EFFECTIVE RATES FOR ITERATIONS INVOLVING BREGMAN STRONGLY NONEXPANSIVE OPERATORS

NICHOLAS PISCHKE AND ULRICH KOHLENBACH

Department of Mathematics, Technische Universität Darmstadt,
Schlossgartenstraße 7, 64289 Darmstadt, Germany,
E-mails: pischke@mathematik.tu-darmstadt.de, kohlenbach@mathematik.tu-darmstadt.de
Corresponding author: Nicholas Pischke

Abstract. We develop the theory of Bregman strongly nonexpansive maps for uniformly Fréchet differentiable Bregman functions from a quantitative perspective. In that vein, we provide moduli witnessing quantitative versions of the central assumptions commonly used in this field on the underlying Bregman function and the Bregman strongly nonexpansive maps. In terms of these moduli, we then compute explicit and effective rates for the asymptotic regularity of Picard iterations of Bregman strongly nonexpansive maps and of the method of cyclic Bregman projections. Further, we also provide similar rates for the asymptotic regularity and metastability of a strongly convergent Halpern-type iteration of a family of such mappings and we use these new results to derive rates for various special instantiations like a Halpern-type proximal point algorithm for monotone operators in Banach spaces as well as Halpern-Mann- and Tikhonov-Mann-type methods.

Keywords: Bregman strongly nonexpansive mappings, Legendre functions, maximal monotone operators, Bregman projections, proof mining

MSC2010 Classification: 47J25, 65J15, 03F10.

1. Introduction

Monotone set-valued operators $A : X \to 2^X$ in a Hilbert space X are usually studied via their resolvents which are firmly nonexpansive and so, in particular, strongly nonexpansive which is crucially used in many asymptotic regularity and convergence results for important algorithms in convex optimization such as forms of the Proximal Point Algorithm. In the case of more general classes of Banach spaces X, where monotone operators have as values subsets of the dual space X^*, the usual notions of resolvent and metric projection are not even nonexpansive. Replacing the norm by the so-called Bregman distance D_f (which is not a metric) one obtains that suitable concepts of resolvent and (Bregman) projection are firmly nonexpansive and (quasi-)strongly nonexpansive w.r.t. this concept of distance. This makes it possible to extend many algorithms approximating e.g. zeros of monotone operators from the Hilbert space setting to Banach spaces.

In this paper we develop the theory of Bregman strongly nonexpansive operators for uniformly Fréchet differentiable Legendre functions f for the first time from a quantitative point of view, taking the situation of (quasi-)strongly nonexpansive operators in the ordinary metric setting as studied quantitatively in [27] by the second author as a point of departure.

In addition to constructing effective rates of asymptotic regularity for Picard iterations of Bregman strongly nonexpansive mappings and of the cyclic projection method for Bregman projections we also study strongly convergent Halpern-type iterations of Bregman strongly nonexpansive mappings and, in particular, provide explicit rates of asymptotic regularity and metastability (in the sense of T. Tao [60, 61]) for (an extension to sequences of mappings of) the main strong convergence result in [58] which - as special instantiations - yields such rates for a Halpern-type Proximal Point Algorithm, Halpern-type variants of the method of cyclic Bregman projections, of an algorithm for approximating common zeros of finitely many monotone operators as well as forms of Tikhonov-Mann iterations. Finally, we even derive a full rate of convergence for the asymptotic regularity of a certain Halpern-type proximal point algorithm.

Our proofs have been found using the logic-based methodology of proof mining (see [26, 28, 29]), in particular using the recent work [44] of the first author that for the first time provided a treatment in the context of proof mining for the dual space of a Banach space as well as various notions surrounding convex functions, their

Date: November 25, 2024.

gradients and conjugates in Banach spaces. However, as common in the context of applications arising from proof mining, the results in this paper are presented in a way which does not refer to any facts from logic which, however, were instrumental in determining the various moduli in which the aforementioned rates are given. These moduli provide quantitative forms of the basic assumptions on the underlying convex function, its gradient and the Bregman strongly nonexpansive mappings in question. With these data, other assumptions such as e.g. the reflexivity of X become redundant and weak convergence arguments can be circumvented. In this way, the framework presented here is also amenable to metric generalizations such as hyperbolic spaces which was crucially exploited in the recent work by the first author [45]. In particular, the results contained in this paper together with this potential for generalizations show not only that the theory of Bregman distances and their applications to nonlinear analysis is a viable and promising new area for proof mining but also show the applicability of the approach for extending proof mining to these areas as recently proposed in [44].

2. Preliminaries

In this section, we discuss the basic notions surrounding convex functions, their gradients and their corresponding Bregman distances. For further expositions about convex analysis in Banach or Hilbert spaces, we refer to the standard works [4, 54, 56, 66].

2.1. **Convex functions and differentiability.** Throughout, let X be a real Banach space with norm $\|\cdot\|$ and let $f : X \to (-\infty, +\infty]$ be a given function with extended real values. In the following, we will assume that

(1) f is proper, i.e.
$$\mathrm{dom} f := \{x \in X \mid f(x) < +\infty\} \neq \emptyset,$$
(2) f is lower-semicontinuous, i.e.
$$\forall x \in \mathrm{dom} f \ \forall y < f(x) \ \exists \delta > 0 \ \forall z \in B_\delta(x) \left(f(z) > y\right),$$
(3) f is convex, i.e.
$$\forall x, y \in \mathrm{dom} f \ \forall \lambda \in [0,1] \left(f\left(\lambda x + (1-\lambda)\, y\right) \leq \lambda f(x) + (1-\lambda)\, f(y)\right).$$

In this work, we will only consider functions f which are also differentiable. For that, we consider the following notions:

Definition 2.1 (Gâteaux and Fréchet differentiability)**.** A function f is called Gâteaux differentiable at x if there exists an element $\nabla f(x) \in X^*$ such that
$$\lim_{t \to 0} \frac{f(x+ty) - f(x)}{t} = \langle y, \nabla f(x) \rangle$$
for any y. It is called Gâteaux differentiable if it is Gâteaux differentiable at every $x \in \mathrm{intdom} f$. Further, f is called Fréchet differentiable if this limit is uniform in $\|y\| = 1$ and uniformly Fréchet differentiable on a set $C \subseteq X$ if this limit is also uniform in $x \in C$.

In most cases, we will assume that the Fréchet derivative is uniformly norm-to-norm continuous on bounded sets. By a result due to Reich and Sabach, this is the case if f is uniformly Fréchet differentiable on bounded sets:

Proposition 2.2 ([48])**.** *Let X be reflexive. If f is uniformly Fréchet differentiable and bounded on bounded sets, then ∇f is uniformly norm-to-norm continuous on bounded sets.*

The main object instigating duality theory for convex functions is the Fenchel conjugate as introduced in [22] (see also [10, 53]).

Definition 2.3. Given f, define $f^* : X^* \to (-\infty, +\infty]$ by
$$f^*(x^*) := \sup_{x \in X} \left(\langle x, x^* \rangle - f(x)\right).$$

It is immediate from the definition that f^* satisfies the so-called Fenchel-Young inequality: for any $x \in X$ and any $x^* \in X^*$, it holds that
$$f(x) + f^*(x^*) \geq \langle x, x^* \rangle.$$

There now is a plethora of correspondence results for the pair of functions f and f^*. Crucial for this paper, the assumption of f^* being bounded on bounded sets relates to coercivity properties of f through the following result.

Proposition 2.4 ([2])**.** *Call f supercoercive (or strongly coercive) if*

$$\lim_{\|x\|\to+\infty} \frac{f(x)}{\|x\|} = +\infty.$$

Then, the following are equivalent:

(1) f is supercoercive.
(2) f^* is bounded on bounded subsets.

In particular, both imply that $\mathrm{dom} f^* = X^*$.

In all iterations studied later, we will be interested in functions f where both f and f^* are differentiable. This relates to the influential notion of a function being Legendre for which we recall the following definition:

Definition 2.5 ([2])**.** A function f is called:

(1) essentially smooth if ∂f is locally bounded and single-valued on its domain,
(2) essentially strictly convex if $(\partial f)^{-1}$ is locally bounded and f is strictly convex on every convex subset of $\mathrm{dom}\partial f$,
(3) Legendre if it is both essentially smooth and essentially strictly convex.

Over reflexive spaces, being Legendre can be recognized as requiring the differentiability of a function and its conjugate in the following sense:

Proposition 2.6 (essentially [2], see Theorem 5.4 and 5.6 therein)**.** *If X is reflexive, then f is Legendre if, and only if*

(1) It holds that $\mathrm{intdom} f \neq \emptyset$*, that f is Gâteaux differentiable on* $\mathrm{intdom} f$*, and* $\mathrm{dom}\nabla f = \mathrm{intdom} f$.
(2) It holds that $\mathrm{intdom} f^* \neq \emptyset$*, that f^* is Gâteaux differentiable on* $\mathrm{intdom} f^*$*, and* $\mathrm{dom}\nabla f^* = \mathrm{intdom} f^*$.

Further, if f is Legendre, then ∇f is a bijection with $\mathrm{ran}\nabla f = \mathrm{dom}\nabla f^*$, $\mathrm{ran}\nabla f^* = \mathrm{dom}\nabla f = \mathrm{intdom} f$ *and*

$$\nabla f = (\nabla f^*)^{-1}.$$

Remark 2.7. While reflexivity features as a key assumption in the above lemma, if further differentiability assumptions are made regarding f and f^* then reflexivity is an inherent property in that context. Concretely, by a result of Borwein and Vanderwerff [7], any space where f and f^* are Fréchet differentiable, f is continuous and $\mathrm{dom} f^* = X^*$ is already reflexive and it follows from results by Borwein, Guirao, Hájek and Vanderwerff [6] that if f and f^* are uniformly Fréchet differentiable and $\mathrm{dom} f^* = X^*$, then X is even superreflexive.

2.2. **Bregman distances.** The fundamental notion of distance in this work is that of the influential Bregman distance:

Definition 2.8 ([9])**.** Let f be Gâteaux differentiable. The Bregman distance associated with f is the function $D_f : \mathrm{dom} f \times \mathrm{intdom} f \to [0, +\infty)$ which is defined as follows:

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

Crucial for many proofs involving the Bregman distance is the use of the following dual function $W_f : \mathrm{dom} f \times \mathrm{dom} f^* \to [0, +\infty)$ defined by

$$W_f(x, x^*) := f(x) - \langle x, x^*\rangle + f^*(x^*)$$

which often provides a medium through which D_f is studied (see e.g. [37, 38]). If $f : X \to \mathbb{R}$ is Legendre and supercoercive and if X is reflexive, one in particular has that

$$W_f(x, \nabla f(y)) = D_f(x, y)$$

for all $x, y \in X$ as well as that W_f is convex in its right argument and satisfies the inequality

$$W_f(x, x^*) \leq W_f(x, x^* + y^*) - \langle \nabla f^*(x^*) - x, y^*\rangle$$

for any $x \in X$ and any $x^*, y^* \in X^*$ (see [33]).

Lastly, we want to mention the following so-called three and four point identities for D_f:

Lemma 2.9 (folklore, see e.g. [50])**.** *The following inequalities are true for all $x, y, z, w \in$* $\mathrm{intdom} f$*:*

(1) $D_f(x, y) + D_f(y, z) - D_f(x, z) = \langle x - y, \nabla f(z) - \nabla f(y)\rangle$.
(2) $D_f(y, x) - D_f(y, z) - D_f(w, x) + D_f(w, z) = \langle y - w, \nabla f(z) - \nabla f(x)\rangle$.

3. Gradients, Bregman distances and their quantitative properties

Throughout this paper, if not indicated otherwise, we will now assume that f and f^* are total (i.e. $\mathrm{dom} f = X$ and $\mathrm{dom} f^* = X^*$) and that both are Fréchet differentiable *everywhere* with gradients ∇f and ∇f^* (and so - by Remark 2.7 - X will be reflexive).

3.1. Quantitative properties of gradients.

Definition 3.1. We say that a function $\omega^{\nabla f} : (0,\infty)^2 \to (0,\infty)$ is a modulus of uniform continuity on bounded sets for ∇f if for any $\varepsilon, b > 0$ and any $x, y \in \overline{B}_b(0)$:

$$\|x - y\| < \omega^{\nabla f}(\varepsilon, b) \to \|\nabla f(x) - \nabla f(y)\| < \varepsilon.$$

Using such a modulus, we can derive quantitative witnesses for various central properties of ∇f and f:

Lemma 3.2. *Assume that ∇f is uniformly continuous on bounded subsets with a modulus $\omega^{\nabla f}$. Then:*

(1) f is uniformly Fréchet differentiable on bounded subsets with modulus

$$\Delta(\varepsilon, b) := \min\{\omega^{\nabla f}(\varepsilon, b+1), 1\},$$

i.e. for all $b, \varepsilon > 0$ and all $x \in \overline{B}_b(0), y \in X$:

$$0 < \|y\| < \Delta(\varepsilon, b) \to \frac{|f(x+y) - f(x) - \langle y, \nabla f(x)\rangle|}{\|y\|} < \varepsilon.$$

(2) ∇f is bounded on bounded subsets with modulus

$$C(b) := \lceil b/\omega^{\nabla f}(1, b)\rceil + \|\nabla f(0)\| + 1,$$

i.e. for all $b > 0$ and all $x \in \overline{B}_b(0)$:

$$\|\nabla f(x)\| \leq C(b).$$

(3) f is uniformly continuous on bounded subsets with modulus

$$\omega^f(\varepsilon, b) := \frac{\varepsilon}{C(b)},$$

i.e. for all $\varepsilon, b > 0$ and all $x, y \in \overline{B}_b(0)$:

$$\|x - y\| < \omega^f(\varepsilon, b) \to |f(x) - f(y)| < \varepsilon,$$

where C is any modulus witnessing that ∇f is bounded on bounded subsets.

(4) f is bounded on bounded sets with modulus

$$D(b) := \lceil b/\omega^f(1, b)\rceil + |f(0)| + 1,$$

i.e. for all $b > 0$ and all $x \in \overline{B}_b(0)$:

$$|f(x)| \leq D(b),$$

where ω^f is any modulus witnessing that f is uniformly continuous on bounded subsets.

For a proof (formulated with errors of the form 2^{-k} instead of ε), see e.g. [44].

Similar results of course also hold for the conjugate f^* if we assume a modulus of uniform continuity on bounded sets for the respective gradient ∇f^*.

If f is Fréchet differentiable, then the associated Bregman-distance is continuous in both arguments and by analyzing the corresponding proof, we can extract a transformation that turns a modulus for the (uniform) continuity of the gradient of f into a modulus for the (uniform) continuity of the associated Bregman-distance. This is collected in the following lemma:

Lemma 3.3. *Assume that ∇f is uniformly continuous on bounded subsets with a modulus $\omega^{\nabla f}$. Let C be a modulus for ∇f being bounded on bounded sets.*[1]

[1] As shown in the previous Lemma 3.2, such a C can actually be constructed from $\omega^{\nabla f}$. We however throughout work with a given C as a black box so that the contributions of the different types of moduli are highlighted.

(1) *For any $\varepsilon, b > 0$ and any $x, y, y' \in \overline{B}_b(0)$:*

$$\|y - y'\| < \xi(\varepsilon, b) \to |D_f(x, y) - D_f(x, y')| < \varepsilon$$

where $\xi : (0, \infty)^2 \to (0, \infty)$ can be explicitly given by

$$\xi(\varepsilon, b) := \min\left\{\frac{\varepsilon}{4C(b)}, \omega^{\nabla f}\left(\frac{\varepsilon}{4b}, b\right)\right\}.$$

(2) *For any $\varepsilon, b > 0$ and any $x, x', y \in \overline{B}_b(0)$:*

$$\|x - x'\| < \xi'(\varepsilon, b) \to |D_f(x, y) - D_f(x', y)| < \varepsilon$$

where $\xi' : (0, \infty)^2 \to (0, \infty)$ can be explicitly given by

$$\xi'(\varepsilon, b) := \frac{\varepsilon}{2C(b)}.$$

Proof. For item (1), note that we have

$$|\langle y, \nabla f y\rangle - \langle y', \nabla f y'\rangle| \leq \|\nabla f y\| \|y - y'\| + \|y'\| \|\nabla f y - \nabla f y'\|.$$

Using that, we derive

$$|D_f(x, y) - D_f(x, y')| \leq |f(y) - f(y')| + \|x\| \|\nabla f y - \nabla f y'\| + \|\nabla f y\| \|y - y'\| + \|y'\| \|\nabla f y - \nabla f y'\|.$$

This yields the claim by the definition of ξ as by Lemma 3.2, we have that $\varepsilon/4C(b) = \omega^f(\varepsilon/4, b)$ for a suitably defined modulus ω^f for f being uniformly continuous on bounded sets.

For item (2), note that

$$|D_f(x, y) - D_f(x', y)| \leq |f(x) - f(x')| + \|x - x'\| \|\nabla f(y)\|$$

and this yields the claim by the definition of ξ' as by Lemma 3.2, we again have that $\varepsilon/2C(b) = \omega^f(\varepsilon/2, b)$ for a suitably defined modulus ω^f for f being uniformly continuous on bounded sets. □

An assumption that is later used in the context of Halpern-type iterations is that f is uniformly strictly convex on bounded subsets in the sense of [16], i.e.

$$\forall \varepsilon, b > 0\ \exists \delta > 0\ \forall x, y \in X\left(\|x\|, \|y\| \leq b \land \|x - y\| \geq \varepsilon \to \left(f\left(\frac{x + y}{2}\right) \leq \frac{1}{2}f(x) + \frac{1}{2}f(y) - \delta\right)\right).$$

In the following, we will occasionally assume a modulus of uniform strict convexity $\eta : (0, \infty)^2 \to (0, \infty)$ for f, i.e. an η witnessing the above quantifier $\exists \delta > 0$ in terms of ε and b. By the equivalent characterization of strictly convex functions f as those where ∇f is strictly monotone, we can translate such a modulus of uniform strict convexity into a modulus witnessing the 'uniform strict monotonicity' of ∇f, i.e. an $\widehat{\eta} : (0, \infty)^2 \to (0, \infty)$ witnessing δ in terms of ε, b in the following condition:

$$\forall \varepsilon, b > 0\ \exists \delta > 0\ \forall x, y \in X(\|x\|, \|y\| \leq b \land \|x - y\| \geq \varepsilon \to (\langle x - y, \nabla f x - \nabla f y\rangle \geq \delta)).$$

This is collected in the following lemma.

Lemma 3.4. *Let $\eta(\varepsilon, b)$ be a modulus of uniform strict convexity for f. Then $\widehat{\eta}(\varepsilon, b) = 4\eta(\varepsilon, b)$ is a modulus of uniform strict monotonicity for ∇f.*

Proof. Note that we have

$$f\left(\frac{x + y}{2}\right) \leq 1/2f(y) + 1/2f(x) - \eta(\varepsilon, b) = f(x) + 1/2(f(y) - f(x)) - \eta(\varepsilon, b)$$

if $\|x - y\| \geq \varepsilon$ as η is a modulus of uniform strict convexity of f. As $\nabla f w$ is a subgradient of f at w, we have

$$\langle z, \nabla f w\rangle \leq \inf_{\alpha > 0} \frac{f(w + \alpha z) - f(w)}{\alpha},$$

for all w, z and from this we get $\langle y - x, \nabla f x\rangle \leq f(y) - f(x) - 2\eta(\varepsilon, b)$. Similarly, we get $\langle x - y, \nabla f y\rangle \leq f(x) - f(y) - 2\eta(\varepsilon, b)$ and this implies $\langle x - y, \nabla f y - \nabla f x\rangle \leq -4\eta(\varepsilon, b)$ which gives that $\widehat{\eta}(\varepsilon, b) = 4\eta(\varepsilon, b)$ is a modulus of uniform strict monotonicity of ∇f. □

Conversely, also from a modulus $\widehat{\eta}$ for the uniform strict monotonicity we can construct a modulus η for the uniform strict convexity but we omit this other direction as, for one, this construction is rather messy and, for another, the one direction presented above suffices to justify that such an $\widehat{\eta}$ exists in the context of the central assumptions featured in the convergence results later on.

3.2. **Sequential consistency and total convexity.** Another central assumption featuring in the convergence results later on is that of the total convexity of f which we want to discuss in the following. For this, we briefly only assume that $f : X \to (-\infty, +\infty]$ is proper, lower-semicontinuous and convex.

Definition 3.5 (see e.g. [14])**.** Given a function f, define its modulus of total convexity $v_f : \mathrm{intdom} f \times [0, +\infty) \to [0, +\infty]$ by

$$v_f(x, t) := \inf\{D_f(y, x) \mid y \in \mathrm{dom} f, \|y - x\| = t\}.$$

The function f is called totally convex at a point $x \in \mathrm{intdom} f$ if $v_f(x, t) > 0$ whenever $t > 0$. It is called totally convex if it is totally convex at every point. Lastly, we call f totally convex on bounded sets if

$$v_f(B, t) := \inf\{v_f(x, t) \mid x \in B \cap \mathrm{intdom} f\} > 0$$

for any $t > 0$ and for any nonempty bounded set $B \subseteq X$.

This notion is intimately connected with the so-called sequential consistency for the function f:

Definition 3.6 ([16])**.** A function f is called sequentially consistent if for all bounded sequences (x_n) and (y_n) in $\mathrm{intdom} f$:

$$D_f(x_n, y_n) \to 0 \ (n \to \infty) \text{ implies } \|x_n - y_n\| \to 0 \ (n \to \infty).$$

Concretely, the main result connecting total convexity and sequential consistency is now the following:

Lemma 3.7 ([14])**.** *A proper, lower-semicontinuous and convex function $f : X \to (-\infty, +\infty]$ whose domain contains at least two points is totally convex on bounded sets if, and only if, it is sequentially consistent.*

In the following, we will rely on a modulus witnessing the sequential consistency of a function quantitatively. To motivate this, we move to another equivalent way of formulating sequential consistency (which is somewhat in spirit of e.g. Proposition 2.5 of [16], see also [51]). For the following, let f now again be total and Fréchet differentiable everywhere like in the previous standing assumptions.

Lemma 3.8. *A function f is sequentially consistent if, and only if, for all $b > 0$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that*

$$(+) \qquad \forall x, y \in X \left(\|x\|, \|y\| \leq b \wedge D_f(x, y) < \delta \to \|x - y\| < \varepsilon\right).$$

Proof. For sufficiency, consider arbitrary sequences $(x_n), (y_n)$ with $\|x_n\|, \|y_n\| \leq b$ for some $b > 0$ and assume that $\lim D_f(x_n, y_n) = 0$. Let $\varepsilon > 0$ be given. By (+), there is a δ such that

$$(++) \qquad \forall m \in \mathbb{N} \left(D_f(x_m, y_m) < \delta \to \|x_m - y_m\| < \varepsilon\right).$$

Then, by $\lim D_f(x_n, y_n) = 0$ there exists $N \in \mathbb{N}$ such that $\forall m \geq N \, (D_f(x_m, y_m) < \delta)$, which by (++) entails that $\|x_m - y_m\| < \varepsilon$, for all $m \geq N$. This means that $\lim \|x_n - y_n\| = 0$, and we conclude the sequential consistency of f.

For necessity, suppose that (+) fails. Then for some $\varepsilon > 0$ and $b > 0$, we have

$$\forall n \in \mathbb{N} \ \exists x_n, y_n \in X \left(\|x_n\|, \|y_n\| \leq b \wedge D_f(x_n, y_n) < \frac{1}{n+1} \wedge \|x_n - y_n\| \geq \varepsilon\right).$$

Then in particular $D_f(x_n, y_n) < \frac{1}{n+1}$ for all $n \in \mathbb{N}$ which entails that $\lim D_f(x_n, y_n) = 0$. However $\|x_n - y_n\|$ is bounded away from zero by ε, and so f can not be sequentially consistent as x_n and y_n are bounded. □

Definition 3.9. Let f be sequentially consistent. A modulus of consistency for f is a function $\rho : (0, \infty)^2 \to (0, \infty)$ such that for all $b \in \mathbb{N}$ and $\varepsilon > 0$:

$$\forall x, y \in X \left(\|x\|, \|y\| \leq b \wedge D_f(x, y) < \rho(\varepsilon, b) \to \|x - y\| < \varepsilon\right).$$

By the above result, a function f is sequentially consistent if, and only if, it has a modulus of consistency.

We call a modulus of this type but for the converse implication, i.e. translating errors for the metric distance into errors for the Bregman distance, a *modulus of reverse consistency.* Further, such a modulus can actually be computed from a modulus of ∇f being bounded on bounded sets.

Lemma 3.10. *Let ∇f be bounded on bounded sets with a modulus C. Then for all $\varepsilon > 0$ and $b > 0$:*

$$\forall x, y \in X \left(\|x\|, \|y\| \leq b \wedge \|x - y\| < P(\varepsilon, b) \to D_f(x, y) < \varepsilon\right)$$

where $P(\varepsilon, b)$ can be given in terms of C via $P(\varepsilon, b) := \varepsilon / 2C(b)$.

Proof. By Lemma 3.2, we have that $\omega^f(\varepsilon, b) = \varepsilon/C(b)$ is a modulus of uniform continuity for f on bounded sets. So for $\|x-y\| < P(\varepsilon, b) = \omega^f(\varepsilon/2, b)$, we have $f(x) - f(y) < \varepsilon/2$ and thus

$$D_f(x,y) = f(x) - f(y) - \langle x-y, \nabla f(y)\rangle < \varepsilon/2 + \|x-y\| \|\nabla f(y)\| \ \leq \varepsilon/2 + \|x-y\| C(b) < \varepsilon$$

which is the claim. □

We want to note that the collection of such a modulus P together with a modulus of consistency ρ are called moduli of consistency in [43]. In particular, as discussed in [43], these moduli can be used to derive a so-called modulus of weak triangularity for D_f, i.e. a function $\theta : (0,\infty)^2 \to (0,\infty)$ such that[2]

$$\forall \varepsilon, b > 0 \ \forall x,y,z \in X \ (\|x\|, \|y\|, \|z\| \leq b \wedge D_f(y,x), D_f(y,z) < \theta(\varepsilon, b) \to D_f(x,z) < \varepsilon).$$

In other words, θ witnesses that although the triangle inequality is not valid for D_f, it locally behaves similar to a distance function with a triangle inequality. To derive such a θ from a given ρ and P as above, set

$$\theta(\varepsilon, b) = \rho(P(\varepsilon, b)/2, b).$$

Then, if $D_f(y,x), D(y,z) < \theta(\varepsilon, b)$ for $\|x\|, \|y\|, \|z\| \leq b$, we have $\|x-y\|, \|z-y\| < P(\varepsilon, b)/2$ using the properties of ρ. This implies $\|x-z\| < P(\varepsilon, b)$ by triangle inequality of $\|\cdot\|$. So, using the properties of P, this yields $D_f(x,z) < \varepsilon$.

Remark 3.11. Note that in the presence of such moduli ρ and P, all moduli introduced later that depend on measuring a distance $\|x-y\|$ in the premise or conclusion could be translated into moduli that depend on measuring the distance $D_f(x,y)$.

Besides sequential consistency, being totally convex on bounded sets can be further recognized to be equivalent to another well-known convexity property for f already mentioned before, at least in the context of the standing assumptions of this paper.

Lemma 3.12 (essentially [16, Theorem 2.10])**.** *Let $f : X \to \mathbb{R}$ be Fréchet differentiable and let ∇f be uniformly continuous on bounded sets. Then f is totally convex on bounded sets if, and only if, f is uniformly strictly convex on bounded sets.*

In that vein, the following remark shortly discusses the relationship between the modulus of consistency and the previous modulus of uniform strict convexity together with other convexity moduli from the literature.

Remark 3.13. Note that it can be easily shown that ρ is a modulus of consistency if

$$v_f(\overline{B}_b(0), t) \geq \rho(t, b)$$

for any $t, b > 0$ (using e.g. Proposition 2.1 from [16]) and conversely, if ρ is a modulus of consistency, then $v_f(\overline{B}_b(0), t) \geq \rho(t, b+t)$ for any $t, b > 0$. In that way, moduli of consistency as defined in this paper actually immediately witness the total convexity of the function f.

Further, define the modulus of uniform convexity $\mu_f(x,t)$ as in [62] (see also [15, 65]), i.e.

$$\mu_f(x,t) := \inf\left\{ \frac{\lambda f(x) + (1-\lambda) f(y) - f(\lambda x + (1-\lambda) y)}{\lambda(1-\lambda)} \mid y \in X, \|y-x\| = t, \lambda \in (0,1) \right\}$$

and write

$$\mu_f(B,t) := \inf\{\mu_f(x,t) \mid x \in B\}$$

for a given set $B \subseteq X$ similar as with v_f. Similarly define

$$\overline{\mu}_f(x,t) := \inf\left\{ f(x) + f(y) - 2f\left(\frac{x+y}{2}\right) \mid y \in X, \|y-x\| = t \right\}$$

as in [14] (see also [15]). Then as shown in [15], we have

$$\mu_f(x,t) \geq \overline{\mu}_f(x,t) \geq \frac{1}{2}\mu_f(x,t)$$

for any $x \in X$ and $t \geq 0$ as well as $v_f(x,t) \geq \mu_f(x,t)$ for any $x \in X$ and $t > 0$ as shown in [14, Proposition 1.2.5]. Now, it is also immediate that η is a modulus of uniform strict convexity of f as defined above if

$$\frac{1}{2}\overline{\mu}_f(\overline{B}_b(0), t) \geq \eta(t, b)$$

[2]Actually, this notion is introduced in a slightly different manner in [43] but this will not matter in this work.

for any $t, b > 0$ where $\overline{\mu}_f(B,t)$, given a set $B \subseteq X$, is defined similarly as $\mu_f(B,t)$. Conversely, if η is a modulus of uniform strict convexity, then $\overline{\mu}_f(\overline{B}_b(0),t) \geq 2\eta(t, b+t)$ for any $t, b > 0$. Thus any modulus η of uniform strict convexity of f induces a modulus of consistency and thus witnesses the total convexity of f.

Conversely, as follows from the above Lemmas 3.7 and 3.12, if f is Fréchet differentiable with a gradient that is uniformly continuous on bounded sets, then f being sequentially consistent implies f being uniformly strictly convex on bounded sets. As shown in [15], both of these items are further equivalent to f^* being uniformly Fréchet differentiable (and thus to ∇f^* being uniformly continuous on bounded sets if f is also supercoercive by Propositions 2.2 and 2.4).

3.3. **Boundedness properties of the Bregman distance.** As is well-known, the distances D_f in general have very weak properties. In particular, a sequence (x_n) such that $D_f(x_n, y)$ is bounded for some y is not necessarily bounded itself. In that way, it is thus a common requirement in the context of Bregman distances to require that the level sets

$$L_1(y,\alpha) := \{x \in X \mid D_f(x,y) \leq \alpha\},\ L_2(x,\alpha) := \{y \in X \mid D_f(x,y) \leq \alpha\},$$

are bounded for every $\alpha > 0$ and $x, y \in X$. In particular, this condition features in the list of conditions exhibited by Eckstein in [21] and by Butnariu and Iusem in [14] regarding Bregman functions and a stronger requirement of these sets being compact already featured in Bregman's seminal work [9] for the conditions imposed on his general distances D.

As shown in [14], in the case that $L_2(x,\alpha)$ is bounded for all x and α and if f is additionally sequentially consistent, then $L_1(y,\alpha)$ is likewise bounded. Further, as shown in [2] (see Lemma 7.3 therein), if f is supercoercive in a reflexive Banach space, then $L_2(x,\alpha)$ is bounded for any α, which is thus guaranteed in essentially all situations in this paper.

In the following, we will rely on so-called moduli of boundedness for D_f that witness a uniform quantitative version of the boundedness of L_2. Concretely, by a modulus of boundedness for D_f we will mean a function $o : (0,\infty)^2 \to (0,\infty)$ such that

$$\forall x, y \in X\ \forall \alpha, b > 0\, (\|x\| \leq b \wedge D_f(x,y) \leq \alpha \to \|y\| \leq o(\alpha, b))\,.$$

We call D_f uniformly bounded if such a modulus exists.

Remark 3.14. Such a modulus of boundedness for D_f actually exists under certain additional assumptions and in fact can be constructed from respective moduli witnessing these assumptions. Assume that f, ∇f^* are bounded on bounded sets with moduli D, F (which can be constructed from moduli of uniform continuity for ∇f, ∇f^*, respectively, using Lemma 3.2). Now, note that using the Fenchel-Moreau theorem, it holds that $f = f^{**}$ and so, as f is bounded on bounded sets, we get that f^* is supercoercive by Proposition 2.4. Let $\alpha^{f^*}(K)$ be a modulus for that, i.e.

$$\|x^*\| > \alpha^{f^*}(K) \text{ implies } f^*(x^*)/\|x^*\| \geq K$$

for any $x^* \in X^*$ and $K > 0$. Then also $f^*(x^*) - \langle x, x^*\rangle$ is supercoercive with a modulus $\alpha^{f^*}(K + b)$ where $b \geq \|x\|$. Then, if $\|x\| \leq b$ and $D_f(x,y) \leq \alpha$, since $D_f(x,y) = W_f(x, \nabla f y) = f(x) + f^*(\nabla f y) - \langle x, \nabla f y\rangle$, we get that

$$f^*(\nabla f y) - \langle x, \nabla f y\rangle = D_f(x,y) - f(x) \leq \alpha + D(b).$$

Thus $\|\nabla f y\| \leq \max\{\alpha^{f^*}(\alpha + D(b) + b + 1), 1\}$ and thus

$$\|y\| \leq F(\max\{\alpha^{f^*}(\alpha + D(b) + b + 1), 1\}).$$

This gives a modulus $o(\alpha, b)$ as above.

4. Bregman strongly nonexpansive mappings and related notions

The main notion of mapping considered in this paper will be that of a Bregman strongly nonexpansive mapping as introduced in [17, 47].

Let $T : X \to X$ be a mapping. We say that a point $p \in X$ is an asymptotic fixed point of T if there is a sequence (x_n) which converges weakly to p and satisfies $\lim_{n\to\infty} \|x_n - Tx_n\| = 0$. We write $\widehat{F}(T)$ for the set of all such asymptotic fixed points and $F(T)$ for the set of ordinary fixed points of T.

Definition 4.1. A map $T : X \to X$ is called

(1) Bregman nonexpansive (see [38]) if
$$D_f(Tx,Ty) \leq D_f(x,y)$$
for any $x,y \in X$,

(2) Bregman quasi-nonexpansive (see [36, 38]) if
$$D_f(p,Tx) \leq D_f(p,x)$$
for any $x \in X$ and $p \in \widehat{F}(T)$,

(3) Bregman strongly nonexpansive (see [17, 47]) if
$$D_f(p,Tx) \leq D_f(p,x)$$
for any $x \in X$, $p \in \widehat{F}(T)$ and if additionally
$$\lim_{n\to\infty} (D_f(p,x_n) - D_f(p,Tx_n)) = 0 \to \lim_{n\to\infty} D_f(Tx_n,x_n) = 0$$
for any bounded sequence $(x_n) \subseteq X$ and any $p \in \widehat{F}(T)$,

(4) Bregman firmly nonexpansive (see e.g. [3]) if
$$\langle Tx - Ty, \nabla fTx - \nabla fTy\rangle \leq \langle Tx - Ty, \nabla fx - \nabla fy\rangle$$
for all $x,y \in X$.

It is rather immediate to see that being Bregman firmly nonexpansive implies being Bregman strongly nonexpansive (see also Lemma 4.9 later) and it is clear that any Bregman strongly nonexpansive mapping is Bregman quasi-nonexpansive.

We want to note that the above notion of Bregman strongly nonexpansive operators is called strictly left Bregman strongly nonexpansive in other parts of the literature (see in particular [37]) since the fixed points occur in the left argument of the Bregman distance and since we used $\widehat{F}(T)$. If $F(T) = \widehat{F}(T)$ is further assumed, then the resulting notion is called fully left Bregman strongly nonexpansive in these parts of the literature. Note also that Bregman firmly nonexpansive maps are called D-firm in [3] and ∇f firmly nonexpansive in [5].

Fundamental for the quantitative results discussed later for iterations involving such mappings are moduli which quantitatively witness the defining properties of Bregman strongly nonexpansive mappings. The whole approach taken here in regard to quantitative moduli witnessing the Bregman strongly nonexpansiveness is modeled after the work of the second author [27] for 'ordinary' quasi-nonexpansive functions. In these quantitative moduli, it will always be $F(T)$ that we use when deriving the moduli which results e.g. in the fact that instead of full fixed points, these moduli will concern approximate fixed points. If it is presumed that $\widehat{F}(T) = F(T)$ and if this assumption features crucially in a given proof, then a uniform quantitative version of this fact will feature necessarily in its analysis (see the discussion before Theorem 4.15 for this uniform quantitative version).

Definition 4.2. A function $\omega : (0,\infty)^2 \to (0,\infty)$ such that
$$\forall \varepsilon, b > 0 \; \forall p \in F(T) \cap \overline{B}_b(0) \; \forall x \in \overline{B}_b(0) \, (D_f(p,x) - D_f(p,Tx) < \omega(\varepsilon,b) \to D_f(Tx,x) < \varepsilon)$$
is called a BSNE-modulus of T.

If we are given a specific element $p \in F(T)$, we will later say that a function $\omega : (0,\infty)^2 \to (0,\infty)$ is a BSNE-modulus w.r.t. p if
$$\forall \varepsilon, b > 0 \; \forall x \in \overline{B}_b(0) \, (D_f(p,x) - D_f(p,Tx) < \omega(\varepsilon,b) \to D_f(Tx,x) < \varepsilon)$$
holds for that specific p.

We will later be concerned with a stronger type of modulus which only requires p to be a sufficiently good approximate fixed point.

Definition 4.3. A function $\omega : (0,\infty)^2 \to (0,\infty)$ is called a strong BSNE-modulus of T if
$$\forall \varepsilon, b > 0 \; \forall x, p \in X \, (\|p\|, \|x\| \leq b \wedge \|Tp - p\| < \omega(\varepsilon,b) \wedge D_f(p,x) - D_f(p,Tx) < \omega(\varepsilon,b) \to D_f(Tx,x) < \varepsilon) \, .$$
We say that T is uniformly Bregman strongly nonexpansive if it has such a modulus.

Clearly, a strong BSNE-modulus is also an ordinary BSNE-modulus.

Remark 4.4 (For logicians). In the context of the nonstandard uniform boundedness principle $\exists\text{-UB}^X$ (see [25]), which can be conservatively (for statements of the logical form required) added to the systems used in the general logical metatheorems of proof mining as in [26, 28, 44], Bregman strong nonexpansivity implies uniform Bregman strong nonexpansivity. It, moreover, follows that even from a (potentially noneffective) proof of the weaker property, a kind of strict Bregman quasi-nonexpansivity,

$$\forall x, p \in X\, (Tp = p \wedge D_f(p,x) - D_f(p,Tx) \leq 0 \to D_f(Tx,x) = 0)\,,$$

in such a formal system, an explicit and effective strong BSNE-modulus can be extracted. A concrete example for this is presented in Lemma 4.9 below. Note the strong similarity of these circumstances to that of the moduli for the ordinary strong nonexpansivity and their extractibility already from proofs of strict nonexpansivity as discussed in [27].

From the following lemma, we get that a uniformly Bregman strongly nonexpansive map T is in particular Bregman strongly nonexpansive whenever $\widehat{F}(T) = F(T)$.

Lemma 4.5. *Let f be such that D_f satisfies $D_f(x,y) = 0 \leftrightarrow \|x-y\| = 0$ for any $x, y \in X$.*[3] *Let $T : X \to X$ be given. If T satisfies that for any $\varepsilon, b > 0$ there exists a $\delta > 0$ such that for any $p \in \overline{B}_b(0)$ with $\|p - Tp\| < \delta$ and any $x \in \overline{B}_b(0)$:*

$$D_f(p,x) - D_f(p,Tx) < \delta \to D_f(Tx,x) < \varepsilon$$

and if $\widehat{F}(T) = F(T)$, then T is Bregman strongly nonexpansive.

The proof is rather immediate and we hence omit it.

If D_f is uniformly bounded with a modulus of boundedness o as introduced in Section 3.3, then any Bregman quasi-nonexpansive map T with a nonempty fixed point set is bounded on bounded sets and we can also construct a witness for that in the following sense:

Lemma 4.6. *Let T be Bregman quasi-nonexpansive and let $p_0 \in F(T) \neq \emptyset$. Let ∇f, f be bounded on bounded sets with moduli C, D, respectively. Let o be a modulus of boundedness for D_f.*

Then T is bounded on bounded sets with

$$\|Tx\| \leq E(b) := o(2D(b) + 2bC(b), b)$$

for $b \geq \|x\|, \|p_0\|$.

Proof. Note that $D_f(p_0, Tx) \leq D_f(p_0, x)$ as T is Bregman quasi-nonexpansive and thus

$$D_f(p_0,Tx) \leq |f(p_0)| + |f(x)| + |\langle p_0 - x, \nabla f(x)\rangle| \leq 2D(b) + 2bC(b)$$

from which the claim follows using the properties of o. ☐

Conceptually, these strong BSNE-moduli are related to the notion of 'quantitative quasiness' as discussed in [57] and, from such a strong BSNE-modulus, one can in particular derive a modulus $\omega' : (0,\infty)^2 \to (0,\infty)$ which satisfies

$$\forall \varepsilon, b > 0\ \forall x, p \in X\, (\|p\|, \|x\| \leq b \wedge \|Tp - p\| < \omega'(\varepsilon, b) \to D_f(p,Tx) - D_f(p,x) < \varepsilon)\,.$$

This is collected in the following lemma:

Lemma 4.7. *Let ξ be a modulus of uniform continuity on bounded sets for D_f in its second argument and let ρ be a modulus of consistency for f. Let E be a modulus for T being bounded on bounded sets and let ω be a strong BSNE-modulus for T.*

Then there exists an ω' such that

$$\forall \varepsilon, b > 0\ \forall x, p \in X\, (\|p\|, \|x\| \leq b \wedge \|Tp - p\| < \omega'(\varepsilon, b) \to D_f(p,Tx) - D_f(p,x) < \varepsilon)\,.$$

which can be moreover constructed as

$$\omega'(\varepsilon, b) := \omega(\rho(\xi(\varepsilon, \widehat{b}), \widehat{b}), b)$$

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

Proof. If $D_f(p,Tx) - D_f(p,x) \leq 0$, then the claim holds trivially. So suppose $D_f(p,Tx) - D_f(p,x) > 0$. Then trivially $D_f(p,x) - D_f(p,Tx) < 0 < \omega'(\varepsilon, b)$ which implies that $D_f(Tx,x) < \rho(\xi(\varepsilon,\widehat{b}),\widehat{b})$. This yields $\|Tx - x\| < \xi(\varepsilon, \widehat{b})$. Thus, we in particular have that $D_f(p,Tx) - D_f(p,x) < \varepsilon$. ☐

[3]Naturally, this is the case if f is strictly convex.

In the following, we will call such an ω' a derived modulus of ω.

As mentioned above, any Bregman firmly nonexpansive map is Bregman strongly nonexpansive. From the proof of this fact, we can immediately extract a (strong) BSNE-modulus for any Bregman firmly nonexpansive map T. Crucial for this is the following equivalent characterization of Bregman firmly nonexpansive mappings:

Lemma 4.8 ([3])**.** *A map $T : X \to X$ is Bregman firmly nonexpansive if, and only if,*

$$D_f(Tx,Ty) + D_f(Ty,Tx) \leq D_f(Tx,y) + D_f(Ty,x) - D_f(Tx,x) - D_f(Ty,y).$$

for all $x, y \in X$.

Lemma 4.9. *Let T be a Bregman firmly nonexpansive map which is bounded on bounded sets with a modulus E and let ξ, ξ' be moduli that D_f is uniformly continuous on bounded sets in its right and left argument, respectively.*

Then T is uniformly Bregman strongly nonexpansive with a strong BSNE-modulus ω defined by

$$\omega(\varepsilon, b) := \min\{\xi(\varepsilon/4, \widehat{b}), \xi'(\varepsilon/4, \widehat{b}), \varepsilon/4\}$$

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

Further, one can choose $\omega(\varepsilon, b) := \varepsilon$ as a BSNE-modulus for any Bregman firmly nonexpansive T.

Proof. For the strong modulus, let x, p be given. Using Lemma 4.8 with $y = p$, we get

$$D_f(Tx,Tp) + D_f(Tp,Tx) \leq D_f(Tx,p) + D_f(Tp,x) - D_f(Tx,x).$$

Rearranging yields

$$\begin{aligned} D_f(Tx,x) &\leq D_f(Tx,p) - D_f(Tx,Tp) + D_f(Tp,x) - D_f(Tp,Tx) \\ &\leq (D_f(Tx,p) - D_f(Tx,Tp)) + (D_f(Tp,x) - D_f(p,x)) \\ &\quad + (D_f(p,x) - D_f(p,Tx)) + (D_f(p,Tx) - D_f(Tp,Tx)). \end{aligned}$$

Thus if $\|p\|, \|x\| \leq b$ and $\|Tp - p\| < \omega(\varepsilon, b)$ as well as $D_f(p,x) - D_f(p,Tx) < \omega(\varepsilon, b) \leq \varepsilon/4$, then we get $D_f(Tx,x) < \varepsilon$.

For the ordinary BSNE-modulus, note that if $p = Tp$, then Lemma 4.8 with $y = p$ even yields $D_f(p,Tx) \leq D_f(p,x) - D_f(Tx,x)$ which is equivalent to $D_f(Tx,x) \leq D_f(p,x) - D_f(p,Tx)$ which yields the given modulus. □

Compare this BSNE-modulus in particular to the modulus extracted in [27] for ordinary (meaning in the usual metric sense) strongly (quasi-)nonexpansive maps which even in the simple case of Hilbert spaces (where the notions of firmly nonexpansive and Bregman firmly nonexpansive for $f = \|\cdot\|^2/2$ coincide) is quadratic in ε. By taking a look at the above proof, this seems due to the fact even in the Hilbert case with the specific choice $f = \|\cdot\|^2/2$, the distance D_f fits closer to the notion of firmly nonexpansive maps and the quadratic increase comes from converting from D_f to the usual norm.

A concrete example for Bregman firmly nonexpansive mappings are the resolvents Res^f_A relative to f for a given monotone operator A in Banach spaces. For this, we first recall the notion of monotone operators.

Definition 4.10 ([11, 12])**.** Let $A : X \to 2^{X^*}$ be a set-valued operator. The operator A is called monotone if

$$\langle x - y, x^* - y^* \rangle \geq 0$$

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

Further, A is called maximally monotone if its graph is not strictly contained in the graph of another monotone operator.

The f-resolvents of A are then defined using ∇f:[4]

Definition 4.11 ([3, 21])**.** Let $A : X \to 2^{X^*}$ be a set-valued operator. Given f, we define the resolvent of A relative to f as the operator $\mathrm{Res}^f_A : X \to 2^X$ with

$$\mathrm{Res}^f_A(x) := \left((\nabla f + A)^{-1} \circ \nabla f\right)(x).$$

[4]The idea of considering the above notion in general Banach spaces is due to [3] (where it was introduced under the name of D-resolvents) but this notion of a resolvent relative to f was already considered by Eckstein in [21] in the context of finite-dimensional spaces.

The following properties are essential for the resolvent relative to f:

Proposition 4.12 ([3]). *Let $f : X \to (-\infty, +\infty]$ be a function which is proper, convex, lower semicontinuous, Gâteaux differentiable and strictly convex on* $\mathrm{intdom} f$ *and let A be a monotone operator such that* $\mathrm{intdom} f \cap \mathrm{dom} A \neq \emptyset$. *Then following statements hold:*

(1) $\mathrm{dom}\,\mathrm{Res}^f_A \subseteq \mathrm{intdom} f$ *and* $\mathrm{ran}\,\mathrm{Res}^f_A \subseteq \mathrm{intdom} f$,
(2) Res^f_A *is single-valued on its domain,*
(3) $F(\mathrm{Res}^f_A) = \mathrm{intdom} f \cap A^{-1}0$,
(4) Res^f_A *is Bregman firmly nonexpansive on its domain.*

Further, the classical result for monotone operators in Hilbert spaces established by Minty [39] that maximal monotonicity is equivalent to the totality of the resolvents extends to these resolvents relative to f under suitable assumptions on f:

Proposition 4.13 ([5]). *Let X be reflexive. Let A be monotone and assume that $f : X \to \mathbb{R}$ is Gâteaux differentiable, strictly convex and cofinite (i.e.* $\mathrm{dom} f^* = X^*$*). Then A is maximal monotone if and only if* $\mathrm{ran}(A + \nabla f) = X^*$.

As we will mostly consider a fixed operator A in the following, we introduce a more compact notation for resolvents with real parameters in such a case: given $\gamma > 0$, we simply write Res^f_γ for $\mathrm{Res}^f_{\gamma A}$.

Important for the study of resolvents are their corresponding Yosida approximates defined by

$$A^f_\gamma(x) := \frac{1}{\gamma}\left(\nabla f(x) - \nabla f \mathrm{Res}^f_\gamma(x)\right)$$

for a given $\gamma > 0$.

It follows essentially by the definitions of Res^f_γ and A^f_γ (see e.g. [49]) that we have $(\mathrm{Res}^f_\gamma x, A^f_\gamma x) \in A$ for any $\gamma > 0$ and any $x \in \mathrm{dom}\,\mathrm{Res}^f_A$.

By the above results, as any Res^f_γ is Bregman firmly nonexpansive, all such resolvents for a maximal monotone A have the same BSNE-modulus (and also the same strong BSNE-modulus if they are bounded on bounded sets with a common modulus).

The resolvents relative to f also include Bregman projections (see [9]) as these can be considered to be special resolvents: If C is a nonempty, closed and convex subset, we may define the indicator function

$$\iota_C(x) := \begin{cases} 0, & \text{if } x \in C, \\ +\infty, & \text{if } x \notin C. \end{cases}$$

It is straightforward to see that this function is proper, lower-semicontinuous and convex. Therefore, the subgradient $\partial\iota_C$ is maximally monotone [52, 55]. The Bregman projection P^f_C is then defined as the resolvent $\mathrm{Res}^f_{\partial\iota_C}$ and in particular is Bregman firmly nonexpansive. Thus also here, the above moduli apply.

In general, already for Bregman firmly nonexpansive mappings, it is not immediately clear which (if any) form of ordinary metric continuity such mappings inherit. However, if one assumes that ∇f is uniformly continuous on bounded subsets as well as uniformly strictly monotone, then at least every Bregman firmly nonexpansive map that is bounded on bounded sets (i.e., by the above lemma, in particular any such map with a fixed point) is indeed uniformly continuous on bounded subsets.

Lemma 4.14. *Let T be Bregman firmly nonexpansive and assume that T is bounded on bounded sets with a modulus E. Assume that ∇f is uniformly continuous on bounded sets with a modulus $\omega^{\nabla f}$ and that it is uniformly strictly monotone with a modulus η, i.e.*

$$\forall \varepsilon, b > 0\ \forall x, y \in X\ (\|x\|, \|y\| \leq b \wedge \langle x - y, \nabla f x - \nabla f y\rangle < \eta(\varepsilon, b) \to \|x - y\| < \varepsilon).$$

Then T is uniformly continuous on bounded sets with

$$\forall \varepsilon, b > 0\ \forall x, y \in \overline{B}_b(0)\ \left(\|x - y\| < \omega^{\nabla f}(\eta(\varepsilon, E(b))/2E(b), b) \to \|Tx - Ty\| < \varepsilon\right).$$

Proof. Let x, y be given with $\|x\|, \|y\| \leq b$. As T is Bregman firmly nonexpansive, we get by definition that

$$\begin{aligned}\langle Tx - Ty, \nabla fTx - \nabla fTy\rangle &\leq \langle Tx - Ty, \nabla fx - \nabla fy\rangle \\ &\leq \|Tx - Ty\| \|\nabla fx - \nabla fy\| \\ &\leq 2E(b) \|\nabla fx - \nabla fy\|.\end{aligned}$$

In particular, if $\|x - y\| < \omega^{\nabla f}(\varepsilon/2E(b), b)$, we have $\langle Tx - Ty, \nabla fTx - \nabla fTy\rangle < \varepsilon$ and so the result follows immediately using the properties of the modulus η. □

A crucial feature of strongly nonexpansive maps (in the usual sense) as compared to e.g. firmly nonexpansive maps is that they are closed under composition. A similar result holds for Bregman strongly nonexpansive maps as established in [37]. We now derive a quantitative variant that allows one to combine (strong) BSNE-moduli for the factors into a (strong) BSNE-modulus for the composition. This result is similar to the corresponding results for 'ordinary' (quasi-)strongly nonexpansive maps given in [27] (see Theorem 2.10 and Theorem 4.6 therein).

However, before we move to this result on moduli for compositions, we first consider a quantitative treatment of the fact that fixed points of compositions of Bregman strongly nonexpansive operators are fixed points of the factors (see e.g. Proposition 3.4 in [37]). This result, however, crucially relies on the fact that $\widehat{F}(T) \subseteq F(T)$ and so here, we will have to rely on a quantitative treatment of this aspect. The inclusion $\widehat{F}(T) \subseteq F(T)$ concretely expresses the closure property

$$\forall x \in X, (x_n) \subseteq X\ (\|x_n - Tx_n\| \to 0 \text{ and } x_n \to x \text{ (weakly) implies } x = Tx)$$

of which the underlying logical methods used in this paper suggest the following uniform quantitative version to be necessary in the analysis:

$$\forall \varepsilon, b > 0\ \exists \kappa > 0\ \forall x, y \in X\ (\|x\|, \|y\| \leq b \wedge \|y - Ty\|, \|y - x\| < \kappa \to \|x - Tx\| < \varepsilon).$$

We call a function $\kappa(\varepsilon, b)$ that provides witness for such a κ in terms of ε, b a modulus of uniform closedness for $F(T)$ (and we call $F(T)$ uniformly closed if such a modulus exists) as this kind of modulus is essentially just a concrete instantiation of the moduli of uniform closedness considered in an abstract context in [31]. In particular, we want to note that this modulus can from a logical perspective be recognized as a quantitative form of a weak extensionality principle for T, namely

$$\forall x, y\ (y = Ty \wedge x = y \to x = Tx)$$

which has previously received attention in proof mining, in particular due to the fact that there are meaningful classes of maps that posses such moduli of uniform closedness but fail to be uniformly continuous (as e.g. maps satisfying Suzuki's (E) condition [23, 59], see also the discussions in [28, 31]).

In the presence of such a modulus, we can now turn to the following quantitative result (which is anyhow analogous to Proposition 4.15 from [27]):

Theorem 4.15. *Let ξ be a modulus of uniform continuity on bounded subsets for D_f in its second argument. Let θ be a modulus of weak triangularity for D_f. Let ρ be a modulus of consistency for f and let P be a modulus for reverse consistency for f. Let $T_1, \dots, T_N : X \to X$ be Bregman strongly nonexpansive with a (not necessarily strong) BSNE-modulus ω w.r.t. some common fixed point $p \in \bigcap_{i=1}^N F(T_i)$. Let κ be a common modulus of uniform closedness of $F(T_1), \dots, F(T_N)$.*

Then for all $\varepsilon > 0$:

$$\|T_N \circ \dots \circ T_1 x - x\| < P(\varphi(\varepsilon, b, N), b) \to \bigwedge_{i=1}^N \|x - T_i x\| < \varepsilon$$

whenever $b \geq \|x\|, \|p\|$ and $b \geq \|T_k \circ \dots \circ T_1 x\|$ for $1 \leq k \leq N$ where $\varphi(\varepsilon, b, N) := \chi_b(N - 1, \varepsilon)$ and, given b, $\chi_b : \mathbb{N} \times (0, \infty) \to (0, \infty)$ is defined by

$$\begin{cases}\chi_b(0, \varepsilon) := \min\{\rho(\kappa(\varepsilon, b), b), \rho(\varepsilon, b)\}, \\ \chi_b(n+1, \varepsilon) := \min\{\rho(\xi(\omega(\min\{\theta(\chi_b(n, \varepsilon), b), \rho(\kappa(\varepsilon, b), b)\}, b), b), b), \chi_b(n, \varepsilon), \theta(\chi_b(n, \varepsilon), b)\}.\end{cases}$$

In particular, if E is a common modulus for $T_1, \dots, T_N$ being bounded on bounded sets, then above claim holds for $b \geq \|x\|, \|p\|$ and $P(\varphi(\varepsilon, \widehat{b}, N), \widehat{b})$ with $\varphi(\varepsilon, b, N) := \chi_{\widehat{b}}(N - 1, \varepsilon)$ and where $\widehat{b} := \max\{b, E(b), \dots, E^{(N)}(b)\}$.

We refer to the appendix for a proof of this result.

We now turn to the following result on moduli for compositions of Bregman strongly nonexpansive maps (which is modeled after Theorem 2.10 and Theorem 4.6 from [27]):

Theorem 4.16. *Let ξ be a modulus of uniform continuity on bounded subsets for D_f in its second argument. Let θ be a modulus of weak triangularity for D_f. Let ρ be a modulus of consistency for f and let P be a modulus of reverse consistency. Let $T_1, \dots, T_n : X \to X$ be uniformly Bregman strongly nonexpansive maps with strong BSNE-moduli $\omega_1, \dots, \omega_n$ and derived moduli $\omega_1', \dots, \omega_n'$ and assume that the T_i's have a common fixed point. Let κ be a common modulus of uniform closedness of $F(T_1), \dots, F(T_N)$.*

Then $T := T_n \circ \dots \circ T_1$ is uniformly Bregman strongly nonexpansive with modulus

$$\omega(\varepsilon, b) := \min\left\{\widehat{\omega}(\varepsilon, b)/2, P(\varphi(\min\{\widehat{\omega}'(\varepsilon, b), \widehat{\omega}(\varepsilon, b)\}, \widehat{b}, n), \widehat{b})\right\}$$

where

$$\widehat{\omega}(\varepsilon, b) := \min\left\{\omega_1(\rho(P(\varepsilon, \widehat{b})/n, \widehat{b}), \widehat{b}), \dots, \omega_n(\rho(P(\varepsilon, \widehat{b})/n, \widehat{b}), \widehat{b})\right\},$$

$$\widehat{\omega}'(\varepsilon, b) := \min\left\{\omega_1'(\widehat{\omega}(\varepsilon, b)/2(n-1), \widehat{b}), \dots, \omega_n'(\widehat{\omega}(\varepsilon, b)/2(n-1), \widehat{b})\right\},$$

and where φ is defined as in Theorem 4.15, where $\widehat{b} := \max\{b, E(b), \dots, E^{(n)}(b)\}$ for b satisfying $b \geq \|q\|$ for a common fixed point q of the T_i's and where E is a common modulus for $T_1, \dots, T_n$ being bounded on bounded sets.

If the ω_i are ordinary BSNE-moduli, then ω defined by

$$\omega(\varepsilon, b) := \min\left\{\omega_1(\rho(P(\varepsilon, \widehat{b})/n, \widehat{b}), \widehat{b}), \omega_2(\rho(P(\varepsilon, \widehat{b})/n, \widehat{b}), \widehat{b}), \dots, \omega_n(\rho(P(\varepsilon, \widehat{b})/n, \widehat{b}), \widehat{b})\right\}$$

is a BSNE-modulus for $T = T_n \circ \dots \circ T_1$ where $\widehat{b}$ is defined as before.

Again, we refer to the appendix for a proof of this result.

The last type of operation on Bregman strongly nonexpansive operators that we consider here is that of the block operator introduced in [37, 38]:

Definition 4.17 ([37, 38])**.** Let T_i, $i = 1, \dots, N$, be finitely many operators and let $w_i \in [0, 1]$, $i = 1, \dots, N$, be finitely many weights with $\sum_{i=1}^N w_i = 1$. Then the associated block operator is defined as

$$Tx := \nabla f^* \left(\sum_{i=1}^N w_i \nabla f T_i x\right).$$

In particular, as shown in [37, 38], such block operators, if composed of Bregman strongly nonexpansive maps, are again Bregman strongly nonexpansive. For a quantitative version of the said result, we consider the following lemmas.

At first, we note that a block operator is bounded on bounded sets if its summands are.

Lemma 4.18. *Let ∇f, ∇f^* be bounded on bounded sets with moduli C, F, respectively. Let T_i, $i = 1, \dots, N$, be finitely many operators which are bounded on bounded sets with a common modulus E and let $w_i \in [0, 1]$, $i = 1, \dots, N$, be finitely many weights with $\sum_{i=1}^N w_i = 1$.*

Then the associated block operator T is bounded on bounded sets with a modulus $E'(b) := F(C(E(b)))$.

Proof. For $\|x\| \leq b$, we clearly have

$$\left\|\sum_{i=1}^N w_i \nabla f T_i x\right\| \leq \sum_{i=1}^N w_i \|\nabla f T_i x\| \leq C(E(b))$$

and thus $\|Tx\| = \left\|\nabla f^* \sum_{i=1}^N w_i \nabla f T_i x\right\| \leq F(C(E(b)))$. □

As shown in [38], one has $F(T) \subseteq F(T_i)$ for a block operator T and a summand T_i. The following lemma gives a quantitative version of this, translating bounds for approximate fixed points.

Theorem 4.19. *Let ξ be a modulus of uniform continuity of D_f in its second argument. Let T_i, $i = 1, \dots, N$, be finitely many Bregman strongly nonexpansive operators with a (not necessarily strong) BSNE-modulus ω and let $w_i \in [0,1]$, $i = 1, \dots, N$, be finitely many weights with $\sum_{i=1}^N w_i = 1$. Let T be the associated block operator. Assume that T and all T_i's are bounded on bounded sets with a common modulus E. Let p_0 be a common fixed point of all T_i's and let $b \geq \|p_0\|$.*

Then for any x with $\|x\| \leq b$ and any $k = 1, \dots, N$:

$$w_k \geq w > 0 \land \|x - Tx\| < \xi\left(w\omega(\rho(\varepsilon, \widehat{b}), b), \widehat{b}\right) \to \|x - T_k x\| < \varepsilon$$

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

Also here, we refer to the appendix for a proof of this result.

The following lemma now provides a map that translates strong BSNE-moduli for the summands into strong BSNE-moduli for the block operator and in that sense is a quantitative version of Proposition 14 in [38].

Theorem 4.20. *Let ξ be a modulus of uniform continuity of D_f in its second argument. Let $\omega^{\nabla f}$ be a modulus of uniform continuity of ∇f on bounded sets and C be a modulus witnessing that ∇f is bounded on bounded sets. Let T_i, $i = 1, \dots, N$, be finitely many uniformly Bregman strongly nonexpansive operators with a common strong BSNE-modulus ω and derived modulus ω' and let $w_i \in [0,1]$, $i = 1, \dots, N$, be finitely many weights with $\sum_{i=1}^N w_i = 1$. Let T be the associated block operator. Assume that T and all T_i's are bounded on bounded sets with a common modulus E. Let $p_0 \in F(T)$ be a common fixed point of all T_i's and let $b \geq \|p_0\|$.*

Then T is uniformly Bregman strongly nonexpansive with a strong BSNE-modulus $\widehat{\omega}$ which can be defined by

$$\widehat{\omega}(\varepsilon, b) := \min\{w^2\omega(\varepsilon', b), \xi(w\omega(\rho(\min\{\omega(\varepsilon', b), \omega'(w\omega(\varepsilon', b), b)\}, \widehat{b}), b), \widehat{b})\}$$

where $\widehat{b} := \max\{b, E(b)\}$ and $\varepsilon' := \rho(\omega^{\nabla f}(\varepsilon/4\widehat{b}, \widehat{b}), \widehat{b})$ and $w := \min\{\varepsilon/8N\widehat{b}C(\widehat{b}), 1\}$.

If ω is only a (not necessarily strong) BSNE-modulus, then we can chose $\widehat{\omega}(\varepsilon, b) := w\omega(\varepsilon', b)$ as a BSNE-modulus for T.

Lastly, we also defer the proof of this result to the appendix.

5. Picard iterations

We now consider the first type of iteration of Bregman strongly nonexpansive mappings: as shown in [36], a Bregman strongly nonexpansive map $T : X \to X$ (in the context of some surrounding assumptions) is asymptotically regular, i.e. it holds that $\|x_n - Tx_n\| \to 0$ where $x_n := T^n x$ is the Picard iteration of T. In this section, we now derive quantitative rates for the above limit. In fact, we will actually first establish a corresponding quantitative result for a more general iteration involving a family of Bregman strongly nonexpansive operators of which the above Picard iteration will be a special case.

For this, we now fix the following moduli abstractly:[5]

(a) Let $\theta : (0,\infty)^2 \to (0,\infty)$ be a modulus of weak triangularity for D_f, i.e.

$$\forall \varepsilon, b > 0\ \forall x, y, z \in X\, (\|x\|, \|y\|, \|z\| \leq b \land D_f(x,y), D_f(z,y) < \theta(\varepsilon, b) \to D_f(x,z) < \varepsilon).$$

(b) Let $\xi : (0,\infty)^2 \to (0,\infty)$ be a modulus for $D_f(x,y)$ being uniformly continuous in y on bounded sets, i.e.

$$\forall \varepsilon, b > 0\ \forall x, y_1, y_2 \in X\, (\|x\|, \|y_1\|, \|y_2\| \leq b \land \|y_1 - y_2\| < \xi(\varepsilon, b) \to |D_f(x,y_1) - D_f(x,y_2)| < \varepsilon).$$

(c) Let $\rho : (0,\infty)^2 \to (0,\infty)$ be a modulus of consistency for f, i.e.

$$\forall \varepsilon, b > 0\ \forall x, y \in X\, (\|x\|, \|y\| \leq b \land D_f(x,y) < \rho(\varepsilon, b) \to \|x - y\| < \varepsilon).$$

We then obtain the following result on rates of metastability and rates of convergence for iterations of families of Bregman strongly nonexpansive mappings. In that vein, the result provides a quantitative version of the respective asymptotic regularity results contained in [37, 38]. Further, the theorem is an adaptation of a similar result (see Theorem 4.7 in [27]) on strongly quasi-nonexpansive mappings in the ordinary sense.

[5]Note the previous sections for how such moduli can be derived from respective moduli for the uniform continuity of ∇f, etc.

Theorem 5.1. *Let $(T_n)_{n\in\mathbb{N}}$ be a sequence of functions $T_n : X \to X$ which are Bregman strongly nonexpansive w.r.t. some $p \in \bigcap_{n\in\mathbb{N}} F(T_n)$ with a common BSNE-modulus $\omega(\varepsilon, b)$. Let $x_0 \in X$, $x_{n+1} := T_n x_n$ and $b \geq D_f(p, x_0), \|p\|, \|x_n\|$.*

Then[6]

$$\forall\varepsilon > 0\ \forall g : \mathbb{N} \to \mathbb{N}\ \exists n \leq \psi_{b,\omega}(\varepsilon, g)\ \forall k \in [n; n + g(n)]\, (D_f(x_{k+1}, x_k) < \varepsilon)$$

where

$$\psi_{b,\omega}(\varepsilon, g) := \widetilde{g}^{(\lceil \frac{b}{\omega(\varepsilon,b)} \rceil)}(0)$$

and $\widetilde{g}(n) := n + g(n) + 1$.

In particular, if o is a modulus of boundedness of D_f, then the above results holds true for $\psi_{\widehat{b},\omega}(\varepsilon, g)$ where $b \geq D_f(p, x_0), \|p\|$ and $\widehat{b} := \max\{o(b, b), b\}$.

Further, if $T_n = T$ for all $n \in \mathbb{N}$ and T, additionally, is also Bregman nonexpansive, then we even have

$$\forall\varepsilon > 0\ \forall k \geq \left\lceil \frac{\widehat{b}}{\omega(\varepsilon, \widehat{b})} \right\rceil (D_f(x_{k+1}, x_k) < \varepsilon)\,.$$

Proof. Since T_n in particular is Bregman quasi-nonexpansive w.r.t. p, we get that

$$0 \leq D_f(p, x_n) \leq D_f(p, x_0) \leq b.$$

Hence by Corollary 2.28 and Remark 2.29 from [26], we get that the function $\varphi(\varepsilon, g) := \widetilde{g}^{(\lceil \frac{b}{\varepsilon} \rceil)}(0)$ satisfies

$$\forall\varepsilon > 0\ \forall g : \mathbb{N} \to \mathbb{N}\ \exists n \leq \varphi(\varepsilon, g)\ \forall i, j \in [n; n + g(n) + 1]\, (|D_f(p, x_i) - D_f(p, x_j)| < \varepsilon)\,.$$

Take ε and g to be given and let, using the above, $n \leq \varphi(\omega(\varepsilon, b), g)$ be such that

$$|D_f(p, x_k) - D_f(p, T_k x_k)| < \omega(\varepsilon, b)$$

for all $k \in [n; n + g(n)]$. Using the fact that T_k is Bregman strongly nonexpansive with modulus ω, we get for any such k that $D_f(T_k x_k, x_k) < \varepsilon$ which proves the first claim. For $g(n) = 0$ for all n, we thus in particular have

$$\forall\varepsilon > 0\ \exists n \leq \widetilde{g}^{(\lceil \frac{b}{\omega(\varepsilon,b)} \rceil)}(0) = \left\lceil \frac{b}{\omega(\varepsilon, b)} \right\rceil (D_f(x_{n+1}, x_n) < \varepsilon)\,.$$

If now $T_k = T$ for all k and T is additionally Bregman nonexpansive, then

$$D_f(x_{k+1}, x_k) = D_f(T^{k+1}x, T^k x) \leq D_f(T^{n+1}x, T^n x) = D_f(x_{n+1}, x_n)$$

for all $k \geq n$ and so the second claim follows. □

From this, we get the following corollary to derive convergence of the norm distance:

Corollary 5.2. In addition to the assumptions in Theorem 5.1, let ρ be a modulus of consistency for f. Then

$$\forall\varepsilon > 0\ \forall g : \mathbb{N} \to \mathbb{N}\ \exists n \leq \psi_{b,\omega}(\rho(\varepsilon, d), g)\ \forall k \in [n; n + g(n)]\, (\|x_k - x_{k+1}\| < \varepsilon)\,.$$

If again $T_k = T$ for all k and T is additionally Bregman nonexpansive *or* nonexpansive (w.r.t. $\|\cdot\|$), then

$$\forall\varepsilon > 0\ \forall k \geq \left\lceil \frac{b}{\omega(\rho(\varepsilon, d), b)} \right\rceil (\|x_k - x_{k+1}\| < \varepsilon)\,.$$

The main application of this Picard process now follows if the iterated map is a composition. Together with Theorem 4.16, we can then obtain the following result giving that the Picard iteration $x_{n+1} = Tx_n$ of a composition $T = T_k \circ \cdots \circ T_1$ is asymptotically regular w.r.t. each T_j (which in particular provides a quantitative perspective on the method of cyclic Bregman projections [47]):

Theorem 5.3. *Let ξ be a modulus of uniform continuity on bounded subsets for D_f in its second argument. Let θ be a modulus of weak triangularity for D_f. Let ρ be a modulus of consistency for f and let P be a modulus for reverse consistency for f. Let o be a modulus of boundedness of D_f. Let ∇f and f be bounded on bounded sets with moduli C, D. Let $T_1, \ldots, T_k : X \to X$ be Bregman strongly nonexpansive w.r.t. some $p \in F(T_1) \cap \cdots \cap F(T_k)$ with a (not necessarily strong) BSNE-modulus ω. Let κ be a common modulus of uniform closedness of $F(T_1), \ldots, F(T_N)$. Define $T := T_k \circ \cdots \circ T_1$ as well as $x_n := T^n x_0$ for some $x_0 \in X$. Let $b \geq D_f(p, x_0), \|p\|$ and define $\widetilde{b} := \max\{o(b, b), b\}$ as well as $\widehat{b} := \max\{\widetilde{b}, E(\widetilde{b}), \ldots, E^{(k)}(\widetilde{b})\}$ for $E(b) := o(2D(b) + 2bC(b), b)$.*

Then

$$\forall\varepsilon > 0\ \forall g : \mathbb{N} \to \mathbb{N}\ \exists n \leq \Phi(\varepsilon, g)\ \forall i \in [n; n + g(n)]\ \forall j \in [1; k]\, (\|T_j x_i - x_i\| < \varepsilon)$$

[6] Here, and in the following, we write $[n; m] = [n, m] \cap \mathbb{N}$.

where Φ is defined by

$$\Phi(\varepsilon,g) := \widetilde{g}^{\left(\left\lceil \frac{\widehat{b}}{\widehat{\omega}(P(\varphi(\varepsilon,\widehat{b},k),\widehat{b}),\widehat{b})}\right\rceil\right)}(0)$$

where $\varphi(\varepsilon,b,k) := \chi_b(k-1,\varepsilon)$ with χ defined by

$$\begin{cases}\chi_b(0,\varepsilon) := \min\{\rho(\kappa(\varepsilon,b),b),\rho(\varepsilon,b)\},\\ \chi_b(n+1,\varepsilon) := \min\{\rho(\xi(\omega(\min\{\theta(\chi_b(n,\varepsilon),b),\rho(\kappa(\varepsilon,b),b)\},b),b),b),\chi_b(n,\varepsilon),\theta(\chi_b(n,\varepsilon),b)\}.\end{cases}$$

and where

$$\widehat{\omega}(\varepsilon,b) := \omega(\rho(P(\varepsilon,\widehat{b})/k,\widehat{b}),\widehat{b}).$$

Proof. The theorem is a straightforward combination of Corollary 5.2, Theorem 4.16, Theorem 4.15 and Lemma 4.6. □

Corollary 5.4. Let Ω_j, $j=1,\dots,k$, be nonempty, closed and convex sets with Bregman projections $P^f_{\Omega_j}$ and assume in addition to the assumptions in Theorem 5.3 that ∇f is uniformly continuous on bounded sets with a modulus $\omega^{\nabla f}$ and that it is uniformly strictly monotone with a modulus η, i.e.

$$\forall \varepsilon, b>0\ \forall x,y\in X\ (\|x\|,\|y\|\leq b \wedge \langle x-y,\nabla fx-\nabla fy\rangle < \eta(\varepsilon,b) \to \|x-y\|<\varepsilon).$$

Then for $T := P^f_{\Omega_k}\circ\dots\circ P^f_{\Omega_1}$ and $x_n := T^n x_0$ for some $x_0\in X$, we have

$$\forall\varepsilon>0\ \forall g:\mathbb{N}\to\mathbb{N}\ \exists n\leq\Phi(\varepsilon,g)\ \forall i\in[n;n+g(n)]\ \forall j\in[1;k]\ \left(\left\|P^f_{\Omega_j}x_i - x_i\right\|<\varepsilon\right)$$

with Φ defined by

$$\Phi(\varepsilon,g) := \widetilde{g}^{\left(\left\lceil \frac{\widehat{b}}{P(\varphi(\varepsilon,\widehat{b},k),\widehat{b})}\right\rceil\right)}(0)$$

with φ and χ defined as in Theorem 5.3, now using

$$\kappa(\varepsilon,b) := \min\{\varepsilon/3,\omega^{\nabla f}(\eta(\varepsilon/3,E(b))/2E(b),b)\}.$$

Proof. The corollary immediately follows from the above Theorem 5.3 where, for the particular case of Bregman projections, one additionally invokes Lemma 4.14 as well as Lemma 4.9 (by which we can use $\omega(\varepsilon,b)=\varepsilon$ as the common BSNE-modulus). □

The following proposition now provides an analogous result in the case that x_{n+1} is not exactly given by $T_n x_n$ but actually is allowed to differ from that point up to some summable error (compare this now to Theorem 4.9 from [27]). For that, we use the following result from [27]:

Lemma 5.5 (Lemma 4.8, [27]). *Let (a_n), (δ_n) be sequences of nonnegative reals with*

$$a_{n+1}\leq a_n+\delta_n,$$

where $\sum\delta_n<\infty$. Let $A,D\in\mathbb{N}$ with $A\geq a_0$ and $D\geq\sum\delta_n$. Define

$$\widetilde{\varphi}_{A,D}(\varepsilon,g) := \widetilde{g}^{(K)}(0),\ \textit{where } K := \left\lceil\frac{4(A+5D)}{\varepsilon}\right\rceil \textit{ and } \widetilde{g}(n) := n+g(n).$$

Then $\widetilde{\varphi}_{A,D}$ is a rate of metastability for (a_n).

Proposition 5.6. *Let $(T_n)_{n\in\mathbb{N}}$ be a sequence of functions $T_n: X\to X$ which are Bregman strongly nonexpansive w.r.t. some $p\in\bigcap_{n\in\mathbb{N}}F(T_n)$ with a common BSNE-modulus $\omega(\varepsilon,b)$. Let ξ be a modulus of uniform continuity of $D_f(p,u)$ in the argument u. Let $(x_n)\subseteq X$ be such that $\|x_{n+1}-T_nx_n\|<\xi(\delta_n,b)$ where $b\geq\|p\|,\|x_k\|,\|T_kx_k\|,D_f(p,x_0)$ for all k and where $(\delta_n)\subseteq[0,\infty)$ with $\sum\delta_n\leq D$. Let α be a rate of convergence for $\delta_n\to 0$, i.e.*

$$\forall\varepsilon>0\ \forall n\geq\alpha(\varepsilon)\ (\delta_n<\varepsilon).$$

Then

$$\forall\varepsilon>0\ \forall g:\mathbb{N}\to\mathbb{N}\ \exists n\leq\widetilde{\psi}_{b,\omega}(\varepsilon,g)\ \forall k\in[n;n+g(n)]\ (D_f(T_kx_k,x_k)<\varepsilon)$$

with

$$\widetilde{\psi}_{b,\omega}(\varepsilon,g) := \widetilde{\varphi}_{b,D}(\omega(\varepsilon,b)/2,g_{\alpha(\omega(\varepsilon,b)/2)}+1)) + \alpha(\omega(\varepsilon,b)/2)$$

where $g_l(n) := g(n+l)+l$ and

$$\widetilde{\varphi}_{b,D}(\varepsilon,g) := \widetilde{g}^{(K)}(0)\ \textit{with } K := \left\lceil\frac{4(b+5D)}{\varepsilon}\right\rceil \textit{ and } \widetilde{g}(n) := n+g(n).$$

In particular, if o is a modulus of boundedness of D_f, then the above results holds true for $\widetilde{\psi}_{\widehat{b},\omega}(\varepsilon,g)$ where $b\geq D_f(p,x_0),\|p\|$ and $\widehat{b} := \max\{o(b+D,b),b\}$.

Proof. Using the definition of ξ, we get for all $n \in \mathbb{N}$:

$$0 \leq D_f(p, x_{n+1}) \leq D_f(p, T_n x_n) + \delta_n \leq D_f(p, x_n) + \delta_n.$$

Hence by Lemma 5.5 applied to $a_n := D_f(p, x_n)$ (note that $b \geq a_0$), we get that

$$\forall \varepsilon > 0\ \forall g : \mathbb{N} \to \mathbb{N}\ \exists n \leq \widetilde{\varphi}_{b,D}(\varepsilon, g+1)\ \forall i, j \in [n; n+g(n)+1]\ (|D_f(p, x_i) - D_f(p, x_j)| < \varepsilon)$$

and so, given ε and g, the above in particular yields the existence of an $n \leq \widetilde{\varphi}_{b,D}(\varepsilon, g_{\alpha(\varepsilon)} + 1)$ such that

$$\forall k \in [n; n + g(n + \alpha(\varepsilon)) + \alpha(\varepsilon)]\ (|D_f(p, x_k) - D_f(p, x_{k+1})| < \varepsilon).$$

By considering $n + \alpha(\varepsilon)$ instead of n, this yields the existence of an $n \leq \widetilde{\varphi}_{b,D}(\varepsilon, g_{\alpha(\varepsilon)} + 1) + \alpha(\varepsilon)$ such that

$$|D_f(p, x_k) - D_f(p, T_k x_k)| < 2\varepsilon$$

for all $k \in [n; n + g(n)]$ since for $k \geq n \geq \alpha(\varepsilon)$, we have

$$|D_f(p, x_k) - D_f(p, T_k x_k)| \leq |D_f(p, x_k) - D_f(p, x_{k+1})| + |D_f(p, x_{k+1}) - D_f(p, T_k x_k)| < \varepsilon + \delta_k \leq 2\varepsilon.$$

From that, the above bound is immediate. □

6. A rate of metastability for a Halpern-type iteration of a family of maps

To obtain a strong convergence result, in [58], the authors defined a suitable Halpern-type iteration of a given Bregman strongly nonexpansive mapping. Concretely, the following result was established:

Theorem 6.1 ([58]). *Let X be a real reflexive Banach space and $f : X \to \mathbb{R}$ be a supercoercive Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets. Let T be a Bregman strongly nonexpansive mapping such that $F(T) = \widehat{F}(T) \neq \emptyset$. Given a $u \in X$, define a sequence x_n by $x_0 := x \in X$ and*

$$x_{n+1} := \nabla f^* \left(\alpha_n \nabla f u + (1 - \alpha_n) \nabla f(T x_n)\right)$$

where $(\alpha_n) \subseteq (0, 1)$ satisfies $\lim \alpha_n = 0$ and $\sum \alpha_n = \infty$. Then (x_n) converges strongly to $P^f_{F(T)}(u)$.

The aim of this section is to provide a quantitative analysis of this result as well as its extension to a family of mappings (T_n) as considered in [58], i.e. given u and x_0, we will consider the sequence

$$(*) \qquad x_{n+1} := \nabla f^*(\alpha_n \nabla f u + (1 - \alpha_n) \nabla f(T_n x_n)).$$

The proof of convergence for Theorem 6.1 as well as its extension to families of maps relies on a Lemma by Xu [63] as well as a subsequence construction due to Maingé [35], both of which have been treated quantitatively before in [34] as well as [30], respectively[7], and we present the quantitative versions of these crucial lemmas below.

Lemma 6.2 ([30], essentially [34]). *Let $b > 0$ and $(a_n) \subseteq [0, b]$ with*

$$a_{n+1} \leq (1 - \alpha_n) a_n + \alpha_n \beta_n + \gamma_n$$

for all n where $(\alpha_n) \subseteq (0, 1]$ with $\sum_{n=0}^{\infty} \alpha_n = +\infty$ (i.e. $\prod_{n=m}^{\infty} (1 - \alpha_n) = 0$ for all $m \in \mathbb{N}$) and $(\beta_n) \subseteq \mathbb{R}$ as well as $(\gamma_n) \subseteq [0, \infty)$. Let $S : (0, \infty) \times \mathbb{N} \to \mathbb{N}$ be nondecreasing in m such that

$$\forall m \in \mathbb{N}, \varepsilon > 0 \left(\prod_{k=m}^{S(\varepsilon, m)} (1 - \alpha_k) \leq \varepsilon \right).$$

For $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$, define

$$\widehat{g}(n) := g^M \left(n + S(\varepsilon/4b, n) + 1\right) + S(\varepsilon/4b, n).$$

Suppose that N satisfies

$$\exists m \leq N\ \forall i \in [m; m + \widehat{g}(m)]\ (\beta_i \leq \varepsilon/4).$$

Then for

$$\Phi(\varepsilon, S, N, b) := N + S(\varepsilon/4b, N) + 1,$$

we get that

$$\sum_{i=0}^{\Phi(\varepsilon, S, N, b) + g^M(\Phi(\varepsilon, S, N, b))} \gamma_i \leq \varepsilon/2 \to \exists n \leq \Phi(\varepsilon, S, N, b)\ \forall i \in [n; n + g(n)]\ (a_i \leq \varepsilon).$$

[7]The quantitative version of Xu's lemma presented in [34] works with slightly stronger assumptions than that presented in [30].

Lemma 6.3 ([30])**.** *Let $b > 0$ and $(a_n) \subseteq [0, b]$.*

(1) Let $\tau : \mathbb{N} \to \mathbb{N}$ be such that

$$(+) \qquad \forall n, k \in \mathbb{N} \, (k \leq n \wedge a_k < a_{k+1} \to k \leq \tau(n)) \, .$$

For $K \in \mathbb{N}$, $g : \mathbb{N} \to \mathbb{N}$, $\varepsilon > 0$ and $\widetilde{g}(n) := n + g(n)$, define

$$\Psi(\varepsilon, g, K, b) := \widetilde{g}^{(\lceil b/\varepsilon \rceil)}(K).$$

Then

$$\tau(\Psi(\varepsilon, g, K, b)) < K \to \exists n \leq \Psi(\varepsilon, g, K, b) \, (n \geq K \wedge \forall i, j \in [n; n + g(n)] \, (|a_i - a_j| \leq \varepsilon)) \, .$$

(2) Let $n_0 \in \mathbb{N}$ be such that $\exists n \leq n_0 \, (a_n < a_{n+1})$. Define

$$\tau(n) := \max\{k \leq \max\{n_0, n\} \mid a_k < a_{k+1}\}.$$

Then τ is well-defined and satisfies (+). Moreover,

(a) $\forall n \in \mathbb{N} \left(a_{\tau(n)} \leq a_{\tau(n)+1}\right)$,

(b) $\forall n \in \mathbb{N} \, (\tau(n) \leq \tau(n+1))$,

(c) $\forall n \geq n_0 \left(a_n \leq a_{\tau(n)+1}\right)$.

Before we move to quantitative results on the iteration considered above, we are first concerned with providing a quantitative account for Bregman projections onto fixed point sets of Bregman strongly nonexpansive maps. For this, the following lemma initially provides a quantitative version of the convexity of $F(T)$ as (essentially) shown in [50].

Lemma 6.4. *Let ρ be a modulus of consistency for f. Let T be uniformly Bregman strongly nonexpansive with strong BSNE-modulus ω and derived modulus ω'. Let T be bounded on bounded sets with a modulus E. Let $\varepsilon, b > 0$ be given and let x, y be such that $\|x\|, \|y\| \leq b$ and let $z := tx + (1-t)y$ for some $t \in [0, 1]$.*

If

$$\|Tx - x\|, \|Ty - y\| < \omega'(\rho(\varepsilon, \max\{b, E(b)\}), b),$$

then we have

$$\|Tz - z\| < \varepsilon.$$

Proof. Note that $\|z\| \leq t\|x\| + (1-t)\|y\| \leq b$. As in [50], we get

$$D_f(z, Tz) = f(z) + tD_f(x, Tz) + (1-t)D_f(y, Tz) - tf(x) - (1-t)f(y).$$

Using ω', we get

$$D_f(x, Tz) - D_f(x, z), D_f(y, Tz) - D_f(y, z) < \rho(\varepsilon, \max\{b, E(b)\})$$

and thus, using the above and the definition of D_f, we get

$$\begin{aligned} D_f(z, Tz) &< f(z) + tD_f(x, z) + (1-t)D_f(y, z) - tf(x) - (1-t)f(y) + \rho(\varepsilon, \max\{b, E(b)\}) \\ &= \rho(\varepsilon, \max\{b, E(b)\}). \end{aligned}$$

As $\|Tz\| \leq E(b)$, we get $\|z - Tz\| < \varepsilon$. □

Now, the following lemma provides a quantitative result on the existence of approximative projections onto fixed point sets of Bregman strongly nonexpansive maps. While the first part is concerned with the definition of said projections in terms of an infimum over Bregman distances, the second part is concerned with the characterization of Bregman projections in terms of a generalized type of variational inequality provided in [16] by which for a nonempty, closed and convex subset C and for a Gâteaux differentiable and totally convex function $f : X \to \mathbb{R}$, it holds that $z = P^f_C(x)$ if, and only if $z \in C$

$$\langle y - z, \nabla fx - \nabla fz \rangle \leq 0 \text{ for all } y \in C.$$

Note for both results that for a Bregman quasi-nonexpansive map T, the set of fixed points $F(T)$ is closed and convex (see e.g. [50][8]) and so $P^f_{F(T)}$ is defined for such a map whenever $F(T) \neq \emptyset$.

Lemma 6.5. *Let ρ be a modulus of consistency for f. Let T be uniformly Bregman strongly nonexpansive with strong BSNE-modulus ω and derived modulus ω'. Let T be bounded on bounded sets with a modulus E. Let $p_0 \in X$ be a fixed point of T with $D_f(p_0, u), \|p_0\| \leq b$.*

[8]Note here, as well as already in the context of Lemma 6.4, that the results from [50], while phrased for Bregman firmly nonexpansive maps, clearly already hold for Bregman quasi-nonexpansive maps.

(1) *For any* $\varepsilon > 0$ *and* $\psi : (0,\infty) \to (0,\infty)$*, let*

$$\varphi(\varepsilon,\psi) := \min\{\psi^{(r)}(1) \mid r \leq \lceil (b+1)/\varepsilon \rceil\}.$$

Then there exists a $p \in X$ *and a* $\delta \geq \varphi(\varepsilon,\psi)$ *with* $\|p\| \leq b$ *and* $\|Tp - p\| < \psi(\delta)$ *and*

$$\forall q \in X\,(\|q\| \leq b \wedge \|Tq - q\| < \delta \to D_f(p,u) < D_f(q,u) + \varepsilon)\,.$$

(2) *Let further* Δ *be a modulus witnessing that* $D_f(\cdot,u)$ *is uniformly Fréchet differentiable on bounded subsets with derivative* $x \mapsto \nabla fx - \nabla fu$*, i.e. for any* $b,\varepsilon > 0$ *and any* $x \in \overline{B}_b(0)$*,* $y \in X$*:*

$$0 < \|y\| < \Delta(\varepsilon,b) \to \frac{|D_f(x+y,u) - D_f(x,u) - \langle y, \nabla fx - \nabla fu\rangle|}{\|y\|} < \varepsilon.$$

For any $\varepsilon > 0$ *and* $\psi : (0,\infty) \to (0,\infty)$*, let*

$$\varphi'(\varepsilon,\psi) := \min\{\omega'(\rho(\psi'^{(r)}(1), \max\{b, E(b)\}), b) \mid r \leq \lceil (b+1)/\varepsilon' \rceil\}$$

with $\varepsilon' := \frac{\varepsilon}{2}\min\left\{\frac{\Delta(\varepsilon/4b,b)}{4b}, 1/2\right\}$ *and with*

$$\psi'(\delta) := \min\{\psi(\omega'(\rho(\delta,\max\{b,E(b)\}),b)), \omega'(\rho(\delta,\max\{b,E(b)\}),b)\}.$$

Then there exists a $p \in X$ *with* $\|p\| \leq b$ *and a* $\delta' \geq \varphi'(\varepsilon,\psi)$ *such that* $\|Tp - p\| < \psi(\delta')$

$$\forall q \in X\,(\|q\| \leq b \wedge \|Tq - q\| < \delta' \to \langle q - p, \nabla f(u) - \nabla f(p)\rangle < \varepsilon)\,.$$

We defer the proof to the appendix.

Remark 6.6. Such a modulus Δ witnessing that $D_f(\cdot,u)$ is uniformly Fréchet differentiable on bounded subsets with derivative $x \mapsto \nabla fx - \nabla fu$ can be computed from $\omega^{\nabla f}$: we have that

$$\|[D_f(\cdot,u)]'(x) - [D_f(\cdot,u)]'(y)\| = \|\nabla fx - \nabla fy\|$$

so that $\omega^{\nabla f}$ is a modulus for $[D_f(\cdot,u)]'$ being uniformly continuous on bounded subsets. Therefore, we can apply Lemma 3.2, (1) to derive that $\Delta(\varepsilon,b) = \min\{\omega^{\nabla f}(\varepsilon,b+1),1\}$ is a suitable such modulus.

For the rest of this section, we are now concerned with quantitative results on the extension of the iteration from Theorem 6.1 to families of mappings discussed before. For the following quantitative results, we again fix some moduli abstractly:

(a) Let (T_n) be a family of uniformly Bregman strongly nonexpansive maps with a common strong BSNE-modulus ω and a common derived modulus ω', i.e.

$$\forall \varepsilon,b > 0\ \forall x,p \in X\,(\|p\|,\|x\| \leq b \wedge \|T_np - p\| < \omega(\varepsilon,b) \wedge D_f(p,x) - D_f(p,T_nx) < \omega(\varepsilon,b) \to D_f(T_nx,x) < \varepsilon)$$

as well as

$$\forall \varepsilon,b > 0\ \forall x,p \in X\,(\|p\|,\|x\| \leq b \wedge \|T_np - p\| < \omega'(\varepsilon,b) \to D_f(p,T_nx) - D_f(p,x) < \varepsilon)$$

for any $n \in \mathbb{N}$.

(b) Let $(\alpha_n) \subseteq (0,1]$ converge to zero with a rate $\sigma : (0,\infty) \to \mathbb{N}$, i.e.

$$\forall n \geq \sigma(\varepsilon)\,(\alpha_n < \varepsilon)\,.$$

(c) Let f be sequentially consistent with a modulus of consistency ρ.

(d) Let $b \in \mathbb{N}^*$ be given and let x_n be defined by $(*)$ such that

$$b \geq \|x_n\|, \|T_nx_n\|, \|\nabla f(T_nx_n)\|, \|\nabla f(x_n)\|, \|u\|, \|\nabla f(u)\|, \|p_0\|, \|\nabla f(p_0)\|, D_f(p_0,x_n), D_f(p_0,T_nx_n), D_f(p_0,u)$$

for all $n \in \mathbb{N}$ where p_0 is some given element of $F(T)$.

(e) Let $\omega^{\nabla f^*} : (0,\infty) \to (0,\infty)$ be a modulus of uniform continuity for ∇f^* on b-bounded sets.

(f) Let $\omega^f : (0,\infty) \to (0,\infty)$ be a modulus of uniform continuity for f on b-bounded sets.

(g) Let $S : (0,\infty) \times \mathbb{N} \to \mathbb{N}$ be nondecreasing in m such that

$$\forall m \in \mathbb{N}, \varepsilon > 0 \left(\prod_{k=m}^{S(\varepsilon,m)} (1 - \alpha_k) \leq \varepsilon\right).$$

(h) For each n, let $\overline{\alpha}_n$ be such that $0 < \overline{\alpha}_n \leq \alpha_n$ and define $\widetilde{\alpha}_n = \min\{\overline{\alpha}_i \mid i \leq n\}$.

Lemma 6.7. *Let $\varepsilon > 0$ and let $p \in X$ be given and let x_n be defined by $(*)$. Define*

$$N := \sigma\left(\min\left\{\frac{\widetilde{\varepsilon}}{8b}, \frac{\widetilde{\varepsilon}}{16b^2}, \frac{1}{2b}\omega^{\nabla f^*}\left(\min\left\{\frac{\widetilde{\varepsilon}}{4b}, \omega^f\left(\frac{\widetilde{\varepsilon}}{4}\right)\right\}\right)\right\}\right)$$

where $\widetilde{\varepsilon} = \omega(\rho(\varepsilon, b), b)$.

For any $n \geq N$ and $p \in X$ with

$$\|p\|, D_f(p, x_n), D_f(p, T_n x_n), D_f(p, u) \leq b$$

for the above b and where $\|T_n p - p\| < \min\{\widetilde{\varepsilon}, \omega'(\widetilde{\varepsilon}/8, b)\}$ as well as

$$D_f(p, x_n) \leq D_f(p, x_{n+1}) \text{ or } |D_f(p, x_{n+1}) - D_f(p, x_n)| < \widetilde{\varepsilon}/4,$$

it holds that

$$\|x_n - T_n x_n\| < \varepsilon.$$

Proof. At first, given an $n \geq N$ with $D_f(p, x_n) \leq D_f(p, x_{n+1})$, we have

$$\begin{aligned}
0 &\leq D_f(p, x_{n+1}) - D_f(p, x_n) \\
&\leq \alpha_n D_f(p, u) + (1 - \alpha_n) D_f(p, T_n x_n) - D_f(p, x_n) \\
&= \alpha_n \left(D_f(p, u) - D_f(p, T_n x_n)\right) + D_f(p, T_n x_n) - D_f(p, x_n) \\
&\leq \alpha_n \left(D_f(p, u) - D_f(p, T_n x_n)\right) + \frac{\widetilde{\varepsilon}}{8} \\
&\leq b\alpha_n + \frac{\widetilde{\varepsilon}}{8} \\
&< b\frac{\widetilde{\varepsilon}}{8b} + \frac{\widetilde{\varepsilon}}{8} = \frac{\widetilde{\varepsilon}}{4}
\end{aligned}$$

where the fourth line follows using (a) and the last line follows using (b) and $n \geq N$. Therefore, the first disjunct of the premise implies the second disjunct. So assume $n \geq N$ and $|D_f(p, x_{n+1}) - D_f(p, x_n)| < \widetilde{\varepsilon}/4$. Now, we have

$$\begin{aligned}
\|\nabla f(x_{n+1}) - \nabla f(T_n x_n)\| &= \alpha_n \|\nabla f(u) - \nabla f(T_n x_n)\| \\
&\leq \alpha_n 2b && \text{(using (d))} \\
&< \min\left\{\frac{\widetilde{\varepsilon}}{8b}, \omega^{\nabla f^*}\left(\min\left\{\frac{\widetilde{\varepsilon}}{4b}, \omega^f\left(\frac{\widetilde{\varepsilon}}{4}\right)\right\}\right)\right\} && \text{(using (b) and } n \geq N\text{)}
\end{aligned}$$

and so by (e) and (d), we obtain

$$\|x_{n+1} - T_n x_n\| = \|\nabla f^*(\nabla f(x_{n+1})) - \nabla f^*(\nabla f(T_n x_n))\| \leq \min\left\{\frac{\widetilde{\varepsilon}}{4b}, \omega^f\left(\frac{\widetilde{\varepsilon}}{4}\right)\right\}.$$

By (f), we get

$$|f(x_{n+1}) - f(T_n x_n)| < \frac{\widetilde{\varepsilon}}{4}$$

and hence we obtain (reasoning similarly to [58])

$$\begin{aligned}
|D_f(p, T_n x_n) - D_f(p, x_n)| &= |f(p) - f(T_n x_n) - \langle p - T_n x_n, \nabla f(T_n x_n)\rangle - D_f(p, x_n)| \\
&= |f(p) - f(x_{n+1}) + f(x_{n+1}) - f(T_n x_n) - \langle p - x_{n+1}, \nabla f(x_{n+1})\rangle \\
&\qquad + \langle p - x_{n+1}, \nabla f(x_{n+1})\rangle - \langle p - T_n x_n, \nabla f(T_n x_n)\rangle - D_f(p, x_n)| \\
&= |D_f(p, x_{n+1}) + f(x_{n+1}) - f(T_n x_n) + \langle p - x_{n+1}, \nabla f(x_{n+1})\rangle \\
&\qquad - \langle p - T_n x_n, \nabla f(T_n x_n)\rangle - D_f(p, x_n)| \\
&= |D_f(p, x_{n+1}) - D_f(p, x_n) + f(x_{n+1}) - f(T_n x_n) \\
&\qquad + \langle p - x_{n+1}, \nabla f(x_{n+1}) - \nabla f(T_n x_n)\rangle - \langle x_{n+1} - T_n x_n, \nabla f(T_n x_n)\rangle| \\
&\leq |D_f(p, x_{n+1}) - D_f(p, x_n)| + |f(x_{n+1}) - f(T_n x_n)| \\
&\qquad + \|\nabla f(x_{n+1}) - \nabla f(T_n x_n)\| \|p - x_{n+1}\| + \|\nabla f(T_n x_n)\| \|x_{n+1} - T_n x_n\| \\
&< \frac{\widetilde{\varepsilon}}{4} + \frac{\widetilde{\varepsilon}}{4} + \frac{\widetilde{\varepsilon}}{8b} 2b + \frac{\widetilde{\varepsilon}}{4b} b = \widetilde{\varepsilon}.
\end{aligned}$$

Hence by (a) and (d), we obtain $D_f(T_n x_n, x_n) < \rho(\varepsilon, b)$ and so, by (c) and (d), we get $\|x_n - T_n x_n\| < \varepsilon$. ☐

Lemma 6.8. *For $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$, assume that we have a value φ and a $p \in X$ such that additionally*
$$\|p\|, \|\nabla f p\|, D_f(p, x_n), D_f(p, T_n x_n), D_f(p, u) \leq b$$
for the above b and
$$\|T_n p - p\| < \min\left\{\widetilde{\varphi}, \omega'\left(\frac{\widetilde{\varphi}}{8}, b\right), \omega'\left(\frac{\rho(\varepsilon, b)}{4(\Phi_\varphi(\varepsilon, g) + g^M(\Phi_\varphi(\varepsilon, g)) + 1)}, b\right), \omega'\left(\frac{\rho(\varepsilon, b)\widetilde{\alpha}_{(\Phi_\varphi(\varepsilon,g)+g'^M(\Phi_\varphi(\varepsilon,g)))}}{2}, b\right)\right\}$$
as well as
$$\forall y \in X\,(\|y\| \leq b \wedge \|T_n y - y\| < \varphi \to \langle y - p, \nabla f(u) - \nabla f(p)\rangle < \rho(\varepsilon, b)/8)$$
for any $n \leq \Phi_\varphi(\varepsilon, g) + g'^M(\Phi_\varphi(\varepsilon, g))$ where we define
$$N := \max\left\{\sigma\left(\frac{\omega^{\nabla f^*}(\rho(\varepsilon, b)/16b, b)}{2b}\right), \sigma\left(\min\left\{\frac{\widetilde{\varphi}}{8b}, \frac{\widetilde{\varphi}}{16b^2}, \frac{1}{2b}\omega^{\nabla f^*}\left(\min\left\{\frac{\widetilde{\varphi}}{4b}, \omega^f\left(\frac{\widetilde{\varphi}}{4}\right)\right\}\right)\right\}\right), 1\right\}$$
with $\widetilde{\varphi} := \omega(\rho(\varphi', b), b)$ and $\varphi' := \min\{\varphi, \rho(\varepsilon, b)/16b\}$ as well as
$$\Phi_\varphi(\varepsilon, g) := K_1 + S(\rho(\varepsilon, b)/8b, K_1) + 1$$
with
$$K_0 := \widetilde{g'}^{(\lceil 4(b+1)/\widetilde{\varphi}\rceil)}(N),\ K_1 := \widetilde{g'}^{(\lceil 8(b+1)/\widetilde{\varphi}\rceil)}(K_0),$$
and $\widetilde{g'}(n) := g'(n) + n$ where $g'(n) := \widehat{g}(n) + 2$ for
$$\widehat{g}(n) := g^M\left(n + S(\rho(\varepsilon, b)/8b, n) + 1\right) + S(\rho(\varepsilon, b)/8b, n).$$

Then it holds that
$$\exists n \leq \Phi_\varphi(\varepsilon, g)\ \forall i \in [n; n + g(n)]\,(\|p - x_i\| < \varepsilon)\,.$$

Proof (compare also [30, 57]). We write $a_i := D_f(p, x_i)$. To establish the claim, we divide between two cases:

Case 1: $\forall i \leq K_0\,(a_{i+1} \leq a_i)$.
Suppose first that
$$\forall i < \lceil 4(b+1)/\widetilde{\varphi}\rceil\left(a_{\widetilde{g'}^{(i+1)}(N)} \leq a_{\widetilde{g'}^{(i)}(N)} - \widetilde{\varphi}/4\right).$$
Then we would get
$$a_{\widetilde{g'}^{(0)}(N)} \geq a_{\widetilde{g'}^{(1)}(N)} + \widetilde{\varphi}/4 \geq \cdots \geq a_{\widetilde{g'}^{\lceil 4(b+1)/\widetilde{\varphi}\rceil}(N)} + \lceil 4(b+1)/\widetilde{\varphi}\rceil\widetilde{\varphi}/4 > b$$
which is a contradiction. Thus, we have
$$\exists i_0 < \lceil 4(b+1)/\widetilde{\varphi}\rceil\left(a_{\widetilde{g'}^{(i_0+1)}(N)} > a_{\widetilde{g'}^{(i_0)}(N)} - \widetilde{\varphi}/4\right).$$
and in particular for $n = \widetilde{g'}^{(i_0)}(N) \leq K_0 \leq K_1$, we have (since also $n + \widehat{g}(n) + 2 = \widetilde{g'}^{(i_0+1)}(N) \leq K_0$):
$$\forall i, j \in [n; n + \widehat{g}(n) + 2]\left(|a_i - a_j| \leq a_n - a_{n+\widehat{g}(n)+2} < \widetilde{\varphi}/4\right).$$

Case 2: $\exists i \leq K_0\,(a_{i+1} > a_i)$.
Then, we define τ as in Lemma 6.3.(2), i.e.
$$\tau(n) := \max\{k \leq \max\{K_0, n\} \mid a_k < a_{k+1}\}.$$
In particular, we have

1. $\forall n\left(a_{\tau(n)} \leq a_{\tau(n)+1}, \tau(n) \leq \tau(n+1)\right)$,
2. $\forall n \geq K_0\left(a_n \leq a_{\tau(n)+1}\right)$.

Case 2.1: $\forall m \in [K_1; K_1 + g'(K_1)]\,(\tau(m) \geq K_0)$.
Using the properties of τ, we in particular have
$$D_f(p, x_{\tau(m)}) = a_{\tau(m)} \leq a_{\tau(m)+1} = D_f(p, x_{\tau(m)+1})$$
for all m. For $m \in [K_1; K_1 + g'(K_1)]$ we have $\tau(m) \geq K_0 \geq N$ and thus we get
$$\left\|x_{\tau(m)} - T_{\tau(m)}x_{\tau(m)}\right\| < \varphi' = \min\{\varphi, \rho(\varepsilon, b)/16b\}$$
using Lemma 6.7. Thus, using the assumption on φ, we get (as $\tau(m) \leq m \leq K_1 + g'(K_1) \leq \Phi_\varphi(\varepsilon, g) + g'^M(\Phi_\varphi(\varepsilon, g))$):
$$\langle x_{\tau(m)} - p, \nabla f(u) - \nabla f(p)\rangle < \rho(\varepsilon, b)/8 < \rho(\varepsilon, b)/4.$$

As in the proof Lemma 6.7, we get

$$\left\|\nabla f(x_{\tau(m)+1}) - \nabla f(T_{\tau(m)}x_{\tau(m)})\right\| \leq \alpha_{\tau(m)}2b.$$

As $\tau(m) \geq K_0 \geq N$, we in particular have $\left\|x_{\tau(m)+1} - T_{\tau(m)}x_{\tau(m)}\right\| < \rho(\varepsilon, b)/16b$. Further, from above we also have $\left\|x_{\tau(m)} - T_{\tau(m)}x_{\tau(m)}\right\| < \rho(\varepsilon, b)/16b$ such that this combined yields $\left\|x_{\tau(m)+1} - x_{\tau(m)}\right\| < \rho(\varepsilon, b)/8b$. Therefore:

$$\begin{aligned}\langle x_{\tau(m)+1} - p, \nabla f(u) - \nabla f(p)\rangle &= \langle x_{\tau(m)+1} - x_{\tau(m)}, \nabla f(u) - \nabla f(p)\rangle + \langle x_{\tau(m)} - p, \nabla f(u) - \nabla f(p)\rangle \\ &< \left\|x_{\tau(m)+1} - x_{\tau(m)}\right\| 2b + \rho(\varepsilon, b)/4 \\ &< \rho(\varepsilon, b)/2.\end{aligned}$$

As in [58] (p. 495), we can derive

$$D_f(p, x_{i+1}) \leq (1 - \alpha_i)D_f(p, T_i x_i) + \alpha_i\langle x_{i+1} - p, \nabla f(u) - \nabla f(p)\rangle$$

for any i. Therefore, we can derive

$$\begin{aligned}&D_f(p, x_{\tau(m)+1}) \\ &\quad\leq (1 - \alpha_{\tau(m)})D_f(p, T_{\tau(m)}x_{\tau(m)}) + \alpha_{\tau(m)}\langle x_{\tau(m)+1} - p, \nabla f(u) - \nabla f(p)\rangle \\ &\quad\leq (1 - \alpha_{\tau(m)})D_f(p, x_{\tau(m)}) + \alpha_{\tau(m)}\langle x_{\tau(m)+1} - p, \nabla f(u) - \nabla f(p)\rangle + \frac{\rho(\varepsilon, b)\widetilde{\alpha}_{(\Phi_\varphi(\varepsilon,g)+g'^M(\Phi_\varphi(\varepsilon,g)))}}{2} \\ &\quad\leq (1 - \alpha_{\tau(m)})D_f(p, x_{\tau(m)+1}) + \alpha_{\tau(m)}\langle x_{\tau(m)+1} - p, \nabla f(u) - \nabla f(p)\rangle + \frac{\rho(\varepsilon, b)\widetilde{\alpha}_{(\Phi_\varphi(\varepsilon,g)+g'^M(\Phi_\varphi(\varepsilon,g)))}}{2}\end{aligned}$$

for all $m \in [K_1; K_1 + g'(K_1)]$ (since $\tau(m) \leq m \leq K_1 + g'(K_1)$). From this, we get

$$D_f(p, x_{\tau(m)+1}) \leq \langle x_{\tau(m)+1} - p, \nabla f(u) - \nabla f(p)\rangle + \frac{\rho(\varepsilon, b)\widetilde{\alpha}_{(\Phi_\varphi(\varepsilon,g)+g'^M(\Phi_\varphi(\varepsilon,g)))}}{2\alpha_{\tau(m)}}$$

for all such m. Again as $\tau(m) \leq m \leq K_1 + g'(K_1)$, we get

$$\widetilde{\alpha}_{(\Phi_\varphi(\varepsilon,g)+g'^M(\Phi_\varphi(\varepsilon,g)))} \leq \overline{\alpha}_{\tau(m)} \leq \alpha_{\tau(m)}$$

for all such m and thus we have

$$D_f(p, x_{\tau(m)+1}) \leq \langle x_{\tau(m)+1} - p, \nabla f(u) - \nabla f(p)\rangle + \frac{\rho(\varepsilon, b)}{2} < \rho(\varepsilon, b).$$

Lastly, we thus have

$$D_f(p, x_m) \leq D_f(p, x_{\tau(m)+1}) < \rho(\varepsilon, b)$$

for all such m by using the properties of τ, which implies $\|p - x_m\| < \varepsilon$ for all such m by (c). So in this case, we have now already established the theorem.

Case 2.2: $\exists m \in [K_1; K_1 + g'(K_1)]\,(\tau(m) < K_0)$.
As $m \geq K_1$, we have

$$\tau\left(\widetilde{g'}^{(\lceil 8(b+1)/\widetilde{\varphi}\rceil)}(K_0)\right) = \tau(K_1) \leq \tau(m) < K_0$$

and thus using Lemma 6.3.(1), we get

$$\exists n \leq \widetilde{g'}^{(\lceil 8(b+1)/\widetilde{\varphi}\rceil)}(K_0)\left(n \geq K_0 \geq N \wedge \forall i, j \in [n; n + \widehat{g}(n) + 2]\,(|a_i - a_j| \leq \widetilde{\varphi}/8 < \widetilde{\varphi}/4)\right).$$

We now establish the theorem in this case as well as in the context of Case 1 as follows: In both cases, we have

$$\exists n \in [N; K_1]\; \forall i, j \in [n; n + \widehat{g}(n) + 2]\,(|a_i - a_j| < \widetilde{\varphi}/4).$$

Therefore, we in particular have for such an n that

$$\forall i \in [n; n + \widehat{g}(n) + 1]\,(|D_f(p, x_{i+1}) - D_f(p, x_i)| < \widetilde{\varphi}/4)$$

in both cases. Using Lemma 6.7, we get

$$\forall i \in [n; n + \widehat{g}(n) + 1]\,(\|x_i - T_i x_i\| < \varphi' \leq \varphi).$$

Using the assumption on φ, we in particular get

$$\forall i \in [n; n + \widehat{g}(n) + 1]\,(\langle x_i - p, \nabla f(u) - \nabla f(p)\rangle < \rho(\varepsilon, b)/8)$$

and thus

$$(*)\qquad \forall i \in [n; n+\widehat{g}(n)]\,(\langle x_{i+1}-p, \nabla f(u)-\nabla f(p)\rangle < \rho(\varepsilon,b)/8)\,.$$

As before (see again p. 495 in [58]), we can derive

$$D_f(p,x_{i+1}) \leq (1-\alpha_i)D_f(p,T_ix_i) + \alpha_i\langle x_{i+1}-p, \nabla f(u)-\nabla f(p)\rangle$$

for any i which implies

$$D_f(p,x_{i+1}) \leq (1-\alpha_i)D_f(p,x_i) + \alpha_i\langle x_{i+1}-p, \nabla f(u)-\nabla f(p)\rangle + \max\{0, D_f(p,T_ix_i)-D_f(p,x_i)\}$$

for all i. As $(*)$ holds for $n \leq K_1$ and since we have

$$\begin{aligned}\sum_{i=0}^{\Phi_\varphi(\varepsilon,g)+g^M(\Phi_\varphi(\varepsilon,g))} \max\{0, D_f(p,T_ix_i)-D_f(p,x_i)\} &\leq \sum_{i=0}^{\Phi_\varphi(\varepsilon,g)+g^M(\Phi_\varphi(\varepsilon,g))} \frac{\rho(\varepsilon,b)}{4(\Phi_\varphi(\varepsilon,g)+g^M(\Phi_\varphi(\varepsilon,g))+1)}\\ &= \rho(\varepsilon,b)/4\end{aligned}$$

by the assumptions on p and ω', Lemma 6.2 with $\beta_n = \langle x_{n+1}-p, \nabla f(u)-\nabla f(p)\rangle$ and $\gamma_n = \max\{0, D_f(p,T_nx_n)-D_f(p,x_n)\}$ can be used to get

$$\exists n \leq K_1 + S(\rho(\varepsilon,b)/8b, K_1) + 1 = \Phi_\varphi(\varepsilon,g)\forall i \in [n; n+g(n)]\left(D_f(p,x_i) = a_i \leq \frac{\rho(\varepsilon,b)}{2} < \rho(\varepsilon,b)\right)$$

which implies $\|p-x_i\| < \varepsilon$ for all such i by (c). □

Together with Lemma 6.5, we thus obtain the following combined result for sequences of Bregman strongly nonexpansive maps. One crucial property that features therein is a uniformized version of the NST condition as e.g. considered in [1] for sequences of strongly nonexpansive maps in the ordinary sense: given a sequence (T_n) of strongly nonexpansive maps and an additional such map T, these are said to satisfy the NST condition if any fixed point of T is a common fixed point for all T_n and if $\|x_n - T_nx_n\| \to 0$ implies $\|x_n - Tx_n\| \to 0$ for any bounded sequence (x_n).

Concretely, the following uniform quantitative variant of this condition will feature crucially in the following combined result: we assume a modulus $\mu : (0,\infty)^2 \times \mathbb{N} \to (0,\infty)$ such that

$$(\dagger)_1\qquad \forall \varepsilon, b > 0\ \forall K \in \mathbb{N}\ \forall p \in X\,(\|p\| \leq b \wedge \|p-Tp\| < \mu(\varepsilon,b,K) \to \forall n \leq K\,(\|p-T_np\| < \varepsilon))$$

as well as a modulus $\nu : (0,\infty)^2 \to (0,\infty)$ such that

$$(\dagger)_2\qquad \forall \varepsilon, b > 0\ \forall n \in \mathbb{N}\ \forall p \in X\,(\|p\| \leq b \wedge \|p-T_np\| < \nu(\varepsilon,b) \to \|p-Tp\| < \varepsilon)\,.$$

If such moduli exist, we say that (T_n) and T satisfy the uniform NST condition.

As we will discuss later, such moduli can in particular be explicitly computed for the resolvents relative to f, thereby allowing applications to a Halpern-type proximal point algorithm.

Theorem 6.9. *Let $(\alpha_n) \subseteq (0,1]$ converge to zero with a rate σ and, for any n, let $\overline{\alpha}_n$ be such that $0 < \overline{\alpha}_n \leq \alpha_n$ and define $\widetilde{\alpha}_n := \min\{\overline{\alpha}_i \mid i \leq n\}$. Let f be sequentially consistent with a modulus of consistency ρ. Let $S : (0,\infty) \times \mathbb{N} \to \mathbb{N}$ be nondecreasing in the right argument such that*

$$\forall m \in \mathbb{N}, \varepsilon > 0\left(\prod_{k=m}^{S(\varepsilon,m)}(1-\alpha_k) \leq \varepsilon\right).$$

Let (T_n) be a sequence of uniformly Bregman strongly nonexpansive maps and T be another uniformly Bregman strongly nonexpansive map with a common strong BSNE-modulus ω and a common derived modulus ω'. Let each T_n and T be bounded on bounded sets with a common modulus E and let $p_0 \in X$ be a common fixed point of all T_n and T. Let o be a modulus of boundedness for D_f. Let ∇f and f be bounded on bounded sets with moduli C, D, respectively. Let $b \in \mathbb{N}^*$ with

$$b \geq \|p_0\|, D_f(p_0,u), \|u\|, D_f(p_0,x_0)$$

and define

$$\begin{aligned}\widehat{b} := \max\{&b, C(b), o(b,b), E(o(b,b)), C(E(o(b,b))), C(o(b,b)),\\ &D(b) + D(E(o(b,b))) + (b+E(o(b,b)))C(E(o(b,b))),\\ &D(b) + D(o(b,b)) + (b+o(b,b))C(o(b,b)), 2D(b) + 2bC(b)\}.\end{aligned}$$

Let $\omega^{\nabla f^*}$, ω^f be moduli of uniform continuity of ∇f^*, f, respectively. Let further Δ be a modulus witnessing that $D_f(\cdot,u)$ is uniformly Fréchet differentiable on bounded subsets with derivative $x\mapsto\nabla fx-\nabla fu$ as in Lemma 6.5. Assume that we have a modulus $\mu:(0,\infty)^2\times\mathbb{N}\to(0,\infty)$ such that

$$\forall\varepsilon,b>0,K\in\mathbb{N},p\in X\left(\|p\|\leq b\wedge\|p-Tp\|<\mu(\varepsilon,b,K)\to\forall n\leq K\left(\|p-T_np\|<\varepsilon\right)\right)$$

as well as a modulus $\nu:(0,\infty)^2\to(0,\infty)$ such that

$$\forall\varepsilon,b>0,n\in\mathbb{N},p\in X\left(\|p\|\leq b\wedge\|p-T_np\|<\nu(\varepsilon,b)\to\|p-Tp\|<\varepsilon\right).$$

For any $\varepsilon>0$ and $g:\mathbb{N}\to\mathbb{N}$ as well as $\varphi>0$, we define

$$\psi(\varphi):=\min\left\{\widetilde{\varphi},\omega'\left(\frac{\widetilde{\varphi}}{8},\widehat{b}\right),\omega'\left(\frac{\rho(\varepsilon,\widehat{b})}{4(\Phi_\varphi(\varepsilon,g)+g^M(\Phi_\varphi(\varepsilon,g))+1)},\widehat{b}\right),\omega'\left(\frac{\rho(\varepsilon,\widehat{b})\widetilde{\alpha}_{(\Phi_\varphi(\varepsilon,g)+g'^M(\Phi_\varphi(\varepsilon,g)))}}{2},\widehat{b}\right)\right\},$$

$$\widehat{\psi}(\varphi):=\mu(\psi(\nu(\varphi,\widehat{b})),\widehat{b},\Phi_{\nu(\varphi,\widehat{b})}(\varepsilon,g)+g'^M(\Phi_{\nu(\varphi,\widehat{b})}(\varepsilon,g))),$$

$$\psi'(\varphi):=\min\{\widehat{\psi}(\omega'(\rho(\varphi,\max\{\widehat{b},E(\widehat{b})\}),\widehat{b})),\omega'(\rho(\varphi,\max\{\widehat{b},E(\widehat{b})\}),\widehat{b})\},$$

with

$$N:=\max\left\{\sigma\left(\frac{\omega^{\nabla f^*}(\rho(\varepsilon,\widehat{b})/16\widehat{b},\widehat{b})}{2\widehat{b}}\right),\sigma\left(\min\left\{\frac{\widetilde{\varphi}}{8\widehat{b}},\frac{\widetilde{\varphi}}{16\widehat{b}^2},\frac{1}{2\widehat{b}}\omega^{\nabla f^*}\left(\min\left\{\frac{\widetilde{\varphi}}{4\widehat{b}},\omega^f\left(\frac{\widetilde{\varphi}}{4},\widehat{b}\right)\right\},\widehat{b}\right)\right\}\right),1\right\}$$

where $\widetilde{\varphi}:=\omega(\rho(\varphi',\widehat{b}),\widehat{b})$ and $\varphi':=\min\{\varphi,\rho(\varepsilon,\widehat{b})/16\widehat{b}\}$ as well as

$$\Phi_\varphi(\varepsilon,g):=K_1+S(\rho(\varepsilon,\widehat{b})/8\widehat{b},K_1)+1$$

with

$$K_0:=\widetilde{g'}^{(\lceil 4(\widehat{b}+1)/\widetilde{\varphi}\rceil)}(N),\ K_1:=\widetilde{g'}^{(\lceil 8(\widehat{b}+1)/\widetilde{\varphi}\rceil)}(K_0),$$

and $\widetilde{g'}(n):=g'(n)+n$ where $g'(n):=\widehat{g}(n)+2$ for

$$\widehat{g}(n):=g^M\left(n+S(\rho(\varepsilon,\widehat{b})/8\widehat{b},n)+1\right)+S(\rho(\varepsilon,\widehat{b})/8\widehat{b},n).$$

Then it holds that

$$\forall\varepsilon>0\ \forall g:\mathbb{N}\to\mathbb{N}\ \exists n\leq\widehat{\Phi}(\varepsilon,g)\ \forall i,j\in[n;n+g(n)]\left(\|x_i-x_j\|<\varepsilon\right)$$

where

$$\widehat{\Phi}(\varepsilon,g):=\max\left\{\Phi_{\nu(\omega'(\rho(\psi'^{(r)}(1),\max\{\widehat{b},E(\widehat{b})\}),\widehat{b}),\widehat{b})}(\varepsilon/2,g)\mid r\leq\left\lceil(\widehat{b}+1)/\varepsilon'\right\rceil\right\}$$

with

$$\varepsilon':=\frac{\rho(\varepsilon/2,\widehat{b})}{16}\min\left\{\frac{\Delta(\rho(\varepsilon/2,\widehat{b})/32\widehat{b},\widehat{b})}{4\widehat{b}},1/2\right\}.$$

Proof. Let ε and g be given. Using (the proof of) Lemma 6.5, (2), we get that for the above $\widehat{\psi}$, there exists a $p\in X$ with $\|p\|\leq b\leq\widehat{b}$ and an $r\leq\left\lceil(\widehat{b}+1)/\varepsilon'\right\rceil$ such that for $\delta=\omega'(\rho(\psi'^{(r)}(1),\max\{\widehat{b},E(\widehat{b})\}),\widehat{b})$ we have $\|Tp-p\|<\widehat{\psi}(\delta)$

$$\forall q\in X\left(\|q\|\leq\widehat{b}\wedge\|Tq-q\|<\delta\to\langle q-p,\nabla f(u)-\nabla f(p)\rangle<\rho(\varepsilon/2,\widehat{b})/8\right).$$

Then, as

$$\|Tp-p\|<\widehat{\psi}(\delta)=\mu(\psi(\nu(\delta,\widehat{b})),\widehat{b},\Phi_{\nu(\delta,\widehat{b})}(\varepsilon/2,g)+g'^M(\Phi_{\nu(\delta,\widehat{b})}(\varepsilon/2,g))),$$

we get

$$\|T_np-p\|<\psi(\nu(\delta,\widehat{b}))$$

for all $n\leq\Phi_{\nu(\delta,\widehat{b})}(\varepsilon/2,g)+g'^M(\Phi_{\nu(\delta,\widehat{b})}(\varepsilon/2,g))$. Further, if $\|q-T_nq\|<\nu(\delta,\widehat{b})$, we have $\|Tq-q\|<\delta$ and thus also

$$\forall q\in X\left(\|q\|\leq\widehat{b}\wedge\|T_nq-q\|<\nu(\delta,\widehat{b})\to\langle q-p,\nabla f(u)-\nabla f(p)\rangle<\rho(\varepsilon/2,\widehat{b})/8\right)$$

for any n. Lemma 6.8 then yields that

$$\exists n\leq\Phi_{\nu(\delta,\widehat{b})}(\varepsilon/2,g)\ \forall i\in[n;n+g(n)]\left(\|p-x_i\|<\varepsilon/2\right)$$

as $\widehat{b}$ bounds all the objects involved. After using the triangle inequality, we get the desired claim. □

In particular, since having a rate of metastability is equivalent to being convergent, the above quantitative result implies the following (non-quantitative) convergence result. For that, we say that (T_n) and T are commonly uniformly Bregman strongly nonexpansive if all T_n and T are uniformly Bregman strongly nonexpansive with a common strong BSNE-modulus and we say that they are commonly bounded on bounded sets if there exists a common modulus witnessing that all T_n and T are bounded on bounded sets.

Theorem 6.10. *Let X be a real reflexive Banach space and $f : X \to \mathbb{R}$ be a supercoercive Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets. Let (T_n) be a sequence of selfmaps and T be a selfmap such that they are commonly uniformly Bregman strongly nonexpansive and commonly bounded on bounded sets. Assume that (T_n) and T satisfy the uniform NST condition and that they posses a common fixed point. Given a $u \in X$, define a sequence x_n by $x_0 := x \in X$ and*

$$x_{n+1} := \nabla f^* \left(\alpha_n \nabla f u + (1 - \alpha_n) \nabla f T_n x_n\right)$$

where $(\alpha_n) \subseteq (0, 1]$ satisfies $\lim \alpha_n = 0$ and $\sum \alpha_n = \infty$. Then (x_n) is Cauchy.

Further, if we have $\overline{F}(T) \subseteq F(T)$ where $\overline{F}(T)$ is the set of all strong asymptotic fixed points (i.e. of all p such that there is a sequence (p_n) with $p_n \to p$ and $\|p_n - Tp_n\| \to 0$ for $n \to \infty$), then (x_n) converges strongly to $P^f_{F(T)}(u)$.

Proof. First, note that under the assumptions presented above, all moduli featured in Theorem 6.9 exist and we shortly discuss this for the assumptions not explicitly covered already: A modulus of consistency ρ exists for f as f is totally convex on bounded sets using Lemmas 3.7 and 3.8. As f is uniformly Fréchet differentiable and bounded on bounded sets, ∇f is uniformly continuous on bounded sets by Proposition 2.2 and thus a corresponding modulus $\omega^{\nabla f}$ exists which allows us to construct a corresponding modulus ω^f for the uniform continuity of f as well as moduli for ∇f, f being bounded on bounded sets using Lemma 3.2. Also, as discussed in Remark 6.6, $\omega^{\nabla f}$ can be used to construct the modulus Δ featured in Theorem 6.9. Now, as discussed in Remark 3.13, f being totally convex on bounded sets implies f^* being uniformly Fréchet differentiable and thus ∇f^* being uniformly continuous as f is supercoercive (again using Proposition 2.2). Thus a corresponding modulus $\omega^{\nabla f^*}$ exists. Lastly, a modulus of boundedness for D_f exists as well and can be constructed as discussed in Remark 3.14.

So Theorem 6.9 applies and we therefore get

$$\forall \varepsilon > 0\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists n \in \mathbb{N}\, \forall i, j \in [n; n + g(n)]\, (\|x_i - x_j\| < \varepsilon).$$

Thus (x_n) is Cauchy as if not, there exists an $\varepsilon > 0$ such that for any n, there exists an m such that $\|x_n - x_{n+m}\| \geq \varepsilon$. Pick $g(n) = m$ for such an m. Then this ε and g refute the above property. Now, as (x_n) is Cauchy, it converges to a limit x.

To see that this limit is indeed the projection $P^f_{F(T)}(u)$, let $\Omega(\varepsilon, b)$ be a modulus of uniform continuity on bounded sets for the function $p \mapsto \langle y - p, \nabla f u - \nabla f p\rangle$ uniform in $\|u\|, \|y\| \leq b$.[9] Now, let $\varepsilon > 0$ be given and let K be so large that

$$\forall m \geq K \left(\|x_m - x\| < \frac{1}{2}\Omega\left(\frac{\varepsilon}{2}, \widehat{b}\right)\right).$$

Now, for $\varepsilon' := 1/2\Omega(\varepsilon/2, \widehat{b})$, we can use Lemma 6.5 to choose a $p \in X$ and a δ with $\|p\| \leq \widehat{b}$ and $\|p - Tp\| < \widehat{\psi}(\delta)$ as well as

$$\forall q \in X \left(\|q\| \leq \widehat{b} \wedge \|q - Tq\| < \delta \to \langle q - p, \nabla f u - \nabla f p\rangle < \rho(\varepsilon', \widehat{b})/8\right).$$

Then, using this p and reasoning as in the proof of Theorem 6.9, we can apply Lemma 6.8 to $g(n) := K$ and ε' which yields an $n \geq K$ such that $\|p - x_n\| < \varepsilon' = 1/2\Omega(\varepsilon/2, \widehat{b})$. That $n \geq K$ holds in particular yields $\|p - x\| < \Omega(\varepsilon/2, \widehat{b})$. Let w.l.o.g. $\rho(\varepsilon, b) \leq \varepsilon$ and $\Omega(\varepsilon, b) \leq \varepsilon$. Then we in particular have

$$\langle q - p, \nabla f u - \nabla f p\rangle < \varepsilon/2$$

for any q with $\|q\| \leq \widehat{b}$ and $\|q - Tq\| < \delta$. Thus

$$\langle q - x, \nabla f u - \nabla f x\rangle < \varepsilon$$

for all such q.

If now $q = Tq$, then we get $\langle q - x, \nabla f u - \nabla f x\rangle < \varepsilon$ for all $\varepsilon > 0$, i.e.

$$\forall q \in F(T)\, (\langle q - x, \nabla f u - \nabla f x\rangle \leq 0).$$

[9] It can be easily seen that such a modulus Ω can actually be constructed from $\omega^{\nabla f}$.

Further, if we assume that $\overline{F}(T) \subseteq F(T)$, then x is also fixed point of T. For this, we reason as in the proof of Lemma 6.7 and derive that

$$\|\nabla f x_{n+1} - \nabla f T_n x_n\| = \alpha_n \|\nabla f u - \nabla f T x_n\| \to 0$$

as $\alpha_n \to 0$ and as $\nabla f T x_n$ is bounded since x_n is bounded and since T and ∇f are bounded on bounded sets. Thus $\|x_{n+1} - T_n x_n\| \to 0$ and therefore also $\|x_n - T_n x_n\| \to 0$ as $x_n \to x$. As (T_n) and T satisfy the uniform NST condition, we get $\|x_n - T x_n\| \to 0$. As $\|x_n - x\| \to 0$, this yields $x \in \overline{F}(T) \subseteq F(T)$. Combined, this yields that $x = P^f_{F(T)}(u)$ (recall the discussion before Lemma 6.5). □

Remark 6.11. The above result in particular contains the previous Theorem 6.1 for uniformly Bregman strongly nonexpansive maps T by picking $T_n = T$. Naturally T is bounded on bounded sets as $F(T) \neq \emptyset$ and as T is Bregman quasi-nonexpansive. However, note that in the context of uniformly Bregman strongly nonexpansive maps T, the assumption that $\widehat{F}(T) \subseteq F(T)$ was properly weakened through the analysis to $\overline{F}(T) \subseteq F(T)$.

Using this theorem, we will in particular be able to derive the strong convergence of the Halpern-type proximal point algorithm together with other interesting instantiations that will be discussed in the following section.

7. Special cases and instantiations

We are now concerned with the range of the above results. For that, this section discusses how the above (quantitative) results can be instantiated in various ways so that they apply to many other well-known methods in the context of Bregman distance. Concretely, we obtain quantitative strong convergence results for Halpern-type variants of the method of cyclic Bregman projections, of the proximal point algorithm and of a method for finding common zeros of maximally monotone operators as discussed by Naraghirad [40] as well as for a Halpern-Mann-type iteration of Bregman strongly nonexpansive maps [67].

In particular, we show how the Halpern-Mann type iteration presented in [67] can be recognized as an instantiation of the Halpern-iteration considered before for a family of uniformly Bregman strongly nonexpansive maps. Further, inspired by the recent considerations [18] on the relationship between modified Halpern methods in the sense of [20, 24] and Tikhonov-Mann type methods as developed by [8, 19, 64], we use this instantiation to even provide a strong convergence result for a new Tikhonov-Mann type iteration of Bregman strongly nonexpansive maps which provides a suitable lift of such iterations to this Bregman context.

Beyond the examples discussed here, we want to mention that the previous results can also be used to obtain similar quantitative strong convergence results for Halpern-type variants of special cases of a method solving operator equations due to Butnariu and Resmerita [16] as well as of special cases of the forward-backward Bregman splitting method discussed by Búi and Combettes [13] (see also Van Nguyen [41]), at least in the context of specific operators induced by uniformly Bregman strongly nonexpansive mappings, but we do not explore this here further at any depth.

In any way, all these results in particular show that the additional requirement in the previous theorems that the maps are even uniformly Bregman strongly nonexpansive is practically of lesser significance as most maps encountered in the literature that are Bregman strongly nonexpansive are already uniformly Bregman strongly nonexpansive.

7.1. **Cyclic projections.** A first readily defined instantiation of Theorem 6.10 on the Halpern-iteration is that obtained by using the cyclic projection operator

$$T := P^f_{\Omega_k} \circ \cdots \circ P^f_{\Omega_1}$$

where $P^f_{\Omega_j}$ is the Bregman projection onto a given nonempty closed convex set Ω_j for $j = 1, \ldots, k$. Assume that $\Omega_1 \cap \cdots \cap \Omega_k \neq \emptyset$. Then this operator T is uniformly Bregman strongly nonexpansive since every projection $P^f_{\Omega_j}$ is even Bregman firmly nonexpansive and moduli for the Bregman strong nonexpansivity of T can be calculated from the moduli of the factors by following Theorem 4.16 as well as Lemma 4.9. For this, note further that by $\Omega_1 \cap \cdots \cap \Omega_k \neq \emptyset$, using Lemma 4.6, each $P^f_{\Omega_j}$ and thus T is bounded on bounded sets. Further, note that any Bregman firmly nonexpansive map that is bounded on bounded sets actually possesses a modulus of uniform closedness if ∇f is uniformly continuous on bounded subsets as well as uniformly strictly monotone (the latter of which, recalling the discussion from Remark 3.13, follows from the assumption that f is totally convex on bounded sets) as by Lemma 4.14, each such map is then uniformly continuous on bounded subsets. Thus $P^f_{\Omega_j}$ is uniformly continuous on bounded subsets. In particular, from a corresponding (common) modulus of uniform continuity, a (common) modulus κ of uniform closedness can be immediately defined. Note that through the uniform continuity of each $P^f_{\Omega_j}$, also T is uniformly continuous on bounded sets and thus also T possesses a modulus of uniform closedness which in particular yields that $\overline{F}(T) \subseteq F(T)$.

Combining this with Theorem 6.10, we get the following corollary on a Halpern-type variant of the method of cyclic projections (where we can identify the limit as the corresponding projection as we have previously established $\overline{F}(T) \subseteq F(T)$).

Theorem 7.1. *Let X be a real reflexive Banach space and $f : X \to \mathbb{R}$ be a supercoercive Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets. Let $\Omega_1, \dots, \Omega_k$ be nonempty closed convex sets and assume that $\Omega_1 \cap \dots \cap \Omega_k \neq \emptyset$. Given a $u \in X$, define a sequence x_n by $x_0 := x \in X$ and*

$$x_{n+1} := \nabla f^* \left(\alpha_n \nabla f u + (1 - \alpha_n) \nabla f P^f_{\Omega_k} \circ \dots \circ P^f_{\Omega_1} x_n \right)$$

where $(\alpha_n) \subseteq (0, 1]$ satisfies $\lim \alpha_n = 0$ and $\sum \alpha_n = \infty$. Then (x_n) converges strongly to $P^f_{F(T)}(u)$ for $T = P^f_{\Omega_k} \circ \dots \circ P^f_{\Omega_1}$.

In particular, a rate of metastability can be calculated using Theorem 6.9 together with Lemmas 4.6 and 4.9 as well as Theorems 4.16 and 4.15.

7.2. **The proximal point algorithm.** We are now concerned with a Halpern-type variant of the proximal point algorithm for a maximally monotone operator A with resolvents Res^f_γ as before. Concretely, for a given u and x_0, we consider the sequence (cf. also [58])

$$(**) \qquad x_{n+1} := \nabla f^* (\alpha_n \nabla f u + (1 - \alpha_n) \nabla f \mathrm{Res}^f_{r_n} x_n)$$

for a given additional sequence r_n that satisfies

$$0 < \bar{r} := \inf\{r_n \mid n \in \mathbb{N}\}.$$

To show that the previous results contained in Theorems 6.9 and 6.10 apply here, we will in the following provide concrete instantiations for the moduli μ and ν for the concrete choices of $T_n = \mathrm{Res}^f_{r_n}$ and $T = \mathrm{Res}^f_{\bar{r}}$.

For this, we will however need some further facts about the resolvent relative to f. It is straightforward to show that the set of fixed points of any Res^f_γ equals to the set of zeros $A^{-1}0$ of the operator A. The following lemma provides a quantitative result for one of the directions of the equivalence.

Lemma 7.2. *Let $\widehat{\eta}$ be a modulus of uniform strict monotonicity of ∇f on bounded sets. Given $\gamma > 0$ and $\varepsilon > 0$, let $(x, y) \in A$ with $b > 0$ such that $b \geq \|x\|, \|\mathrm{Res}^f_\gamma x\|, \gamma$. If we have $\|y\| < \widehat{\eta}(\varepsilon, b)/2b^2$, then $\|x - \mathrm{Res}^f_\gamma x\| < \varepsilon$.*

Proof. By monotonicity of A, we have $\langle \mathrm{Res}^f_\gamma x - x, A^f_\gamma x - y \rangle \geq 0$ and thus

$$\langle x - \mathrm{Res}^f_\gamma x, \nabla f x - \nabla f \mathrm{Res}^f_\gamma x \rangle \leq \gamma \langle x - \mathrm{Res}^f_\gamma x, y \rangle \leq \gamma (\|x\| + \|\mathrm{Res}^f_\gamma x\|) \|y\| \leq 2b^2 \|y\|.$$

Thus $\|y\| < \widehat{\eta}(\varepsilon, b)/2b^2$ implies $\|x - \mathrm{Res}^f_\gamma x\| < \varepsilon$ by the assumptions on $\widehat{\eta}$. □

The following lemma due to Reich and Sabach provides a crucial relation between the resolvent relative to f and the Bregman distance associated with f.

Lemma 7.3 ([49, 50])**.** *Let A be maximally monotone and assume that $A^{-1}0 \neq \emptyset$. Then*

$$D_f(u, \mathrm{Res}^f_\gamma x) + D_f(\mathrm{Res}^f_\gamma x, x) \leq D_f(u, x)$$

for all $\gamma > 0$, $u \in A^{-1}0$ and $x \in X$.

In particular, we will in the following rely on a quantitative version of this result as given in the next lemma.

Lemma 7.4. *Let $\omega^{\nabla f}(\varepsilon, b) \leq \varepsilon$ be a modulus of uniform continuity of ∇f on bounded subsets. Let $x, y \in X$ and $r, s > 0$ be given such that $b \geq \|x\|, \|\mathrm{Res}^f_s x\|, \|y\|, \|\mathrm{Res}^f_r y\|$. Then for any $\varepsilon > 0$, if*

$$\|x - Res^f_s x\| < \omega^{\nabla f}\left(\frac{\varepsilon}{2E}, b\right) \text{ for } E \geq \max\{2b, rs^{-1}2b\},$$

then we have

$$D_f(x, \mathrm{Res}^f_r y) + D_f(\mathrm{Res}^f_r y, y) < D_f(x, y) + \varepsilon.$$

A proof of the above result can be found in the appendix.

As a concrete instantiation of Theorem 6.10, we now obtain the following (cf. Theorem 5.1 in [58]):

Theorem 7.5. *Let X be a real reflexive Banach space and $f : X \to \mathbb{R}$ be a supercoercive Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets. Let A be a maximally monotone operator with resolvents Res^f_γ and assume that $A^{-1}0 \neq \emptyset$. Given a $u \in X$, define a sequence x_n by $x_0 := x \in X$ and*

$$x_{n+1} := \nabla f^*(\alpha_n \nabla f u + (1-\alpha_n)\nabla f \mathrm{Res}^f_{r_n} x_n)$$

where $(\alpha_n) \subseteq (0,1]$ satisfies $\lim \alpha_n = 0$ and $\sum \alpha_n = \infty$ and where $(r_n) \in (0,\infty)$ satisfies $0 < \bar{r} := \inf\{r_n \mid n \in \mathbb{N}\}$. Then (x_n) converges strongly to $P^f_{A^{-1}0}(u)$.

In particular, a rate of metastability can be calculated using Theorem 6.9 together with Lemmas 4.6 and 4.9 and with moduli

$$\mu(\varepsilon, b, K) := \omega^{\nabla f}(\rho(\varepsilon, \widehat{b})/2E'(K), \widehat{b}) \text{ and } \nu(\varepsilon, b) := \omega^{\nabla f}(\rho(\varepsilon, \widehat{b})/4\widehat{b}, \widehat{b})$$

for the uniform NST condition where $E'(K) := \max\{2\widehat{b}, R(K)\bar{r}^{-1}2\widehat{b}\}$ and $\widehat{b} := \max\{b, E(b)\}$ as well as $R(n) := \max\{r_k \mid k \leq n\}$ and where E is a modulus for Res^f_γ being bounded on bounded sets.

Proof. Note that using Lemmas 4.6 and 4.9 as well as $A^{-1}0 = \mathrm{F}(\mathrm{Res}^f_r)$ for any $r > 0$, it is immediate that the $\mathrm{Res}^f_{r_n}$ and $\mathrm{Res}^f_{\bar{r}}$ are commonly uniformly Bregman strongly nonexpansive and commonly bounded on bounded sets and corresponding moduli can be calculated. This also yields that a modulus of uniform closedness exists for $F(\mathrm{Res}^f_{\bar{r}})$.

The only thing left to prove is that the constructed μ and ν witness the uniform NST condition for $T_n = \mathrm{Res}^f_{r_n}$ and $T = \mathrm{Res}^f_{\bar{r}}$. By Lemma 7.4, we get that $\|x - \mathrm{Res}^f_s x\| < \omega^{\nabla f}(\varepsilon/2E', \widehat{b})$ for $\|x\| \leq b$ implies that $D_f(x, \mathrm{Res}^f_r x) < \varepsilon$ for $E' \geq \max\{2\widehat{b}, rs^{-1}2\widehat{b}\}$ and $\widehat{b} = \max\{b, E(b)\}$. In particular, we have $\|x - \mathrm{Res}^f_r x\| < \varepsilon$ for any x with $\|x\| \leq b$ and $\|x - \mathrm{Res}^f_s x\| < \omega^{\nabla f}(\rho(\varepsilon, \widehat{b})/2E', \widehat{b})$. So, for $s \geq r$ we get for $\|x\| \leq b$ and $\|x - \mathrm{Res}^f_s x\| < \omega^{\nabla f}(\rho(\varepsilon, \widehat{b})/4\widehat{b}, \widehat{b})$ that $\|x - \mathrm{Res}^f_r x\| < \varepsilon$. Therefore, as $\bar{r} \leq r_n$ for all n, we get that ν indeed satisfies $(\dagger)_2$ for the given T_n and T.

Further, assuming that $\|x - \mathrm{Res}^f_{\bar{r}} x\| < \mu(\varepsilon, b, K)$, we get by the above that $\|x - \mathrm{Res}^f_{r_n} x\| < \varepsilon$ as $E'(K) = \max\{2\widehat{b}, R(K)\bar{r}^{-1}2\widehat{b}\} \geq \max\{2\widehat{b}, r_n\bar{r}^{-1}2\widehat{b}\}$ for $n \leq K$. Thus μ satisfies $(\dagger)_1$ for the given T_n and T. ☐

7.3. **Finding common zeros of maximally monotone operators.** Another readily defined instantiation of Theorem 6.10 on the Halpern-iteration is that of finding common zeros of a finite collection $(A_i)_{i=1,\dots,N}$ of maximally monotone operators with $A_1^{-1}0 \cap \dots \cap A_N^{-1}0 \neq \emptyset$. Similar to the idea in [40], we in that context can consider a composite operator

$$Tx := \nabla f^* \sum_{i=1}^N w_i \nabla f \mathrm{Res}^f_{A_i} x$$

for weights $w_i \in (0,1)$ such that $\sum_{i=1}^N w_i = 1$. Then T is a block operator in the sense of [37, 38] (as also discussed in the previous sections) and moduli for the uniform Bregman strong nonexpansivity for this operator can be calculated from the moduli of the summands following Theorem 4.20. From Lemma 4.18, also a modulus for T being bounded on bounded sets can be calculated from corresponding moduli for $\nabla f, \nabla f^*$ and $\mathrm{Res}^f_{A_i}$ being bounded on bounded sets (using Lemma 4.6, the latter of which in particular exists as $A_1^{-1}0 \cap \dots \cap A_N^{-1}0 \neq \emptyset$ and as any $\mathrm{Res}^f_{A_i}$ is Bregman firmly nonexpansive and thus Bregman quasi-nonexpansive). Lastly, note that by Lemma 4.14, each $\mathrm{Res}^f_{A_i}$ is uniformly continuous on bounded sets and it is easy to see that, since ∇f, ∇f^* are also uniformly continuous, this extends to T as well. Therefore, a corresponding modulus of uniform closedness exists for $F(T)$.

Combining this with Theorem 6.10, we get the following corollary on the approximation of common zeros:

Theorem 7.6. *Let X be a real reflexive Banach space and $f : X \to \mathbb{R}$ be a supercoercive Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets. Let $A_1, \dots, A_N$ be maximally monotone operators with resolvents $\mathrm{Res}^f_{A_i}$ at parameter 1. Assume that $A_1^{-1}0 \cap \dots \cap A_N^{-1}0 \neq \emptyset$.*

Given a $u \in X$*, define a sequence* x_n *by* $x_0 := x \in X$ *and*

$$x_{n+1} := \nabla f^* \left(\alpha_n \nabla f u + (1-\alpha_n) \sum_{i=1}^{N} w_i \nabla f \mathrm{Res}^{\mathrm{f}}_{\mathrm{A_i}} x_n \right)$$

where $(\alpha_n) \subseteq (0,1]$ *satisfies* $\lim \alpha_n = 0$ *and* $\sum \alpha_n = \infty$ *and where the* $w_i \in (0,1)$ *are such that* $\sum_{i=1}^N w_i = 1$*. Then* (x_n) *converges strongly to* $P^f_{F(T)}(u)$ *for* T *defined as above.*

In particular, a rate of metastability can be calculated using Theorem 6.9 together with Lemmas 4.6 and 4.9 as well as Theorems 4.20 and 4.19.

7.4. **Modified-Halpern, Tikhonov-Mann and Halpern-Mann type methods.** In this last subsection, we are concerned with a few generalizations of Halpern-type iterations that incorporate elements from Mann-type iterations. The first such generalization that we consider is the modified Halpern iteration as introduced in [24] (see also [20])

$$x_{n+1} := \gamma_n u + (1-\gamma_n)(\alpha_n x_n + (1-\alpha_n)Tx_n)$$

where (γ_n) and (α_n) are sequences in $[0,1]$ and $T : X \to X$ is a given mapping. Such a type of iteration has been considered for Bregman strongly nonexpansive maps in [67] under the name of *Halpern-Mann iterations.* Concretely, in [67] the authors proved the strong convergence of the iteration

$$x_{n+1} := \nabla f^*(\alpha_n \nabla f u + (1-\alpha_n)(\beta_n \nabla f x_n + (1-\beta_n)\nabla f T x_n))$$

under the scalar conditions that (α_n) and (β_n) are sequences in $(0,1)$ satisfying

(1) $\alpha_n \to 0$ for $n \to \infty$,
(2) $\sum \alpha_n = +\infty$,
(3) $0 < \liminf \beta_n \leq \limsup \beta_n < 1$.

We begin by showing that for uniformly Bregman strongly nonexpansive maps, the convergence of this iteration can be derived by our previous result for families of maps. For this, note first that the above iteration is nothing else but a usual Halpern-type iteration of the family of operators

$$T_n x := \nabla f^*(\beta_n \nabla f x + (1-\beta_n)\nabla f T x)$$

Assume $F(T) \neq \emptyset$. Then these operators, being block operators, are uniformly Bregman strongly nonexpansive and using Theorem 4.20, we can construct even a common strong BSNE-modulus from a strong BSNE-modulus of T using the assumption of a fixed point for T. Note however that for this, condition (3) is not needed at all and $\beta_n \in [0,1]$ can be permitted. Also note that the T_n together with T are commonly bounded on bounded sets by using Lemmas 4.18 and 4.6 together with the assumption of a fixed point for T.

To see that this sequence is permissible for our Halpern-type iteration for families of maps, we need to again provide concrete instantiations of the moduli μ and ν witnessing the uniform NST condition for the choice of these T_n together with the map T. For this, it is rather immediately clear that given moduli E, C, F for T, ∇f, ∇f^* being bounded on bounded sets as well as a modulus of consistency ρ and a modulus of reverse consistency P, one has that

$$\mu(\varepsilon, b, K) := P(\rho(\varepsilon, \max\{b, F(C(E(b)))\}), \max\{b, E(b)\})$$

suffices as we immediately have for given $\varepsilon, b > 0$ and $p \in X$ with $\|p - Tp\| < \mu(\varepsilon, b, K)$ that

$$D_f(p, T_n p) \leq (1-\beta_n) D_f(p, Tp) \leq D_f(p, Tp) < \rho(\varepsilon, \max\{b, F(C(E(b)))\})$$

so that $\|p - T_n p\| < \varepsilon$.
For ν, assume that we have an $N_{\overline{\beta}}$ and a $\overline{\beta} < 1$ with $\beta_n \leq \overline{\beta}$ for all $n \geq N_{\overline{\beta}}$ (witnessing $\limsup_n \beta_n < 1$), a modulus of consistency ρ, a modulus of uniform continuity of D_f in its second argument ξ, a BSNE-modulus ω for T, moduli E, C, F for T, ∇f, ∇f^* being bounded on bounded sets, and a fixed point of T named p_0 with $b \geq \|p_0\|$. Then by Theorem 4.19: for any x with $\|x\| \leq b$, we have

$$\|x - T_n x\| < \xi\left((1-\overline{\beta})\omega(\rho(\varepsilon, \widehat{b}), b), \widehat{b}\right) \to \|x - Tx\| < \varepsilon$$

for $n \geq N_{\overline{\beta}}$ where $\widehat{b} := \max\{b, E(b), F(C(E(b)))\}$ so that

$$\nu(\varepsilon, b) := \xi\left((1-\overline{\beta})\omega(\rho(\varepsilon, \widehat{b}), b), \widehat{b}\right)$$

suffices (after suitably shifting the sequence with $N_{\overline{\beta}}$). Combined, we thus derive the following result from Theorem 6.10.

Theorem 7.7. *Let X be a real reflexive Banach space and $f : X \to \mathbb{R}$ be a supercoercive Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets. Let T be uniformly Bregman strongly nonexpansive with $F(T) \neq \emptyset$. Given a $u \in X$, define a sequence x_n by $x_0 := x \in X$ and*

$$x_{n+1} := \nabla f^*(\alpha_n \nabla fu + (1-\alpha_n)(\beta_n \nabla fx_n + (1-\beta_n)\nabla fTx_n))$$

where $(\alpha_n) \subseteq (0,1]$ satisfies $\lim \alpha_n = 0$ and $\sum \alpha_n = \infty$ and where $(\beta_n) \subseteq [0,1)$ satisfies $\limsup \beta_n < 1$. If $\overline{F}(T) \subseteq F(T)$, then (x_n) converges strongly to $P^f_{F(T)}(u)$.

In particular, a rate of metastability can be obtained by suitably instantiating the rate given in Theorem 6.9 using Theorems 4.20 and 4.19 as well as the above moduli for μ and ν.

In particular, with this theorem we reobtain the strong convergence result for this iteration established in [67] (recall for this Remark 6.11) for uniformly Bregman strongly nonexpansive maps. However, the assumption (3) presented above which features in [67] could be substantially weakened to $\limsup \beta_n < 1$ which in particular allows $\beta_n = 0$ for all n. Thus, in the above iteration, the Mann-part can be 'deactivated' and the original Halpern-type result can be reobtained, contrary to [67].

The other generalization of Halpern's method which we consider is an iteration of Tikhonov-Mann type. In the usual metric case, this type of iteration takes the form of

$$y_{n+1} := (1-\lambda_n)((1-\beta_n)u + \beta_n x_n) + \lambda_n T((1-\beta_n)u + \beta_n x_n)$$

as defined in [19] where (λ_n), (β_n) are sequences in $[0,1]$ and $T : X \to X$ is a again a given mapping. In particular, for $u = 0$, this iteration becomes the modified Mann iteration as studied in [64] and rediscovered in the seminal work by Boţ, Csetnek and Meier [8]. For these types of iterations, we can now prove a (new) strong convergence result for the following natural analog in the context of uniformly Bregman strongly nonexpansive maps:

$$y_{n+1} := \nabla f^*(\beta_n \nabla fu_n + (1-\beta_n)\nabla fTu_n) \text{ with } u_n := \nabla f^*(\alpha_n \nabla fu + (1-\alpha_n)\nabla fy_n).$$

As discussed in [18], methods of a modified Halpern type as well as methods of a Tikhonov-Mann type in both a normed and a hyperbolic context are closely related and in fact can be translated into each other.

By suitably adapting the arguments from [18] to this Bregman case, we arrive at the following result (which is similar to Proposition 3.2 in [18]):

Lemma 7.8. *Define the iterations*

$$x_{n+1} := \nabla f^*(\alpha_{n+1}\nabla fu + (1-\alpha_{n+1})\nabla fv_n) \text{ with } v_n := \nabla f^*(\beta_n \nabla fx_n + (1-\beta_n)\nabla fTx_n))$$

as well as

$$y_{n+1} := \nabla f^*(\beta_n \nabla fu_n + (1-\beta_n)\nabla fTu_n) \text{ with } u_n := \nabla f^*(\alpha_n \nabla fu + (1-\alpha_n)\nabla fy_n).$$

If $x_0 = \nabla f^*(\alpha_0 \nabla fu + (1-\alpha_0)\nabla fy_0)$, then for any $n \in \mathbb{N}$: $u_n = x_n$ and $y_{n+1} = v_n$.

Proof. The proof is by induction on n. For $n = 0$, it follows by the definition of u_0 as well as the assumption on x_0 that $x_0 = u_0$. From that, we get

$$y_1 = \nabla f^*(\beta_0 \nabla fu_0 + (1-\beta_0)\nabla fTu_0) = \nabla f^*(\beta_0 \nabla fx_0 + (1-\beta_0)\nabla fTx_0) = v_0.$$

For the induction step, suppose now that $u_n = x_n$ and $y_{n+1} = v_n$. Then

$$x_{n+1} = \nabla f^*(\alpha_{n+1}\nabla fu + (1-\alpha_{n+1})\nabla fv_n) = \nabla f^*(\alpha_{n+1}\nabla fu + (1-\alpha_{n+1})\nabla fy_{n+1}) = u_{n+1},$$

where the second equality follows by induction hypothesis. Further, we thus have

$$y_{n+2} = \nabla f^*(\beta_{n+1}\nabla fu_{n+1} + (1-\beta_{n+1})\nabla fTu_{n+1}) = \nabla f^*(\beta_{n+1}\nabla fx_{n+1} + (1-\beta_{n+1})\nabla fTx_{n+1}) = v_{n+1}.$$

□

Together with the above theorem, this allows us to derive the following new strong convergence result:

Theorem 7.9. *Let X be a real reflexive Banach space and $f : X \to \mathbb{R}$ be a supercoercive Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets. Let T be uniformly Bregman strongly nonexpansive with $F(T) \neq \emptyset$. Given a $u \in X$, define a sequence y_n by $y_0 := y \in X$ and*

$$y_{n+1} := \nabla f^*(\beta_n \nabla fu_n + (1-\beta_n)\nabla fTu_n) \text{ with } u_n := \nabla f^*(\alpha_n \nabla fu + (1-\alpha_n)\nabla fy_n)$$

where $(\alpha_n) \subseteq (0,1]$ satisfies $\lim \alpha_n = 0$ and $\sum \alpha_n = \infty$ and where $(\beta_n) \subseteq [0,1)$ satisfies $\limsup \beta_n < 1$. If $\overline{F}(T) \subseteq F(T)$, then (y_n) converges strongly to $P^f_{F(T)}(u)$.

In particular, a rate of metastability can be obtained by suitably translating the rate from Theorem 7.7.

Proof. It suffices to show that given a rate of metastability Ω for the sequence x_n as defined in Theorem 7.7 (with α_{n+1} instead of α_n), i.e. Ω satisfying

$$\forall\varepsilon > 0, g : \mathbb{N} \to \mathbb{N}\ \exists n \leq \Omega(\varepsilon, g)\ \forall i, j \in [n; n + g(n)]\ (\|x_i - x_j\| < \varepsilon),$$

we can construct a rate of metastability for y_n.

For this, note first that $\|y_n - u_n\| \to 0$ for $n \to \infty$ and we can witness this limit even by a rate of convergence. To see this, let $\bar{b}$ be such that $\bar{b} \geq D_f(y_n, u), \|y_n\|, \|u_n\|$ for all n.[10] Let σ be a rate of convergence for $\alpha_n \to 0$ as before. Then we get

$$D_f(y_n, u_n) \leq \alpha_n D_f(y_n, u) + (1 - \alpha_n)D_f(y_n, y_n) = \alpha_n D_f(y_n, u)$$

so that for $n \geq \sigma(\varepsilon/\bar{b})$, we have $D_f(y_n, u_n) < \varepsilon$. In particular, for $n \geq \sigma(\rho(\varepsilon, \bar{b})/\bar{b})$ we get $\|y_n - u_n\| < \varepsilon$.

We can now construct a rate of metastability for y_n given one for x_n. At first, using Lemma 7.8, we get $u_n = x_n$ for all n so that Ω is also a rate of metastability for u_n. Then

$$\|y_i - y_j\| \leq \|y_i - u_i\| + \|u_i - u_j\| + \|u_j - y_j\|$$

and by reasoning similar to [18], it can be rather immediately seen that Ω' defined by

$$\Omega'(\varepsilon, g) := \widetilde{\Omega}(\varepsilon/3, g, \sigma(\rho(\varepsilon/3, \bar{b})/\bar{b})),$$
$$\widetilde{\Omega}(\varepsilon, g, q) := \Omega(\varepsilon, g_q) + q \text{ with } g_q(n) := g(n + q) + q,$$

is therefore a rate of metastability for y_n. □

8. A rate of convergence for the asymptotic regularity of the Halpern-type proximal point algorithm

We now turn back to the Halpern-type proximal point algorithm studied before. For that, we again fix a maximally monotone operator A with resolvents Res^f_γ and for given u and x_0, we consider the sequence (x_n) previously defined in $(**)$, i.e.

$$x_{n+1} := \nabla f^*(\alpha_n \nabla f u + (1 - \alpha_n)\nabla f \mathrm{Res}^f_{r_n} x_n),$$

where r_n is assumed to satisfy $0 < \bar{r} := \inf\{r_n \mid n \in \mathbb{N}\}$. The convergence proof of this algorithm as studied before relies on an argument revolving around a case distinction and (essentially) because of this, we are not able to derive full rates of convergence for the asymptotic regularity relative to the resolvents, i.e. rates for the convergence

$$\left\|x_n - \mathrm{Res}^f_\gamma x_n\right\| \to 0 \quad (n \to \infty)$$

for $\gamma > 0$. In this final section, we consider this method under the additional condition $r_n \to \infty$ (conceptually similar to the work of Kohsaka and Takahashi [32]) for which we are able to derive full rates of convergence for the asymptotic regularity relative to the resolvents in the above sense. In the case of Hilbert spaces with the ordinary Halpern-type proximal point algorithm induced by a maximally monotone operator, such a rate of convergence (in the context of the assumption of $r_n \to \infty$ similar to here) was first given by Pinto in [42].

For this, we begin with the following preliminary result:

Lemma 8.1. *Let $b \geq \|u\|, \|x_n\|, \left\|\mathrm{Res}^f_{r_n} x_n\right\|$ for all n with (x_n) defined as above and let σ be a rate of convergence for $\alpha_n \to 0$ as $n \to \infty$. Let further C be a modulus for ∇f being bounded on bounded sets. Then, for any $\varepsilon > 0$:*

$$\forall n \geq \sigma\left(\frac{\varepsilon}{2C(b)}\right)\left(\left\|\nabla f x_{n+1} - \nabla f \mathrm{Res}^f_{r_n} x_n\right\| < \varepsilon\right).$$

In particular, if $\omega^{\nabla f^*}$ is a modulus of uniform continuity for ∇f^* on bounded subsets, then for any $\varepsilon > 0$:

$$\forall n \geq \sigma\left(\frac{\omega^{\nabla f^*}(\varepsilon, C(b))}{2C(b)}\right)\left(\left\|x_{n+1} - \mathrm{Res}^f_{r_n} x_n\right\| < \varepsilon\right).$$

[10] Such a $\bar{b}$ can naturally be constructed from a $b \geq D_f(p, u), D_f(p, y_0)$ for a given fixed point p together with a modulus of boundedness for D_f and moduli for ∇f, f being bounded on bounded sets. We omit the details.

Proof. As before, we have

$$\left\|\nabla f x_{n+1} - \nabla f \mathrm{Res}^f_{r_n} x_n\right\| = \alpha_n \left\|\nabla f u - \nabla f \mathrm{Res}^f_{r_n} x_n\right\| \leq \alpha_n 2C(b).$$

It immediately follows from the assumption on σ that $\|\nabla f x_{n+1} - \nabla f \mathrm{Res}^f_{r_n} x_n\| < \varepsilon$ for all $n \geq \sigma(\varepsilon/2C(b))$. The second part of the lemma is immediate. □

Theorem 8.2. *Let $\gamma > 0$ be given. Assume that $b > 0$ is such that*

$$b \geq \|u\|, \|x_n\|, \|\mathrm{Res}^f_{r_n} x_n\|, \gamma, \|\mathrm{Res}^f_\gamma \mathrm{Res}^f_{r_n} x_n\|, \|\mathrm{Res}^f_\gamma x_n\|$$

for all n with (x_n) defined as above. Let P be a modulus of reverse consistency. Assume that $\widehat{\eta}$ is a modulus of uniform strict monotonicity of ∇f on bounded sets and let ρ be a modulus of consistency for f. Let further C be a modulus for ∇f being bounded on bounded sets and let σ be a rate of convergence for $\alpha_n \to 0$ as $n \to \infty$. Let τ be a rate of divergence for $r_n \to \infty$ as $n \to \infty$, i.e. $\forall E > 0\ \forall n \geq \tau(E)\,(r_n > E)$. Let $\omega^{\nabla f^*}$ be a modulus of uniform continuity for ∇f^* on bounded subsets and let $\omega^{\nabla f}(\varepsilon, b) \leq \varepsilon$ be a modulus of uniform continuity of ∇f on bounded subsets. Then for any $\varepsilon > 0$:

$$\forall n \geq \Phi(\varepsilon) \left(\left\|x_{n+1} - \mathrm{Res}^f_\gamma x_{n+1}\right\| < \varepsilon\right)$$

where

$$\Phi(\varepsilon) := \max\left\{\tau\left(\frac{2C(b)}{\chi\left(\omega^{\nabla f}\left(\frac{\rho(\omega^{\nabla f}(\omega^{\nabla f^*}(\varepsilon, C(b))/2, b), b)}{8b}, b\right)\right)}\right),\right.$$
$$\left.\varphi\left(P\left(\frac{\rho(\omega^{\nabla f}(\omega^{\nabla f^*}(\varepsilon, C(b))/2, b), b)}{2}, b\right)\right), \sigma\left(\frac{\omega^{\nabla f^*}(\varepsilon, C(b))/2}{2C(b)}\right)\right\}$$

and

$$\chi(\varepsilon) := \frac{\widehat{\eta}(\varepsilon, b)}{2b^2},\ \varphi(\varepsilon) := \sigma\left(\frac{\omega^{\nabla f^*}(\varepsilon, C(b))}{2C(b)}\right).$$

Proof. Note that

$$\|A^f_{r_n} x_n\| = \frac{1}{r_n}\left\|\nabla f x_n - \nabla f \mathrm{Res}^f_{r_n} x_n\right\| \leq \frac{2C(b)}{r_n}$$

and thus for any $\varepsilon > 0$ and any $n \geq \tau(2C(b)/\varepsilon)$, we have $\|A^f_{r_n} x_n\| < \varepsilon$. Therefore, since $A^f_{r_n} x_n \in A(\mathrm{Res}^f_{r_n} x_n)$, we have that $n \geq \tau(2C(b)/\chi(\varepsilon))$ implies $\|\mathrm{Res}^f_{r_n} x_n - \mathrm{Res}^f_\gamma \mathrm{Res}^f_{r_n} x_n\| < \varepsilon$ by Lemma 7.2. Therefore, we have for

$$n \geq \tau\left(\frac{2C(b)}{\chi(\omega^{\nabla f}\left(\frac{\varepsilon}{8b}, b\right))}\right)$$

that $\|\mathrm{Res}^f_{r_n} x_n - \mathrm{Res}^f_\gamma \mathrm{Res}^f_{r_n} x_n\| < \omega^{\nabla f}(\varepsilon/8b, b)$, and thus $D_f(\mathrm{Res}^f_{r_n} x_n, \mathrm{Res}^f_\gamma x_{n+1}) \leq D_f(\mathrm{Res}^f_{r_n} x_n, x_{n+1}) + \varepsilon/2$ in that case by Lemma 7.4 (with $s := r := \gamma$ and using $E := 2b$). Now, for

$$n \geq \max\left\{\tau\left(\frac{2C(b)}{\chi(\omega^{\nabla f}\left(\frac{\varepsilon}{8b}, b\right))}\right), \varphi(P(\varepsilon/2, b))\right\}$$

we get $D_f(\mathrm{Res}^f_{r_n} x_n, \mathrm{Res}^f_\gamma x_{n+1}) \leq D_f(\mathrm{Res}^f_{r_n} x_n, x_{n+1}) + \varepsilon/2$ from before as well as that $D_f(\mathrm{Res}^f_{r_n} x_n, x_{n+1}) < \varepsilon/2$ by Lemma 8.1 and the assumption on P. Thus in that case, we also have $D_f(\mathrm{Res}^f_{r_n} x_n, \mathrm{Res}^f_\gamma x_{n+1}) < \varepsilon$. Thus for

$$n \geq \max\left\{\tau\left(\frac{2C(b)}{\chi\left(\omega^{\nabla f}\left(\frac{\rho(\varepsilon, b)}{8b}, b\right)\right)}\right), \varphi(P(\rho(\varepsilon, b)/2, b))\right\}$$

we get $\|\mathrm{Res}^f_{r_n} x_n - \mathrm{Res}^f_\gamma x_{n+1}\| < \varepsilon$ using the assumption on ρ. Now, note that

$$\begin{aligned}\|\nabla f x_{n+1} - \nabla f \mathrm{Res}_\gamma x_{n+1}\| &\leq \alpha_n \left\|\nabla f u - \nabla f \mathrm{Res}^f_{r_n} x_n\right\| + \left\|\nabla f \mathrm{Res}^f_{r_n} x_n - \nabla f \mathrm{Res}^f_\gamma x_{n+1}\right\| \\ &\leq \alpha_n 2C(b) + \left\|\nabla f \mathrm{Res}^f_{r_n} x_n - \nabla f \mathrm{Res}^f_\gamma x_{n+1}\right\|\end{aligned}$$

Thus, for

$$n \geq \max \left\{ \tau \left(\frac{2C(b)}{\chi \left(\omega^{\nabla f} \left(\frac{\rho(\omega^{\nabla f}(\varepsilon/2,b),b)}{8b}, b \right) \right)} \right), \varphi \left(P \left(\frac{\rho(\omega^{\nabla f}(\varepsilon/2,b),b)}{2}, b \right) \right), \sigma \left(\frac{\varepsilon/2}{2C(b)} \right) \right\}$$

we get $\|\nabla f x_{n+1} - \nabla f \mathrm{Res}^f_\gamma x_{n+1}\| < \varepsilon$. This gives the claim using $\omega^{\nabla f^*}$. □

Corollary 8.3. In the context of the assumptions of the above Theorem 8.2, we have for any $\varepsilon > 0$ that

$$\forall n \geq \Phi(\omega^{\nabla f}(\varepsilon, b))\ \exists z \in X \left(\|z\| < \varepsilon \text{ and } z \in A(\mathrm{Res}^f_1 x_{n+1}) \right)$$

where Φ is as in Theorem 8.2 (with $\gamma = 1$).

Proof. Using Theorem 8.2, we get $\|\nabla f x_{n+1} - \nabla f \mathrm{Res}^f_1 x_{n+1}\| < \varepsilon$ for any $n \geq \Phi(\omega^{\nabla f}(\varepsilon, b))$. By definition of the relativized resolvent Res^f_1, we get $\nabla f x_{n+1} - \nabla f \mathrm{Res}^f_1 x_{n+1} \in A\mathrm{Res}^f_1 x_{n+1}$ and this yields the above result with $z = \nabla f x_{n+1} - \nabla f \mathrm{Res}^f_1 x_{n+1}$. □

Remark 8.4. Some of the moduli featuring in the assumptions of the above Theorem 8.2 can be derived from a smaller set of core moduli via very simple transformations: From a modulus $\omega^{\nabla f}$ for the uniform continuity of ∇f on bounded subsets, we can construct a modulus C for ∇f being bounded on bounded sets using Lemma 3.2. Using Lemma 3.10, we can derive a modulus of reverse consistency P from C. Alternatively to assuming a modulus $\widehat{\eta}$ for the uniform strict monotonicity of ∇f, we could assume a modulus η for the uniform strict convexity of f from which we can immediately reobtain an $\widehat{\eta}$ (using Lemma 3.4) and from which one can easily derive a modulus of consistency ρ (as discussed in Remark 3.13). So one can define the resulting rate already in terms of the moduli $b, \eta, \sigma, \tau, \omega^{\nabla f}$ and $\omega^{\nabla f *}$. As before, a b bounding all objects involved can be constructed using the range of moduli discussed before together with some simple initial bounds but we refrain from spelling this out in more detail.

Acknowledgments: The results of this paper form the main part of Chapter 9 of the doctoral dissertation [46] submitted by the first author on October 10, 2023, to the Department of Mathematics of TU Darmstadt and written under the supervision of the second author.

Funding and/or Conflicts of interests Both authors were supported by the 'Deutsche Forschungsgemeinschaft' Project DFG KO 1737/6-2. The authors have no relevant financial or non-financial interests to disclose.

Data Availability Data sharing not applicable to this article as no data sets were generated or analyzed.

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, 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.

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

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

[5] 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.

[6] J.M. Borwein, A.J. Guirao, P. Hájek, and J. Vanderwerff. Uniformly convex functions on Banach spaces. *Proceedings of the American Mathematical Society*, 137(3):1081–1091, 2009.

[7] J.M. Borwein and J. Vanderwerff. Fréchet-Legendre functions and reflexive Banach spaces. *Journal of Convex Analysis*, 17(3–4):915–924, 2010.

[8] R.I. Boţ, E.R. Csetnek, and D. Meier. Inducing strong convergence into the asymptotic behaviour of proximal splitting algorithms in Hilbert spaces. *Optimization Methods and Software*, 34:489–514, 2019.

[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] A. Brøndsted. Conjugate convex functions in topological vector spaces. *Mat. Fys. Medd. Danske Vid. Selsk.*, 34, 1964.

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

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

[13] M.N. Búi and P.L. Combettes. Bregman forward-backward operator splitting. *Set-Valued and Variational Analysis*, 29:583–603, 2021.

[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, A.N. Iusem, and C. Zălinescu. On uniform convexity, total convexity and convergence of the proximal point and outer Bregman projection algorithm in Banach spaces. *Journal of Convex Analysis*, 10(1):35–62, 2003.

[16] 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.

[17] Y. Censor and S. Reich. Iterations of paracontractions and firmly nonexpansive operators with applications to feasibility and optimization. *Optimization*, 37:323–339, 1996.

[18] H. Cheval, U. Kohlenbach, and L. Leuştean. On modified Halpern and Tikhonov-Mann iterations. *Journal of Optimization Theory and Applications*, 197:233–251, 2023.

[19] H. Cheval and L. Leuştean. Quadratic rates of asymptotic regularity for the Tikhonov-Mann iteration. *Optimization Methods and Software*, 37:2225–2240, 2022.

[20] A. Cuntavenapit and B. Panyanak. Strong convergence of modified Halpern iterations in CAT(0) spaces. *Journal of Fixed Point Theory and Applications*, 16:Article no. 869458, 2011.

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

[22] W. Fenchel. On conjugate convex functions. *Canadian Journal of Mathematics*, 1:73–77, 1949.

[23] J. García-Falset, E. Llorens-Fuster, and T. Suzuki. Fixed point theory for a class of generalized nonexpansive mappings. *Journal of Mathematical Analysis and Applications*, 375:185–195, 2011.

[24] T.-H. Kim and H.-K. Xu. Strong convergence of modified Mann iterations. *Nonlinear Analysis*, 61:51–60, 2005.

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

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

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

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

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

[30] U. Kohlenbach. Quantitative analysis of a Halpern-type Proximal Point Algorithm for accretive operators in Banach spaces. *Journal of Nonlinear and Convex Analysis*, 21(9):2125–2138, 2020.

[31] U. Kohlenbach, L. Leuştean, and A. Nicolae. Quantitative results on Fejér monotone sequences. *Communications in Contemporary Mathematics*, 20:Article no. 1750015, 2018.

[32] F. Kohsaka and W. Takahashi. Strong convergence of an iterative sequence for maximal monotone operators in a Banach space. *Abstract and Applied Analysis*, 3:239–249, 2004.

[33] F. Kohsaka and W. Takahashi. Proximal point algorithms with Bregman functions in Banach spaces. *Journal of Nonlinear and Convex Analysis*, 6(3):505–523, 2005.

[34] D. Körnlein. Quantitative results for Halpern iterations of nonexpansive mappings. *Journal of Mathematical Analysis and Applications*, 428:1161–1172, 2015.

[35] P.-E. Maingé. Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization. *Set-Valued Analysis*, 16:899–912, 2008.

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

[37] 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.

[38] 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.

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

[40] E. Naraghirad. Compositions and convex combinations of Bregman weakly relatively nonexpansive operators in reflexive Banach spaces. *Journal of Fixed Point Theory and Applications*, 22, 2020. Article no. 65.

[41] Q. Van Nguyen. Forward-backward splitting with Bregman distances. *Vietnam Journal of Mathematics*, 45(3):519–539, 2017.

[42] P. Pinto. A rate of metastability for the Halpern type Proximal Point Algorithm. *Numerical Functional Analysis and Optimization*, 42(3):320–343, 2021.

[43] N. Pischke. Generalized Fejér monotone sequences and their finitary content. *Optimization*, 2024. to appear, doi:10.1080/02331934.2024.2390114.

[44] N. Pischke. Proof mining for the dual of a Banach space with extensions for uniformly Fréchet differentiable functions. *Transactions of the American Mathematical Society*, 377(10):7475–7517, 2024.

[45] N. Pischke. Duality, Fréchet differentiability and Bregman distances in hyperbolic spaces. *Israel Journal of Mathematics*, 2024. to appear, manuscript available at `https://sites.google.com/view/nicholaspischke/notes-and-papers`.

[46] N. Pischke. *Proof-Theoretical Aspects of Nonlinear and Set-Valued Analysis*. PhD thesis, TU Darmstadt, 2024. Thesis available at `https://doi.org/10.26083/tuprints-00026584`

[47] 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.

[48] 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.

[49] 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.

[50] 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.
[51] E. Resmerita. On total convexity, Bregman projections and stability in Banach spaces. *Journal of Convex Analysis*, 11:1–16, 2004.
[52] R.T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific Journal of Mathematics*, 17:497–510, 1966.
[53] R.T. Rockafellar. Level sets and continuity of conjugate convex functions. *Transactions of the American Mathematical Society*, 123(1):46–63, 1966.
[54] R.T. Rockafellar. *Convex analysis*. Princeton university press, 1970.
[55] R.T. Rockafellar. On the maximal monotonicity of subdifferential mappings. *Pacific Journal of Mathematics*, 33:209–216, 1970.
[56] R.T. Rockafellar and R.J.B. Wets. *Variational Analysis*. Grundlehren der mathematischen Wissenschaften. Springer Berlin, Heidelberg, 1998.
[57] A. Sipoş. Abstract strongly convergent variants of the proximal point algorithm. *Computational Optimization and Applications*, 83:349–380, 2022.
[58] S. Suantai, Y.J. Cho, and P. Cholamjiak. Halpern's iteration for Bregman strongly nonexpansive mappings in reflexive Banach spaces. *Computers & Mathematics with Applications*, 64(4):489–499, 2012.
[59] T. Suzuki. Fixed point theorems and convergence theorems for some generalized nonexpansive mappings. *Journal of Mathematical Analysis and Applications*, 340:1088–1095, 2008.
[60] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory and Dynamical Systems*, 28(2):657–688, 2008.
[61] 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.
[62] A.A. Vladimirov, Y.E. Nesterov, and Y.N. Cěkanov. Uniformly convex functionals. *Vestnik Moskovskogo Universiteta. Seriya XV. Vychislitel'naya Matematika i Kibernetika*, 3:12–23, 1978.
[63] H.-K. Xu. An iterative approach to quadratic optimization. *Journal of Optimization Theory and Applications*, 116:659–678, 2003.
[64] Y. Yao, H. Zhou, and Y.-C. Liou. Strong convergence of a modified Krasnoselski-Mann iterative algorithm for non-expansive mappings. *Journal of Applied Mathematics and Computing*, 29:383–389, 2009.
[65] C. Zălinescu. On uniformly convex functions. *Journal of Mathematical Analysis and Applications*, 95(2):344–374, 1983.
[66] C. Zălinescu. *Convex analysis in general vector spaces*. World scientific, 2002.
[67] J.H. Zhu and S.S. Chang. Halpern-Mann's iterations for Bregman strongly nonexpansive mappings in reflexive Banach spaces with applications. *Journal of Inequalities and Applications*, page Article no. 146, 2013.

Appendix

Proof of Theorem 4.15 (compare also [27]). Note first that

$$\chi_b(n,\varepsilon) \leq \min\{\rho(\kappa(\varepsilon,b),b),\rho(\varepsilon,b)\}. \tag{0}$$

Also note that every T_k is in particular Bregman quasi-nonexpansive w.r.t. p. We show by induction on $1 \leq k \leq N$ that $D_f(T_k \circ \cdots \circ T_1 x, x) < \chi_b(k-1,\varepsilon)$ implies $\|x - T_i x\| < \varepsilon$ for $1 \leq i \leq k$. For $k = 1$, the statement trivially holds since $\chi_b(0,\varepsilon) \leq \rho(\varepsilon,b)$. So let $1 < k \leq N$ and assume that the claim holds for $k-1$ and that

$$\begin{aligned}&D_f(T_k \circ \cdots \circ T_1 x, x) < \chi_b(k-1,\varepsilon)\\ &\quad = \min\{\rho(\xi(\omega(\min\{\theta(\chi_b(k-2,\varepsilon),b),\rho(\kappa(\varepsilon,b),b)\},b),b),b),\chi_b(k-2,\varepsilon),\theta(\chi_b(k-2,\varepsilon),b)\}.\end{aligned} \tag{1}$$

For $y = T_{k-1} \circ \cdots \circ T_1 x$, we have

$$\|x - T_k y\| < \xi(\omega(\min\{\theta(\chi_b(k-2,\varepsilon),b),\rho(\kappa(\varepsilon,b),b)\},b),b). \tag{2}$$

Hence by (2), the assumption on ξ and $p \in \bigcap_{i=1}^k F(T_i)$, we derive

$$\begin{aligned}D_f(p,y) - \omega(\min\{\theta(\chi_b(k-2,\varepsilon),b),\rho(\kappa(\varepsilon,b),b)\},b) &\leq D_f(p,x) - \omega(\min\{\theta(\chi_b(k-2,\varepsilon),b),\rho(\kappa(\varepsilon,b),b)\},b)\\ &< D_f(p,T_k y)\end{aligned}$$

where we in particular used that $D_f(p,y) \leq D_f(p,x)$. Thus, since ω is a BSNE-modulus for T_k:

$$D_f(T_k y, y) < \min\{\theta(\chi_b(k-2,\varepsilon),b),\rho(\kappa(\varepsilon,b),b)\}. \tag{3}$$

By (1) and (3) together with the assumption on θ, we thus obtain

$$D_f(T_{k-1} \circ \cdots \circ T_1 x, x) = D_f(y,x) < \chi_b(k-2,\varepsilon) \tag{4}$$

from which we derive $\|x - T_i x\| < \varepsilon$ for all $i = 1,\dots,k-1$ using the induction hypothesis. From (0) and (4) together with the definition of ρ, we also get $\|x - T_{k-1} \circ \cdots \circ T_1 x\| < \kappa(\varepsilon,b)$ and so by (3), we obtain $\|x - T_k x\| < \varepsilon$. □

Proof of Theorem 4.16 (compare also [27]). Define

$$\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b) := \min\left\{\omega_1(\rho(\varepsilon_1,\widehat{b}),\widehat{b}),\dots,\omega_n(\rho(\varepsilon_n,\widehat{b}),\widehat{b})\right\},$$

$$\widehat{\omega}'(\varepsilon_1,\dots,\varepsilon_n,b) := \min\left\{\omega_1'(\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)/2(n-1),\widehat{b}),\dots,\omega_n'(\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)/2(n-1),\widehat{b})\right\}$$

as well as

$$\omega(\varepsilon_1,\dots,\varepsilon_n,b) := \min\left\{\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)/2, P(\varphi(\min\{\widehat{\omega}'(\varepsilon_1,\dots,\varepsilon_n,b),\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)\},\widehat{b},n),\widehat{b})\right\}.$$

Now, suppose

$$\|p-Tp\|, D_f(p,x)-D_f(p,Tx) < \omega(\varepsilon_1,\dots,\varepsilon_n,b),$$

for points x,p with $\|x\|,\|p\|\leq b$. Then Theorem 4.15 yields that

$$\|p-T_ip\| < \min\{\widehat{\omega}'(\varepsilon_1,\dots,\varepsilon_n,b),\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)\}.$$

Therefore, we get

$$\begin{aligned}
D_f(p,Tx) &= D_f(p,T_n\circ T_{n-1}\circ\dots\circ T_1x)\\
&\leq D_f(p,T_{n-1}\circ\dots\circ T_1x)+\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)/2(n-1)\\
&\leq \dots\\
&\leq D_f(p,T_1x)+(n-1)\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)/2(n-1)\\
&\leq D_f(p,x)+n\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)/2(n-1)
\end{aligned}$$

and therefore

$$\begin{aligned}
D_f(p,T_{i-1}\circ\dots\circ T_1x)-D_f(p,T_i\circ T_{i-1}\circ\dots\circ T_1x) &\leq D_f(p,x)-D_f(p,Tx)+(n-1)\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)/2(n-1)\\
&< \widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)\\
&\leq \omega_i(\rho(\varepsilon_i,\widehat{b}),\widehat{b})
\end{aligned}$$

for any $i=1,\dots,n$. This, together with $\|p-T_ip\|<\omega_i(\rho(\varepsilon_i,\widehat{b}),\widehat{b})$, yields

$$D_f(T_i\circ T_{i-1}\circ\dots\circ T_1x, T_{i-1}\circ\dots\circ T_1x) < \rho(\varepsilon_i,\widehat{b})$$

as ω_i is a strong BSNE-modulus for T_i. In particular, we have

$$\|T_i\circ T_{i-1}\circ\dots\circ T_1x - T_{i-1}\circ\dots\circ T_1x\| < \varepsilon_i$$

so that we get $\|x-Tx\|<\varepsilon_1+\dots+\varepsilon_n$. Now, for $\varepsilon_i = P(\varepsilon,\widehat{b})/n$, we then get $\|x-Tx\|<P(\varepsilon,\widehat{b})$ so that $D_f(Tx,x)<\varepsilon$.

If the ω_i's are BSNE-moduli and if p is a real fixed point of T (and thus a common fixed point of the T_i's as $\widehat{F}(T)\subseteq\bigcap_{i=1}^n\widehat{F}(T_i)$, see [37]), then it is clear that the second term involving φ can be dropped. ☐

Proof of Theorem 4.19. If

$$\|x-Tx\| < \xi\left(w\omega(\rho(\varepsilon,\widehat{b}),b),\widehat{b}\right),$$

then we get $D_f(p_0,x)-D_f(p_0,Tx)<w\omega(\rho(\varepsilon,\widehat{b}),b)$. Fix $k=1,\dots,N$. Then

$$\begin{aligned}
D_f(p_0,Tx) &\leq w_kD_f(p_0,T_kx)+\sum_{i\neq k}w_iD_f(p_0,T_ix)\\
&\leq w_kD_f(p_0,T_kx)+\sum_{i\neq k}w_iD_f(p_0,x)\\
&= w_kD_f(p_0,T_kx)+(1-w_k)D_f(p_0,x)\\
&\leq w_k(D_f(p_0,T_kx)-D_f(p_0,x))+D_f(p_0,x)
\end{aligned}$$

and thus in particular

$$w_k\left(D_f(p_0,x)-D_f(p_0,T_kx)\right) \leq D_f(p_0,x)-D_f(p_0,Tx) < w\omega(\rho(\varepsilon,\widehat{b}),b)$$

which implies $D_f(p_0,x)-D_f(p_0,T_kx)<\omega(\rho(\varepsilon,\widehat{b}),b)$. As ω is a BSNE-modulus for T_k, we get $D_f(T_kx,x)<\rho(\varepsilon,\widehat{b})$ which yields $\|x-T_kx\|<\varepsilon$. ☐

Proof of Theorem 4.20. Let x, p be given with $\|x\|, \|p\| \leq b$, $\|p - Tp\| < \widehat{\omega}(\varepsilon, b)$ as well as $D_f(p,x) - D_f(p,Tx) < \widehat{\omega}(\varepsilon, b)$. Then in particular

$$\|p - Tp\| < \xi(w\omega(\rho(\min\{\omega(\varepsilon', b), \omega'(w\omega(\varepsilon', b), b)\}, \widehat{b}), b), \widehat{b})$$

and by Theorem 4.19, we have $\|p - T_k p\| < \min\{\omega(\varepsilon', b), \omega'(w\omega(\varepsilon', b), b)\}$ for any k with $w_k \geq w$.

We further have

$$D_f(p, Tx) \leq \sum_{i=1}^{N} w_i D_f(p, T_i x)$$

and, therefore,

$$\sum_{i=1}^{N} w_i \left(D_f(p,x) - D_f(p, T_i x)\right) \leq D_f(p,x) - D_f(p,Tx) < \widehat{\omega}(\varepsilon, b)$$

which implies

$$\begin{aligned} w_k \left(D_f(p,x) - D_f(p, T_k x)\right) &< \widehat{\omega}(\varepsilon, b) + \sum_{i \neq k} w_i \left(D_f(p, T_i x) - D_f(p,x)\right) \\ &< \widehat{\omega}(\varepsilon, b) + (1 - w_k) w\omega(\varepsilon', b) \\ &\leq w^2 \omega(\varepsilon', b) + (1 - w) w\omega(\varepsilon', b) \\ &= w\omega(\varepsilon', b) \end{aligned}$$

and thus $D_f(p,x) - D_f(p, T_k x) < \omega(\varepsilon', b)$ for any k with $w_k \geq w$. As ω is a strong BSNE-modulus for T_k, this gives $D_f(T_k x, x) < \varepsilon'$ for any such k. Thus in particular $\|x - T_k x\| < \omega^{\nabla f}(\varepsilon/4\widehat{b}, \widehat{b})$ which yields $\|\nabla f x - \nabla f T_k x\| < \varepsilon/4\widehat{b}$. As we have

$$\nabla f T x - \nabla f x = \sum_{i=1}^{N} w_i (\nabla f T_i x - \nabla f x)$$

the above yields

$$\begin{aligned} \|\nabla f T x - \nabla f x\| &\leq \sum_{i=1}^{N} w_i \|\nabla f x - \nabla f T_i x\| \\ &= \sum_{i : w_i \geq w} w_i \|\nabla f x - \nabla f T_i x\| + \sum_{i : w_i < w} w_i \|\nabla f x - \nabla f T_i x\| \\ &< \sum_{i : w_i \geq w} w_i \varepsilon / 4\widehat{b} + \sum_{i : w_i < w} w_i 2C(\widehat{b}) \\ &< \varepsilon / 4\widehat{b} \sum_{i : w_i \geq w} w_i + \sum_{i : w_i < w} w 2C(\widehat{b}) \\ &\leq \varepsilon / 2\widehat{b}. \end{aligned}$$

Now using Lemma 2.9, we have

$$D_f(Tx, x) + D_f(x, Tx) = \langle Tx - x, \nabla f(Tx) - \nabla f(x) \rangle \leq \|\nabla f T x - \nabla f x\| \|Tx - x\| \leq \|\nabla f T x - \nabla f x\| 2\widehat{b} < \varepsilon$$

which in particular yields $D_f(Tx, x) < \varepsilon$.

It is immediate to see that if ω is just a (not necessarily strong) BSNE-modulus and p is a fixed point of T, that $w\omega(\varepsilon', b)$ suffices. □

Proof of Lemma 6.5. (1) Assume the contrary, i.e. that there are ε and ψ such that for any $p \in X$ and any $\delta \geq \varphi(\varepsilon, \psi)$ with $\|p\| \leq b$ and $\|Tp - p\| < \psi(\delta)$:

$$\exists q \in X \left(\|q\| \leq b \wedge \|Tq - q\| < \delta \wedge D_f(p, u) - \varepsilon \geq D_f(q, u)\right).$$

Let $r = \lceil (b+1)/\varepsilon \rceil$ and pick $q_0 = p_0$. Then clearly $\|q_0\| \leq b$ and

$$\|Tq_0 - q_0\| < \psi^{(r+1)}(1) = \psi(\psi^{(r)}(1)).$$

By definition, we have $\psi^{(r)}(1) \geq \varphi(\varepsilon, \psi)$ so that there exists a q_1 with $\|q_1\| \leq b$ and $\|Tq_1 - q_1\| < \psi^{(r)}(1)$ as well as $D_f(q_0, u) - \varepsilon \geq D_f(q_1, u)$. Iterating this up to r yields a q_r such that

$$0 > D_f(q_0, u) - (b+1) \geq D_f(q_0, u) - \lceil (b+1)/\varepsilon \rceil \varepsilon = D_f(q_0, u) - r\varepsilon \geq D_f(q_r, u)$$

which is a contradiction.

(2) Using (the proof of) (1), let $p \in X$ and $\delta = {\psi'}^{(i)}(1)$ for $i \leq \lceil (b+1)/\varepsilon' \rceil$ be such that $\|p\| \leq b$, $\|Tp - p\| < \psi'(\delta)$ and

$$\forall q \in X \left(\|q\| \leq b \wedge \|Tq - q\| < \delta \to D_f(p,u) < D_f(q,u) + \varepsilon' \right).$$

Let $\delta' = \omega'(\rho(\delta, \max\{b, E(b)\}), b)$. Then at first

$$\|Tp - p\| < \psi'(\delta) = \min\{\psi(\delta'), \delta'\} \leq \psi(\delta').$$

Now let q be such that $\|q\| \leq b$ and $\|Tq - q\| < \delta'$. If $q = p$, the claim is trivial. So suppose $q \neq p$. Then we can now reason along the lines of [14]: write $p(\alpha)$ for $p + \alpha(q-p)$. Using Lemma 6.4, as $\|Tp - p\| < \delta'$, we have $\|Tp(\alpha) - p(\alpha)\| < \delta$. Therefore, for any $\alpha \in [0,1]$: $D_f(p,u) < D_f(p(\alpha),u) + \varepsilon'$. Now, using the fact that D_f is convex and differentiable in its left argument with

$$[D_f(\cdot, x)]'(y) = \nabla f(y) - \nabla f(x),$$

we get

$$\frac{|D_f(p(\alpha),u) - D_f(p,u) - \langle \alpha(q-p), \nabla f p - \nabla f u\rangle|}{\|\alpha(q-p)\|} < \varepsilon/4b$$

if $\|\alpha(q-p)\| < \Delta(\varepsilon/4b, b)$, i.e. in particular if $\alpha < \Delta(\varepsilon/4b, b)/2b$. Thus in particular

$$\frac{\langle q-p, \nabla f u - \nabla f p\rangle}{\|q-p\|} = \frac{-\langle \alpha(q-p), \nabla f p - \nabla f u\rangle}{\|\alpha(q-p)\|} < \frac{D_f(p,u) - D_f(p(\alpha),u)}{\|\alpha(q-p)\|} + \varepsilon/4b$$

which implies

$$\langle q-p, \nabla f u - \nabla f p\rangle < \frac{D_f(p,u) - D_f(p(\alpha),u)}{\alpha} + (\varepsilon/4b)\,\|q-p\| \leq \frac{\varepsilon'}{\alpha} + \varepsilon/2$$

for any $\alpha < \min\left\{\frac{\Delta(\varepsilon/4b,b)}{2b}, 1\right\}$. In particular, for $\alpha = \min\left\{\frac{\Delta(\varepsilon/4b,b)}{4b}, 1/2\right\}$, we get

$$\langle q-p, \nabla f u - \nabla f p\rangle < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

□

Proof of Lemma 7.4. Using the three-point-identity for D_f, we get

$$\begin{aligned} D_f(x,y) &= D_f(x, \mathrm{Res}^f_r y) + D_f(\mathrm{Res}^f_r y, y) + \langle x - \mathrm{Res}^f_r y, \nabla f \mathrm{Res}^f_r y - \nabla f y\rangle \\ &= D_f(x, \mathrm{Res}^f_r y) + D_f(\mathrm{Res}^f_r y, y) + r\langle x - \mathrm{Res}^f_r y, -A^f_r y\rangle. \end{aligned}$$

Using the monotonicity of A, we further derive that

$$\begin{aligned} \langle x - \mathrm{Res}^f_r y, -A^f_r y\rangle &= \langle x - \mathrm{Res}^f_s x, -A^f_r y\rangle + \langle \mathrm{Res}^f_s x - \mathrm{Res}^f_r y, -A^f_r y\rangle \\ &= \langle x - \mathrm{Res}^f_s x, -A^f_r y\rangle + \langle \mathrm{Res}^f_s x - \mathrm{Res}^f_r y, s^{-1}(\nabla f x - \nabla f \mathrm{Res}^f_s x) - A^f_r y\rangle \\ &\quad + \langle \mathrm{Res}^f_s x - \mathrm{Res}^f_r y, -s^{-1}(\nabla f x - \nabla f \mathrm{Res}^f_s x)\rangle \\ &\geq \langle x - \mathrm{Res}^f_s x, -A^f_r y\rangle + s^{-1}\langle \mathrm{Res}^f_s x - \mathrm{Res}^f_r y, \nabla f \mathrm{Res}^f_s x - \nabla f x\rangle \\ &\geq -\|x - \mathrm{Res}^f_s x\|\|A^f_r y\| - s^{-1}\|\mathrm{Res}^f_s x - \mathrm{Res}^f_r y\|\|\nabla f \mathrm{Res}^f_s x - \nabla f x\| \\ &\geq -\|x - \mathrm{Res}^f_s x\| r^{-1}\left(\|\mathrm{Res}^f_r y\| + \|y\|\right) - s^{-1}\left(\|\mathrm{Res}^f_s x\| + \|\mathrm{Res}^f_r y\|\right)\|\nabla f \mathrm{Res}^f_s x - \nabla f x\|. \end{aligned}$$

Combined with the above, this yields

$$\begin{aligned} D_f(x,y) &\geq D_f(x, \mathrm{Res}^f_r y) + D_f(\mathrm{Res}^f_r y, y) - \|x - \mathrm{Res}^f_s x\|\left(\|\mathrm{Res}^f_r y\| + \|y\|\right) \\ &\quad - rs^{-1}\left(\|\mathrm{Res}^f_s x\| + \|\mathrm{Res}^f_r y\|\right)\|\nabla f \mathrm{Res}^f_s x - \nabla f x\| \\ &\geq D_f(x, \mathrm{Res}^f_r y) + D_f(\mathrm{Res}^f_r y, y) - 2b\|x - \mathrm{Res}^f_s x\| - rs^{-1}2b\|\nabla f \mathrm{Res}^f_s x - \nabla f x\| \\ &\geq D_f(x, \mathrm{Res}^f_r y) + D_f(\mathrm{Res}^f_r y, y) - E\left(\|x - \mathrm{Res}^f_s x\| + \|\nabla f \mathrm{Res}^f_s x - \nabla f x\|\right) \end{aligned}$$

and therefore, for x such that

$$\left\|x - \mathrm{Res}^f_s x\right\| < \omega^{\nabla f}\left(\frac{\varepsilon}{2E}, b\right),$$

we get that $D_f(x,y) > D_f(x, \mathrm{Res}^f_r y) + D_f(\mathrm{Res}^f_r y, y) - \varepsilon$ which is the claim. □