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-orthogonal setting. The results 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, which in particular contain the case of polyhedra in Euclidean spaces. 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 suitable infinite dimensional spaces.

Keywords: Convex Feasibility; Dykstra's Algorithm; Bregman Projections; 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 $C_1, \ldots, C_m \subseteq X$ be finitely many closed and convex sets such that

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

Finding a point $c \in C$ is referred to as the convex feasibility problem and it has been the subject of extensive research due to its wide-ranging applications in applied mathematics, including statistics, partial differential equations, signal restoration, and computed tomography. In order to solve this problem, a wide range of methods have been developed throughout the course of the development of (modern) convex analysis and we refer to [6, 19, 20] as well as the surveys [2, 22], and the references therein, for further discussions.

The most well-known algorithm for solving the convex feasibility problem is the method of alternating projections, introduced by von Neumann [53], who established its strong convergence to the optimal solution (i.e. the point in C closest to the initial guess) in the context of two closed vector subspaces of a Hilbert space. Halperin [29] later extended this result to the case of an arbitrary finite number of closed vector subspaces. In the more general setting, where the sets C_i are merely closed convex sets, Bregman [10] proved that, in the context of Hilbert spaces, the method of alternating projections converges weakly to a point in the intersection.

Date: June 24, 2025.
Funding: This research was supported through the program "Oberwolfach Leibniz Fellows" by the Mathematisches Forschungsinstitut Oberwolfach in 2024. Further, both authors were supported by the DFG Project KO 1737/6-2 and the first author was also supported by the DFG Project PI 2070/1-1.

In this broader context, as shown by Hundal [32] (see also [48]), this result is indeed the best one can expect. Moreover, even in finite dimensional settings, where weak convergence gets upgraded to strong convergence, there are simple examples where the method of alternating projections fails to find the optimal solution and instead just converges to some other solution point.

In this paper, we are concerned with a more sophisticated approach which was originally introduced by Dykstra [25] and which takes the following form: Over the Euclidean space $\mathbb{R}^d$, set $q_{-(m-1)} = \cdots = q_0 := 0$, define $C_n := C_{n \bmod m}$ and let P_n be the metric projection onto C_n. Fix $x_0 \in \mathbb{R}^d$ and simultaneously define

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

for $n \geq 1$. Then (x_n) converges to $P_C x_0$. More concretely, Dykstra [25] first proved the convergence of this iteration in the case where all the sets are closed convex cones and Boyle and Dykstra [9] 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 $P_C x_0$. The method was rediscovered by Han [30] in 1988 and, both in Han's work [30] and by Iusem and De Pierro [33], it was also observed that in the polyhedral case, Dykstra's algorithm becomes Hildreth's algorithm [31] (see also [21] and [40] where this method is further extended). Dykstra's method in inconsistent cases was studied in particular in the work of Iusem and De Pierro [33] over Euclidean spaces and, for two sets, in infinite dimensional Hilbert spaces by Bauschke and Borwein [1]. In the polyhedral case, Deutsch and Hundal [24] derived convergence rates for Dykstra's algorithm. Notably, in the case where $m = 2$, the convergence rate is independent of the initial point, depending only on an upper bound of its distance to the intersection. For further discussion on convergence rates and their applications, we refer to [23] and the references therein.

It immediately arises as a natural question if Dykstra's method is limited to metric projections or whether more general types of projection operators can be allowed. One specific way of introducing such a meaningful class of generalized projections was developed in the pivotal work of Bregman [11], where projections along a certain distance relative to a convex function are considered instead.

Concretely, over a reflexive Banach space X, 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 [4], i.e. if f is both

(1) essentially smooth, i.e. ∂f is locally bounded and single-valued on its domain,
(2) essentially strictly convex, i.e. $(\partial f)^{-1}$ is locally bounded on its domain and f is strictly convex on every convex subset of the domain of ∂f,

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 [4] for more details on this).

By substituting the metric projections by these Bregman projections in the previous method of Dykstra in Euclidean spaces, one obtains Dykstra's algorithm with Bregman projections. This method was first proposed in the seminal work by Censor and Reich [17] where the sets C_i are all halfspaces and then extended to general closed convex nonempty sets by Bauschke and Lewis [7]. Using Bregman projections is a particularly nice choice for an extension of Dykstra's method as this class of projections not only broadens the range of different permissible projections considerably, but it also has the potential to make the method more feasible by easing the computational load of the projection through a carefully chosen function f to which the projection is relativized (see e.g. the discussion in [14]). In any case, one further main advantage of studying Dykstra's method with Bregman projections is that this scheme, by means of suitable instantiations of the associated functions and sets, encompasses or is related to a considerable variety of different projection methods (beyond the ones already related to the usual version of Dykstra's algorithm as mentioned before). For example, as shown by Bregman, Censor and Reich [12], Dykstra's method with Bregman projections can be seen as a nonlinear extension of Bregman's optimization method, which was already shown in a special case in [17]. Also, Bauschke and Lewis [7] highlighted an intimate connection between Dykstra's method with Bregman projections and the seminal work of Tseng [52]. We refer to [7, 12, 17] for further discussions on these issues.

The convergence proof for Dykstra's method with Bregman projections given by Bauschke and Lewis in [7] requires 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 [4], 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 ([7]). *Let X be the Euclidean space $\mathbb{R}^d$ and let f be closed, convex, proper, Legendre, co-finite and very strictly convex. Let $C_1, \dots, C_m$ be finitely many closed and convex subsets of X with $C \cap \mathrm{intdom} f \neq \emptyset$ where $C := \bigcap_i C_i$. Set $q_{-(m-1)} = \dots = q_0 := 0$ as well as $C_n := C_{n \bmod m}$ and let P^f_n be the Bregman projection onto C_n relative to f. Given $x_0 \in \mathrm{intdom} f$, simultaneously define*

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

for $n \geq 1$. 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 [50, 51]. By results from computability theory due to Specker [49] (see also [41]), 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 but very general construction for this at the end of the paper, the extend of which we illustrate by a particular instance that allows for the derivation of computable and highly uniform rates of convergence of low-complexity in the context of Euclidean spaces where all C_i are basic semi-algebraic convex sets by utilizing a deep result of Borwein, Li and Yao [8]. This in particular covers the case where all C_i are halfspaces and so provides, to our knowledge, the first general rates of convergence for the circumstances from the work of Censor and Reich [17] where Dykstra's algorithm with Bregman projections was originally considered.

Further, all our quantitative results are valid in general normed spaces where a suitable corresponding function f exists, the assumptions on which arising as a suitable lift to infinite dimensional spaces of the assumptions presented in [7] (recall Theorem 1.1) in the finite dimensional case (as will be discussed in detail in the next section). 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 a usual (that is non-quantitative) convergence result of Dykstra's algorithm with Bregman projections for the first time in the setting of infinite dimensional spaces. The proofs of the quantitative results that we give here arise as generalizations of the recent quantitative analysis of Dykstra's method in Hilbert spaces by the first author in [43][1] (see also [42]) in combination with recent work of the second author and Kohlenbach [46] 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 [34, 36] and in particular the recent [45] 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 [7] which we discuss in this section. These preliminary results are taken, or adapted, from either [46] or [43], 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 [7], this assumption in particular entails the existence of certain moduli regarding the associated Bregman distance and the gradient:

[1] In particular, one can obtain rates of similar complexity as derived in [43] by instantiating the results given here with the function $f = \|\cdot\|^2/2$ and the respective moduli in a given (pre-)Hilbert space.

Proposition 2.1 ([7]). *Let X be the Euclidean space $\mathbb{R}^d$ and let f be very strictly convex. Then, for any convex and compact set $K \subseteq \mathrm{int dom} f$, 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) $\|\nabla f(x) - \nabla f(y)\| \leq \Theta(b) \|x - y\|$,

for any $b > 0$ and $x, y \in \overline{B}_b(0) \cap \mathrm{int dom} f$. 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 [7] 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 $(x_n), (y_n) \subseteq \mathrm{int dom} f$. In particular, by the results from [15], this implies that f is totally convex on $\mathrm{int dom} f$.

As (essentially) shown in [46], 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

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

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 $\mathrm{int dom} f$, i.e. that $C \cap \mathrm{int dom} f \neq \emptyset$. From a quantitative perspective, we will in the following fix some data relating to this condition:

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

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

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

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

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

are bounded for every $\alpha > 0$ and $x \in \mathrm{int dom} f$. 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 [15, 27]). In the context of finite-dimensional spaces, as shown in [3, 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 $x \in \mathrm{int dom} f$ and thus

$L(x, \alpha)$ is bounded. In reflexive Banach spaces, as shown in [4, 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 of this property for $x = p$.

We now shortly want to discuss a selection of functions which naturally satisfy our standing assumption (C1) – (C4):

Example 2.2. *In any Hilbert space X, taking $f(x) = \|x\|^2/2$ yields $\nabla f = \mathrm{Id}$ and clearly f is proper, convex and co-finite and Legendre and satisfies (C1). Further, f is clearly supercoercive and so all level sets $L(x, \alpha)$ are bounded as discussed before. Lastly, it is easy to see that in this case $D_f(x, y) = \|x - y\|^2$ and so f satisfies (C2). In that case, the Bregman projections are the usual metric projections and so it naturally allows one to recover the usual Dykstra's method in Hilbert spaces* [25].

Example 2.3 ([7])**.** *Let X be the usual Euclidean space $\mathbb{R}^d$ and let $f : \mathbb{R}^d \to (-\infty, +\infty]$ be separable, i.e.*

$$f(x) = \sum_{k=1}^{d} f_k(x_k)$$

for $f_k : \mathbb{R} \to (-\infty, +\infty]$, $k = 1, \ldots, d$. As highlighted in [7]*, f is a closed, convex, proper Legendre function that is co-finite and very strictly convex if, and only if, all f_k are so as well. As discussed above, any such function f naturally satisfies the assumptions (C1) – (C4). A list of examples of different function permissible for the f_k in that vein is e.g. the following, taken from* [7]*:*

(1) $g(x) := x^2/2$ on $\mathbb{R}$,
(2) $g(x) := x \ln x - x$ on $[0, +\infty)$ (with $0 \ln 0 := 0$),
(3) $g(x) := -\sqrt{1 - x^2}$ on $[-1, +1]$,
(4) $g(x) := x \ln x + (1 - x) \ln(1 - x)$ on $[0, 1]$,
(5) $g(x) := x^2/2 + 2x + 1/2$ if $x \leq -1$; $-1 - \ln(-x)$ if $-1 \leq x < 0$; $+\infty$ otherwise.

As highlighted in [7]*, the fact that all these functions are closed, convex, proper, Legendre, co-finite and very strictly convex follows by the comprehensive discussions in the seminal work* [3]*. It should be noted that the separable function arising from taking every f_k to be the function in (2), the so-called Boltzmann/Shannon entropy, naturally allows one to recover Dykstra's method for finding I-projections* [26]*, as discussed in detail in* [17].

Remark 2.4. *We are currently not aware of an example of a space together with a function that together satisfy the assumptions (C1) – (C4) and where the space is not a Hilbert space. To our knowledge, it nevertheless might be conceivable that a suitable reflexive Banach space together with a function on it satisfying the standing assumptions actually exist. Consequently, we decided to write this paper in the framework of reflexive Banach spaces which allows all proofs to be carried out and so highlights that Dykstra's method with Bregman projections is, potentially, not limited to Hilbert spaces.*

In any case, the assumptions (C1) – (C4) entail a further crucial quantitative property on the associated Bregman distance D_f:

Lemma 2.5. *The distance D_f is reverse consistent as defined in* [46]*, i.e.*

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

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

Proof. Let $x, y \in \overline{B}_r(0) \cap \mathrm{int}\,\mathrm{dom} f$. 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

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

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 2.6 (folklore, see e.g. [18])**.** *For any $x, y, z, w \in \mathrm{int}\,\mathrm{dom} f$, 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 2.7 (essentially [16])**.** *Let S be a closed convex subset of X such that $S \cap \mathrm{int}\,\mathrm{dom} f \neq \emptyset$. Consider $y \in \mathrm{int}\,\mathrm{dom} f$. Then the Bregman projection $P^f_S y$ is characterized by*

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

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 [46] (which in turn is adapted from [28, 35]). From here on out, we write $[n; m] = [n, m] \cap \mathbb{N}$ for $n, m \in \mathbb{N}$.

Proposition 2.8 (essentially [28, 35, 46])**.** *Let $r > 0$ and $u \in \mathrm{int}\,\mathrm{dom} f$ as well as $q \in C \cap \mathrm{int}\,\mathrm{dom} f$ 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 $x \in \overline{B}_r(0) \cap \mathrm{int}\,\mathrm{dom} f$ such that $\bigwedge_{j=1}^m \|x - P^f_j(x)\| \leq \delta(\eta)$ and*

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

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 $x \in \overline{B}_r(0) \cap \mathrm{int}\,\mathrm{dom} f$ such that $\|x - P^f_j(x)\| \leq \delta(\eta)$ for all $j \in [1; m]$, we have

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

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, $y_0 \in \overline{B}_r(0) \cap \mathrm{int}\,\mathrm{dom} f$ 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 [43].

Lemma 2.9 ([43])**.** *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 2.10 ([43])**.** *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 [6, Lemma 30.6].

3. Main results

We recall 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:

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

For the remaining sections, unless stated otherwise, we consider (x_n) to be the iteration generated by (DB), and we assume 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 [7], the proofs of which naturally transfer to the generalized setting here.

Lemma 3.1 (essentially [7])**.** *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 $n \in \mathbb{N}$:

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

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

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

and in particular

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

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 2.10, we derive the following lim inf-rate (akin to Proposition 3.5 in [43]).

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 2.10.

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 2.10 (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 as well as Lemma 3.1.(v) and Lemma 3.3, 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 2.9 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 2.9.

Therefore, the sequence (x_n) is asymptotically regular in the sense of [13]. 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 2.9.

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 2.5. 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 (‡), we have

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

for any $i \in [0; m-1]$. Since $\|x_{k+i}\|, \|x_k\| \le o(b)$, Lemma 2.5 for the function P gives

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

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

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

By Lemma 3.3 and (C1), we get $\|P^f_{j_{k+i}}(x_k) - x_k\| \le \varepsilon$, and the conclusion now follows from observing that for any $k \in \mathbb{N}$, $\{P^f_{j_{k+i}} \mid i \in [0; m-1]\} = \{P^f_1, \cdots, P^f_m\}$. $\square$

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 [43]).

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

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

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

$$\overline{\beta} := \beta\left(o(b), \frac{\varepsilon}{3}, \delta\right),$$
$$\delta(\eta) := \min\left\{ \frac{\varepsilon}{6o(b)\Theta_0\big(\overline{\alpha}(\eta) + \Phi_\varepsilon(\overline{\alpha}(\eta))\big)}, \widetilde{\Delta}\Big(\overline{\alpha}(\eta) + \Phi_\varepsilon\big(\overline{\alpha}(\eta)\big)\Big) \right\},$$
$$\overline{\alpha}(\eta) := \alpha(b, m, \eta, \Phi_\varepsilon),$$
$$\Phi_\varepsilon\big(N\big) := \Phi\Big(b, m, \frac{\varepsilon}{3}, N\Big),$$
$$\widetilde{\Delta}(k) := \min\{\Delta(k') \mid k' \le k\},$$

and α, β, Φ are as in Propositions 3.6, 2.8 and 3.4, respectively.

Proof. By Proposition 2.8 with $u = x_0$ and $q = p$, noting that $o(b) \ge b$ and $o(b) \ge \|p\|$ since $b > 0$, there are $\eta_0 \ge \overline{\beta}$ and $x \in \overline{B}_{o(b)}(0) \cap \mathrm{int dom} f$ such that $\|x - P^f_j(x)\| \le \delta(\eta_0)$ for all $j \in [1; m]$, and

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

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

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

Since $(x_n) \subseteq \overline{B}_{o(b)}(0) \cap \mathrm{int dom} f$, 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 $(*)$, we get

$$\begin{aligned} D_f(x, x_{n_0}) &= \langle x - x_{n_0}, \nabla f(x_0) - \nabla f(x_{n_0}) \rangle + D_f(x, x_0) - D_f(x_{n_0}, x_0) \\ &\leq \langle x - x_{n_0}, \nabla f(x_0) - \nabla f(x_{n_0}) \rangle + \frac{\varepsilon}{3}. \end{aligned}$$

Using Lemma 3.1.(ii) and $(**)$, we get

$$\begin{aligned} \langle x - x_{n_0}, \nabla f(x_0) - \nabla f(x_{n_0}) \rangle &= \sum_{k=n_0-m+1}^{n_0} \langle x - x_{n_0}, q_k \rangle \\ &= \sum_{k=n_0-m+1}^{n_0} \langle x - x_k, q_k \rangle + \sum_{k=n_0-m+1}^{n_0} \langle x_k - x_{n_0}, q_k \rangle \\ &\leq \sum_{k=n_0-m+1}^{n_0} \langle x - P_k^f(x), q_k \rangle + \sum_{k=n_0-m+1}^{n_0} \langle P_k^f(x) - x_k, q_k \rangle + \frac{\varepsilon}{3}. \end{aligned}$$

As $P_k^f(x) \in C_k$, we get $\sum_{k=n_0-m+1}^{n_0} \langle P_k^f(x) - x_k, q_k \rangle \leq 0$ by Lemma 3.1.(iii). Therefore, we in particular have

$$\langle x - x_{n_0}, \nabla f(x_0) - \nabla f(x_{n_0}) \rangle \leq \sum_{k=n_0-m+1}^{n_0} \langle x - P_k^f(x), q_k \rangle + \frac{\varepsilon}{3}.$$

Now, we can further estimate the former term by

$$\begin{aligned}\sum_{k=n_0-m+1}^{n_0}\langle x-P_k^f(x),q_k\rangle &\leq \sum_{k=n_0-m+1}^{n_0}\|x-P_k^f(x)\|\cdot\|q_k\| \\ &\leq \delta(\eta_0)\sum_{k=0}^{n_0-1}\|\nabla f(x_k)-\nabla f(x_{k+1})\| \\ &\leq \delta(\eta_0)\sum_{k=0}^{n_0-1}\Theta_0\,\|x_k-x_{k+1}\| \\ &\leq \delta(\eta_0)\cdot n_0\cdot 2o(b)\Theta_0.\end{aligned}$$

using Lemma 3.1.(v) and $\|x-P_j^f(x)\|\leq\delta(\eta_0)$ as well as Lemma 3.3. Combined, we have

$$D_f(x,x_{n_0})\leq\delta(\eta_0)\cdot n_0\cdot 2o(b)\Theta_0+\frac{2\varepsilon}{3}\leq\frac{\varepsilon}{6o(b)\Theta_0 n_0}\cdot n_0\cdot 2o(b)\Theta_0+\frac{2\varepsilon}{3}=\varepsilon$$

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* [43] *(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* [43, Proposition 2.5] *to*

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

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

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

Then [43, 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* $\mathrm{int\,dom}f\neq\emptyset$ *with gradient ∇f. Let $C_1,\cdots,C_m$ be $m\geq 2$ convex sets such that $C\cap\mathrm{int\,dom}f\neq\emptyset$ 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 $g:\mathbb{N}\to\mathbb{N}$,*

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

where $\Omega(b,m,\varepsilon,g):=\gamma(b,m,\tilde{\varepsilon},\Delta_{\varepsilon,g})$ with γ defined as in Proposition 3.7 and

$$\tilde{\varepsilon}:=\min\left\{\left(\frac{\overline{\rho}}{12o(b)\Theta_0}\right)^2\cdot\theta_0,\frac{\overline{\rho}}{6}\right\},\quad\overline{\rho}:=\frac{\varepsilon^2}{4}\cdot\theta_0$$

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

as well as $\Theta_0:=\Theta(o(b))$ and $\theta_0:=\theta(o(b))$.

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

(a) $\bigwedge_{j=1}^m \|x - P_j^f(x)\| \leq \Delta_{\varepsilon,g}(n_0)$,

(b) $D_f(x, x_{n_0}) \leq \tilde{\varepsilon} \leq \min\left\{\left(\frac{\overline{\rho}}{12o(b)\Theta_0}\right)^2 \cdot \theta_0, \frac{\overline{\rho}}{3}\right\}$,

(c) $\sum_{k=n_0-m+1}^{n_0} \langle x_k - x_{n_0}, q_k\rangle \leq \tilde{\varepsilon} \leq \frac{\overline{\rho}}{6}$.

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 $g(n_0) \geq 1$, and thus $\max\{n_0 + g(n_0), 1\} = n_0 + g(n_0)$, otherwise the result trivially holds. Since $i \geq n_0$, by Lemma 3.1.(vii) and using (b), we have

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

We naturally get

$$t_1 = \sum_{k=n_0-m+1}^{n_0} \langle x_k - x_{n_0}, q_k\rangle + \sum_{k=n_0-m+1}^{n_0} \langle x_{n_0} - x, q_k\rangle.$$

Using (c) yields

$$\sum_{k=n_0-m+1}^{n_0} \langle x_k - x_{n_0}, q_k\rangle \leq \frac{\overline{\rho}}{6}$$

and Lemma 3.1.(ii) as well as Lemma 3.3 together with (b) and (C2) imply

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

which together implies $t_1 \leq \overline{\rho}/3$. On the other hand, we naturally derive

$$t_2 = \sum_{k=i-m+1}^{i} \langle x - P_k^f(x), q_k\rangle + \sum_{k=i-m+1}^{i} \langle P_k^f(x) - x_k, q_k\rangle$$

and since $P_k^f(x) \in C_k$, we get

$$\sum_{k=i-m+1}^{i} \langle P_k^f(x) - x_k, q_k\rangle \leq 0$$

by Lemma 3.1.(iii). Now, we can estimate the former term using (a) as well as Lemma 3.1.(v) and Lemma 3.3 by

$$\begin{aligned}\sum_{k=i-m+1}^{i} \langle x - P_k^f(x), q_k\rangle &\leq \sum_{k=i-m+1}^{i} \|x - P_k^f(x)\|\|q_k\| \\ &\leq \Delta_{\varepsilon,g}(n_0) \cdot \sum_{k=i-m+1}^{i} \|q_k\| \\ &\leq \Delta_{\varepsilon,g}(n_0) \cdot \sum_{k=0}^{i-1} \|\nabla f(x_k) - \nabla f(x_{k+1})\| \\ &\leq \Delta_{\varepsilon,g}(n_0) \cdot \sum_{k=0}^{i-1} \Theta_0 \|x_k - x_{k+1}\| \\ &\leq \Delta_{\varepsilon,g}(n_0) \cdot 2o(b)\Theta_0 \cdot i.\end{aligned}$$

By definition of $\Delta_{\varepsilon,g}$, we thus have

$$\sum_{k=i-m+1}^{i} \langle x - P_k^f(x), q_k\rangle \leq \Delta_{\varepsilon,g}(n_0) \cdot 2o(b)\Theta_0 \cdot i \leq \frac{\overline{\rho}}{6o(b)\Theta_0(n_0 + g(n_0))} \cdot 2o(b)\Theta_0 \cdot i \leq \frac{\overline{\rho}}{3},$$

using in the last inequality the fact that $i \leq n_0 + g(n_0)$. This together implies $t_2 \leq \overline{\rho}/3$. Overall, we conclude that

$$D_f(x, x_i) \leq \frac{\overline{\rho}}{3} + \frac{\overline{\rho}}{3} + \frac{\overline{\rho}}{3} = \frac{\varepsilon^2}{4} \cdot \theta_0,$$

and, by Lemma 3.3 and (C1), we get $\|x_i - x\| \leq \varepsilon/2$, which entails the result by triangle inequality. □

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 [43].

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 $P_C^f(x_0)$.*

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 [47] 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. $P^f_C(x_0)$.

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

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

and let $b' \geq D_f(z,x_0)$. With $b > 0$ and p as in (C3) and θ_0, Θ_0 as in Lemma 3.3, for an arbitrary $\varepsilon > 0$ we define $\overline{\rho} := \rho(\varepsilon/2, \max\{o(b), o'(b')\})$ with $\rho(\varepsilon, b) := \theta(b)\varepsilon^2$ as in Section 2. Since $z = \lim x_n$, consider $N_0 \in \mathbb{N}$ such that

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

where $P(\varepsilon, r) := \sqrt{\varepsilon/\Theta(r)}$ as in Lemma 2.5. 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 $(x_n) \subseteq \overline{B}_{o(b)}(0) \cap \mathrm{int dom} f$, we also have $\|z\| \leq o(b)$. Moreover, by Proposition 2.7, we have

$$D_f(z, P^f_C(x_0)) \leq D_f(z,x_0) - D_f(P^f_C(x_0), x_0) \leq D_f(z,x_0) \leq b',$$

and hence by assumption on o', we have $\|P^f_C(x_0)\| \leq o'(b')$. Since $\|x_{n_0} - z\| \leq P(\overline{\rho}/4, o(b))$, by Lemma 2.5 it follows that $D_f(z, x_{n_0}) \leq \overline{\rho}/4$. As $z \in C$, by the definition of P^f_C, we have

$$D_f(P^f_C(x_0), x_0) - D_f(x_{n_0}, x_0) \leq D_f(z, x_0) - D_f(x_{n_0}, x_0),$$

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

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

which, by the properties of ρ, i.e. (C1), together with $\|x_{n_0} - z\| \leq \varepsilon/2$, entails $\|P^f_C(x_0) - z\| \leq \varepsilon$ and so, as ε is arbitrary, $z = P^f_C(x_0)$. $\square$

3.4. **Regularity and rates 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 [43] and respectively from [38].

Definition 3.11. *We call a function* $\mu : (0,\infty)^2 \to (0,\infty)$ *satisfying*

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

for all $\varepsilon, r > 0$ *a* modulus of regularity *for the sets* $C_1, \cdots, C_m$.

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 3.12, modeled after Theorem 4.2 from [43] which provided such a result in the context of Hilbert spaces for the classical version of Dykstra's algorithm. Further details on general circumstances where such moduli of regularity can actually be obtained will be discussed in the next part of this section, where we in particular show that such a modulus can be conveniently given in the case where all C_i are basic semi-algebraic convex sets (and so in particular where the sets C_i are all halfspaces).

Theorem 3.12. *Let* f *be a proper, convex and co-finite Legendre function which is Fréchet differentiable on* $\mathrm{intdom} f \neq \emptyset$ *with gradient* ∇f. *Let* $C_1, \cdots, C_m$ *be* $m \geq 2$ *convex sets such that* $C \cap \mathrm{intdom} f \neq \emptyset$ *for* $C := \bigcap_{j=1}^m C_j$. *Assume that the conditions (C1) – (C4) hold and let* μ *be a modulus of regularity for the sets* $C_1, \cdots, C_m$. *Then, the sequence* (x_n) *generated by* (DB) *satisfies*

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

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* α, Φ *as in Propositions 3.6 and 3.4, respectively, and*

$$\tilde{\varepsilon} := \frac{\varepsilon^2 \theta_0}{16 o(b) \Theta_0} \text{ and } \Phi_\varepsilon(N) := \Phi\left(b, m, \frac{\varepsilon^2 \theta_0}{8}, N\right),$$

and where $\theta_0 := \theta(o(b))$ *and* $\Theta_0 := \Theta(o(b))$.

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

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

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

$$(\circ) \qquad \forall n \in [N_0; N_0 + \Phi_\varepsilon(N_0)] \; \exists z \in C \left(\|x_n - z\| \leq \tilde{\varepsilon} = \frac{\varepsilon^2 \theta_0}{16 o(b) \Theta_0} \right).$$

Applying Proposition 3.4 (with $\varepsilon = \frac{\varepsilon^2 \theta_0}{8}$ and $N = N_0$), we have

$$\exists n_0 \in [N_0; N_0 + \Phi_\varepsilon(N_0)] \left(\sum_{k=n_0-m+1}^{n_0} \langle x_k - x_{n_0}, q_k \rangle \leq \frac{\varepsilon^2 \theta_0}{8} \right).$$

By $(\circ)$, let $\hat{z} \in C$ be such that $\|\hat{z} - x_{n_0}\| \leq \frac{\varepsilon^2\theta_0}{16o(b)\Theta_0}$. Thus, for any $i \geq n_0$:

$$\begin{aligned}
\sum_{k=i-m+1}^{i} \langle x_k - x_{n_0}, q_k\rangle &= \underbrace{\sum_{k=i-m+1}^{i} \langle x_k - \hat{z}, q_k\rangle}_{\geq 0,\ \text{by Lemma 3.1.(iii)}} + \sum_{k=i-m+1}^{i} \langle \hat{z} - x_{n_0}, q_k\rangle \\
&\geq \langle \hat{z} - x_{n_0}, \sum_{k=i-m+1}^{i} q_k\rangle \\
&= \langle \hat{z} - x_{n_0}, \nabla f(x_0) - \nabla f(x_i)\rangle && \text{(by Lemma 3.1.(ii))} \\
&\geq -\|\hat{z} - x_{n_0}\|\|\nabla f(x_0) - \nabla f(x_i)\| \\
&\geq -\|\hat{z} - x_{n_0}\|\Theta_0 \|x_0 - x_i\| && \text{(by Lemma 3.3 and (C2))} \\
&\geq -\frac{2o(b)\Theta_0\varepsilon^2\theta_0}{16o(b)\Theta_0} = -\frac{\varepsilon^2\theta_0}{8}.
\end{aligned}$$

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

$$D_f(x_{n_0}, x_i) \leq \sum_{k=n_0-m+1}^{n_0} \langle x_k - x_{n_0}, q_k\rangle - \sum_{k=i-m+1}^{i} \langle x_k - x_{n_0}, q_k\rangle \leq \frac{\varepsilon^2\theta_0}{4},$$

and by Lemma 3.3 and (C1), we get $\|x_i - x_{n_0}\| \leq \varepsilon/2$ which entails the result by triangle inequality. $\square$

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* [39] *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. In an upcoming work, an adaptation of the generalized Fejér monotonicity from* [39] *to incorporate more general distance functions than metrics along the line of the recent work* [44] *of the second author (which simultaneously extends works on Fejér monotone sequences in metric spaces as considered in* [37, 38] *and Bregman monotone methods as considered in* [5]*) will be discussed, which provides a similar Fejér-type perspective also for Dykstra's method with Bregman projections as considered here.*

3.5. **The special case of basic semi-algebraic convex sets.** In this last section, we want to discuss concrete cases where moduli of regularity for the sets $C_1, \ldots, C_m$ actually can be conveniently given. For that, we first begin with a rather general observation.

Say X is a uniformly convex Banach space where, correspondingly, the usual metric projections P_i onto the sets C_i exist for any $i \in [1; m]$. In that case, since $P_i^f(x) \in C_i$, we naturally have

$$\|x - P_i(x)\| \leq \|x - P_i^f(x)\|$$

for any $x \in \mathrm{intdom} f$ and so any function $\mu : (0, \infty)^2 \to (0, \infty)$ satisfying

$$(\star\star) \qquad \forall x \in \overline{B}_r(0) \left(\bigwedge_{j=1}^{m} \|x - P_j(x)\| \leq \mu_r(\varepsilon) \to \exists z \in C\, (\|x - z\| \leq \varepsilon) \right),$$

for all $\varepsilon, r > 0$, which is a modulus of regularity for the sets $C_1, \cdots, C_m$ in the usual sense of [43] and [38], naturally also satisfies $(\star)$ and so is a modulus of regularity for the sets $C_1, \cdots, C_m$ relative to the Bregman projections as defined here.

Conversely, as of course also $P_i(x) \in C_i$, we naturally have

$$\theta(r)\|x - P_i^f(x)\|^2 \leq D_f(P_i^f(x), x) \leq D_f(P_i(x), x) \leq \Theta(r)\,\|x - P_i(x)\|^2$$

using (the proof of) Lemma 2.5 as well as assumption (C1), assuming $x, P_i(x), P_i^f(x) \in \overline{B}_r(0) \cap \mathrm{int}\,\mathrm{dom} f$. So, given a function μ as above satisfying ($\star$), the function

$$\mu_r'(\varepsilon) := \sqrt{\frac{\Theta(r)}{\theta(r)}}\mu_r(\varepsilon)$$

satisfies ($\star\star$), at least in the context where $x, P_i(x), P_i^f(x) \in \overline{B}_r(0) \cap \mathrm{int}\,\mathrm{dom} f$.

Such conversions between moduli of regularity formulated for the ordinary metric projection and the Bregman projection are of special significance in the context of results akin to the above Theorem 3.12 since thereby, any modulus of regularity utilized to provide a rate of convergence for the usual Dykstra's algorithm in Hilbert spaces using the results contained in [43], or utilized to provide rates of convergence for an even further wider variety of projection methods as surveyed in [38], can immediately be brought to bear in the context of the variant of Dykstra's method with Bregman projections as well.

This becomes even more prevalent by noting Proposition 4.7 from [43] whereby any suitably uniform rate of convergence for Dykstra's algorithm in Hilbert spaces allows one to define a modulus of regularity in the sense of ($\star\star$). So, in conjunction with this, the above conversions allow for a sort of "automatic upgrade" by which any suitably uniform rate of convergence proved for Dykstra's method in Hilbert spaces allows one to derive a corresponding rate of convergence for Dykstra's method with Bregman projections.

This might be rather surprising to a certain degree as, naturally, the arguments used to obtain such moduli of regularity might depend heavily on geometrical arguments or the like tailored to the special situation of, say, Hilbert spaces and usual metric projections with no clear sign of how one might approach generalizations to Bregman projections.

We end this section by discussing one particular such case in more detail. For that, let X be the usual Euclidean space $\mathbb{R}^d$. Recall that $C \subseteq \mathbb{R}^d$ is a basic semi-algebraic convex set if there is a $\gamma \geq 1$ and convex polynomial functions g_i on $\mathbb{R}^d$ for $i = 1, \dots, \gamma$ such that

$$C = \{x \in \mathbb{R}^d \mid g_i(x) \leq 0 \text{ for all } i \in [1;\gamma]\}.$$

As was observed in [38], a well-known and deep result due to Borwein, Li and Yao [8] (see Theorem 3.6 therein) implies the following regularity property for the intersection of such sets:

Example 3.14 (essentially [8]). *Let $C_1, \dots, C_m \subseteq \mathbb{R}^d$ be basic semi-algebraic convex sets with $C := \bigcap_{i=1}^m C_i \neq \emptyset$ and with describing polynomials of degree $\leq \delta$. Then for any $r \in \mathbb{N}$, there is a constant $c_r > 0$ such that*

$$\mu_r(\varepsilon) := \frac{(\varepsilon/c_r)^\sigma}{m} \text{ with } \sigma := \min\left\{\frac{(2\delta-1)^d+1}{2}, B(d-1)\delta^d\right\},$$

where $B(d) = \binom{d}{\lfloor d/2 \rfloor}$ is the central binomial coefficient, is a modulus of regularity for $C_1, \dots, C_m$ in the usual sense, i.e. it satisfies ($\star\star$).

Naturally, by the above discussion, this modulus $\mu_r(\varepsilon)$ therefore also satisfies ($\star$) and so Theorem 3.12 can be used to derive corresponding rates of convergence for Dykstra's method with Bregman projections over $\mathbb{R}^d$.

To illustrate the resulting rate, we now further assume that all C_i are halfspaces (and hence focus on the original setting from [17] where Dykstra's algorithm with Bregman projections was originally introduced). In that case, all describing polynomials of all sets C_i naturally have degree ≤ 1 and so we can easily calculate

$$\sigma = \min\left\{\frac{(2\delta-1)^d+1}{2}, B(d-1)\delta^d\right\} = \min\left\{1, B(d-1)\right\} = 1$$

in that case since $B(d-1) \geq 1$. Thus, in that case, for any $r \in \mathbb{N}$, there is a constant $c_r > 0$ such that

$$\mu_r(\varepsilon) := \frac{\varepsilon}{mc_r}$$

is a modulus of regularity and using Theorem 3.12, we in particular obtain the following rate of convergence for Dykstra's method with Bregman projections as originally defined by Censor and Reich over halfspaces in $\mathbb{R}^d$:

Theorem 3.15. *Let $X = \mathbb{R}^d$ and let f be a proper, convex and co-finite Legendre function which is Fréchet differentiable on* $\mathrm{intdom} f \neq \emptyset$ *with gradient ∇f. Let $C_1, \cdots, C_m \subseteq \mathbb{R}^d$ be $m \geq 2$ halfspaces such that $C \cap \mathrm{intdom} f \neq \emptyset$ for $C := \bigcap_{j=1}^m C_j$. Assume that the conditions (C1) – (C4) hold. Then, the sequence (x_n) generated by* (DB) *satisfies*

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

where

$$\Theta(b, m, \varepsilon) := \alpha\left(b, m, \frac{\tilde{\varepsilon}}{mc_{o(b)}}, \Phi_\varepsilon\right) + \Phi_\varepsilon\left(\alpha\left(b, m, \frac{\tilde{\varepsilon}}{mc_{o(b)}}, \Phi_\varepsilon\right)\right)$$

with α, Φ as in Propositions 3.6 and 3.4, respectively, and

$$\tilde{\varepsilon} := \frac{\varepsilon^2\theta_0}{16o(b)\Theta_0} \text{ and } \Phi_\varepsilon(N) := \Phi\left(b, m, \frac{\varepsilon^2\theta_0}{8}, N\right),$$

and where $\theta_0 := \theta(o(b))$ and $\Theta_0 := \Theta(o(b))$.

4. Conclusions

This work provided a study on the asymptotic behaviour of Dykstra's algorithm with Bregman projections. Specifically, we derived very general computational information in the form of a computable rate of metastability, which is highly uniform on the parameters of the problem. Indeed, the rate obtained depends only on the number of convex sets, the distance from the initial point to the solution set, and other minor quantitative data related to the underlying conditions of the result, but is otherwise independent from the specifics of the space and the sets.

Moreover, even disregarding the quantitative information obtained, our approach is valid in general reflexive Banach spaces, provided that a suitable function exists which satisfies the specific conditions (C1) – (C4) outlined in Section 2. In particular, this for the first time shows the convergence of Dykstra's algorithm with Bregman projections in infinite dimensional spaces (while it however remains to be seen in future work whether there is actually an instance of a genuine reflexive Banach space, which is not a Hilbert space, together with a function on it satisfying these conditions).

We further showed that the algorithm indeed allows for a uniform rate of convergence provided that a suitable, rather general, regularity assumption holds. In this regard, we also discussed an abstract transfer result between this regularity condition, formulated using Bregman projections, and its counterpart involving the usual metric projections which in particular enabled us to establish a uniform rate of convergence for Dykstra's algorithm with Bregman projections in the polyhedral case.

Conflicts of interests: 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] H.H. Bauschke and J.M. Borwein. Dykstra's alternating projection algorithm for two sets. *Journal of Approximation Theory*, 79:418–443, 1994.
[2] H.H. Bauschke and J.M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Review*, 38(3):367–426, 1996.
[3] 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.
[4] 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.
[5] 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.
[6] H.H. Bauschke and P.L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces.* CMS Books in Mathematics. Springer, 2nd edition, 2017.
[7] H.H. Bauschke and A.S. Lewis. Dykstras algorithm with Bregman projections: A convergence proof. *Optimization*, 48(4):409–427, 2000.
[8] J.M. Borwein, G. Li, and L. Yao. Analysis of the convergence rate for the cyclic projection algorithm applied to basic semialgebraic convex sets. *SIAM Journal on Optimization*, 24(1):498–527, 2014.
[9] 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.
[10] L.M. Bregman. Finding the common point of convex sets by the method of successive projection. *Doklady Akademii Nauk SSSR*, 162:487–490, 1965.
[11] 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.
[12] L.M. Bregman, Y. Censor, and S. Reich. Dykstra's algorithm as the nonlinear extension of Bregman's optimization method. *Journal of Convex Analysis*, 6:319–333, 1999.
[13] 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.
[14] D. Butnariu and A.N. Iusem. On a proximal point method for convex optimization in Banach spaces. *Numerical Functional Analysis and Optimization*, 18:723–744, 1997.
[15] 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.
[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*, 2006. Art. ID 84919, 39pp.
[17] Y. Censor and S. Reich. The Dykstra algorithm with Bregman projections. *Communications in Applied Analysis*, 2(3):407–419, 1998.
[18] 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.
[19] P.L. Combettes. The convex feasibility problem in image recovery. *Advances in imaging and electron physics*, 95:155–270, 1996.
[20] P.L. Combettes. Hilbertian convex feasibility problem: Convergence of projection methods. *Applied Mathematics and Optimization*, 35:311–330, 1997.
[21] D.A. D'Esopo. A convex programming procedure. *Naval Research Logistics Quarterly*, 6:33–42, 1959.

[22] F. Deutsch. The method of alternating orthogonal projections. In S.P. Singh, editor, *Approximation theory, spline functions and applications*, volume 356 of *NATO Advanced Science Institutes Series C: Mathematical and Physical Sciences*, pages 105–121. Kluwer Academic Publishers, Dordrecht, 1992.

[23] F. Deutsch. Dykstra's cyclic projections algorithm: the rate of convergence. In S.P. Singh, editor, *Approximation theory, wavelets and applications*, volume 454 of *NATO Advanced Science Institutes Series C: Mathematical and Physical Sciences*, pages 87–94. Kluwer Academic Publishers, Dordrecht, 1995.

[24] F. Deutsch and H. Hundal. The rate of convergence of Dykstra's cyclic projections algorithm: the polyhedral case. *Numerical Functional Analysis and Optimization*, 15(5-6):537–565, 1994.

[25] R.L. Dykstra. An algorithm for restricted least squares regression. *Journal of the American Statistical Association*, 78(384):837–842, 1983.

[26] R.L. Dykstra. An iterative procedure for obtaining I-projections onto the intersection of convex sets. *The Annals of Probability*, 13:975–984, 1985.

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

[28] F. Ferreira, L. Leuştean, and P. Pinto. On the removal of weak compactness arguments in proof mining. *Advances in Mathematics*, 354:106728, 55pp., 2019.

[29] I. Halperin. The product of projection operators. *Acta Universitatis Szegediensis. Acta Scientiarum Mathematicarum*, 23:96–99, 1962.

[30] S.-P. Han. A successive projection method. *Mathematical Programming*, 40:1–14, 1988.

[31] C. Hildreth. A quadratic programming procedure. *Naval Research Logistics Quarterly*, 4:79–85, 1957.

[32] H. Hundal. An alternating projection that does not converge in norm. *Nonlinear Analysis. Theory, Methods & Applications*, 57(1):35–61, 2004.

[33] A.N. Iusem and A.R. De Pierro. On the convergence of Han's method for convex programming with quadratic objective. *Mathematical Programming*, 52:265–284, 1991.

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

[35] U. Kohlenbach. On quantitative versions of theorems due to F.E. Browder and R. Wittmann. *Advances in Mathematics*, 226:2764–2795, 2011.

[36] 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.

[37] U. Kohlenbach, L. Leuştean, and A. Nicolae. Quantitative results on Fejér monotone sequences. *Communications in Contemporary Mathematics*, 20(2), 2018. 1750015, 42pp.

[38] 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.

[39] U. Kohlenbach and P. Pinto. Fejér monotone sequences revisited. *Journal of Convex Analysis*, 2025. to appear.

[40] A. Lent and Y. Censor. Extensions of Hildreth's row-action method for quadratic programming. *SIAM Journal on Control and Optimization*, 18:444–454, 1980.

[41] E. Neumann. Computational problems in metric fixed point theory and their Weihrauch degrees. *Logical Methods in Computer Science*, 11(4):20, 44pp., 2015.

[42] P. Pinto. Proof mining and the convex feasibility problem: the curious case of Dykstra's algorithm. 2024. 31pp., Oberwolfach Preprints, doi:10.14760/OWP-2024-06.

[43] P. Pinto. On the finitary content of Dykstra's cyclic projections algorithm. *Zeitschrift für Analysis und ihre Anwendungen*, 44(1/2):165–192, 2025.

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

[45] 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.

[46] N. Pischke and U. Kohlenbach. Effective rates for iterations involving Bregman strongly nonexpansive operators. *Set-valued and Variational Analysis*, 32(4), 2024. 33, 58pp.

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

[48] E. Matoušková and S. Reich. The Hundal example revisited. *Journal of Nonlinear and Convex Analysis*, 4(3):411–427, 2003.

[49] E. Specker. Nicht konstruktiv beweisbare Sätze der Analysis. *The Journal of Symbolic Logic*, 14(3):145–158, 1949.

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

[51] 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.

[52] P. Tseng. Dual coordinate ascent methods for non-strictly convex minimization. *Mathematical Programming*, 59:231–247, 1993.

[53] J. von Neumann. *Functional Operators. II. The Geometry of Orthogonal Spaces*, volume 22 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1950.