Mathematisches
Forschungsinstitut
Oberwolfach

Member of Leibniz Leibniz Association

Oberwolfach Preprints

Number 2024-04

PEDRO PINTO
NICHOLAS PISCHKE

On Dykstra's Algorithm with Bregman Projections

April 2024

Oberwolfach Preprints (OWP)

The MFO publishes the Oberwolfach Preprints (OWP) as a series which mainly contains research results related to a longer stay in Oberwolfach, as a documentation of the research work done at the MFO. In particular, this concerns the Oberwolfach Research Fellows program (and the former Research in Pairs program) and the Oberwolfach Leibniz Fellows (OWLF), but this can also include an Oberwolfach Lecture, for example.

All information about the publication process can be found at
https://www.mfo.de/scientific-programme/publications/owp

All published Oberwolfach Preprints can be found at https://publications.mfo.de

ISSN 1864-7596

Mathematisches Forschungsinstitut Oberwolfach gGmbH (MFO)
Schwarzwaldstrasse 9-11
77709 Oberwolfach-Walke
Germany
https://www.mfo.de

DOI 10.14760/OWP-2024-04

ON DYKSTRA'S ALGORITHM WITH BREGMAN PROJECTIONS

PEDRO PINTO AND NICHOLAS PISCHKE

Department of Mathematics, Technische Universität Darmstadt,
Schlossgartenstraße 7, 64289 Darmstadt, Germany,
E-mail: {pinto,pischke}@mathematik.tu-darmstadt.de

Abstract. We provide quantitative results on the asymptotic behavior of Dykstra's algorithm with Bregman projections, a combination of the well-known Dykstra's algorithm and the method of cyclic Bregman projections, designed to find best approximations and solve the convex feasibility problem in a non-Hilbertian setting. The result we provide arise through the lens of proof mining, a program in mathematical logic which extracts computational information from non-effective proofs. Concretely, we provide a highly uniform and computable rate of metastability of low complexity and, moreover, we also specify general circumstances in which one can obtain full and effective rates of convergence. As a byproduct of our quantitative analysis, we also for the first time establish the strong convergence of Dykstra's method with Bregman projections in infinite dimensional (reflexive) Banach spaces.

Keywords: Convex Feasibility; Best Approximation; Projection Methods; Dykstra's Algorithm; Bregman Projections; Legendre Functions; Rates of Convergence; Metastability; Proof Mining
MSC2020 Classification: 41A65; 90C25; 03F10; 41A29; 65J05

1. Introduction

Let X, if not stated otherwise, be a real reflexive Banach space with norm $\|\cdot\|$ and let $C_1, \dots, C_N \subseteq X$ be finitely many closed and convex sets such that

$$C := \bigcap_{i=1}^{N} C_i \neq \emptyset.$$

Finding a point $c \in C$ is referred to as the convex feasibility problem and, in order to solve this problem, a wide range of methods have been developed throughout the course of convex analysis (see e.g. [1, 5, 14, 15]).

In this paper, we are concerned with the particular case of Dykstra's algorithm which, say over $\mathbb{R}^d$, takes the following form: set $q_{-(N-1)} = \cdots = q_0 := 0$, define $C_n := C_{n \bmod N}$ and let P_n be the orthogonal projection onto C_n. Simultaneously define

$$x_n := P_n(x_{n-1} + q_{n-N}) \text{ and } q_n := x_{n-1} + q_{n-N} - x_n.$$

Then (x_n) converges to $P_C x_0$. More concretely, Dykstra [16] first proved the convergence of this iteration in the case where all the sets are closed convex cones and Boyle and Dykstra [7] later extended this convergence result to arbitrary closed and convex sets as well as to infinite dimensional Hilbert spaces, where now (x_n) converges strongly to $P_C x_0$.

Date: April 9, 2024.

Funding: This research was supported through the program "Oberwolfach Leibniz Fellows" by the Mathematisches Forschungsinstitut Oberwolfach in 2024. Further, the authors were supported by the DFG Project KO 1737/6-2.

Beyond Hilbert spaces, complications start to arise already at the level of the orthogonal projections which may not exist in general Banach spaces. One specific way of introducing a meaningful class of projections in Banach spaces was considered in the pivotal work of Bregman [8], where projections along a certain distance relative to a convex function are considered instead.

Concretely, let $f : X \to (-\infty, \infty]$ be a proper, lower-semicontinuous and convex function which is Fréchet differentiable on $\mathrm{intdom} f \neq \emptyset$, i.e. for any $x \in \mathrm{intdom} f$, there exists some $\nabla f(x) \in X^*$ such that

$$\lim_{\|h\| \to 0} \frac{|f(x+h) - f(x) - \langle h, \nabla f(x) \rangle|}{\|h\|} = 0.$$

Relative to f, we can now define the corresponding Bregman distance $D_f : \mathrm{dom} f \times \mathrm{intdom} f \to [0, +\infty)$ via

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

If f is Legendre as defined in [3], then there naturally exists a unique minimizer of $D_f(\cdot, y)$ over $S \cap \mathrm{intdom} f$ for a given closed and convex set $S \subseteq X$ with $S \cap \mathrm{intdom} f \neq \emptyset$. We call this unique element the Bregman projection of y relative to f onto S and denote it by $P^f_S y$ (see [3] for more details on this).

By substituting the orthogonal projections by these Bregman projections in the previous method of Dykstra in Hilbert spaces, one obtains Dykstra's algorithm with Bregman projections. This method was first proposed in the seminal work by Censor and Reich [12] where the sets C_i are all halfspaces and then extended to general closed convex nonempty sets by Bauschke and Lewis [6] in finite dimensional spaces under suitable additional conditions on the function f which we will shortly discuss in the following:

The first additional assumption made, besides that that f is Legendre, is that f is co-finite, i.e. that $\mathrm{dom} f^* = X^*$ where $f^* : X^* \to (-\infty, +\infty]$ is the conjugate function to f defined by

$$f^*(x^*) := \sup_{x \in X} \left(\langle x, x^* \rangle - f(x) \right).$$

Further, as shown in [3], f being Legendre is equivalent to the function f^* being Legendre and hence implies that f^* is Gateaux differentiable on $\mathrm{intdom} f^* \neq \emptyset$.

The second assumption made is that f is also very strictly convex, i.e. that f is twice continuously differentiable on $\mathrm{intdom} f \neq \emptyset$ and that its second derivative $\nabla^2 f(x)$ is positive definite for any $x \in \mathrm{intdom} f$.

Under these assumptions, Bauschke and Lewis obtained the following result:

Theorem 1.1 ([6])**.** *Let X be finite dimensional and let f be closed, convex, proper, Legendre, co-finite and very strictly convex. Let $C_1, \ldots, C_N$ be finitely many closed and convex sets with $C \cap \mathrm{intdom} f \neq \emptyset$ where $C := \bigcap_i C_i$. Set $q_{-(N-1)} = \cdots = q_0 := 0$ as well as $C_n := C_{n \bmod N}$ and let P^f_n be the Bregman projection onto C_n relative to f. Given $x_0 \in \mathrm{intdom} f$, simultaneously define*

$$x_n := P^f_n \nabla f^*(\nabla f(x_{n-1}) + q_{n-N}) \text{ and } q_n := \nabla f(x_{n-1}) + q_{n-N} - \nabla f(x_n).$$

Then (x_n) converges to $P^f_C x_0$.

In this paper, we provide a quantitative version of this result in the form a computable and highly uniform rate of metastability in the sense of Tao [32, 33]. By results from computability theory due to Specker [31] (see also [25]), such a rate of metastability is in general the best one can hope for in the context of many methods from nonlinear optimization when aiming for computational information. In particular, a computable rate of convergence in general does not exist. However, under suitable regularity assumptions, such computable rates do exist and we provide an abstract construction for this at the end of the paper.

Further, our quantitative version is valid in general normed spaces where a suitable corresponding function f exists. By "forgetting" about the quantitative aspects and using that having a rate of metastability is equivalent to the convergence of a sequence, we are therefore able to establish an "infinitary" convergence result of Dykstra's algorithm with Bregman projections which extends the above also to infinite dimensional Banach spaces. The proof of the quantitative result that we give here arises as a generalization of the recent quantitative analysis of Dykstra's method in Hilbert spaces by the first author in [26][1] in combination with recent work of the second author and Kohlenbach [29] on quantitative results for iterations in the context of Bregman distances and Legendre functions. Both of these works were obtained, similar to the results presented here, by methods from proof mining, a program in mathematical logic that aims at the extraction of computational information from prima facie non-computational proofs (see [19, 21] and in particular the recent [28] where the underlying logical methods have been extended to also cover the dual of a Banach space, gradients of convex functions and Bregman distances, etc.). However, as usual for results from the proof mining program, this whole paper requires no logical background.

2. (Quantitative) assumptions and lemmas

The construction of the rate of metastability that we give in the following section relies on certain (quantitative) assumptions on the function f together with quantitative reformulations of the central lemmas used in [6] which we discuss in this section. These preliminary results are taken, or adapted, from either [29] or [26], where the former recently provided the first general quantitative treatment of methods related to Bregman distances and Legendre functions from the perspective of proof mining, and the latter analyzed the computational content of Dykstra's method in Hilbert spaces from that perspective.

Throughout, we will assume that $f : X \to (-\infty, +\infty]$ is a proper, convex and co-finite Legendre function that is Fréchet differentiable on $\mathrm{intdom} f \neq \emptyset$ with a gradient ∇f.

The main assumption on f used in Theorem 1.1 that warrants for a quantitative treatment is that of very strict convexity. As shown in [6], this assumption in particular entails the existence of certain moduli regarding the associated Bregman distance and the gradient:

Proposition 2.1 ([6]). *Let X be finite dimensional and f be very strictly convex. Then, for any convex and compact set $K \subseteq \mathrm{intdom} f$, there exist reals $0 < \theta$ and $\Theta < +\infty$ such that for every $x, y \in K$:*

(1) $D_f(x, y) \geq \theta \|x - y\|^2$,
(2) $\|\nabla f(x) - \nabla f(y)\| \leq \Theta \|x - y\|$.

In the following, instead of considering very strict convexity, we will immediately assume that there exist two monotone non-decreasing functions $\theta, \Theta : (0, \infty) \to (0, \infty)$ such that

(C1) $D_f(x, y) \geq \theta(b) \|x - y\|^2$,
(C2) $\|\nabla f(x) - \nabla f(y)\| \leq \Theta(b) \|x - y\|$,

[1] In particular, one can obtain rates of similar complexity as derived in [26] by instantiating the results given here with the function $f = \|\cdot\|^2 / 2$ and the respective moduli in a given (pre-)Hilbert space.

for any $b > 0$ and $x, y \in \overline{B}_b(0) \cap \mathrm{intdom} f$. That only the assumption of such moduli, witnessing the conclusion of Proposition 2.1, suffices for carrying out the proof of Theorem 1.1 was already mentioned in [6] and we will find that the same holds also in the infinite-dimensional case.

The existence of such moduli θ, Θ in particular guarantees that f is uniformly continuous on bounded sets and, even further, that it is sequentially consistent, i.e. that

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

for any two bounded sequences $(x_n), (y_n) \subseteq \mathrm{intdom} f$. In particular, by the results from [10], this implies that f is totally convex on $\mathrm{intdom} f$.

As (essentially) shown in [29], sequential consistency is equivalent to the existence of a so-called modulus of consistency for f, i.e. a function $\rho : (0, \infty)^2 \to (0, \infty)$ such that

$$\forall \varepsilon > 0 \ \forall b > 0 \ \forall x, y \in \overline{B}_b(0) \cap \mathrm{intdom} f \ (D_f(x, y) \leq \rho(\varepsilon, b) \to \|x - y\| \leq \varepsilon),$$

and such a function can easily be constructed from θ by just setting $\rho(\varepsilon, b) = \theta(b)\varepsilon^2$.

We will also always assume that the convex feasibility problem is consistent on $\mathrm{intdom} f$, i.e. that $C \cap \mathrm{intdom} f \neq \emptyset$. From a quantitative perspective, we will in the following fix some data relating to this condition:

$$\text{(C3)} \qquad \begin{cases} p \in C \cap \mathrm{intdom} f \text{ and } x_0 \in \mathrm{intdom} f \\ \text{as well as } b \in \mathbb{N} \setminus \{0\} \text{ such that } b \geq D_f(p, x_0). \end{cases}$$

Lastly, next to θ and Θ, we will assume the existence of a function $o : (0, \infty) \to (0, \infty)$ satisfying

$$\text{(C4)} \qquad \forall y \in \mathrm{intdom} f \ \forall \alpha > 0 \, (D_f(p, y) \leq \alpha \to \|y\| \leq o(\alpha))$$

with the p fixed in (C3). Without loss of generality, we assume that $o(\alpha) \geq \alpha$ and that o is monotone non-decreasing.

In Theorem 1.1, besides guaranteeing that the main iteration is well-defined, the assumption that f is co-finite is mainly used to derive that the level sets

$$L(x, \alpha) = \{y \in \mathrm{intdom} f \mid D_f(x, y) \leq \alpha\}$$

are bounded for every $\alpha > 0$ and $x \in \mathrm{intdom} f$. This boundedness of the level sets is a common requirement on Bregman distances (e.g. featuring in the list of conditions regarding so-called Bregman functions exhibited in [10, 17]). In the context of finite-dimensional spaces, as shown in [2, Theorem 3.7], if f is essentially strictly convex and $\mathrm{dom} f^*$ is open (which in particular is true when f is Legendre and co-finite), then $D_f(x, \cdot)$ is coercive for any $x \in \mathrm{intdom} f$ and thus $L(x, \alpha)$ is bounded. In reflexive Banach spaces, as shown in [3, Lemma 7.3], the boundedness of all these level sets is in particular implied by f being supercoercive.

However, for our quantitative result, it will suffice to assume the above (C4) which is just a quantitative rendering for this property for $x = p$.

These assumptions in turn entail a further crucial quantitative property on the associated Bregman distance D_f:

Lemma 2.2. *The distance D_f is reverse consistent as defined in* [29]*, i.e.*

$$\forall r, \varepsilon > 0 \ \forall x, y \in \overline{B}_r(0) \cap \mathrm{intdom} f \ (\|x - y\| \leq P(\varepsilon, r) \to D_f(x, y) \leq \varepsilon),$$

with a modulus $P(\varepsilon, r) := \sqrt{\varepsilon / \Theta(r)}$.

Proof. Let $x, y \in \overline{B}_r(0) \cap \mathrm{intdom} f$. As $\nabla f(x)$ is a subgradient of f at x, we get

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

for any y, and so

$$\begin{aligned} D_f(x,y) &= f(x) - f(y) - \langle x - y, \nabla f(y) \rangle \\ &\leq \langle x - y, \nabla f(x) - \nabla f(y) \rangle \\ &\leq \Theta(r) \|x - y\|^2 \end{aligned}$$

and this immediately yields the claim with the given modulus. □

We now move to further properties of the Bregman distance and the associated Bregman projections. The first main properties of the Bregman distance are the so-called *3- and 4-point identities.*

Proposition 2.3 (folklore, see e.g. [13]). *For any $x, y, z, w \in \mathrm{intdom} f$, it holds that*

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

as well as

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

Crucial for Dykstra's method and the accompanying convergence proof in Hilbert spaces is a characterization of the projection using the inner product. An analogous result also holds for Bregman projections and it will similarly play an important role in this paper.

Proposition 2.4 ([11]). *Let S be a closed convex subset of X such that $S \cap \mathrm{intdom} f \neq \emptyset$. Consider $y \in \mathrm{intdom} f$. Then the Bregman projection $P^f_S y$ is characterized by*

$$P^f_S(y) \in S \cap \mathrm{intdom} f \quad \text{and} \quad \forall x \in S \left(\langle x - P^f_S(y), \nabla f(y) - \nabla f(P^f_S(y)) \rangle \leq 0 \right).$$

Moreover, it holds that

$$\forall x \in S \cap \mathrm{dom}(f) \left(D_f\left(P^f_S(y), y\right) \leq D_f(x,y) - D_f\left(x, P^f_S(y)\right) \right).$$

The following quantitative projection result is adapted from [29] (which in turn is adapted from [18, 20]). From here on out, we write $[n;m] = [n,m] \cap \mathbb{N}$ for $n, m \in \mathbb{N}$.

Proposition 2.5 (essentially [18, 20, 29]). *Let $r > 0$ and $u \in \mathrm{intdom} f$ as well as $q \in C \cap \mathrm{intdom} f$ be such that $r \geq \|u\|, \|q\|, D_f(q,u)$. Then for any $\varepsilon > 0$ and function $\delta : (0,\infty) \to (0,\infty)$, there exists $\eta \geq \beta(r,\varepsilon,\delta)$ and $x \in \overline{B}_r(0) \cap \mathrm{intdom} f$ such that $\bigwedge_{j=1}^m \|x - P^f_j(x)\| \leq \delta(\eta)$ and*

$$\forall y \in \overline{B}_r(0) \cap \mathrm{intdom} f \left(\bigwedge_{j=1}^m \|y - P^f_j(y)\| \leq \eta \ \to \ D_f(x,u) \leq D_f(y,u) + \varepsilon \right),$$

where $\beta(r,\varepsilon,\delta) := \min\{\delta^{(i)}(1) \mid i \leq \lceil r/\varepsilon \rceil\}$.

Proof. Let $\varepsilon > 0$ and a function δ be given. Assume towards a contradiction that for all $\eta \geq \beta(r,\varepsilon,\delta)$ and $x \in \overline{B}_r(0) \cap \mathrm{intdom} f$ such that $\|x - P^f_j(x)\| \leq \delta(\eta)$ for all $j \in [1;m]$, we have

$$(\dagger) \qquad \exists y \in \overline{B}_r(0) \cap \mathrm{intdom} f \left(\bigwedge_{j=1}^m \|y - P^f_j(y)\| \leq \eta \ \wedge \ D_f(y,u) < D_f(x,u) - \varepsilon \right).$$

We define a sequence $y_0, \ldots, y_R$, for $R := \lceil r/\varepsilon \rceil$ in the following way. We take $y_0 := q$. Then, $y_0 \in \overline{B}_r(0) \cap \mathrm{intdom} f$ and clearly $\|y_0 - P^f_j(y_0)\| \leq \delta^{(R)}(1)$ for all $j \in [1;m]$. Assume that for

$i \leq R-1$, we have $y_i \in \overline{B}_r(0) \cap \mathrm{intdom} f$ such that $\|y_i - P_j^f(y_i)\| \leq \delta^{(R-i)}(1)$ for all $j \in [1;m]$. Since $\delta^{(R-i-1)}(1) \geq \beta(r,\varepsilon,\delta)$, by (†) there exists some $y \in \overline{B}_r(0) \cap \mathrm{intdom} f$ such that

$$\bigwedge_{j=1}^{m} \|y - P_j^f(y)\| \leq \delta^{(R-i-1)}(1) \text{ and } D_f(y,u) < D_f(y_i,u) - \varepsilon,$$

and we take y_{i+1} to be one such y. Hence, by construction, we have

$$\forall i \leq R-1 \left(D_f(y_{i+1},u) < D_f(y_i,u) - \varepsilon\right),$$

which in turn, using (C3), entails the contradiction that

$$D_f(y_R,u) < D_f(q,u) - R\varepsilon \leq D_f(q,u) - r \leq 0. \qquad \square$$

We require the following two technical lemmas from [26].

Lemma 2.6 ([26]). *Let $(a_n) \in \ell^1_+(\mathbb{N})$ and consider $B \in \mathbb{N}$ such that $\sum a_n \leq B$. Then,*

$$\forall \varepsilon > 0 \ \forall g \in \mathbb{N}^{\mathbb{N}} \ \exists n \leq \Psi(B,\varepsilon,g) \ \forall i \in [n; n+g(n)] \left(a_i \leq \varepsilon\right),$$

where $\Psi(B,\varepsilon,g) := \breve{g}^{(R)}(0)$ with $\breve{g}(p) := p + g(p) + 1$ and $R := \lfloor \frac{B}{\varepsilon} \rfloor$.

Lemma 2.7 ([26]). *Let $(a_n) \in \ell^2_+(\mathbb{N})$ and consider $B \in \mathbb{N}$ such that $\sum a_n^2 \leq B$. For all $n \in \mathbb{N}$, set $s_n := \sum_{k=0}^{n} a_k$, and let $m \geq 2$ be given. Then,*

$$\liminf s_n(s_n - s_{n-m-1}) = 0 \textit{ with } \liminf\textit{-rate } \phi_B(m,\varepsilon,N) := \left\lfloor e^{\left(\frac{(m+1)B}{\varepsilon}\right)^2} \right\rfloor \cdot (N+1),$$

i.e.

$$\forall \varepsilon > 0 \ \forall N \in \mathbb{N} \ \exists n \in [N; N + \phi_B(m,k,N)] \left(s_n(s_n - s_{n-m-1}) \leq \varepsilon\right).$$

Note that the above lemma is a quantitative version of [5, Lemma 30.6].

3. Main results

We recall here the definition of Dykstra's method. Let $C_1, \cdots, C_m$ be $m \geq 2$ closed convex subsets of X such that $C \cap \mathrm{intdom} f \neq \emptyset$ where $C := \bigcap_{j=1}^m C_j$. For any $n \geq 1$, let $j_n := [n-1]+1$ with $[r] := r \mod m$. For $n \geq 1$, we consider $C_n := C_{j_n}$ and denote with P_n^f the Bregman projection onto C_n. Dykstra's algorithm with Bregman projections is defined by the following equations:

$$\text{(DB)} \qquad \begin{cases} x_0 \in \mathrm{intdom} f, \\ q_{-m+1} = \cdots = q_0 := 0. \end{cases} \qquad \forall n \geq 1 \begin{cases} x_n := P_n^f \nabla f^*\left(\nabla f(x_{n-1}) + q_{n-m}\right), \\ q_n := \nabla f(x_{n-1}) + q_{n-m} - \nabla f(x_n). \end{cases}$$

For the remaining sections, unless stated otherwise, we consider (x_n) to be the iteration generated by (DB), and that the conditions (C1) – (C4) hold.

3.1. Fundamental identities and bounds.

We start by stating some facts that follow easily from the definition of the algorithm, all of which are proven (explicitly or in passing) in [6].

Lemma 3.1 ([6]). *For all $n \geq 1$:*

(i) $\nabla f(x_{n-1}) - \nabla f(x_n) = q_n - q_{n-m}$,
(ii) $\nabla f(x_0) - \nabla f(x_n) = \sum_{k=n-m+1}^{n} q_k$,
(iii) $x_n \in C_n \cap \mathrm{intdom} f$ *and* $\forall z \in C_n \left(\langle x_n - z, q_n \rangle \geq 0\right)$,
(iv) $\langle x_n - x_{n+m}, q_n \rangle \geq 0$.

Further, for all $n \in \mathbb{N}$:

$$\text{(v)} \qquad \sum_{k=n-m+1}^{n} \|q_k\| \leq \sum_{k=0}^{n-1} \|\nabla f(x_k) - \nabla f(x_{k+1})\|.$$

Lastly, for all $z \in \operatorname{intdom} f$ and $i, n \in \mathbb{N}$ with $i \geq n$ and arbitrary $x_{-(m-1)}, \cdots, x_{-1} \in \operatorname{intdom} f$, we have

$$\begin{aligned} D_f(z, x_n) =& D_f(z, x_i) + \sum_{k=n}^{i-1} \left(D_f(x_{k+1}, x_k) + \langle x_{k-m+1} - x_{k+1}, q_{k-m+1} \rangle \right) \\ \text{(vi)} \qquad & + \sum_{k=i-m+1}^{i} \langle x_k - z, q_k \rangle - \sum_{k=n-m+1}^{n} \langle x_k - z, q_k \rangle, \end{aligned}$$

and in particular

$$\text{(vii)} \qquad D_f(z, x_i) \leq D_f(z, x_n) + \sum_{k=n-m+1}^{n} \langle x_k - z, q_k \rangle - \sum_{k=i-m+1}^{i} \langle x_k - z, q_k \rangle.$$

The first quantitative result is the following lemma which provides a bound on the Bregman distances of the sequence.

Lemma 3.2. *For all $n \in \mathbb{N}$:*

$$D_f(p, x_n), \ \sum_{k=0}^{n} D_f(x_{k+1}, x_k) \leq b.$$

Proof. By (C3), we have $b \geq D_f(p, x_0)$. The first bound is now immediate from Lemma 3.1.(vii) with $z = p$ and $n = 0$ using Lemma 3.1.(iii) and the fact that $\sum_{k=-(m-1)}^{0} \langle x_k - p, q_k \rangle = 0$. The second bound similarly follows from Lemma 3.1.(vi) (using also Lemma 3.1.(iv)). ☐

The following lemma is then immediate:

Lemma 3.3. *Define $\theta_0 := \theta(o(b))$ and $\Theta_0 := \Theta(o(b))$. Then, for all $k \in \mathbb{N}$:*

(1) $\|x_k\| \leq o(b)$,
(2) $D_f(x_{k+1}, x_k) \geq \theta_0 \|x_{k+1} - x_k\|^2$,
(3) $\|\nabla f(x_{k+1}) - \nabla f(x_k)\| \leq \Theta_0 \|x_{k+1} - x_k\|$.

Using Lemma 2.7, we derive the following lim inf-rate (akin to Proposition 3.5 in [26]).

Proposition 3.4. *We have $\liminf_n \sum_{k=n-m+1}^{n} |\langle x_k - x_n, q_k \rangle| = 0$, and moreover, for all $\varepsilon > 0$ and $N \in \mathbb{N}$*

$$\exists n \in [N; N + \Phi(b, m, \varepsilon, N)] \left(\sum_{k=n-m+1}^{n} |\langle x_k - x_n, q_k \rangle| \leq \varepsilon \right),$$

where $\Phi(b, m, \varepsilon, N) := \phi_{b/\theta_0}(m, \varepsilon/\Theta_0, N)$, with ϕ as defined in Lemma 2.7.

Proof. Let $\varepsilon > 0$ and $N \in \mathbb{N}$ be given. As we have seen $\sum D_f(x_{k+1}, x_k) \leq b$ and so, by the previous lemma, $\sum \|x_{k+1} - x_k\|^2 \leq b/\theta_0$. Hence, we can apply Lemma 2.7 (with $a_n = \|x_n - x_{n+1}\|$ and $B = b/\theta_0$) to conclude that there exists $n \in [N; N + \Phi(b, m, \varepsilon, N)]$ such that

$$\left(\sum_{k=n-m+1}^{n} \|x_k - x_{k+1}\| \right) \cdot \left(\sum_{k=0}^{n} \|x_k - x_{k+1}\| \right) = (s_n - s_{n-m}) s_n \leq \frac{\varepsilon}{\Theta_0}.$$

By triangle inequality, for all $k \in [n-m+1;n]$,

$$\|x_k - x_n\| \leq \sum_{\ell=k}^{n-1} \|x_\ell - x_{\ell+1}\| \leq \sum_{\ell=n-m+1}^{n-1} \|x_\ell - x_{\ell+1}\|,$$

and thus using the definition of the dual norm and Lemma 3.1.(v), we get

$$\begin{aligned}
\sum_{k=n-m+1}^{n} |\langle x_k - x_n, q_k\rangle| &\leq \sum_{k=n-m+1}^{n} \|x_k - x_n\| \cdot \|q_k\| \\
&\leq \left(\sum_{k=n-m+1}^{n} \|q_k\|\right)\left(\sum_{\ell=n-m+1}^{n-1} \|x_\ell - x_{\ell+1}\|\right) \\
&\leq \left(\sum_{k=0}^{n} \|\nabla f(x_k) - \nabla f(x_{k+1})\|\right)\left(\sum_{k=n-m+1}^{n} \|x_k - x_{k+1}\|\right) \\
&\leq \Theta_0 \left(\sum_{k=0}^{n} \|x_k - x_{k+1}\|\right)\left(\sum_{k=n-m+1}^{n} \|x_k - x_{k+1}\|\right) \leq \varepsilon.
\end{aligned}$$

□

Note that the above function Φ is monotone non-decreasing in N.

3.2. **Asymptotic regularity.** Here we discuss the asymptotic regularity of the sequence (x_n). Since we can obtain the bound $\sum_{k=0}^{n} \|x_{k+1} - x_k\|^2 \leq b/\theta_0$ using Lemma 3.2 and Lemma 3.3, by Lemma 2.6 we have the following result:

Proposition 3.5. *We have* $\lim \|x_k - x_{k+1}\| = 0$ *and, moreover,*

$$\forall \varepsilon > 0\ \forall g \in \mathbb{N}^{\mathbb{N}}\ \exists n \leq \Psi(b/\theta_0, \varepsilon^2, g)\ \forall k \in [n; n+g(n)]\ (\|x_k - x_{k+1}\| \leq \varepsilon)\,,$$

where Ψ *is as defined in Lemma 2.6.*

Therefore, the sequence (x_n) is asymptotically regular in the sense of [9]. Furthermore, the sequence (x_n) is asymptotically regular with respect to the individual Bregman projection maps in the following sense:

Proposition 3.6. *For all* $j \in [1;m]$, *we have* $\lim \|x_n - P^f_j(x_n)\| = 0$ *and, moreover,*

$$\forall \varepsilon > 0\ \forall g \in \mathbb{N}^{\mathbb{N}}\ \exists n \leq \alpha(b,m,\varepsilon,g)\ \forall k \in [n; n+g(n)]\left(\bigwedge_{j=1}^{m} \|x_k - P^f_j(x_k)\| \leq \varepsilon\right)$$

where $\alpha(b,m,\varepsilon,g) := \Psi\left(b/\theta_0, \frac{\varepsilon^2\theta_0}{(m-1)^2\Theta_0}, \hat{g}_m\right)$, *with* $\hat{g}_m(n) = g(n) + m - 2$ *and with* Ψ *as defined in Lemma 2.6.*

Proof. For given $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$, by Proposition 3.5 there is $n \leq \alpha(b,m,\varepsilon,g)$ such that

$$(\ddagger) \qquad \forall k \in [n; n+g(n)+m-2]\left(\|x_k - x_{k+1}\| \leq \sqrt{\frac{\varepsilon^2\theta_0}{(m-1)^2\Theta_0}} = \frac{P(\varepsilon^2\theta_0, o(b))}{m-1}\right)$$

for P as in Lemma 2.2. Consider $k \in [n; n+g(n)]$. By the definition, we have $x_k \in C_{j_k}$ with $j_k := [k-1]+1$. Then as $[k; k+m-2] \subset [n; n+g(n)+m-2]$ by ($\ddagger$), we have

$$\|x_{k+i} - x_k\| \leq \sum_{\ell=k}^{k+i-1} \|x_\ell - x_{\ell+1}\| \leq \sum_{\ell=k}^{k+m-2} \|x_\ell - x_{\ell+1}\| \leq P(\varepsilon^2\theta_0, o(b))$$

for any $i \in [0; m-1]$. Since $\|x_{k+i}\|, \|x_k\| \leq o(b)$, Lemma 2.2 for the function P gives

$$D_f(x_{k+i}, x_k) \leq \varepsilon^2 \theta_0.$$

Hence, by the fact that $x_{k+i} \in C_{j_k+i}$ and using the definition of the projection $P^f_{j_k+i}$, we derive

$$D_f(P^f_{j_k+i}(x_k), x_k) \leq D_f(x_{k+i}, x_k) \leq \varepsilon^2 \theta_0.$$

By Lemma 3.3, we get $\|P^f_{j_k+i}(x_k) - x_k\| \leq \varepsilon$, and the conclusion now follows from observing that for any $k \in \mathbb{N}$, $\{P^f_{j_k+i} \mid i \in [0; m-1]\} = \{P^f_1, \cdots, P^f_m\}$. □

Note that the above function α is monotone non-increasing in ε.

3.3. **Metastability and strong convergence.** The following is the fundamental combinatorial lemma of the convergence analysis of Dykstra's method with Bregman distances presented here (and in that way is modeled after Proposition 3.10 in [26]).

Proposition 3.7. *Let $\varepsilon > 0$ and a function $\Delta : \mathbb{N} \to (0, \infty)$ be given. Then,*

$$\exists n \leq \gamma(b, m, \varepsilon, \Delta)\ \exists x \in \overline{B}_{o(b)}(0) \cap \mathrm{intdom} f$$
$$\left(\bigwedge_{j=1}^{m} \|x - P^f_j(x)\| \leq \Delta(n) \wedge D_f(x, x_n) \leq \varepsilon \wedge \sum_{k=n-m+1}^{n} \langle x_k - x_n, q_k \rangle \leq \varepsilon \right),$$

where $\gamma(b, m, \varepsilon, \Delta) := \overline{\alpha}(\overline{\beta}) + \Phi_\varepsilon\big(\overline{\alpha}(\overline{\beta})\big)$ with

$$\overline{\beta} := \beta\left(o(b), \frac{\varepsilon}{3}, \delta\right),$$

$$\delta(\eta) := \min\left\{ \frac{\varepsilon}{6o(b)\Theta_0\big(\overline{\alpha}(\eta) + \Phi_\varepsilon(\overline{\alpha}(\eta))\big)}, \widetilde{\Delta}\Big(\overline{\alpha}(\eta) + \Phi_\varepsilon\big(\overline{\alpha}(\eta)\big)\Big) \right\},$$

$$\overline{\alpha}(\eta) := \alpha(b, m, \eta, \Phi_\varepsilon),$$

$$\Phi_\varepsilon\big(N\big) := \Phi\Big(b, m, \frac{\varepsilon}{3}, N\Big),$$

$$\widetilde{\Delta}(k) := \min\{\Delta(k') \mid k' \leq k\},$$

and α, β, Φ are as in Propositions 3.6, 2.5 and 3.4, respectively.

Proof. By Proposition 2.5 with $u = x_0$ and $q = p$, noting that $o(b) \geq b$ and $o(b) \geq \|p\|$ since $b > 0$, there are $\eta_0 \geq \overline{\beta}$ and $x \in \overline{B}_{o(b)}(0) \cap \mathrm{intdom} f$ such that $\|x - P^f_j(x)\| \leq \delta(\eta_0)$ for all $j \in [1; m]$, and

$$(*) \qquad \forall y \in \overline{B}_{o(b)}(0) \cap \mathrm{intdom} f \left(\bigwedge_{j=1}^{m} \|y - P^f_j(y)\| \leq \eta_0 \to D_f(x, x_0) \leq D_f(y, x_0) + \frac{\varepsilon}{3} \right).$$

Considering Proposition 3.6 with $\varepsilon = \eta_0$ and $g = \Phi_\varepsilon$, we obtain

$$\exists N_0 \leq \overline{\alpha}(\eta_0)\ \forall i \in [N_0; N_0 + \Phi_\varepsilon\big(N_0\big)] \left(\bigwedge_{j=1}^{m} \|x_i - P^f_j(x_i)\| \leq \eta_0 \right).$$

Since $(x_n) \subseteq \overline{B}_{o(b)}(0) \cap \mathrm{intdom} f$, by $(*)$ we have

$$\forall i \in [N_0; N_0 + \Phi_\varepsilon(N_0)] \left(D_f(x, x_0) \leq D_f(x_i, x_0) + \frac{\varepsilon}{3} \right).$$

On the other hand, from Proposition 3.4 (with $\varepsilon = \varepsilon/3$ and $N = N_0$) and the definition of the function Φ_ε, there exists $n_0 \in [N_0; N_0 + \Phi_\varepsilon(N_0)]$ such that

$$(**) \qquad \sum_{k=n_0-m+1}^{n_0} \langle x_k - x_{n_0}, q_k \rangle \leq \frac{\varepsilon}{3}.$$

At this point, we remark that $n_0 \leq \gamma(b, m, \varepsilon, \Delta)$. Indeed, as α and Φ are monotone and using the fact that $\eta_0 \geq \overline{\beta}$:

$$n_0 \leq N_0 + \Phi_\varepsilon(N_0) \leq \overline{\alpha}(\eta_0) + \Phi_\varepsilon\big(\overline{\alpha}(\eta_0)\big) \leq \overline{\alpha}(\overline{\beta}) + \Phi_\varepsilon\big(\overline{\alpha}(\overline{\beta})\big) = \gamma(b, m, \varepsilon, \Delta).$$

The definition of the functions δ and $\widetilde{\Delta}$ then entail

$$\delta(\eta_0) \leq \widetilde{\Delta}\big(\overline{\alpha}(\eta_0) + \Phi_\varepsilon(\overline{\alpha}(\eta_0))\big) \leq \Delta(n_0).$$

It remains to verify that $D_f(x, x_{n_0}) \leq \varepsilon$. Note that the definition of δ also entails

$$\delta(\eta_0) \leq \frac{\varepsilon}{6o(b)\Theta_0(\overline{\alpha}(\eta_0) + \Phi_\varepsilon(\overline{\alpha}(\eta_0)))} \leq \frac{\varepsilon}{6o(b)\Theta_0(N_0 + \Phi_\varepsilon(N_0))} \leq \frac{\varepsilon}{6o(b)\Theta_0 n_0}.$$

Thus, by the 3-point identity and Lemma 3.1, we get

$$\begin{aligned}
D_f(x, x_{n_0}) &= \langle x - x_{n_0}, \nabla f(x_0) - \nabla f(x_{n_0}) \rangle + D_f(x, x_0) - D_f(x_{n_0}, x_0) \\
&\leq \langle x - x_{n_0}, \nabla f(x_0) - \nabla f(x_{n_0}) \rangle + \frac{\varepsilon}{3} \\
&= \sum_{k=n_0-m+1}^{n_0} \langle x - x_{n_0}, q_k \rangle + \frac{\varepsilon}{3} \\
&= \sum_{k=n_0-m+1}^{n_0} \langle x - x_k, q_k \rangle + \sum_{k=n_0-m+1}^{n_0} \langle x_k - x_{n_0}, q_k \rangle + \frac{\varepsilon}{3} \\
&\leq \sum_{k=n_0-m+1}^{n_0} \langle x - P_k^f(x), q_k \rangle + \sum_{k=n_0-m+1}^{n_0} \langle \underbrace{P_k^f(x)}_{\in C_k} - x_k, q_k \rangle + \frac{2\varepsilon}{3} \\
&\leq \sum_{k=n_0-m+1}^{n_0} \langle x - P_k^f(x), q_k \rangle + \frac{2\varepsilon}{3} \\
&\leq \sum_{k=n_0-m+1}^{n_0} \|x - P_k^f(x)\| \cdot \|q_k\| + \frac{2\varepsilon}{3} \\
&\leq \delta(\eta_0) \sum_{k=0}^{n_0-1} \|\nabla f(x_k) - \nabla f(x_{k+1})\| + \frac{2\varepsilon}{3} \\
&\leq \delta(\eta_0) \sum_{k=0}^{n_0-1} \Theta_0 \|x_k - x_{k+1}\| + \frac{2\varepsilon}{3} \\
&\leq \delta(\eta_0) \cdot n_0 \cdot 2o(b)\Theta_0 + \frac{2\varepsilon}{3} \leq \frac{\varepsilon}{6o(b)\Theta_0 n_0} \cdot n_0 \cdot 2o(b)\Theta_0 + \frac{2\varepsilon}{3} = \varepsilon,
\end{aligned}$$

which concludes the proof. □

Remark 3.8. *Note that in the above proposition, the use of the Bregman distance actually revealed that the analogous argument from* [26] *(for Proposition 3.10 therein) in the context of*

Hilbert spaces can be slightly optimized. Namely, we can change the definition of the function β in [26, Proposition 2.5] *to*

$$\beta(r,\varepsilon,\delta) := \min\{\delta^{(i)}(1) \mid i \le \lceil r^2/\varepsilon\rceil\},$$

and now instead conclude that there exist $\eta \ge \beta(r,\varepsilon,\delta)$ and $x \in \overline{B}_r(p)$ such that for all $j \in [1;m]$, it holds that $\|x - P_j(x)\| \le \delta(\eta)$ and

$$\forall y \in \overline{B}_r(p) \left(\bigwedge_{j=1}^{m} \|y - P_j(y)\| \le \eta \to \|x-u\|^2 \le \|y-u\|^2 + \varepsilon \right).$$

Then [26, Proposition 3.10] *is adapted to this new β instead, and the argument there proceeds similarly, now instead making use of the identity*

$$\|x - x_{n_0}\|^2 = 2\langle x - x_{n_0}, x_0 - x_{n_0}\rangle + \|x - x_0\|^2 - \|x_{n_0} - x_0\|^2,$$

which is analogous to the use of the 3-point identity in the above Proposition 3.7.

We are now ready to prove our central result.

Theorem 3.9. *Let f be a proper, convex and co-finite Legendre function which is Fréchet differentiable on $\operatorname{intdom} f \neq \emptyset$ with gradient ∇f. Let $C_1, \cdots, C_m$ be $m \ge 2$ convex sets such that $C \cap \operatorname{intdom} f \neq \emptyset$ for $C := \bigcap_{j=1}^m C_j$. Assume that the conditions (C1) – (C4) hold. Then, the sequence (x_n) generated by* (DB) *is a Cauchy sequence and, moreover, for all $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$,*

$$\exists n \le \Omega(b,m,\varepsilon,g)\ \forall i,j \in [n; n+g(n)]\,(\|x_i - x_j\| \le \varepsilon),$$

where $\Omega(b,m,\varepsilon,g) := \gamma(b,m,\tilde{\varepsilon},\Delta_{\varepsilon,g})$ with

$$\tilde{\varepsilon} := \min\left\{\rho\left(\frac{\overline{\rho}}{12o(b)\Theta_0}, o(b)\right), \frac{\overline{\rho}}{6}\right\}, \quad \overline{\rho} := \rho\left(\frac{\varepsilon}{2}, o(b)\right)$$

$$\Delta_{\varepsilon,g}(k) := \frac{\overline{\rho}}{6o(b)\Theta_0 \cdot \max\{k + g(k), 1\}},$$

and γ is defined as in Proposition 3.7.

Proof. Let $\varepsilon > 0$ and a function $g : \mathbb{N} \to \mathbb{N}$ be given. Using Proposition 3.7, there exist $n_0 \le \Omega(b,m,\varepsilon,g)$ and $x \in \overline{B}_{o(b)}(0) \cap \operatorname{intdom} f$ such that

(*a*) $\bigwedge_{j=1}^m \|x - P_j^f(x)\| \le \Delta_{\varepsilon,g}(n_0)$,
(*b*) $D_f(x, x_{n_0}) \le \tilde{\varepsilon} \le \min\left\{\rho\left(\frac{\overline{\rho}}{12o(b)\Theta_0}, o(b)\right), \frac{\overline{\rho}}{3}\right\}$,
(*c*) $\sum_{k=n_0-m+1}^{n_0} \langle x_k - x_{n_0}, q_k\rangle \le \tilde{\varepsilon} \le \frac{\overline{\rho}}{6}$.

In order to verify that the result holds for such an n_0, we consider $i \in [n_0; n_0 + g(n_0)]$. We assume that $g(n_0) \ge 1$, and thus $\max\{n_0 + g(n_0), 1\} = n_0 + g(n_0)$, otherwise the result trivially holds. Since $i \ge n_0$, by Lemma 3.1.(vii) and using (*b*), we have

$$\begin{aligned} D_f(x, x_i) &\le D_f(x, x_{n_0}) + \sum_{k=n_0-m+1}^{n_0} \langle x_k - x, q_k\rangle - \sum_{k=i-m+1}^{i} \langle x_k - x, q_k\rangle \\ &\le \frac{\overline{\rho}}{3} + \underbrace{\sum_{k=n_0-m+1}^{n_0} \langle x_k - x, q_k\rangle}_{t_1} + \underbrace{\sum_{k=i-m+1}^{i} \langle x - x_k, q_k\rangle}_{t_2}. \end{aligned}$$

Using (b), (c) and Lemma 3.1.(ii), we get

$$\begin{aligned} t_1 &= \sum_{k=n_0-m+1}^{n_0} \langle x_k - x_{n_0}, q_k\rangle + \sum_{k=n_0-m+1}^{n_0} \langle x_{n_0} - x, q_k\rangle \\ &\leq \frac{\overline{\rho}}{6} + \langle x_{n_0} - x, \nabla f(x_0) - \nabla f(x_{n_0})\rangle \\ &\leq \frac{\overline{\rho}}{6} + \|x_{n_0} - x\| \cdot 2o(b)\Theta_0 \leq \frac{\overline{\rho}}{6} + \frac{\overline{\rho}}{12o(b)\Theta_0} \cdot 2o(b)\Theta_0 = \frac{\overline{\rho}}{3}, \end{aligned}$$

and, using (a) as well as Lemma 3.1.(v), we get

$$\begin{aligned} t_2 &= \sum_{k=i-m+1}^{i} \langle x - P^f_k(x), q_k\rangle + \underbrace{\sum_{k=i-m+1}^{i} \langle \underbrace{P^f_k(x)}_{\in C_k} - x_k, q_k\rangle}_{\leq 0} \\ &\leq \sum_{k=i-m+1}^{i} \langle x - P^f_k(x), q_k\rangle \leq \sum_{k=i-m+1}^{i} \|x - P^f_k(x)\|\|q_k\| \\ &\leq \Delta_{\varepsilon,g}(n_0) \sum_{k=i-m+1}^{i} \|q_k\| \leq \Delta_{\varepsilon,g}(n_0) \cdot \sum_{k=0}^{i-1} \|\nabla f(x_k) - \nabla f(x_{k+1})\| \\ &\leq \Delta_{\varepsilon,g}(n_0) \cdot \sum_{k=0}^{i-1} \Theta_0 \|x_k - x_{k+1}\| \leq \Delta_{\varepsilon,g}(n_0) \cdot 2o(b)\Theta_0 \cdot i \\ &= \frac{\overline{\rho}}{6o(b)\Theta_0(n_0 + g(n_0))} \cdot 2o(b)\Theta_0 \cdot i \leq \frac{\overline{\rho}}{3}, \end{aligned}$$

using in the last inequality the fact that $i \leq n_0 + g(n_0)$. Overall, we conclude that

$$D_f(x, x_i) \leq \frac{\overline{\rho}}{3} + \frac{\overline{\rho}}{3} + \frac{\overline{\rho}}{3} = \rho\left(\frac{\varepsilon}{2}, o(b)\right),$$

and, by the properties of ρ, we get $\|x_i - x\| \leq \varepsilon/2$, which entails the result by triangle inequality. $\square$

In particular, from this result we obtain rates of metastability for Dykstra's method in Hilbert spaces by instantiating the above result (with all its moduli) to the special case of $f = \|\cdot\|^2/2$. These rates are of a similar complexity to those obtained in [26].

As a byproduct of our analysis, we then also obtain the following "infinitary" convergence result. Also, this result in particular entails Theorem 1.1.

Theorem 3.10. *Let X be a reflexive Banach space and f be a proper, convex and co-finite Legendre function that is Fréchet differentiable on* $\mathrm{intdom} f \neq \emptyset$ *with gradient ∇f and assume that the conditions (C1) – (C2) hold. Assume further that all level sets $L(x, \alpha)$ for $x \in$* $\mathrm{intdom} f$ *and $\alpha > 0$ are bounded. Let $C_1, \cdots, C_m$ be $m \geq 2$ closed and convex subsets of X such that $C \cap \mathrm{intdom} f \neq \emptyset$ for $C := \bigcap_{j=1}^m C_j$. Then, the sequence (x_n) defined by* (DB) *is norm convergent towards $P^f_C(x_0)$.*

Proof. By assumption, all level sets $L(x, \alpha)$ for $x \in \mathrm{intdom} f$ and $\alpha > 0$ are bounded. Therefore, in particular, there exists an o satisfying (C4) and since we have assumed $C \cap \mathrm{intdom} f \neq \emptyset$, (C3) is easily satisfied with corresponding p and b. Therefore, Theorem 3.9 entails the strong convergence of (x_n). Indeed, as the sequence (x_n) satisfies the metastability property it is a Cauchy sequence, and by completeness it converges to some point of the space, say $z = \lim x_n$.

Now, as f is totally convex on $\mathrm{intdom} f$ (as discussed in Section 2), Proposition 4.3 of [30] implies the continuity of the projection maps P_j^f on $\mathrm{intdom} f$. Thus, by Proposition 3.6, we conclude that z must be a common fixed point for all projections, and so $z \in C$. It only remains to argue that the limit point is actually the feasible point D_f-closest to x_0, i.e. $P_C^f(x_0)$.

For this, let o' be a modulus of boundedness for the level sets $L(z, \alpha)$, i.e.

$$\forall y \in \mathrm{intdom} f\ \forall \alpha > 0\, (D_f(z, y) \leq \alpha \to \|y\| \leq o'(\alpha))$$

and let $b' \geq D_f(z, x_0)$. With $b > 0$ and p as in (C3), for an arbitrary $\varepsilon > 0$ we define $\overline{\rho} := \rho(\varepsilon/2, \max\{o(b), o'(b')\})$. Since $z = \lim x_n$, consider $N_0 \in \mathbb{N}$ such that

$$\forall n \geq N_0 \left(\|x_n - z\| \leq \min\left\{ P\left(\frac{\overline{\rho}}{4}, o(b)\right), \frac{\overline{\rho}}{8o(b)\Theta_0}, \frac{\varepsilon}{2} \right\} \right).$$

As per Proposition 3.4, we may consider some $n_0 \geq N_0$ such that

$$\sum_{k=n_0-m+1}^{n_0} \langle x_k - x_{n_0}, q_k \rangle \leq \frac{\overline{\rho}}{2}.$$

First note that, since $(x_n) \subseteq \overline{B}_{o(b)}(0) \cap \mathrm{intdom} f$, we also have $\|z\| \leq o(b)$. Moreover, by Proposition 2.4, we have

$$D_f(z, P_C^f(x_0)) \leq D_f(z, x_0) - D_f(P_C^f(x_0), x_0) \leq D_f(z, x_0) \leq b',$$

and hence by assumption on o', we have $\|P_C^f(x_0)\| \leq o'(b')$. Since $\|x_{n_0} - z\| \leq P(\overline{\rho}/4, o(b))$, by Lemma 2.2 it follows that $D_f(z, x_{n_0}) \leq \overline{\rho}/4$. As $z \in C$, by the definition of P_C^f, we have

$$D_f(P_C^f(x_0), x_0) - D_f(x_{n_0}, x_0) \leq D_f(z, x_0) - D_f(x_{n_0}, x_0),$$

and using the 3-point identity, we get

$$\begin{aligned} D_f(z, x_0) - D_f(x_{n_0}, x_0) &= D_f(z, x_{n_0}) + \langle x_{n_0} - z, \nabla f(x_0) - \nabla f(x_{n_0}) \rangle \\ &\leq D_f(z, x_{n_0}) + \|x_{n_0} - z\| \cdot 2o(b)\Theta_0 \\ &\leq \frac{\overline{\rho}}{4} + \frac{2o(b)\Theta_0\overline{\rho}}{8o(b)\Theta_0} = \frac{\overline{\rho}}{2}. \end{aligned}$$

Using again the 3-point identity, we now obtain

$$\begin{aligned} D_f(P_C^f(x_0), x_{n_0}) &= D_f(P_C^f(x_0), x_0) - D_f(x_{n_0}, x_0) + \langle P_C^f(x_0) - x_{n_0}, \nabla f(x_0) - \nabla f(x_{n_0}) \rangle \\ &\leq \frac{\overline{\rho}}{2} + \langle P_C^f(x_0) - x_{n_0}, \nabla f(x_0) - \nabla f(x_{n_0}) \rangle \\ &= \frac{\overline{\rho}}{2} + \sum_{k=n_0-m+1}^{n_0} \langle P_C^f(x_0) - x_{n_0}, q_k \rangle \qquad \text{by Lemma 3.1.(ii)} \\ &= \frac{\overline{\rho}}{2} + \sum_{k=n_0-m+1}^{n_0} \langle P_C^f(x_0) - x_k, q_k \rangle + \sum_{k=n_0-m+1}^{n_0} \langle x_k - x_{n_0}, q_k \rangle \\ &\leq \frac{\overline{\rho}}{2} + \frac{\overline{\rho}}{2} = \overline{\rho}, \end{aligned}$$

which, by the properties of ρ together with $\|x_{n_0} - z\| \leq \varepsilon/2$, entails $\|P_C^f(x_0) - z\| \leq \varepsilon$ and so, as ε is arbitrary, $z = P_C^f(x_0)$. □

3.4. **A rate of convergence.** Lastly, we study full rates of convergence for Dykstra's algorithm where, as discussed in the introduction, we provide an abstract construction of such rates under an additional quantitative regularity assumption in the form of a certain modulus adapted from [26].

Definition 3.11 (essentially [26]). *We call a function $\mu : (0,\infty)^2 \to (0,\infty)$ satisfying for all $\varepsilon > 0$ and $r > 0$:*

$$(\star) \quad \forall x \in \overline{B}_r(0) \cap \mathrm{intdom} f \left(\bigwedge_{j=1}^{m} \|x - P^f_j(x)\| \leq \mu_r(\varepsilon) \to \exists z \in C \cap \mathrm{intdom} f \left(\|x - z\| \leq \varepsilon\right) \right),$$

a modulus of regularity *for the sets $C_1, \cdots, C_m$.*

We refer to [26] for a discussion on when and how such moduli can be obtained and how they further relate to rates of convergence (in the context of Hilbert spaces).

In the case where a modulus of regularity is available, we can actually give highly uniform rates of convergence and the construction of such a rate is contained in the following theorem, modeled after Theorem 4.2 from [26] which provided such a result in the context of Hilbert spaces.

Theorem 3.12. *Let μ be a modulus of regularity for the sets $C_1, \cdots, C_m$. Then,*

$$\forall \varepsilon > 0 \; \forall i, j \geq \Theta(b, m, \varepsilon) \left(\|x_i - x_j\| \leq \varepsilon\right),$$

where $\Theta(b, m, \varepsilon) := \alpha(b, m, \mu_{o(b)}(\tilde{\varepsilon}), \Phi_\varepsilon) + \Phi_\varepsilon(\alpha(b, m, \mu_{o(b)}(\tilde{\varepsilon}), \Phi_\varepsilon))$ with

$$\tilde{\varepsilon} := \frac{\rho(\varepsilon/2, o(b))}{4o(b)\Theta_0} \text{ and } \Phi_\varepsilon(N) := \Phi\left(b, m, \frac{\rho(\varepsilon/2, o(b))}{2}, N\right),$$

and α, Φ are as in Propositions 3.6 and 3.4, respectively.

Proof. By Proposition 3.6, there is $N_0 \leq \alpha(b, m, \mu_{o(b)}(\tilde{\varepsilon}), \Phi_\varepsilon)$ such that

$$\forall n \in [N_0; N_0 + \Phi_\varepsilon(N_0)] \left(\bigwedge_{j=1}^{m} \|x_n - P^f_j(x_n)\| \leq \mu_{o(b)}(\tilde{\varepsilon}) \right).$$

Since $(x_n) \subseteq \overline{B}_{o(b)}(0) \cap \mathrm{intdom} f$, by the assumption $(\star)$ on μ it follows that

$$(\circ) \qquad \forall n \in [N_0; N_0 + \Phi_\varepsilon(N_0)] \; \exists z \in C \cap \mathrm{intdom} f \left(\|x_n - z\| \leq \tilde{\varepsilon} = \frac{\rho(\varepsilon/2, o(b))}{4o(b)\Theta_0} \right).$$

Applying Proposition 3.4 (with $\varepsilon = \frac{\rho(\varepsilon/2, o(b))}{2}$ and $N = N_0$), we have

$$\exists n_0 \in [N_0; N_0 + \Phi_\varepsilon(N_0)] \left(\sum_{k=n_0-m+1}^{n_0} \langle x_k - x_{n_0}, q_k \rangle \leq \frac{\rho(\varepsilon/2, o(b))}{2} \right).$$

By $(\circ)$, let $\hat{z} \in C$ be such that $\|\hat{z} - x_{n_0}\| \leq \frac{\rho(\varepsilon/2, o(b))}{4o(b)\Theta_0}$. Thus, for any $i \geq n_0$

$$\begin{aligned}
\sum_{k=i-m+1}^{i} \langle x_k - x_{n_0}, q_k \rangle &= \underbrace{\sum_{k=i-m+1}^{i} \langle x_k - \hat{z}, q_k \rangle}_{\geq 0,\ \text{by Lemma 3.1.(iii)}} + \sum_{k=i-m+1}^{i} \langle \hat{z} - x_{n_0}, q_k \rangle \\
&\geq \langle \hat{z} - x_{n_0}, \sum_{k=i-m+1}^{i} q_k \rangle \\
&= \langle \hat{z} - x_{n_0}, \nabla f(x_0) - \nabla f(x_i) \rangle \qquad \text{by Lemma 3.1.(ii)} \\
&\geq -\|\hat{z} - x_{n_0}\| \|\nabla f(x_0) - \nabla f(x_i)\| \\
&\geq -\|\hat{z} - x_{n_0}\| \Theta_0 \, \|x_0 - x_i\| \\
&\geq -\frac{2o(b)\Theta_0 \rho(\varepsilon/2, o(b))}{4o(b)\Theta_0} = -\frac{\rho(\varepsilon/2, o(b))}{2}.
\end{aligned}$$

Now by Lemma 3.1.(vii) (with $n = n_0$ and $z = x_{n_0}$),

$$D_f(x_{n_0}, x_i) \leq \sum_{k=n_0-m+1}^{n_0} \langle x_k - x_{n_0}, q_k \rangle - \sum_{k=i-m+1}^{i} \langle x_k - x_{n_0}, q_k \rangle \leq \rho(\varepsilon/2, o(b)),$$

and by the properties of ρ, we get $\|x_i - x_{n_0}\| \leq \varepsilon/2$ which entails the result by triangle inequality. □

Remark 3.13. *In the context of Dykstra's method in Hilbert spaces, it was recently recognized in a work of the first author together with Kohlenbach* [24] *that the fact that such moduli of regularity suffice to construct a rate of convergence is due to Dykstra's method being Fejér monotone in a certain generalized sense. If the generalized Fejér monotonicity from* [24] *would be adapted to incorporate more general distance functions than metrics, e.g. along the line of the recent work* [27] *of the second author (which simultaneously extends works on Fejér monotone sequences in metric spaces as considered in* [22, 23] *and Bregman monotone methods as considered in* [4]*), then a respective combined result should be attainable which provides a similar Fejér-type perspective also for Dykstra's method with Bregman projections as considered here.*

References

[1] H.H. Bauschke and J.M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Review*, 38(3):367–426, 1996.

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

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

[4] H.H. Bauschke, J.M. Borwein, and P.L. Combettes. Bregman monotone optimization algorithms. *SIAM Journal on Control and Optimization*, 42(2):596–636, 2003.

[5] H.H. Bauschke and P.L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces.* CMS Books in Mathematics. Springer, 2nd edition, 2017.

[6] H.H. Bauschke and A.S. Lewis. Dykstras algorithm with Bregman projections: A convergence proof. *Optimization*, 48(4):409–427, 2000.

[7] J.P. Boyle and R.L. Dykstra. A method for finding projections onto the intersection of convex sets in Hilbert spaces. In R.L. Dykstra, T. Robertson, and F.T. Wright, editors, *Advances in order restricted statistical inference (Iowa City, Iowa, 1985)*, volume 37 of *Lecture Notes in Statistics*, pages 28–47. Springer, Berlin, 1986.

[8] L.M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *U.S.S.R. Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967.
[9] F.E. Browder and W.V. Petryshyn. The solution by iteration of nonlinear functional equations in Banach spaces. *Bulletin of the American Mathematical Society*, 72:571–575, 1966.
[10] 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.
[11] D. Butnariu and E. Resmerita. Bregman distances, totally convex functions and a method for solving operator equations in Banach spaces. *Abstract and Applied Analysis*, 2006. Art. ID 84919, 39pp.
[12] Y. Censor and S. Reich. The Dykstra algorithm with Bregman projections. *Communications in Applied Analysis*, 2(3):407–419, 1998.
[13] G. Chen and M. Teboulle. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM Journal on Optimization*, 3(3):538–543, 1993.
[14] P.L. Combettes. The convex feasibility problem in image recovery. *Advances in imaging and electron physics*, 95:155–270, 1996.
[15] P.L. Combettes. Hilbertian convex feasibility problem: Convergence of projection methods. *Applied Mathematics and Optimization*, 35:311–330, 1997.
[16] R.L. Dykstra. An algorithm for restricted least squares regression. *Journal of the American Statistical Association*, 78(384):837–842, 1983.
[17] J. Eckstein. Nonlinear Proximal Point Algorithms Using Bregman Functions, with Applications to Convex Programming. *Mathematics of Operations Research*, 18(1):202–226, 1993.
[18] F. Ferreira, L. Leuştean, and P. Pinto. On the removal of weak compactness arguments in proof mining. *Advances in Mathematics*, 354:106728, 55pp., 2019.
[19] U. Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer-Verlag Berlin Heidelberg, 2008.
[20] U. Kohlenbach. On quantitative versions of theorems due to F.E. Browder and R. Wittmann. *Advances in Mathematics*, 226:2764–2795, 2011.
[21] U. Kohlenbach. Proof-theoretic Methods in Nonlinear Analysis. In B. Sirakov, P.N. de Souza, and M. Viana, editors, *Proceedings of the International Congress of Mathematicians 2018*, volume 2, pages 62–82. World Scientific, 2019.
[22] U. Kohlenbach, L. Leuştean, and A. Nicolae. Quantitative results on Fejér monotone sequences. *Communications in Contemporary Mathematics*, 20(2), 2018. 1750015, 42pp.
[23] U. Kohlenbach, G. López-Acedo, and A. Nicolae. Moduli of regularity and rates of convergence for Fejér monotone sequences. *Israel Journal of Mathematics*, 232:261–297, 2019.
[24] U. Kohlenbach and P. Pinto. Fejér monotone sequences revisited. 2023. 13pp., submitted manuscript available at `https://www2.mathematik.tu-darmstadt.de/~pinto/`.
[25] E. Neumann. Computational problems in metric fixed point theory and their Weihrauch degrees. *Logical Methods in Computer Science*, 11(4):20, 44pp., 2015.
[26] P. Pinto. On the finitary content of Dykstra's cyclic projections algorithm. 2023. 25pp., submitted manuscript available at `https://arxiv.org/abs/2306.09791`.
[27] N. Pischke. Generalized Fejér monotone sequences and their finitary content. 2023. 40pp., submitted manuscript available at `https://arxiv.org/abs/2312.01852`.
[28] N. Pischke. Proof mining for the dual of a Banach space with extensions for uniformly Fréchet differentiable functions. 2023. 30pp., submitted manuscript available at `https://sites.google.com/view/nicholaspischke/notes-and-papers`.
[29] N. Pischke and U. Kohlenbach. Effective rates for iterations involving Bregman strongly nonexpansive operators. 2024. 42pp., submitted manuscript available at `https://sites.google.com/view/nicholaspischke/notes-and-papers`.
[30] E. Resmerita. On total convexity, Bregman projection and stability in Banach spaces. *Journal of Convex Analysis*, 11(1):1–16, 2004.
[31] E. Specker. Nicht konstruktiv beweisbare Sätze der Analysis. *The Journal of Symbolic Logic*, 14(3):145–158, 1949.
[32] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory and Dynamical Systems*, 28(2):657–688, 2008.

[33] 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.