\overrightarrow{xy}) = -f(\overrightarrow{yx}))\,, \\ \forall x, y \in X\, (f(\overrightarrow{xy}) + f(\overrightarrow{yz}) = f(\overrightarrow{xz}))\,, \\ \exists C \geqslant 0 \forall x, y \in X\, (|f(\overrightarrow{xy})| \leqslant C d_X(x,y))\,. \end{cases} \right\}.$$

When we set

$$\|f\|_{X^\times} := \inf\{C \geqslant 0 \mid |f(\overrightarrow{xy})| \leqslant C d_X(x,y) \text{ for all } x, y \in X\}$$

for $f \in X^\times$ as well as $d_{X^\times}(f,g) = \|f-g\|_{X^\times}$ where $f-g$ is understood to be defined pointwise, then (after moving to equivalence classes under $\|\cdot\|_{X^\times}$), $X^\times$ becomes a metric space and $X$ together with $X^\times$ satisfy the axioms of a dual system with a pairing simply defined by function application.

## 4. Subdifferentiability and Fréchet derivatives in dual systems

Let $(X, d_X, W_X)$ now be a hyperbolic space. In this and the following sections, we will care for carrying out convex analysis on dual systems in analogy to how convex analysis is carried out on normed spaces and their duals. For this, let $f : X \to (-\infty, +\infty]$ be a proper convex function, with convex meaning

$$f(W_X(x,y,\lambda)) \leqslant (1-\lambda)f(x) + \lambda f(y)$$

for all $x, y \in X$ and $\lambda \in [0,1]$ and proper meaning

$$\mathrm{dom} f := \{x \in X \mid f(x) < +\infty\} \neq \varnothing.$$

In normed spaces, we have a sensible and rather beautiful theory for the differentiability of such functions, in particular regarding the central notion of Fréchet derivatives (we refer to [61] for a standard reference on the subject in the context of Banach spaces and to [5] for a standard reference in Hilbert spaces). The theory, as it is usually developed, however crucially relies on the linear structure of the underlying spaces which is already reflected in the way Fréchet derivatives are defined using certain limits. So if one naively tries to transfer these notions and the corresponding theory to a nonlinear setting, one immediately encounters a wide range of issues. However, we will in the following see how a rather nice theory of Fréchet differentiability can be developed using the above dual systems.

For this, we start with the main object in the context of differentiability of convex functions: the subgradient. This poses no immediate difficulties, as was already observed in [22] in the context of the concrete dual $X^*$ for a CAT(0)-space $X$ (recall Example 3.4), and for the following definition, we essentially just abstracted the notion of subgradient introduced there to our notion of dual systems:

**Definition 4.1.** Let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ be a dual system and $f : X \to (-\infty, +\infty]$ be any function. We define the subgradient of $f$ relative to $\mathcal{D}$ as follows:

$$\partial_{\mathcal{D}} f(x) := \{x^* \in Y \mid f(y) - f(x) \geqslant \langle \overrightarrow{xy}, x^* \rangle \text{ for all } y \in X\}$$

for $x \in \mathrm{dom} f$, and $\partial_{\mathcal{D}} f(x) := \varnothing$, otherwise.

We write $x \in \mathrm{dom}\partial_{\mathcal{D}} f$ if $\partial_{\mathcal{D}} f(x) \neq \varnothing$. Clearly we have $\mathrm{dom}\partial_{\mathcal{D}} f \subseteq \mathrm{dom} f$ by definition.

To arrive at a useful notion of Fréchet differentiability, we turn to normed linear spaces for inspiration. In Banach spaces with the usual notion of a subgradient $\partial f$ for a convex function, Fréchet differentiability is uniquely characterized as follows:

**Theorem 4.2** (folklore (essentially [54]), see also [61, Theorem 3.3.2] and [5, Proposition 17.41] (for Hilbert spaces))**.** *Let $(X, \|\cdot\|)$ be a Banach space and let $f : X \to (-\infty, \infty]$ be a proper, lower-semicontinuous and convex function which is continuous at $x \in \mathrm{dom} f$. Then $f$ is Fréchet differentiable at $x$ if, and only if, there exists a selection $X \to X^*$ of $\partial f$ which is norm-to-norm continuous in $x$.*

We use this equivalent characterization to introduce the following analogous notion of Fréchet differentiability in the context of nonlinear spaces and dual systems:

**Definition 4.3.** Let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ be a dual system and let $f$ be a proper and convex function. Then $f$ is called Fréchet differentiable in $x \in \mathrm{dom}\partial_{\mathcal{D}} f$ if there exists a function $G : \mathrm{dom}\partial_{\mathcal{D}} f \to Y$ such that $G(y) \in \partial_{\mathcal{D}} f(y)$ for any $y \in \mathrm{dom}\partial_{\mathcal{D}} f$ and $G$ is $d_X$-$d_Y$-continuous in $x$. We call $G$ *a (Fréchet) derivative of* $f$ *at* $x$.

*Remark* 4.4 (For logicians). As a logical indication for why this choice for a definition of a gradient is particularly fruitful, we in particular want to mention evidence coming from the proof mining program (see in particular [29] and [31] as well as the references mentioned in the introduction) where recently in [46] so-called general logical metatheorems were proven which for the first time allowed for a tame treatment of dual spaces of Banach spaces together with gradients of convex functions using methods from proof mining which in particular relied on an axiomatization of Fréchet derivatives using the characterization of Theorem 4.2 to provide a proof-theoretically tame approach to these objects. It is this proof-theoretic tameness of this approach from [46] that is observed in normed spaces that leads us to believe that the notion introduced in Definition 4.3 might provide a fruitful generalization of the normed case.

To illustrate this definition with an example, we consider the dual $X^*$ of a CAT(0)-space $X$ from [22] as already discussed in Example 3.4 before.

*Example* 4.5. Consider a CAT(0)-space $(X, d, W)$, the dual $X^*$ as defined in [22] and the dual system $\mathcal{D}^* = (X, X^*, \langle\cdot,\cdot\rangle)$ as defined in Example 3.4. Using the arbitrary, but fixed, point $o \in X$, we define

$$f(x) = \frac{d^2(o,x)}{2}.$$

First note that $f$ is (uniformly) convex as, using the Bruhat-Tits CN-inequality, we have

$$d^2(o,x_\lambda)/2 \leqslant (1-\lambda)d^2(o,x_0)/2 + \lambda d^2(o,x_1)/2 - \lambda(1-\lambda)d^2(x_0,x_1)/2$$

for given $x_0, x_1 \in X$ and where we write $x_\lambda = W(x_0, x_1, \lambda)$. This $f$ is then Fréchet differentiable everywhere and to see this, we first show that $\partial_{\mathcal{D}^*} f(x) \neq \varnothing$ for any $x \in X$. For this, note that in the case of $\mathcal{D}^*$, we have

$$\partial_{\mathcal{D}^*} f(x) = \{[t\,\overrightarrow{ab}] \in X^* \mid f(y) - f(x) \geqslant t\langle \overrightarrow{xy}, \overrightarrow{ab}\rangle \text{ for all } y \in X\}$$

and thus $[t\,\overrightarrow{ab}] \in \partial_{\mathcal{D}^*} f(x)$ holds if, and only if,

$$d^2(o,y) - d^2(o,x) \geqslant t(d^2(a,y) + d^2(b,x) - d^2(a,x) - d^2(b,y))$$

holds for any $y \in X$. If we pick $a = o$, $b = x$ and $t = 1$, we get that $[1\,\overrightarrow{ox}] \in \partial_{\mathcal{D}^*} f(x)$ as clearly

$$d^2(o,y) - d^2(o,x) \geqslant d^2(o,y) - d^2(o,x) - d^2(x,y)$$

holds true for any $y \in X$. Thus $[1\,\overrightarrow{ox}]$ is a selection of $\partial_{\mathcal{D}^*} f(x)$ which we denote by $\overrightarrow{oI}(x)$. Now we get that $\overrightarrow{oI}(x)$ is continuous at every $x$ as we have (similar to Example 3.4 and using the Cauchy-Schwarz inequality for $X$):

$$\begin{aligned} D([1\,\overrightarrow{ox}], [1\,\overrightarrow{oy}]) &= \sup_{a\neq b} \frac{|\langle \overrightarrow{ba}, \overrightarrow{ox}\rangle - \langle \overrightarrow{ba}, \overrightarrow{oy}\rangle|}{d(a,b)} \\ &= \sup_{a\neq b} \frac{|\langle \overrightarrow{ba}, \overrightarrow{yx}\rangle|}{d(a,b)} \\ &\leqslant \sup_{a\neq b} \frac{d(x,y)d(a,b)}{d(a,b)} \\ &= d(x,y). \end{aligned}$$

So $\overrightarrow{oI}(x)$ is a Fréchet derivative of $f$ at any $x$.

Note that the above Fréchet derivative $\overrightarrow{oI}(x)$ of $d^2(o,x)/2$ is not only continuous everywhere but even uniformly $d$-$D$-continuous (and actually even nonexpansive). Abstracting from this, we introduce a notion of uniform Fréchet differentiability which similarly does not rely on limits (as in normed spaces) but on properties of selections of the subgradient.

**Definition 4.6.** Let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ be a dual system and let $f$ be a proper and convex function. Then $f$ is called uniformly Fréchet differentiable on $D \subseteq \mathrm{dom}\partial_{\mathcal{D}}f$ if there exists a function $G : \mathrm{dom}\partial_{\mathcal{D}}f \to Y$ such that $G(y) \in \partial_{\mathcal{D}}f(y)$ for any $y \in \mathrm{dom}\partial_{\mathcal{D}}f$ and $G$ is uniformly $d_X$-$d_Y$-continuous on every $o$-bounded subset of $D$.

If $D = \mathrm{dom}\partial_{\mathcal{D}}f$, we just say that $f$ is uniformly Fréchet differentiable.

Again this is in some sense abstracted from the theory of Fréchet derivatives in normed spaces as it can be easily seen that over normed spaces, if a Fréchet derivative of a function is uniformly continuous on bounded sets, then the function is uniformly Fréchet differentiable on bounded sets in the usual sense (see e.g. [61]). Conversely, as shown by Reich and Sabach [49], in a reflexive Banach space, if a function is uniformly Fréchet differentiable and bounded on bounded sets, then its Fréchet derivative is uniformly norm-to-norm continuous on bounded sets.

## 5. Bregman distances in hyperbolic spaces

As discussed in the introduction, one of the main motivations for studying gradients of convex functions also in this hyperbolic context is that these objects "unlock" the treatment of Bregman distances in this nonlinear setting. Originally, Bregman distances were introduced in the seminal work [9] in the context of Banach spaces $X$ via

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

where $f$ is a given proper, lower-semicontinuous, convex and Gateaux differentiable function and where $\nabla f(y)$ is the corresponding Gateaux derivative of $f$ at $y$. Intuitively, these functions assign a distance to two points $x, y$ by comparing $f(x)$ with the value of a linearized approximation of $f$ around $y$ using the gradient (see e.g. the discussion in [17]). Since the work of Bregman, these distances have become a main tool in convex analysis (recall the discussion and the references in the introduction).

In the presence of the previous Fréchet derivatives on nonlinear dual systems, it is now fairly straightforward to introduce the following nonlinear analogue of this notion:

**Definition 5.1.** Let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ be a dual system and let $f$ be Fréchet differentiable on $\mathrm{dom}\partial_{\mathcal{D}}f$ with a gradient $G$. Then the Bregman distance associated with $f$ and $G$ is defined by

$$D^G_f(x,y) := f(x) - f(y) - \langle\overrightarrow{yx}, G(y)\rangle$$

for all $x \in \mathrm{dom}f$ and $y \in \mathrm{dom}\partial_{\mathcal{D}}f$.

Clearly $x = y \in \mathrm{dom}\partial_{\mathcal{D}}f$ implies $D^G_f(x,y) = 0$. Also, from $G$ being a selection of $\partial_{\mathcal{D}}f$, it follows that $D^G_f(x,y) \geqslant 0$ for all $x \in \mathrm{dom}f$ and all $y \in \mathrm{dom}\partial_{\mathcal{D}}f$.

*Example* 5.2. As discussed in Example 4.5, in the context of CAT(0)-spaces $(X, d, W)$ with the dual $X^*$ as defined in [22], a Fréchet derivative of $f(x) = d^2(o,x)/2$ is the function $\overrightarrow{oI}(x) :=$

$[1\,\overrightarrow{ox}]$. By analogy to the normed case (see e.g. [9]), we now show that the corresponding Bregman distance reduces (essentially) to the metric in this case. Concretely:

$$\begin{aligned}
D_f^{\overrightarrow{oI}}(x,y) &= \frac{d^2(o,x)}{2} - \frac{d^2(o,y)}{2} - \langle \overrightarrow{yx}, \overrightarrow{oy} \rangle \\
&= \frac{\langle \overrightarrow{ox}, \overrightarrow{ox} \rangle}{2} - \frac{\langle \overrightarrow{oy}, \overrightarrow{oy} \rangle}{2} - \langle \overrightarrow{yx}, \overrightarrow{oy} \rangle \\
&= \frac{\langle \overrightarrow{ox}, \overrightarrow{ox} \rangle}{2} - \frac{\langle \overrightarrow{yx}, \overrightarrow{oy} \rangle}{2} - \left( \frac{\langle \overrightarrow{oy}, \overrightarrow{oy} \rangle}{2} + \frac{\langle \overrightarrow{yx}, \overrightarrow{oy} \rangle}{2} \right) \\
&= \frac{\langle \overrightarrow{ox}, \overrightarrow{ox} \rangle}{2} + \frac{\langle \overrightarrow{xy}, \overrightarrow{ox} \rangle}{2} + \frac{\langle \overrightarrow{xy}, \overrightarrow{xy} \rangle}{2} - \frac{\langle \overrightarrow{ox}, \overrightarrow{oy} \rangle}{2} \\
&= \frac{\langle \overrightarrow{oy}, \overrightarrow{ox} \rangle}{2} + \frac{\langle \overrightarrow{xy}, \overrightarrow{xy} \rangle}{2} - \frac{\langle \overrightarrow{ox}, \overrightarrow{oy} \rangle}{2} \\
&= \frac{\langle \overrightarrow{xy}, \overrightarrow{xy} \rangle}{2} \\
&= \frac{d^2(x,y)}{2}.
\end{aligned}$$

In the next section, we will use this new Bregman distance over hyperbolic spaces to introduce a range of nonexpansivity notions for selfmaps on the space $X$ for which we will then derive convergence results for fixed point iterations.

The first main property of Bregman distances that we will need for that throughout is an analogue of the four point identity (see e.g. [4] for this result in a normed setting).

**Lemma 5.3.** *$D_f^G$ satisfies the four point identity, i.e. for any $x, y, z, w \in \mathrm{dom}\partial_{\mathcal{D}} f$ it holds that*

$$D_f^G(y,x) - D_f^G(y,z) - D_f^G(w,x) + D_f^G(w,z) = \langle \overrightarrow{wy}, Gz - Gx \rangle.$$

*Proof.* Unraveling the definition of $D_f^G$, we get

$$\begin{aligned}
&D_f^G(y,x) - D_f^G(y,z) - D_f^G(w,x) + D_f^G(w,z) \\
&\quad = f(y) - f(x) - \langle \overrightarrow{xy}, Gx \rangle - (f(y) - f(z) - \langle \overrightarrow{zy}, Gz \rangle) \\
&\qquad - (f(w) - f(x) - \langle \overrightarrow{xw}, Gx \rangle) + (f(w) - f(z) - \langle \overrightarrow{zw}, Gz \rangle) \\
&\quad = -\langle \overrightarrow{xy}, Gx \rangle + \langle \overrightarrow{zy}, Gz \rangle + \langle \overrightarrow{xw}, Gx \rangle - \langle \overrightarrow{zw}, Gz \rangle \\
&\quad = \langle \overrightarrow{yw}, Gx \rangle + \langle \overrightarrow{wy}, Gz \rangle \\
&\quad = \langle \overrightarrow{wy}, Gz - Gx \rangle.
\end{aligned}$$

□

In the convergence results, we will assume that $f$ is actually uniformly Fréchet differentiable. Naturally, from this strengthened assumption we can infer various further (uniform) properties of $f$ and the associated Bregman distance which we want to discuss in the following. In that way, these next results are closely modeled after quantitative results obtained for uniformly Fréchet differentiable functions and their gradients in [47]. We begin with properties of $f$ and a corresponding uniformly continuous Fréchet derivative $G$.

**Lemma 5.4.** *Let $f : X \to \mathbb{R}$ be uniformly Fréchet differentiable with a gradient $G : X \to Y$ and let $\omega^G : (0,\infty)^2 \to (0,\infty)$ be a modulus witnessing the uniform continuity of $G$ on o-bounded sets, i.e. for all $\varepsilon, b > 0$ and all $x, y \in \overline{B}_b(o)$:*

$$d_X(x,y) < \omega^G(\varepsilon, b) \text{ implies } d_Y(Gx, Gy) < \varepsilon.$$

*Then, we have the following:*

(1) *$G$ is $\mathcal{O}$-bounded on o-bounded sets, i.e. for all $b > 0$ and all $x \in X$:*
$$d_X(x,o) \leqslant b \text{ implies } d_Y(Gx,\mathcal{O}) \leqslant C_n(b)$$
*where*
$$C_n(b) := b/\omega^G(1,b) + 1 + n$$
*for $n \geqslant d_Y(G(o),\mathcal{O})$.*

(2) *$f$ is uniformly continuous on o-bounded subsets, i.e. for all $\varepsilon, b > 0$ and all $x, y \in \overline{B}_b(o)$:*
$$d_X(x,y) < \omega^f(\varepsilon,b) \text{ implies } |f(x) - f(y)| < \varepsilon$$
*where*
$$\omega^f(\varepsilon,b) := \frac{\varepsilon}{C_n(b)}$$
*with $C_n(b)$ as in item (1).*

(3) *$f$ is bounded on o-bounded sets, i.e. for all $b > 0$ and all $x \in X$:*
$$d_X(x,o) \leqslant b \text{ implies } |f(x)| \leqslant D_n(b)$$
*where*
$$D_m(b) := b/\omega^f(1,b) + 1 + n$$
*for $m \geqslant |f(o)|$.*

*Proof.* (1) For $x$ with $d_X(x,o) \leqslant b$, using $W_X$ we can pick $b/\omega^G(1,b)$ many $z_1, \dots, z_{k-1}$ such that
$$d_X(o,z_1), d_X(z_1,z_2), \dots, d_X(z_{k-1},x) < \omega^G(1,b).$$
In particular $d_X(o,z_i) \leqslant b$ for all $i$ and thus, using the properties of $\omega^G$, we get
$$d_Y(G(o),G(z_1)), d_Y(G(z_1),G(z_2)), \dots, d_Y(G(z_{k-1}),G(x)) < 1.$$
Using the triangle inequality in $Y$, we get $d_Y(G(o),G(x)) \leqslant b/\omega^G(1,b) + 1$ and so, using $d_Y(G(o),\mathcal{O}) \leqslant n$, we get $d_Y(\mathcal{O},G(x)) \leqslant b/\omega^G(1,b) + 1 + n$.

(2) Let $x, y$ be given with $d_X(x,o), d_X(y,o) \leqslant b$ and
$$d_X(x,y) < \frac{\varepsilon}{C_n(b)}.$$
As $G$ is a Fréchet derivative of $f$, we get
$$f(x) - f(y) \leqslant \langle \overrightarrow{yx}, Gx \rangle \leqslant d_X(x,y) d_Y(Gx,\mathcal{O}) \leqslant d_X(x,y) C_n(b).$$
Similarly, we have
$$f(y) - f(x) \leqslant \langle \overrightarrow{xy}, Gy \rangle \leqslant d_X(x,y) d_Y(Gy,\mathcal{O}) \leqslant d_X(x,y) C_n(b).$$
Combined, we get
$$|f(x) - f(y)| \leqslant d_X(x,y) C_n(b) < \varepsilon.$$

(3) This can be shown analogously to item (1).

□

In a normed context, the relationship between a Bregman distance and the norm is rather sparse without additional requirements. One common additional assumption thus often placed on Bregman distances, especially in the context of convergence results for iterations defined in terms of Bregman distances, is that of sequential consistency (see e.g. [14, 15, 51]), i.e. that
$$\lim_{n\to\infty} D_f(x_n,y_n) = 0 \text{ implies } \lim_{n\to\infty} \|x_n - y_n\| = 0$$

for all for all bounded sequences $(x_n), (y_n) \subseteq X$.

As shown in [47], this form of sequential consistency is equivalent to the existence of a modulus $\rho : (0,\infty)^2 \to (0,\infty)$ such that

$$\|x\|, \|y\| \leqslant b \text{ and } D_f(x,y) < \rho(\varepsilon, b) \text{ implies } \|x-y\| < \varepsilon$$

for all $\varepsilon, b > 0$ and $x, y \in X$.

This latter equivalent rephrasing is what we now take as a basis for introducing the following notion of consistency in this nonlinear context.

**Definition 5.5.** We call $D_f^G$ consistent if there exists a function $\rho : (0,\infty)^2 \to (0,\infty)$ such that for any $\varepsilon, b > 0$ and any $x, y \in \mathrm{dom}\partial_{\mathcal{D}} f \cap \overline{B}_b(o)$:

$$D_f^G(x,y) < \rho(\varepsilon, b) \text{ implies } d_X(x,y) < \varepsilon.$$

We call $\rho$ a modulus of consistency.

*Example* 5.6. Let $(X, d, W)$ be a CAT(0)-space with the dual $X^*$ as defined in [22]. Let $f(x) = d^2(o,x)/2$ and consider the corresponding Fréchet derivative $\overrightarrow{oI}(x) := [1\,\overrightarrow{ox}]$. As shown in Example 5.2, it holds that

$$D_f^{\overrightarrow{oI}}(x,y) = \frac{d^2(x,y)}{2}$$

and so clearly $\rho(\varepsilon, b) := \varepsilon^2/2$ is a modulus of consistency for $D_f^{\overrightarrow{oI}}$ as if

$$\frac{d^2(x,y)}{2} = D_f^{\overrightarrow{oI}}(x,y) < \rho(\varepsilon, b) = \frac{\varepsilon^2}{2},$$

then $d(x,y) < \varepsilon$.

By inspecting the definition of $D_f^G$, it is rather immediately clear that a modulus for the converse can in particular be given if $G$ is uniformly continuous:

**Lemma 5.7.** *Let $f : X \to \mathbb{R}$ be uniformly Fréchet differentiable with a gradient $G : X \to Y$ and let $\omega^G$ be a modulus witnessing the uniform continuity of $G$ on $o$-bounded sets. Then $D_f^G$ is reverse consistent with a modulus $P$, i.e. for any $\varepsilon, b > 0$ and any $x, y \in \overline{B}_b(o)$:*

$$d_X(x,y) < P(\varepsilon, b) \text{ implies } D_f^G(x,y) < \varepsilon$$

*where*

$$P(\varepsilon, b) := \frac{\varepsilon}{2C_n(b)}$$

*with $C_n$ defined as in Lemma 5.4.*

*Proof.* Note that $P(\varepsilon, b) = \omega^f(\varepsilon/2, b)$ with $\omega^f$ as in Lemma 5.4. Thus we get

$$\begin{aligned} D_f^G(x,y) &= f(x) - f(y) - \langle \overrightarrow{yx}, Gy \rangle \\ &< \varepsilon/2 + d_X(x,y) d_Y(Gy, \mathcal{O}) \\ &\leqslant \varepsilon, \end{aligned}$$

as $d_Y(Gy, \mathcal{O}) \leqslant C_n(b)$ by Lemma 5.4. □

Further, also the continuity of $D_f^G$ in both arguments is an immediate consequence of the assumption of uniform Fréchet differentiability of $f$ and we can give the following lemma spelling this out quantitatively.

**Lemma 5.8.** *Let $f$ be uniformly Fréchet differentiable with a gradient $G$ and let $\omega^G$ be a modulus witnessing the uniform continuity of $G$ on o-bounded sets. Then the Bregman distance $D^G_f$ is uniformly continuous on o-bounded subsets in both arguments, i.e. for any $\varepsilon, b > 0$ and any $x, x', y, y' \in \overline{B}_b(o)$:*

$$d_X(x, x') < \xi_1(\varepsilon, b) \text{ implies } |D^G_f(x, y) - D^G_f(x', y)| < \varepsilon$$

*as well as*

$$d_X(y, y') < \xi_2(\varepsilon, b) \text{ implies } |D^G_f(x, y) - D^G_f(x, y')| < \varepsilon$$

*where*

$$\xi_1(\varepsilon, b) := \frac{\varepsilon}{2C_n(b)} \text{ and } \xi_2(\varepsilon, b) := \min\left\{\frac{\varepsilon}{3C_n(b)}, \omega^G(\varepsilon/6b, b)\right\},$$

*respectively, with $C_n(b)$ as in Lemma 5.4.*

*Proof.* For the former, note that $\xi_1(\varepsilon, b) = \omega^f(\varepsilon/2, b)$ for $\omega^f$ defined as in Lemma 5.4. Thus we get

$$\begin{aligned}
|D^G_f(x, y) - D^G_f(x', y)| &= |f(x) - f(y) - \langle \overrightarrow{yx}, Gy\rangle - (f(x') - f(y) - \langle \overrightarrow{yx'}, Gy\rangle)| \\
&\leqslant |f(x) - f(x')| + |\langle \overrightarrow{xx'}, Gy\rangle| \\
&\leqslant \frac{\varepsilon}{2} + d_X(x, x')d_Y(Gy, \mathcal{O}) \\
&\leqslant \frac{\varepsilon}{2} + d_X(x, x')C_n(b) \\
&< \varepsilon
\end{aligned}$$

if $d_X(x, x') < \xi_1(\varepsilon, b)$ as $d_Y(Gy, \mathcal{O}) \leqslant C_n(b)$ by Lemma 5.4.

For the latter, note that similarly $\varepsilon/3C_n(b) = \omega^f(\varepsilon/3, b)$ for $\omega^f$ defined as in Lemma 5.4, and so we get

$$\begin{aligned}
|D^G_f(x, y) - D^G_f(x, y')| &= |f(x) - f(y) - \langle \overrightarrow{yx}, Gy\rangle - (f(x) - f(y') - \langle \overrightarrow{y'x}, Gy'\rangle)| \\
&\leqslant |f(y') - f(y)| + |\langle \overrightarrow{y'x}, Gy'\rangle - \langle \overrightarrow{yx}, Gy\rangle| \\
&\leqslant |f(y') - f(y)| + |\langle \overrightarrow{y'x}, Gy'\rangle - \langle \overrightarrow{yx}, Gy'\rangle| + |\langle \overrightarrow{yx}, Gy'\rangle - \langle \overrightarrow{yx}, Gy\rangle| \\
&\leqslant |f(y') - f(y)| + |\langle \overrightarrow{y'y}, Gy'\rangle| + |\langle \overrightarrow{yx}, Gy' - Gy\rangle| \\
&\leqslant |f(y') - f(y)| + d_X(y, y')d_Y(Gy', \mathcal{O}) + d_X(x, y)d_Y(Gy', Gy) \\
&\leqslant |f(y') - f(y)| + d_X(y, y')C_n(b) + 2bd_Y(Gy', Gy) \\
&< \varepsilon
\end{aligned}$$

if $d_X(y, y') < \xi_2(\varepsilon, b)$ as $d_Y(Gy', \mathcal{O}) \leqslant C_n(b)$ by Lemma 5.4. □

In normed spaces, another relationship between the norm and a given Bregman distance that is inherently under-determined in general is that between boundedness under the norm and boundedness under the Bregman distance. This situation is similar in the nonlinear context here and we introduce the following notion as a requirement to remedy this (defined in analogy to [47]).

**Definition 5.9.** We call $D^G_f$ uniformly $o$-bounded if for any $\alpha, b > 0$, there exists a $b'$ such that for any $x, y \in \mathrm{dom}\partial_{\mathcal{D}} f$:

$$D^G_f(x, y) \leqslant \alpha \text{ and } d_X(x, o) \leqslant b \text{ implies } d_X(y, o) \leqslant b'.$$

We call a function $\beta(\alpha, b)$ witnessing such a $b'$ in terms of $\alpha, b$ a modulus for the uniform $o$-boundedness of $D^G_f$.

*Example* 5.10. Let $(X, d, W)$ be a CAT(0)-space with the dual $X^*$ as defined in [22]. Let $f(x) = d^2(o, x)/2$ and consider the corresponding Fréchet derivative $\overrightarrow{oI}(x) := [1\,\overrightarrow{ox}]$. As shown in Example 5.2, it holds that

$$D^{\overrightarrow{oI}}_f(x, y) = \frac{d^2(x, y)}{2}$$

and so clearly $\beta(\alpha, b) := \sqrt{2\alpha} + b$ is a modulus for the uniform $o$-boundedness of $D^{\overrightarrow{oI}}_f$ as if $d(x, o) \leqslant b$ and

$$\frac{d^2(x, y)}{2} = D^{\overrightarrow{oI}}_f(x, y) \leqslant \alpha,$$

we have $d(x, y) \leqslant \sqrt{2\alpha}$ and so

$$d(y, o) \leqslant d(x, y) + d(x, o) \leqslant \sqrt{2\alpha} + b = \beta(\alpha, b).$$

## 6. Nonexpansivity in the context of Bregman distances

We now introduce different notions of nonexpansivity relative to our Bregman distance for mappings $T : K \to K$ with $K \subseteq X$ by adapting the various nonexpansivity notions relative to Bregman distances originally introduced in normed spaces as e.g. in the seminal works [39, 40, 41, 48, 51].

Since, in the context of these works, many of the central properties of these mappings rely on the totality of the underlying function $f$ as well as its derivative, we for simplicity assume from this point onwards that $\mathrm{dom}f = X$ and that $f$ is Fréchet differentiable on all of $X$, i.e. there is a derivative $G$ with $\mathrm{dom}G = X$.

The first central nonexpansivity notion is what we will call a Bregman quasi-nonexpansive map. For this, we write $\mathrm{Fix}(T) \subseteq K$ for the set of fixed points of $T$.

**Definition 6.1.** A map $T : K \to K$ with $K \subseteq X$ is called Bregman quasi-nonexpansive (BQNE) if

$$D^G_f(p, Tx) \leqslant D^G_f(p, x)$$

for any $p \in \mathrm{Fix}(T)$ and $x \in K$.

In regard to self-iterations of maps, this class is too broad to carry sensible convergence theorems for the approximation of fixed points. As in the case of normed spaces, we thus now consider a notion of a Bregman strongly quasi-nonexpansive map:

**Definition 6.2.** A map $T : K \to K$ with $K \subseteq X$ is called Bregman strongly quasi-nonexpansive (BSNE) if it is BQNE and additionally

$$\lim_{n\to\infty} D^G_f(p, x_n) - D^G_f(p, Tx_n) = 0 \text{ implies } \lim_{n\to\infty} D^G_f(Tx_n, x_n) = 0$$

for any $p \in \mathrm{Fix}(T)$ and any $o$-bounded sequence $(x_n) \subseteq K$.

This definition is the natural generalization to this nonlinear context of the notion of so-called properly L-BSNE maps considered for normed spaces in [40]. For our convergence results later on, we are also simultaneously concerned with quantitative results on the convergence. Through that perspective, we here consider the following uniform variant of this notion:

**Definition 6.3.** A map $T : K \to K$ with $K \subseteq X$ is called uniformly Bregman strongly quasi-nonexpansive (uniformly BSNE) if it is BQNE and additionally for any $\varepsilon, b > 0$, there exists a $\delta > 0$ such that for all $p \in \mathrm{Fix}(T)$ and all $x \in K \cap \overline{B}_b(o)$:

$$D^G_f(p, x) - D^G_f(p, Tx) < \delta \text{ implies } D^G_f(Tx, x) < \varepsilon.$$

We call a function $\omega(\varepsilon, b)$ witnessing such a $\delta$ in terms of $\varepsilon, b$ a uniform BSNE modulus for $T$.

In the plain metric context of ordinary strong quasi-nonexpansive mappings, the uniform variant analogous to the above Definition 6.3 was introduced by Kohlenbach in [30] (with a "fully" uniform variant, i.e. where $\delta$ is even independent of $b$, already considered by Bruck in [12]). In the context of the ordinary Bregman strongly quasi-nonexpansive mappings over normed spaces, an analogous notion of a modulus was recently considered in [47] (note, however, that a uniform BSNE mapping in [47] is an even stronger notion). In that way, the notion of uniform BSNE mappings considered in Definition 6.3 is a proper generalization of both these notions and, in particular, this will be a suitable notion for the quantitative convergence results provided later on.

Before we move on, we want to remark that a uniform BSNE mapping is nothing else but a BQNE mapping satisfying the above uniform variant of the property that

$$D^G_f(p, x) - D^G_f(p, Tx) \leqslant 0 \text{ implies } D^G_f(Tx, x) = 0$$

for all $x \in K$ and $p \in \mathrm{Fix}(T)$. We call a BQNE mapping that only satisfies this restricted property a strict BQNE mapping. A simple compactness argument yields that, over proper spaces where $D^G_f$ is uniformly continuous on $o$-bounded subsets in both arguments, any continuous and strictly BQNE mapping is already uniformly BSNE in the sense of Definition 6.3.

Also, while this notion outside of compact spaces is generally stronger than the plain notion of a BSNE mapping, we want to further argue that the price paid by moving to this uniform variant is rather low in most practical cases as many natural BSNE maps are already uniformly BSNE. In fact, there is an underlying logical reason for this circumstance as the following remark briefly discusses.

*Remark* 6.4 (For logicians). In suitable formal systems for treating these dual systems and metric Fréchet derivatives as well as Bregman distances and that enjoy general logical metatheorems in the style of proof mining (which might be developed analogously to [46]), being BSNE is equivalent to being uniformly BSNE in the context of a nonstandard uniform boundedness principle $\exists$-$\mathrm{UB}^X$ (see [28]) that can be conservatively added to such systems. Further, already from a (noneffective) proof of the strict BQNE property in such a formal system, one can extract an effective uniform BSNE modulus. Note also the similarity of these circumstances to those of both the works [30] and [47].

One example of a class of BSNE maps that are immediately also recognized as being uniformly BSNE is that of Bregman firmly nonexpansive maps, a Bregman distance analogue of the classical and influential notion of firm nonexpansivity introduced in a normed setting at various points in the literature under different names (e.g. being called D-firm in [4], Bregman firmly nonexpansive in [51] and a certain subclass was introduced in [35, 36] under the name of mappings of firmly nonexpansive type).

**Definition 6.5.** A map $T : K \to K$ with $K \subseteq X$ is called Bregman firmly nonexpansive (BFNE) if

$$\langle \overrightarrow{TyTx}, Gx - Gy\rangle \geqslant \langle \overrightarrow{TyTx}, GTx - GTy\rangle$$

for any $x, y \in K$.

Using the four point identity, we can derive another characterization of BFNE maps (akin to [51]).

**Lemma 6.6.** *A map $T : K \to K$ with $K \subseteq X$ is BFNE if, and only if, it holds that*

$$D_f^G(Tx, y) - D_f^G(Tx, x) - D_f^G(Ty, y) + D_f^G(Ty, x) \geqslant D_f^G(Tx, Ty) + D_f^G(Ty, Tx)$$

*for all $x, y \in K$.*

*Proof.* Using the four point identity, we get that

$$\begin{aligned}\langle \overrightarrow{TyTx}, GTx - GTy \rangle &= D_f^G(Tx, Ty) - D_f^G(Tx, Tx) - D_f^G(Ty, Ty) + D_f^G(Ty, Tx) \\ &= D_f^G(Tx, Ty) + D_f^G(Ty, Tx).\end{aligned}$$

Similarly, we also get

$$\langle \overrightarrow{TyTx}, Gx - Gy \rangle = D_f^G(Tx, y) - D_f^G(Tx, x) - D_f^G(Ty, y) + D_f^G(Ty, x)$$

and combined with the definition of BFNE maps, we get the result. □

Then, as we can see now, every BFNE map is also uniformly BSNE with a particularly simple modulus.

**Lemma 6.7.** *If $T$ is BFNE, then $T$ is also uniformly BSNE with a modulus $\omega(\varepsilon, b) = \varepsilon$.*

*Proof.* By Lemma 6.6, we have that $T$ satisfies

$$D_f^G(Tx, y) - D_f^G(Tx, x) - D_f^G(Ty, y) + D_f^G(Ty, x) \geqslant D_f^G(Tx, Ty) + D_f^G(Ty, Tx).$$

If now $y = p \in \mathrm{Fix}(T)$, then

$$D_f^G(p, x) - D_f^G(p, Tx) \geqslant D_f^G(Tx, x).$$

This immediately gives the modulus. □

A last property of BQNE maps that will be handy later on is that, in the presence of a fixed point and under the condition of a uniformly continuous gradient $G$ such that the associated Bregman distance $D_f^G$ is uniformly $o$-bounded, these mappings are $o$-bounded on $o$-bounded sets.

**Lemma 6.8.** *Let $f$ be uniformly Fréchet differentiable with a gradient $G$ and let $\omega^G$ be a modulus witnessing the uniform continuity of $G$ on $o$-bounded sets. Let $T$ be BQNE with $\mathrm{Fix}(T) \neq \emptyset$ and let $D_f^G$ be uniformly $o$-bounded with a modulus $\beta$. Then $T$ is $o$-bounded on $o$-bounded sets, i.e. for all $b > 0$ and all $x \in X$:*

$$d_X(x, o) \leqslant b \text{ implies } d_X(Tx, o) \leqslant \beta(E_l(b), l)$$

*where*

$$E_l(b) := D_m(l) + D_m(b) + (b + l)C_n(b)$$

*for $p \in \mathrm{Fix}(T)$ with $l \geqslant d_X(o, p)$ and $D_m, C_n$ as in Lemma 5.4.*

*Proof.* For $p$ as above, we have

$$\begin{aligned}D_f^G(p, Tx) &\leqslant D_f^G(p, x) \\ &= f(p) - f(x) - \langle \overrightarrow{xp}, Gx \rangle \\ &\leqslant f(p) + D_m(b) + d_X(x, p) d_Y(Gx, \mathcal{O}) \\ &\leqslant D_m(l) + D_m(b) + (d_X(x, o) + d_X(o, p))C_n(b) \\ &\leqslant D_m(l) + D_m(b) + (b + l)C_n(b).\end{aligned}$$

Then $d_X(Tx, o) \leqslant \beta(E_l(b), l)$ as $l \geqslant d_X(o, p)$. □

## 7. Monotone operators and resolvents in dual systems

We now introduce monotone operators relative to dual systems. In the context of inner product spaces, monotonicity arose in the 1960's through the seminal work of Minty [43, 44] and was subsequently extended to general normed spaces. Concretely, two main generalizations emerged: For one, based on an equivalent variant of monotonicity written solely in terms of the underlying norm of the space, Kato in [23] introduced the notion of accretive operators which have become seminal tools e.g. in semigroup theory. For another, the monotonicity condition was generalized by Browder (see [10, 11]) to normed spaces by replacing the inner product with the application of functionals from the corresponding dual space. This latter definition is what we extend here to the context of our dual systems, where we want to emphasize that this new notion essentially just arises by abstracting the main approach from [24] away from the choice of $X^*$ for a CAT(0)-space $X$.

**Definition 7.1.** Let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ be a dual system. An operator $A \subseteq X \times Y$ is called monotone if

$$\langle\overrightarrow{yx}, x^* - y^*\rangle \geqslant 0$$

for any $(x, x^*), (y, y^*) \in A$.

In analogy to the theory in normed spaces, the convexity of $f$ is also here linked to the monotonicity of $\partial_{\mathcal{D}} f$ and thus also of any selection $G$ which we collect in the following lemma:

**Lemma 7.2.** *Let $f$ be convex. Then operator $\partial_{\mathcal{D}} f$ is monotone. A fortiori any Fréchet derivative $G$ of $f$ is monotone.*

*Proof.* Let $(x, x^*), (y, y^*) \in \partial_{\mathcal{D}} f$, i.e.

$$f(z) - f(x) \geqslant \langle\overrightarrow{xz}, x^*\rangle \text{ and } f(z) - f(y) \geqslant \langle\overrightarrow{yz}, y^*\rangle$$

for all $z \in X$. This implies

$$f(x) - f(y) \leqslant \langle\overrightarrow{yx}, x^*\rangle \text{ and } f(x) - f(y) \geqslant \langle\overrightarrow{yx}, y^*\rangle$$

Now, we can derive

$$\langle\overrightarrow{yx}, x^* - y^*\rangle \geqslant f(x) - f(y) - (f(x) - f(y)) = 0.$$

□

A stronger notion of monotonicity that we crucially rely on later is that of strict monotonicity which we also define in analogy to [22].

**Definition 7.3.** Let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ be a dual system. An operator $A \subseteq X \times Y$ is called strictly monotone if it is monotone and for any $(x, x^*), (y, y^*) \in A$, it holds that

$$\langle\overrightarrow{yx}, x^* - y^*\rangle \leqslant 0 \text{ implies } x = y.$$

For a quantitative and uniform version of the above definition, we introduce the following notion:

**Definition 7.4.** Let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ be a dual system. An operator $A \subseteq X \times Y$ is called uniformly strictly monotone with a modulus $\eta : (0, \infty)^2 \to (0, \infty)$ if it is monotone and for any $\varepsilon, b > 0$ and any $(x, x^*), (y, y^*) \in A$, it holds that

$$d_X(x, o), d_X(y, o) \leqslant b \text{ and } \langle\overrightarrow{yx}, x^* - y^*\rangle < \eta(\varepsilon, b) \text{ implies } d_X(x, y) < \varepsilon.$$

*Example* 7.5. In a CAT(0)-space $(X,d,W)$ with dual $X^*$ as defined in [22], we have that $\overrightarrow{oI}(x) := [1\,\overrightarrow{ox}]$ is uniformly strictly monotone with the simple modulus $\eta(\varepsilon,b) = \varepsilon^2$. To see this, let $x,y \in X$ be given. Then

$$\langle \overrightarrow{yx}, \overrightarrow{oI}(x) - \overrightarrow{oI}(y)\rangle = \langle \overrightarrow{yx}, \overrightarrow{ox}\rangle - \langle \overrightarrow{yx}, \overrightarrow{oy}\rangle = \langle \overrightarrow{yx}, \overrightarrow{yx}\rangle = d^2(x,y)$$

and so, if $\langle \overrightarrow{yx}, \overrightarrow{oI}(x) - \overrightarrow{oI}(y)\rangle < \varepsilon^2$, then $d^2(y,x) < \varepsilon^2$, i.e. $d(x,y) < \varepsilon$.

In the context of such a modulus $\eta$ for $G$, we can then in particular show that essentially any BFNE map $T$ is uniformly continuous on $o$-bounded sets.

**Lemma 7.6.** *Let $\mathcal{D} = (X,Y,\langle\cdot,\cdot\rangle)$ be a dual system. Let $f$ be uniformly Fréchet differentiable with a gradient $G$, let $\omega^G$ be a modulus witnessing the uniform continuity of $G$ on o-bounded sets and let $G$ be uniformly strictly monotone with a modulus $\eta$. Let $T : K \to K$ for $K \subseteq X$ be BFNE and be o-bounded on o-bounded sets with a modulus $E$. Then $T$ is uniformly continuous on o-bounded sets with a simple modulus $\Delta$, i.e. for all $\varepsilon, b > 0$ and all $x,y \in K \cap \overline{B}_b(o)$:*

$$d_X(x,y) < \Delta(\varepsilon,b) \text{ implies } d_X(Tx,Ty) < \varepsilon$$

*with*

$$\Delta(\varepsilon,b) := \omega^G\left(\frac{\eta(\varepsilon,E(b))}{2E(b)}, b\right).$$

*Proof.* Note that we have

$$\begin{aligned}\langle \overrightarrow{TyTx}, GTx - GTy\rangle &\leqslant \langle \overrightarrow{TyTx}, Gx - Gy\rangle \\ &\leqslant d_X(Tx,Ty)d_Y(Gx,Gy) \\ &\leqslant 2E(b)d_Y(Gx,Gy).\end{aligned}$$

This yields that

$$\langle \overrightarrow{TyTx}, GTx - GTy\rangle < \eta(\varepsilon, E(b))$$

for $d_X(x,y) < \Delta(\varepsilon,b)$ and so $d_X(Tx,Ty) < \varepsilon$ by using the uniform strict monotonicity of $G$. □

The main object associated with monotone operators is the so-called resolvent. We here now abstract the notion of a resolvent for monotone operators on CAT(0)-spaces as introduced in [24] (see also Example 7.8 later on) and simultaneously generalize said notion by defining a resolvent relativized to a gradient $G$ for monotone operators on dual systems for hyperbolic spaces. This definition for a resolvent can thus also be considered to be a generalization of the notion of resolvents relative to gradients from normed spaces as e.g. introduced originally in the seminal paper of Eckstein [17] but also introduced in [4] under the name of "$D$-resolvents" and considered in [50] under the name of "resolvents of $A$ relative to $f$".

**Definition 7.7.** Let $\mathcal{D} = (X,Y,\langle\cdot,\cdot\rangle)$ be a dual system and $A \subseteq X \times Y$ be an operator. Let $G : \mathrm{dom}G \subseteq X \to Y$ be a function. The resolvent relative to $G$ is defined as

$$J^G_{\lambda A} := (G + \lambda A)^{-1} \circ G$$

where addition in $Y$ is to be understood as addition of suitable equivalence classes of functionals on $X^2$, i.e.

$$\begin{aligned}J^G_{\lambda A}(x) &:= \{z \in \mathrm{dom}G \mid \exists z^* \in Az \text{ such that } z^* =_{\mathcal{D}} \lambda^{-1}(Gx - Gz)\} \\ &:= \{z \in \mathrm{dom}G \mid \exists z^* \in Az \text{ such that } \langle \overrightarrow{ab}, z^*\rangle = \langle \overrightarrow{ab}, \lambda^{-1}(Gx - Gz)\rangle \text{ for all } a,b \in X\}\end{aligned}$$

for $x \in \mathrm{dom}G$, and $J^G_{\lambda A}(x) := \varnothing$, otherwise.

Clearly, for a given $G : \mathrm{dom}G \subseteq X \to Y$, we in general have $\mathrm{dom}J^G_{\lambda A} \subseteq \mathrm{dom}G$ and $\mathrm{ran}J^G_{\lambda A} \subseteq \mathrm{dom}G$.

*Example* 7.8. In the case of $\overrightarrow{oI} := [1\,\overrightarrow{ox}]$ on a CAT(0)-space $(X, d, W)$ with dual companion $X^*$ as in [22], we get

$$J^{\overrightarrow{oI}}_{\lambda A} = (\overrightarrow{oI} + \lambda A)^{-1} \circ \overrightarrow{oI}$$

which, as discussed in [18], reduces to the resolvent notion introduced in [24], i.e.

$$J^{\overrightarrow{oI}}_{\lambda A}(x) = \{z \in X \mid [\lambda^{-1}\,\overrightarrow{zx}] \in Az\}.$$

Concretely, to see this equality, note first that $\lambda^{-1}([1\,\overrightarrow{ox}] - [1\,\overrightarrow{oz}]) =_{\mathcal{D}^*} [\lambda^{-1}\,\overrightarrow{zx}]$ with $\mathcal{D}^*$ defined as in Example 3.4 as we have

$$\begin{aligned}
\langle \overrightarrow{ab}, [\lambda^{-1}\,\overrightarrow{zx}]\rangle &= \lambda^{-1}\langle \overrightarrow{ab}, \overrightarrow{zx}\rangle \\
&= \lambda^{-1}(\langle \overrightarrow{ab}, \overrightarrow{zo}\rangle + \langle \overrightarrow{ab}, \overrightarrow{ox}\rangle) \\
&= \lambda^{-1}(\langle \overrightarrow{ab}, \overrightarrow{ox}\rangle - \langle \overrightarrow{ab}, \overrightarrow{oz}\rangle) \\
&= \langle \overrightarrow{ab}, \lambda^{-1}([1\,\overrightarrow{ox}] - [1\,\overrightarrow{oz}])\rangle
\end{aligned}$$

for all $a, b \in X$. Thus, in particular, if $[\lambda^{-1}\,\overrightarrow{zx}] \in Az$, then $z \in J^{\overrightarrow{oI}}_{\lambda A}(x)$. Conversely, if we have $z \in J^{\overrightarrow{oI}}_{\lambda A}(x)$, i.e. we have a $z^* \in Az$ with $z^* =_{\mathcal{D}^*} \lambda^{-1}(\overrightarrow{oI}(x) - \overrightarrow{oI}(z))$, we also have $z^* =_{\mathcal{D}^*} [\lambda^{-1}\,\overrightarrow{zx}]$. Now, as shown in Example 3.4, the system $\mathcal{D}^*$ is in particular (uniformly) non-degenerate and so we derive $z^* = [\lambda^{-1}\,\overrightarrow{zx}]$ in the sense of $X^*$ (i.e. $D(z^*, [\lambda^{-1}\,\overrightarrow{zx}]) = 0$) from this. Thus also $[\lambda^{-1}\,\overrightarrow{zx}] \in Az$.

In particular, as already discussed in [24], if in that case $A$ is instantiated with $\partial f$ in the sense of [22], i.e. $\partial_{\mathcal{D}^*} f$ in our notation, for a proper, convex and continuous function $f$ on a complete CAT(0)-space $X$, one obtains the resolvents of $f$ in the sense of Jost [21] (see also e.g. [7]), i.e.

$$J^{\overrightarrow{oI}}_{\lambda \partial_{\mathcal{D}^*} f}(x) = \mathrm{argmin}_{z \in X}\left\{f(z) + \frac{1}{2\lambda}d^2(z, x)\right\}.$$

We now discuss some nice properties of $J^G_{\lambda A}$, the first of which concerning the single-valuedness of $J^G_{\lambda A}$ which crucially uses the assumption that $G$ is strictly monotone (similar to the normed case, see e.g. [4]).

**Lemma 7.9.** *Let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ and let $G : \mathrm{dom}G \subseteq X \to Y$ be a function which is strictly monotone. Let $A \subseteq X \times Y$ be a monotone operator. Let $\lambda > 0$ be arbitrary. Then $J^G_{\lambda A}$ is either empty or a singleton.*

*Proof.* Let $z, z' \in J^G_{\lambda A}x$ for $x \in \mathrm{dom}G$ with $z^* \in Az$ and $z'^* \in Az'$ such that

$$z^* =_{\mathcal{D}} \lambda^{-1}(Gx - Gz) \text{ and } z'^* =_{\mathcal{D}} \lambda^{-1}(Gx - Gz').$$

By the monotonicity of $A$, we get

$$\begin{aligned}
\langle \overrightarrow{z'z}, z^* - z'^*\rangle \geqslant 0 &\Rightarrow \langle \overrightarrow{z'z}, z^*\rangle \geqslant \langle \overrightarrow{z'z}, z'^*\rangle \\
&\Rightarrow \langle \overrightarrow{z'z}, \lambda^{-1}(Gx - Gz)\rangle \geqslant \langle \overrightarrow{z'z}, \lambda^{-1}(Gx - Gz')\rangle \\
&\Rightarrow \langle \overrightarrow{z'z}, Gz' - Gz\rangle \geqslant 0 \\
&\Rightarrow \langle \overrightarrow{z'z}, Gz - Gz'\rangle \leqslant 0
\end{aligned}$$

which in turn implies $z = z'$ by strict monotonicity of $G$. $\square$

In the context of a strictly monotone map $G$, we thus write $J^G_{\lambda A}$ for the canonical selection map. The first property that we derive for this resolvent is that it is actually Bregman firmly nonexpansive w.r.t. $G$.

**Lemma 7.10.** *Let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ and let $G : \mathrm{dom}G \subseteq X \to Y$ be a function which is strictly monotone. Let $A \subseteq X \times Y$ be a monotone operator. Let $\lambda > 0$ be arbitrary. Then $J^G_{\lambda A}$ satisfies*

$$\langle \overrightarrow{J^G_{\lambda A}yJ^G_{\lambda A}x}, Gx - Gy\rangle \geqslant \langle \overrightarrow{J^G_{\lambda A}yJ^G_{\lambda A}x}, GJ^G_{\lambda A}x - GJ^G_{\lambda A}y\rangle$$

*for any $x, y \in \mathrm{dom}J^G_{\lambda A}$.*

*Proof.* Let $x, y \in \mathrm{dom}J^G_{\lambda A}$ and $z^* \in AJ^G_{\lambda A}x$ as well as $w^* \in AJ^G_{\lambda A}y$ such that

$$z^* =_{\mathcal{D}} \lambda^{-1}(Gx - GJ^G_{\lambda A}x) \text{ and } w^* =_{\mathcal{D}} \lambda^{-1}(Gy - GJ^G_{\lambda A}y).$$

By monotonicity, we get

$$\langle \overrightarrow{J^G_{\lambda A}yJ^G_{\lambda A}x}, z^* - w^*\rangle \geqslant 0$$

and thus

$$\langle \overrightarrow{J^G_{\lambda A}yJ^G_{\lambda A}x}, \lambda^{-1}(Gx - GJ^G_{\lambda A}x) - \lambda^{-1}(Gy - GJ^G_{\lambda A}y)\rangle \geqslant 0.$$

Multiplying by $\lambda > 0$ and rearranging yields the result. □

*Remark* 7.11. The above result, showing that $J^G_{\lambda A}$ is BFNE w.r.t. $G$, extends that of Bauschke, Borwein and Combettes [4] for the analogous result in normed spaces. In the particular case of $G = \vec{oI}$ on a CAT(0)-space $(X, d, W)$ with dual companion $X^*$ as in [22], i.e. for $J^{\vec{oI}}_{\lambda A}$, we in particular reobtain the result from [24] that this mapping is firmly nonexpansive as it satisfies

$$\begin{aligned}
&\langle \overrightarrow{J^{\vec{oI}}_{\lambda A}yJ^{\vec{oI}}_{\lambda A}x}, \vec{oI}x - \vec{oI}y\rangle \geqslant \langle \overrightarrow{J^{\vec{oI}}_{\lambda A}yJ^{\vec{oI}}_{\lambda A}x}, \vec{oI}J^{\vec{oI}}_{\lambda A}x - \vec{oI}J^{\vec{oI}}_{\lambda A}y\rangle \\
&\quad\Leftrightarrow \langle \overrightarrow{J^{\vec{oI}}_{\lambda A}yJ^{\vec{oI}}_{\lambda A}x}, \overrightarrow{ox}\rangle - \langle \overrightarrow{J^{\vec{oI}}_{\lambda A}yJ^{\vec{oI}}_{\lambda A}x}, \overrightarrow{oy}\rangle \geqslant \langle \overrightarrow{J^{\vec{oI}}_{\lambda A}yJ^{\vec{oI}}_{\lambda A}x}, \overrightarrow{oJ^{\vec{oI}}_{\lambda A}x}\rangle - \langle \overrightarrow{J^{\vec{oI}}_{\lambda A}yJ^{\vec{oI}}_{\lambda A}x}, \overrightarrow{oJ^{\vec{oI}}_{\lambda A}y}\rangle \\
&\quad\Leftrightarrow \langle \overrightarrow{J^{\vec{oI}}_{\lambda A}yJ^{\vec{oI}}_{\lambda A}x}, \overrightarrow{ox}\rangle - \langle \overrightarrow{J^{\vec{oI}}_{\lambda A}yJ^{\vec{oI}}_{\lambda A}x}, \overrightarrow{oJ^{\vec{oI}}_{\lambda A}x}\rangle \geqslant \langle \overrightarrow{J^{\vec{oI}}_{\lambda A}yJ^{\vec{oI}}_{\lambda A}x}, \overrightarrow{oy}\rangle - \langle \overrightarrow{J^{\vec{oI}}_{\lambda A}yJ^{\vec{oI}}_{\lambda A}x}, \overrightarrow{oJ^{\vec{oI}}_{\lambda A}y}\rangle \\
&\quad\Leftrightarrow \langle \overrightarrow{J^{\vec{oI}}_{\lambda A}xJ^{\vec{oI}}_{\lambda A}y}, \overrightarrow{J^{\vec{oI}}_{\lambda A}xx}\rangle + \langle \overrightarrow{J^{\vec{oI}}_{\lambda A}yJ^{\vec{oI}}_{\lambda A}x}, \overrightarrow{J^{\vec{oI}}_{\lambda A}yy}\rangle \leqslant 0
\end{aligned}$$

for all $x, y \in \mathrm{dom}J^G_{\lambda A}$ which in turn is equivalent to the usual notion of firm nonexpansivity in CAT(0)-spaces, i.e. that

$$d^2(J^{\vec{oI}}_{\lambda A}x, J^{\vec{oI}}_{\lambda A}y) \leqslant \langle \overrightarrow{J^{\vec{oI}}_{\lambda A}xJ^{\vec{oI}}_{\lambda A}y}, \overrightarrow{xy}\rangle$$

for all $x, y \in \mathrm{dom}J^G_{\lambda A}$, as shown in [24].

Using Lemma 6.6, we get the following corollary:

*Corollary* 7.12. Let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ and let $G : \mathrm{dom}G \subseteq X \to Y$ be a function which is strictly monotone. Let $A \subseteq X \times Y$ be a monotone operator. Let $\lambda > 0$ be arbitrary. Then for any $x, y \in \mathrm{dom}J^G_{\lambda A}$:

$$D^G_f(J^G_{\lambda A}x, y) - D^G_f(J^G_{\lambda A}x, x) - D^G_f(J^G_{\lambda A}y, y) + D^G_f(J^G_{\lambda A}y, x) \geqslant D^G_f(J^G_{\lambda A}x, J^G_{\lambda A}y) + D^G_f(J^G_{\lambda A}y, J^G_{\lambda A}x).$$

We now turn to the notion of a zero of a monotone operator. By abstracting from the case of CAT(0)-spaces as e.g. considered in [24], we consider the following notion of a zero:

**Definition 7.13.** Let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ be a dual system and let $A \subseteq X \times Y$ be an operator. We call a point $x \in X$ a zero of $A$ if

$$\exists z^* \in Ax\,(z^* =_{\mathcal{D}} \mathcal{O}).$$

We write $\mathrm{zer}A$ for the set of zeros of $A$.

As in the normed case (and in the context of CAT(0)-spaces for that matter, see [24]), the zeros of a monotone operator correspond to the fixed points of (relativized) resolvents:

**Lemma 7.14.** *Let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ and let $G : \mathrm{dom}G \subseteq X \to Y$ be a function which is strictly monotone. Let $A \subseteq X \times Y$ be a monotone operator. Let $J^G_{\lambda A}$ be its resolvent relative to $G$ for $\lambda > 0$. Then*

$$\mathrm{Fix}(J^G_{\lambda A}) = \mathrm{zer}A \cap \mathrm{dom}G.$$

*Proof.* Let $x \in \mathrm{dom}G$ be a zero of $A$ and let $z^* \in Ax$ with $z^* =_{\mathcal{D}} \mathcal{O}$. By monotonicity of $A$, we get

$$\langle \overrightarrow{xJ^G_{\lambda A}x}, \lambda^{-1}(Gx - GJ^G_{\lambda A}x)\rangle = \langle \overrightarrow{xJ^G_{\lambda A}x}, \lambda^{-1}(Gx - GJ^G_{\lambda A}x) - z^*\rangle \geqslant 0.$$

Multiplying by $-\lambda$ yields

$$\langle \overrightarrow{xJ^G_{\lambda A}x}, GJ^G_{\lambda A}x - Gx\rangle \leqslant 0$$

and the strict monotonicity of $G$ yields $x = J^G_{\lambda A}x$, i.e. $x \in \mathrm{Fix}(J^G_{\lambda A})$.

For the converse, let $x$ be a fixed point of $J^G_{\lambda A}$. Then by definition $x \in \mathrm{dom}G$ and there exists a $z^* \in A(J^G_{\lambda A}x) = Ax$ such that

$$\lambda^{-1}(Gx - GJ^G_{\lambda A}x) =_{\mathcal{D}} z^*.$$

Further, we have

$$|\langle \overrightarrow{ab}, z^*\rangle| = |\langle \overrightarrow{ab}, \lambda^{-1}(Gx - GJ^G_{\lambda A}x)\rangle| \leqslant \lambda^{-1} d_X(a,b) d_Y(Gx, GJ^G_{\lambda A}x) = 0.$$

Thus $z^* =_{\mathcal{D}} \mathcal{O}$ and thus $x$ is a zero of $A$. □

## 8. The proximal point algorithm and (quantitative) convergence

Having monotone operators and resolvents at our disposal, we now want to consider an associated proximal point type method. As already discussed in the introduction, the proximal point method belongs both in normed as well as in nonlinear spaces to the most seminal and well-studied methods in nonlinear optimization and we refer for more information to the references given there.

Concretely, we now want to establish the convergence of a proximal point method associated with the relativized resolvents of a monotone operator $A$, i.e. of the iteration

$$(\dagger) \qquad x_0 \in X,\ x_{n+1} := J^G_{r_n A} x_n$$

for a given sequence $(r_n) \subseteq (0, \infty)$ with $\inf\{r_n \mid n \in \mathbb{N}\} =: r > 0$ and where $G : X \to Y$ is a uniformly continuous and uniformly strictly monotone Fréchet derivative of a convex function $f : X \to \mathbb{R}$ such that $D^G_f$ is consistent. In order to sustain this iteration, we have to make the common assumption that the resolvents of the operator $A$ have a large enough domain. For that, we in the following assume the condition that

$$\mathrm{dom}J^G_{\lambda A} = X \text{ for all } \lambda > 0$$

which we call the *G-relativized range condition* (which is just a relativized version of the range condition already considered in [24]). By the results from [6] it follows that if $f$ is a Fréchet differentiable function in a reflexive Banach space which is strictly convex and cofinite, then any maximally monotone operator satisfies the $\nabla f$-relativized range condition. Further, since in a complete CAT(0)-space $X$ with the usual dual $X^*$, the mapping $J^{\overrightarrow{oI}}_{\lambda\partial_{\mathcal{D}^*}f}(x)$ for a proper, convex and continuous mapping $f$ is just the resolvent as defined by Jost [21] (recall [24] and Example 7.8), any such subdifferential in a complete CAT(0)-space satisfies the $\overrightarrow{oI}$-relativized range condition.

We will show the convergence of this method over dual systems $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ where a suitable subset of $X$ is totally bounded and to do that, we reduce this convergence result to that of a more general iteration of a family $(T_n)$ of uniform BSNE maps $T_n : X \to X$, namely the iteration

$$(*) \qquad x_0 \in X,\ x_{n+1} := T_n x_n,$$

where the family $(T_n)$ relates to another uniform BSNE map $T$ in a suitable way.

The main feature exploited in the corresponding convergence proof given later is that the sequence $(x_n)$ defined by $(*)$ is then Fejér monotone w.r.t. to the distance $D^G_f$ and the set $\bigcap_n \mathrm{Fix}(T_n)$, i.e. for $p \in \bigcap_n \mathrm{Fix}(T_n)$ and any $n$, it holds that

$$D^G_f(p, T_n x_n) \leqslant D^G_f(p, x_n)$$

since $T_n$ is BQNE w.r.t. $\mathrm{Fix}(T_n)$, which actually follows from $T_n$ being uniformly BSNE as if $D^G_f(p, T_n x) > D^G_f(p, x)$, i.e. $0 > D^G_f(p, x) - D^G_f(p, T_n x)$, then $D^G_f(T_n x, x) = 0$ by being uniformly BSNE which yields $T_n x = x$ by consistency of $D^G_f$ and this entails $D^G_f(p, T_n x) = D^G_f(p, x)$ which is a contradiction.

For Fejér monotone sequences, very general convergence results exist which in compact spaces guarantee the "strong" convergence of such methods if they in addition satisfy a suitable asymptotic condition (essentially amounting in one form or another to having approximate solutions along the iteration). Such general results deal e.g. with Fejér monotone sequences in general metric spaces [33, 34], sequences in Banach spaces which are Fejér monotone w.r.t. Bregman distances [4] or sequences in Hilbert spaces which are Fejér monotone w.r.t. to variable metrics [16]. All of these general results however do not cover the present case as we are not set in normed spaces and since $D^G_f$ here is not a metric. However, in the recent work on generalized Fejér monotone sequences [45], the results from [33, 34] have been generalized to allow for very general classes of distances (e.g. in particular also containing consistent metric Bregman distances $D^G_f$ as defined here as well as consistent Bregman distances in normed spaces), all in a metric setting. As such, our convergence results for the methods discussed above will arise as special instances of the work [45] and in that way this paper actually presents the first instantiation of the general scheme of iterations considered in [45] that goes beyond the previous literature on Fejér monotone sequences. As these general results arose from the proof mining program, we in particular also obtain quantitative information for these convergence results in the form of a highly uniform and computable rate of metastability, i.e. a bound on the $n$ in the expression

$$\forall k \in \mathbb{N}, g \in \mathbb{N}^{\mathbb{N}} \exists n \in \mathbb{N} \forall i, j \in [n; n + g(n)] \left( d_X(x_i, x_j) \leqslant \frac{1}{k+1} \right)$$

in terms of $k$ and $g$. This notion of so-called metastability is, albeit noneffectively, equivalent to the usual Cauchy property of the sequence $(x_n)$ in question and represents a particularly fruitful finitary phrasing of that property, as in particular also highlighted by Tao [59, 60]. In the context of quantitative aspects of convergence for Fejér monotone sequences, it however gains even further importance as already in the most simple cases of ordinary Fejér monotonicity, e.g. taking the real numbers as the underlying space and assuming that all the involved data are computable, there, in general, are no computable rates of convergence as one can show using methods from computability theory (essentially reducing to a seminal paper of Specker [57], see [33] for a discussion on this). In that way, if one aims at computable information on the quantitative aspects of the convergence of such iterations, a rate of metastability is in this generality the best one can hope to attain. However, under a general additional metric regularity assumption, we will nevertheless also be able to obtain explicit, uniform and computable constructions of full rates of convergence for the iterations considered here, which will similarly arise by instantiating the general results from [45].

Let us now fix the remaining data surrounding the iteration defined by $(*)$: With the iteration in $(*)$, we in the end want to approximate fixed points of another mapping $T : X \to X$ and for this, we need to relate the $T_n$'s to this mapping $T$. We do this via requiring that there exists a pair of moduli $\nu(\varepsilon, b)$ and $\mu(\varepsilon, b, K)$ such that for any $\varepsilon, b > 0$ and $K \in \mathbb{N}$ as well as any $p \in \overline{B}_b(o)$:

$$\forall n \in \mathbb{N}\, (d_X(p, T_n p) < \nu(\varepsilon, b) \text{ implies } d_X(p, Tp) < \varepsilon)$$

as well as

$$d_X(p, Tp) < \mu(\varepsilon, b, K) \text{ implies } \forall n \leqslant K\, (d_X(p, T_n p) < \varepsilon)\,.$$

We call such a pair of moduli *fixed point moduli*[1] for $(T_n)$ w.r.t. $T$ and we say that $\mathrm{Fix}(T_n) = \mathrm{Fix}(T)$ holds uniformly if such moduli exist.

Further, we in the following assume that all $T_n$ are uniformly BSNE with a common modulus $\omega(\varepsilon, b)$ and we pick $b$ with

$$b \geqslant D^G_f(p_0, x_0), d_X(p_0, o), d_X(x_n, o)$$

where $p_0 \in \bigcap_n \mathrm{Fix}(T_n) \neq \varnothing$ is an arbitrary but fixed solution. We call a family of mappings $(T_n)$ where such a common modulus $\omega$ exists *jointly uniformly BSNE*. We also fix a modulus $\omega^G$ of uniform continuity for $G$ on $o$-bounded sets and a modulus $\eta$ of uniform strict monotonicity for $G$ on $o$-bounded sets. Lastly, we fix a modulus of consistency $\rho$ for $D^G_f$.

*Remark* 8.1. Note that from $b \geqslant D^G_f(p_0, x_0), d_X(p_0, o)$ as well as a modulus $\beta$ for the uniform $o$-boundedness of $D^G_f$, an upper bound for $d_X(x_n, o)$ can be calculated as we have

$$D^G_f(p_0, x_n) \leqslant D^G_f(p_0, x_0) \leqslant b$$

and so we in particular would get $\beta(b, b) \geqslant d_X(x_n, o)$.

The concrete convergence result for the iteration in $(*)$ that we then obtain is the following:

**Theorem 8.2.** *Let $X$ be a hyperbolic space, $Y$ be a metric space and let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ be a dual system. Assume that $X$ is proper. Let $f : X \to \mathbb{R}$ be Fréchet differentiable with a gradient $G$ which is strictly monotone. Further assume that $D^G_f$ is consistent and uniformly $o$-bounded. Let $T_n : X \to X$ together with $T : X \to X$ be mappings such that $\bigcap_n \mathrm{Fix}(T_n) \neq \varnothing$, such that $\mathrm{Fix}(T_n) = \mathrm{Fix}(T)$ holds uniformly, such that $(T_n)$ is jointly uniformly BSNE and such that $T$ is uniformly continuous on $o$-bounded sets.*

*Then $(x_n)$ defined by $(*)$ converges to a fixed point of $T$.*

[1]These moduli are defined in analogy to [47] where they were considered for a uniform quantitative analogue of the influential NST condition (see e.g. [1]).

In particular, as mentioned before, we obtain this theorem through a quantitative variant giving a highly uniform and computable rate of metastability as discussed before. Further instantiating this theorem will then allow us to also derive the following analogous convergence result for the iteration in (†):

**Theorem 8.3.** *Let $X$ be a hyperbolic space, $Y$ be a metric space and let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ be a dual system. Assume that $X$ is proper. Let $f : X \to \mathbb{R}$ be Fréchet differentiable with a gradient $G$ which is strictly monotone. Further assume that $D^G_f$ is consistent and uniformly o-bounded. Let $A$ be a monotone operator that satisfies the $G$-relativized range condition and assume that $\mathrm{zer}A \neq \emptyset$.*

*Then $(x_n)$ defined by (†) converges to a zero of $A$.*

Again, also this theorem comes equipped with a quantitative variant, giving a computable and uniform rate of metastability.

In particular, both the proximal point algorithm and the iteration of families of uniform BSNE maps defined above provide analogues in hyperbolic spaces for the previous results on a proximal point algorithm involving relativized resolvents and iterations of Bregman strongly nonexpansive maps in normed spaces as e.g. considered in [17, 39, 41, 47, 49]. Further, the proximal point method considered here in particular includes that devised in the seminal work [24] in the context of monotone operators in CAT(0)-spaces and, as a special case, we therefore reobtain the "strong" convergence of that iteration in proper CAT(0)-spaces by setting $f = d^2(o, x)/2$ and $G = \overrightarrow{oI}$ (recall for this in particular the Examples 3.4, 4.5, 5.2, 5.6 and 5.10).

The proofs of both results with their preceding quantitative variants will be given in the next subsection, with constructions for rates of convergence using the additional assumption of a modulus of regularity in the subsequent subsection.

8.1. **Uniform Fejér monotonicity and rates of metastability.** We now want to apply the abstract results of [45] to the particular instance of the iteration given by (∗). These results given in [45] rely on uniform reformulations of the respective properties (like Fejér monotonicity) in terms of approximations instead of full solutions (see [45] for details), and we consequently consider such approximate variants here as well. Concretely, we write $X_0 = \{x \in X \mid d(o, x) \leqslant b\}$ for the $b$ fixed before and define

$$F := \mathrm{Fix}(T) \cap X_0 = \{p \in X_0 \mid Tp = p\}$$

as the designated set of solutions and we set

$$AF_k := \left\{ p \in X_0 \mid d_X(p, Tp) \leqslant \frac{1}{k+1} \right\}$$

as a choice for a set of "$1/(k+1)$-approximate solutions".

Then, we can now provide the relevant results on the asymptotic behavior and Fejér monotonicity of the iteration and translate these into the corresponding bounds and moduli required in the general abstract setup presented in [45] in the context of these uniform reformulations (which we outline in the respective results so that essentially no precise familiarity with [45] is presupposed).

The first such quantitative result is that of the asymptotic regularity of the iteration $x_n$ generated by $(*)$. For this, we first show the following:

**Lemma 8.4.** *For any $\varepsilon > 0$ and any $g : \mathbb{N} \to \mathbb{N}$:*

$$\exists n \leqslant \psi_{b,\omega}(\varepsilon, g) \forall k \in [n; n+g(n)] \left( D^G_f(x_{k+1}, x_k) < \varepsilon \right)$$

*where*

$$\psi_{b,\omega}(\varepsilon, g) = \tilde{g}^{(\lceil b/\omega(\varepsilon,b) \rceil)}(0)$$

*with $\tilde{g}(n) = n + g(n) + 1$.*

*Proof.* We have

$$0 \leqslant D^G_f(p_0, x_{n+1}) \leqslant D^G_f(p_0, x_n) \leqslant D^G_f(p_0, x_0) \leqslant b$$

as so similar as in Proposition 2.27 of [29], we get for

$$\varphi(\varepsilon, g) = \tilde{g}^{\lceil b/\varepsilon \rceil}(0)$$

that for any $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$:

$$\exists n \leqslant \varphi(\varepsilon, g) \forall i, j \in [n; n+g(n)+1] \left( |D^G_f(p, x_i) - D^G_f(p, x_j)| < \varepsilon \right).$$

Thus we in particular have

$$\exists n \leqslant \varphi(\varepsilon, g) \forall k \in [n; n+g(n)] \left( |D^G_f(p, x_k) - D^G_f(p, T_k x_k)| < \varepsilon \right).$$

Using $\omega$, we thus get the result as $\psi_{b,\omega}(\varepsilon, g) = \varphi(\omega(\varepsilon, b), g)$ □

*Remark* 8.5. The whole result and construction is similar to an analogous result on Bregman strongly nonexpansive maps in Banach spaces discussed in [47], which in turn is similar to preceding results on strongly quasi-nonexpansive maps discussed in [30].

Combined with fixed point moduli for $(T_n)$ w.r.t. $T$ as well as a modulus of consistency for $D^G_f$, we then get the following result:

**Lemma 8.6.** *For any $\varepsilon > 0$ and any $g : \mathbb{N} \to \mathbb{N}$:*

$$\exists n \leqslant \psi_{b,\omega}(\rho(\nu(\varepsilon, b), b), g) \forall k \in [n; n+g(n)] \left( d_X(x_k, Tx_k) < \varepsilon \right)$$

*where $\psi_{b,\omega}$ is defined as in Lemma 8.4. In particular, $(d_X(x_k, Tx_k))_k$ converges to 0 for $k \to \infty$.*

*Proof.* Using Lemma 8.4, we get an $n \leqslant \psi_{b,\omega}(\rho(\nu(\varepsilon, b), b), g)$ such that for any $k \in [n; n+g(n)]$:

$$D^G_f(x_{k+1}, x_k) < \rho(\nu(\varepsilon, b), b).$$

Thus, for any such $k$, we have first

$$d_X(x_k, T_k x_k) = d_X(x_{k+1}, x_k) < \nu(\varepsilon, b)$$

by the properties of $\rho$ and thus we get $d_X(x_k, Tx_k) < \varepsilon$ by the properties of $\nu$. To see that $(d_X(x_k, Tx_k))_k$ converges to 0, assume the contrary, i.e.

$$\exists \varepsilon > 0 \forall n \in \mathbb{N} \exists k \in \mathbb{N} \left( d_X(x_{n+k}, Tx_{n+k}) \geqslant \varepsilon \right).$$

Define $g(n)$ to be such a $k$ for any given $n$. Then we have

$$\forall n \in \mathbb{N} \left( d_X(x_{n+g(n)}, Tx_{n+g(n)}) \geqslant \varepsilon \right).$$

But this is a contradiction to the above as for this $\varepsilon$ and $g$, we have to have that

$$\exists n \leqslant \psi_{b,\omega}(\rho(\nu(\varepsilon, b), b), g) \forall k \in [n; n+g(n)] \left( d_X(x_k, Tx_k) < \varepsilon \right).$$

□

We now fit this general asymptotic regularity result for $(x_n)$ into the framework of [45]. For this, it actually suffices to only have a modulus witnessing that $(x_n)$ contains approximate fixed points of $T$ in the sense of the $AF_k$. However, because of the setup from [45], we have to convert to errors of the form $1/(k+1)$ and consider the sequence $x_{2n}$ instead of $x_n$.[2]

**Lemma 8.7.** *The sequence $(x_{2n})$ has asymptotic $F$-points in the sense of* [33, 45] *with a respective bound $\Phi$, i.e for any $k \in \mathbb{N}$:*

$$\exists n \leqslant \Phi(k)\left(x_{2n} \in AF_k\right),$$

*where the so-called "approximate $F$-point bound" $\Phi$ is defined by*

$$\Phi(k) = 2\left\lceil \frac{b}{\omega(\rho(\nu(1/(k+1), b), b), b)} \right\rceil + 1.$$

*Proof.* This follows immediately from Lemma 8.6 by taking $g(n) := 1$ for all $n \in \mathbb{N}$ and $\varepsilon := 1/(k+1)$. □

As discussed before, since all the $T_n$'s are uniformly BSNE, they are also BQNE w.r.t. $\mathrm{Fix}(T_n)$ and therefore the sequence $(x_n)$ is Fejér monotone w.r.t. $D^G_f$ and $\bigcap_n \mathrm{Fix}(T_n) = \mathrm{Fix}(T)$. The quantitative results from [45] then rely on a uniform quantitative version of this Fejér monotonicity relative to the sets $AF_k$ in the form of a so-called *modulus of uniform Fejér monotonicity* which is a bound (viz. witness) on the $k$ in the expression

$$\forall r, n, m \in \mathbb{N} \exists k \in \mathbb{N} \forall p \in AF_k \forall l \leqslant m \left( D^G_f(p, x_{n+l}) < D^G_f(p, x_n) + \frac{1}{r+1} \right)$$

in terms of the $r, n, m$. By a simple compactness argument, such a modulus always exists for $(x_n)$ defined by $(*)$ if $X_0$ is compact (and $F$ is suitably closed relative to the $AF_k$ in a way that will be discussed further below) but there is a rather small enrichment of the main assumptions on the operators $T_n$ that allows one to immediately derive such a modulus for the iteration given by $(*)$:

**Definition 8.8.** We call $T$ uniformly BQNE with a modulus $\varpi : (0,\infty)^2 \to (0,\infty)$ if for all $\varepsilon, b > 0$ and all $p, x \in \overline{B}_b(o)$:

$$d_X(p, Tp) < \varpi(\varepsilon, b) \text{ implies } D^G_f(p, Tx) < D^G_f(p, x) + \varepsilon.$$

Then, if all $T_n$ are actually uniformly BQNE, we can immediately derive a rather simple modulus of uniform Fejér monotonicity:

**Lemma 8.9.** *Let $T_n$ be uniformly BQNE with a respective modulus $\varpi_n$ for every $n \in \mathbb{N}$. Then, the sequence $(x_n)$ is uniformly Fejér monotone w.r.t. $F$, $(AF_k)$ and $D^G_f$ in the sense of* [45] *with a respective modulus $\chi_0$, i.e for any $r, n, m \in \mathbb{N}$ and all $p \in AF_{\chi_0(n,m,r)}$:*

$$\forall l \leqslant m \left( D^G_f(p, x_{n+l}) < D^G_f(p, x_n) + \frac{1}{r+1} \right)$$

*where the "modulus of uniform Fejér monotonicity" $\chi_0$ is defined by*

$$\chi_0(n, m, r) := \mu(\cdot, b, \widehat{n+m} \dot{-} 1)\left( \varpi^M_{n+m \dot{-} 1}\left( \frac{1}{m(r+1)+1}, b \right) \right)$$

*where $\varpi^M_i(\varepsilon, b) := \min\{\varpi_j(\varepsilon, b) \mid j \leqslant i\}$.*

[2]That $x_{2n}$ has to be considered instead of $x_n$ is due to the fact that [45] allows for the treatment of sequences which are only partially Fejér monotone in a certain sense and $x_{2n}$ instead of $x_n$ has to be considered to translate from the setup of ordinary Fejér monotonicity to this partial notion.

*Proof.* Let $p \in AF_{\chi_0(n,m,r)}$, i.e.

$$d_X(p, Tp) \leqslant \frac{1}{\chi_0(n,m,r)+1} < \mu\left(\varpi^M_{n+m\dot{-}1}\left(\frac{1}{m(r+1)+1}, b\right), b, n+m \dot{-} 1\right).$$

Thus, for any $l \leqslant m$, we get

$$d_X(p, T_{n+l\dot{-}1}p) < \varpi^M_{n+m\dot{-}1}\left(\frac{1}{m(r+1)+1}, b\right) \leqslant \varpi_{n+l\dot{-}1}\left(\frac{1}{m(r+1)+1}, b\right).$$

So, if w.l.o.g. $l \geqslant 1$, we have

$$\begin{aligned} D^G_f(p, x_{n+l}) &< D^G_f(p, x_{n+l-1}) + \frac{1}{m(r+1)+1} \\ &< \dots \\ &< D^G_f(p, x_n) + \frac{l}{m(r+1)+1} \\ &\leqslant D^G_f(p, x_n) + \frac{1}{r+1} \end{aligned}$$

which was the claim. □

The second to last quantitative ingredient that we need in order to apply the results from [45] are moduli witnessing that $F$ is closed in a suitably nice sense w.r.t. the $AF_k$.

**Lemma 8.10.** *Let $\Delta$ be a modulus of uniform continuity for $T$ on o-bounded subsets. Then the set $F$ is uniformly closed w.r.t. $AF_k$ in the sense of* [33, 45] *with respective moduli $\delta_1, \delta_2$, i.e. for any $p, q \in X_0$:*

$$q \in AF_{\delta_1(k)} \text{ and } d_X(q,p) \leqslant \frac{1}{\delta_2(k)+1} \text{ implies } p \in AF_k,$$

*where the so-called "moduli of uniform closedness" $\delta_1, \delta_2$ are defined by*

$$\delta_1(k) := 3k+2 \text{ and } \delta_2(k) := \max\left\{3k+2, \widehat{\Delta(\cdot,b)}(3k+2)\right\}.$$

*Proof.* Let $q \in AF_{\delta_1(k)}$, i.e. $d_X(q, Tq) \leqslant 1/3(k+1)$ and let $d_X(q,p) \leqslant 1/(\delta_2(k)+1)$. We have

$$d_X(p,Tp) \leqslant d_X(Tp,Tq) + d_X(q,Tq) + d_X(q,p) \leqslant \frac{1}{3k+3} + d_X(Tp,Tq) + d_X(p,q).$$

As $\delta_2(k) \geqslant 3k+2$, we further derive $d_X(p,q) \leqslant 1/(3k+3)$ so that

$$d_X(p,Tp) \leqslant \frac{2}{3k+3} + d_X(Tp,Tq)$$

and as

$$\delta_2(k) \geqslant \widehat{\Delta(\cdot,b)}(3k+2) = \left\lceil \frac{1}{\Delta(\frac{1}{3k+3}, b)} \right\rceil$$

we have

$$d_X(p,q) < \Delta\left(\frac{1}{3k+3}, b\right)$$

and thus $d_X(Tp,Tq) < 1/(3k+3)$ which yields $d_X(p,Tp) \leqslant 1/(k+1)$. □

As hinted on in the initial remarks of this section, we need to assume that $X_0$ is totally bounded to guarantee the convergence of the iteration in the sense of the metric. Quantitatively, this total boundedness is witnessed by a corresponding modulus as introduced by Gerhardy in [19]:

**Definition 8.11** ([19])**.** A set $A \subseteq X$ of a metric space $(X, d_X)$ is totally bounded with a modulus of total boundedness $\gamma$ if for any $k \in \mathbb{N}$ and any $(x_n) \subseteq A$:

$$\exists 0 \leqslant i < j \leqslant \gamma(k) \left( d_X(x_i, x_j) \leqslant \frac{1}{k+1} \right).$$

In particular, note that a set is totally bounded if, and only if, it has a modulus of total boundedness.

Using this modulus, we can now formulate our main quantitative result:

**Theorem 8.12.** *Assume that $X_0$ is totally bounded with a modulus of total boundedness $\gamma$ and that $(x_n)$ is uniformly Fejér monotone w.r.t. $F$, $(AF_k)$ and $D_f^G$ in the sense of* [45] *with a respective modulus $\chi_0$. Then $(x_n)$ is Cauchy and for all $k \in \mathbb{N}$ and $g \in \mathbb{N}^{\mathbb{N}}$:*

$$\exists N \leqslant \Psi(k, g) \forall i, j \in [N; N + g(N)] \left( d_X(x_i, x_j) \leqslant \frac{1}{k+1} \text{ and } x_i \in AF_k \right),$$

*where $\Psi(k, g) := 2\Psi_0(n_0, k, g)$ with $n_0 := \gamma'(16\tilde{\theta}(k) + 20)$ for $\gamma'(k) := \gamma(\Lambda(k))$ where $\lambda(k) := \widehat{\rho(\cdot, b)}(k)$ and $\Lambda(k) := \widehat{P(\cdot, b)}(k)$ with $P$ as in Lemma 5.7 as well as*

$$\theta(k) := \lambda(2k+1, b) \text{ and } \tilde{\theta}(k) := \max\{\theta(k), \widehat{\Delta(\cdot, b)}(3k+2), 3k+2\}.$$

*Here, $\Psi_0$ is defined by*

$$\begin{cases} \Psi_0(0, k, g) := 0, \\ \Psi_0(n+1, k, g) := \Phi^M(\tilde{\eta}_k^M(\Psi_0(n, k, g), 4\tilde{\theta}(k) + 12)), \end{cases}$$

*where*

$$\tilde{\eta}_k(n, r) := \max\left\{ 3k+2, \chi_0\left(2n, 2\left\lfloor \frac{g(2n)}{2} \right\rfloor, r\right), \chi_0\left(2n, 2\left\lfloor \frac{g(2n)}{2} \right\rfloor + 1, r\right), \chi_0(2n, 0, 4r+7) \right\}$$

*together with*

$$\Phi(k) := 2\left\lceil \frac{b}{\omega(\rho(\nu(1/(k+1), b), b), b)} \right\rceil + 1$$

*and where $\Phi^M(k) := \max\{\Phi(j) \mid j \leqslant k\}$.*

*In particular, if each $T_n$ is uniformly BQNE with a respective modulus $\varpi_n$, then $\chi_0$ can be taken to be*

$$\chi_0(n, m, r) := \widehat{\mu(\cdot, b, n + m} \dot{-} 1) \left( \varpi^M_{n+m \dot{-} 1} \left( \frac{1}{m(r+1)+1}, b \right) \right)$$

*where $\varpi_i^M(\varepsilon, b) := \min\{\varpi_j(\varepsilon, b) \mid j \leqslant i\}$.*

*Proof.* The result arises as an instantiation of Theorem 5.5 from [45]. For this, $\phi$ as well as $\psi$ are instantiated with $D_f^G$. All other moduli are set to be the identity. The modulus $\chi$ in Theorem 5.5 from [45] is instantiated with $\chi_0(2n, 2m, r)$ (note for this Proposition 4.26 from [45]). Similarly, $\zeta$ in Theorem 5.5 from [45] is instantiated with $\chi_0(2n, 2m+1, r)$ (note again Proposition 4.26 from [45]). The modulus $\Phi$ in Theorem 5.5 from [45] is just instantiated with $\Phi$ from Lemma 8.7. Lastly, $\omega$ and $\delta$ are instantiated with $\delta_1$ and $\delta_2$ from Lemma 8.10, respectively.

The bounds given here just result from those given in [45] by obvious simplifications and instantiations and the claim regarding the special case when all $T_n$ are uniformly BQNE follows immediately from Lemma 8.9. □

From this quantitative result, we can now in particular derive the new convergence result given in Theorem 8.2:

*Proof of Theorem 8.2.* Note that through the assumptions of Theorem 8.2, all the moduli required for Theorem 8.12 actually exist: A suitable $b$ exists by Remark 8.1. Further, by simple compactness arguments using that $X$ is proper, we get that as $f$ is Fréchet differentiable with a gradient $G$, it is even uniformly Fréchet differentiable with the gradient $G$ and as $G$ is strictly monotone, it is similarly even uniformly strictly monotone. Also, since the sequence $(x_n)$ is Fejér monotone w.r.t. $D^G_f$ and $F$, and since $F$ is uniformly closed w.r.t. $AF_k$, another simple compactness argument gives that $(x_n)$ is actually uniformly Fejér monotone w.r.t. $F$, $(AF_k)$ and $D^G_f$. Therefore, we get that

$$(+) \qquad \forall k \in \mathbb{N}, g \in \mathbb{N}^{\mathbb{N}} \exists N \in \mathbb{N} \forall i, j \in [N; N + g(N)] \left( d_X(x_i, x_j) \leqslant \frac{1}{k+1} \text{ and } x_i \in AF_k \right)$$

by Theorem 8.12. This now implies the Cauchyness of $(x_n)$. To see this, suppose $(x_n)$ is not Cauchy. Then there exists a $k \in \mathbb{N}$ such that for any $N \in \mathbb{N}$, there exists an $n \in \mathbb{N}$ with

$$d_X(x_N, x_{N+n}) > \frac{1}{k+1}.$$

Given any $N$, define $g(N) = n$ for such an $n$. Then this $k$ and $g$ are a contradiction to the above $(+)$. Therefore $(x_n)$ is Cauchy and as every proper space is complete, $(x_n)$ converges to a limit $\widehat{x}$. We show now $\widehat{x} \in AF_k$ for all $k$ which yields $\widehat{x} \in F = \mathrm{Fix}(T) \cap X_0$. For this, let $k$ be given. Then as $\widehat{x}$ is the limit of $x_n$, we get

$$\exists N_0 \forall n \geqslant N_0 \left( d_X(x_n, \widehat{x}) \leqslant \frac{1}{\delta_2(k)+1} \right).$$

for $\delta_2$ from Lemma 8.10. Take $g(N) = N_0$. Then $(+)$ yields

$$\exists N_1 \forall i \in [N_1; N_1 + N_0] \left( x_i \in AF_{\delta_1(k)} \right)$$

for $\delta_1$ from Lemma 8.10. Combined, we have

$$x_{N_1+N_0} \in AF_{\delta_1(k)} \text{ and } d_X(x_{N_1+N_0}, \widehat{x}) \leqslant \frac{1}{\delta_2(k)+1}$$

which yields $\widehat{x} \in AF_k$ by Lemma 8.10. □

The above results for the iteration defined by $(*)$ are still relatively abstract, working with many moduli assumed for some arbitrary but fixed family $(T_n)$. We now instantiate this further with picking $T_n := J^G_{r_nA}$ and $T := J^G_{rA}$ for $r := \inf\{r_n \mid n \in \mathbb{N}\} > 0$ for a monotone operator $A$ with $\mathrm{zer}A \neq \varnothing$ satisfying the $G$-relativized range condition as before, yielding a similar result for the iteration in (†). For this, we still fix a modulus $\omega^G$ of uniform continuity on $o$-bounded sets for $G$ and a modulus $\eta$ of uniform strict monotonicity for $G$ on $o$-bounded sets. Lastly, we fix a modulus of consistency $\rho$ for $D^G_f$ and a modulus of uniform $o$-boundedness $\beta$ for $D^G_f$. We pick a modulus of reverse consistency $P$ as in Lemma 5.7.

Now, note that by Lemma 6.7, all $J^G_{r_nA}$ are immediately uniformly BSNE with the particularly simple modulus $\omega(\varepsilon, b) = \varepsilon$. We fix $\underline{b}$ with

$$\underline{b} \geqslant D^G_f(p_0, x_0), d_X(p_0, o)$$

where $p_0 \in \bigcap_n \mathrm{Fix}(J^G_{r_nA})$. Using Lemma 6.8, we get that all the $T_n$ and $T$ are $o$-bounded on $o$-bounded sets with a common modulus $E(b) := \beta(E'(b), \underline{b})$ where $E'$ can be defined by

$$E'(b) := D_m(\underline{b}) + D_m(b) + (b + \underline{b})C_n(b)$$

with $D_m, C_n$ as in Lemma 5.4. Lastly, we pick

$$\Delta(\varepsilon, b) := \omega^G\left(\frac{\eta(\varepsilon, E(b))}{2E(b)}, b\right)$$

as in Lemma 7.6 as a modulus of uniform continuity for $T$ (and in fact for all $T_n$'s too, though we do not need this fact here) and we define

$$b := \max\{\underline{b}, \beta(\underline{b}, \underline{b})\}$$

so that we have $b \geqslant d_X(x_n, o), D_f^G(p_0, x_0), d_X(p_0, o)$.

So, the only things left to derive for this particular choice of $T_n$ and $T$ is, for one, a modulus of uniform Fejér monotonicity for the sequence defined by (†) and, for another, the fixed point moduli $\nu$ and $\mu$. We begin with the former and in that context actually utilize Lemma 8.9 by which it suffices to show that each $T_n$ is uniformly BQNE. This in turn follows from the next lemma, by which this modulus can actually even be chosen jointly for all $T_n$:

**Lemma 8.13.** *Let $\lambda > 0$. Then $J^G_{\lambda A}$ is uniformly BQNE with a modulus*

$$\varpi(\varepsilon, b) := \min\{\xi_1(\varepsilon/3, \widehat{b}), \xi_2(\varepsilon/3, \widehat{b})\}$$

*where $\xi_1$, $\xi_2$ are defined as in Lemma 5.8 and $\widehat{b} := \max\{b, E(b)\}$ for $E$ as above.*

*Proof.* Note that by Corollary 7.12, we have for any $x$ and $y$ that

$$D_f^G(J^G_{\lambda A}x, y) - D_f^G(J^G_{\lambda A}x, x) - D_f^G(J^G_{\lambda A}y, y) + D_f^G(J^G_{\lambda A}y, x) \geqslant D_f^G(J^G_{\lambda A}x, J^G_{\lambda A}y) + D_f^G(J^G_{\lambda A}y, J^G_{\lambda A}x).$$

This in particular implies that

$$D_f^G(J^G_{\lambda A}y, J^G_{\lambda A}x) \leqslant D_f^G(J^G_{\lambda A}y, x) + D_f^G(J^G_{\lambda A}x, y) - D_f^G(J^G_{\lambda A}x, J^G_{\lambda A}y).$$

So let $y = p$ and $d_X(J^G_{\lambda A}p, p) < \varpi(\varepsilon, b)$. As thus $d_X(J^G_{\lambda A}p, p) < \xi_1(\varepsilon/3, \widehat{b})$, we get

$$D_f^G(p, J^G_{\lambda A}x) - \varepsilon/3 < D_f^G(p, x) + \varepsilon/3 + D_f^G(J^G_{\lambda A}x, p) - D_f^G(J^G_{\lambda A}x, J^G_{\lambda A}p)$$

and as further $d_X(J^G_{\lambda A}p, p) < \xi_2(\varepsilon/3, \widehat{b})$, we get

$$D_f^G(p, J^G_{\lambda A}x) < D_f^G(p, x) + 2\varepsilon/3 + \varepsilon/3 = \varepsilon.$$

□

Now we turn to the fixed point moduli. For this, we first prove the following rather general lemma on fixed points of relativized resolvents:

**Lemma 8.14.** *Let $r, s > 0$. If $d_X(x, J^G_{sA}x) < \widehat{\nu}_c(\varepsilon, B) := \nu_c(\rho(\varepsilon, \widehat{B}), B)$, then $d_X(x, J^G_{rA}x) < \varepsilon$ where*

$$\nu_c(\varepsilon, B) := \min\left\{\frac{\varepsilon}{4C_n(\widehat{B})}, \omega^G\left(\frac{\varepsilon}{4cE(B)}, \widehat{B}\right)\right\}$$

*where $\widehat{B} := \max\{B, E(B)\}$, $B \geqslant d(x, o)$ and $c \geqslant rs^{-1}$.*

*Proof.* Using the four point identity, we have

$$0 = D_f^G(x, x) = D_f^G(x, J^G_{rA}x) + D_f^G(J^G_{rA}x, x) + \langle \overrightarrow{J^G_{rA}x x}, GJ^G_{rA}x - Gx\rangle.$$

Therefore, we in particular have

$$\begin{aligned}
D^G_f(x, J^G_{rA}x) &\leqslant \langle \overrightarrow{J^G_{rA}xx}, Gx - GJ^G_{rA}x\rangle \\
&= \langle \overrightarrow{J^G_{sA}xx}, Gx - GJ^G_{rA}x\rangle + \langle \overrightarrow{J^G_{rA}xJ^G_{sA}x}, Gx - GJ^G_{rA}x\rangle \\
&= \langle \overrightarrow{J^G_{sA}xx}, Gx - GJ^G_{rA}x\rangle + \langle \overrightarrow{J^G_{rA}xJ^G_{sA}x}, -s^{-1}(Gx - GJ^G_{sA}x) + r^{-1}(Gx - GJ^G_{rA}x)\rangle \\
&\quad + r\langle \overrightarrow{J^G_{rA}xJ^G_{sA}x}, s^{-1}(Gx - GJ^G_{sA}x)\rangle \\
&\leqslant \langle \overrightarrow{J^G_{sA}xx}, Gx - GJ^G_{rA}x\rangle + rs^{-1}\langle \overrightarrow{J^G_{rA}xJ^G_{sA}x}, Gx - GJ^G_{sA}x\rangle \\
&\leqslant d_X(J^G_{sA}x, x)d_Y(Gx, GJ^G_{rA}x) + rs^{-1}d_X(J^G_{rA}x, J^G_{sA}x)d_Y(Gx, GJ^G_{sA}x) \\
&\leqslant d_X(J^G_{sA}x, x)(d_Y(Gx, \mathcal{O}) + d_Y(GJ^G_{rA}x, \mathcal{O})) \\
&\quad + rs^{-1}(d_X(J^G_{rA}x, o) + d_X(J^G_{sA}x, o))d_Y(Gx, GJ^G_{sA}x) \\
&\leqslant d_X(J^G_{sA}x, x)2C_n(\widehat{B}) + rs^{-1}2E(B)d_Y(Gx, GJ^G_{sA}x)
\end{aligned}$$

where for the third to the fourth line, we have used the monotonicity of $A$ and the definition of the resolvent. If $d_X(x, J^G_{sA}x) < \nu_c(\varepsilon, B)$, this then gives that $D^G_f(x, J^G_{rA}x) < \varepsilon$. The result now follows from the assumptions on $\rho$. □

**Lemma 8.15.** *Let $(r_n)$ be given with $\inf\{r_n \mid n \in \mathbb{N}\} =: r > 0$. Define $T_n = J^G_{r_nA}$ and $T = J^G_{rA}$ as well as $\overline{r}_K := \max\{r_n \mid n \leqslant K\}$. Let $\widehat{\nu}_c(\varepsilon, B)$ be defined as in Lemma 8.14. Then $\mathrm{Fix}(T_n) = \mathrm{Fix}(T)$ holds uniformly with fixed point moduli $\overline{\nu}(\varepsilon, B)$ and $\overline{\mu}(\varepsilon, B, K)$ defined by*

$$\overline{\nu}(\varepsilon, B) := \widehat{\nu}_1(\varepsilon, B) \text{ and } \overline{\mu}(\varepsilon, B, K) := \widehat{\nu}_{\overline{r}_K r^{-1}}(\varepsilon, B)$$

*where $\widehat{B} := \max\{B, E(B)\}$.*

*Proof.* The correctness of both moduli $\overline{\nu}$ follows immediately from Lemma 8.14. For $\overline{\nu}$, note that $1 \geqslant rr_n^{-1}$ since $r_n \geqslant r$ and for $\overline{\mu}$, note that $\overline{r}_K r^{-1} \geqslant r_n r^{-1}$ for any $n \leqslant K$. □

As an instantiation of the previous Theorem 8.12, we thus now get the following very concrete result for our analogue of the proximal point algorithm defined in (†).

**Theorem 8.16.** *Assume that $X_0$ is totally bounded with a modulus of total boundedness $\gamma$. Then $(x_n)$ defined by (†) is Cauchy and for all $k \in \mathbb{N}$ and $g \in \mathbb{N}^{\mathbb{N}}$:*

$$\exists N \leqslant \Psi(\lambda(k), g)\forall i, j \in [N; N + g(N)]\left(d_X(x_i, x_j) \leqslant \frac{1}{k+1} \text{ and } x_i \in AF_k\right),$$

*where $\Psi(k, g) := 2\Psi_0(n_0, k, g)$ with $n_0 := \gamma'(16\tilde{\theta}(k) + 20)$ for $\gamma'(k) := \gamma(\Lambda(k))$ where $\lambda(k) := \widehat{\rho(\cdot, b)}(k)$ and $\Lambda(k) := \widehat{P(\cdot, b)}(k)$ with $P$ as in Lemma 5.7 as well as*

$$\theta(k) := \lambda(2\Lambda(k, b) + 1, b) \text{ and } \tilde{\theta}(k) := \max\left\{\theta(k), \left\lceil \frac{1}{\omega^G\left(\frac{\eta(1/(3k+3), E(b))}{2E(b)}, b\right)} \right\rceil, 3k + 2\right\}.$$

*Here, $\Psi_0$ is defined by*

$$\begin{cases} \Psi_0(0, k, g) := 0, \\ \Psi_0(n + 1, k, g) := \Phi^M(\tilde{\eta}^M_k(\Psi_0(n, k, g), 4\tilde{\theta}(k) + 12)), \end{cases}$$

*where*

$$\tilde{\eta}_k(n, r) := \max\left\{3k + 2, \chi_0\left(2n, 2\left\lfloor \frac{g(2n)}{2} \right\rfloor, r\right), \chi_0\left(2n, 2\left\lfloor \frac{g(2n)}{2} \right\rfloor + 1, r\right), \chi_0(2n, 0, 4r + 7)\right\}$$

*for*

$$\chi_0(n,m,r) := \overline{\mu}(\cdot, b, \widehat{n+m} \dot- 1)\left(\varpi\left(\frac{1}{m(r+1)+1}, b\right)\right)$$

*and*

$$\varpi(\varepsilon, b) := \min\{\xi_1(\varepsilon/3, \widehat{b}), \xi_2(\varepsilon/3, \widehat{b})\}$$

*with* $\xi_1$, $\xi_2$ *defined as in Lemma 5.8 and together with*

$$\Phi(k) := 2\left\lceil \frac{b}{\rho(\overline{\nu}(1/(k+1), b), b)} \right\rceil + 1$$

*where* $\Phi^M(k) := \max\{\Phi(j) \mid j \leqslant k\}$ *and where*

$$\overline{\nu}(\varepsilon, B) := \widehat{\nu}_1(\varepsilon, B) \text{ and } \overline{\mu}(\varepsilon, B, K) := \widehat{\nu}_{\overline{r}_K r^{-1}}(\varepsilon, B)$$

*for*

$$\widehat{\nu}_c(\varepsilon, B) := \min\left\{\frac{\rho(\varepsilon, \widehat{B})}{4C_n(\widehat{B})}, \omega^G\left(\frac{\rho(\varepsilon, \widehat{B})}{4cE(B)}, \widehat{B}\right)\right\}$$

*with* $\widehat{B} := \max\{B, E(B)\}$ *and* $\overline{r}_K := \max\{r_n \mid n \leqslant K\}$.

Theorem 8.3 now follows from this quantitative result in a similar way that Theorem 8.2 followed from Theorem 8.12.

8.2. **Moduli of regularity and rates of convergence.** As mentioned before, for Fejér monotone sequences, computable rates of convergence are in general ruled out. However, we can provide such rates under additional assumptions. A large class of such assumptions in the context of Fejér monotone sequences, generalizing various concepts known from nonlinear analysis and optimization such as error bounds and metric subregularity, among others, was introduced and studied in [34] under the name of *moduli of regularity.* In our context, we consider the following instantiation of this notion:

**Definition 8.17.** Let $T : X \to X$ be given and define $F_T(x) := d_X(Tx, x)$. A function $\varphi : (0,\infty) \to (0,\infty)$ is called a modulus of regularity for $T$ w.r.t. Fix$T$ and $S \subseteq X$ if for all $\varepsilon > 0$ and all $x \in S$:

$$F_T(x) < \varphi(\varepsilon) \text{ implies } \mathrm{dist}_X(x, \mathrm{Fix}T) < \varepsilon$$

where $\mathrm{dist}_X(x, K) := \inf_{y\in K} d_X(x, y)$.

The abstract results from [45] can now be used to provide a transformation that combines such a modulus of regularity together with an approximate fixed point bound (and some other minor quantitative data) into a full rate of convergence for the iteration defined by (∗). For this, we just briefly note that the work [34] (and similarly [45] where the results from [34] are extended beyond metrics) is written in the context of a formal setup where, instead of using sets $F/AF_k$ as above to formulate the solutions and approximations, a function $F : X \to [0, +\infty]$ is employed and the sets $F$ and $AF_k$ are (conceptually) replaced by zer$F$ and $\{x \mid F(x) \leqslant \varepsilon\}$ for general $\varepsilon > 0$, respectively. The above notion arises from the general definition given in [34, 45] by using $F_T(x)$ in place of such an $F$. However, in the following, we mostly suppress this whole setup from [34, 45] as we did before with the construction of the rate of metastability and introduce the notions only as needed.

Now, in the context of a modulus of regularity for $T$, the general results from [45] concretely yield the following construction of a rate of convergence for the iteration defined via (∗) if $\mathrm{Fix}(T_n) = \mathrm{Fix}(T)$ holds uniformly.

**Theorem 8.18.** *Let $X$ be a hyperbolic space, $Y$ be a metric space and let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ be a dual system. Let $f : X \to \mathbb{R}$ be Fréchet differentiable with a gradient $G$ which is uniformly continuous on o-bounded sets with a modulus $\omega^G$ and which is uniformly strictly monotone with a modulus $\eta$. Further assume that $D^G_f$ is consistent with a modulus of consistency $\rho$. Let $P$ be as in Lemma 5.7. Let $T_n : X \to X$ together with $T : X \to X$ be mappings such that $p_0 \in \bigcap_n \mathrm{Fix}(T_n) \neq \emptyset$, such that $\mathrm{Fix}(T_n) = \mathrm{Fix}(T)$ holds uniformly with fixed point moduli $\nu$ (and $\mu$) and such that $(T_n)$ is jointly uniformly BSNE with a common modulus $\omega$. Let $(x_n)$ be defined by $(*)$ and let $\varphi$ be a modulus of regularity for $F$ w.r.t. $\mathrm{Fix}(T)$ and $S$ where $(x_n) \subseteq S$. Let $b \geqslant D^G_f(p_0, x_0), d_X(p_0, o), d_X(x_n, o)$.*

*Then $(x_n)$ is Cauchy and*

$$\forall \delta > 0 \forall n, m \geqslant 2\tau\left(\varphi'\left(\frac{\theta(\delta)}{4}\right)\right)(d_X(x_n, x_m) < \delta)$$

*with $\theta(\varepsilon) := \rho(\varepsilon/2, b)$ and $\varphi'(\varepsilon) := \varphi(P(\varepsilon/2, b))$ as well as*

$$\tau(\delta) := 2\left\lceil \frac{b}{\omega(\rho(\nu(\delta, b), b), b)} \right\rceil + 1.$$

*Proof.* The result arises as an instantiation of Theorem 6.6 from [45]. For this, $\phi$ as well as $\psi$ are instantiated with $D^G_f$ as before and the other minor moduli are set to be the identity. Further, $\lambda(\varepsilon)$ is instantiated with $\rho(\varepsilon, b)$ and $\Lambda(\varepsilon)$ is instantiated with $P(\varepsilon, b)$. By the previous discussions on the Fejér monotonicity of the method $(*)$, it is clear that $(x_n)$ is partially $D^G_f$-$(\mathrm{id}, \mathrm{id})$-id-Fejér monotone w.r.t. $\mathrm{Fix}(T)$ in the sense of [45]. Lastly, note that by similar arguments as in Lemma 8.7, $\tau$ as defined above satisfies

$$\forall \delta > 0 \exists n \leqslant \tau(\delta)(d_X(Tx_{2n}, x_{2n}) < \delta)$$

as required by Theorem 6.6 from [45].

The bounds given here just result from those given in [45] by obvious simplifications and instantiations. □

*Remark* 8.19. A set $S$ such that $(x_n) \subseteq S$ can for example be simply given as the closed ball $\overline{B}_b(o)$.

As before, we now explicitly give the instantiation of the above result to the proximal point algorithm (†).

**Theorem 8.20.** *Let $X$ be a hyperbolic space, $Y$ be a metric space and let $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ be a dual system. Let $f : X \to \mathbb{R}$ be Fréchet differentiable with a gradient $G$ which is uniformly continuous on o-bounded sets with a modulus $\omega^G$ and which is uniformly strictly monotone with a modulus $\eta$. Further assume that $D^G_f$ is consistent with a modulus of consistency $\rho$. Let $P$ be as in Lemma 5.7. Let $A$ be a monotone operator that satisfies the $G$-relativized range condition and assume that $\mathrm{zer}A \neq \emptyset$. Let $(x_n)$ be defined by (†) and let $\varphi$ be a modulus of regularity for $J^G_{rA}$ w.r.t. $\mathrm{Fix}(J^G_{rA})$ and $S$ where $(x_n) \subseteq S$. Let $b \geqslant D^G_f(p_0, x_0), d_X(p_0, o), d_X(x_n, o)$.*

*Then $(x_n)$ is Cauchy and*

$$\forall \delta > 0 \forall n, m \geqslant 2\tau\left(\varphi'\left(\frac{\theta(\delta)}{4}\right)\right)(d_X(x_n, x_m) < \delta)$$

*with $\theta(\varepsilon) := \rho(\varepsilon/2, b)$ and $\varphi'(\varepsilon) := \varphi(P(\varepsilon/2, b))$ as well as*

$$\tau(\delta) := 2\left\lceil \frac{b}{\rho(\overline{\nu}(\delta, b), b)} \right\rceil + 1.$$

*where* $\overline{\nu}(\varepsilon, B) := \widehat{\nu}_1(\varepsilon, B)$ *for*

$$\widehat{\nu}_c(\varepsilon, B) := \min\left\{\frac{\rho(\varepsilon, \widehat{B})}{4C_n(\widehat{B})}, \omega^G\left(\frac{\rho(\varepsilon, \widehat{B})}{4cE(B)}, \widehat{B}\right)\right\}$$

*and where* $\widehat{B} := \max\{B, E(B)\}$.

*Remark* 8.21. Note that in the above Theorems 8.18 and 8.20, in the context of a modulus of regularity, the convergence in particular holds without any compactness assumption.

Using a modulus of regularity relative to the resolvents may seem rather counterintuitive in the context of a monotone operator $A$. For that reason, we here want to also consider another function which captures the same regularity notion but which is maybe more naturally defined just in terms of $A$. Concretely, given a dual system $\mathcal{D} = (X, Y, \langle\cdot,\cdot\rangle)$ and an operator $A \subseteq X \times Y$, we define

$$F_A(x) = \inf_{z^* \in Ax} \sup_{a \neq b} \frac{|\langle \overrightarrow{ab}, z^* \rangle|}{d_X(a, b)}.$$

We say that a point $x \in X$ is an $\varepsilon$-approximate zero of $A$ if

$$\exists z^* \in Ax \forall a \neq b \in X \left(|\langle \overrightarrow{ab}, z^* \rangle| < d_X(a, b) \cdot \varepsilon\right).$$

Then it is easy to see that $F_A(x) < \varepsilon$ if, and only if, $x$ is an $\varepsilon$-approximate zero.

Using $F_A$, we get the following notion of a modulus of regularity for an operator $A$:

**Definition 8.22.** Let $A \subseteq X \times Y$ be given and define $F_A$ as above. A function $\varphi : (0, \infty) \to (0, \infty)$ is called a modulus of regularity for $A$ w.r.t. zer$A$ and $S \subseteq X$ if for all $\varepsilon > 0$ and all $x \in S$:

$$F_A(x) < \varphi(\varepsilon) \to \mathrm{dist}_X(x, \mathrm{zer}A) < \varepsilon.$$

Further, $F_A$ relates to the function

$$F_{J^G_{rA}}(x) = d_X(x, J^G_{rA}x)$$

for a given $r > 0$ by the following lemma:

**Lemma 8.23.** *Let $A$ be a monotone operator with* zer$A \neq \emptyset$ *that satisfies the $G$-relativized range condition and let $J^G_{\lambda A}$ be its resolvent relative to $G$ for $\lambda > 0$. Assume that $J^G_{\lambda A}$ is o-bounded on o-bounded sets with a modulus $E$. Let $\Lambda_1 \geqslant \lambda \geqslant \Lambda_0 > 0$ and assume that $G$ is uniformly strictly monotone with a modulus $\eta$.*

(1) *If $\varphi$ is a modulus of regularity for $J^G_{\lambda A}$ w.r.t.* Fix$(J^G_{\lambda A})$ *and $\overline{B}_b(o)$, then*

$$\varphi_1(\varepsilon) := \frac{\eta(\varphi(\varepsilon), \widehat{b})}{2\Lambda_1 \widehat{b}}$$

*where $\widehat{b} := \max\{b, E(b)\}$ is a modulus of regularity for $A$ w.r.t.* zer$A$ *and $\overline{B}_b(o)$.*

(2) *If $\varphi$ is a modulus of regularity for $A$ w.r.t.* zer$(A)$ *and $\overline{B}_{\widehat{b}}(o)$, then*

$$\varphi_0(\varepsilon) := \min\left\{\frac{\varepsilon}{2}, \omega^G\left(\varphi\left(\frac{\varepsilon}{2}\right)\Lambda_0, \widehat{b}\right)\right\}$$

*where $\widehat{b} := \max\{b, E(b)\}$ is a modulus of regularity for $J^G_{\lambda A}$ w.r.t.* Fix$(J^G_{\lambda A})$ *and $\overline{B}_b(o)$.*

*Proof.* (1) Assume $F_A(x) < \varphi_1(\varepsilon)$. Then there is a $z^* \in Ax$ such that
$$|\langle \overrightarrow{ab}, z^* \rangle| < d_X(a,b) \cdot \varphi_1(\varepsilon)$$
for all $a, b \in X$. By monotonicity of $A$, we get
$$\begin{aligned}0 &\leqslant \langle \overrightarrow{xJ^G_{\lambda A}x}, \lambda^{-1}(Gx - GJ^G_{\lambda A}x) - z^* \rangle \\ &= \langle \overrightarrow{xJ^G_{\lambda A}x}, \lambda^{-1}(Gx - GJ^G_{\lambda A}x) \rangle - \langle \overrightarrow{xJ^G_{\lambda A}x}, z^* \rangle.\end{aligned}$$
Rearranging and multiplying by $-\lambda$ yields
$$\begin{aligned}\langle \overrightarrow{xJ^G_{\lambda A}x}, GJ^G_{\lambda A}x - Gx \rangle &\leqslant \lambda \langle \overrightarrow{J^G_{\lambda A}xx}, z^* \rangle \\ &< \lambda d_X(x, J^G_{\lambda A}x)\varphi_1(\varepsilon) \\ &\leqslant 2\lambda\widehat{b}\frac{\eta(\varphi(\varepsilon), \widehat{b})}{2\Lambda_1\widehat{b}} \\ &\leqslant \eta(\varphi(\varepsilon), \widehat{b}).\end{aligned}$$
The uniform strict convexity of $G$ with modulus $\eta$ gives $d_X(x, J^G_{\lambda A}x) < \varphi(\varepsilon)$. As $\varphi$ is a modulus of regularity for $J^G_{\lambda A}$ w.r.t. $\mathrm{Fix}(J^G_{\lambda A})$ and $S$, we get $\mathrm{dist}_X(x, \mathrm{Fix}(J^G_{\lambda A})) < \varepsilon$. By Lemma 7.14, we thus have
$$\mathrm{dist}_X(x, \mathrm{zer}A) \leqslant \mathrm{dist}_X(x, \mathrm{zer}A \cap \mathrm{dom}G) = \mathrm{dist}_X(x, \mathrm{Fix}(J^G_{\lambda A})) < \varepsilon.$$

(2) Assume that $d_X(x, J^G_{\lambda A}x) = F_{J^G_{\lambda A}}(x) < \varphi_0(\varepsilon)$. By definition of $J^G_{\lambda A}$, there exists a $z^* \in A(J^G_{\lambda A}x)$ with
$$z^* =_{\mathcal{D}} \lambda^{-1}(Gx - GJ^G_{\lambda A}x).$$
As $d_X(x, J^G_{\lambda A}x) < \varphi_0(\varepsilon)$, we have for any $a, b \in X$ that
$$\begin{aligned}|\langle \overrightarrow{ab}, z^* \rangle| &= |\langle \overrightarrow{ab}, \lambda^{-1}(Gx - GJ^G_{\lambda A}x) \rangle| \\ &\leqslant \lambda^{-1} d_X(a,b) d_Y(Gx, GJ^G_{\lambda A}x) \\ &\leqslant \lambda^{-1} d_X(a,b)\varphi\left(\frac{\varepsilon}{2}\right)\Lambda_0 \\ &\leqslant d_X(a,b)\varphi\left(\frac{\varepsilon}{2}\right)\end{aligned}$$
and so we get $F_A(J^G_{\lambda A}x) < \varphi(\varepsilon/2)$ which yields $\mathrm{dist}_X(J^G_{\lambda A}x, \mathrm{zer}A) < \varepsilon/2$ and as $d_X(x, J^G_{\lambda A}x) < \varphi_0(\varepsilon) \leqslant \varepsilon/2$, we get
$$\mathrm{dist}_X(x, \mathrm{zer}A) < \varepsilon.$$
□

Note also that $F_A$ reduces to a maybe more intuitive or expected form if the system $\mathcal{D}$ is non-degenerate, as the following lemma shows:

**Lemma 8.24.** *Let $\mathcal{D} = (X, Y, \langle \cdot, \cdot \rangle)$ be a dual system and let $A \subseteq X \times Y$ be an operator. Define*
$$\widehat{F}_A(x) := \mathrm{dist}_Y(\mathcal{O}, Ax) := \inf_{z^* \in Ax} d_Y(\mathcal{O}, z^*).$$
*Then for any $x \in X$ and $\varepsilon > 0$, we have*
$$\widehat{F}_A(x) < \varepsilon \text{ implies } F_A(x) < \varepsilon.$$
*If $\mathcal{D}$ is uniformly non-degenerate with a modulus of uniform non-degenerateness $\Delta(\varepsilon)$, then we also have for any $x \in X$ and $\varepsilon > 0$ that*
$$F_A(x) < \Delta(\varepsilon) \text{ implies } \widehat{F}_A(x) < \varepsilon.$$

*Proof.* Let $x \in X$ and $\varepsilon > 0$ be given and assume that $\widehat{F}_A(x) < \varepsilon$. Then, there exists a $z^* \in Ax$ such that $d_Y(\mathcal{O}, z^*) < \varepsilon$. Therefore, we have

$$\sup_{a \neq b} \frac{|\langle \overrightarrow{ab}, z^* \rangle|}{d_X(a,b)} \leqslant \sup_{a \neq b} \frac{d_X(a,b) d_Y(z^*, \mathcal{O})}{d_X(a,b)} < \varepsilon$$

and thus also $F_A(x) < \varepsilon$.

Now let $\mathcal{D}$ be uniformly non-degenerate with a modulus $\Delta$ and assume $F_A(x) < \Delta(\varepsilon)$. Then there exists a $z^* \in Ax$ such that

$$\sup_{a \neq b} \frac{|\langle \overrightarrow{ab}, z^* \rangle|}{d_X(a,b)} < \Delta(\varepsilon).$$

Using the uniform non-degenerateness, we have $d_Y(\mathcal{O}, z^*) < \varepsilon$ and thus $\widehat{F}_A(x) < \varepsilon$. □

We want to end this section with a few examples of mappings and operators where a modulus of regularity immediately exists.

At first, if $T$ is continuous and if $X$ is proper, a modulus of regularity for $T$ w.r.t. $\mathrm{Fix}T$ and $\overline{B}_r(o)$ always exists for any $r > 0$ (see [34], Corollary 3.5). We refer to [34] for a further extensive discussion of metric notions for $T$ that guarantee a modulus of regularity.

Turning to operators, we can immediately guarantee a modulus of regularity for $A$ if the operator satisfies the following uniform monotonicity condition defined in analogy to the notion from linear spaces:

**Definition 8.25.** Let $\mathcal{D} = (X, Y, \langle \cdot, \cdot \rangle)$ be a dual system. An operator $A \subseteq X \times Y$ is called $\psi$-uniformly monotone if

$$\langle \overrightarrow{yx}, x^* - y^* \rangle \geqslant \psi(d_X(x,y)) d_X(x,y)$$

holds for all $(x, x^*), (y, y^*) \in A$ where $\psi : [0, \infty) \to [0, \infty)$ is a strictly increasing map with $\psi(0) = 0$.

A particular instance of this are so-called $\alpha$-strongly monotone operators which we define by abstracting the corresponding definition from [24] from the context of CAT(0)-spaces.

**Definition 8.26.** Let $\mathcal{D} = (X, Y, \langle \cdot, \cdot \rangle)$ be a dual system. An operator $A \subseteq X \times Y$ is called $\alpha$-strongly monotone if

$$\langle \overrightarrow{yx}, x^* - y^* \rangle \geqslant \alpha d_X^2(x,y)$$

holds for all $(x, x^*), (y, y^*) \in A$ where $\alpha > 0$.

Clearly, an $\alpha$-strongly monotone operator is just $\psi$-uniformly monotone for $\psi(\varepsilon) = \alpha\varepsilon$. Now, for a $\psi$-uniformly monotone operator $A$, we get that $\psi$ itself is a modulus of regularity for $F_A$:

**Lemma 8.27.** *Let $A$ be a $\psi$-uniformly monotone operator with* $\mathrm{zer}A \neq \emptyset$*. Then $\psi$ is a modulus of regularity for $A$.*

*Proof.* Let $x$ be given with $F_A(x) < \psi(\varepsilon)$. Then there exists a $z^* \in Ax$ with

$$|\langle \overrightarrow{ab}, z^* \rangle| < d_X(a,b) \cdot \psi(\varepsilon).$$

Let $y$ be the (unique) zero of $A$ and let $w^* \in Ay$ with $w^* =_{\mathcal{D}} \mathcal{O}$. Then we get

$$\begin{aligned} \psi(d_X(x,y)) d_X(x,y) &\leqslant \langle \overrightarrow{yx}, z^* - w^* \rangle \\ &= \langle \overrightarrow{yx}, z^* \rangle \\ &< d_X(x,y) \cdot \psi(\varepsilon). \end{aligned}$$

Thus we have

$$\psi(d_X(x,y)) < \psi(\varepsilon)$$

and as $\psi$ is strictly increasing, we get $d_X(x,y) < \varepsilon$. As $y \in \text{zer}A$, we get $\text{dist}_X(x, \text{zer}A) < \varepsilon$. □

**Acknowledgments:** I want to thank Ulrich Kohlenbach and Pedro Pinto for stimulating conversations that lead to this work. Further, I want to thank the anonymous referee for the valuable comments that improved the paper at various places. The author was supported by the 'Deutsche Forschungsgemeinschaft' Project DFG KO 1737/6-2.

## References

[1] K. Aoyama and M. Toyoda. Approximation of common fixed points of strongly nonexpansive sequences in a Banach space. *Journal of Fixed Point Theory and Applications*, 21, 2019. Article no. 35.

[2] H.H. Bauschke and J.M. Borwein. Legendre functions and the method of random Bregman projections. *Journal of Convex Analysis*, 4(1):27–67, 1997.

[3] H.H. Bauschke, J.M. Borwein, and P.L. Combettes. Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces. *Communications in Contemporary Mathematics*, 3:615–647, 2001.

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

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

[6] H.H. Bauschke, X. Wang, and L. Yao. General resolvents for monotone operators: characterization and extension. In Y. Censor, M. Jiang, and G. Wang, editors, *Biomedical Mathematics: Promising Directions in Imaging, Therapy Planning and Inverse Problems*, pages 57–74. Medical Physics Publishing, Madison, WI, USA, 2010.

[7] M. Bačák. Convergence of nonlinear semigroups under nonpositive curvature. *Transactions of the American Mathematical Society*, 367:3929–3953, 2015.

[8] I.D. Berg and I.G. Nikolaev. Quasilinearization and curvature of Aleksandrov spaces. *Geometriae Dedicata*, 133(1):195–218, 2008.

[9] L.M. Bregman. The relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7:200–217, 1967.

[10] F.E. Browder. Nonlinear monotone operators and convex sets in Banach spaces. *Bulletin of the American Mathematical Society*, 71(5):780–785, 1965.

[11] F.E. Browder. Nonlinear maximal monotone operators in Banach space. *Mathematische Annalen*, 175:89–113, 1968.

[12] R.E. Bruck. Random products of contractions in metric and Banach Spaces. *Journal of Mathematical Analysis and Applications*, 88(2):319–332, 1982.

[13] F. Bruhat and J. Tits. Groupes réductifs sur un corps local. *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, 41:5–251, 1972.

[14] D. Butnariu and A.N. Iusem. *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization*, volume 40 of *Applied Optimization.* Springer Dordrecht, 2000.

[15] D. Butnariu and E. Resmerita. Bregman distances, totally convex functions and a method for solving operator equations in Banach spaces. *Abstract and Applied Analysis*, pages 1–39, 2006. Art. ID 84919.

[16] P.L. Combettes and B.C. V u. Variable metric quasi-Fejér monotonicity. *Nonlinear Analysis: Theory, Methods & Applications*, 78:17–31, 2013.

[17] J. Eckstein. Nonlinear Proximal Point Algorithms Using Bregman Functions, with Applications to Convex Programming. *Mathematics of Operations Research*, 18(1):202–226, 1993.

[18] G.Z. Eskandani and M. Raeisi. On the zero point problem of monotone operators in Hadamard spaces. *Numerical Algorithms*, 80:1155–1179, 2019.

[19] P. Gerhardy. Proof Mining in Topological Dynamics. *Notre Dame Journal of Formal Logic*, 49(4):431–446, 2008.

[20] K. Goebel and W.A. Kirk. Iteration processes for nonexpansive mappings. In S.P. Singh, S. Thomeier, and B. Watson, editors, *Topological Methods in Nonlinear Functional Analysis*, volume 21 of *Contemporary Mathematics*, pages 115–123. AMS, Providence, 1983.

[21] J. Jost. Convex functionals and generalized harmonic maps into spaces of non positive curvature. *Commentarii Mathematici Helvetici*, 70:659–673, 1995.

[22] B.A. Kakavandi and M. Amini. Duality and subdifferential for convex functions on complete metric spaces. *Nonlinear Analysis: Theory, Methods & Applications*, 73(10):3450–3455, 2010.

[23] T. Kato. Nonlinear semigroups and evolution equations. *Journal of the Mathematical Society of Japan*, 19:508–520, 1967.

[24] H. Khatibzadeh and S. Ranjbar. Monotone operators and the proximal point algorithm in complete CAT (0) metric spaces. *Journal of the Australian Mathematical Society*, 103(1):70–90, 2017.

[25] W.A. Kirk. Krasnosel'skii iteration process in hyperbolic spaces. *Numerical Functional Analysis and Optimization*, 4:371–381, 1982.

[26] W.A. Kirk. Geodesic geometry and fixed point theory II. In J. García-Falset, E. Llorens-Fuster, and B. Sims, editors, *Proceedings of the International Conference on Fixed Point Theory (Valencia 2003)*, pages 113–142. Yokohama Press, 2004.

[27] U. Kohlenbach. Some logical metatheorems with applications in functional analysis. *Transactions of the American Mathematical Society*, 357(1):89–128, 2005.

[28] U. Kohlenbach. A logical uniform boundedness principle for abstract metric and hyperbolic spaces. *Electronic Notes in Theoretical Computer Science*, 165:81–93, 2006.

[29] U. Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer Berlin, Heidelberg, 2008.

[30] U. Kohlenbach. On the quantitative asymptotic behavior of strongly nonexpansive mappings in Banach and geodesic spaces. *Israel Journal of Mathematics*, 216:215–246, 2016.

[31] U. Kohlenbach. Recent progress in proof mining in nonlinear analysis. *IFCoLog Journal of Logics and their Applications*, 10(4):3361–3410, 2017.

[32] U. Kohlenbach. Proof-theoretic Methods in Nonlinear Analysis. In B. Sirakov, P. Ney de Souza, and M. Viana, editors, *Proc. ICM 2018*, volume 2, pages 61–82. World Scientific, 2019.

[33] U. Kohlenbach, L. Leuştean, and A. Nicolae. Quantitative results on Fejér monotone sequences. *Communications in Contemporary Mathematics*, 20:42pp., DOI: 10.1142/S0219199717500158, 2018.

[34] U. Kohlenbach, G. López-Acedo, and A. Nicolae. Moduli of regularity and rates of convergence for Fejér monotone sequences. *Israel Journal of Mathematics*, 232:261–297, 2019.

[35] F. Kohsaka and W. Takahashi. Existence and approximation of fixed points of firmly nonexpansive-type mappings in Banach spaces. *SIAM Journal on Control and Optimization*, 19:824–835, 2008.

[36] F. Kohsaka and W. Takahashi. Fixed point theorems for a class of nonlinear mappings related to maximal monotone operators in Banach spaces. *Archiv der Mathematik*, 21:166–177, 2008.

[37] G. Kreisel. On the Interpretation of Non-Finitist Proofs–Part I. *The Journal of Symbolic Logic*, 16(4):241–267, 1951.

[38] G. Kreisel. On the Interpretation of Non-Finitist Proofs–Part II. Interpretation of Number Theory. Applications. *The Journal of Symbolic Logic*, 17(1):43–58, 1952.

[39] V. Martín-Márquez, S. Reich, and S. Sabach. Right Bregman nonexpansive operators in Banach spaces. *Nonlinear Analysis: Theory, Methods & Applications*, 75:5448–5465, 2012.

[40] V. Martín-Márquez, S. Reich, and S. Sabach. Bregman strongly nonexpansive operators in reflexive Banach spaces. *Journal of Mathematical Analysis and Applications*, 400(2):597–614, 2013.

[41] V. Martín-Márquez, S. Reich, and S. Sabach. Iterative methods for approximating fixed points of Bregman nonexpansive operators. *Discrete and Continuous Dynamical Systems-Series S*, 6(4):1043–1063, 2013.

[42] B. Martinet. Régularisation dinéquations variationnelles par approximations successives. *Revue française d'informatique et de recherche opérationnelle*, 4:154–158, 1970.

[43] G.J. Minty. Monotone networks. *Proceedings of the Royal Society A*, 257:194–212, 1960.

[44] G.J. Minty. Monotone (nonlinear) operators in Hilbert spaces. *Duke Mathematical Journal*, 29:341–346, 1962.

[45] N. Pischke. Generalized Fejér monotone sequences and their finitary content. 2023. Submitted manuscript available at `https://arxiv.org/abs/2312.01852`.

[46] N. Pischke. Proof mining for the dual of a Banach space with extensions for uniformly Fréchet differentiable functions. *Transactions of the American Mathematical Society*, 2024. To appear, doi:10.1090/tran/9226.

[47] N. Pischke and U. Kohlenbach. Effective rates for iterations involving Bregman strongly nonexpansive operators. 2024. Submitted manuscript available at `https://sites.google.com/view/nicholaspischke/notes-and-papers`.

[48] S. Reich. A weak convergence theorem for the alternating method with Bregman distances. In A.G. Kartsatos, editor, *Theory and Applications of Nonlinear Operators of Accretive and Monotone Type*, pages 313–318. Marcel Dekker, New York, 1996.

[49] S. Reich and S. Sabach. A strong convergence theorem for a proximal-type algorithm in reflexive Banach spaces. *Journal of Nonlinear and Convex Analysis*, 10:471–485, 2009.

[50] S. Reich and S. Sabach. Two strong convergence theorems for a proximal method in reflexive Banach spaces. *Numerical Functional Analysis and Optimization*, 31(1):22–44, 2010.

[51] S. Reich and S. Sabach. Existence and approximation of fixed points of Bregman firmly nonexpansive mappings in reflexive Banach spaces. In H.H. Bauschke, R.S. Burachik, P.L. Combettes, V. Elser, D.R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 301–316. Springer, New York, 2011.

[52] S. Reich and I. Shafrir. Nonexpansive iterations in hyperbolic spaces. *Nonlinear Analysis: Theory, Methods & Applications*, 15:537–558, 1990.

[53] E. Resmerita. On total convexity, Bregman projections and stability in Banach spaces. *Journal of Convex Analysis*, 11:1–16, 2004.

[54] R.T. Rockafellar. *Convex analysis*. Princeton university press, 1970.

[55] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal of Control and Optimization*, 14:877–898, 1976.

[56] H.H. Schaefer and M.P. Wolff. *Topological Vector Spaces*. Graduate Texts in Mathematics. Springer New York, 1999. 2nd ed.

[57] E. Specker. Nicht Konstruktiv Beweisbare Sätze Der Analysis. *The Journal of Symbolic Logic*, 14(3):145–158, 1949.

[58] W. Takahashi. A convexity in metric space and nonexpansive mappings I. *Kodai Mathematical Seminar Reports*, 22:142–149, 1970.

[59] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory and Dynamical Systems*, 28(2):657–688, 2008.

[60] T. Tao. *Structure and Randomness: Pages from Year One of a Mathematical Blog*, chapter Soft Analysis, Hard Analysis, and the Finite Convergence Principle, pages 17–29. American Mathematical Society, Providence, RI, 2008.

[61] C. Zălinescu. *Convex analysis in general vector spaces*. World scientific, 2002.